

ICAP 09

Aug 31 - Sep 4, 2009
San Francisco

[general](#)[author info](#)[registration](#)[program](#)[contact](#)[proceedings](#)[exhibitor program](#)[welcome](#)[hotel info](#)[getting there](#)[committees](#)[exhibitors](#)[confirmed
attendees](#)[photos](#)

10th International Computational Accelerator Physics Conference

ICAP'09 currently has 168 confirmed attendees, the most of any in the conference series' 21 year history.

For a list of confirmed speakers, see the [Program page](#)

The 2009 International Computational Accelerator Conference will be held Aug 31 - Sept 4, 2009 in the San Francisco Bay Area.

The conference venue will be the [Mark Hopkins Intercontinental Hotel](#) located at 1 Nob Hill in the heart of San Francisco.

ICAP focuses on the latest advances in computational accelerator physics. ICAP 2009 will be the 10th in the conference series following meetings in La Jolla, California (1988), Los Alamos, New Mexico (1990), Pleasanton, California (1993), Williamsburg, Virginia (1996), Monterey, California (1998), Darmstadt, Germany (2000), East Lansing, Michigan (2002), St. Petersburg, Russia (2004), and Chamonix, France (2006).

ICAP 2009 is being organized by the Lawrence Berkeley National Laboratory and the Stanford Linear Accelerator Center.

Important Dates:

Abstract Submission Available:	Friday Feb 13
Registration Opens:	Friday May 8
Abstract Submission Deadline:	Monday June 1 (extended to Friday June 26 for posters only)
Notification of Abstract Acceptance:	Monday June 8
Early Registration Deadline:	Monday June 29
Hotel Room Block Reservation Deadline:	August 2
Conference begins:	8:00 AM, Monday August 31
Conference Ends:	Noon, Friday Sept 4
Publication Submission Deadline:	Friday Sept 4

PREFACE

The 2009 International Computational Accelerator Conference was held Aug 31 - Sept 4, 2009 in San Francisco, California. Hosted by Lawrence Berkeley National Laboratory and the SLAC National Accelerator Laboratory, ICAP'09 was the 10th in the conference series, following meetings in La Jolla, California (1988), Los Alamos, New Mexico (1990), Pleasanton, California (1993), Williamsburg, Virginia (1996), Monterey, California (1998), Darmstadt, Germany (2000), East Lansing, Michigan (2002), St. Petersburg, Russia (2004), and Chamonix, France (2006).

The combination of an exciting technical program and a world-class venue (the Mark Hopkins InterContinental Hotel in the heart of San Francisco) resulted in a conference that attracted 168 registered attendees, the most in the conference series history, from North America, Europe, and Asia. Accepted for the program were 115 talks and 32 posters on a variety of topics, including computational beam dynamics, electromagnetics, advanced accelerator modeling, numerical algorithms, and visualization. A special session was organized by researchers from the National Energy Research Scientific Computing Center (NERSC) on Trends in High Performance Computing.

This conference would not have been possible without the dedicated efforts of many people. The Local Organizing Committee consisted of Martha Condon, Conference Administrator; Laurie O'Brien, Financial Officer; Sarah Poon, Conference Webmaster; Joe Chew, Conference Editor; and myself, Conference Chair. The International Organizing Committee consisted of Martin Berz (MSU), Oliver Boine-Frankenheim (GSI), Kwok Ko (SLAC), Kyoko Makino (MSU), Cho Ng (SLAC), Kazuhito Ohmi (KEK), Dmitri Ovsyannikov (St. Petersburg State University), Ursula van Rienen (U. Rostock), Stephan Russenschuck (CERN), Frank Schmidt (CERN), Tsumoru Shintake (RIKEN), Chuan-Xiang Tang (Tsinghua University), and Thomas Weiland (TEMF/Darmstadt).

I want to thank Christine Petit-Jean-Genaz for providing Joint Accelerator Conferences Working Group (JACoW) support for ICAP'09, and Matt Arena for assistance with the JACoW database.

I especially want to thank Joe Chew, Conference Editor, whose efforts before, during, and after the conference have made these proceedings a reality.

Lastly, thanks to the participants of ICAP'09. Their involvement and enthusiasm are ultimately responsible for the success of ICAP.

Robert D. Ryne
ICAP'09 Conference Chair

Contents

Preface	i
Foreword	iii
Contents	v
Committees	vii
Pictures	viii
MO3IOPK03 – Calculation of Realistic Charged-Particle Transfer Maps	1
MO3IOPK04 – Construction of Large-Period Symplectic Maps by Interpolative Methods	6
MO4IOPK02 – Highly Scalable Numerical Methods for Simulation of Space Charge Dominated Beams	12
MO4IOPK04 – Overview of (Some) Computational Approaches in Spin Studies	18
MO4IOPK05 – An Efficient 3D Space Charge Routine with Self-Adaptive Discretization	23
MO3IODN01 – Impedance Estimation by Parabolic Partial Differential Equation for Rectangular Taper	27
MO4IODN02 – Applying an hp-Adaptive Discontinuous Galerkin Scheme to Beam Dynamics Simulations	30
MO4IODN03 – Portable High Performance Computing for Microwave Simulation by FDTD/FIT Machines	35
MO4IODN05 – High-Order Differential Algebra Methods for PDEs Including Rigorous Error Verification	38
TU1IOPK01 – Computational Beam Dynamics for a High Intensity Ring: Benchmarking with Experiment in the SNS	42
TU1IOPK02 – Comparison of Different Simulation Codes with UNILAC Measurements for High Beam Currents	48
TU1IOPK04 – Benchmarking Different Codes for the High Frequency RF Calculation	53
TU2IOPK02 – Simulation Studies & Code Validation For The Head-Tail Instability With Space Charge	58
TU3IOPK03 – Progress with Understanding and Control of Nonlinear Beam Dynamics At The Diamond Storage Ring	64
TU3IOPK04 – Design and Control of Ultra Low Emittance Light Sources	68
TU4IOPK02 – Novel Methods for Simulating Relativistic Systems Using an Optimal Boosted Frame	73
TU3IODN03 – Modeling Techniques for Design and Analysis of Superconducting Accelerator Magnets	77
TU3IODN05 – Transient, Large-Scale 3D Finite Element Simulations of the SIS100 Magnet	83
TU4IODN01 – A Parallel Hybrid Linear Solver for Accelerator Cavity Design	89
WE2IOPK01 – Hard- and Software-based Acceleration Techniques for Field Computation	93
WE2IOPK03 – Graphical Processing Unit-Based Particle-In-Cell Simulations	96
WE2IOPK05 – VizSchema - A Standard Approach for Visualization of Computational Accelerator Physics Data	101
WE3IOPK01 – The Object Oriented Parallel Accelerator Library (OPAL), Design, Implementation and Application	107
WE3IOPK02 – Recent Progress and Plans for the Code ELEGANT	111
WE3IOPK04 – Update on MAD-X and Future Plans	117
WE4IOPK02 – High-Fidelity Injector Modeling with Parallel Finite Element 3D Electromagnetic PIC Code Pic3P	122
WE4IOPK04 – Beam Dynamics In The Low Energy Part Of The Low Emittance Gun (LEG)	125
WE3IODN01 – The XAL Infrastructure for High Level Control Room Applications	131
WE3IODN03 – Improvement Plans for the RHIC/AGS On-Line Model Environments	137
WE4IODN01 – Beam-Beam Simulations for KEKB and Super-B Factories	141
WE4IODN03 – Recent Advances of Beam-Beam Simulation in BEPCII	147
TH1IOPK02 – Modeling of Ultra-cold and Crystalline Ion Beams	151
TH1IOPK04 – Developing the Physics Design for NDCX-II, a Unique Pulse-Compressing Ion Accelerator	157
TH2IOPK01 – Self Field of Sheet Bunch: A Search for Improved Methods	163
TH2IOPK02 – Simulation of Microwave Instability in LER of KEKB And SuperKEKB	169
TH2IOPK04 – Study of Beam-Scattering Effects for a Proposed APS ERL Upgrade	173
TH3IOPK01 – The Simulation of the Electron Cloud Instability in BEPCII and CSNS/RCS	179
TH3IOPK03 – Modeling Laser Stripping with the Python ORBIT Code	184
TH3IOPK04 – Using Geant4-based Tools to Simulate a Proton Extraction and Transfer Line	190
TH4IOPK02 – End To End Simulations of the GSI Linear Accelerator Facility	196
TH4IOPK03 – Aperture and Beam-Tube Models for Accelerator Magnets	202
TH1IODN01 – A Fast and Universal Vlasov Solver for Beam Dynamics Simulations in 3D	208
TH1IODN04 – Discretizing Transient Current Densities in the Maxwell Equations	212
TH2IODN01 – Simulation and Commissioning of J-PARC Linac Using the IMPACT Code	218

TH2IODN04 – Physics Problem Study For A 100 MeV, 500 Microamp H^- Beam Compact Cyclotron . . .	224
TH3IODN02 – Space Charge Simulations for ISIS	229
TH4IODN02 – An Integrated Beam Optics-Nuclear Processes Framework in COSY Infinity and Its Applica- tions to FRIB	235
TH4IODN04 – The Study on the Space Charge Effects of RCS/CSNS	239
FR1IOPK01 – Optimization Algorithms for Accelerator Physics Problems	245
FR1IOPK02 – Application of Multiobjective Genetic Algorithm in Accelerator Physics	251
FR1IOPK09 – Application of Direct Methods of Optimizing Storage Ring Dynamic and Momentum Apertures	255
THPSC003 – RadTrack: A User-Friendly, Modular Code to Calculate the Emission Processes from High- Brightness Electron Beams	259
THPSC004 – Tomographic Reconstruction of a Beam Phase Space from Limited Projection Data	262
THPSC006 – Particle-In-Cell Simulation of Electron-Helium Plasma in Cyclotron Gas Stopper	266
THPSC010 – Including Partial Siberian Snakes Into the AGS Online Model	270
THPSC011 – A Fast Point to Point Interaction Model for Charged Particle Bunches By Means of Nonequi- spaced Fast Fourier Transform (NFFT)	273
THPSC012 – TRIUMF-VECC Electron Linac Beam Dynamics Optimization	277
THPSC013 – Design of 10 GeV Laser Wakefield Accelerator Stages with Shaped Laser Modes	281
THPSC017 – Multipole Effects in the RF Gun for the PSI Injector	285
THPSC018 – An Application of Differential Algebraic Methods and Liouville's Theorem: Uniformization of Gaussian Beams	289
THPSC019 – COSY Extensions for Beam-Material Interactions	292
THPSC020 – Optimizing SRF Gun Cavity Profiles in a Genetic Algorithm Framework	296
THPSC021 – Computational Models for μ Channel Plate Simulations	300
THPSC022 – Recent Improvement of Tracking Code BBSIMC	304
THPSC023 – A New Model-Independent Method for Optimization of Machine Settings and Electron Beam Parameters	308
THPSC026 – RF-Kick Caused by the Couplers in the ILC Acceleration Structure	311
THPSC028 – Computation of a Two Variable Wake Field Induced by an Electron Cloud	314
THPSC030 – A High-Level Interface for the ANKA Control System	318
THPSC031 – PteqHI Development and Code Comparing	322
THPSC035 – Tracy#	326
THPSC036 – Modeling Single Particle Dynamics in Low Energy and Small Radius Accelerators	330
THPSC037 – Possibility of Round Beam Formation in RIBF Cyclotrons	333
THPSC041 – Set Code Development and Space Charge Studies on ISIS	337
THPSC047 – Complete RF Design of the HINS RFQ with CST MWS and HFSS	340
THPSC049 – H5PartRoot - A Visualization And Post-Processing Tool For Accelerator Simulations	343
THPSC050 – Parallel SDDS: A Scientific High-Performance I/O Interface	347
THPSC052 – The Python Shell for the ORBIT Code	351
THPSC054 – Recent Progress on Parallel ELEGANT	355
THPSC056 – Beam Fields in an Integrated Cavity, Coupler, and Window Configuration	359
THPSC057 – BPM Breakdown Potential in the PEP-II B-factory Storage Ring Collider	363
THPSC058 – Recycler Lattice for Project X at Fermilab	367
THPSC059 – Array Based Truncated Power Series Package	371
THPSC061 – Molecular Dynamics Simulation of Crystalline Beams Extracted from a Storage Ring	374
Appendices	379
List of Authors	379
Institutes List	383
Participants List	387

COMMITTEES

Conference Chairman

Robert Ryne (LBNL)

Local Organizing Committee

Sarah Poon (LBNL), webmaster

Joe Chew (LBNL), editor

Martha Condon (LBNL), conference administrator

Thanh Ly (SLAC), SLAC visit coordinator

Laurie O'Brien (LBNL), finance officer

International Organizing Committee

Martin Berz (MSU)

Oliver Boine-Frankenheim (GSI)

Kwok Ko (SLAC)

Kyoko Makino (MSU)

Cho Ng (SLAC)

Kazuhito Ohmi (KEK)

Dmitri Ovsyannikov (St. Petersburg State University)

Ursula van Rienen (U. Rostock)

Stephan Russenschuck, CERN

Robert Ryne (LBNL), Chair

Frank Schmidt (CERN)

Tsumoru Shintake (RIKEN)

Chuan-Xiang Tang (Tsinghua University)

Thomas Weiland (TEMF/Darmstadt)

CALCULATION OF REALISTIC CHARGED-PARTICLE TRANSFER MAPS*

C E. Mitchell and A. J. Dragt, University of Maryland, USA

Abstract

The stability of orbits in storage and damping rings can depend sensitively on nonlinear fringe-field and high-order-multipole effects in the various beam-line elements. The inclusion of these effects requires a detailed and realistic model of the interior and fringe electric and magnetic fields, including their high spatial derivatives. In the case of magnetic elements a collection of surface fitting methods has been developed for extracting this information accurately from 3-dimensional magnetic field data on a grid, as provided by various 3-dimensional finite element field codes. The virtue of surface methods is that they exactly satisfy the Maxwell equations and are relatively insensitive to numerical noise in the data. These techniques can be used both to compute realistic design orbits and realistic high-order transfer maps about these orbits. An exactly soluble but numerically challenging model field is used to provide a rigorous collection of performance benchmarks.

BACKGROUND

For the design of high-performance storage or damping rings it is essential to have realistic electric and magnetic field information for the various beam-line elements, in order to compute accurate design orbits and high-order transfer maps about the design orbits. Realistic field data can be provided on a grid with the aid of various 3-dimensional finite element codes, sometimes spot checked against measured data. But the computation of high-order transfer maps based on this data appears to pose an insurmountable problem: the calculation of high-order transfer maps requires a knowledge of high derivatives of the field data. The direct calculation of high derivatives based only on grid data is intolerably sensitive to noise (due to truncation or round-off) in the grid data. We will see that this problem can be overcome by the use of surface methods. The effect of numerical noise can be overcome by fitting onto a bounding surface far from the beam axis and continuing inward using the Maxwell equations. While the process of differentiation serves to amplify the effect of numerical noise, the process of continuing inward using the Maxwell equations is *smoothing*. This smoothing is related to the fact that harmonic functions take their extrema on boundaries. When using surface methods, all fits are made to such boundaries. Therefore if these fits are accurate, interior data based on these fits will be even more accurate.

In this paper we will devote our attention to magnetic beam-line elements. (For a treatment of RF cav-

ities, see [1].) Two cases have been treated separately: straight and curved. For straight beam-line elements such as quadrupoles, sextupoles, octupoles, and wigglers, it is convenient to employ cylindrical surfaces. These surfaces may have circular, elliptical, or rectangular cross sections. We will describe the use of elliptical cylinders. The use of circular and rectangular cylinders is described elsewhere [2, 3, 4]. For the case of curved magnetic elements such as dipoles with large design-orbit sagitta, we will employ the surface of a bent box with straight ends. In all cases the bounding surface will surround the design orbit within the beam-line element and will extend into the fringe-field regions outside the beam-line element, thus taking into account all fringe-field effects as well as all effects within the body of the beam-line element.

For the case of straight beam-line elements it is convenient to describe the magnetic field in terms of a magnetic scalar potential ψ . Then, if one wishes to compute transfer maps in terms of canonical coordinates, one can proceed with the aid of an associated vector potential \mathbf{A} computed from ψ . Alternatively, if one wishes to integrate noncanonical equations employing the magnetic field \mathbf{B} , it can be obtained from the relation $\mathbf{B} = \nabla\psi$.

For the case of curved beam-line elements it is convenient to work directly with the vector potential. Its use in the case of canonical coordinates is then immediate. If instead one wishes to integrate noncanonical equations employing the magnetic field \mathbf{B} , it can be obtained from the relation $\mathbf{B} = \nabla \times \mathbf{A}$.

In this paper we will first treat the case of straight beam-line elements. For this case a cylindrical multipole expansion for ψ is convenient. In Sections II-III we will describe such an expansion and how it can be computed based on \mathbf{B} data provided on a grid and interpolated onto the surface of an elliptical cylinder. In Section IV we will treat the case of curved beam-line elements. In this case \mathbf{A} will be computed based on both \mathbf{B} and ψ data provided on a grid and interpolated onto the surface of a bent box.

CYLINDRICAL HARMONIC EXPANSIONS

In a current-free region the magnetic field \mathbf{B} is curl free, and can therefore be described in terms of a magnetic scalar potential. Because \mathbf{B} is also divergence free, ψ must obey the Laplace equation $\nabla^2\psi = 0$. A general solution ψ satisfying the Laplace equation in cylindrical coordinates

* Work supported by U.S. Department of Energy Grant DE-FG02-96ER40949.

and analytic near the axis $\rho = 0$ takes the form

$$\psi(\rho, \phi, z) = \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk I_m(k\rho) e^{ikz} [G_{m,s}(k) \sin m\phi + G_{m,c}(k) \cos m\phi]. \quad (1)$$

By utilizing the Taylor series of the modified Bessel function I_m we may write ψ in the form of a cylindrical harmonic (multipole) expansion:

$$\psi(\rho, \phi, z) = \sum_{m=0}^{\infty} [\psi_{m,s}(\rho, z) \sin m\phi + \psi_{m,c}(\rho, z) \cos m\phi], \quad (2)$$

where for $\alpha = s, c$,

$$\psi_{m,\alpha}(\rho, z) = \sum_{l=0}^{\infty} \frac{(-1)^l m!}{2^{2l} l! (l+m)!} C_{m,\alpha}^{[2l]}(z) \rho^{2l+m}. \quad (3)$$

The functions $C_{m,\alpha}^{[n]}$, known as generalized gradients, are defined by

$$C_{m,\alpha}^{[n]}(z) = \frac{i^n}{2^m m!} \int_{-\infty}^{\infty} dk e^{ikz} k^{m+n} G_{m,s}(k). \quad (4)$$

Once they are known, ψ is specified. Note that $C_{m,\alpha}^{[n]}(z) = d^n C_{m,\alpha}^{[0]}(z)/dz^n$.

FITTING USING AN ELLIPTICAL CYLINDER

We will now describe how the generalized gradients can be computed based on \mathbf{B} data provided on a grid and interpolated onto the surface of an elliptical cylinder. Elliptic coordinates in the x, y plane are described by the relations

$$x = f \cosh(u) \cos(v), \quad y = f \sinh(u) \sin(v). \quad (5)$$

Contours of constant u , with $u \in [0, \infty]$, are nested ellipses with common foci located at $(x; y) = (\pm f; 0)$. Contours of constant v , with $v \in [0, 2\pi]$, are hyperbolae. Together these contours form an orthogonal coordinate system.

Suppose we are provided with the three components of the magnetic field on a regular 3-d Cartesian grid. Consider an elliptical cylinder surrounding the axis of the magnetic element, which lies within all iron and other magnetic sources. Such a surface is obtained by setting $u = U$ and allowing the coordinates v, z to vary. This data can then be interpolated onto the surface of the elliptical cylinder to obtain the normal component B_u of the field on this surface.

Define the functions F_ℓ and G_ℓ in terms of the surface data as

$$F_\ell(U, k) = \int_{-\pi}^{\pi} \sqrt{J(U, v)} \tilde{B}_u(U, v, k) se_\ell(v, q) dv, \quad (6a)$$

$$G_\ell(U, k) = \int_{-\pi}^{\pi} \sqrt{J(U, v)} \tilde{B}_u(U, v, k) ce_\ell(v, q) dv. \quad (6b)$$

Here $\tilde{B}_u(U, v, k)$ is the Fourier transform of $B_u(U, v, z)$, se_ℓ and ce_ℓ are Mathieu functions [5, 6], $J(u, v)$ is the Jacobian of the mapping from Cartesian to elliptic coordinates, and $q = -k^2 f^2/4$. The on-axis gradients are now given by

$$C_{m,\alpha}^{[n]}(z) = \frac{i^n}{2^m m!} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k^{m+n} e^{ikz} \beta_m^\alpha(U, k) dk \quad (7)$$

where

$$\beta_m^s(U, k) = \sum_{\ell=0}^{\infty} g_s^\ell(k) E_m^{(\ell)}(k) \left[\frac{F_\ell(U, k)}{S e_\ell'(U, q)} \right], \quad (8a)$$

$$\beta_m^c(U, k) = \sum_{\ell=0}^{\infty} g_c^\ell(k) D_m^{(\ell)}(k) \left[\frac{G_\ell(U, k)}{C e_\ell'(U, q)} \right]. \quad (8b)$$

Here $S e_\ell$ and $C e_\ell$ are modified Mathieu functions, and $g_s^\ell, D_m^{(\ell)}, E_m^{(\ell)}$ are known functions that relate Mathieu and Bessel functions [3, 4].

Benchmarks

Here we describe an exactly-soluble but numerically challenging model field to be used to numerically benchmark the procedure just described. Suppose two magnetic monopoles having strengths $\pm g$ are placed at the (x, y, z) locations

$$r^+ = (0, a, 0), \quad r^- = (0, -a, 0). \quad (9)$$

These monopoles generate a scalar potential $\psi(x, y, z)$ given by the relation

$$\psi(x, y, z) = \psi_+(x, y, z) + \psi_-(x, y, z) = -g[x^2 + (y - a)^2 + z^2]^{-1/2} + g[x^2 + (y + a)^2 + z^2]^{-1/2}. \quad (10)$$

Due to the symmetries of the field, it can be shown that the only nonvanishing associated generalized gradients $C_{m,\alpha}^{[n]}$ are those with $\alpha = s$ and m odd. They have the values

$$C_{m,s}^{[0]}(z) = (-1)^{(m-1)/2} \frac{g}{a^{m+1}} \frac{(2m)!}{2^{2m-2} (m!)^2} \beta^{2m+1}(z) \quad (11a)$$

where

$$\beta(z) = \frac{a}{\sqrt{z^2 + a^2}}. \quad (11b)$$

This result has been used to benchmark the technique described in the previous section in the case that $a = 2.4$ cm and $g = 1$ Tesla-(cm)². We set up a regular grid in x, y, z space, where we let each variable range over the intervals $x \in [-4.4, 4.4]$ with spacing $h_x = 0.1$, $y \in [-2.4, 2.4]$ with $h_y = 0.1$, and $z \in [-300, 300]$ with $h_z = 0.125$ (in units of cm). The values of the magnetic field at each grid point are computed using $\mathbf{B} = \nabla\psi$.

Consider an elliptical cylinder of semimajor axis $x_{max} = 4$ cm, semiminor axis $y_{max} = 2$ cm, and length 600 cm. We use bicubic interpolation to interpolate \mathbf{B} at

these grid points onto 120 selected angular points on the cylinder, for each of the 4801 selected values of z . The angular integration in (6) is performed using a Riemann sum with $N = 120$. (This is necessary to ensure sufficient convergence of the angular integrals to within 10^{-4} .) We evaluate the Fourier transform at 401 values of k in the range $[-K_c, K_c]$ with $K_c = 20$, using a spline-based Fourier transform algorithm. We use these same points in k space to evaluate the inverse Fourier transform, providing a set of numerically determined functions $C_{m,\alpha}^{[n]}(z)$.

Suppose, for example, we wish to obtain the transfer map for the monopole doublet through terms of degree 7. We then require the generalized gradient functions $C_{m,s}^{[n]}(z)$ with $(m+n) \leq 7$. For each of these functions, we find that the relative difference between the numerical results and (11) is on the order of 10^{-4} or smaller.

Smoothing

The key feature of this technique is that results are relatively insensitive to surface errors due to smoothing. That is, each kernel multiplying the surface functions F_ℓ and G_ℓ in (8) falls off rapidly with spatial frequency k . As a result, high frequency noise appearing in the boundary data has little effect on the functions $C_{m,\alpha}^{[n]}$ of (7).

To illustrate this effect, let $B_y(0, 0, z)$ denote the on-axis monopole doublet field and make the noise model

$$B_x^{noise}(x_j, y_j, z_j) = \epsilon B_y(0, 0, z_j) \delta_x(j), \quad (12)$$

$$B_y^{noise}(x_j, y_j, z_j) = \epsilon B_y(0, 0, z_j) \delta_y(j), \quad (13)$$

where $\delta_x(j)$ and $\delta_y(j)$ are uniformly distributed random variables taking values in the interval $[-1, 1]$. We take $\epsilon = 0.01$, obtaining a random distribution of field values that are proportional at the 1% level to the on-axis vertical field of the monopole doublet. Figs. 1-2 illustrate the generalized gradients $C_{7,c}^{[0]}$ and $C_{1,c}^{[6]}$ as computed from these values using (7). The solid line illustrates the rms value of the generalized gradient at each value of z , as computed using a circular cylinder of radius $R = 2$ cm. The dashed line illustrates the rms value of the generalized gradient at each value of z as computed using a circumscribed elliptical cylinder of semiminor axis 2 cm and semimajor axis 4 cm. In each case we find that a 1% noise in the surface data produces only a 0.01% error or less in the generalized gradients. These figures also illustrate that fitting to the surface of an elliptical cylinder leads to greater suppression of noise than fitting to an inscribed circular cylinder. This is consistent with what we expect if errors in the surface data are suppressed with increasing distance from the bounding surface. We see that it is advantageous to use a fitting surface which is as far from the axis as possible.

Applications

A less stringent test of the accuracy of this procedure (but also a test of the quality of the magnetic data on the Beam Dynamics, Other

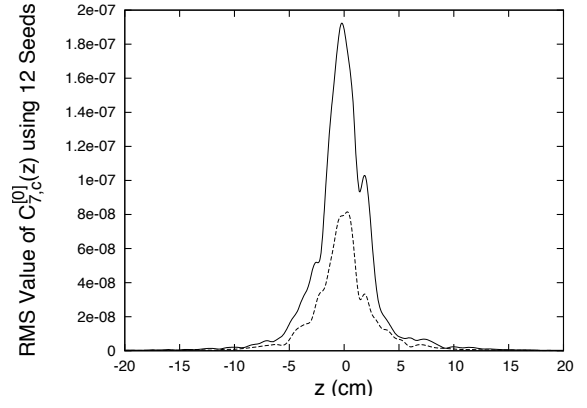


Figure 1: The quantity $C_{7,c}^{[0]}(z)$ computed from a field consisting of random noise. (Solid line) Values computed using a circular cylinder of radius $R = 2$ cm. (Dashed line) Values computed using an elliptical cylinder of semimajor axis $x_{max} = 4$ cm and semiminor axis $y_{max} = 2$ cm.

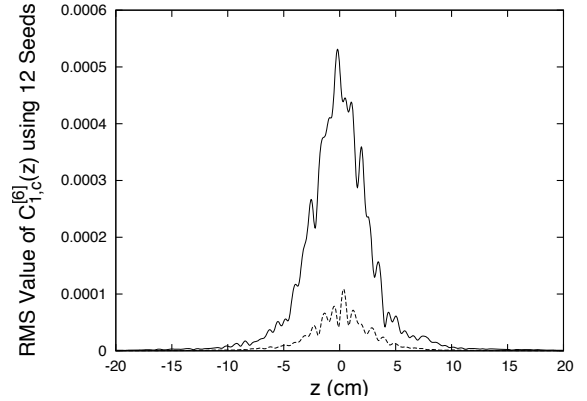


Figure 2: The quantity $C_{1,c}^{[6]}(z)$ computed from a field consisting of random noise. (Solid line) Values computed using a circular cylinder of radius $R = 2$ cm. (Dashed line) Values computed using an elliptical cylinder of semimajor axis $x_{max} = 4$ cm and semiminor axis $y_{max} = 2$ cm.

mesh) is that the magnetic field computed from the surface data should reproduce the magnetic field at the interior grid points. We computed such an interior fit for the modified CESR-c design of the Cornell wiggler, which has been adopted as the design prototype for use in International Linear Collider studies. Cornell provided data obtained from the 3-dimensional finite element modeling code OPERA-3d for the field components B_x , B_y , and B_z on a grid of spacing $0.4 \times 0.2 \times 0.2$ cm in a volume $10.4 \times 5.2 \times 480$ cm, extending beyond the fringe-field region. An elliptic cylinder with semimajor axis 4.4 cm and semiminor axis 2.4 cm was placed in the domain of the data, and the field on the elliptic cylinder boundary was constructed using nearest-neighbor interpolation with cubic splines. See Fig. 3.

The interior field was computed using generalized gradients through terms of degree 6 in x, y over the domain

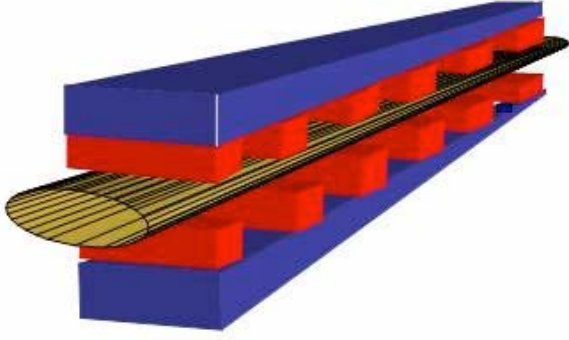


Figure 3: Schematic of the ILC wiggler and an elliptic cylinder centered on the z -axis, fitting within the bore of the wiggler, and extending beyond the fringe-field regions at the ends of the wiggler.

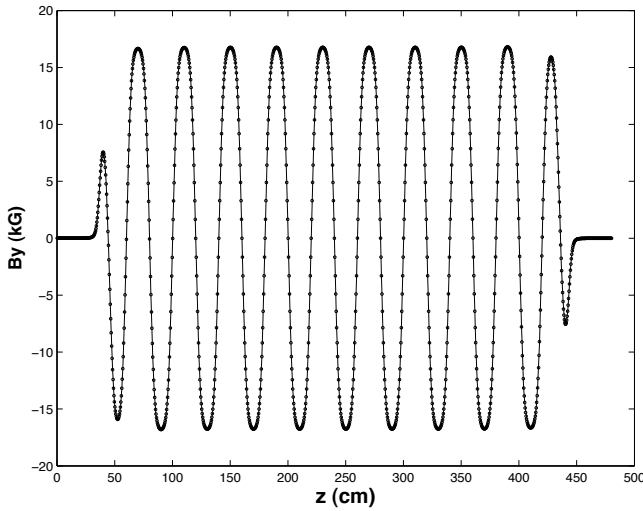


Figure 4: Fit obtained to proposed ILC wiggler vertical field using the elliptic cylinder of Fig. 3. The solid line is computed from surface data; dots are numerical data provided by OPERA-3d.

of the original data. This solution for the interior field was then compared to the original data at each grid point. Fig. 4 displays the fit to the vertical field B_y off-axis at $(x, y) = (0.4, 0.2)$ cm along the length of the wiggler. Note that the fitted field captures the fringe-field behavior. The RMS error obtained was $|\mathbf{B}_{data} - \mathbf{B}_{fit}|/|\mathbf{B}_{peak}| = 3.5 \times 10^{-4}$. All other field components are fit equally well at all interior points.

FITTING USING A BENT BOX

In this section we consider magnetic elements with large sagitta. We employ a bent box with straight ends surrounding the region of the beam, but excluding all iron or other Beam Dynamics, Other

sources (Fig. 5). Suppose \mathbf{B} and ψ are given on a grid, and these data are then interpolated onto the boundary of the box, which we call Γ . Given such values on Γ , the vector potential is given as the sum of two terms, $\mathbf{A} = \mathbf{A}^n + \mathbf{A}^t$, where

$$\begin{aligned} \mathbf{A}^n(\mathbf{r}) &= \int_{\Gamma} [\mathbf{n}(\mathbf{r}') \cdot \mathbf{B}(\mathbf{r}')] \mathbf{G}^n(\mathbf{r}; \mathbf{r}', \mathbf{m}(\mathbf{r}')) dS', \\ \mathbf{A}^t(\mathbf{r}) &= \int_{\Gamma} \psi(\mathbf{r}') \mathbf{G}^t(\mathbf{r}; \mathbf{r}', \mathbf{n}(\mathbf{r}')) dS'. \end{aligned} \quad (14)$$

Here $\mathbf{n}(\mathbf{r}')$ denotes the unit normal and $\mathbf{m}(\mathbf{r}') = \mathbf{r}'/r'$ at each point \mathbf{r}' on the surface of the box. The vector-valued integration kernels \mathbf{G}^n and \mathbf{G}^t are given by [3, 7]

$$\begin{aligned} \mathbf{G}^n(\mathbf{r}; \mathbf{r}', \mathbf{m}(\mathbf{r}')) &= \frac{\mathbf{m} \times (\mathbf{r} - \mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|(|\mathbf{r} - \mathbf{r}'| - \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}'))}, \\ \mathbf{G}^t(\mathbf{r}; \mathbf{r}', \mathbf{n}(\mathbf{r}')) &= \frac{\mathbf{n}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|^3}. \end{aligned} \quad (15)$$

It can be verified that each kernel \mathbf{G}^α has the two properties $\nabla \cdot \mathbf{G}^\alpha = 0$ and $\nabla \times \nabla \times \mathbf{G}^\alpha = 0$ within the region of interest, where derivatives are taken with respect to the variable \mathbf{r} . As a result, the vector potential \mathbf{A} given by (14) shares these properties. It follows that $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{B} = 0$, and \mathbf{A} satisfies the Coulomb gauge condition, for any surface data $\mathbf{n} \cdot \mathbf{B}$ and ψ , even if the data are noisy and the surface integrals are only evaluated approximately. Furthermore, \mathbf{G}^n and \mathbf{G}^t are analytic within the region of interest, and therefore \mathbf{A} is also analytic in this region. By expanding the kernels \mathbf{G}^α as power series in the transverse variables x and y , we may obtain corresponding power series for the vector potential \mathbf{A} .

This method has been implemented in a Fortran 90 routine [3]. This routine uses efficient truncated power series algebra (TPSA) algorithms to compute the Taylor series of the integral kernels (15) about each point on the design orbit. Each Taylor coefficient of the vector potential \mathbf{A} at a fixed value of z requires a single integration of the field and potential data over the surface Γ . The resulting coefficients may then be utilized to find design orbits and high-order transfer maps about these orbits.

Benchmarks

Consider the monopole doublet field described earlier. We interpolate \mathbf{B} and ψ onto the surface of a bent box with a bending angle of 30 degrees, and then employ the method described in the previous section to compute the vector potential \mathbf{A} and its Taylor coefficients about each point along a path through the center of the box using (14). See Fig. 5. The resulting vector potential is shown in Fig. 6 as a function of path length through the center of the box.

Using the power series for \mathbf{A} , power series for the components of \mathbf{B} about each reference point \mathbf{r}_d are computed using $\mathbf{B} = \nabla \times \mathbf{A}$. These results can then be compared to the known Taylor coefficients of the field. We find that all computed coefficients through terms of degree 4 are accurate to 10^{-6} in the arc section of the box, in the region

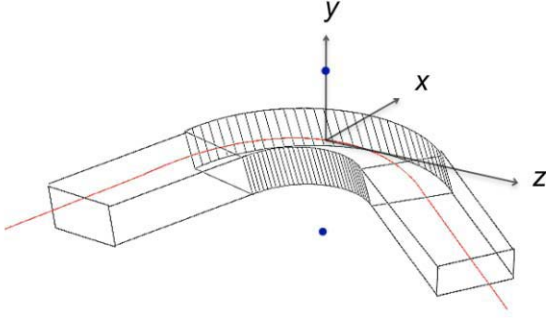


Figure 5: Illustration of a monopole doublet and a bent box with straight ends. The two dots denote equal and opposite magnetic charges. The red curve denotes a path through the center of the box.

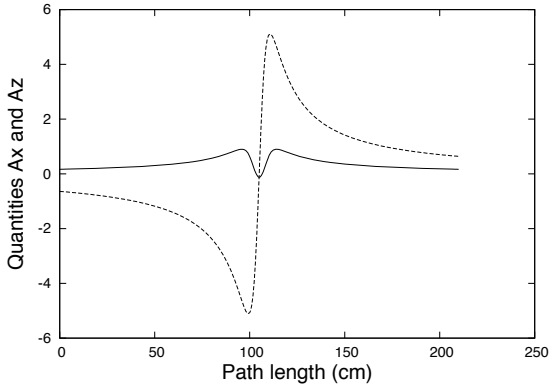


Figure 6: The vector potential of the monopole doublet illustrated in Fig. 5, computed from surface data, along a curved trajectory through the center of the bent box. In this case $A_y = 0$ and the quantities A_x (dashed line) and A_z (solid line) are plotted versus path length.

where the field is rapidly varying. This error increases very near the ends of the straight sections to 10^{-4} . The error at the ends is due to the fact that the field of a monopole doublet falls off quite slowly. The error would have been smaller had the straight ends of the box been made longer. The smoothing of numerical noise, as described earlier, can also be shown to occur in the case of the bent box. A detailed discussion of this issue can be found in [3].

CONCLUSIONS

A collection of surface fitting methods has been developed for providing accurate interior field data in analytic form based on 3-dimensional magnetic field data on a grid, as provided by various 3-dimensional finite element field codes. Each of these methods involves fitting field data onto a boundary surface and continuing inward to obtain ψ and/or \mathbf{A} and their Taylor coefficients in a neighborhood of the beam. These surface-fitting procedures have several

Beam Dynamics, Other

distinct advantages:

- The Maxwell equations are exactly satisfied.
- The results are manifestly analytic in all variables.
- The error is globally controlled. Both the exact and computed fields satisfy the Laplace equation. Therefore their difference, the error field, also satisfies the Laplace equation, and must take its extrema on the boundary. The fitting error on the boundary is controlled, and the interior error must therefore be even smaller.
- Interior values inferred from surface data are relatively insensitive to errors/noise in the surface data. In general, the sensitivity to noise in the data decreases rapidly (as some high inverse power of distance) with increasing distance from the surface, and this property improves the accuracy of the high-order interior derivatives needed to compute high-order transfer maps.

As a result one can, for the first time, obtain a realistic high-order transfer map for an entire accelerator or storage ring without the uncertainties associated with the use of only approximate field models.

REFERENCES

- [1] D. Abell, "Computation of High Order Maps for RF Cavities from Surface Data," these proceedings.
- [2] M. Venturini and A. Dragt, Nucl. Inst. and Meth. A 427, 387 (1999).
- [3] C. Mitchell, "Calculation of Realistic Charged-Particle Transfer Maps," Ph. D. thesis, University of Maryland, College Park (2007), <http://www.physics.umd.edu/dsat/>.
- [4] A. Dragt, *Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics*, University of Maryland Technical Report (2009), <http://www.physics.umd.edu/dsat/>.
- [5] M. Abramowitz and I. Stegun, *Handbook of Mathematical Formulas and Integrals* (Academic Press, 1995).
- [6] N. McLachlan, *Theory and Application of Mathieu Functions* (Dover Publications, Inc., 1964).
- [7] A. Dragt, T. Stashevich, and P. Walstrom, "Computation of charged-particle transfer maps for general fields and geometries using electromagnetic boundary-value data," PAC'01, Chicago, IL, June 2001, p. 1776 (2001).

CONSTRUCTION OF LARGE-PERIOD SYMPLECTIC MAPS BY INTERPOLATIVE METHODS *

Robert L. Warnock[†] and Yunhai Cai[‡]

SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA

James A. Ellison[§]

Dept. of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA

Abstract

The goal is to construct a symplectic evolution map for a large section of an accelerator, say a full turn of a large ring or a long wiggler. We start with an accurate tracking algorithm for single particles, which is allowed to be slightly non-symplectic. By tracking many particles for a distance S one acquires sufficient data to construct the mixed-variable generator of a symplectic map for evolution over S , given in terms of interpolatory functions. Two ways to find the generator are considered: (i) Find its gradient from tracking data, then the generator itself as a line integral. (ii) Compute the action integral on many orbits. A test of method (i) has been made in a difficult example: a full turn map for an electron ring with strong nonlinearity near the dynamic aperture. The method succeeds at fairly large amplitudes, but there are technical difficulties near the dynamic aperture due to oddly shaped interpolation domains. For a generally applicable algorithm we propose method (ii), realized with meshless interpolation methods.

1. INTRODUCTION

The method of differential algebra, giving automatic differentiation of functions defined by complex algorithms, allows the construction of the truncated Taylor series of a map defined by a tracking code. After this method was implemented by Martin Berz [1], the option of producing a Taylor map eventually became a feature of several tracking codes. The Taylor map is not symplectic, but some codes use the Taylor coefficients to produce the mixed-variable generator of a symplectic map, itself represented as a truncated power series [2]. Another way is to use the Taylor coefficients to form the symplectic “jolt factorization” of Irwin, Abell, and Dragt [3]. The Taylor map, symplectified or not, is good at small phase space amplitudes, but has a range of usefulness at larger amplitudes that varies with the type of accelerator lattice considered. It appears to be fairly useful for hadron rings, but can fail badly for electron rings with stronger nonlinearity near the dynamic aperture. In this paper we choose such a lattice for a demanding test of mapping methods.

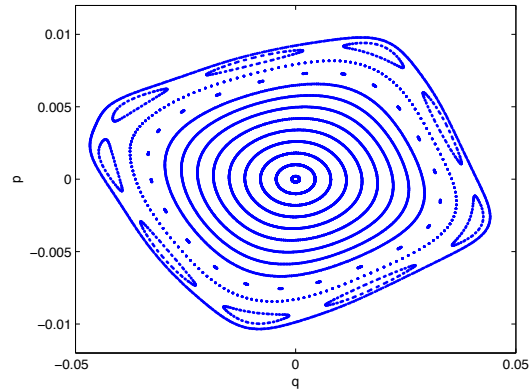


Figure 1: Phase plot from element-by-element tracking, 1000 turns, $\nu_x = 16.23$.

For our example the Taylor map fails at large amplitudes. For a striking illustration we choose a tune $\nu_x = 16.23$. Element-by-element tracking gives the plot of Fig.1 on a Poincaré section at a fixed position in the ring; p is dimensionless and q is in meters. The corresponding plot from iteration of the 10th order Taylor map is shown in Fig.2. The prominent 9th order island chain is only vaguely visible, and there is spurious stochasticity. The result is not improved by going to 13th order. The symplectified Taylor map [2] shows islands and gets rid of the stochasticity, but the shape of phase contours is all wrong. Changing to a better tune of $\nu_x = 15.81$, for which the lattice has a much larger dynamic aperture, we find that the Taylor map still breaks down at about the same amplitude.

One can easily see, however, that producing a more successful map when the Taylor series fails is not out of the question. A spline fit to one-turn tracking data on a grid of initial conditions gives a map which produces the plot of Fig.3 in 1000 iterations. To graphical accuracy it agrees with the tracking map of Fig.1, but the symplectic condition is badly violated at large amplitudes: the determinant of the map’s Jacobian differs from 1 at some points by as much as 0.004. Nevertheless it does not do badly over 10^5 iterations, as is seen in Fig.4. The main features of the phase plot persist correctly, but fuzziness appears near ends of the islands. By 10^6 turns there is a clear failure, with spurious damping, whereas phase curves including the islands are sharply defined in tracking for 10^6 turns. The spline is a tensor product B-spline interpolating tracking

* Work supported in part by U.S. Department of Energy contracts DE-AC02-76SF00515 and DE-FG-99ER41104.

[†] warnock@slac.stanford.edu; also affiliated with LBNL and UNM.

[‡] yunhai@slac.stanford.edu

[§] ellison@math.unm.edu

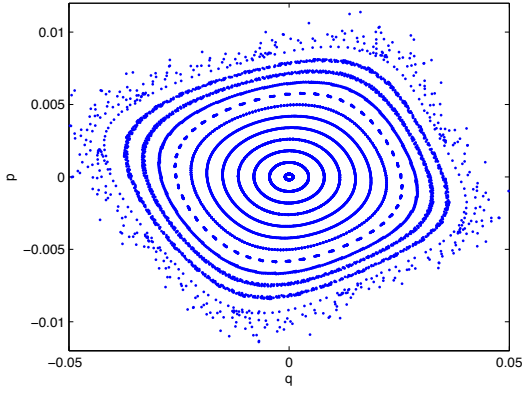


Figure 2: Phase plot from 10th order Taylor map, 1000 turns, $\nu_x = 16.23$.

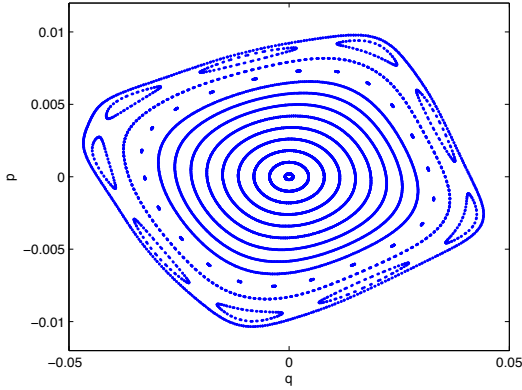


Figure 3: Phase plot from Spline Map, 1000 turns, $\nu_x = 16.23$.

data on a 40×40 uniform mesh; the spline coefficients were determined in 0.6 sec, and the time for 10^5 iterations was 0.14 sec. Computation times are for a single 2.66 GHz processor.

This example supports our belief that interpolative methods can produce maps that are both accurate and fast, even in cases where power series are not useful. Actually, the promise of interpolative map construction was evident long ago, in the work of Refs.[4, 5, 6, 7]. That work resulted in the generator of a fast symplectic map represented in polar coordinates in a hybrid Fourier-spline basis. It was applied successfully to an early LHC lattice, but in a slightly restricted region of phase space owing to a coordinate singularity arising from the polar coordinates. In work going back to 1984, G. Wüstefeld and collaborators have applied generating functions to maps for complicated magnetic fields, sometimes using interpolative methods [8].

In 1999 two of the authors proposed a method to find the map generator in Cartesian coordinates, with splines in all variables [9]. (Earlier, Berz had described a formally similar construction of the generator, but to be realized through power series rather than splines [10].) The

Linear and Nonlinear Orbit Theory

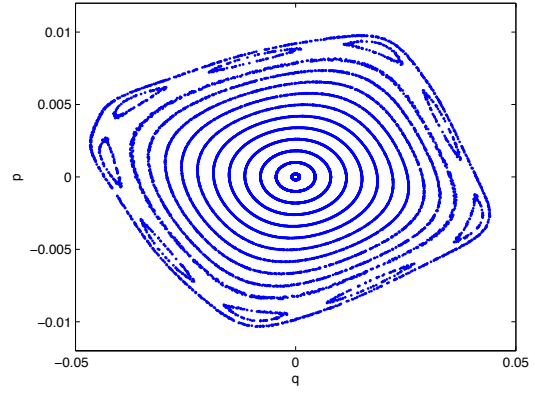


Figure 4: Phase plot from Spline Map, 10^5 turns, $\nu_x = 16.23$.

present paper gives the first numerical implementation of the scheme of Ref.[9]. Another idea, which came to light when one of the authors was preparing an article on the Hamilton-Jacobi equation [11], is to compute the generator directly as Hamilton's Principal Function, which is the integral of the Lagrangian regarded as a function of initial and final spatial coordinates.

2. RELATION OF THE GENERATOR TO THE TRACKING MAP

We denote the map defined through an element-by-element tracking code as follows:

$$q = Q(q_0, p_0), \quad (1)$$

$$p = P(q_0, p_0), \quad (2)$$

where $z = (q, p) = (z_1, \dots, z_{2n})$ and $z_0 = (q_0, p_0)$ are final and initial phase space points for a system with n degrees of freedom. The map can refer to an arbitrary period in a circular or linear machine, but in the present application it is for one turn of a circular machine. In the theory of canonical transformations [11] the map is expressed implicitly in terms of a generating function or generator, which we take to be a "Type 1" generator of the form $F(q, q_0)$. The implicit map is defined by the equations

$$p = F_q(q, q_0), \quad (3)$$

$$p_0 = -F_{q_0}(q, q_0), \quad (4)$$

where subscripts denote partial derivatives: $F_q = (\partial F / \partial q_1, \dots, \partial F / \partial q_n)$. We suppose that for q and q_0 in the region of interest, the $n \times n$ Hessian matrix F_{qq_0} is non-singular:

$$\det F_{qq_0}(q, q_0) \neq 0. \quad (5)$$

Then one can solve (4) for $q = q(q_0, p_0)$ (at least locally) and substitute the result in (3) to obtain also $p(q_0, p_0) = F_q(q(q_0, p_0), q_0)$ [12]. We wish to determine F so that the functions $q(q_0, p_0), p(q_0, p_0)$ can be identified with the map components $Q(q_0, p_0), P(q_0, p_0)$.

One can show that *any* $F(q, q_0)$ that has continuous second derivatives and satisfies (5) defines a symplectic transformation $(q(q_0, p_0), p(q_0, p_0))$. This fact is important for our map construction since it shows that symplecticity can be ensured even if F does not precisely reproduce the map (1),(2).

To derive the relation of F to the map (Q, P) , let us suppose that in the region of interest

$$\det Q_{p_0}(q_0, p_0) \neq 0, \quad (6)$$

in which case we can solve (1) for $p_0(q, q_0)$, provided that q is in the range of Q . Then, if F actually generates the map, Eqs. (3),(4), (2) show that

$$F_q(q, q_0) = P(q_0, p_0(q, q_0)) =: \gamma_1(q, q_0), \quad (7)$$

$$F_{q_0}(q, q_0) = -p_0(q, q_0) =: \gamma_2(q, q_0). \quad (8)$$

Symplecticity of the map guarantees that the vector $\gamma = (\gamma_1, \gamma_2)$ is actually a gradient; i.e., that it has zero curl. The proof, which is not obvious, is given in Ref.[13]. Thus we have $\nabla F = \gamma$ from which we can obtain F itself as a path-independent line integral,

$$F(\zeta) = \int_{\zeta_0}^{\zeta} \gamma(\zeta') \cdot d\zeta', \quad \zeta = (q, q_0). \quad (9)$$

3. APPROXIMATION OF THE GENERATOR AS A SPLINE

Our first realization of the scheme of the previous section is based on interpolation of data by spline functions, with the help of the B-spline basis [14]. Let $s(x)$ be any spline function specified by a knot sequence $[t_i]_{i=1}^{n+k}$ where k is the order (degree +1) of the local polynomials that are joined to make up the spline. The B-splines $B_i(x)$ determined by that knot sequence form a basis, so that for some λ_i

$$s(x) = \sum_{i=1}^n \lambda_i B_i(x). \quad (10)$$

The λ_i are fixed so that $s(x)$ interpolates data at distinct sites x_j , $j = 1, \dots, n$. This is done efficiently thanks to the banded nature of the interpolation matrix. The banded structure arises because at any x only k of the B-splines are non-zero. At a given x all of the non-zero B_i are computed at once by de Boor's stable recursive method; their derivatives can be obtained similarly. This gives a fast evaluation of (10) or its derivatives in a time that increases only mildly with n ; the increase is due only to a higher cost of searching for the knot interval in which x lies. Thus when the B-spline representation of a function is refined the evaluation time hardly changes, in marked contrast to a representation by Taylor series. For all B-spline operations we use standard Fortran software available at netlib.org [15].

For multidimensional interpolation the simplest approach is through a tensor product of B-splines. In 2D this is

$$s(x_1, x_2) = \sum_{i,j} \lambda_{ij} B_i^{(1)}(x_1) B_j^{(2)}(x_2), \quad (11)$$

Linear and Nonlinear Orbit Theory

a simple iteration of the 1D interpolation. The superscripts indicate possibly different knot sequences for the two dimensions. The tensor product interpolation requires data on a Cartesian grid, which can be used in the present study only at fairly small phase space amplitudes. At large amplitudes we have oddly shaped interpolation domains, which we handle by completing the array of real data with reasonable but arbitrary values. An item on our agenda is to look at more local interpolation methods that not only can handle general domains but also are more efficient in high-dimensions, for instance radial basis functions [16] or Shepard interpolation [17]

Our construction of F proceeds in the following steps:

1. Make a spline $\tilde{Q}(q_0, p_0)$ of $Q(q_0, p_0)$ from values on a Cartesian mesh $\{q_{0i}, p_{0j}\}$.
2. Solve (1) for $p_0(q, q_0)$ on a similar mesh $\{q_i, q_{0j}\}$. This is done by Newton's method, taking a first guess for p_0 from the linear part of the map. The required Jacobian is approximated by \tilde{Q}_{p_0} .
3. Using this solution, evaluate γ_1 in (7) on the same mesh, and then make a spline of (γ_1, γ_2) .
4. Integrate this spline on some convenient path to find $F(q_i, q_{0j})$ through (9).
5. Finally, make a spline to represent F from the values $F(q_i, q_{0j})$. This spline must be of order $k \geq 4$ (cubic or higher degree), to ensure that $F \in C^2$.

The map defined implicitly by F comes from (3),(4) and is evaluated as follows:

- a. For any initial point (q_0, p_0) solve (4) for q by Newton's method, taking the spline value $\tilde{Q}(q_0, p_0)$ as a first guess.
- b. Substitute the solution from (a) in (3) to obtain p as well.

In all the above steps the derivatives and integrals of splines are expressed analytically then evaluated numerically by means of BSPLVD and FPINTB [15]. Note that for a cubic spline the function $F_{q_0}(q, q_0)$ is C^2 in q , which makes it suitable for Newton's method in step (a).

4. APPLICATION TO AN ELECTRON RING LATTICE

We have carried out the map construction just described for horizontal motion in a lattice for an electron ring that was part of a study for an ILC damping ring. For a full description see [18]. It has racetrack form and entails 64 cells, primarily 90° FODO cells. The energy is 5 GeV, the length 960m, and the x -emittance is 47nm. A scaling to include more cells gives a lower emittance and a candidate for the damping ring. Tracking is done by the code LEGO, which integrates equations of motion based on the Hamiltonian in the local frame of each lattice component [19].

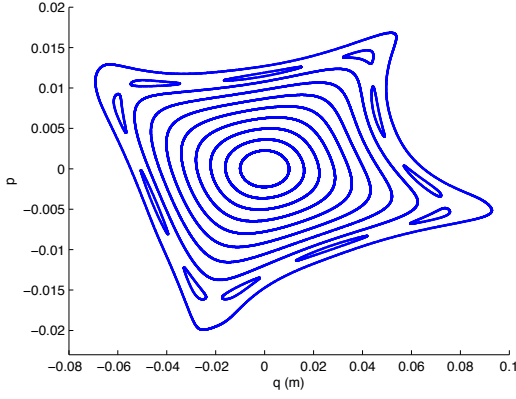


Figure 5: Phase plot from tracking, $\nu_x = 15.81$. Dynamic aperture (2000 turns) just beyond outer curve.

The motion of this example becomes very nonlinear as the dynamic aperture is approached. This makes it a challenging example for map construction. A hadron machine such as the LHC has weaker sextupoles and weaker nonlinearity out to the dynamic aperture, notwithstanding important effects of random higher order multipoles which contribute to long-term diffusion. On the basis of earlier work [5, 6, 7] we expect that the present algorithm will be useful for the LHC over a bigger range of amplitudes than it is in the present example. Nevertheless, the difficulties of this example have been very informative, and the lessons learned will certainly be relevant to other cases.

We now choose a tune, $\nu_x = 15.81$, for which the short term dynamic aperture is large, near the orbit with $(q_0, p_0) = (7\text{cm}, 0)$. There are no easily found resonances except for an 11th order one very close to the dynamic aperture. A phase plot is shown in Fig.5.

We try to construct a map on a rectangular region $|q_0| \leq r_1$, $|p_0| \leq r_2$, and expect that some sub-region will be mapped into itself by the constructed map. We therefore seek to construct $F(q, q_0)$ in a square domain, $|q|, |q_0| \leq r_1$. For $r_1 = 2.5\text{cm}$ everything goes according to the plan of the previous section. This is in accord with the mathematical analysis of Ref.[9] which showed that the scheme should work at small amplitude. Passing to $r_1 = 4.5\text{cm}$, we encounter a new situation. The solution for $p_0(q, q_0)$ does not exist in one corner of the domain, as is seen in Fig.(6). Let us call this triangular region \mathcal{B} , for Bermuda Triangle. The Newton iteration, which converges beautifully elsewhere, diverges in \mathcal{B} in a mode such that iterates get larger and larger. To generate the plot of successful solutions in Fig.6 the Newton iteration was declared a failure when iterates $p_0^{(n)}$ got bigger than a small multiple of r_2 , or when it failed to converge to desired accuracy within a set maximum number of iterations. As expected, the boundary of \mathcal{B} corresponds closely to the curve along which the condition of Eq.(6) first fails.

In order to make a tensor product spline of $\gamma(q, q_0)$ in spite of the missing values, we have to fill in the array Linear and Nonlinear Orbit Theory

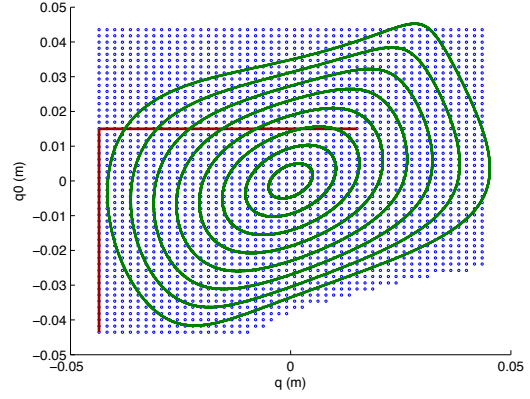


Figure 6: Points at which $p_0(q, q_0)$ exists, in blue. The curves represent values of (q, q_0) on the orbits of Fig.7. The path of integration for Eq.(9) is shown in red.

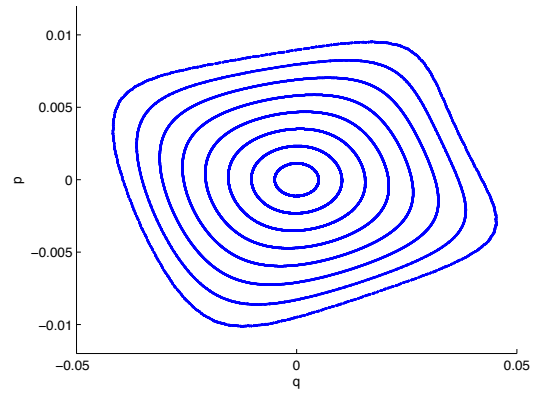


Figure 7: Symplectic map from generator, 10^7 turns. Iteration time 20-25 sec. for 10^7 turns, per orbit.

in some reasonable manner. We choose to put in values close to those on the boundary of \mathcal{B} , line-by-line in the q -direction. Of course, the path in the line integral must avoid the region \mathcal{B} . We choose the path shown in Fig.6.

This procedure with a 50×50 mesh leads to a symplectic map that gives the phase plot of Fig.7, where each orbit is followed for 10^7 turns in 20-25 seconds. We have also followed the outer orbit for 10^9 turns, finding no change to graphical accuracy from the result of 10^7 turns. Mapping time per turn is 20-200 times faster than the underlying tracking code (depending on the tracking integrator chosen), but this is not a fair comparison since our tracking code is not optimal for 2D tracking; it routinely computes the time of flight which is not needed in 2D. Nevertheless, the timing suggests a good outlook for speed in higher dimensions. The Newton iteration to solve (4) converged to machine precision in two or three steps.

The plot of Fig.7 is graphically indistinguishable from the corresponding plot from tracking over 10^5 turns. The quantitative agreement is not spectacular, however. At the

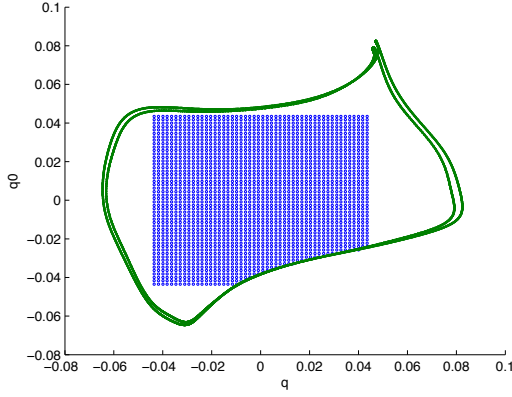


Figure 8: Two orbits in (q, q_0) plane close to the boundary of the region in which $F(q, q_0)$ exists. The dotted points are the same as those of Fig.6.

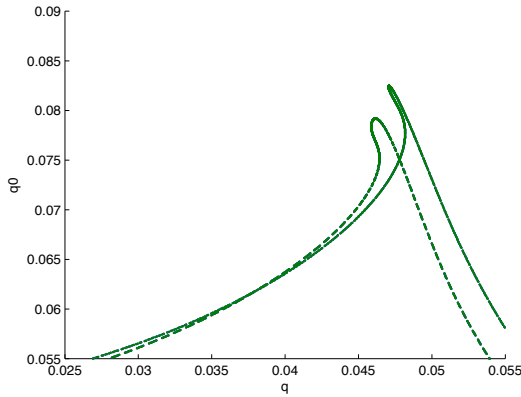


Figure 9: An enlargement of the top of Fig.8 showing intersecting orbits in (q, q_0) plane.

one turn level we have agreement to 5 digits at small amplitudes, declining to 4 digits at the outermost curve. As the map is iterated a lot of phase error quickly builds up, even though the iterates adhere to a clearly defined invariant curve. This is in accord with much earlier experience in approximating maps. For a quantitative test in spite of phase error one has to make an accurate representation of an invariant curve from tracking, then see how well the map follows it; this can be done with the method of [20].

When we try to go beyond the outermost curve of Fig.7, the method abruptly breaks down. The reason can be understood by using the tracking code to plot orbits in the (q, q_0) coordinates for increasingly large amplitudes. At some point two curves for neighboring initial conditions cross, as is seen in the upper right corner of Fig.8. There are two crossings, as seen in the enlargement in Fig.9. After this happens, the pair (q, q_0) does not determine a unique orbit, hence $p_0(q, q_0)$ is not a single-valued function and the construction of $F(q, q_0)$ must fail. The curve with smaller initial condition ($q_0 = 0.0604m$, $p_0 = 0$) roughly

Linear and Nonlinear Orbit Theory

defines the boundary of existence of F . As seen in Fig.8, the divergence of the Newton method defines part of the boundary quite accurately.

We should be able with sufficient care to construct the generator in its full region of existence. The reason that our attempted construction failed abruptly with increasing amplitude is seen by comparing Fig.6 and Fig.8, both of which show the points at which $p_0(q, q_0)$ was determined by Newton's method. The orbits of higher amplitude that we seek go into the cusp-like region in the upper right corner of Fig.8, but the path of integration used in the code can reach such points only by penetrating the boundary of the allowed region, thus giving nonsensical F . This could be avoided by a better choice of path, but that would be difficult if not impossible to generalize to higher dimensions. Even with a better path, we have to face the problem of interpolation, which becomes increasingly awkward for the tensor product spline.

One might ask whether a generator $F_2(q, p_0)$ exists when $F(q, q_0) = F_1(q, q_0)$ does not. Plotting the same two orbits of Fig.8 in (q, p_0) coordinates, we find again two intersections but at negative $q \approx -.0605, -.025$, rather than the positive $q \approx .0375, .0475$ of the intersections in (q, q_0) coordinates. Thus when (q, q_0) fails to specify an orbit (q, p_0) succeeds, and vice versa. Each of the generators fails to exist at some points on each of the two orbits. The symplectic condition in 2D phase space is $Q_{q_0}P_{p_0} - Q_{p_0}P_{q_0} = 1$ which shows that Q_{p_0} and Q_{q_0} cannot vanish simultaneously, hence either F_1 or F_2 must exist *locally*. A theorem in [21] generalizes this statement to higher dimensions.

5. GENERATOR CONSTRUCTION BY ACTION INTEGRAL

We have shown one way to construct the generator, through a partial map inversion to give $p_0(q, q_0)$ and a line integral. Another method goes back to Hamilton's original work of 1830-1832 [11]. He obtained the generator through the integral of the Lagrangian. Namely,

$$S(q_0, p_0, t) = \int_0^t [p(\tau, z_0) \cdot \dot{q}(\tau, z_0) - H(z(\tau, z_0), \tau)] d\tau, \quad (12)$$

where H is the Hamiltonian and the second argument of the orbit $z(\tau, z_0)$ indicates its initial condition, $z_0 = (q_0, p_0)$. The time-like variable t would normally be path length s in accelerator physics. Hamilton's essential idea was to regard the action as a function of (q, q_0) rather than (q_0, p_0) . Then the generator, which is a solution of the Hamilton-Jacobi equation called Hamilton's Principal Function, is

$$F(q, q_0, t) = S(q_0, p_0(q, q_0, t), t). \quad (13)$$

This is identical to the function of the previous sections when $t = T$ is the period of the map.

The advantage of this formula for a practical construction is that it avoids the line integral (9), which could be almost impossible to deal with in higher dimensions with excluded regions. A disadvantage is that the tracking code must be augmented to calculate the action integral in the course of tracking. Fortunately, LEGO is able to furnish the Hamiltonian needed in the integrand.

6. LOCAL INTERPOLATION AND THE USE OF SCATTERED DATA

Tensor product splines may be adequate in many problems but in general we need a more local sort of interpolation or approximation that can be used in non-rectangular domains. One possibility is a generalized Shepard method [17] based on the formula

$$F(\zeta) = \sum_i P_i(\zeta) \frac{w_i(\zeta)}{\sum_j w_j(\zeta)},$$

$$w_i(\zeta) = c(\zeta - \zeta_i) \|\zeta - \zeta_i\|^{-n}, \quad (14)$$

where n is a positive integer such as 4 or 6. Here $P_i(\zeta)$ is a polynomial that interpolates or approximates values of F at ζ_i and a few nearby sites. The factor $c(\zeta - \zeta_i)$ is a smooth cutoff that restricts the sum at any evaluation. For ζ close to ζ_i this behaves like $P_i(\zeta)$ with small corrections from other terms in the sum. The formula is globally smooth, with the degree of smoothness controlled by that of c .

This formula works when the sites ζ_i lie on a mesh or even when they are scattered. In the latter case one would normally use a least squares fit to determine the coefficients of P_i , rather than strict interpolation, since the interpolation matrix could be nearly singular for some dispositions of sites. Interpolation could be done by replacing the polynomial by a radial basis function [16].

As suggested by Fasshauer (private communication) it may be advantageous to use scattered sites from a quasi-random (low discrepancy) sequence. This might enjoy the advantage that quasi - Monte Carlo quadrature has over mesh based quadrature in high dimensions [22]. Such a scheme was tried in Ref.[23], in connection with solution of a Vlasov equation. In 2D it was possible to reduce the number of data sites by a factor of 8 in comparison to mesh based interpolation. A much bigger advantage is expected in higher dimensions, and this augurs well for construction of maps in 4D or 6D phase space using quasi-random data.

7. CONCLUSION

After analysis of a difficult example in 2D phase space, we have a plan for building fast symplectic maps in higher dimensions. The method described in Sec.3 will probably succeed in many cases, but in general we shall need a more powerful approach using local interpolation and Hamilton's Principal Function. The use of scattered interpolation sites from quasi-random sequences may increase the efficiency of interpolation in high dimensions.

Linear and Nonlinear Orbit Theory

ACKNOWLEDGMENT

We thank J. Scott Berg and Ronald Ruth for helpful remarks and encouragement.

REFERENCES

- [1] M. Berz, Particle Accelerators **24**, 109 (1989).
- [2] D. Douglas, É. Forest, and R. V. Servranckx, IEEE Trans. Nuc. Sci., **NS-32**, 2279 (1985); Y. Yan, P. Channell, and M. Syphers, SSCL-Preprint-157 (1992).
- [3] D. T. Abell, E. McIntosh, and F. Schmidt, Phys. Rev. ST Accel. Beams **6**, 064001 (2003) and references therein.
- [4] J. S. Berg and R. L. Warnock, Proc. IEEE 1991 Part. Accel. Conf., San Francisco, p.1654.
- [5] J. S. Berg, R. L. Warnock, R. D. Ruth, and É. Forest, Phys. Rev. E **49**, 722-739 (1994).
- [6] R. L. Warnock and J. S. Berg, Particle Accelerators **54**, 213-222 (1996).
- [7] R. L. Warnock and J. S. Berg, AIP Conf. Proc. **395**, 423-445 (1997).
- [8] G. Wüsfeld, "A Generating Function for Particle Tracking", note obtained from wuestefeld@bessy.de; M. Scheer, Ph.D. Thesis, Humboldt-Universität, Berlin (2008).
- [9] R. L. Warnock and J. A. Ellison, Applied Numerical Math. **29**, 89-98 (1999).
- [10] M. Berz, in *Nonlinear Problems in Future Accelerators*, (World Scientific, Singapore, 1991), pp. 288-296.
- [11] R. L. Warnock, "The Hamilton-Jacobi Equation", under Dynamical Systems at scholarpedia.com.
- [12] Although a local solution does not imply a single-valued global solution, we can proceed assuming (5) and try to verify a global solution.
- [13] R. L. Warnock and J. A. Ellison, AIP Conf. Proc. **405**, 41-45 (1997).
- [14] C. de Boor, "A Practical Guide to Splines", Revised Edition, (Springer, New York, 2001)
- [15] See libraries pppack and dierckx at netlib.org. In pppack find SPLINT for interpolation, BSPLVD with INTERV for evaluation and differentiation. In dierckx find FPINTB for integration.
- [16] G. Fasshauer, "Meshless Approximation Methods with Matlab", (World Scientific, Singapore, 2007).
- [17] R. Farwig, Math. of Computation **46**, 577-590 (1986).
- [18] Y. Cai, SLAC-PUB-11084 (2005).
- [19] Y. Cai, SLAC-PUB-8011 (1998).
- [20] R. L. Warnock and R. D. Ruth, Physica D **56**, 188-215 (1992).
- [21] V. I. Arnol'd, "Mathematical Methods of Classical Mechanics", (Springer, New York, 1978), Section 48-B.
- [22] H. Niederreiter, "Random Number Generation and Quasi - Monte Carlo Methods", (SIAM, Philadelphia, 1992).
- [23] R. L. Warnock, J. A. Ellison, K. Heinemann, and G. Q. Zhang, Proc. EPAC'08, Genova, paper TUPP109.

HIGHLY SCALABLE NUMERICAL METHODS FOR SIMULATION OF SPACE CHARGE DOMINATED BEAMS *

J. Xu[#], B. Mustapha, P. N. Ostroumov, and J. A. Nolen, Argonne National Laboratory, Argonne, IL 60439, USA

ABSTRACT

Beam dynamic simulations with kinetic model have been conducted. We have successfully parallelized a PIC solver, TRACK, and developed new Vlasov solvers. For the PIC solver, particles are distributed evenly on different processors and space charge effect has been counted by solving Poisson's equation on a finite mesh. Several Poisson solvers have been developed using Fourier method in Cartesian coordinate system, Fourier Spectral Element in Cylindrical coordinate system, Wavelet method, Spectral Element Method (SEM) on structured and unstructured grids. Domain decomposition (DD) has been used to parallelize these solvers. Different Poisson solvers have been developed for simulating space charge dominated beams. These solvers have been incorporated into PTRACK and Vlasov solvers. PTRACK has now widely been used for large scale beam dynamics simulations in linear accelerators. For the Vlasov solver, Semi-Lagrangian method and time splitting scheme have been employed to solve Vlasov equation directly in 1P1V and 2P2V phase spaces. Similarly, DD has been used for parallelization of Vlasov solvers.

INTRODUCTION

Plasma and charged particle simulations have great importance in science. There are three different approaches to simulate plasmas: the microscopic model, the kinetic model and the fluid model. In the microscopic model, each charged particle is described by 6 variables (x, y, z, v_x, v_y, v_z). Therefore, for N particles, there are $6N$ variables in total. This requires solving the Vlasov equation in $6N$ dimensions, which exceeds the capability of current supercomputers for large N . On the other end is the fluid model which is the simplest because it treats the plasma as a conducting fluid with electromagnetic forces exerted on it. This leads to solving the Magneto-hydrodynamics (MHD) equations in 3D (x, y and z). MHD solves for the average quantities, such as density and charge, which makes it difficult to describe the fine structure in the plasma. Between these two models is the kinetic model, which solves for the charge density function by solving the Boltzmann or Vlasov equations in 6 dimensions (x, y, z, v_x, v_y, v_z). The Vlasov equation describes the evolution of a system of particles under the effects of self-consistent electromagnetic fields. This paper deals with the kinetic model.

There are two different ways to solve the kinetic model. The most popular one is to represent the beam bunch by macro particles and push the macro particles along the characteristics of the Vlasov equation. This is the so called Particle-In-Cell (PIC) method, which utilizes the motion of the particles along the characteristics of the Vlasov equation using a Lagrange-Euler approach [1, 2]. The PIC method has the advantages of speed and easy implementation, but similar to MHD, it is hard to calculate fine structures in the plasma. Furthermore, there is noise associated with the finite number of particles in the simulation. This noise decreases very slowly, as $1/\sqrt{N}$, when the number of particles N is increased. The other way to solve the kinetic model is to solve the Vlasov equation directly. This can overcome the shortcomings of the PIC method. We have applied SEM which can achieve high order accuracy and developed scalable Poisson and Vlasov solvers. This paper reports our work using both models. In order to describe space charge effects, several Poisson solvers have been developed.

BEAM DYNAMIC SIMULATION WITH PIC SOLVER

In the last several years, we have parallelized a PIC solver, TRACK, which has been developed in physics division at ANL. Parallel algorithm and detailed benchmark results can be in [2, 3, 4]. Recently PTRACK has been used for an one-to-one RFQ simulation of FNAL proton driver. Totally 865M charged particles have been simulated from 50 keV to 2.5 MeV in 325 MHz radio frequency quadrupole of a proton driver at FNAL. Figure 1 is the comparison in $(\phi, \Delta W/W)$ plane. This result provides much more accurate information and useful to the design optimizations. Now PTRACK has been used as workhorse for large scale optimizations.

PARALLEL POISSON SOLVERS

Fourier Method

$$\phi(x, y, z, t) =$$

$$\sum_{m=-M/2}^{M/2-1} \sum_{p=-P/2}^{P/2-1} \sum_{n=-N/2}^{N/2-1} \phi(m, p, n, t) e^{-iamx} e^{-i\beta py} e^{-i\gamma nz}$$

This is the most standard method for solving the Poisson's equation in Cartesian coordinate system. The potential has been expanded in Fourier series in all three directions. Periodic and Dirichlet zero boundary conditions have been applied in all three directions. Three different domain decomposition methods have been implemented as shown in Fig.2. Using model C, it is easy to use tens of thousands of processors with relatively small grid for space charge calculation. Since

* This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357.
[#]jin_xu@anl.gov

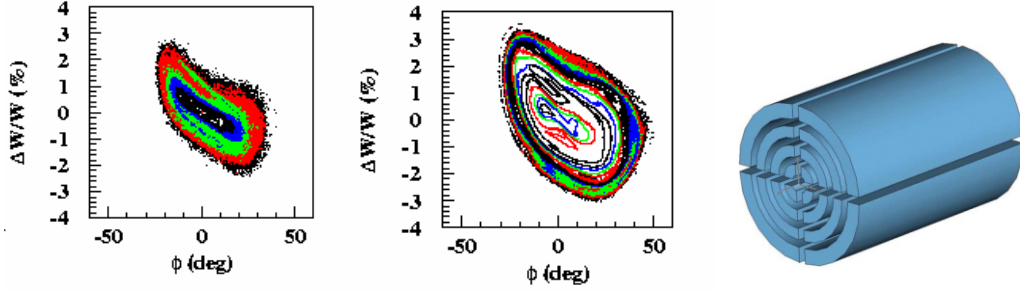


Fig. 1: Phase contour comparison in $(\phi, \Delta W/W)$ plane and parallel model in cylinder coordinate system

relatively small grid can be used for space charge calculation, good scaling has been obtained and can be found in [3, 4].

Fourier Spectral Element Method

This solver is developed for cylinder coordinate system. The potential is expanded in Fourier series in the axial and circumferential directions, while it uses spectral element expansion in the radial direction.

$$\phi(x, y, z, t) = \sum_{m=-M/2-1}^{M/2-1} \sum_{n=-N/2}^{N/2-1} \sum_{p=0}^P \phi(m, p, n, t) e^{-i\alpha m x} e^{-i\beta n z} P_p(y)$$

$$P_p(y) = \begin{cases} \left(\frac{1-y}{2}\right) & p=0 \\ \left(\frac{1-y}{2}\right)\left(\frac{1+y}{2}\right)P_{p-1}^{1,1}(y) & 0 < p < P \\ \left(\frac{1+y}{2}\right) & p=P \end{cases}$$

Domain decomposition in the radial and circumferential directions has been implemented as shown in the right of Fig. 1. Periodic B.C. has been applied in the axial and circumferential directions and zero Dirichlet B.C. have been applied in the cylinder wall. Detailed method and benchmark results can be found in [2].

Wavelet Method

Recently, multi-resolution analysis (MRA) emerges

$$V_j = \{\phi_{j,k}, k \in \mathbb{Z}\} \quad \phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$$

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2(R)$$

Since $V_j \supset V_{j+1}$, we can define $W_j \perp V_j$ and $W_j \supset V_{j+1}$, such that $V_j \oplus W_j = V_{j+1}$, where

$$W_j = \{\psi_{j,k}, k \in \mathbb{Z}\} \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$$

$$\begin{aligned} \text{Then } V_2 &= W_1 \oplus V_1 = W_1 \oplus W_0 \oplus V_0 \\ &= W_1 \oplus W_0 \oplus W_{-1} \oplus V_{-1} \end{aligned}$$

There is an efficient preconditioning matrix for the linear system generated by the wavelet expansion. We use following analytical solution to test this wavelet Poisson solver.

$$\begin{aligned} \phi(x, y, z) &= (x + \pi)(x - 3\pi) \\ &\quad (y + \pi)(y - 3\pi)(z + \pi)(z - 3\pi) \\ \Delta \phi(x, y, z) &= -\rho(x, y, z) \\ \rho(x, y, z) &= 2(x + \pi)(x - 3\pi)(y + \pi)(y - 3\pi) \\ &\quad + 2(x + \pi)(x - 3\pi)(z + \pi)(z - 3\pi) \\ &\quad + 2(y + \pi)(y - 3\pi)(z + \pi)(z - 3\pi) \end{aligned}$$

The computation domain is $[-\pi, 3\pi]^3$, and zero Dirichlet B.C. has been used. The history of the converging error is shown in the left of Fig. 3, the dash line corresponding to no preconditioning, and the solid line corresponding to with preconditioning. Using preconditioning the speed is about three times faster as

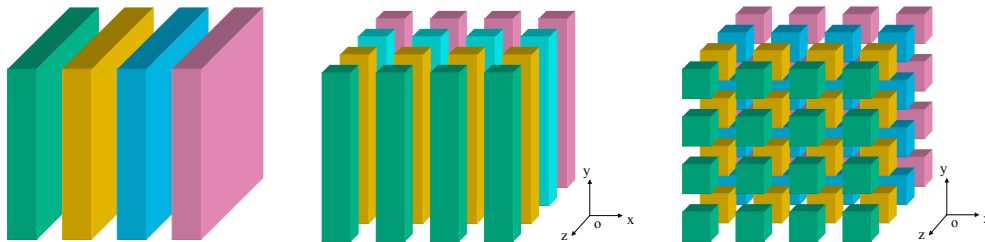


Fig. 2: Three parallel models (A, B and C) for solving Poisson's equation using Fourier method

as a powerful tool to analyses multi-scale phenomena. We have developed a Poisson solver using wavelet expansions. A MRA of consists of a chain of closed subspaces

before. Using wavelet, there are only few percent of total coefficients are larger than some critical value, as shown on the middle of Fig. 3.

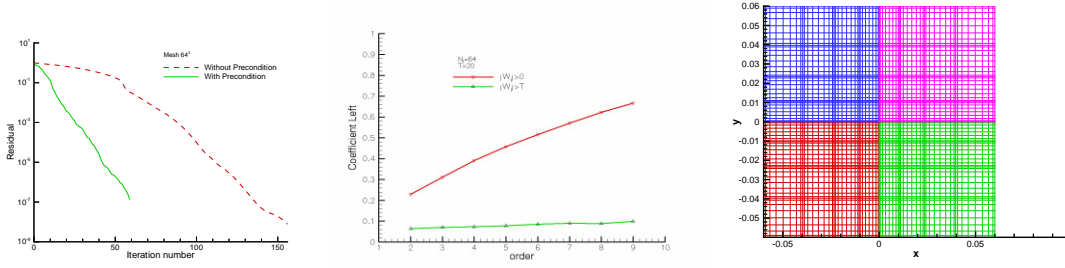


Fig. 3: Convergence history with/out preconditioning (left), wavelet coefficients (middle) and 2D structured grid (right)

Domain decomposition has also been used for parallelization, and relatively good scaling has been achieved. It has also been incorporated into PTRACK code, nearly the same results have been obtained.

Spectral Element Method on Structured Grid

The Spectral Element Method (SEM) originated in the 1980's [6, 7, 8], and has been applied in many different areas. In our Vlasov solvers, a parallel Poisson solver based on Spectral Element Method (SEM) on structured grid has been constructed. 2D structured grid has been shown on the right of Fig. 3. 2D bases have been shown on the left of the Fig. 4. Similarly, domain decomposition has been used for parallelization with Dirichlet boundary conditions. Continuous Galerkin (CG) method has been used and zero Dirichlet B.C. has been imposed. Due to the memory limitation, only the iterative solver can be used for solving boundary modes of the 2D Poisson's equation when the mesh is large. Interior modes in each element have been solved directly according to the Shur complement. The discrete system of Poisson's equation can be written as: (b and i correspond to boundary and interior variables)

$$\begin{pmatrix} A_{bb} & C_{bi} \\ C_{bi}^T & A_{ii} \end{pmatrix} \begin{pmatrix} u_b \\ u_i \end{pmatrix} = \begin{pmatrix} f_b \\ f_i \end{pmatrix}$$

$$(A_{bb} - C_{bi} A_{ii}^{-1} C_{bi}^T) u_b = f_b - C_{bi} A_{ii}^{-1} f_i$$

$$u_i = A_{ii}^{-1} (f_i - C_{bi}^T u_b)$$

Figure 4 (middle) and (right) show the charge and

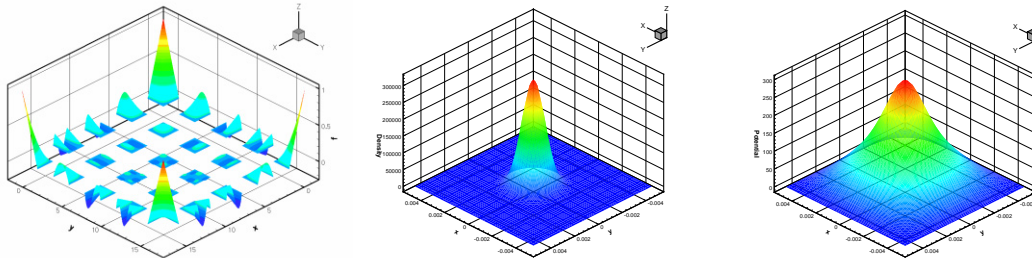


Fig. 4: Modal bases (left) on 2D structured grid, charge (middle) and potential (right) distributions

potential distributions. Since zero Dirichlet B.C. has been used, the potential is nonzero close to the boundary. This means the domain size need large enough to obtain the right potential distribution.

Spectral Element Method on Unstructured Grid

In order to solve Poisson in complex geometries, a parallel Poisson solver using SEM on an unstructured grid has been developed recently. Since finite element method (FEM) can handle complex geometry easily, and spectral method can achieve high order accuracy. Combine these two, SEM can handle the complex geometry and also achieve high order accuracy at the same time.

The potential can be expressed as

$$f(r, s) = \sum_{i+j \leq p} \psi_{ij}(r, s) \hat{f}_{ij}$$

The continuous Galerkin formula for solving the Poisson's equation is

$$\begin{aligned} \int_V f \cdot \phi dv &= \int_V \nabla^2 u \cdot \phi dv \\ &= \int_V [\nabla(\nabla u \phi) - \nabla u \cdot \nabla \phi] dv \\ &= \int_{\partial V} \bar{n} \cdot (\nabla u \phi) dS - \int_V \nabla u \cdot \nabla \phi dv \\ \int_V \nabla u \cdot \nabla \phi dv &= \int_{\partial V} \bar{n} \cdot (\nabla u \phi) dS - \int_V f \cdot \phi dv \\ \int_{\partial V} \bar{n} \cdot (\nabla u \phi) dS &= \int_{\partial V} (n_x \phi \frac{\partial u}{\partial x} + n_y \phi \frac{\partial u}{\partial y}) dS \end{aligned}$$

A mesh partition for 4 processors has been shown on the left of Fig. 5. Middle and right plots in Fig. 5 show the charge and potential distributions on a circular domain.

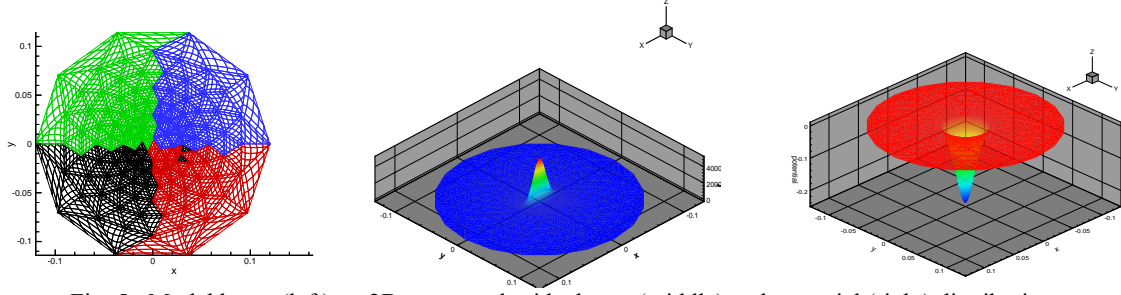


Fig. 5: Modal bases (left) on 2D structured grid, charge (middle) and potential (right) distributions

DYNAMIC SIMULATION WITH VLASOV SOLVERS

In order to overcome the shortcoming of the PIC solvers, we have developed direct Vlasov solvers. The distribution function $f(\vec{x}, \vec{v}, t)$ in phase space is governed by the Vlasov equation.

Vlasov equation in 1P1V phase space

In 1P1V phase space, the non-dimensional Vlasov equation can be written as following:

$$\frac{\partial f(x, v, t)}{\partial t} + v(x, t) \frac{\partial f(x, v, t)}{\partial x} + E(x, t) \frac{\partial f(x, v, t)}{\partial v} = 0$$

$$E(x, t) = -\frac{\partial \phi(x, t)}{\partial x}, -\Delta \phi(x, t) = \frac{\partial E(x, t)}{\partial x} = \rho(x, t) - 1$$

$$\rho(x, t) = \int_{-\infty}^{\infty} f(x, v, t) dv$$

Vlasov equation in 2P2V phase space

In beam dynamics, a simplified model can be deduced in 2P2V form as a paraxial model based on the following assumptions:

- The beam is in a steady-state: All partial derivatives with respect to time vanish;
- The beam is sufficiently long so that the longitudinal self-consistent forces can be neglected;
- The beam is propagating at a constant velocity v_b

along the propagation axis z ;

- Electromagnetic self-forces are included;
- $\vec{p} = (p_x, p_y, p_z)$, $p_z \sim p_b$ and $p_x, p_y \ll p_b$

where $p_b = \gamma m v_b$ is the beam momentum. It follows in particular that

$$\beta \approx \beta_b = (v_b / c)^2, \gamma \approx \gamma_b = (1 - \beta_b^2)^{-1/2}$$

- The beam is narrow: the transverse dimensions of the beam are small compared to the characteristic

longitudinal dimension. The paraxial model can be written as:

$$\frac{\partial f}{\partial z} + \frac{\vec{v}}{v_b} \cdot \nabla_{\vec{x}} f + \frac{q}{\gamma_b m v_b} \left(-\frac{1}{\gamma_b^2} \nabla \Phi^s + \vec{E}^e + (\vec{v}, v_b)^T \times \vec{B}^e \right) \cdot \nabla_{\vec{v}} f = 0$$

Coupled with Poisson's equation

$$-\Delta_{\vec{x}} \Phi^s = \frac{q}{\epsilon_0} \int_{R^2} f(z, \vec{x}, \vec{v}) d\vec{v}$$

, where Φ^s is the self-consistent electric potential due to charges. \vec{E}^e and \vec{B}^e are external electric and magnetic fields. v_b is the reference beam velocity.

Numerical Algorithm

The Semi-Lagrangian Method (SLM) [9] has been used for time integration. A plot explains the idea has been shown on the left of Fig. 7. The time splitting scheme has been used for time integration as proposed by Cheng and Knorr [10].

```

.....
Do istep=1, nstep:
- Compute  $\rho(x, t) = \int f(x, v, t) dv$ ;
- Compute  $E^{pred} = E^n - j^n \Delta t$  from Ampere's law;
- Do until  $|E^{N=1} - E^{pred}| < \epsilon$ 
    · Substep1:  $v^{n+1/2} = v^{n+1} - E^{pred}(x^{n+1}) \Delta t / 2$ 
    · Substep2:  $x^n = x^{n+1} - v^{n+1/2} \Delta t$ ;
    · Substep3:  $v^n = v^{n+1/2} - E^n(x^n) \Delta t / 2$ ;
    · Interpolate to compute charge density;
    · Solve Poisson's equation for  $E^{n+1}$ ;
    · Update new  $E^{pred} = E^{n+1}$ .
- Enddo
Enddo
.....
    
```

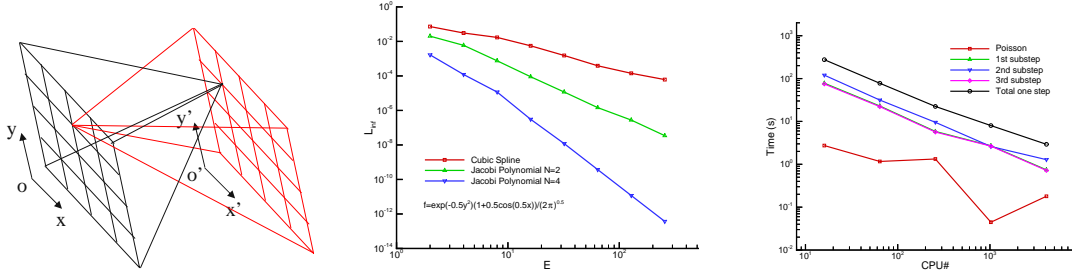


Fig. 6: 4D domain decomposition (left), Interpolation errors vs. element number (middle) and strong scaling in 2P2V simulation (right)

The algorithm for 2P2V simulation is similar to the above, but with advancing the first and the last substeps in physical space and the second substep in the velocity space.

Table 1. Scaling 2D Poisson solver ($E=64$, $P=4$)

CPU	16	64	256	1024	4096
Time (s)	286	68	17.2	4.08	1.66
PE	1.0	1.0	1.0	1.0	0.673

Benchmarks and Simulation Results

The code comprises two major parts: interpolation and space charge (SC) calculation. The SLM performs back tracking and interpolation respectively in the physical and velocity spaces. Each processor has only part of the global mesh for the space charge calculations. The field mesh and space charge mesh are different. This scheme has the advantage of easy implementation and no communication for particle tracking is required. However, this method requires large memory in each processor and intense communication for the parallel Poisson solver. Figure 6 (left) shows the domain decomposition in 4D for 2P2V simulations.

Table 1 shows the benchmark results for the 2D Poisson solver. Good scaling has been achieved. Figure 6 (middle) compares the interpolation errors with cubic spline, Jacobi polynomial with $P=2$ and 4. Clearly using a Jacobi polynomial gives much better results, which is good to use in the Semi-Lagrangian scheme. The right plot in Fig. 6 shows the strong scaling results for both the Poisson and Vlasov solvers in 2P2V simulations. It

shows that the Vlasov solver can have good scaling because the most time consuming part is the interpolation. And since the interpolations are local on each processor, there is no communication between different processors. So even when the scaling of the Poisson solver becomes worse with 4k processors, the overall scaling is still good.

The middle and right figures in the Fig. 7 show the time history of $\log(E_x)$ for linear and strong Landau damping. The initial particle distribution function and the related parameters are shown in following:

$$f(0, x, v) = \frac{1}{\sqrt{2\pi}} \exp(-v^2/2)(1 + \alpha \cos(kx))$$

$$\forall (x, v) \in [0, L] \times R, \quad \alpha = 0.01, \quad k = 0.5$$

$$L = 4\pi, \quad R = [-6, 6], \quad \Delta t = 1/8,$$

$$P = 16, \quad E = 64, \quad 1024 \times 1024$$

$$CPU = 256, \quad T \sim 10 \text{ min } s$$

For the linear Landau damping, $\alpha=0.01$, and for the strong Landau damping, α is 0.5. Clearly they represent different dynamics. The decreasing and increasing rate can be measured and are consistent with theoretical predictions and other researchers.

2P2V Simulations

In 2P2V simulations, a proton beam has been simulated through alternating hard edge electric quadrupole channel. The initial emittance is $\varepsilon = 200\pi$ mm mrad, and the energy is $W=0.2$ MeV. The current of the beam is 0.1 A, and the reference velocity is $v_b = 6.19 \times 10^6$ m/s. The transverse physical space is $[-0.12, 0.12]$ by $[-0.12, 0.12]$, and the velocity space is

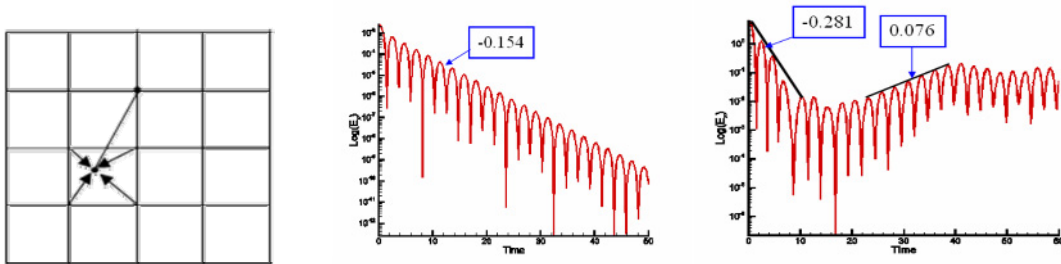


Fig. 7: Semi-Lagrange Scheme (left), Linear Landau Damping (middle), Strong Landau Damping (right)

$[-8 \times 10^5, 8 \times 10^5]$ by $[-8 \times 10^5, 8 \times 10^5]$ m/s. The alternating electric quadrupole field is defined as $\vec{E}^e(x, y, z) = (k_0(z)x, -k_0(z)y)$ and shown in Fig. 8.

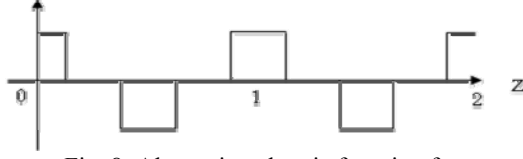


Fig. 8: Alternating electric focusing force

Since the initial beam distribution is Gaussian (not a KV distribution), the RMS envelope is not periodic with the amplitude fluctuating from one period to the next. Figure 9 shows the beam contours in (x, y) , (x, x') , (y, y') and (x', y') phase planes at $z=0$ and 192 steps. Detailed information on 1P1V and 2P2V Vlasov simulations can be found in [2].

SUMMARY

This paper presents our researches on beam dynamic simulations with PIC method, different parallel Poisson solvers, and beam dynamic simulations with direct Vlasov solvers. Domain decomposition has been adopted for parallelization of TRACK code, and several parallel Poisson solvers have been developed and incorporated into PTRACK. PTRACK has now been used for large scale beam dynamic optimization and simulations. Several numerical techniques have been used to solve Poisson's equation in different conditions, such as using Cartesian and Cylindrical coordinate systems, using structure and unstructured grids, etc. Direct Vlasov solvers have been developed with a high-order SEM. The advantages and effectiveness of the SEM have been demonstrated. The Vlasov solvers have adopted the Semi-Lagrangian method. Similarly domain decomposition has been used for parallelization of these solvers. Scalable Poisson solvers have been developed within. Benchmarks of the parallel models have shown good scaling on BlueGene/P at ANL with up to 4k processors. The SEM shows its advantages in these direct Vlasov solvers, such as local interpolation, easy parallelization and long time integration. These explorations are encouraging, and more investigation will be done.

REFERENCES

- [1] P.N. Ostroumov, V. N. Aseev, and B. Mustapha. Phys. Rev. ST. Accel. Beams 7, 090101 (2004).
- [2] J. Xu and P.N. Ostroumov, Computer Physics Communications, 178 (2008) 290-300.
- [3] J. Xu, B. Mustapha, V.N. Aseev and P.N. Ostroumov, Phys. Rev. ST. Accel. Beams 10, 014201 (2007).
- [4] J. Xu, B. Mustapha, V. N. Aseev and P.N. Ostroumov, HB 2008, Nashville, Tennessee, USA, Aug. 25- 29, 2008.
- [4] M. Gutnic, M. Haefele, I. Paun and E. Sonnendrücker, Comput. Phys. Comm., 164, 214-219 (2004).

- [5] E. Sonnendrücker, J. Roche, P. Bertrand and A. Ghizzo, J. Comput. Phys., 149, 201 (1998).
- [6] F. Filbet, E. Sonnendrücker, Vol.16, No.5, 763-791 (2006).
- [6] M.O. Deville, P.F. Fischer and E.H. Mund, High-Order Methods for Incompressible Fluid Flow, Cambridge University Press, Cambridge (2002).
- [7] J.S. Hesthaven and T. Warburton, Nodal Discontinuous Galerkin Methods: Algorithms, Analysis and Applications, Springer, New York (2008).
- [8] G.E. Karniadakis and S.J. Sherwin, Spectral/hp element methods for CFD, Oxford University Press, London (1999).
- [9] A. Robert, Atmos. Ocean. 19, 35 (1981).
- [10] C.Z. Cheng and G. Knorr, J. Sci. Comp., 22, 330 (1976).

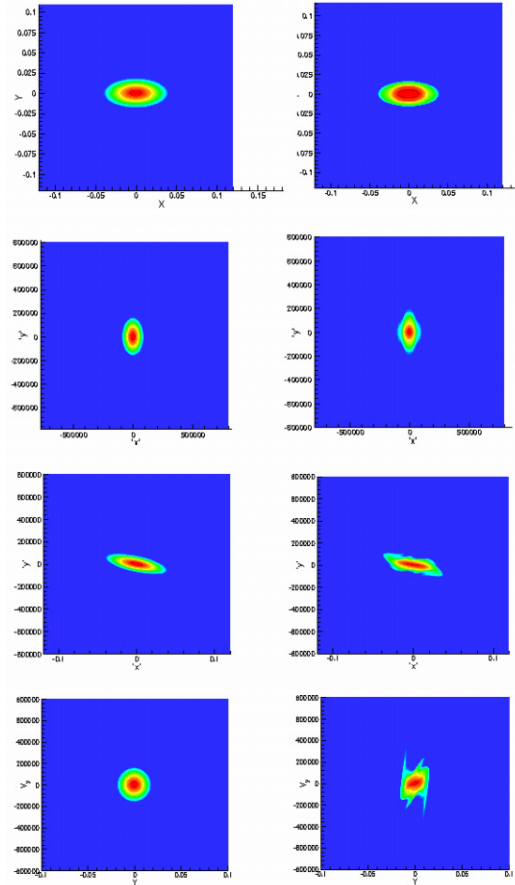


Fig. 9: From top to bottom are contours in the (x, y) , (x, x') , (y, y') and (x', y') planes, from left to right correspond to $z=0$ and 192 time steps.

OVERVIEW OF (SOME) COMPUTATIONAL APPROACHES IN SPIN STUDIES

F. Lin, A.U. Luccio, N.D. Malitsky, W.M. Morse, Y.K. Semertzidis, BNL, Upton, NY, USA
 C.J.G. Onderwater, University of Groningen, NL-9747AA Groningen, the Netherlands
 Y.F. Orlov, R. Talman, Cornell University, Ithaca, NY, USA

Abstract

In the proposed electric dipole moment (EDM) experiment, with an estimated spin coherence time of 1000 s, the spin precession due to an EDM of 10^{-29} e.cm will produce a change in the vertical spin component of approximately $10 \mu\text{rad}$ during the storage time. Such high sensitivity needs an highly accurate and reliable simulation environment of the beam and spin behavior during the storage time. Therefore, several spin-related accelerator simulation programs have been considered. The paper surveys the computational algorithms of these approaches and discusses their comprehensive analysis from multiple perspectives.

INTRODUCTION

Introduced by Uhlenbeck and Goudsmit to explain the result of Stern-Gerlach experiments, spin has become a fundamental concept and plays an important role in the interactions of elementary particles. To study the various related phenomena, different experiment environments are required. For example, to study spin dependence in the interactions at the quark and gluon level, one employs collision of intense beams of polarized protons at high energy. The Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Lab provides a unique facility for this study. Here, polarized protons can be collided with 50 to 500 GeV center of mass energy. The design calls for an intensity of 2×10^{11} protons per bunch with a polarization of 70%. When the polarized beam is produced from the source, accelerated by several pre-acceleration facilities, injected and ramped in RHIC to the required energy, numerous spin resonances due to the interaction of the magnetic moment and external electromagnetic fields can deteriorate the polarization. Hence, the spin dynamics have to be understood and solutions have to be proposed to preserve the polarization during the acceleration and storage.

Recently, another quest for physics beyond the Standard Model (SM) represents a major effort in basic physics research. A non-vanishing EDM is a violation of Time-Reversal (T) and Parity (P) symmetries, and under the assumption of CPT invariance also violates the CP symmetry. Because the Electric Dipole Moment (EDM) values predicted by most extensions to the SM are many orders of magnitude larger than those of the SM itself and close to present experimental sensitivity levels, EDM experiments have become very sensitive probes for new physics, such Computer Codes (Design, Simulation, Field Calculation)

as new sources of CP violation.

A completely new approach to EDM studies is based on a charged polarized particle storage ring [1, 2]. A non-zero EDM will affect the observed spin precession, resulting in the eventual change of the polarization. This technique promises a significant sensitivity improvement, reaching down to 10^{-29} e.cm in 10^7 s of physics running time. In such a long time running, the polarization of the beam has to be maintained, which requires the spin dynamics systematic errors need to be tightly controlled.

The most general description of spin motion under the influence of external electromagnetic fields is

$$\frac{d\vec{S}}{dt} = \mu\vec{S} \times \vec{F}_\mu(\vec{B}, \vec{\beta} \times \vec{E}) + d\vec{S} \times \vec{F}_d(\vec{E}, \vec{\beta} \times \vec{B}). \quad (1)$$

Here, the spin vector \vec{S} is in the particle rest frame, \vec{B} and \vec{E} stand for the laboratory magnetic field and electric field, respectively. The first term, representing the spin precession due to the magnetic dipole moment $\mu = g\frac{e}{2mc}$, has been explored in previous accelerator experiments, for example at RHIC. The second term is the spin precession due to the electric dipole moment $d = \eta\frac{e}{2mc}$, which is proposed for study in the EDM experiment.

The design, optimization, and commissioning of modern accelerator complexes rely on dedicated beam studies based on advanced numerical approaches. Analysis of the spin motion required further development of conventional accelerator codes by augmenting positional coordinates with spin coordinates and combining the Lorentz and Thomas-BMT equations. However, such composite spin-orbital applications do not affect the basic computational framework, especially given that particle orbits are essentially independent of spin orientation (Stern-Gerlach forces are negligible at the high particle energies considered). The same numerical approaches are applicable and can be divided into two major categories: tracking and mapping. The list of successful implementations of these approaches is very long and this paper does not presume to cover all of them. Its primary goal is rather practical: to build an open simulation environment addressing the challenging EDM experiment. The correction of spin (g-2) frequency was selected as an initial benchmark application.

EQUATIONS OF MOTION

Present beam simulation programs have usually considered only the first term of Eq.(1) dealing with the spin precession due to the magnetic dipole moment. This term with

the magnetic field \vec{B} and electric fields \vec{E} defined in the laboratory frame is described by the Thomas-BMT equation [3]

$$\frac{d\vec{S}}{dt} = \vec{S} \times \vec{F}, \quad (2)$$

where

$$\vec{F} = \frac{ev_0}{p_0 c} \left[(a\gamma + 1)\vec{B} - \frac{a\gamma^2}{\gamma+1}(\vec{\beta} \cdot \vec{B})\vec{\beta} - (a\gamma + \frac{\gamma}{\gamma+1})\vec{\beta} \times \vec{E} \right]. \quad (3)$$

Here $a = (g - 2)/2$ (also called G in much of the accelerator literature) is the anomalous magnetic moment, $\vec{\beta} = \vec{v}/c$ and γ is the Lorentz factor.

In a circular accelerator, the particle motion is conveniently described relative to the trajectory of a reference particle with momentum p_0 since all machine magnets are positioned relative to this trajectory. In this coordinate system, called Frénet-Serret plane coordinate system, the Thomas-BMT equation can be rewritten as

$$\begin{aligned} \frac{dS_x}{ds} &= \left[\vec{S} \times \vec{F} \right]_x \frac{dt}{ds} + hS_z \\ \frac{dS_y}{ds} &= \left[\vec{S} \times \vec{F} \right]_y \frac{dt}{ds} \\ \frac{dS_z}{ds} &= \left[\vec{S} \times \vec{F} \right]_z \frac{dt}{ds} - hS_x \end{aligned} \quad (4)$$

where

$$\frac{dt}{ds} = \frac{1 + hx}{\frac{p_z}{p_0}} \frac{1}{v_0}. \quad (5)$$

Similarly, the Lorentz equations describing the particle behavior in an external electromagnetic field in the Frénet-Serret coordinate system are expressed in the following form:

$$\begin{aligned} \frac{d}{ds} \left(\frac{p_x}{p_0} \right) &= \frac{e}{p_0 c} \left[\frac{\vec{E}}{\beta_0} + \frac{\vec{p}}{p_0} \times \vec{B} \right]_x \cdot \left(\frac{dt}{ds} \cdot v_0 \right) + h \frac{p_z}{p_0} \\ \frac{d}{ds} \left(\frac{p_y}{p_0} \right) &= \frac{e}{p_0 c} \left[\frac{\vec{E}}{\beta_0} + \frac{\vec{p}}{p_0} \times \vec{B} \right]_y \cdot \left(\frac{dt}{ds} \cdot v_0 \right) \\ \frac{d}{ds} \left(\frac{\mathcal{E}}{p_0 c} \right) &= \left(\frac{e\vec{E}}{p_0 c} \right) \cdot \frac{\vec{p}}{p_0} \cdot \left(\frac{dt}{ds} \cdot v_0 \right) \end{aligned} \quad (6)$$

where the third component of momentum is directly calculated as

$$\frac{p_z}{p_0} = \sqrt{\frac{\mathcal{E}^2 - (m_0 c^2)^2}{p_0^2 c^2} - \left(\frac{p_x}{p_0} \right)^2 - \left(\frac{p_y}{p_0} \right)^2}. \quad (7)$$

The above equations are given in the canonical coordinates used in the MAD [4] and MAD-X program [5]. This set of coordinates is not unique and is chosen differently in different accelerator codes.

Computer Codes (Design, Simulation, Field Calculation)

TRACKING APPROACH

Introduced in the early days of the SSC, the code TEAPOT [6] had the narrowly defined purpose of investigating proton emittance growth (and lifetime reduction) due to magnetic imperfections. Symplecticity was therefore of paramount importance. Because of the difficulty of preserving symplecticity with thick elements, the code accepts only zero length elements, making it a “kick” code, or a “finite element” code, or a “symplectic integrator”.

TEAPOT, like all tracking codes, is thus a numerical differential equation solver (of the Lorentz force equation). Finite-length elements are sliced as narrowly as is required for the desired accuracy. Unlike some symplectic integrators, however, negative drift lengths are disallowed. This is consistent with the philosophy that zero length elements are “better than” finite length elements, and that the sliced accelerator, though idealized, is otherwise physical.

Conceptually the accelerator consists of a sequence of zero length elements alternating with finite element drifts. All orbits are therefore straight lines with kinks. All straight lines and kinks are calculated exactly and symplectically. One therefore calculates “exactly in an approximate accelerator”, rather than “approximately in an exact accelerator”.

All longitudinal evolution proceeds by straight line segments. A single slice of a magnet is represented by a “multipole plane” which includes both intended and unintended (error) fields. After solving for the intersection of an individual particle trajectory with this plane, the orbit kink is calculated and the new straight line determined. Particle tracking consists of iterating this process for all lattice elements and for each particle.

Unlike most beam dynamics codes, neither linear nor nonlinear transfer matrices are used for particle propagation. But transfer maps and lattice functions are calculated numerically and made available as optional output. This capability has been completely superseded by arbitrary order truncated Taylor series determination in the TEAPOT module of UAL [7]. Numerous other simulation features such as bunch generation, orbit smoothing, tune and chromaticity adjustment, and decoupling are also supported, along with even more specialized features.

By augmenting the six phase space coordinates with three spin coordinates, and solving also the BMT equation using the same finite element approach, it was relatively straightforward for S. Mane, in 1994, to extend the code to include spin tracking.

Another approach has been implemented in the spin tracking code SPINK [8]. This program was written for the RHIC project at Brookhaven National Laboratory and employed for years to study the behavior of polarized protons in all stages of the accelerator complex. SPINK uses a composite approach. Its orbit module is based on the first order matrices and the second order Transport maps produced by the MAD program. Spin propagation was implemented by the additional thin elements, spin kicks, which

rotate the spin in each accelerator element using the spin rotation matrix \mathcal{M} :

$$\mathbf{S} = \mathcal{M} \mathbf{S}_0 \quad (8)$$

This procedure has the advantage of very high computational speed using matrices that by construction represent a rotation of the spin vector by an angle $\delta\mu$ around an axis defined by its latitude θ and longitude ϕ , with respect to the reference coordinate axes.

Eqs.(4) yield three third order formally identical linear equations for the three components of the spin

$$S''' + \omega^2 S' = 0, \quad (9)$$

with

$$\omega^2 = f_x^2 + \left(f_y - \frac{1}{\rho}\right)^2 + f_z^2, \quad \vec{f} = \vec{F} \cdot \frac{dt}{ds}. \quad (10)$$

The general integral is

$$S = C_1 + C_2 \cos(\delta\mu) + C_3 \sin(\delta\mu). \quad (11)$$

with

$$\delta\mu = \omega \delta s \quad (12)$$

the spin rotation angle (spin kick) in the machine element. δs is the path length through the element along the reference orbit. The constants of integration, C_1 , C_2 , C_3 , can be found as a linear function of the initial and the final values of the spin components using S , S' , S'' and the original system. The algebra is tedious but straightforward and the resulting linear relation between spin after and before kick, expressed in the form of the 3×3 matrix \mathcal{M} , is

$$\begin{pmatrix} 1 - (B^2 + C^2)c & ABc + Cs & ACc - Bs \\ ABc - Cs & 1 - (A^2 + C^2)c & BCc + As \\ ACc + Bs & BCc - As & 1 - (A^2 + B^2)c \end{pmatrix}$$

with

$$c = 1 - \cos(\delta\mu) \quad s = \sin(\delta\mu) \quad (13)$$

and

$$A = \frac{f_x}{\omega}, \quad B = \frac{f_y - \frac{1}{\rho}}{\omega}, \quad C = \frac{f_z}{\omega} \quad (14)$$

\mathcal{M} is a parametric linear transformation, whose elements are function of the values of the particle coordinates in the laboratory, and has a determinant

$$\det(\mathcal{M}) = 1. \quad (15)$$

The matrix represents a rotation of the spin vector by an angle $\delta\mu$ around an axis defined by the two angles θ and ϕ . The coefficient of Eq.(14) can be expressed as a function of these angles, namely

$$A = \cos \theta \sin \phi, \quad B = \sin \theta, \quad C = \cos \theta \cos \phi \quad (16)$$

Computer Codes (Design, Simulation, Field Calculation)

MAPPING APPROACH

The transformation of spin-orbit variables after passage through given beam line, which could be a single pass transport channel or one turn in circular accelerator, can be represented in the form of the map

$$\begin{cases} \vec{z}_f &= \vec{F}(\vec{z}_i) \\ \vec{S}_f &= A(\vec{z}_i) \cdot \vec{S}_i \end{cases} \quad (17)$$

where \vec{z} denotes the combined vector of six orbital variables and 3×3 matrix $A(\vec{z})$ belonging to the $SO(3)$ group. Assuming that $\vec{F}(\vec{0}) = \vec{0}$ (i.e. orbital variables are deviations from reference orbit) and elements of the vector function \vec{F} and matrix A are differentiable functions, one can make an approximation of the map in Eq.(17) expanding its right hand side in Taylor series with respect to \vec{z} and truncating this expansion at the certain order.

There is a variety of ways to calculate the numerical values of the coefficients of this truncated map. A.W.Chao program SLIM, the first in this field, utilizes explicit formulas and allows the treatment of the linear expansion of the matrix A and assumes linear orbit motion [11]. The code SPINLIE also uses explicit formulas obtained by formula manipulators and extends the treatment of spin-orbit motion up to third order [12]. It is clear that the complexity of this semi analytical approach grows rapidly with the truncation order making further development difficult.

Another possible way is to use some spin-orbit tracking code with subsequent application of numerical differentiation technique. This approach has its own difficulties and will not be discussed here.

It seems that the most reliable way lies in the area of analytical manipulations with polynomials. This allows obtaining spin-orbit map in a way which is independent from the truncation order. The first program with such possibilities for spin motion was FMN [10] written in 1992 under the name VasiLIE. The computer code FMN was applied for several projects, for example, to the investigation of schemes for preserving the polarization in the TRIUMF KAON Booster, to the study of the possibilities to accelerate polarized proton beams in the Nuclotron ring in Dubna, in the HERA proton ring, and others. Later on similar possibilities were included into COSY INFINITY [13].

Both codes, FMN and COSY INFINITY, use such powerful methods as numerical integration using the differential algebra techniques and the direct summing of Lie exponent series for s -independent elements. Note that Taylor map coefficients obtained by these methods accumulate two types of errors: computer rounding errors and errors defined by the size of the integration step (or by the number of terms in the Lie series taken for summation). There is an approach free from the defects of the second type in which the number of the steps does not depend on a precision (like in the Gauss method of matrix inversion, where the number of operations depends on the matrix dimension only), which is also used in the FMN code [14].

UAL FRAMEWORK

The Unified Accelerator Libraries (UAL) attempt to "manage the complexity" of accelerators by providing an environment for simulating a variety of properties of a variety of accelerators using a variety of simulation codes and methods. The intended value of the environment is to provide homogeneous access to these resources while masking their diversity yet assuring their consistency. This allows different methods to be consistently applied to the same accelerator and the same methods to be applied to different accelerators.

To facilitate such unification UAL has introduced an open architecture in which diverse accelerator codes are connected together via common accelerator objects such as *Element*, *Bunch*, *Twiss*, etc. In this architecture each accelerator code is implemented as an object-oriented library of C++ classes. There is a very natural identification of physical elements, such as magnets, with their representation by computer objects. UAL supports considerable flexibility in the attributes of all objects - certainly enough that all attributes of objects contained in modules included so far have been describable without constraint. Such flexibility has made it practical to evaluate, compare, and integrate a variety of design models and to build heterogeneous, project-specific applications.

The accumulated experience has motivated the development of the *Element-Algorithm-Probe* framework [15], a uniform mechanism for combining diverse modules to simulate complex combinations of the physical effects and dynamic processes. *Elements* are things like bending magnets, RF cavities, collimators, and so on. *Algorithms* are mathematical formulas capable of evolving quantities known at element inputs to their corresponding values at element outputs. So both element and algorithm are terms of common usage, likely to be understood unambiguously by all workers in the field. The term *probe*, because it is less standard, requires more explanation. A similar term is *observable*. As used in UAL a particle or beam bunch probes the lattice elements. Other things also probe the lattice. For example, the transfer map (identity at the origin) evolves into the transfer map from origin to element output as it evolves through an element. So a probe is anything whatsoever for which continuous evolution is meaningful and the evolution is unambiguously caused by the elements making up the lattice. It therefore makes sense to evaluate the evolution of a probe caused by the lattice. Examples of probes are 6D particle coordinates of all particles in a bunch, spin components, moments of a bunch of particles, lattice functions such as Twiss functions and dispersion functions, transfer matrices (i.e. linear maps), nonlinear maps (which are represented by truncated power series), wake fields, and so on.

Several years ago, The BNL team extended UAL with the space charge module SIMBAD [16] as a part of the European Union contract. At this time, addressing the special requirements of the EDM experiment, UAL accom-

modated another BNL code, SPINK, described in the previous section on tracking approaches. This development solved three major tasks. First, UAL allowed the upgrade of the original SPINK version with the TEAPOT symplectic tracking engine and provided access to other modules, for example, the differential algebra-based integrator and others. Second, the SPINK Fortran code has been refactored and transformed into a few dedicated C++ classes. As a result, it added flexibility for developing new extensions, such as new types of elements and enhancing the accuracy of the original approach with the slicing technique. Finally, the SPINK algorithm became a part of the simulation infrastructure which can be applied to other projects.

(G-2) EXPERIMENT

In the studies of spin motion, the (g-2) frequency is defined as the frequency of the spin precession relative to the momentum vector, describing the rate of change of the component of spin \vec{S} parallel to the velocity. For a particle with momentum p_0 moving in a plane perpendicular to the magnetic field \vec{B} , the (g-2) frequency is given by Thomas-BMT equation as

$$\omega_0 = \frac{eB}{p_0 c} \gamma v_0 \cdot \frac{g-2}{2}, \quad (18)$$

where B is the magnitude of the field. If the orbit is not exactly perpendicular to the magnetic field, a small correction to the (g-2) frequency will appear. This pitch correction has been addressed and examined in a few papers [17, 18] when the pitch angle (the angle with respect to the plane perpendicular to the magnetic field) is varying due to axial focusing force, presented by an almost uniform and only particle linear position dependent magnetic field.

As part of a calibration effort of particle tracking codes for the storage ring of electric dipole moment (EDM) experiments (proton and deuteron), a study of the pitch correction has been carried out by simulating a circulating particle in a continuous ring with weak magnetic focusing. Given the field focusing index $n = -[\frac{\rho_0}{B_0} \frac{\partial B_y}{\partial x}]$, the pitch correction is extended [19] because of the quadratic dependence on y^2 in the curvilinear coordinate system. In the limit of fast pitch change ($\omega_y \gg \omega_0$, with ω_y the frequency of the vertical betatron oscillation), the corrected (g-2) frequency is given as

$$\omega_a = \omega_0 \left[1 - \frac{a}{2} \psi_0^2 + \frac{n}{2(1-n)} \psi_0^2 \right]. \quad (19)$$

Here, ω_0 is the uncorrected (g-2) frequency given in Eq.(18), $a = (g-2)/2$ and ψ_0 is the amplitude of the pitch angle with respect to the xz -plane perpendicular to the y direction. This analytical formula has been confirmed quantitatively by the integration of the differential equations of motion using the subroutine TDHPCG [20] based on the fourth order corrector-predictor method of Runge-Kutta.

We applied this result to a simple muon ring defined by the parameters in table 1. For n equals 0.13, $\omega_y/\omega_0 \approx 225$

Table 1: Testing muon particle and ring parameters.

Muon momentum p (GeV/c)	0.1
Focusing Index n	0.13
Radius ρ (m)	5
Anomalous magnetic moment a	0.00116592
Initial pitch angle ψ_0 (mrad)	1
Initial spin vector Sx, Sy, Sz	0, 0, 1

close to the limit of fast pitch change. The results from the analytical formula Eq.(19) and the UAL simulation are listed in table 2. The number of slices each element broken up into in UAL is listed in the first column. The last column shows the CPU time of per particle per turn.

Table 2: List of pitch correction from analytical formula and simulation of UAL.

Approach	$\frac{\omega_a - \omega_0}{\omega_0} (\times 10^{-8})$	CPU time (ms)
Analytical formula	7.42	
UAL, slices = 8	6.48 ± 0.01	0.86
UAL, slices = 32	7.34 ± 0.02	2.97
UAL, slices = 64	7.38 ± 0.02	5.81
UAL, slices = 128	7.39 ± 0.02	10.66
UAL, slices = 256	7.40 ± 0.02	21.72

SUMMARY

This paper has given a brief summary of the most prominent sorts of accelerator simulation codes. The motivation for this is the need for extraordinarily precise spin tracking calculations required for interpreting high precision experimental measurements of magnetic and, especially, electric dipole moments. The presentation has attempted to be neutral as regards physical methods, but partial as regards establishing an environment in which diverse methods of calculation can be applied to multiple physical effects.

Many component errors and imperfections are capable of producing behavior that can be misinterpreted as being due to the particular dipole moment being measured. Examples are fringe fields, errors, nonlinearities, etc. In this report just one (particularly simple) such effect has been investigated and compared with an analytic calculation. This has been intended only as a prototype of the sorts of calculations and comparisons that will be required. This initial step needs to be extended in many directions: new benchmarks, new programs, especially from the mapping category (like FMN and COSY), and new extensions encompassing all relevant physical effects.

ACKNOWLEDGEMENT

One of the authors (N.Malitsky) would like to thank V.Balandin, N.Golubeva for their invaluable input.

Computer Codes (Design, Simulation, Field Calculation)

REFERENCES

- [1] F.J.M.Farley et al. Phys.Rev.Lett. 93, 052001 (2004).
- [2] "AGS Proposal: Search for a permanent electric dipole moment of the deuteron nucleus at the 10^{29} e.cm level." Spokesperson Y.K.Semertzidis (2008)
- [3] L.H.Thomas, Phil. Mag. 3, 1 (1927) ; V.Bargmann, L.Michel, V.L.Telegdi, Phys. Rev. Lett. 2, 435 (1956).
- [4] H.Grote and F.Iselin, The MAD Program (Methodical Accelerator Design), User's Reference Manual, CERN/SL/SO-13(A:P) (1990).
- [5] P.K.Skowronski, F.Schmidt, E.Forest. "Advances in MAD-X using PTC", Proceedings of PAC07, p.3381, Albuquerque, New Mexico, USA (2007).
- [6] L.Schachinger and R. Talman, "TEAPOT: A Thin-Element Accelerator Program For Optics and Tracking", Particle Accelerators, Vol.22, pp.35-56 (1987).
- [7] N.Malitsky and R.Talman, Unified Accelerator Libraries. AIP 391 <http://code.google.com/p/ual> (1996).
- [8] A.U.Luccio, Proceedings of the Adriatico Research Conference on Trends in Collider Spin Physics p.235 Trieste, Italy (1995).
- [9] V.Balandin and N.Golubeva, *Hamiltonian Methods for the Study of Polarized Proton Beam Dynamics in Accelerators and Storage Rings*, DESY 98-016, February 1998.
- [10] V.Balandin and N.Golubeva, *Nonlinear Spin Dynamics*, Proc. XV Int. Conf. High Energy Accelerators, Hamburg, 1992, Int. J. Mod. Phys. A, **2B**, 998, (1992).
- [11] A.W.Chao, Nucl. Instr. Meth., **180** (1981) 29.
- [12] Yu. Eidelman and V. Yakimenko, Proc 1991 IEEE Part. Accel. Conf., 1991, San-Francisco.
- [13] V.Balandin, M.Berz and N.Golubeva, *Computation and Analysis of Spin Dynamics*, Computational Accelerator Physics Conference, Williamsburg, Virginia, USA, (1996).
- [14] V.Balandin, *One of Methods to Extract Truncated Taylor Map for Orbital and Spin Motion in Proton Storage Rings*, Particle Accelerator Conference, Washington, USA, (1993).
- [15] N.Malitsky and R.Talman, Framework of Unified Accelerator Libraries. ICAP'98 (1998).
- [16] N.L.D'Imperio, A.U.Luccio, N.Malitsky, O.Boine-Frankenheim, Parallel 3-D Space Charge Calculations in the Unified Accelerator Library. EPAC'04 (2004).
- [17] S.Granger, G.W.Ford, "Electron Spin Motion in a Magnetic Mirror Trap", Phys. Rev. Lett. 28, 1479 (1972).
- [18] F.J.N.Farley, "Pitch Correction in (g-2) Experiments", Phys. Lett. 42B, 1, 66 (1972).
- [19] F.Lin, etc "Extended Pitch Correction with Weak Magnetic Focusing", EDM internal note, (2009)
- [20] This program is the differential equation solver-DHPCG which uses the fourth order corrector-predictor method. This is a modification of subroutine DHPCG from the scientific subroutine package (SSP) published by IBM.

AN EFFICIENT 3D SPACE CHARGE ROUTINE WITH SELF-ADAPTIVE DISCRETIZATION *

G. Pöplau[†], U. van Rienen, Rostock University, Rostock, Germany

Abstract

Precise and fast 3D space charge calculations for bunches of charged particles are still of growing importance in recent accelerator designs. A widespread approach is the particle-mesh method computing the potential of a bunch in the rest frame by means of Poisson's equation. Whereas an adaptive discretization of a bunch is often required for efficient space charge calculations in practice, such a technique is not implemented in many computer codes.

In this paper we present a new approach to an adaptive discretization which is based on the multigrid technique. The goal is that the error estimator needed for the adaptive distribution of mesh lines can be calculated directly from the multigrid procedure. The algorithm was implemented in the software package MOEVE and investigated for several particle distributions. It turns out that the adaptive discretization technique performs very efficiently.

INTRODUCTION

The simulation of the dynamics of high-brightness charged particle bunches demand the fast calculation of 3D non-linear space charge fields with an accuracy that matches the quality of the bunch. The particle-mesh method is a widespread model for space charge calculations. Here, adaptive discretization techniques are often required in order to satisfy both computational demands: accuracy and fast performance. Nevertheless, adaptive discretizations are implemented only in a few software packages together with space charge calculations. For instance, the FFT Poisson solver that is often applied allows only an equidistant mesh. An adaptive discretization following the particle density distribution is implemented in the GPT tracking code (General Particle Tracer, Pulsar Physics) together with a multigrid Poisson solver of the software package MOEVE (Multigrid for non-equidistant grids to solve Poisson's equation) [5, 11]. The disadvantage of this approach is that it does not provide a hierarchical construction of meshes which could be used directly by the multigrid algorithm.

In this paper we present a new approach to an adaptive discretization which is based on the multigrid technique. The goal is that the error estimator needed for the adaptive distribution of mesh lines can be calculated directly from the multigrid procedure. The algorithm has been implemented within the framework of the software package MOEVE. It will be investigated for several particle distributions among them a particle distribution which occurred during simulations for the European XFEL [1].

The space charge model we consider here is the particle-mesh method. The distribution of particles in a bunch is modeled as distribution of macro particles. Assuming that the energy of the macro particles is within the same range the space charge field is calculated in the rest frame of the bunch. After the transformation into the rest frame a mesh is constructed around the particles of the bunch and the charge of the particles is assigned to the mesh points. Now, the potential φ can be obtained from Poisson's equation given by

3D SPACE CHARGE MODEL

where ϱ the space charge distribution, ε_0 the dielectric constant and r the distance between the centre of the bunch and the boundary. Usually, the domain Ω is a rectangular box constructed around the bunch. On the surface $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$ ($\partial\Omega_1 \cap \partial\Omega_2 = \emptyset$) perfectly conducting boundaries ($\partial\Omega_1$) or open boundaries ($\partial\Omega_2$) can be applied. For space charge calculations within a beam pipe the domain Ω is assumed to be a cylinder with elliptical cross section. A detailed description of the 3D space charge model can be found in [8] and the model with elliptical shaped beam pipe in [4], respectively.

For the solution of the Poisson equation we applied the discretization with second order finite differences. This leads to a linear system of equations of the form

$$L_h u_h = f_h, \quad (2)$$

where u_h denotes the vector of the unknown values of the potential and f_h the vector of the given space charge density at the grid points. The step size h indicates a certain refinement level and the operator L_h is the discretization of the Laplacian.

THE POISSON SOLVERS OF MOEVE

The software package MOEVE has been developed for space charge calculations. It involves several iterative Poisson solvers among them the state-of-the-art multigrid Poisson solvers MG (multigrid) and MG-PCG (multigrid preconditioned conjugate gradients). These algorithms provide optimal convergence, i. e. the number of iteration steps

* Work supported by DFG under contract number RI 814/18-1

[†] gisela.poeplau@uni-rostock.de

needed to obtain a certain accuracy is independent of the step size h .

The multigrid algorithm operates on a certain number of grids starting with the mesh given by the discretization of Poisson's equation which is related to equation (2). This mesh is referred to as the fine grid. Then a sequence of coarser grids is generated by cutting mesh lines. On an equidistant mesh where the number of mesh lines equals $N = 2^t + 1$, every second mesh line can be removed. Of course this process can be also performed vice versa, i. e. we start with a coarse mesh and refine the mesh (adaptively) until a certain fine level is achieved. More details of the Poisson solvers of MOEVE can be found elsewhere, for instance in [5, 6].

Since the bunch is located only in a certain part of Ω , an adaptive discretization strategy is required for an efficient solution method. Successfully applied for space charge calculations, one possible method is the adaptive distribution of mesh lines according to the distribution of the particles inside the bunch as implemented in the tracking code GPT [10]. However, the main drawback of this approach is that the mesh has no natural relation to a multigrid hierarchy. Together with the multigrid Poisson solver of MOEVE, a special coarsening strategy has been developed for these adaptive meshes in order to achieve optimal multigrid convergence [9].

SELF-ADAPTIVE MULTIGRID

The generation of adaptive meshes providing both an appropriate approximation of the bunch and a hierarchy of meshes for the multigrid Poisson solver is an important task for the development of efficient 3D space charge algorithms. The grid refinement should be self-adaptive, i. e. the grid refinement is carried out dynamically during the solution process. It is controlled by some refinement criterion.

A common criterion which will be tested in this paper is the τ -criterion [12]. Before this criterion can be defined some notations are necessary. The step sizes h and $2h$ refer to the step sizes on the fine and the next coarser grid (usually with double mesh size), respectively. The operators I_h^{2h} and \hat{I}_h^{2h} denote different restriction operators. For the numerical tests of the next section the injection was chosen for \hat{I}_h^{2h} and the full weighting restriction for I_h^{2h} . The τ -criterion is based on the so-called $(h, 2h)$ relative truncation error τ_h^{2h} with respect to the restriction operators I_h^{2h} and \hat{I}_h^{2h} . It is defined by

$$\tau_h^{2h} := L_{2h} \hat{I}_h^{2h} u_h - I_h^{2h} L_h u_h. \quad (3)$$

The values of τ for the "XFEL bunch" (see section on numerical investigations below) are represented in Figure 1.

By means of the refinement criterion a hierarchy of locally refined grids can be generated. The self-adaptive multigrid scheme is given as follows:

Computer Codes (Design, Simulation, Field Calculation)

Algorithm: Self-Adaptive Multigrid

1. Start on a relatively coarse mesh.
2. Perform a few multigrid cycles on equation (2).
3. Calculate τ_h^{2h} .
4. Add mesh lines locally, where $|\tau_h^{2h}| > \varepsilon$.
5. Proceed from 2) as long as $|\tau_h^{2h}| > \varepsilon$.

The algorithm is constructed such that the mesh is refined at locations where $|\tau_h^{2h}| > \varepsilon$. Hence, the structure of the resulting mesh corresponds to the tensor product mesh implemented in MOEVE. This strategy allows the direct application of the MOEVE MG Poisson solvers on the refined mesh. This algorithm was first given in [6].

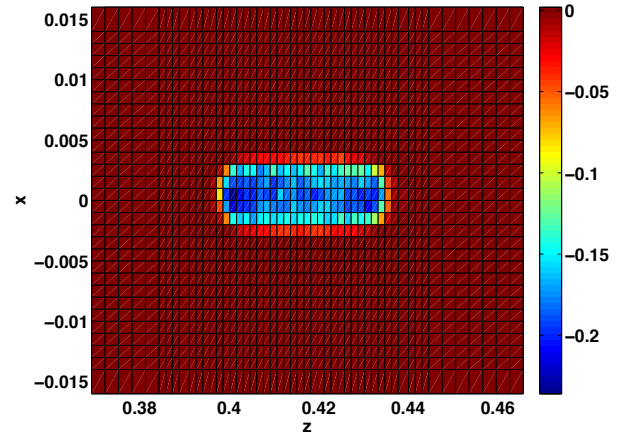


Figure 1: Values of τ for the XFEL bunch. The values are shown in the (x, z) -plane at $x=0.0$ m.

NUMERICAL INVESTIGATIONS

In this section the performance of the adaptive discretization strategy is investigated. We applied both MG (multigrid) and MG-PCG (multigrid preconditioned conjugate gradients) as multigrid Poisson solvers implemented in MOEVE. Although MG-PCG takes more time for one single step because of additional calculations it is often more stable in real life applications. The results for the adaptively refined grid were compared to those obtained with an equivalent equidistant grid. All Poisson solvers were applied until the relative residual had achieved a value of less than 10^{-2} . This value seems to be rather low, but further iterations would not improve the numerical error, because the space charge densities of the following particle distributions are discontinues.

Cylindrical Particle Distribution

First, the refinement algorithm was applied to a cylindrical particle distribution with 100,000 macro-particles and

$N_x \times N_y \times N_z$	MG steps	CPU time
$17 \times 17 \times 33$	2	0.05 s
$33 \times 33 \times 65$	3	0.16 s
$65 \times 65 \times 129$	4	1.36 s
$129 \times 129 \times 257$	4	10.80 s
$257 \times 257 \times 513$	4	88.90 s
$N_x \times N_y \times N_z$	MG-PCG steps	CPU time
$17 \times 17 \times 33$	2	0.05 s
$33 \times 33 \times 65$	4	0.20 s
$65 \times 65 \times 129$	4	1.38 s
$129 \times 129 \times 257$	4	11.10 s
$257 \times 257 \times 513$	4	92.2 s

Table 1: Performance of MG and MG-PCG for a cylindrical particle distribution on equidistant grids.

a radius of $R = 2$ mm and a length of 17.3 mm. Such a cigar-like bunch shape was chosen because these long distributions usually pose a problem to the Poisson solvers [7]. The macro particles of the bunch were uniformly distributed. The bunch had a total charge of $Q = -1$ nC. The computational domain Ω was constructed as $\Omega = [-0.0085, 0.0085] \times [-0.0085, 0.0085] \times [-0.017, 0.017]$ and open boundary conditions were applied. For the adaptive algorithm we started with an equidistant mesh i. e. the step size at the coarsest level is approximately equal for all coordinate directions. This is not necessary in general but for long bunches it improves the performance.

In the following the performance of the straight-forward multigrid technique on equidistant grids was compared to the performance on the adaptive grid. The CPU times include all calculations for space charge i. e. also the assignment of the particle's charge to the grid points.

In Table 1 the CPU time is given for calculations with the MG and MG-PCG solver on equidistant grids. The results approve the linear behaviour of the MG and MG-PCG algorithms.

Table 2 represents the results for the self-adaptive multigrid procedure. The algorithm starts with a grid of $17 \times 17 \times 33$ mesh points. Then the mesh will be refined successively until all values of τ are less than 10^{-2} . The refinement of each level is given in Table 2. The smallest step size of each level coincides with the step size of the corresponding level of the equidistant discretization. Since the adaptive discretization requires only a fraction of the mesh points of the equidistant grid the performance time for MG-PCG reduces to a quarter of the time required for the same refinement level on an equidistant grid.

XFEL bunch

The particle distribution in this section was taken from simulations for the European XFEL [1]. The bunch had a total charge of -1 nC. At the cathode it was started with a radius of 3 mm and the laser was modeled as a flat top shape with 20 ps full width half maximum and a rise and fall time of 2 ps. The bunch was tracked through the rf gun Computer Codes (Design, Simulation, Field Calculation)

$N_x \times N_y \times N_z$	MG steps	CPU time
$17 \times 17 \times 33$	2	0.04 s
$33 \times 33 \times 59$	3	0.20 s
$47 \times 47 \times 97$	4	0.87 s
$65 \times 65 \times 163$	7	4.36 s
$95 \times 95 \times 283$	18	41.10 s
$N_x \times N_y \times N_z$	MG-PCG steps	CPU time
$17 \times 17 \times 33$	2	0.04 s
$33 \times 33 \times 59$	4	0.23 s
$47 \times 47 \times 97$	4	0.93 s
$65 \times 65 \times 163$	6	4.01 s
$85 \times 85 \times 283$	8	21.80 s

Table 2: Performance of MG and MG-PCG for a cylindrical particle distribution on the adaptively constructed mesh.

$N_x \times N_y \times N_z$	MG steps	CPU time
$33 \times 33 \times 65$	3	0.15 s
$65 \times 65 \times 129$	3	1.4 s
$129 \times 129 \times 257$	3	3.19 s
$257 \times 257 \times 513$	2	51.20 s
$N_x \times N_y \times N_z$	MG-PCG steps	CPU time
$33 \times 33 \times 65$	4	0.12 s
$65 \times 65 \times 129$	4	1.34 s
$129 \times 129 \times 257$	4	10.70 s
$257 \times 257 \times 513$	2	53.00 s

Table 3: Performance of MG and MG-PCG for the simulated XFEL bunch at 0.07 m on an equidistant mesh.

with the tracking code ASTRA [3] with the settings given in the technical desing report [2].

For this numerical test the particle distribution at the longitudinal position of 0.07 m after the cathode was chosen. Here, the bunch had a radius of 3.0 mm, a length of 8.0 mm and an energy of 2.5 MeV. For the simulation the bunch contained 100,000 macro particles. The bounding box in the rest frame was constructed as $\Omega = [-0.016, 0.016] \times [-0.016, 0.016] \times [0.037, 0.047]$ and Dirichlet boundary conditions were applied.

Table 3 and 4 present the CPU time for the equidistant and for the adaptively constructed grids, respectively. It turns out that the computational effort can be reduced enormously by the application of the adaptive discretization strategy. Considering the finest discretization the CPU time of the MG method was reduced nearly to a tenth on the adaptive mesh.

Figure 2 and 3 show the adaptive discretization with $45 \times 45 \times 85$ grid points.

CONCLUSIONS

In this paper we constructed an adaptive discretization technique for space charge calculations which is based on the τ -criterion. By means of this approach a hierarchy of grids can be generated which can be directly used by the multigrid scheme of the software package MOEVE. The

$N_x \times N_y \times N_z$	MG steps	CPU time
$31 \times 31 \times 53$	2	0.06 s
$45 \times 45 \times 85$	3	0.52 s
$59 \times 59 \times 141$	2	2.00 s
$83 \times 83 \times 243$	3	5.96 s
$N_x \times N_y \times N_z$	MG-PCG steps	CPU time
$31 \times 31 \times 53$	2	0.14 s
$45 \times 45 \times 85$	4	0.65 s
$59 \times 59 \times 141$	6	2.16 s
$83 \times 83 \times 243$	6	7.96 s

Table 4: Performance of MG and MG-PCG the XFEL bunch at 0.07 m on the adaptively constructed mesh.

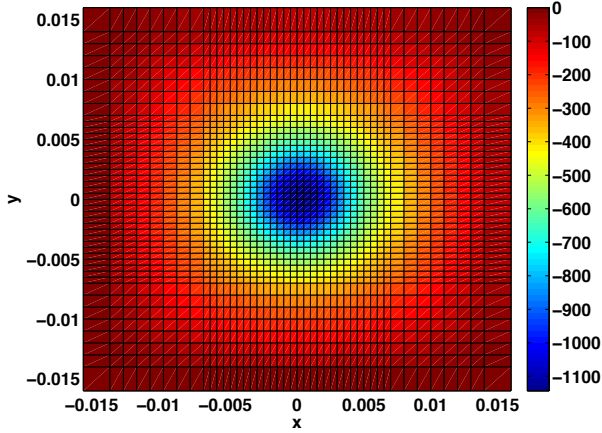


Figure 2: Adaptive discretization (transversal) for the XFEL bunch. The potential in the rest frame is plotted in the transversal plane through the longitudinal centre of the bunch.

implementation into the framework of MOEVE allowed a relatively simple approach. The numerical tests showed that the adaptive discretization reduced the number of mesh points enormously. Hence, the space charge calculations were performed very efficiently.

REFERENCES

- [1] M. Clemens, M. Dohlus, S. Lange, G. Pöplau, T. Limberg, and U. van Rienen. Microbunch amplification in the european XFEL. TESLA-FEL-2009-02, DESY, Hamburg (flash.desy.de/reports_publications/tesla_fel_reports/tesla_fel_2009/), 2009.
- [2] M. Altarelli et al., editor. *The European X-Ray Free-Electron Laser Technical design report*. DESY 2006-097, 2007.
- [3] K. Flöttmann. *ASTRA*. DESY, Hamburg, www.desy.de/~mpyflo, 2000.
- [4] A. Marković. Numerical computation of space-charge fields of electron bunches in a beam pipe of elliptical shape. TESLA-Report 2005-21, DESY, Hamburg (tesla.desy.de/new_pages/TESLA/TTFnot05.html), 2005.
- [5] G. Pöplau. *MOEVE 2.0: Multigrid Poisson Solver for Non-Equidistant Tensor Product Meshes*. Universität Rostock, 2006.
- [6] G. Pöplau and U. van Rienen. A self-adaptive multigrid technique for 3D space charge calculations. *IEEE Transactions on Magnetics*, 44(6):1242–1245, 2008.
- [7] G. Pöplau, U. van Rienen, and K. Flöttmann. 3D space charge routines: The software package MOEVE and FFT Poisson solvers. In H. Fukuma, E.S. Kim, and K. Ohmi, editors, *Proceedings of the International Workshop on Electron-Cloud Effects (ELOUD'07)*, Daegu, Korea, 9-12 April 2007, KEK Proceedings 2007-10, pages 181–188, 2007.
- [8] G. Pöplau, U. van Rienen, S.B. van der Geer, and M.J. de Loos. Fast calculation of space charge in beam line tracking by multigrid techniques. In W.H.A. Schilders, E.J.W. ter Marten, and S.H.M. Houben, editors, *Scientific Computing in Electrical Engineering*, number 4 in Mathematics in Industry, pages 329–336, Berlin, 2004. Springer-Verlag.
- [9] G. Pöplau, U. van Rienen, S.B. van der Geer, and M.J. de Loos. Multigrid algorithms for the fast calculation of space-charge effects in accelerator design. *IEEE Transactions on Magnetics*, 40(2):714–717, 2004.
- [10] G. Pöplau, U. van Rienen, S.B. van der Geer, and M.J. de Loos. A multigrid based 3D space-charge routine in the tracking code GPT. In M. Berz and K. Makino, editors, *Computational Accelerator Physics 2002 (Proceedings of the 7th International Computational Accelerator Physics Conference, East Lansing, Michigan, USA)*, number 175 in Institute of Physics Conference Series, pages 281–288, Bristol and Philadelphia, 2005. Institute of Physics Publishing.
- [11] Pulsar Physics, Burghstraat 47, 5614 BC Eindhoven, The Netherlands, www.pulsar.nl/gpt. *General Particle Tracer (GPT)*, release 2.70 edition, 2004.
- [12] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, San Diego, 2001.

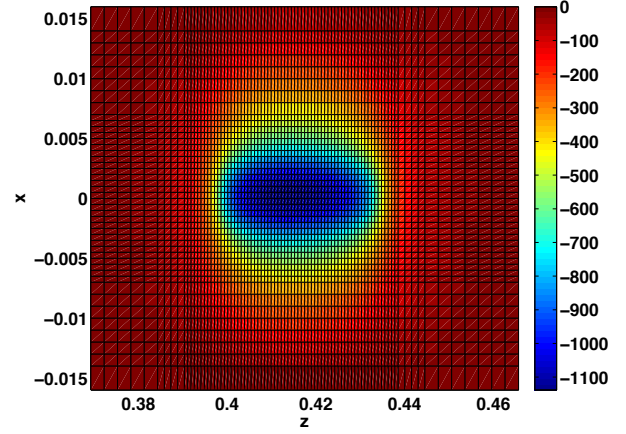


Figure 3: Adaptive discretization (longitudinal) for the XFEL bunch. The potential in the rest frame is plotted in the (x,z) plane at $y=0.0$ m.

IMPEDANCE ESTIMATION BY PARABOLIC PARTIAL DIFFERENTIAL EQUATION FOR RECTANGULAR TAPER*

N. Okuda, The University of Tokyo, Tokyo, Japan
K. Yokoya, KEK, Tsukuba, Japan

Abstract

The mesh calculation based on the paraxial approximation can be much faster than ordinary methods when the bunch is very short. There are two reasons. One is to be able to choose the longitudinal mesh size independent of the bunch length. The other is that the problem can be solved as an initial-value problem in spite of frequency domain calculation.

However, the accuracy of the results by the approximation is not clear generally. It will be shown that the approximation is valid for rectangular tapered chamber in some frequency range.

INTRODUCTION

Recently, the calculation of wake field and the impedance has become more important because new accelerators require high current and much required fineness. In many cases they are usually calculated numerically by simulation using a mesh.

There are many methods of mesh calculation. The finite-difference time domain (FDTD) [1] and the finite integration technique (FIT) [2] are popular.

The mesh computation based on the paraxial approximation [3] can be much faster than ordinary methods if the bunch length is very short. The approximation has been used in geometrical optics. Since several years ago, it has also been used for beam field. The calculation of Coherent Synchrotron Radiation (CSR) by paraxial approximation gives good results[4][5]. In Ref.[3], the analytical solution of geometric wake impedances by paraxial approximation is shown. They are that for axisymmetric geometry.

In these proceedings, numerical 3D calculation will be shown. The vertical impedance for rectangular tapered chamber is computed. It agrees with the analytic solution in the appropriate frequency.

The smaller angle taper should be better because the wave at small angle is dominant.

We will focus only on short range wake and completely conducting wall in this proceeding. Resistive wall is not considered. MKSA unit is used in these proceedings.

IMPEDANCE

In these proceedings, special transformed fields for arbitrary function $\tilde{f}(x, y, z, t)$ are defined by

$$\tilde{f}(x, y, z, t) = \frac{1}{2\pi} \int \hat{f}(x, y, z, k) e^{-ik(ct-z)} dk. \quad (1)$$

Tilde means original value and ‘grave’ (\hat{f}) means the transformed value satisfying the equation above. This is like Fourier transform. However, the factor e^{ikz} has to be noticed.

The wake effect is able to be represented by impedance. Conventional vertical impedance is

$$Z_y = \frac{-i}{cqy_s} \int_{-\infty}^{\infty} dz \left[\hat{E}_y + c\hat{B}_x \right]_{\mathbf{r}_w = \mathbf{r}_s = (0, y_s)}, \quad (2)$$

where q is a charge, $\mathbf{r}_s = (x_s, y_s)$ is an offset of source particle, $\mathbf{r}_w = (x_w, y_w)$ is an offset of witness particle, y_s is small, and \hat{E}_y, \hat{B}_x are the special transformed fields defined by Eq.(1). We will omit grave ($\hat{f} \rightarrow f$) from now on.

PARAXIAL APPROXIMATION

In this study, paraxial approximation is used. It is valid if the wave propagates at small angle θ from z axis, as shown in Figure 1.

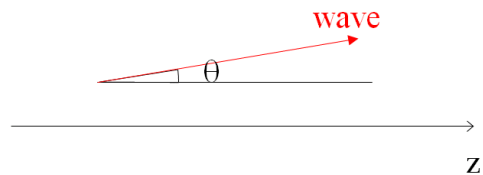


Figure 1: ‘Paraxial’ means the wave angle θ is small.

We will consider $c\tau \ll g$, where $\tau = (z_s - z_w)/c$ and g is the transverse minimum distance from beam axis. For rectangular chamber whose height is smaller than width, g is the smallest half height. In this range, backward or large angle propagating waves cannot take effect. Therefore, $\theta \ll 1$ can be assumed. Suppose $c\tau (\ll g)$ is bunch length, the waves of $\theta \geq 1$ which generated by a bunch don’t catch up with the same bunch. When considering wake effect for bunch itself, paraxial approximation is valid for very short bunch length.

* Work supported by KEK and by Global COE Program “the Physical Sciences Frontier”, MEXT, Japan.

We divide electromagnetic fields into 2 terms as

$$\mathbf{E} = \mathbf{E}^{(b)} + \mathbf{\mathcal{E}}, \quad (3)$$

$$\mathbf{B} = \mathbf{B}^{(b)} + \mathbf{\mathcal{B}}, \quad (4)$$

where \mathbf{E} , \mathbf{B} is net electric, magnetic field, and $\mathbf{E}^{(b)}$, $\mathbf{B}^{(b)}$ is vacuum fields, which is the solution of no wall. $\mathbf{\mathcal{E}}$, $\mathbf{\mathcal{B}}$ defined by above equation is the radiated fields.

The vacuum fields are found by solving Maxwell equation with charge density

$$\tilde{\rho}(\mathbf{r}, z, t) = q\delta(t - z/c)\delta(\mathbf{r} - \mathbf{r}_s). \quad (5)$$

Therefore

$$\tilde{\mathbf{E}}^{(b)}(\mathbf{r}, z, t) = \frac{q\delta(t - z/c)}{2\pi\epsilon_0|\mathbf{r} - \mathbf{r}_s|^2}(\mathbf{r} - \mathbf{r}_s). \quad (6)$$

$$\tilde{\mathbf{B}}^{(b)} = \frac{1}{c}\mathbf{e}_z \times \tilde{\mathbf{E}}^{(b)} \quad (7)$$

From Eq(1),

$$\mathbf{E}^{(b)}(\mathbf{r}, z, k) = \frac{cq}{2\pi\epsilon_0|\mathbf{r} - \mathbf{r}_s|^2}(\mathbf{r} - \mathbf{r}_s) \quad (8)$$

$$\mathbf{B}^{(b)} = \frac{1}{c}\mathbf{e}_z \times \mathbf{E}^{(b)} \quad (9)$$

Accordingly, Maxwell equation with respect to $\mathbf{\mathcal{E}}$ is

$$(\nabla_{\perp}^2 + 2ik\partial_z + \partial_z^2)\mathbf{\mathcal{E}} = 0. \quad (10)$$

The source term is cancelled.

We will show 3rd term of Eq.(10) is negligible. In time domain, $\tilde{\mathcal{E}}$ is superposition of the plain wave

$$\tilde{\mathcal{E}}_{\eta,k}(x, y, z, t) = \hat{\mathcal{E}}(\eta, k)e^{-ik(ct - \eta_x x - \eta_y y - \eta_z z)}, \quad (11)$$

where $\|\eta\| = 1$, $\hat{\mathcal{E}}$ is a function of (η, k) . η can be represented by

$$\eta = \mathbf{e}_x \sin \theta \cos \phi + \mathbf{e}_y \sin \theta \sin \phi + \mathbf{e}_z \cos \theta. \quad (12)$$

θ is the angle of wave propagating from z axis. Neglecting the order of θ ,

$$\eta \approx \mathbf{e}_x \theta \cos \phi + \mathbf{e}_y \theta \sin \phi + \mathbf{e}_z(1 - \theta^2/2). \quad (13)$$

Therefore,

$$\begin{aligned} & \tilde{\mathcal{E}}_{\eta,k}(x, y, z, t) \\ &= \hat{\mathcal{E}}e^{-ik(-x\theta \cos \phi - y\theta \sin \phi + z\theta^2/2)}e^{-ik(ct - z)}. \end{aligned} \quad (14)$$

\mathcal{E} is the superposition of

$$\mathcal{E}_{\eta}(x, y, z, k) = \hat{\mathcal{E}}e^{ik(x\theta \cos \phi + y\theta \sin \phi - z\theta^2/2)}. \quad (15)$$

If θ^3 is not omitted but $O()$ is used,

$$\begin{aligned} & \mathcal{E}_{\eta}(x, y, z, k) \\ &= \hat{\mathcal{E}}e^{ik[(\theta + O(\theta^3))(\cos \phi x + \sin \phi y) - (\theta^2/2 + O(\theta^4))z]} \end{aligned} \quad (16)$$

Computer Codes (Design, Simulation, Field Calculation)

Substituting it for \mathcal{E} in Eq.(10), it is found that the first and second term are order of $k^2\theta^2$ and the third term is order of $k^2\theta^4$. Accordingly, it is negligible. Eq.(10) is approximately the following equation,

$$[\nabla_{\perp}^2 + 2ik\partial_z]\mathcal{E} = 0. \quad (17)$$

This is parabolic equation.

The parabolic equation and boundary conditions give transverse electric fields. From them, we obtain the other fields as

$$\mathbf{\mathcal{B}}_{\perp} \approx \frac{1}{c}\mathbf{e}_z \times \mathbf{\mathcal{E}}_{\perp}, \quad (18)$$

$$\mathcal{E}_z = \frac{ci}{k}\mathbf{e}_z \cdot (\nabla_{\perp} \times \mathbf{\mathcal{B}}_{\perp}), \quad (19)$$

$$\mathcal{B}_z = -\frac{i}{ck}\nabla_{\perp} \times \mathbf{\mathcal{E}}_{\perp}. \quad (20)$$

The impedances are calculated from these fields.

Paraxial approximation has two advantages. One is to be able to choose the transverse mesh size independent of the bunch length. The other is a problem can be solved as an initial-value problem in spite of frequency domain calculation. The reason is Eq.(10) is the first order with respect to z . Common frequency domain calculations are eigenvalue problem. Therefore, it must be faster than FDTD.

TAPERED RECTANGULAR CHAMBER

It is important to calculate the impedances of collimators because they are designed depending on the wake field. When the tapered angle of a collimator is small, bunch length is short, and the collimator is not axisymmetric, this impedance is hard to calculate by ordinary method. However, it can be fast calculated by paraxial approximation. Both small tapered angle and short bunch are good for paraxial approximation. When the tapered angle of a collimator is small, the wave at small angle is dominant. For short bunch length, only small τ is need if the wake in the same bunch is considered. Therefore, we computed the impedance for collimator which is not axisymmetric by paraxial approximation. As an example of the collimator, tapered rectangular structure was choiced.

Geometry

The half height $b(z)$ depends on z . The half width w is 10mm. Figure 2 is up side view of the calculated geometry from $-x$ direction.

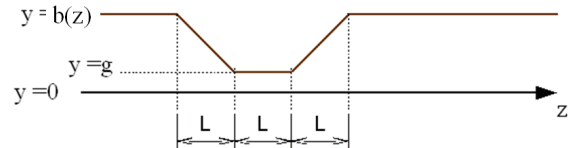


Figure 2: Side view of the upper half of the collimator. The maximum of $b(z)$ is 5 mm, and the minimum g is 3 mm. $L = 200$ mm.

Mesh

Figure 3 shows how to mesh about vertical and longitudinal direction. Mesh sizes Δx , $\Delta \eta$, $\Delta \zeta$ is constant. Here $\eta = y/b(z)$, $\zeta = z/k$.

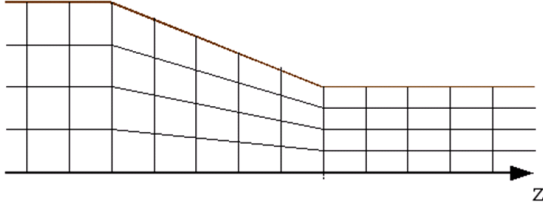


Figure 3: How to mesh.

Analytic Formula

Stupakov derived analytic formula,

$$Z_y(k \rightarrow 0) = -\frac{i\omega Z_0}{4} \int dz \frac{(b')^2}{b^3} \quad (21)$$

in Ref [6]. it is assumed $b \ll g \ll L$, and

$$k \ll \frac{b}{\alpha w^2}, \quad (22)$$

where α is the tapered angle.

Result

Figures 4 and 5 show comparisons of the simulation with the analytical value. In Fig. 4, the horizontal axis represents frequency f and the vertical axis represents real part of vertical impedance. Two kinds of data with different longitudinal mesh size are plotted. Blue diamonds show $\Delta \zeta = 10^{-9} \text{m}^2$. Green triangles show $\Delta \zeta = 10^{-8} \text{m}^2$. Orange line shows analytic solution. It is applicable while f is much smaller than 190 GHz because of condition (22). On the other hand, paraxial approximation requires that f is much greater than 16 GHz because $c\tau \ll g$ is corresponding to $f \gg c/(2\pi g) \approx 16 \text{ GHz}$. Therefore, the simulation agrees with the analytic solution while $16 \text{ GHz} \ll f \ll 190 \text{ GHz}$ although it fluctuates very much.

The imaginary impedance is shown in Fig. 5. Red squares show $\Delta \zeta = 10^{-9} \text{m}^2$. Purple \times 's show $\Delta \zeta = 10^{-8} \text{m}^2$. Yellow line shows analytic solution. As is mentioned, it is valid while f is much smaller than 190 GHz. On the other hand, paraxial approximation requires $f \gg 16 \text{ GHz}$. As you can see, the simulation agrees with the analytic solution while f is applicable range.

CONCLUSIONS AND DISCUSSIONS

In summary, for rectangular tapered chamber, the vertical impedance by the simulation agrees with the analytic solution in the appropriate frequency. Accordingly, the simulation is effective when bunch length σ_z is much smaller than g .

Computer Codes (Design, Simulation, Field Calculation)

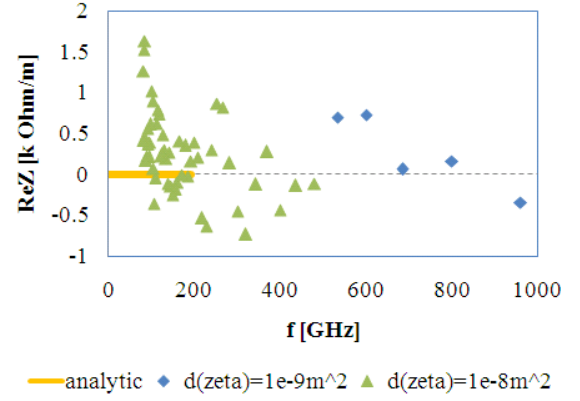


Figure 4: Real part of the impedance for rectangular taper.

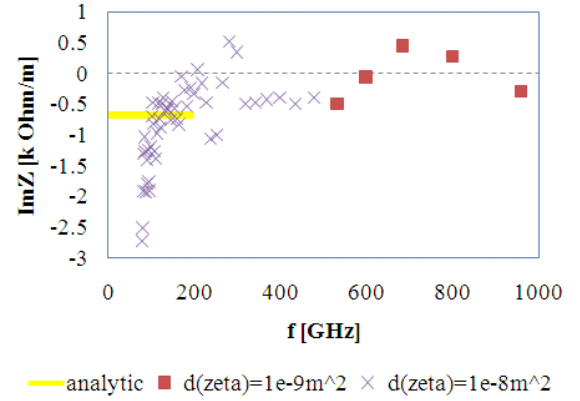


Figure 5: Imaginary part of the impedance for rectangular taper.

There are at least three future tasks. First is to improve the code. Second is to compare calculated impedances with that of other simulation codes, then to know how accurate and how fast the code is. Third is to calculate the impedance of ILC collimator. It is hard to calculate.

REFERENCES

- [1] K.S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", IEEE Trans. Antennas Propag., vol.AP-14, no.8, pp.302-307, May 1966.
- [2] T. Weiland, "A Discretization Method for the Solution of Maxwell's Equations for Six-Component Fields", Electronics and Communications AEUE, vol.31, no.3, pp. 116-120, 1977.
- [3] G. Stupakov, "Using the parabolic equation for calculation of beam impedance", New J. Phys. 8, 280 (2006).
- [4] T. Agoh, "Dynamics of Coherent Synchrotron Radiation by Paraxial Approximation", Doctoral Thesis, U. Tokyo (2004).
- [5] T. Agoh, K. Yokoya, "Calculation of coherent synchrotron radiation using mesh", Phys. Rev. ST Accel. Beams 7 054403, 2004.
- [6] G. V. Stupakov, "Impedance of Small-Angle Collimators in High-Frequency Limit", SLAC-PUB-8857, SLAC (June 2001).

APPLYING AN HP-ADAPTIVE DISCONTINUOUS GALERKIN SCHEME TO BEAM DYNAMICS SIMULATIONS*

S. Schnepf[†], Graduate School of Computational Engineering, TU Darmstadt, Germany
 E. Gjonaj[‡], T. Weiland[§], Institut für Theorie Elektromagnetischer Felder, TU Darmstadt, Germany

Abstract

An adaptive high order discontinuous Galerkin (DG) scheme for performing beam dynamics simulations is presented. We elaborate on h - and p -adaptations, the former modifying the actual size of computational elements and the latter the dimension of the associated approximation space. The efficiency and stability of the adaptation techniques are emphasized. The scheme is applied in order to perform hp -adaptive beam dynamics simulations. We compare the results with the analytical solution and demonstrate that the adaptive scheme requires significantly less computational resources for obtaining a certain accuracy.

INTRODUCTION

The problem of self-consistent simulations of short relativistic particle bunches in long accelerator structures exhibits a pronounced multi-scale character. The adequate resolution of the THz space charge fields excited by short ultra-relativistic bunches requires mesh spacings in the micrometer range. On the other hand, the discretization of complete accelerator sections using such fine meshes results in a prohibitive number of Degrees of Freedom (DoF). Due to the spatial concentration of the particles and the excited space charge fields, the application of time-adaptive mesh refinement is an emerging idea. We reported on the implementation of time-adaptive mesh refinement for the Finite Integration Technique (FIT) [1]. Based on this work, an adaptive discontinuous Galerkin (DG) code was implemented. Within the DG method, the electromagnetic field solution is approximated elementwise, employing a set of basis functions. This provides two options for adapting the local accuracy of the DG approximation. First, the size of the grid elements can be varied. This is referred to as h -adaptation. Additionally, the maximum order of the employed basis functions can be modified, which is referred to as p -adaptation. Combining both options yields an hp -adaptive method. The twofold refinement mechanisms of the hp -adaptive DG method offer maximum freedom for the approximation of the electromagnetic field solution.

DISCONTINUOUS GALERKIN METHOD FOR MAXWELL'S EQUATIONS

Spatial Discretization Procedure

Given a decomposition of the computational domain Ω into N non-overlapping, hexahedral elements $\{C_i\}$, $i = 1..N$, a set of linearly independent basis functions $\{\varphi_i^p\}$, $p = 0..P$ for every cell is defined, where P denotes the highest order employed. The basis functions are required to be continuous within the cell C_i and vanish otherwise

$$\varphi_i^p(\mathbf{r}) = \begin{cases} \varphi_i^p(\mathbf{r}), & \mathbf{r} \in C_i, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Subsequently, the space and time continuous electromagnetic field quantities \mathbf{E} and \mathbf{H} are approximated in the form

$$\mathbf{E}(\mathbf{r}, t) \approx \tilde{\mathbf{E}}(\mathbf{r}, t) = \sum_i \tilde{\mathbf{E}}_i(\mathbf{r}, t) = \sum_{i,p} e_i^p(t) \varphi_i^p(\mathbf{r}), \quad (2)$$

$$\mathbf{H}(\mathbf{r}, t) \approx \tilde{\mathbf{H}}(\mathbf{r}, t) = \sum_i \tilde{\mathbf{H}}_i(\mathbf{r}, t) = \sum_{i,p} h_i^p(t) \varphi_i^p(\mathbf{r}), \quad (3)$$

where $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ denote the approximate field vectors. The numerical DoF are denoted by e_i^p and h_i^p . They are gathered in the vectors \mathbf{e} and \mathbf{h} .

Substituting the electromagnetic quantities by their approximations in Faraday's and Ampère's law and applying the Galerkin procedure yields the weak DG formulation [2]. Due to the cell-wise compact support of the basis functions (1), the approximations (2) and (3) will, in general, be discontinuous at element interfaces. Continuity is enforced only in the weak sense via numerical fluxes. Among the different flux definitions, we chose centered fluxes. As demonstrated in [2, 3] this ensures the strict conservation of the electromagnetic energy. Using the naming convention given there and vector notation for all terms, the semidiscrete formulation reads

$$\frac{d}{dt} \begin{pmatrix} \mathbf{M}_\epsilon \mathbf{e} \\ \mathbf{M}_\mu \mathbf{h} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & -\mathbf{C} \\ \mathbf{C}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{e} \\ \mathbf{h} \end{pmatrix} = - \begin{pmatrix} \mathbf{j} \\ \mathbf{0} \end{pmatrix}. \quad (4)$$

The terms \mathbf{M}_ϵ and \mathbf{M}_μ are the mass matrices and \mathbf{C} denotes the weak DG curl operator. The vector \mathbf{j} represents the convective currents.

In the particular case of particle accelerator problems, the issue of charge conservation is specially relevant due to the existence of freely moving charges. In [3] it was shown that strict charge conservation is guaranteed if, and only if, a tensor product basis on conforming Cartesian grids is

* The work of S. Schnepf is supported by the 'Initiative for Excellence' of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt.

[†] schnepf@gsc.tu-darmstadt.de

[‡] gjonaj@temf.tu-darmstadt.de

[§] weiland@temf.tu-darmstadt.de

applied for the approximations (2) and (3). The continuous current density is given by

$$\mathbf{J}(\mathbf{r}, t) = \sum_{k=1}^{N_p} Q_k \mathbf{v}_k S(|\mathbf{r}_k - \mathbf{r}|), \quad (5)$$

where Q_k is the charge carried by the k -th of N_p particles, \mathbf{r}_k and \mathbf{v}_k are its position and velocity vector and S is a distribution function describing the assumed particle shape. In case S corresponds to the Kronecker delta, the particles are modeled as point charges. Other shape functions are discussed in [4]. Integrating the current density over one time step and projecting to the basis functions yields the DG approximation

$$(j_i^p)^n = \frac{(\varphi_i^p, \int_{(n-1)\Delta t}^{n\Delta t} \mathbf{J}(\mathbf{r}, t) dt)_{C_i}}{(\varphi_i^p, \varphi_i^p)_{C_i}}. \quad (6)$$

Above, $(a, b)_{C_i}$ denotes the inner product of a and b on C_i , Δt is the time step and n the time step number.

Suitability for hp -Adaptation

The discontinuous Galerkin method is highly suited for the application of h - and p -adaptation techniques. Independently of the approximation order, the method always preserve the same high degree of locality. Due to the element-wise compact support, individual elements communicate only via interface fluxes. The computation of these, in turn, involves only the interface values of direct neighbors. Splitting and merging of elements, hence, is a purely local operation, which affects none but the adapted elements. For our particular setup of Cartesian grids and a tensor product basis, this adaptation can be performed very efficiently, as it will be shown in the following section. Since additionally, assigning different approximation orders throughout the elements of the computational domain is readily possible, the method is well-suited for an hp -adaptive procedure, i.e., adapting the local grid step size, as well as the local approximation order in a problem oriented fashion.

ADAPTATION TECHNIQUES

h -Adaptation

When performing h -adaptation the size of dedicated elements is modified locally by splitting them into subelements or, conversely, by merging them into larger ones. The grid topology is, thus, altered while the approximation orders of the involved elements are kept constant.

In the case of h -refinement, the approximations within the subelements are obtained by projecting the existing approximation onto the support of the subelements. We restrict ourselves to the bisection of elements and do not consider other division ratios. Continuous basis functions $\psi_{i,l}$ and $\psi_{i,r}$ are defined within the interior of the left and right

Computer Codes (Design, Simulation, Field Calculation)

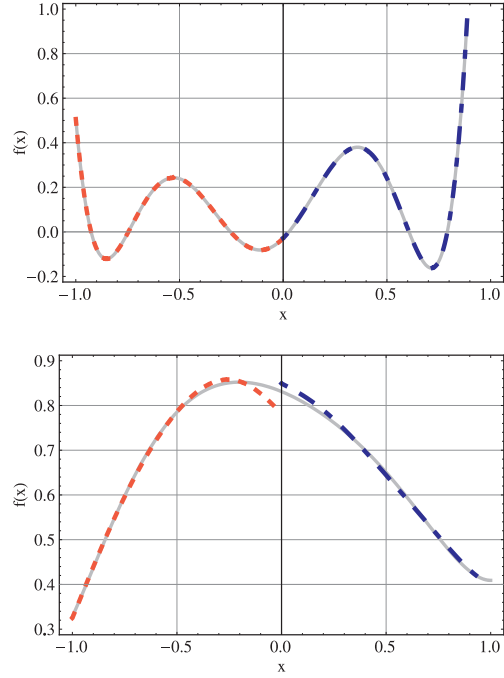


Figure 1: For the splitting of elements (top), the approximate solution within the element C_i (gray, solid) is accurately projected onto its subelements $C_{i,l}$ (red, dotted) and $C_{i,r}$ (blue, dash-dotted). Due to the discontinuity the approximation within the merged element cannot be exact (bottom).

subelement $C_{i,l}$ and $C_{i,r}$ and set identically zero elsewhere. The integral terms

$$\mathbf{P}_l^{qp} = \frac{(\varphi_l^p, \psi_l^q)_{C_{i,l}}}{(\psi_l^q, \psi_l^q)_{C_{i,l}}} \quad \text{and} \quad \mathbf{P}_r^{qp} = \frac{(\varphi_r^p, \psi_r^q)_{C_{i,r}}}{(\psi_r^q, \psi_r^q)_{C_{i,r}}}, \quad (7)$$

for $p, q = 1..P$ determine the entries of the projection matrices \mathbf{P}_l and \mathbf{P}_r . Each term describes the contribution of the basis function of order p of the element C_i to the basis function of order q of its subelements. The numerical DoF of the subelements, $\mathbf{e}_{i,l}$, $\mathbf{e}_{i,r}$ and $\mathbf{h}_{i,l}$, $\mathbf{h}_{i,r}$, are obtained by means of the matrix-vector multiplications

$$\mathbf{e}_{i,l} = \mathbf{P}_l \mathbf{e}_i, \quad \mathbf{e}_{i,r} = \mathbf{P}_r \mathbf{e}_i, \quad (8)$$

$$\mathbf{h}_{i,l} = \mathbf{P}_l \mathbf{h}_i, \quad \mathbf{h}_{i,r} = \mathbf{P}_r \mathbf{h}_i. \quad (9)$$

These operations are exact in the sense that the approximation within the subelements is identical to the approximation within the original element in every point. This is shown in Figure 1 (top).

For the converse process, i.e., merging of elements, it is not possible to obtain an exact representation within the merged element due to the discontinuous character of the approximations. The approximation within C_i is considered to be given piecewise on the subelements with the DoF $\mathbf{e}_{i,l}$, $\mathbf{e}_{i,r}$ and $\mathbf{h}_{i,l}$, $\mathbf{h}_{i,r}$. The entries of the projection matrix

\mathbf{P}_m are specified by

$$\begin{aligned}\mathbf{P}_m^{qp} &= \frac{(\psi_l^q + \psi_r^q, \varphi^p)_{C_i}}{(\varphi^p, \varphi^p)_{C_i}} = \frac{(\psi_l^q, \varphi^p)_{C_i}}{(\varphi^p, \varphi^p)_{C_i}} + \frac{(\psi_r^q, \varphi^p)_{C_i}}{(\varphi^p, \varphi^p)_{C_i}} \\ &= \mathbf{P}_{m,l}^{qp} + \mathbf{P}_{m,r}^{qp}.\end{aligned}\quad (10)$$

Summing up the contributions of the subelements yields the DoF of the merged element

$$\mathbf{e}_i = \mathbf{P}_{m,l}\mathbf{e}_{i,l} + \mathbf{P}_{m,r}\mathbf{e}_{i,r}, \quad (11)$$

$$\mathbf{h}_i = \mathbf{P}_{m,l}\mathbf{h}_{i,l} + \mathbf{P}_{m,r}\mathbf{h}_{i,r}. \quad (12)$$

The matrices $\mathbf{P}_{m,l}$ and $\mathbf{P}_{m,r}$, as well as \mathbf{P}_l and \mathbf{P}_r , do not depend upon the actual approximation but only on the basis functions. They can be evaluated analytically and stored for repeated use.

p-Adaptation

P-adaptation refers to a local or global modification of the approximation order P while keeping the mesh unaltered. In time-domain simulations *p*-adaptation can be efficiently implemented if a set of hierarchical basis functions is employed. In this case, the DoF associated with the different approximation orders do not depend upon each other. Thus, increasing the approximation order is as simple as preserving the current coefficients and attaching those of the next higher order basis functions to the DoF vectors \mathbf{e}_i and \mathbf{h}_i for the respective element C_i . The new coefficients are initialized to zero. In order to reduce the order, the DoF associated with the higher order basis functions are dismissed.

hp-Adaptation

For smooth solutions, the approximation order is directly linked to the asymptotic convergence order. It is, thus, desirable to work with high order elements. However, the convergence order breaks down if the solution is non-smooth. In Fig. 2 the error of the DG approximation to a Gaussian and a trapezoidal field distribution is shown. The error is plotted versus the grid step size, both of them on a logarithmic scale. The approximation error of the Gaussian (top) reduces with the expected rate for increased P . For the trapezoidal distribution (bottom), however, the convergence rate stagnates for higher approximation orders. An important key point of an *hp*-adaptation algorithm is, hence, the detection of regions with a low degree of solution smoothness. Also, regions of low electromagnetic field energy should be identified since they probably do not require a high resolution.

Efficiency and Stability

We employ the orthogonal set of Legendre polynomials as basis functions and define a tensor product basis for conforming Cartesian grids. This particular setting allows for Computer Codes (Design, Simulation, Field Calculation)

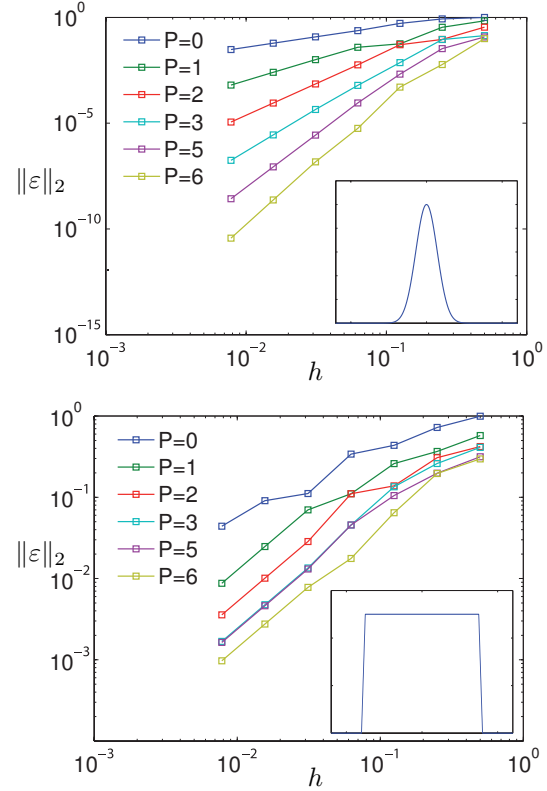


Figure 2: Approximation error in the L^2 -norm of a Gaussian (top) and a trapezoidal distribution (bottom) versus the grid step. For the smooth function in the top graph the expected convergence rate is observed. For the non-smooth function in the bottom graph the convergence rate breaks down.

a very efficient implementation of the DG method since all inner products of basis and test functions reduce to zero if their order differs and to a factor which depends only on the dimensions of the respective element, otherwise.

The projection based *h*-adaptation can also be implemented in a very efficient manner since the projection matrices do not vary with time. They have to be evaluated only once, which can be done analytically. During the actual simulation, we additionally benefit from the Cartesian grids and the tensor product basis ansatz since the bisection of elements and the calculation of the coefficients within the modified elements can also be performed very efficiently.

It remains to address the stability of the adaptation techniques. The adaptations are stable if their application does not increase the electromagnetic energy within the modified elements. The energy of the element C_i reads

$$W_i = \frac{1}{2}|C_i|(\epsilon_i\|\mathbf{e}_i\|_2^2 + \mu_i\|\mathbf{h}_i\|_2^2), \quad (13)$$

where $|C_i|$ denotes the volume of C_i and ϵ_i, μ_i are its permittivity and permeability.

It is readily seen that a *p*-reduction induces a loss of the discrete energy associated with the respective coefficients.

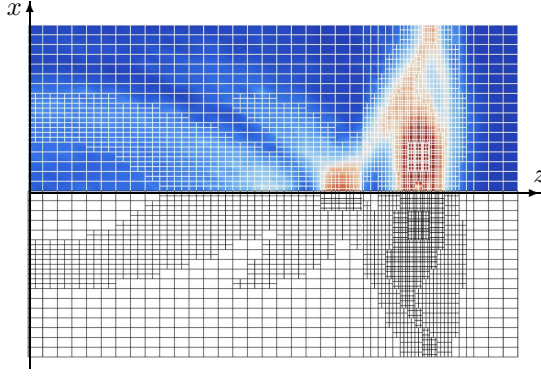


Figure 3: Snapshot of the total electric field excited by a bunch of particles in cut view. The upper half shows the electric field and the computational grid, the lower half shows the hp -adapted grid exclusively.

The coefficients, which are added during p -enrichment are initialized by zero. In this case W_i is preserved.

We will not present formal derivations of the stability of the h -adaptations but rather state their results. Details are found in [5]. In the case of h -refinement, the sum of the discrete energies of the subelements is equal to the energy of the parent element. Since the approximations within the subelements exactly represent the original approximation this is consequential. For h -coarsening the energy is at most preserved. The application of p -enrichment or h -refinement, hence, conserves the discrete energy exactly while p -reduction or h -coarsening usually induce a loss of energy.

APPLICATIONS

The method has been applied to the simulation of a particle bunch drifting with the constant velocity $v = 0.9c_0$ in a cylindrical perfectly electrically conducting tube, where c_0 is the speed of light in vacuum. The electromagnetic field solution for this situation at an arbitrary point in time is described in [6], where the tube is considered semi-infinite in the direction of bunch propagation. The length and radius of the tube are set to 120 mm and 40 mm respectively. The particle distribution of the bunch is Gaussian with an RMS length and radius of 5 mm and 3 mm, respectively.

We apply hp -adaptation for the simulation of the problem. The p -adaptation is controlled by the field amplitude. In addition, h -refinement is applied to the elements containing particles and all neighboring elements with the same index along the z -axis. This ensures a conforming grid refinement. Fig. 3 shows the total electric field along with the hp -adapted grid. For the p -adaptation a range of one through four for the approximation orders was applied. In order to visualize the approximation order of an element, the respective number of collocation points are plotted. Along the z -direction an additional conforming h -refinement is applied to the elements, which contain particles and all neighboring elements with the same z -index.

Table 1: Relative error and total variation of solutions obtained by the non-adaptive and adaptive scheme using various settings.

L	P_{\min}	P_{\max}	$\approx \text{DoF} / 1e6$	Time / sec	ϵ^{rel}	TV
0	1	1	6.80	310	0.081	8.71
0	2	2	22.96	2020	0.042	4.15
1	1	3	9.00	1170	0.043	2.68
1	1	4	10.50	1650	0.028	1.00

The relative L^∞ -error and the total variation (TV) [7] of the electric field along the cylinder axis are given as measures for the quality of the numerical solutions. The TV is a measure for the smoothness of a function. It is defined as

$$TV(\mathbf{E}(z)) = \int |\mathbf{E}'(z)| dz. \quad (14)$$

Table 1 summarizes the results. The number of h -refinement levels is denoted by L . The grid step sizes and the h -refinement level are chosen such that the minimum step size remains identical. The total variation is normalized to the smallest value obtained. The adaptive simulations consume more time per DoF. Partly, the extra time is spent for the adaptation routines, and partly it is due to the reduced maximally stable time step connected with higher approximation orders.

As a second application, we have simulated a part of the PITZ injector (Photo Injector Test Facility at DESY Zeuthen) [8]. The PITZ project was initiated in order to test and optimize sources of high brightness electron beams for future free electron lasers (FELs) and linear colliders. Fig. 4 shows the total electric field and hp -adapted grid in a y -cut through the three-dimensional domain. The simulated section consists of the 1.5 cell RF gun and has a length of 25 cm. This grid was generated for illustration purposes only. Accurate simulations require a higher resolution. Nevertheless, it demonstrates the ability of the scheme to handle more complex situations.

OUTLOOK ON ERROR BASED AUTOMATIC HP-ADAPTATION

For the examples presented above, we used a naive algorithm for controlling the hp -adaptation. The approximation order p of every element, e.g., is determined from a comparison of its field magnitude with the maximum field magnitude. However, the adaptation should be based on the approximation error, which is for general examples not linked to the field magnitude. In [9] it is shown that the residual relates to the size of the jump across element interfaces. Performing an error estimation is, hence, a trivial task for the DG method. After deciding whether to adapt an element, the next step is the decision between h - and p -adaptation. This requires the estimation of the local smoothness of the

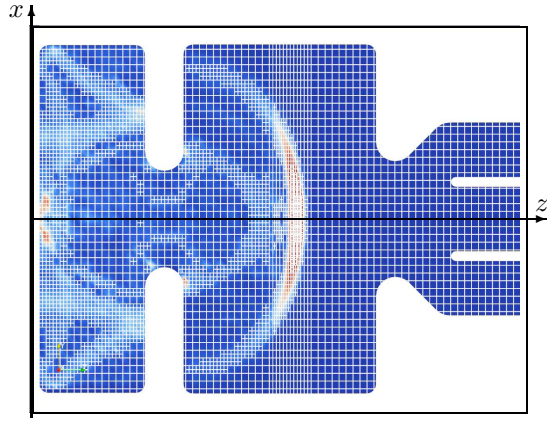


Figure 4: Snapshot of the total electric field excited by a bunch of particles in cut view. The upper half shows the electric field and the computational grid, the lower half shows the hp -adapted grid exclusively.

approximation. In [5, 10] smoothness indicators are presented.

We implemented an algorithm for performing automatic hp -adaptations for the scalar wave equation, which was tested on a Gaussian and trapezoidal wave packet as depicted in Fig. 5. There, the blue and green line represent the electric and magnetic field approximations. The gray dashed lines depict the position of the grid points and the height of the red circles indicate the approximation order employed for the respective element divided by ten. The snapshots were recorded after several hundreds of time steps. For the Gaussian packet the adaptation algorithm chooses medium sized to big elements and the preset maximum approximation order of five in the vicinity of the packet. For the non-smooth trapezoidal packet, the algorithm does not employ an approximation order above two throughout the simulation. In the vicinity of the pulse edges a high degree of h -refinement is applied, thus showing the desired behavior for the second packet as well.

CONCLUSIONS

Adaptation techniques for the high order DG method have been presented, which we apply on Cartesian grids using a tensor product basis of orthogonal basis functions. This particular setting allows us to perform h - and p -adaptations in a very efficient manner. Details of both kinds of adaptation were presented, and we showed that their application does not induce any instability. We performed hp -adaptive simulations and compared the results with the analytical solution. The hp -adaptive scheme was shown to require significantly less computational resources and time, yielding a higher accuracy as the non-adaptive method.

REFERENCES

[1] S. Schnepf, E. Gjonaj, and T. Weiland, “A Time-Adaptive Mesh Approach for the Self-Consistent Simulation of Particle Computer Codes (Design, Simulation, Field Calculation)

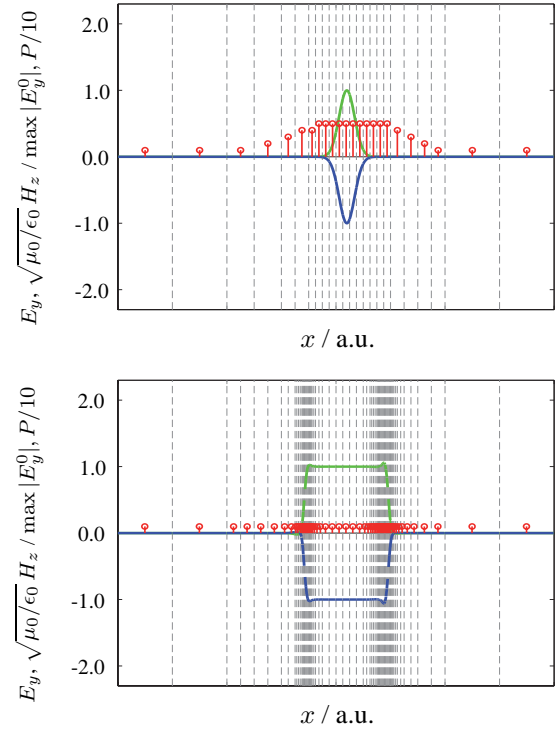


Figure 5: Simulation of a Gaussian and a trapezoidal wave packet using automated hp -adaptation. The electric field is plotted in blue and the magnetic field in green. The grid is indicated by gray dashed lines and the red circles indicate the approximation order used in the respective element.

Beams”, Int. Comp. Acc. Phys. Conf. (ICAP), 2006.

[2] E. Gjonaj, T. Lau, S. Schnepf, F. Wolfheimer, and T. Weiland, “Accurate Modelling of Charged Particle Beams in Linear Accelerators”, New Journal of Physics (8), 2006.

[3] E. Gjonaj, T. Lau, and T. Weiland, “Conservation Properties of the Discontinuous Galerkin Method for Maxwell Equations”, Int. Conf. Electromagnetics in Advanced Applications (ICEAA), 2007.

[4] J.S. Hesthaven, and G.B. Jacobs, “High-order nodal discontinuous Galerkin particle-in-cell method on unstructured grids”, J. of Comp. Phys. (214), 2005.

[5] S. Schnepf, “Space-Time Adaptive Methods for Beam Dynamics Simulations”, PhD thesis, TU Darmstadt, 2009.

[6] I.N. Onishchenko, D.Y. Sidorenko, and G.V. Sotnikov, “Structure of electromagnetic field excited by an electron bunch in a semi-infinite dielectric-filled waveguide”, Phys. Rev. E (65), 2006.

[7] A. Harten, “High Resolution Schemes for Hyperbolic Conservation Laws”, J. Comp. Phys. (49), 1983.

[8] K. Flöttmann, and F. Stephan, “BMBF Proposal for a RF Photoinjectors as Sources for Electron Bunches of Extremely Short Length and Small Emittance”, 1999.

[9] B. Cockburn, “Discontinuous Galerkin Methods”, Z. Angewandte Mathematische Mechanik (83), 2003.

[10] L. Krivodonova, N. Chevaugneon, and J. Flaherty, “Shock Detection and Limiting with Discontinuous Galerkin Methods for Hyperbolic Conservation Laws”, Appl. Numer. Math., 2004.

PORTABLE HIGH PERFORMANCE COMPUTING FOR MICROWAVE SIMULATION BY FDTD/FIT MACHINES*

Yuya Fujita and Hideki Kawaguchi[#], Muroran Institute of Technology, Muroran, 050-8585, Japan

Abstract

This paper presents a development of fully customized printed circuit board of a dedicated computer for FDTD/FIT method which is aiming to portable high performance computation for microwave simulation. In the construction of dedicated computer hardware, it is very important to carefully consider parallel properties hidden in the target application scheme of the dedicated computer to achieve its maximum performance. In addition, judgement on tread-off between calculation performance and flexibility for various target applications is also very important in concrete design of the dedicated computer architecture. In this paper, basic concept and concrete design of the FDTD/FIT dedicated computer architecture are described, and a printed circuit board which is manufactured according to these concepts are presented.

INTRODUCTION

Due to strong requirements of high performance computation (HPC) for electromagnetic microwave simulation in not only science but also industry, a method of FDTD or FIT dedicated computer has been actively investigated in the last several years [1-8]. One of most important advantages of the method of dedicated computer in a regime of HPC technologies is possibility of portable computing. That is to say, in most cases, HPC hardware such as supercomputer and PC cluster is installed at apart from user computers because such the HPC hardware are basically multi-user system, therefore those kinds of HPC environment are not familiar with industrial applications such as microwave simulation connecting with CAD system which are usually carried out in user personal computers. Then the method of the dedicated computer is one of powerful solutions to those requirements. Especially appearance of large size rewritable LSI hardware such as FPGA and GPU, very useful LSI design tool and low price Printed Circuit Board (PCB) development service enhance such the activities of dedicated computer researches. Authors also have been working in development of dedicated computers of the FDTD or FIT method for high performance computation in microwave simulation [5-8]. The method of dedicated computer allows us to construct highly optimized hardware architecture and high performance computation dedicated into the calculation

scheme of the target application by comparably small size hardware. Especially a problem of Neumann bottleneck related to memory access overhead, which is essential problem for achievement of HPC in Neumann architecture based computer system, can be avoidable in the method of dedicated computer. This paper presents a design of the dedicated computer specialized into FDTD / FIT method and its hardware implement.

PARALLEL PROPETIES HIDDEN IN FDTD METHOD AND CONCEPTUAL DESIGN OF DEDICATED COMPUTER

Parallel properties hidden in FDTD method

To achieve efficient computation in use of dedicated computer, we need to embed parallel properties hidden in the FDTD method ((1) and (2)) into the hardware architecture of the dedicated computer.

$$E_{x i, j, k}^{n+1} = E_{x i, j, k}^{n-1} + \frac{\Delta t}{\epsilon \Delta l} \left[H_{z i, j, k}^n - H_{z i, j-1, k}^n - H_{y i, j, k}^n + H_{y i, j, k-1}^n \right], \quad (1)$$

$$E_{y i, j, k}^{n+1} = E_{y i, j, k}^{n-1} + \frac{\Delta t}{\epsilon \Delta l} \left[H_{x i, j, k}^n - H_{x i, j, k-1}^n - H_{z i, j, k}^n + H_{z i-1, j, k}^n \right]$$

$$E_{z i, j, k}^{n+1} = E_{z i, j, k}^{n-1} + \frac{\Delta t}{\epsilon \Delta l} \left[H_{y i, j, k}^n - H_{y i-1, j, k}^n - H_{x i, j, k}^n + H_{x i, j-1, k}^n \right]$$

$$H_{x i, j, k}^n = H_{x i, j, k}^{n-1} - \frac{\Delta t}{\mu \Delta l} \left[E_{z i, j+1, k}^n - E_{z i, j, k}^n - E_{y i, j, k+1}^n + E_{y i, j, k}^n \right], \quad (2)$$

$$H_{y i, j, k}^n = H_{y i, j, k}^{n-1} - \frac{\Delta t}{\mu \Delta l} \left[E_{x i, j, k+1}^n - E_{x i, j, k}^n - E_{z i+1, j, k}^n + E_{z i, j, k}^n \right]$$

$$H_{z i, j, k}^n = H_{z i, j, k}^{n-1} - \frac{\Delta t}{\mu \Delta l} \left[E_{y i+1, j, k}^n - E_{y i, j, k}^n - E_{x i, j+1, k}^n + E_{x i, j, k}^n \right]$$

There are roughly three kinds of parallel properties in the FDTD method,

- dataflow property in the FDTD calculation structure
- parallel calculation of x-, y-, z- three field components
- parallel calculation of different grid field components

The FDTD dedicated computer implementing all of these parallel properties (we call this "full dataflow architecture FDTD/FIT machine") indeed gives us extremely high performance computation, however its hardware size is quite large and especially 3D FDTD machine will be impossible in the next several years even beyond remarkable progress of recent LSI technology [6]. Accordingly the FDTD/FIT dedicated computer based on the first two parallel properties is practical solution in the present LSI technology (we call this "memory architecture FDTD/FIT machine").

Design of memory architecture FDTD machine

The figure 1 shows an overview of the FDTD/FIT dedicated computer architecture. The hardware mainly consists of two parts, calculation and memory modules. All components of electromagnetic field values, material

*Work supported by JSPS Grants-in-Aid for Scientific Research (No. 21560345)

[#]kawa@mmm.muroran-it.ac.jp

constant distributions, boundary conditions are stored in the memory module, and x-, y-, z- three component FDTD scheme calculation for one grid are executed in the calculation module. The fundamental operation of this machine is as follows, (a) downloading electromagnetic field components from the memory module to the calculation module, (b) execution of one grid FDTD calculation ((1) or (2)) and PML calculation at the calculation module, (c) uploading the calculated field components from the calculation module to the memory module.

Then calculation module is composed of three same dataflow circuits for both of normal FDTD grid (see Fig.2 for (1)) and PML grid to perform three components FDTD calculation of (1) or (2) in single computer clock cycle. This structure of the calculation module realizes the first two parallel properties of the FDTD method.

Parallel memory access architecture

The FDTD method is heavy memory access scheme, for example, at least 12 different field component values are needed in three components calculation as in the right hand side of (1) or (2). This mean that 12 times memory access are required if the memory module is composed by simply single RAM. To avoid this quite inefficient machine operation for the FDTD calculation, we adopt parallel memory access architecture for the FDTD/FIT dedicated computer as shown in Fig.1. That is, in the memory module, each field component value is stored individually in four physically different RAMs. In this memory structure, all field components which are needed in three component calculation of (1) or (2) can be loaded from the memory module to the calculation module within only single memory access cycle.

Automatic boundary condition setting

Implementation of boundary condition setting in the dedicated computer is also one of most important matter for practical use of the FDTD/FIT machine. There are mainly three kinds of fundamental boundary conditions,

material boundary condition, perfect conductor boundary condition and absorbing boundary condition such as PML. These boundary conditions can be automatically imposed by hardware circuit of the dedicated computer [see ref. 5 and 6]. Especially various kind of numerical models can be treated by rewriting the contents of memory module, then it is not necessary to change hardware circuit.

FULL CUSTOM PCB OF FDTD/FIT DEDICATED COMPUTER

The figure 3 shows fully custom made printed circuit board (PCB) for the FDTD/FIT dedicated computer according to the memory architecture of Fig.1. The memory module is composed by 16 SRAMs and all other complicated circuits such as calculation module, address generator, machine operation scheduler, etc. are embedded in one FPGA. As an example of microwave simulation based on this FDTD / FIT dedicated computer, microwave radiation from rectangular wave guide is simulated here (see Fig.4). The numerical model is stored in one of SRAMs and then machine hardware automatically executes FDTD microwave simulation for this numerical model without any software program code. It is confirmed that simulation results show good agreement with software simulations.

CONCLUSION

This paper has presented a development of the FDTD/FIT dedicated computer aiming portable high performance computation of electromagnetic microwave simulation. Especially highly optimized memory access architecture dedicated into the FDTD scheme are introduced to achieve high throughput performance including memory access. And then real PCB hardware of the dedicated computer has been shown.

Based on the development of the prototype FDTD/FIT machine, we are now planning to proceed to further improvements such as introduction of DDR like high speed SDRAM, parallel calculation by interlocking operation of the dedicated computer PCBs, etc.

MEMORY MODULE

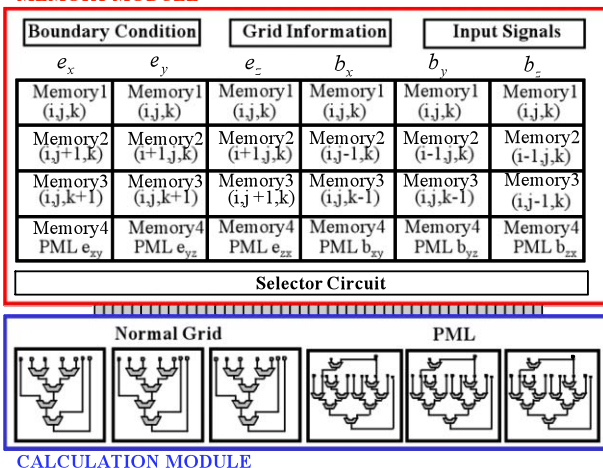


Fig.1 Hardware of architecture FDTD/FIT machine
Computer Codes (Design, Simulation, Field Calculation)

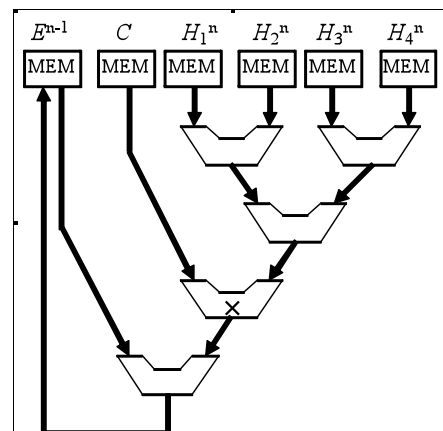


Fig.2 Dataflow circuit for FDTD/FIT basic calculation

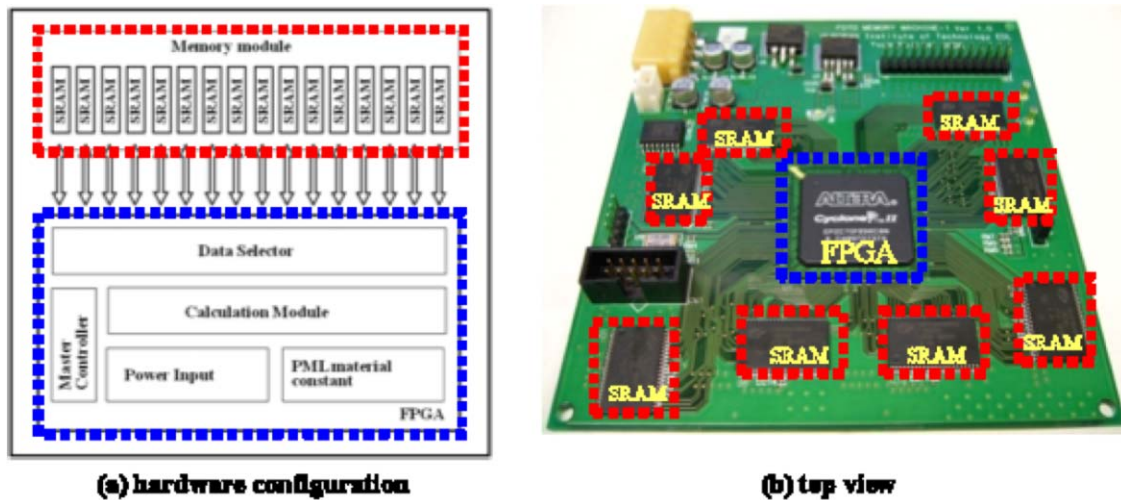


Fig.3 Full custom printed circuit board of FDTD / FIT dedicated computer

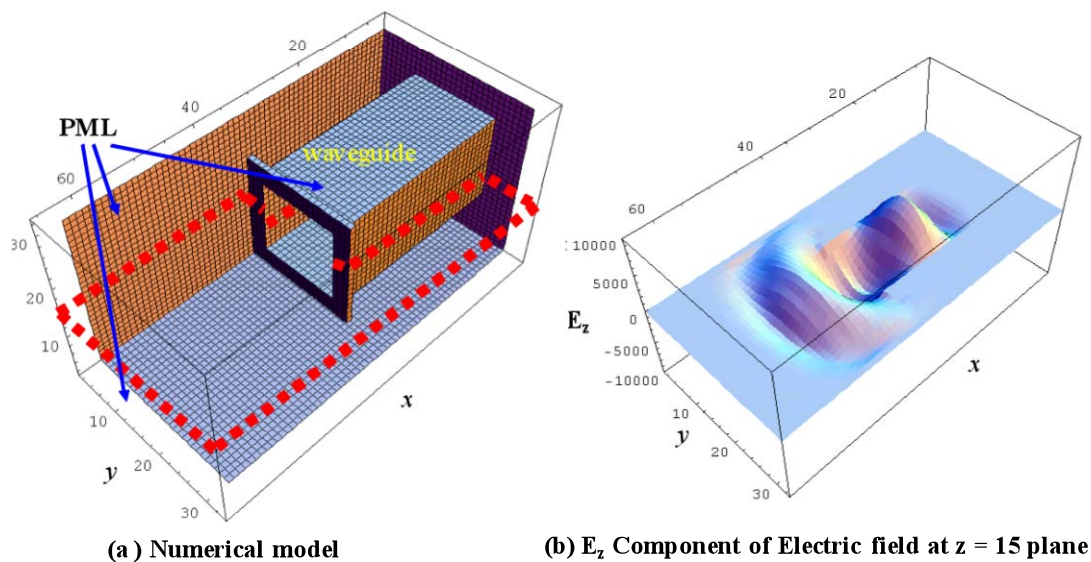


Fig.4 Numerical example of FDTD/FIT machine operation for radiation from waveguide

REFERENCES

- [1] R.N.Schneider, L.E.Turner, and M.M.Okoniewski, "Application of FPGA technology to accelerate the finite-difference time-domain(FDTD) method," Proc. 10th Int. Symp. Field Programmable Gate Arrays, Feb. pp.24-26, 2002.
- [2] P.Placidi, L.Verducci, G.Matrella, L.Roselli, and P.Ciampolini, "A custom VLSI architecture for the solution of FDTD equations," IEICE Trans. Electron., Vol. E85-C, No.3, pp. 572-577, -2002
- [3] W.Chen, P.Kosmas, M.Leeser, and C.Rappaport, "An FPGA Implementation of the Two-Dimensional Finite-Difference Time-Domain (FDTD) Algorithm," presented at the 12th ACM International Symposium on Field-Programmable Gate Arrays, 2004.
- [4] J.P.Durbano, F.E.Ortiz, J.R.Humphrey, and D.W.Prather, "FPGA based acceleration of the 3D finite difference time domain method," presented at the 12th Annu. IEEE Symp. Field-Programmable, Custom Computing Machines (FCCM), Napa Valley, CA, 2004.
- [5] H.Kawaguchi, K.Takahara, and D.Yamauchi, "Design study of ultra high speed microwave simulator engine," IEEE Trans. Magn., Vol. 38, No. 2, pt. 1, pp. 689-692, 2002.
- [6] S.Matsuoka and H.Kawaguchi, Study of a Microwave Simulation Dedicated Computer: FDTD/FIT Data Flow Machine, IEICE Trans. Electron., Vol.E86-c, No.11 (2003), pp.2199-2206.
- [7] H.Kawaguchi, Y.Fujita, Y.Fujishima and S.Matsuoka, Improved Architecture of FDTD/FIT Dedicated Computer for Higher Performance Computation, IEEE Trans. on Magn., Vol.44, No.6, p1226-1229, 2008
- [8] Y.Fujita, H.Kawaguchi, Full Custom PCB Implementation of FDTD/FIT Dedicated Computer, IEEE Transactions on Magnetics, Volume: 45, Issue: 3, p1100-1103, March 2009.

HIGH-ORDER DIFFERENTIAL ALGEBRA METHODS FOR PDEs INCLUDING RIGOROUS ERROR VERIFICATION

Shashikant Manikonda[†] Martin Berz[‡] and Kyoko Makino[‡]

[†]Physics Division, Argonne National Laboratory, Argonne, IL 60439

[‡]Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824

Abstract

Many processes in Physics can be described by Partial Differential equations (PDE's). For various practical problems, very precise and verified solutions of PDE are required; but with conventional finite element or finite difference codes this is difficult to achieve because of the need for an exceedingly fine mesh which leads to often prohibitive CPU time. We present an alternative approach based on high-order quadrature and a high-order finite element method. Both of the ingredients become possible through the use of Differential Algebra techniques. Further the method can be extended to provide rigorous error verification by using the Taylor model techniques. Application of these techniques and the precision that can be achieved will be presented for the case of 3D Laplace's equation. Using only around 100 finite elements of order 7, verified accuracies in the range of 1E-7 can be obtained.

INTRODUCTION

Many problems in physics and engineering require the solution of the three dimensional (3D) Laplace equation

$$\Delta\psi(\vec{r}) = 0 \text{ in the bounded volume } \Omega \subset \mathbb{R}^3 \quad (1)$$

It is well known that under mild smoothness conditions for the boundary $\partial\Omega$ of Ω , the Laplace equation admits unique solutions if either ψ or its derivative normal to $\partial\Omega$ are specified on the entire boundary surface $\partial\Omega$. In many typical applications, not only the normal derivative of ψ but indeed the entire gradient $\vec{\nabla}\psi$ is known on the surface; for example, in the magnetostatic case the entire field $\vec{B} = \vec{\nabla}\psi$ is measured, and not merely whatever component happens to be normal to the surface under consideration. The corresponding problem of determining ψ based on the knowledge of the field $\vec{\nabla}\psi(\vec{r}) = \vec{f}(\vec{r})$ on the surface $\partial\Omega$ is referred to as the Helmholtz problem.

Analytic closed form solutions for the 3D case can usually only be found for special problems with certain regular geometries where a separation of variables can be performed. However, in most practical 3D cases, numerical methods are the only way to proceed. Frequently the finite difference or finite element approaches are used to find the approximations of the solution on a set of points in the region of interest. But because of their relatively low approximation order, for the problem of precise solution of PDEs, the methods have very limited success because of the prohibitively large number of mesh points required. For Computer Codes (Design, Simulation, Field Calculation)

reference, codes like the frequently used TOSCA [1, 2] can usually solve 3D Laplace problems with a relative accuracy of 10^{-4} with meshes of size about 10^{-6} [3]. Furthermore, direct validation of such methods is often very difficult.

In the following we develop a new method based on the Helmholtz theorem and the Taylor model methods[4, 5] and the corresponding tools in the code COSY Infinity [6, 7] to find a validated solution of the Laplace equation starting from the field boundary data. The final solution is provided as a set of local Taylor models, each of which represents an enclosure of a solution for a sub-box of the volume of interest.

THEORY AND IMPLEMENTATION

The Helmholtz Approach

We begin by representing the solution of the Laplace equation via the Helmholtz vector decomposition theorem [8, 9, 10, 11, 12, 13], which states that any vector field \vec{B} which vanishes at infinity can inside Ω be written as the sum of two terms

$$\vec{B}(\vec{x}) = \vec{\nabla} \times \vec{A}_t(\vec{x}) + \vec{\nabla} \phi_n(\vec{x}), \text{ where} \quad (2)$$

$$\phi_n(\vec{x}) = \frac{1}{4\pi} \int_{\partial\Omega} \frac{\vec{n}(\vec{x}_s) \cdot \vec{B}(\vec{x}_s)}{|\vec{x} - \vec{x}_s|} ds - \frac{1}{4\pi} \int_{\Omega} \frac{\vec{\nabla} \cdot \vec{B}(\vec{x}_v)}{|\vec{x} - \vec{x}_v|} dV$$

$$\vec{A}_t(\vec{x}) = -\frac{1}{4\pi} \int_{\partial\Omega} \frac{\vec{n}(\vec{x}_s) \times \vec{B}(\vec{x}_s)}{|\vec{x} - \vec{x}_s|} ds + \frac{1}{4\pi} \int_{\Omega} \frac{\vec{\nabla} \times \vec{B}(\vec{x}_v)}{|\vec{x} - \vec{x}_v|} dV.$$

Here $\partial\Omega$ is the surface which bounds the volume Ω . \vec{x}_s denotes points on the surface $\partial\Omega$, and \vec{x}_v denotes points within Ω . \vec{n} is the unit vector perpendicular to $\partial\Omega$ that points away from Ω , and $\vec{\nabla}$ denotes the gradient with respect to \vec{x}_v .

The first term is usually referred to as the solenoidal term, and the second term as the irrotational term. Because of the apparent similarity of these two terms to the well-known vector- and scalar potentials to \vec{B} , we note that in the above representation, it is in general not possible to utilize only one of them; for a given problem, in general both ϕ_n and \vec{A}_t will be nonzero.

For the special case that $\vec{B} = \vec{\nabla}V$, we have $\vec{\nabla} \times \vec{B} = 0$; furthermore, if V is a solution of the Laplace equation $\vec{\nabla}^2 V = 0$, we have $\vec{\nabla} \cdot \vec{B} = 0$. Thus in this case, all the volume integral terms vanish, and $\phi_n(\vec{x})$ and $\vec{A}_t(\vec{x})$ are

completely determined from the normal and the tangential components of \vec{B} on the surface $\partial\Omega$ via

$$\phi_n(\vec{x}) = \frac{1}{4\pi} \int_{\partial\Omega} \frac{\vec{n}(\vec{x}_s) \cdot \vec{B}(\vec{x}_s)}{|\vec{x} - \vec{x}_s|} ds$$

$$\vec{A}_t(\vec{x}) = -\frac{1}{4\pi} \int_{\partial\Omega} \frac{\vec{n}(\vec{x}_s) \times \vec{B}(\vec{x}_s)}{|\vec{x} - \vec{x}_s|} ds$$

For any point within the volume Ω , the scalar and vector potentials and consequently the solution of the Laplace equation depend only on the field on the surface $\partial\Omega$.

Using the fact that if $\vec{x} \neq \vec{x}_s$, we have $\vec{\nabla}(1/|\vec{x} - \vec{x}_s|) = -(\vec{x} - \vec{x}_s)/|\vec{x} - \vec{x}_s|^3$, and similar relationships, it is possible to explicitly obtain the gradient of the scalar potential, and with some more work the curl of the vector potential; the results have the explicit form

$$\vec{\nabla}\phi_n(\vec{x}) = -\frac{1}{4\pi} \int_{\partial\Omega} \frac{(\vec{x} - \vec{x}_s) (\vec{n}(\vec{x}_s) \cdot \vec{B}(\vec{x}_s))}{|\vec{x} - \vec{x}_s|^3} ds \quad (3)$$

$$\vec{\nabla} \times \vec{A}_t(\vec{x}) = \frac{1}{4\pi} \int_{\partial\Omega} \frac{(\vec{x} - \vec{x}_s) \times (\vec{n}(\vec{x}_s) \times \vec{B}(\vec{x}_s))}{|\vec{x} - \vec{x}_s|^3} ds \quad (4)$$

From eq.(2) we know that the field inside the volume of interest is just a sum of the irrotational(eq.(3)) and the solenoidal(eq.(4)) part. This is then the solution for the magnetic field as surface integrals. But to numerically integrate the kernel and get the validated solution as the local Taylor model we need a specialized numerical scheme. In the next section we introduce one such scheme based on the Taylor models of the code COSY Infinity[6, 7]. We quickly introduce the definition of the Taylor model and discuss briefly the anti-derivation operation on the Taylor models which will be extensively used in implementation of the scheme. We then proceed to explain the numerical scheme to perform the surface integration.

Solution of the Helmholtz Problem using Taylor Models

In the following, we develop a validated method based on Taylor model methods to determine sharp enclosures of the field \vec{B} and the potential ψ utilizing the Helmholtz method.

Definition (Taylor Model) Let $f : D \subset R^v \rightarrow R$ be a function that is $(n + 1)$ times continuously partially differentiable on an open set containing the v -dimensional domain D . Let x_0 be a point in D and P the n -th order Taylor polynomial of f around x_0 . Let I be an interval such that

$$f(x) \in P(x - x_0) + I \text{ for all } x \in D$$

Computer Codes (Design, Simulation, Field Calculation)

and that has the property that I scales with the $(n + 1)$ st power of the width of D . Then we call the pair (P, I) an n -th order Taylor model of f around x_0 on D .

A full theory of Taylor model arithmetic for elementary operations, intrinsic functions, initial value problems and functional inversion problems has been developed; see [5, 4] and references therein. Details about the validated implementation of arithmetic operation in COSY can be found in [14, 4]. For the purposes of the further discussion, one particular "intrinsic" function, the so-called antiderivation, plays an important role. We note that a Taylor model for the integral with respect to variable i of a function f can be obtained from the Taylor model (P, I) of the function by merely integrating the part P_{n-1} of order up to order $n - 1$ of the polynomial, and bounding the n -th order into the new remainder bound. Specifically, we have

$$\partial_i^{-1}(P, I) = \left(\int_0^{x_i} P_{n-1}(x) dx_i, (B(P - P_{n-1}) + I) \cdot (b_i - a_i) \right)$$

More details about the implementation of the anti-derivation operation can be found in [15].

Utilizing Taylor model arithmetic, the following algorithm now allows to solve the Laplace equation for the Helmholtz problem.

1. Discretize the surface $\partial\Omega$ into individual surface cells S_i with centers s_i and the volume Ω into volume cells V_j with centers v_j .
2. Pick a volume cell V_j .
3. For each surface cell S_i , evaluate the integrands in eq. (3) and (4), the so-called "kernels", in Taylor model arithmetic to obtain a Taylor model representations in BOTH the surface variables of S_i AND the volume variables of V_j , i.e. in a total of five variables.
4. Use the Taylor model anti-derivation operation twice to perform integration over the surface variables of each cell S_i .
5. Add up all results to obtain a three dimensional Taylor model enclosing the field \vec{B} over the volume cell V_j .
6. If a validated enclosure of the potential ψ to \vec{B} over the volume cell V_j is desired, integrate the field \vec{B} over any path using the anti-derivation operation.

As a result, for each of the volume cells V_j , Taylor model enclosures for the fields \vec{B} and potentials ψ are obtained. All the mathematical operations to evaluate these Taylor Models and surface integration are implemented using the Taylor Model tools available in the code COSY Infinity[6, 7].

Apparently the computational expense scales with the product of the number of volume elements and the number of surface elements; of these, the number of volume elements is more significant because of their larger number. In practice one observes that when using high-order

Taylor models, a rather small number of volume elements is required, in particular compared to the situation in conventional field solvers discussed above.

AN EXAMPLE: THE BAR MAGNET

The Example Field

As a reference problem to study the behavior of the method, we consider the magnetic field of rectangular iron bars of a uniformly magnetized material with inner surfaces ($y = \pm y_0$) parallel to the mid-plane $y = 0$ as shown in fig.1. The geometry of these uniformly magnetized bars, which are assumed to be infinitely extended in the $\pm y$ -directions, is defined by: $x_1 \leq x \leq x_2, |y| \geq y_0$, and $z_1 \leq z \leq z_2$. From this bar magnet one can obtain an analytic solution for the magnetic field $\vec{B}(x, y, z)$ - see for example [16, 17, 18] - and the result is given by

$$B_y(x, y, z) = \frac{B_0}{4\pi} \sum_{i,j=1}^2 (-1)^{i+j} \left[\arctan \left(\frac{X_i \cdot Z_j}{Y_+ \cdot R_{ij}^+} \right) + \arctan \left(\frac{X_i \cdot Z_j}{Y_- \cdot R_{ij}^-} \right) \right]$$

$$B_x(x, y, z) = \frac{B_0}{4\pi} \sum_{i,j=1}^2 (-1)^{i+j} \left[\ln \left(\frac{Z_j + R_{ij}^-}{Z_j + R_{ij}^+} \right) \right]$$

$$B_z(x, y, z) = \frac{B_0}{4\pi} \sum_{i,j=1}^2 (-1)^{i+j} \left[\ln \left(\frac{X_j + R_{ij}^-}{X_j + R_{ij}^+} \right) \right]$$

where $X_i = x - x_i, Y_{\pm} = y_0 \pm y, Z_i = z - z_i$, and $R_{\pm} = (X_i^2 + Y_{\pm}^2 + Z_i^2)^{\frac{1}{2}}$.

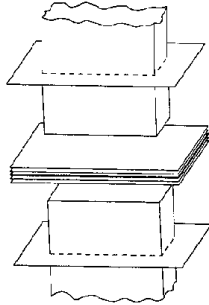


Figure 1: Geometric layout of the bar magnet, consisting of two bars of magnetized material.

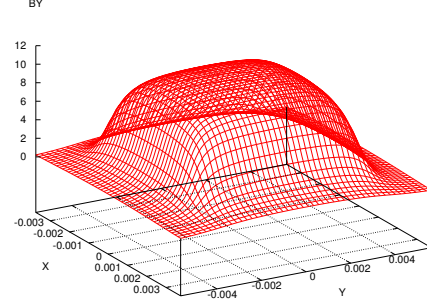


Figure 2: The magnetic field component B_y on the center plane of the bar magnet

Results and Analysis As a first step in the analysis of the influence of the discretization of the surface and volume on the result, we study the contributions of the surface elements towards the remainder interval part of the total integral. The volume expansion point is chosen as $\vec{r} = (.1, .1, .1)$, and the size of the volume box around it is chosen zero. Thus after the surface integration, the polynomial part of the dependence on volume vanishes except for the constant term, and the accuracy is only limited by the width of the surface element, which after integration over the surface variables influences the width of the remainder bound. We plot the width of the remainder interval versus surface element length for the scalar potential Fig.3. The center of the surface element is chosen as $\vec{r}_s = (.034, .011, .5)$. It is observed that for high orders, the method quickly reaches an accuracy of around 10^{-16} for about 2^5 surface subdivisions, which correspond to about $2^{10} \approx 1000$ surface element cells per surface. Under the assumption that each of these surface cells brings a similar contribution, the accuracy due to the surface discretization will be in the range of approximately $6 \cdot 1000 \cdot 10^{-16} < 10^{-12}$.

We now study the dependency of the polynomial part and width of the remainder interval of the magnetic field on the volume element length. In all these plots the surface element length is kept fixed at $1/128$. Figure 4 shows the remainder interval width for the y component of the magnetic field versus volume element lengths for different orders of computation. The other components of the magnetic field exhibit a similar behavior.

We see that a validated accuracy in the range of 10^{-4} can be achieved for a volume element width of around 10^{-1} , corresponding to a total of around 1000 volume elements. This number compares very favorably to the above-mentioned numbers for the commercial code TOSCA [1, 2]. An accuracy in the range of 10^{-7} can be achieved for a width of around $10^{-1.4}$, corresponding to a total of around 200,000 volume elements.

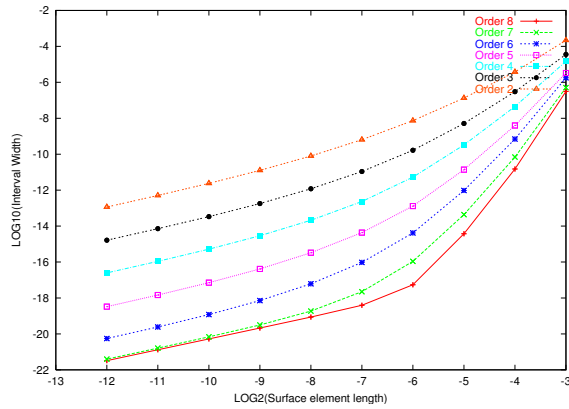


Figure 3: Remainder interval width versus surface element length for integration over a single surface element and vanishing volume size

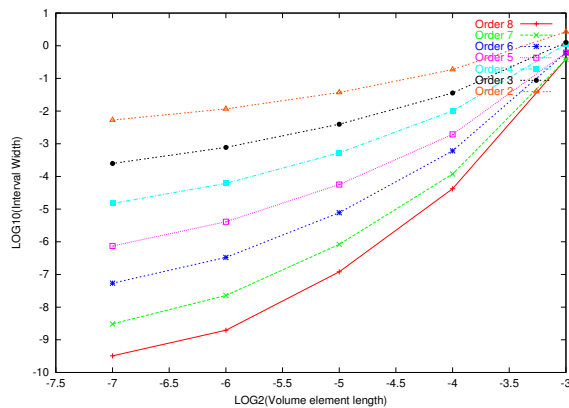


Figure 4: Remainder interval width vs length of volume element for y component of the magnetic field.

Overall, we see that the method of simultaneous surface and volume expansion of the Helmholtz integrals leads to validated tools for the solutions of ODEs which when executed in Taylor model arithmetic can lead to very sharp enclosures. It is obvious that the method can be generalized to other surface-integral based approaches to the solution of PDEs.

Acknowledgement

This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357.

REFERENCES

- [1] The toscia reference manual. Technical report.
- [2] The toscia user guide. Technical report.
- [3] W. Wan. Private communication.
- [4] K. Makino and M. Berz. Taylor models and other validated functional inclusion methods. *International Journal of Pure Computer Codes (Design, Simulation, Field Calculation)*

and *Applied Mathematics*, 6,3:239–316, 2003. available at <http://bt.pa.msu.edu/pub>.

- [5] K. Makino. *Rigorous Analysis of Nonlinear Motion in Particle Accelerators*. PhD thesis, Michigan State University, East Lansing, Michigan, USA, 1998. Also MSUCL-1093.
- [6] M. Berz, J. Hoefkens, and K. Makino. COSY INFINITY Version 8.1 - programming manual. Technical Report MSUHEP-20703, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2002. see also <http://cosy.pa.msu.edu>.
- [7] M. Berz and K. Makino. COSY INFINITY Version 8.1 - user's guide and reference manual. Technical Report MSUHEP-20704, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2002. see also <http://cosy.pa.msu.edu>.
- [8] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics, Part I and II*. 1953.
- [9] P.L. Walstrom. Soft-edged magnet models for higher-order beam-optics map codes. *Nucl. Instrum. Meth.*, A519, Issues 1-2:216–221, 2004.
- [10] M. Venturini, D. Abell, and A. Dragt. Map computation from magnetic field data and application to the lhc high-gradient quadrupoles. *eConf*, C980914:184–188, 1998.
- [11] P. Walstrom, A. Dragt, and T. Stasevich. Computation of charged-particle transfer maps for general fields and geometries using electromagnetic boundary-value data. Presented at IEEE Particle Accelerator Conference (PAC2001), Chicago, Illinois, 18-22 Jun 2001.
- [12] M. Venturini and A. Dragt. Computing transfer maps from magnetic field data. Contributed to IEEE Particle Accelerator Conference (PAC 99), New York, 29 Mar - 2 Apr 1999.
- [13] M. Venturini and A. J. Dragt. Accurate computation of transfer maps from magnetic field data. *Nucl. Instrum. Meth.*, A427:387–392, 1999.
- [14] N. Revol, K. Makino, and M. Berz. Taylor models and floating-point arithmetic: Proof that arithmetic operations are validated in COSY. *Journal of Logic and Algebraic Programming*, in print, 2004. University of Lyon LIP Report RR 2003-11, MSU Department of Physics Report MSUHEP-30212, <http://bt.pa.msu.edu/pub>.
- [15] M. Berz and K. Makino. Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliable Computing*, 4(4):361–369, 1998.
- [16] M. M. Gordon and T. Taivassalo. The z^4 orbit code and the focusing bar fields used in beam extraction calculations for superconducting cyclotrons. *Nuclear Instruments and Methods*, 247:423, 1986.
- [17] R. Degenhardt and M. Berz. High accuracy description of the fringe field in particle spectrographs. *Nuclear Instruments and Methods*, A427:151–156, 1999.
- [18] M. Berz. *Modern Map Methods in Particle Beam Physics*. Academic Press, San Diego, 1999. Also available at <http://bt.pa.msu.edu/pub>.

COMPUTATIONAL BEAM DYNAMICS FOR A HIGH INTENSITY RING: BENCHMARKING WITH EXPERIMENT IN THE SNS*

J. Holmes, S. Cousineau, and V. Danilov, ORNL, Oak Ridge, TN 37831, U.S.A.
Z. Liu, Indiana University, Bloomington, IN 47405, U.S.A.

Abstract

As the Spallation Neutron Source (SNS) continues to ramp toward full intensity, we are acquiring a wealth of experimental data. Much effort is being applied to understand the details of the beam accumulation process under a variety of experimental conditions. An important part of this effort is the computational benchmarking of the experimental observations. In order to obtain quantitative agreement between the calculations and the observations, and hence a full understanding of the machine, a great deal of care must be taken to incorporate all the relevant experimental parameters into the calculation. These vary from case to case, depending upon what is being studied. In some of these cases, the benchmarks have been critical in unearthing flaws in the machine and in guiding their mitigation. In this paper, we present the results of benchmarks with a variety of experiments, including coupling in beam distributions at low intensities, space charge effects at moderate intensities, and a transverse instability driven by the impedance of the ring extraction kickers.

INTRODUCTION

The Spallation Neutron Source continues to make impressive progress toward its full operating power of 1.44 MW. In the most recent run, SNS operated at a sustained power of 865 kW during production. At the current applied energy of 930 MeV, this corresponds to nearly 10^{14} protons on target per pulse. In recent dedicated high intensity studies, 1.55×10^{14} protons (24.8 μC) were injected stably into the ring, extracted, and transported to the target. This is the first time the SNS ring has exceeded its design beam intensity of 1.5×10^{14} protons per pulse. Although we are able to operate in production mode at 865 kW with acceptably low losses ($< 10^{-3}$ total beam loss and 10^{-4} uncontrolled beam loss), losses in the high intensity studies are much higher. In order to achieve acceptable losses as we continue to increase the beam intensity, we must gain a detailed understanding of the underlying beam dynamics.

Another reason to thoroughly understand the SNS ring beam dynamics is to avoid instabilities. In several studies, including the one that achieved the record beam intensity, we found that we can easily induce instabilities in the ring. In order to do so, a number of measures are typically taken. These include various combinations of the following: ring RF buncher voltages are modified, or

turned off altogether, so that coasting beams are accumulated; the choppers may be turned off to provide continuous beam with no gap; the chromatic sextupoles are activated in order to zero the ring chromaticity; and the ring tunes may be altered to induce the resistive wall instability. So far, three independent instabilities have been observed. The frequency signatures of these instabilities strongly suggest 1) a low frequency resistive wall instability at ~ 100 kHz, 2) a transverse (extraction kicker) impedance-induced instability in the 4–10 MHz range, and 3) a broad e-p instability in the 20–100 MHz range. The slow-growing resistive wall instability occurs only when the tunes are set below integer values, such as $5.8 < \nu_{x,y} < 6$. Because SNS is operated nominally with $\nu_x = 6.23$ and $\nu_y = 6.20$, the resistive wall instability will not be a problem. The extraction kicker instability has been observed only for a continuous coasting beam under the condition of zero chromaticity and with the beam stored for several milliseconds. It is not expected to be a problem for SNS as currently designed, but it could arise at the higher powers being considered for an SNS upgrade. The e-p instability has been observed the most frequently and under a wide variety of conditions. It is likely to present the greatest challenge as we push the intensity frontier. The observations of these instabilities have been discussed in Refs. [1,2], and preliminary simulation results have been shown in Refs. [3-5].

In order to gain a quantitative understanding of the beam dynamics in the SNS ring, it is necessary to apply careful numerical simulation. We carry this out using the ORBIT Code [6], which was written with high intensity rings and transfer lines in mind. We now present the results of benchmarks with a variety of experiments, including coupling in beam distributions at low intensities, space charge effects at moderate intensities, and the transverse instability driven by the impedance of the ring extraction kickers. We begin with a low-intensity study that revealed the presence of x-y coupling in the extraction septum magnet and follow this work through to its present state of benchmarking with low and intermediate intensity beams with injection painting. We then present the results of a careful study of the observed extraction kicker induced instability. Lacking from this paper will be any results of e-p instability simulation. We are actively pursuing this work on a number of fronts, but have not yet obtained results beyond those previously presented [3,4]. Throughout this paper we stress the necessity of attention to detail as well as the interplay between theory, experiment, and computation required for successful and accurate simulation.

* ORNL/SNS is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

ACCUMULATION

One of the first benchmarks attempted for the SNS ring was to simulate the accumulation process. The object of the benchmark was to compare measured wire scanner profiles of the extracted beam in the Ring to Target Beam Transport (RTBT) line with corresponding profiles from the simulations. The simulations meticulously matched the parameters used in the experiments. These included the injection stripper foil position and incident beam emittances, position, and angles; the painting waveforms of the injection kickers; the ring lattice settings and tunes; and the applied ring RF focusing. The runs were carried out using ORBIT's foil scattering model, a complete set of apertures and collimators, symplectic tracking, 1D longitudinal space charge, and 2.5D transverse space charge.

The initial tracking studies were conducted at low intensities, and were thus designed to test the accuracy of our single particle model of the ring, rather than collective effects. The simulation results bore no similarity to the measured profiles, which changed shape significantly, in both planes, over the several wire scanners in the RTBT. This, together with an unexplained tilt of the beam at the target view screen, suggested that there was x-y coupling somewhere upstream of the wire scanners. Because care had been taken to correct x-y coupling in the ring using the available skew quadrupole magnets, we focused on the large and complicated extraction septum magnet, which simultaneously bends the beam vertically down and horizontally left from the ring into the RTBT. J.G. Wang then modeled this magnet using the OPERA 3D Code [7]. He found many higher order multipoles, including strong sextupole components [8].

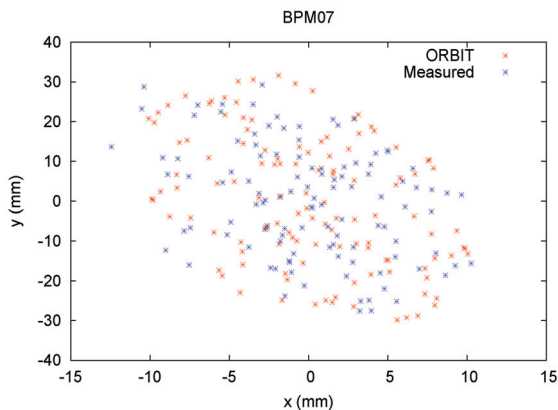


Figure 1. Measured and simulated distributions of single turn injected pulses stored for varying times.

These calculated multipoles were then included in the ORBIT simulation, and a careful experiment was performed. Experimentally, single turn injection was carried out and the pulses were stored for various durations up to 120 turns. They were then extracted and their positions were recorded on the several beam position

monitors (BPMs) in the RTBT. In this way, the signals from the single pulses were used to build up a distribution at each BPM. The resulting experimental distributions were then compared with one from an ORBIT simulation in which a single macroparticle was injected each turn for 120 turns. Figure 1 shows a comparison between the experimental and simulation results at BPM07 in the RTBT. Both distributions have substantially the same size, shape, and orientation, thus confirming both the presence of coupling and the extraction septum magnet as its source. Without the coupling multipoles, the distribution is a rectangle without tilt in the x-y plane. Similar results were found at BPMs throughout the RTBT. As a result of this initial benchmark, the extraction septum magnet was modified to remove the higher order multipoles and the consequent coupling [8]. This illustrates the value of careful benchmarking in finding and diagnosing problems with real machine components.

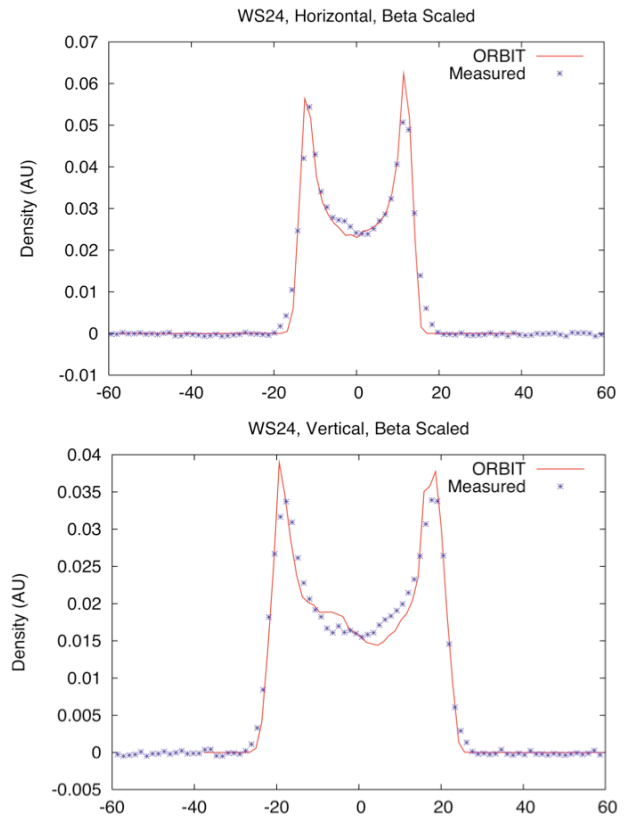


Figure 2. Measured and calculated horizontal and vertical beam profiles at wire scanner WS24 in the RTBT for a low intensity beam of 8.6×10^{12} protons per pulse.

After the modification of the extraction septum, the benchmark of the injection process was repeated. The simplest cases used flat-topped injection kicker waveforms, so that the painting was constant. In both the experiments and the benchmarks the injected beam spot at the stripper foil was 18 mm, both horizontally and vertically, from the closed orbit of the circulating beam. Beams were accumulated in the ring for various times,

ranging from 80 to 460 turns, and corresponding intensities, ranging from 8.6×10^{12} to 5.3×10^{13} protons per pulse. Thus, these benchmarks were carried out at low to moderate beam intensity. The beams were extracted after two turns of storage and then measured in the RTBT. Figure 2 shows a comparison of the measured and calculated horizontal and vertical beam profiles at the low intensity of 8.6×10^{12} protons per pulse taken at wire scanner WS24 in the RTBT. The agreement is very good, with ORBIT matching the widths and hollow shape, due to off-axis injection, of the actual profiles. Minor observed differences include slight tails on the experimental data and slightly different density fluctuations in the profiles at the beam center. Comparably good agreement was obtained at the other wire scanners in the RTBT.

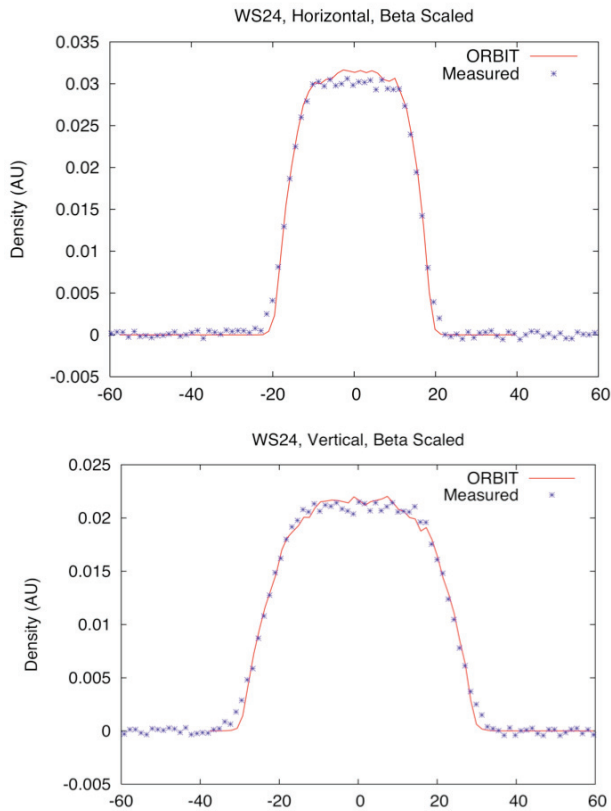


Figure 3. Measured and calculated horizontal and vertical beam profiles at wire scanner WS24 in the RTBT for a moderate intensity beam of 5.3×10^{13} protons per pulse.

Good agreement between experiment and simulation was found over the entire range of intensities. Figure 3 shows the comparison of measured and calculated beam profiles at wire scanner WS24 for 5.3×10^{13} protons per pulse. In comparison with the low intensity profiles, accumulated over 80 turns, the moderate intensity profiles, accumulated for 460 turns, fill in the centers and broaden slightly due primarily to space charge. The broadening is not substantial, however, because the

maximum incoherent space charge tune shifts, estimated using ORBIT, are only about 0.07, so the coherent tune shift is perhaps 0.04. With the bare tunes of the ring set at $\nu_x = 6.23$ and $\nu_y = 6.20$, there is not sufficient space charge in the beam to activate the half integer resonance at $\nu = 6$ [9]. In addition to profiles, measurements were also taken of the RMS emittances of the beams. These are compared with the calculated RMS emittances from the simulations in Table 1, in units of millimeter-milliradians. As with the beam profiles, the agreement is quite good, falling within 5% for all cases. It is interesting that, in spite of the different shapes of the low and high intensity beams, the RMS emittances change little with intensity.

Table 1: RMS Emittances, Experiment and Simulation

Case	Expmnt (H,V)	ORBIT (H,V)
80 Turns	(13.3,13.1)	(13.7,12.6)
460 Turns	(13.8,13.1)	(13.2,12.5)

The studies for injection with painting were carried out in a somewhat different fashion than those with flat-topped injection kickers. In all cases with painting, 640 turns of beam accumulation and two turns of storage were performed prior to extraction. The intensity was varied by beam decimation, in which pulses are injected only on a fraction of available turns. In this way, intensities ranging from 8.2×10^{12} to 7.5×10^{13} protons per pulse were obtained. As in the flat top case, these benchmarks are still in the range of low to moderate intensity. The initial results showed less agreement between experiment and simulation than for the flat-topped case, especially at low beam intensity, where the simulated case resulted in more hollow and slightly narrower profiles than the experimental case. At moderate intensity, the agreement was better, but the disagreement at low intensity was a cause for concern. Because the agreement in the flat top injection cases was so good, we focused on the injection kicker waveforms. The simulations faithfully implemented the programmed waveforms from the control room, so an experiment was carried out to test the experimental waveforms. Single turn injection was carried out with varying delay times between the initiation of the injection kickers and the injected pulse. Then the BPMs were used to measure the displacement between the injected pulse and the closed orbit in the ring at the time of injection. In this way, the actual injection waveforms were determined. Not surprisingly, these did not match the programmed waveforms. The reason turned out to be that the timings of the injection kickers were not synchronized with each other or with the beam, resulting in delays of as many as 60 turns. This problem is now being rectified. Once again, careful benchmarking provided guidance in finding and diagnosing faulty hardware performance.

In order to “close the loop” on this injection benchmark, the painting experiment was repeated, but this time the actual measured kicker waveforms were used in the simulations. Figure 4 shows the resulting profiles in the vertical direction at wire scanner WS24, both for low and moderate intensities. The moderate intensity results

are good, with the measured profiles just slightly broader than the calculations. At low intensity the agreement, although better than before (compare red and green curves with blue experimental data), is not yet perfect. As we continue to improve these injection benchmarks, our next efforts will be to repeat the painting experiment after the injection kickers have been corrected, and also to repeat the flat top experiment with different kicker amplitudes.

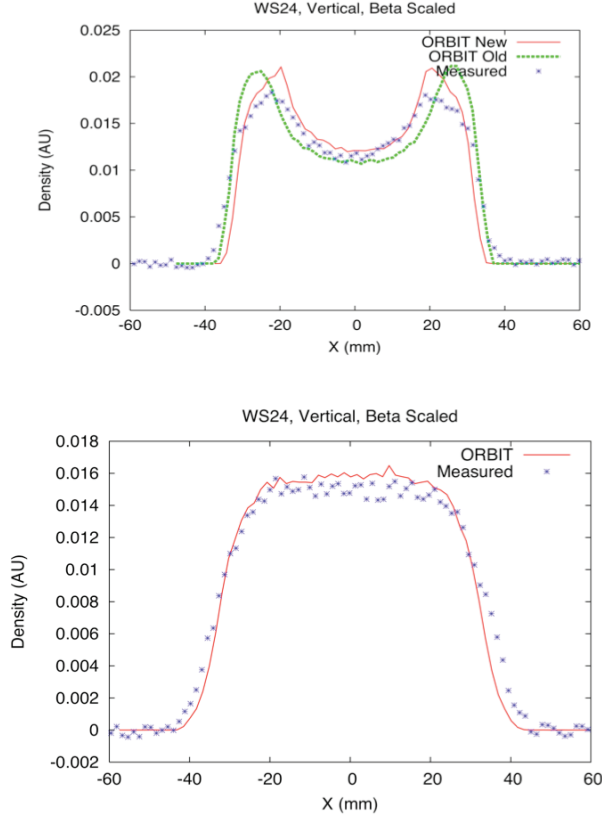


Figure 4. Measured and calculated vertical beam profiles at wire scanner WS24 in the RTBT for low (8.2×10^{12} protons per pulse) and moderate (7.5×10^{13} protons per pulse) intensity beams.

Before leaving this subject, we again remark on the difference between the low and moderate intensity beam profiles. Calculation of the incoherent space charge tune shifts using ORBIT shows maximum shifts of about 0.07 for the moderate intensity case. As in the flat top injection, this is insufficient to excite the half integer resonance. Accordingly, we see neither substantial beam broadening nor an increase in emittances with increasing intensity. Rather, space charge tends to “smooth out” the profiles by filling in the hollow center and adding slightly to the width. As we push to higher intensities and greater tune shifts, we expect to see broadening of the profiles due to space charge [9].

EXTRACTION KICKER INSTABILITY

The Spallation Neutron Source accumulator ring was designed and constructed to be stable at the full intensity

of 1.5×10^{14} protons. Because early estimates predicted that the extraction kickers would present the dominant ring impedance, they were carefully designed to minimize that impedance. As a result, stability calculations for the extraction kicker impedance showed longitudinal stability up to 8×10^{14} protons, while transverse stability at 1.5×10^{14} protons was predicted for up to 3 to 4 times the known impedance [10]. These results were obtained computationally for bunched beams using ORBIT. However, for coasting beams in SNS, analytic and ORBIT calculations for mode number $n = 10$ and 1.5×10^{14} protons predict vertical instability when $\text{Re}(Z) > 0 \text{ k}\Omega/\text{m}$ at zero chromaticity and when $\text{Re}(Z) \geq 250 \text{ k}\Omega/\text{m}$ at natural chromaticity. The measured impedance of the extraction kickers in the vicinity of $n = 10$ is $\text{Re}(Z) \sim 25\text{--}30 \text{ k}\Omega/\text{m}$. It is therefore appropriate to use coasting beams and corrected chromaticity to look for this instability.

The extraction kicker instability has been observed in the course of high intensity beam studies. The scenario was the following: The ring tunes were set at $Q_x = 6.23$ and $Q_y = 6.20$. The chromaticity was corrected to zero and the RF buncher cavities were turned off. The choppers were also turned off so that a continuous coasting beam was accumulated. An 860 MeV beam of 7.7×10^{13} protons, more than $12 \mu\text{C}$, was injected for 850 turns and subsequently stored until the beam was lost in the ring. The evolution of the beam was followed for 10000 turns. The observed instability began at about 1200 turns and saturated somewhat beyond 4000 turns. It was active in the transverse vertical direction with dominant harmonic at 6 MHz and noticeable excitation in the $4 \rightarrow 10$ MHz range. Interpreting this to be a “slow” mode, the frequency is consistent with dominant harmonic $n = 12$, and excitation in the range $10 \leq n \leq 16$. This agrees well with the predicted range of dominant unstable mode numbers from the extraction kicker impedance.

The experimental results for this case have been presented in Ref. [1], and the growth rate of the 6 MHz ($n = 12$) harmonic was used to theoretically infer the extraction kicker impedance at that frequency. The resulting prediction of $28 \text{ k}\Omega/\text{m}$ is in excellent agreement with the laboratory-measured impedance of $25 \text{ k}\Omega/\text{m}$. The intent of this paper is to precisely simulate this case, using ORBIT, and to match all known experimental details as closely as possible. The parameters used in the earlier simulations, presented in Refs. [3–5], differ somewhat from those of the experiment, so that quantitative comparison is not appropriate for those cases.

The present simulation was carried out using the ORBIT code [6]. We used the actual ring settings with $Q_x = 6.23$ and $Q_y = 6.20$ and zero chromaticity. The ring RF cavities were turned off and a continuous coasting beam of 7.7×10^{13} protons at 860 MeV was injected for 850 turns and stored up to 10000 turns. The injected beam RMS energy spread was taken to be 0.5 MeV, consistent with observation, and the nominal SNS transverse injection painting was employed. Tracking was carried out using symplectic single particle transport, the laboratory-measured longitudinal and transverse

impedances for the extraction kickers, and the 3D space charge model. In addition, the ORBIT foil scattering model was activated and a complete set of apertures was included to incorporate beam losses during accumulation and storage. The number of macroparticles in the simulation was 4.25 million.

One of the more impressive results presented in Ref. [1] was the agreement between the extraction kicker impedance calculated from the growth rate of the instability and that measured in the laboratory. The relationship between the impedance and the growth rate is

$$\text{Re}(Z) = \frac{2\gamma\beta^2 E_0}{\tau\beta_{\text{twiss}} I_{\text{avg}}} \quad (1)$$

where Z is the impedance in Ω/m , γ and β are the relativistic factors, E_0 is the proton mass in eV, τ is the growth time in turns, β_{twiss} is the beta function at the location of the impedance, and I_{avg} is the average beam current. There are 14 extraction kickers distributed over a length of about 10 m in the SNS ring. The vertical beta function at the kickers varies from a minimum of 6.4 m to a maximum value of 13.5 m, with an average value of 9.3 m. The measurement gave an experimental growth time of 1036 turns. In estimating the value of 28 k Ω/m using Eq. (1), a value $\beta_{\text{twiss}} = 7$ m, close to the average beta of the ring, was assumed. If, instead, the average value of $\beta_{\text{twiss}} = 9.3$ m for the kickers is used, the estimated impedance is 21 k Ω/m , a bit lower than the lab value of 25 k Ω/m . However, Eq. (1) was derived under the assumptions that we are far from threshold and that the energy distribution is a delta function, thus ignoring Landau damping. Both these assumptions overestimate the growth of a real beam with energy spread. Therefore, the experimental growth time should be longer than that from Eq. (1), and its use in Eq. (1) should result in an impedance prediction that is somewhat lower than the actual value.

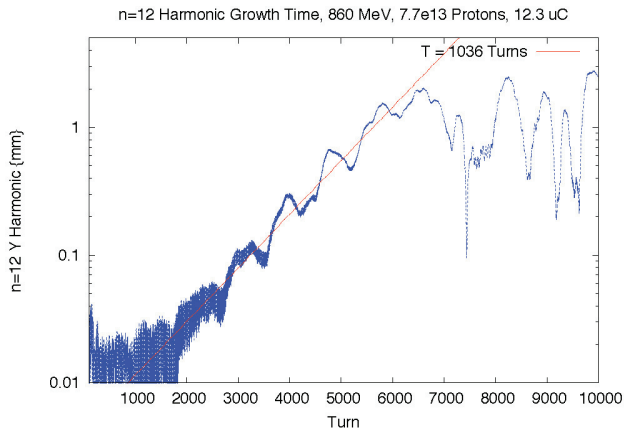


Figure 4. Vertical $n=12$ harmonic versus turn number in ORBIT extraction kicker instability simulation.

The ORBIT simulation was carried out with a single extraction kicker impedance node using the experimentally measured impedance values and placed at a position among the extraction kickers where the beta

function satisfies $\beta_{\text{twiss}} = 9.3$ m. The result, shown in Fig. 4, is an exponential growth time for the $n=12$ harmonic that is completely consistent with the measured time of 1036 turns. This impressive result is an important testimony to the necessity of getting all the details correct when performing a quantitative comparison between experiment and simulation. In reaching the result of this simulation, we made a number of false starts [3-5]. Erroneous assumptions included the use of chopped beams, the use of (too large) impedances from a previous design of the extraction kickers, and the placement of the impedance node at the geometric center of the extraction kickers rather than at a location with the average beta function. With hindsight, mistakes such as these appear to be foolish. However, in simulating a complicated particle accelerator, there are many details, each of which can affect the results. As each of these errors was rectified, the simulation improved, until we now achieve the correct growth rate for the instability.

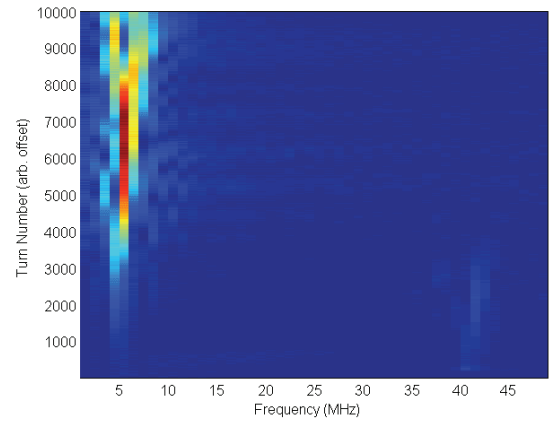


Figure 5. Evolution of experimental turn-by-turn vertical harmonic spectrum of the extraction kicker instability.

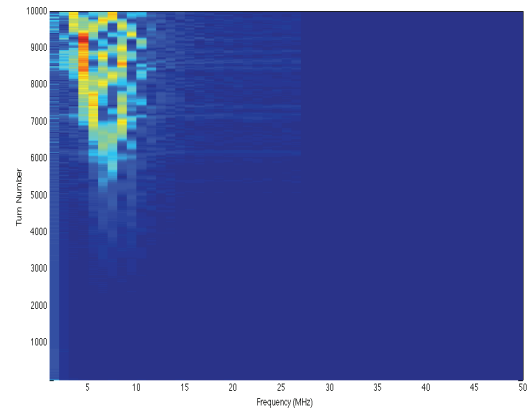


Figure 6. Evolution of simulated turn-by-turn vertical harmonic spectrum of the extraction kicker instability.

We have now completed the calculation to 10000 turns. The evolution of the simulated turn-by-turn spectrum shown in Fig. 6 displays activity over a similar range of frequencies as the experimental spectrum shown in Fig. 5, although the simulation shows somewhat more spreading than the experiment. Also, most of the activity occurs

after 5000 turns, which is after the completion of the linear growth. Should we wish to pursue this comparison of the nonlinear stage further, there are a few uncertainties we will need to address. The experimental signals were extracted from an FFT of high bandwidth BPM data. The simulated signals are harmonics of the beam centroid oscillations, in millimeters. Are they directly comparable? We will also need to compare the beam loss over 10000 turns in the experiment, which can vary from shot to shot, with that in the simulations, which is calculated using ORBIT's aperture and collimator routines. These are possibilities for future work.

CONCLUSIONS

As we ramp to full power operation in SNS, we are simulating much experimental data in order to gain a thorough understanding of the machine's beam dynamical behavior and to diagnose problems when they arise. Benchmarking has already allowed us to identify significant x-y coupling in the original ring extraction septum magnet and also improper performance of the injection kicker painting waveforms. Significantly, we are obtaining good agreement between simulated and experimental results. Thus far, our benchmarks of the ring injection process have been limited to low and medium intensities. In these cases, the main effect of space charge is to fill in the hollow central region of the beam. So far, the space charge tune shifts are insufficient to cause beam broadening through the half integer resonance. We have also completed a careful benchmark of the extraction kicker impedance instability, and have found that the calculated and experimental growth rates are in perfect agreement. Comparison of the spectral evolution of the experiment and simulation out to 10000 turns shows qualitatively similar results. However, the detailed evolution in the nonlinear stage of the instability after 5000 turns is somewhat different. Finally, although not presented here, we have begun simulation of a wealth of e-p instability data acquired during dedicated high intensity shifts. This data includes systematic variation of e-p activity with longitudinal beam profile, controlled by varying the relative phases of the ring RF cavities.

REFERENCES

- [1] V. Danilov, S. Cousineau, A. Aleksandrov, S. Assadi, W. Blokland, C. Deibele, S. Henderson, J. Holmes, M. Plum, and A. Shishlo, "Accumulation of High Intensity Beam and Observations of First Instabilities in the SNS Accumulator Ring," in *Proceedings of the 41st ICFA Advanced Beam Dynamics Workshop on High-Intensity, High-Brightness Hadron Beams (HB2006)*, Tsukuba, Japan, May, 2006.
- [2] S. Cousineau, "Instability Observations in the SNS Accumulator Ring," in *Proceedings of the 42nd ICFA Advanced Beam Dynamics Workshop on High-Intensity, High-Brightness Hadron Beams (HB2008)*, Nashville, August, 2008.
- [3] J. A. Holmes, S. Cousineau, V. Danilov, and A. Shishlo, "Computational Beam Dynamics Studies of Collective Instabilities Observed in SNS," in *Proceedings of the European Particle Accelerator Conference (EPAC08)*, Genoa, Italy, 2008.
- [4] J. A. Holmes, S. Cousineau, V. Danilov, M. Plum, and A. Shishlo, "High Intensity Effects in the SNS Accumulator Ring," in *Proceedings of the 42nd ICFA Advanced Beam Dynamics Workshop on High-Intensity, High-Brightness Hadron Beams*, Nashville, 2008,
- [5] J. A. Holmes, S. Cousineau, V. Danilov, and Z. Liu, "ORBIT Benchmark of Extraction Kicker Instability Observed in SNS," in *Proceedings of the Particle Accelerator Conference (PAC09)*, Vancouver, Canada, 2009.
- [6] J.A. Holmes, S. Cousineau, V.V. Danilov, S. Henderson, A. Shishlo, Y. Sato, W. Chou, L. Michelotti, and F. Ostiguy, in *The ICFA Beam Dynamics Newsletter*, Vol. 30, 2003.
- [7] Vector Fields: Software for Electromagnetic Design, "TOSCA", Cobham Group Company, 2005.
- [8] J.G. Wang, *Phys. Rev. ST Accel. Beams* 12, (2009) 042402.
- [9] S. Cousineau, S. Y. Lee, J. Holmes, V. Danilov, and A. Fedotov, *Phys. Rev. ST Accel. Beams* 6, (2003) 034205.
- [10] J. A. Holmes, V. Danilov, and L. Jain, "Transverse Stability Studies of the SNS Ring," in *Proceedings of the Particle Accelerator Conference (PAC05)*, Knoxville, June 2005, TPAT032, (2005).

COMPARISON OF DIFFERENT SIMULATION CODES WITH UNILAC MEASUREMENTS FOR HIGH BEAM CURRENTS

L. Groening, W. Barth, W. Bayer, G. Clemente, L. Dahl, P. Forck, P. Gerhard,
I. Hofmann, M.S. Kaiser, M. Maier, S. Mickat, T. Milosic, G. Riehl, H. Vormann,
S. Yaramyshev, GSI, D-64291 Darmstadt, Germany
D. Jeon, SNS, ORNL, Oak Ridge, TN 37831, USA
D. Uriot, CEA IRFU, F-91191 Gif-sur-Yvette, France
R. Tiede, Goethe University, Frankfurt a.M., Germany

Abstract

The GSI Universal Linear Accelerator UNILAC can accelerate all ion species from protons to uranium. Hence its DTL section is equipped with e.m. quadrupoles allowing for a wide range of field strength along the section. During the last years various campaigns on the quality of high current beams at the DTL exit as a function of the applied transverse focusing have been performed. Measurements were compared with up to four different high intensity beam dynamics codes. Those comparisons triggered significant improvement of the final beam quality. The codes were used to prepare an ambitious and successful beam experiment on the first observation of a space charge driven octupolar resonance in a linear accelerator.

INTRODUCTION

In the last decades many beam dynamics computer codes were developed [1] in order to simulate emittance growth along a linac. Several benchmarking studies among codes have been performed [2, 3, 4] generally assuming idealized conditions as initial Gaussian distributions, equal transverse emittances, matched injection into a periodic lattice, and small longitudinal emittance with respect to the rf-bucket size. In case of an operating linac generally not all of these conditions are met. To apply simulation codes to a realistic environment a benchmark activity was started aiming at simulations of beam emittance measurements performed at a DTL entrance and exit, respectively. The studies were performed at the GSI UNILAC [5]. For the simulations four different codes were used: DYNAMION [6], PARMILA [7], TraceWin [8], and LORASR [9].

The first benchmarking was done with moderate mismatch with respect to the periodic DTL solution. The zero current transverse phase advance σ_o has been varied from 35° to 90° . A detailed description of this first campaign is given in [10]. A second campaign suggested in [11] aimed at exploration of the 90° stop-band by varying σ_o from 60° to 130° . In this campaign the mismatch was minimized in order to mitigate the effects of the envelope instability. Ref. [12] is dedicated to this campaign.

EXPERIMENT SET-UP AND PROCEDURE

Intense beams are provided by a MUCIS source at low charge states with the energy of 2.2 keV/u. An RFQ fol-
Beam Dynamics, Other

lowed by two IH-cavities (HSI section) accelerates the ions to 1.4 MeV/u using an rf-frequency of 36 MHz. A subsequent gas-stripper increases the average charge state of the ion beam. Final acceleration to 11.4 MeV/u is done in the Alvarez DTL section operated at 108 MHz. The increase of rf-frequency by a factor of three requires a dedicated matching section preceding the DTL. It comprises a 36 MHz re-buncher for longitudinal bunch compression, a 108 MHz re-buncher for final bunch rotation, a quadrupole doublet for transverse compression, and a quadrupole triplet for final transverse beam matching.

The Alvarez DTL comprises five independent rf-tanks. Transverse beam focusing is performed by quadrupoles in the F-D-D-F mode. Each drift tube houses one quadrupole. The periodicity of the lattice is interrupted by four inter-tank sections, where D-F-D focusing is applied. Acceleration is done -30° from crest in the first three tanks and -25° from crest in the last two tanks.

Figure 1 presents the schematic set-up of the experiments. Beam current transformers are placed in front of

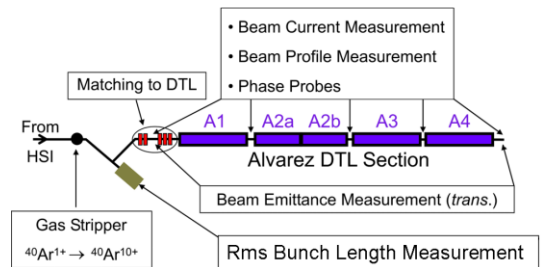


Figure 1: Schematic set-up of the experiments.

and behind the DTL as well as horizontal and vertical slit/grid emittance measurement devices. The total accuracy of each rms emittance measurement including its evaluation is estimated to be 10%. A set-up to measure the longitudinal rms bunch length is available in front of the DTL [13]. It measures the time of impact of a single ion on a foil. This time is related to a 36 MHz master oscillator. The resolution is 0.3° (36 MHz). Prior to the high intensity measurements a scan with very low beam current was done, demonstrating that no emittance growth occurs in absence of space charge forces. Afterwards the HSI was set to obtain 7.1 mA of 40Ar^{10+} in front of the DTL. Horizontal and vertical phase space distributions were measured in front of the DTL. The longitudinal rms bunch length was

measured at the entrance to the DTL. The DTL quadrupoles were set to the required zero current transverse phase advance σ_o . Due to space charge the phase advances in all three dimensions were depressed. The transverse depression reached from 14% (130°) to 43% (35°). Afterwards the quadrupoles and re-bunchers preceding the DTL were set to obtain full transmission and to minimize low energy tails of the beam. For each value of σ_o horizontal and vertical beam emittances were measured at the exit of the DTL with a resolution of 0.8 mm in space and 0.5 mrad in angle.

Each emittance measurement delivers a two dimensional matrix of discrete slit-positions and discrete angles. The data are processed by the measurement & evaluation program PROEMI [14]. Simulations deliver a set of six dimensional particle coordinates. This ensemble is projected onto a pixel-grid having the same characteristics as the slit/grid device used for the measurements. The grid is read by the measurement evaluation program PROEMI such that data reduction was done in the same way as for measured data.

INPUT FOR SIMULATIONS

The reconstruction of the initial distribution was done in two steps. First the rms Twiss parameters were determined. In the second step the type of distribution was reconstructed. The transverse rms measurements and the longitudinal rms measurements of the initial distribution, done at different locations along the beam line (Fig. 2) were combined in a self-consistent way based on envelope tracking of an rms equivalent KV-distribution [15]. The recon-

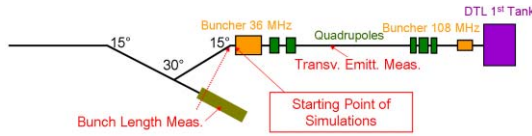


Figure 2: Matching section to the DTL including the reference points used for reconstruction of the initial phase space distribution.

struction of the type of distribution is based on evaluation of the brilliance curve, i.e. the fractional rms emittance as a function of the fraction [10]. Different amounts of halo have been found in the two transverse plane. For proper modelling of the initial distribution, both brilliance curves must be reproduced simultaneously. This was achieved by using a distribution function as

$$f(R) = \frac{a}{2.5 \cdot 10^{-4} + R^{10}}, \quad R \leq 1 \quad (1)$$

and $f(R) = 0$ for $R > 1$ with

$$R^2 = X^2 + X'^2 + Y^2 + Y'^2 + \Phi^2 + (\delta P/P)^2, \quad (2)$$

where a is the normalization constant and the constant in the denominator results from the cut off condition at $R = 1$. By defining the radius R using different powers for different sub phase spaces the halos within the planes could be Beam Dynamics, Other

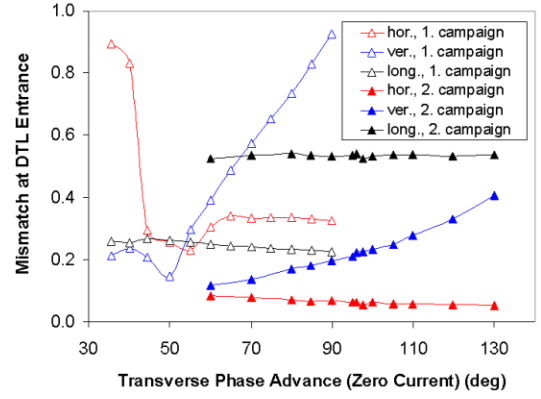


Figure 3: Mismatch between beam rms Twiss parameters and periodic Twiss parameters at injection into the DTL as a function of the phase advance σ_o for the two campaigns.

modelled. Since for the longitudinal phase space distribution no measurement but on the rms bunch length is available, a Gaussian distribution cut at 4σ is assumed. This can be approximated by setting the respective powers in the definition of R to a value of 2. It must be mentioned that the applied method is not sensitive for eventual correlations among different planes. Such correlations have been assumed to be zero.

MEASUREMENTS WITH MODERATE MISMATCH

The reconstructed initial distribution together with the applied setting along the matching section to the DTL was used to estimate the amount of mismatch [16] at the DTL entry using the DYNAMION code. As shown in Fig 3 the mismatch has been very small for intermediate σ_o and moderate in one plane for low and for high values of σ_o . For the applied σ_o from 35° to 90° full beam transmission was observed through the DTL in the experiment. The codes predicted losses of about 2%. Figure 4 displays final horizontal phase space distributions at the DTL exit as obtained from measurements and from simulations for three different values of σ_o . The simulated final distributions look quite similar. Simulations could reproduce the wings attached to the core measured at highest phase advances. But the codes did not show the asymmetric distributions measured at lowest phase advances. The simulated longitudinal phase spaces showed filamentation and emittance growth due to slight rf-bucket overflow at the DTL injection. Final transverse rms emittances are presented in Fig. 5 to Fig. 7 as a function of the transverse phase advance σ_o . The measurements and the simulations revealed lowest emittances at $\sigma_o \approx 60^\circ$. In general good agreement among the codes was found for the sum of the two transverse emittances but codes slightly underestimate the emittance growth. This is reasonable since the codes assume a machine without errors causing additional growth [17]. However, within single planes considerable

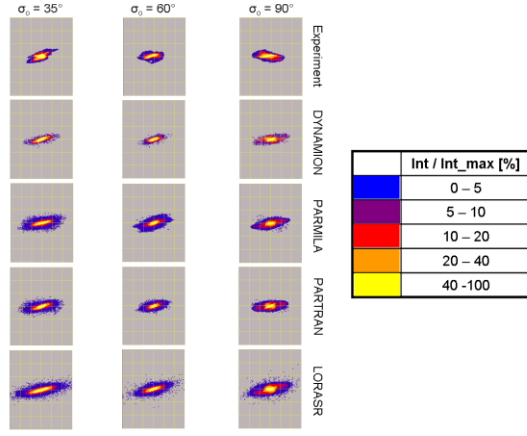


Figure 4: Top to bottom: horizontal phase space distributions at the DTL exit. Left: $\sigma_o = 35^\circ$; center: $\sigma_o = 60^\circ$; right: $\sigma_o = 90^\circ$. The scaling is ± 24 mm (horizontal axis) and ± 24 mrad (vertical axis), respectively.

differences among the codes have been found.

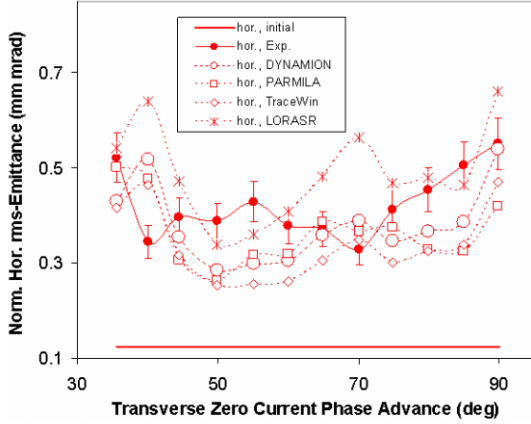


Figure 5: Horizontal rms emittance at the end of the DTL as a function of σ_o .

RMS MATCHED DTL-INJECTION

In order to minimize the mismatch to the DTL a dedicated computer code was created working in two steps. In the first step the periodic solution at the entrance to the DTL is calculated. Acceleration is neglected and the gap voltage and phase are adjusted such that the transverse defocusing is equal to the accelerating case. The DTL starts at a symmetry point of the periodic envelope and accordingly all three α -parameters, i.e. envelope slopes, are assumed to be zero. In the following an iterative procedure as depicted in Fig. 8 is applied. Initial values for the β -functions are used to track the envelope through one period. This is done by numerically solving the coupled system of differential equations from Ref. [15] describing the propagation of the beam's rms envelopes with space charge. As a solver the routine *rkqs* from [18] has been used. The next iteration Beam Dynamics, Other

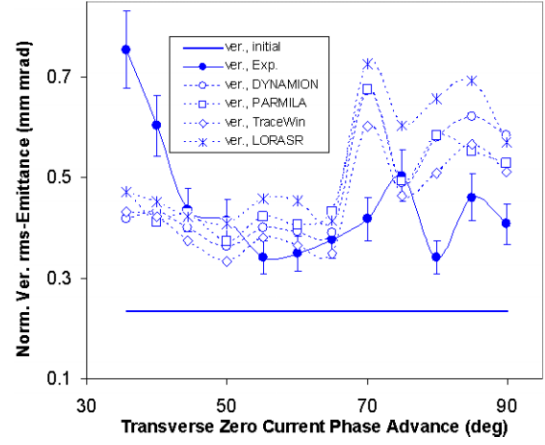


Figure 6: Vertical rms emittance at the end of the DTL as a function of σ_o .

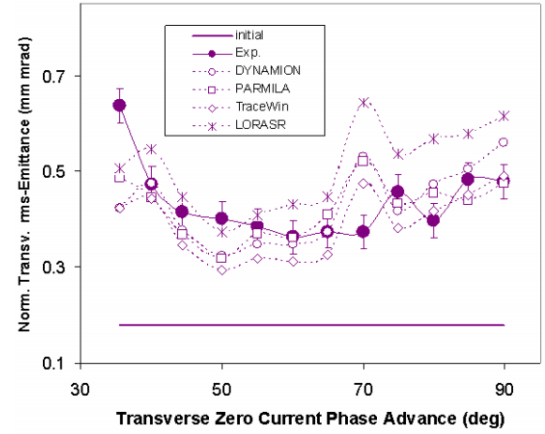


Figure 7: Mean value of horizontal and vertical rms emittance at the end of the DTL as a function of σ_o .

uses the mean of initial and final β -function as starting condition. Convergence is reached in less than 30 iterations. The second step of matching determines the settings of the focusing elements of the matching section (Fig. 2). Using the rms parameters of the reconstructed distribution at the entrance to the section as initial condition for the tracking equations, the final rms parameters at the DTL entrance depend on the focusing strengths f_i of the two bunchers and the five quadrupoles of the section. The final rms parameters together with the periodic solution define the mismatch M_i in each plane [16]. Matched injection is achieved if M is zero in each of the three planes. Defining

$$F(f_1 \dots f_7) = M_x^3(\dots) + M_y^3(\dots) + M_z^3(\dots) \quad (3)$$

as a function of the seven focusing strengths in the matching section, the matching is optimized for the setting $f_1 \dots f_7$ that minimizes F . Minimization of F is done with the routine *powell* from [19]. It turned out that the sum of the third powers gives the best results w.r.t. convergence and final mismatch. During all experiments $M_i = 0$ has been reached within less than three minutes.

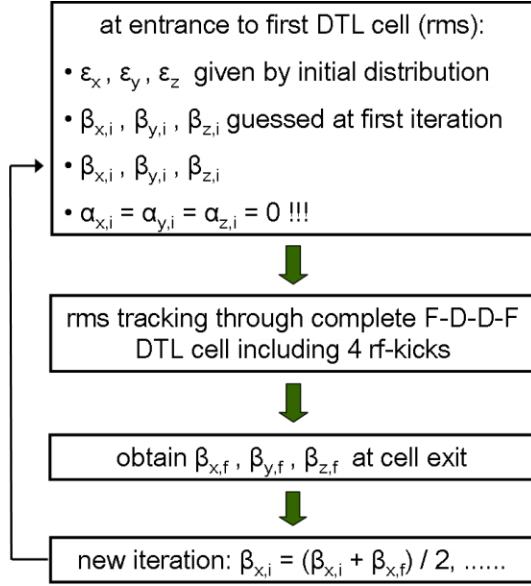


Figure 8: Flux diagram to obtain the periodic solution of the beam envelope inside the periodic DTL.

The emittance growth scan was repeated with matched injection by re-measuring rms emittance growth rates at three different phase advances. It was demonstrated that the application of the matching routine reduced the emittance growth significantly. Fig. 9 plots the measured growth rates as well as results from simulations using the DYNAMION code. In addition to reduced growth, matching leads to

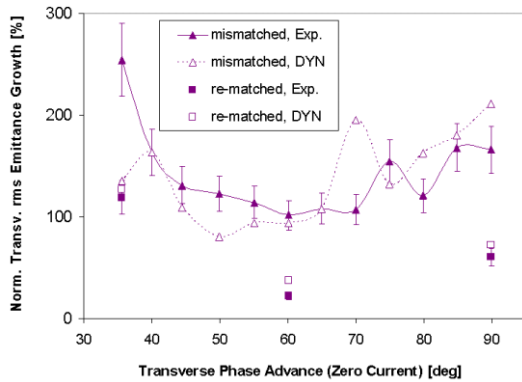


Figure 9: Relative growth of mean value of horizontal and vertical rms emittance at the end of the DTL as a function of σ_o . Data are shown for moderate mismatch and for the case of minimized mismatch.

an improved agreement between measurements and simulations. In order to further investigate this correlation all recent UNILAC experiments were evaluated w.r.t. accuracy of code predictions as a function of mismatch. The agreement improves with the quality of the matching to the DTL. This is plotted in Fig. 10, where the relative deviation between the measured emittances and the code predictions is plotted versus the mean transverse mismatch. For a Beam Dynamics, Other

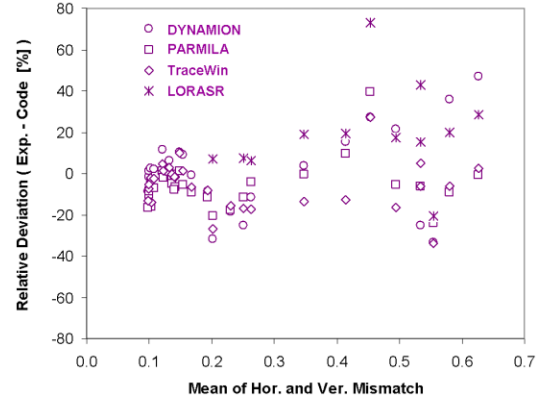


Figure 10: Relative deviation between final transverse rms emittances as measured and predicted by the codes versus the mean of horizontal and vertical mismatch to the DTL.

mean mismatch of less than 20% the codes accuracies are always better than 20%, while for larger mismatches the codes might deliver final rms emittances that differ of up to 80% from the experimental values.

OCTUPOLAR RESONANCE IN A LINAC

The parabolic term of the space charge density rises a field term of third order, i.e. a octupolar term. Since matched beam envelopes perform quasi-periodic oscillations, this term acts as a periodic third-order perturbation. The perturbation is resonant if the envelope phase advance per period, i.e. 360° , is four times higher than the single particles phase advance. Accordingly, a resonance is expected to occur at a depressed transverse phase advance of 90° . Simulations using the PARMILA and DYNAMION codes suggested to perform the measurement at the UNILAC [11]. To exclude any inter-tank mismatch along the DTL just the first tank was used and the zero current phase advance σ_o was varied from 60° to 130° . A characteristic feature of this resonance is the formation of four wings in the transverse phase space distributions being attached to the beam core.

Measured and simulated transverse phase space distributions at the tank exit are plotted in Fig. 11, as well as the corresponding rms emittances. Distributions corresponding to phase advances far away from the resonance have elliptical shapes. At $\sigma_o \approx 100^\circ$, i.e. at a depressed phase advance of 90° , instead the measurements and the simulations clearly revealed four arms, which are typical for a resonant octupolar interaction. A detailed description of the campaign on the octupolar resonance can be found in [11, 12] pointing out that the observed emittance growth is not caused by the well-known envelope instability but by the octupolar space charge potential term.

We summarize that the codes are reliable tools for machine optimization and for the preparation of beam experiments. However, qualitative accuracy strongly depends on the quality of beam matching. Good matching is re-

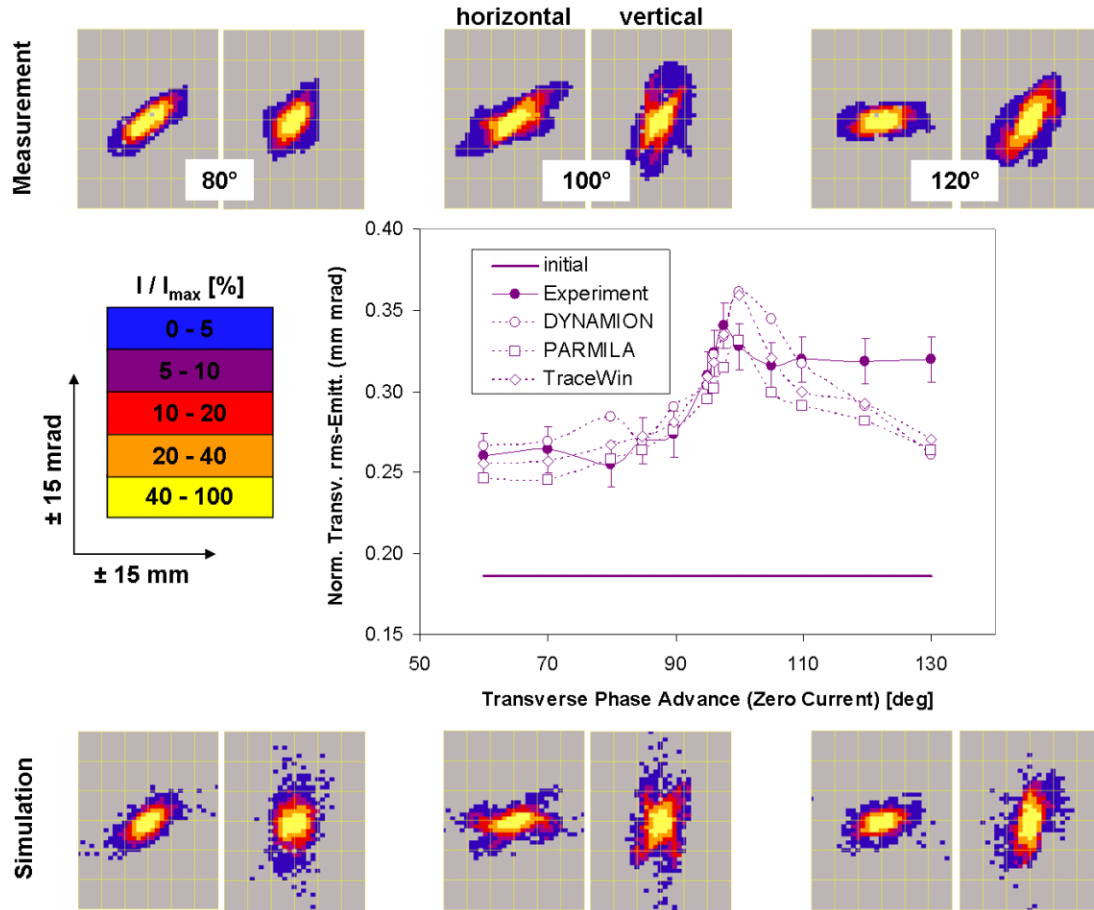


Figure 11: Upper and lower: phase space distributions at the exit of the first DTL tank as obtained from measurements and from the DYNAMION code for phase advances σ_o of 80° , 100° , and 120° . Left (right) side distributions refer the horizontal (vertical) plane. The scale is ± 15 mm and ± 15 mrad. Fractional intensities refer to the phase space element including the highest intensity. Center: Mean of horizontal and vertical normalized rms emittance behind the first DTL tank as a function of σ_o .

quired for useful predictions on the sum of transverse emittances. Considerable deviations within single planes were observed also for well matched beams.

REFERENCES

- [1] R. Ryne et al., Proc. of the HB2008 Workshop, Nashville, U.S.A. (2008).
- [2] S. Nath et al., Proc. of the 2001 Part. Accel. Conf., Chicago, U.S.A., (2001).
- [3] A. Franchi et al., Proc. of the XXIII Linac Conf., Knoxville, U.S.A., (2006).
- [4] A. Franchi et al., <http://www-dapnia.cea.fr/Phocea/file.php?class=std&&file=Doc/Care/note-2006-011-HIPPI.pdf>
- [5] W. Barth et al., Proc. of the XXII Linac Conf., Lübeck, Germany, (2004).
- [6] S. Yaramyshev et al., Nucl. Instrum. Methods Phys. Res., A **558**, 90 (2006).
- [7] J.H. Billen and H. Takeda, PARMILA Manual, Report LAUR-98-4478, Los Alamos, 1998 (Revised 2004).
- [8] <http://irfu.cea.fr/Sacm/logiciels/index.php>.
- [9] R. Tiede, PhD thesis, Goethe University Frankfurt a.M., to be published.
- [10] L. Groening et al., Phys. Rev. ST Accel. Beams **11**, 094201 (2008).
- [11] D. Jeon et al., Phys. Rev. ST Accel. Beams **12**, 054204 (2009).
- [12] L. Groening et al., Phys. Rev. Lett. **102**, 234801 (2009).
- [13] P. Forck et al., Proc. of the XX Linac Conf., Monterey, U.S.A., (2000).
- [14] G. Riehl, *PROEMI: An emittance measurement and evaluation code*. The manual is available from the code author on request (g.riehl@gsi.de).
- [15] T.P. Wangler, "Rf Linear Accelerators", John Wiley & Sons Inc., New York, p. 278, (1998).
- [16] see Ref. [15], p. 217.
- [17] K.R. Crandall, AOT Division Technical Note LA-CP-96-16, January 22, 1966.
- [18] W.H. Press et al., "Numerical Recipes in C", Cambridge University Press, Cambridge, p. 719, (1992).
- [19] see Ref. [18], p. 419.

BENCHMARKING DIFFERENT CODES FOR THE HIGH FREQUENCY RF CALCULATION*

K. Tian, G. Cheng, F. Marhauser, H. Wang

Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, U.S.A.

Abstract

In this paper, we present benchmarking results for high-class 3D electromagnetic (EM) codes in designing RF cavities today. These codes include Omega3P [1], VORPAL [2], CST Microwave Studio [3], Ansoft HFSS [4], and ANSYS [5]. Two spherical cavities are selected as the benchmark models. We have compared not only the accuracy of resonant frequencies, but also that of surface EM fields, which are critical for superconducting RF cavities. By removing degenerated modes, we calculate all the resonant modes up to 10 GHz with similar mesh densities, so that the geometry approximation and field interpolation error related to the wavelength can be observed.

INTRODUCTION

Numerical EM simulations are very important for designing and optimizing new cavity structures, investigating the RF breakdown fields of cavity operation, and studying beam dynamics in RF cavities. Hence, it is very important to understand the accuracy, limitation, and capability of an EM code before applying an EM code in analyzing such problems. For many advanced problems in superconducting RF cavities, high accuracy is not only demanded for the calculation of resonant frequencies but also for the surface electromagnetic fields. For example, multipacting is still an important factor that limits the performance of a superconducting cavity. To correctly predict the process of multipacting in a cavity, the second emission yields of each impact, which is dependent on the impact energy, has to be calculated accurately. Therefore, an accurate calculation of the surface EM fields is a natural requirement for simulating the multipacting process. Another example is the Lorentz force detuning [6] resulting from the interaction of the rf magnetic field with the rf wall current in superconducting cavities. Because the superconducting cavity wall is relatively thin, at high accelerating fields, the cavity shape could be significantly deformed by the inward radiation pressure on the iris wall and the outward radiation pressure on the equator. It hence is a simulation challenge to simulate the frequency shift due to this effect. Since the radiation pressure is directly calculated from the surface EM fields, the key for an accurate simulation relies on the correct prediction of the surface EM fields.

Several EM codes, either developed by commercial companies or non profit research institutes, are utilized

for simulations related to RF cavities. Most of these codes provide good benchmarking results against a simple pillbox cavity for the accuracy of frequency and EM fields. However, a real RF cavity is usually much more complex than the pillbox structure, and normally featured with curved 3-D surface. On the other hand, due to measurement errors and unpredictable operating complexity, it is often difficult to conduct benchmarking comparison between different simulation codes using measured experimental data. In a word, the ideal benchmarking model should have 3-D curved boundaries and can be solved analytically. Spherical cavities are such candidates for the benchmarking study. In this paper, we compare the integrity of results from different EM codes using same spherical cavity models.

ANALYTICAL SOLUTIONS

In this paper, we use r , θ , and ϕ to denote the radial distance, zenith angle, and azimuth angle in a spherical coordinate system, respectively. As shown in Fig.1, we use two different spherical models for the benchmarking: one is a simple cavity bounded by the perfect conductor at $r=a$; the other model is formed by subtracting two cones from a concentric sphere ($d < r < b$). For convenience, throughout this paper, we call the first cavity single sphere, and the second cavity double sphere. We have chosen $a=b=10$ cm and $d=5$ cm.

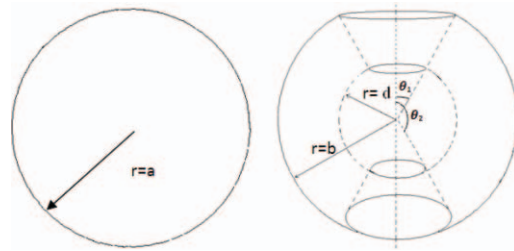


Figure1: Two spherical cavities.

The electromagnetic fields inside a spherical cavity can be obtained by solving Helmholtz equations in spherical coordinates using the Borgnis technique as shown in Ref [7]. For simplicity, we assume that the EM fields, rotationally symmetric, are independent of the azimuthal angle ϕ . Under this assumption, if we choose the radial direction as the longitudinal direction in a spherical cavity, the EM fields can be classified into TM and TE modes, whose general solutions of EM fields are shown in Equations. (1) and (2), respectively:

* Authored by Jefferson Science Associates, LLC under U. S. DOE Contract No. DE-AC05-06OR23177.

#ktian@jlab.org

$$\begin{cases} E_\theta = \frac{1}{r} \frac{\partial^2 U}{\partial \theta \partial r} \\ E_r = \frac{\partial^2 U}{\partial r^2} + k^2 U \\ H_\phi = -\frac{j\omega\epsilon}{r} \frac{\partial U}{\partial \theta} \end{cases} \quad (1)$$

$$\begin{cases} E_\phi = -\frac{j\omega\mu}{r} \frac{\partial V}{\partial \theta} \\ H_\theta = \frac{1}{r} \frac{\partial^2 V}{\partial \theta \partial r} \\ H_r = \frac{\partial^2 V}{\partial r^2} + k^2 V \end{cases} \quad (2)$$

where μ is the vacuum permeability; ϵ is the electric permittivity; ω is the angular frequency; U and V , the electrical and magnetic Borgnis functions [7], have different forms for these two cavities.

After analyzing the analytical solutions of both single sphere and double sphere, we found out the following facts. The maximum surface electric fields of TM modes for the single sphere only occur at the two poles, which are two points. We characterize the numerical accuracy of the surface EM fields by comparing the peak values on the surface. Therefore, in order to avoid the singularity in numerical simulation, we have chosen not to use the TM modes of the single sphere for benchmarking surface electric fields. As to TE modes, the surface electric fields vanish in both models. Hence, in order to benchmark the analytical results of maximum surface electric fields against simulations, we need to solve TM modes of the double spherical cavity. For benchmarking of maximum magnetic fields, we have chosen TE modes in the single sphere due to its relatively simple analytical computation. In the following, we will solve these two families of modes of our benchmarking interest.

TE modes of the single sphere

According to the Borgnis technique, in this case, $U=0$ and:

$$V = B_n \sqrt{r} J_{n+1/2}(kr) P_n(\cos \theta) \quad (3)$$

where B_n is a constant, and k is the wave number. Substituting Eq. (3) to (2), we can obtain the fields expressions with the only undefined parameter B_n . The metal boundary condition on the spherical surface leads to Eq. (4):

$$V|_{r=a} = 0 \quad (4)$$

,which can be transformed to the dispersion equation:

$$\sqrt{y} J_{n+1/2}(y)|_{y=ka} = 0 \quad (5)$$

By solving Eq. (5), we can obtain the resonant frequencies of the mode TE_{n0p} :

$$\omega_{n0p} = \frac{y_{np}}{a \sqrt{\mu_0 \epsilon_0}} \quad (6)$$

where y_{np} is the p^{th} root of Eq. (5).

For TE modes, the only component of electric field is

E_ϕ :

$$E_\phi = -\frac{j\omega\mu_0 B_n}{\sqrt{r}} J_{n+1/2}(kr) \frac{dP_n(\cos \theta)}{d\theta} \quad (7)$$

so the stored energy of the whole cavity can be calculated by integrating E_ϕ throughout the whole spherical volume:

$$W = \frac{1}{2} \epsilon_0 \int_0^{\pi} |E_\phi|^2 dv \quad (8)$$

On the surface of the sphere, all other field components vanish except

$$H_\theta|_{r=a} = -\frac{B_n}{\sqrt{a}} k J'_{n+1/2}(ka) P'_n(\cos \theta) \quad (9)$$

which is only dependent on θ .

For the purpose of benchmarking, we define the normalized surface magnetic field as

$$h = \frac{H_\theta|_{r=a}}{\sqrt{W}} \quad (10)$$

By substituting Eqs (7)-(9) to the above expression:

$$h = \sqrt{\frac{2}{a\mu_0}} \frac{J'_{n+1/2}(ka)}{\sqrt{\int_0^{\pi} [J_{n+1/2}(kr) \frac{dP_n(\cos \theta)}{d\theta}]^2 dv}} P'_n(\cos \theta) \quad (11)$$

Using numerical techniques implanted in MATLAB [8], the maximum values of h for different modes can be determined. The computing precision of MATLAB is 10^{-16} , which is enough for benchmarking. We list the solutions for the first 10 modes in Table 1.

Table 1: The first 10 TE_{n0p} modes

Frequency (Hz)	n	p	$h_{\max}(\text{A/m/J}^{1/2})$
2.1439607465E+09	1	1	-19492.383
2.7499453139E+09	2	1	-21793.147
3.3341835523E+09	3	1	-25108.575
3.6859842956E+09	1	2	19492.383
3.9041824976E+09	4	1	-28186.443
4.3395438284E+09	2	2	21793.147
4.4639808826E+09	5	1	-31009.186
4.9703668154E+09	3	2	25108.575
5.0160366334E+09	6	1	-33617.568
5.2027328095E+09	1	3	-19492.383

TM modes of the double sphere

To solve TM modes of the double sphere cavity, we arbitrarily choose $\theta_1 = \tan^{-1}(8/15)$ and $\theta_2 = \pi - \theta_1$. The Borgnis function $V=0$, and

$$U = [A\sqrt{r}J_{\nu+1/2}(kr) + B\sqrt{r}N_{\nu+1/2}(kr)][CP_{\nu}(\cos\theta) + DP_{\nu}(-\cos\theta)] \quad (12)$$

where A, B, C, D are constants. For this cavity, field expressions contain both Bessel and Neumann functions, and ν is not an integer any more. Field components can be derived by substituting Eqn. (7) in (1). The boundary conditions at four external surfaces are: $E_r=0$ at $\theta=\theta_1$ and θ_2 ; $E_{\theta}=0$ at $r=d$ and b , these conditions lead to the following two equations:

$$U|_{\theta=\theta_1, \theta_2} = 0 \quad (13)$$

$$\frac{\partial U}{\partial r}|_{r=b, d} = 0 \quad (14)$$

The condition of the existence of non zero A, B, C, and D to satisfy Eq (1) leads to the two eigen equations:

$$P_{\nu}(\cos\theta_1)P_{\nu}(-\cos\theta_2) - P_{\nu}(-\cos\theta_1)P_{\nu}(\cos\theta_2) = 0 \quad (15)$$

$$\begin{aligned} & \left[\frac{1}{2\sqrt{a}} J_{\nu+1/2}(ka) + k\sqrt{a} J'_{\nu+1/2}(ka) \right] \left[\frac{1}{2\sqrt{b}} N_{\nu+1/2}(kb) + k\sqrt{b} N'_{\nu+1/2}(kb) \right] \\ & = \left[\frac{1}{2\sqrt{b}} J_{\nu+1/2}(kb) + k\sqrt{b} J'_{\nu+1/2}(kb) \right] \left[\frac{1}{2\sqrt{a}} N_{\nu+1/2}(ka) + k\sqrt{a} N'_{\nu+1/2}(ka) \right] \end{aligned} \quad (16)$$

For a given ν , the only variable in this equation is k , which is the wave number. Hence, we can obtain the resonant frequency of any $TM_{\nu 0 p}$ mode by solving Eq. (16).

Furthermore, we define the normalized surface electric field on the two spherical surfaces as:

$$e_1 = \frac{E_r|_{r=b}}{\sqrt{W}} \quad (17)$$

and

$$e_2 = \frac{E_r|_{r=d}}{\sqrt{W}} \quad (18)$$

Table 2: The frequency and normalized maximum surface magnetic field of the first 6 TE_{n0p} modes

Frequency (Hz)	$e_{1\max}(\text{V/m/J}^{1/2})$	$e_{2\max}(\text{V/m/J}^{1/2})$
8.377783223E+08	-22032489	-6016071
1.832902930E+09	19975209	7675141
2.713403917E+09	15231237	9805581
3.123146964E+09	8868395	-2162046
3.518547956E+09	10047403	11835720
3.598480132E+09	19323910	-4314285

Using the similar technique as the single sphere, the maximum values of e_1 and e_2 can be derived after some algebra. Results calculated from analytical solutions using MATLAB for the first six modes are listed in Table 2.

BENCHMARK RESULTS

Since only azimuthally uniform modes are of interest, we can simulate partial volumes with proper configuration of boundary conditions. This not only improves the computing efficiency, but helps us to suppress the degenerated modes. As shown in Fig. 2, we simulated a sector of 18 degree azimuthally for both cavities. For TE modes of the single sphere, we further cut it at its equatorial symmetry plane, generating a simulated volume of 1/40 of the whole sphere. By setting two side planes as electric boundaries and alternating the equatorial plane as electric or magnetic boundary, we can solve all TE modes. For the double sphere we did not cut the 18 degree sector into half, so the simulated model is 1/20 of the whole double sphere. By setting two side planes as magnetic boundaries, we can obtain all TM modes.

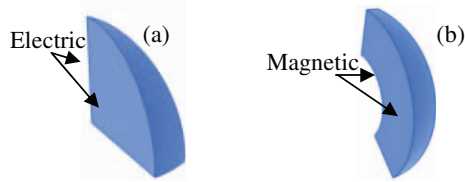


Figure 2: Simulation volumes for (a) single sphere; (b) double sphere

We have simulated modes of these two cavities up to 10 GHz using Omega3P 7.2.1, VORPAL 4.0, CST Microwave Studio (MWS) 2009 and 2008, HFSS 11.0, and ANSYS 11.0. Omega3P, HFSS, and ANSYS utilize Eigen solvers for EM problems based on the Finite Element Method (FEM). The Eigen solver of CST MWS uses the Finite Integration Method and meshing is done by a Perfect Boundary Approximation® method. VORPAL implements the finite-difference-time-domain integration of the EM field on a Yee mesh. Among all these codes, Omega3P, VORPAL, MWS, and ANSYS are capable of parallel computation. In the following, we summarize the simulation conditions of each code.

Omega3P

We automatically meshed both cavities with unstructured 10-point tetrahedral meshes with a mesh size of 2mm using CUBIT [9]. For the single sphere cavity, we ran the code with one CPU for 2 hours on a local server. The double sphere was simulated on Bassi at NERSC, with rather massive parallel computing environment, in about 20 minutes with 56 CPUs. The accuracy is recorded by the given residual number, typically less than $1e-7$.

VORPAL

The single sphere is simulated as 1/20 of the whole sphere without a cut at its equator. VORPAL time-domain simulations are split into 3 runs to cover different frequency range. The filter-diagonalization method (FDM) [10] is used to extract the frequency and construct the field pattern of each TE mode. For a 2mm uniform mesh size, each VORPAL run takes about 3 hours with 16 CPUs on a local Linux cluster. The time for FDM analysis is negligible with integrated frequency analysis tools in VORPAL VIEW. The benchmarking against the double sphere is still in progress.

ANSYS

Both cavities were run on a local windows server in parallel processing mode with four Intel Xeon CPUs and took about 1 hour for each cavity. A uniform 2mm tetrahedral mesh was constructed in ANSYS models.

MWS

The simulation was divided in several runs on a PC with 1 CPU. The average mesh size of 2mm was fixed by setting the upper frequency limit. We used the AKS solver for the lower frequency band and the JDM solver for the higher frequency band. The CPU time varied from a few minutes to 6 hours depending on the solver settings. The accuracy of each solution is checked after the calculation by using Maxwell equations.

HFSS

The simulation was also split into several runs, with each run only covering 10 modes. The mesh scheme was based on a 2 mm average mesh length for both the volume and the sphere surface. Adaptive meshing was used during the iterations. The CPU time varied from a few minutes to about 1 hour depending on the converging speed of different modes. For sufficient accuracy, the maximum deviation in the mode frequency per simulation pass was set at 0.01%.

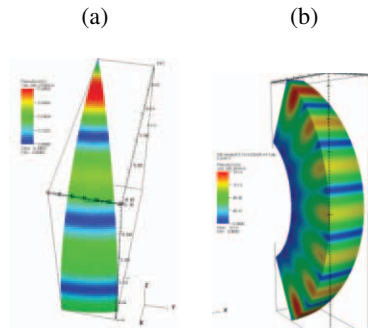


Figure 3: Examples of strap-like surface magnetic fields: (a) single sphere; (b) double sphere.

To exclude non physical (poor accuracy) or degenerated modes from simulations, the rule of thumb is to choose only those modes with strap-like pattern for the amplitude of surface EM fields. Fig. 3 shows the magnetic field distribution of a desired TM mode of the

single sphere. For each mode from different simulation codes, we checked their field distribution and selected only those meeting our criteria for benchmarking.

To quantify the accuracy of simulation results, we define relative errors as the following:

$$Rel. Error = \frac{|X_S - X_A|}{|X_A|} \quad (19)$$

where X_S can represent the simulation result of resonant frequency or the maximum EM field on surface; X_A represents the corresponding analytical result. Fig. 4 presents results of resonant frequencies for both the single and double sphere. For the double sphere, HFSS could not provide solutions for modes higher than about 6 GHz within a reasonable CPU time. The comparison indicates that solutions from Omega3P are consistently more accurate than those from other codes. Its accuracy is well below 10^{-5} . With increasing frequency, the mesh number per wavelength decreases, therefore relative errors for most codes increase as expected. An interesting exception is that relative errors achieved with ANSYS kept rather constant at about 2×10^{-5} .

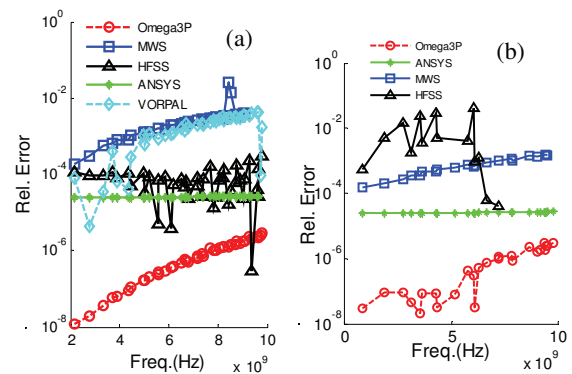


Figure 4: Relative errors of resonant frequencies achieved with various codes for (a) single sphere and (b) double sphere.

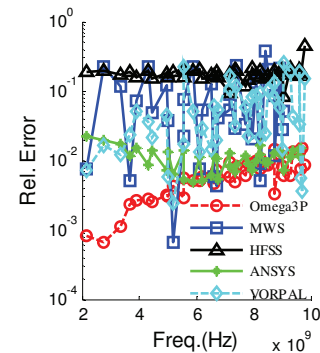


Figure 5: Relative errors of the maximum surface ($r=10\text{cm}$) magnetic fields of the single sphere.

Relative errors for surface EM fields are plotted in Fig. 5 and Fig. 6. Omega3P has overall the best accuracy, ranging from 10^{-4} to 10^{-2} for the magnetic fields and from

10^{-5} to 10^{-2} for the electric fields. Results of ANSYS are not consistent, better for the single sphere, and on the inner surface of the double sphere, but not as good on its outer surface. VORPAL is still comparable with MWS, but HFSS shows very poor accuracy for all cases.

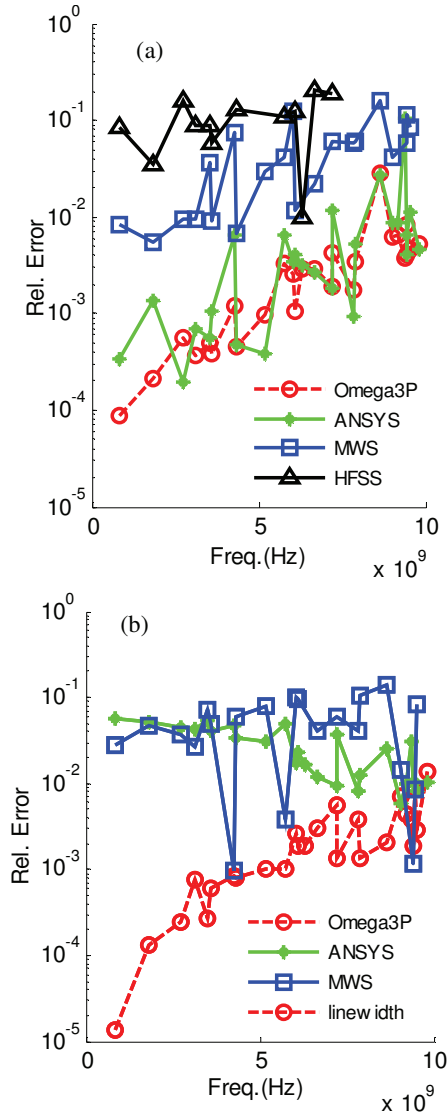


Figure 6: Relative errors of the maximum surface electric fields for the double sphere. (a) Inner surface; (b) outer surface.

CONCLUSION

The accuracy of a numerical simulation is affected by many factors, such as the meshing, algorithms, EM solver types, and interpolation techniques etc. In addition, each different code has its own unique advantages and disadvantages. The benchmarking results in this paper do not intend to favor a specific code to another. However, by choosing a unique analytical model, common simulation settings for all different codes, and most importantly, the same cavity geometry, the results provide better understanding of the performance and limitations of different EM codes, especially when calculating surface fields, which are of high importance for many cavity related phenomena.

ACKNOWLEDGEMENTS

We wish to acknowledge J. Delayen, R. Geng, and R. Rimmer from Jefferson Lab for helpful discussions. We also want to thank the ACD department in SLAC for providing the Omega3P, and TechX corp for providing VORPAL. We also want to specially thank C. D. Zhou for providing the benchmarking data of VORPAL. This work is supported by DOE COMPASS SCIDAC-2 project.

REFERENCES

- [1] K. Ko *et al.*, *Physica C* **441**, 258 (2006).
- [2] C. Nieter, *et. al.*, *J. of Comp. Phys.* 196 (2004) 448-473
- [3] <http://www.cst.de>
- [4] <http://www.ansoft.com>
- [5] <http://www.ansys.com>
- [6] H. Padamsee, J. Knobloch, and T. Hays, "RF Superconductivity for Accelerators", Wiley-VCH, 2008
- [7] K. Zhang, D. Li, *Electromagnetic Theory for Microwaves and Optoelectronics*, Springer, 2007
- [8] <http://www.mathworks.com/>
- [9] <http://cubit.sandia.gov>
- [10] G.R. Werner, *et. al.*, *J. of Comp. Phys.*, 227, (2008), 5200-5214

SIMULATION STUDIES & CODE VALIDATION FOR THE HEAD-TAIL INSTABILITY WITH SPACE CHARGE

Vladimir Kornilov and Oliver Boine-Frankenheim
GSI, Planckstr. 1, 64291 Darmstadt, Germany

Abstract

The head-tail instability represents a potential intensity limitation for bunched beams in the synchrotrons of the FAIR project. Parametrical studies with numerical simulations over very long time scales are necessary in order to understand the effect of direct space charge, nonlinear synchrotron oscillations and image charges, which are all important in the FAIR synchrotrons. Existing analytic approaches either neglect space charge or describe simplified models, which require a numerical or experimental validation. For our simulation studies we use two different computer codes, HEADTAIL and PATRIC. In this work we verify models for wake-field kicks and space-charge effect using the analytic solution for head-tail mode frequencies and growth rates from the barrier airbag model.

INTRODUCTION

Modern synchrotrons, as SIS-100 and SIS-18 of the FAIR [1] complex, will operate with ion bunches under conditions of strong space charge, $\Delta Q_{sc} \gg Q_s$, or moderate space charge, $\Delta Q_{sc} \gtrsim Q_s$, where ΔQ_{sc} is the shift of the betatron tune due to space charge and Q_s is the synchrotron tune. Transverse head-tail instability, which is one of the main concerns for the high-intensity operation, can be strongly modified by the effect of space charge. Classical theories, such as the model of Sacherer [2], do not include interactions of a head-tail mode with any incoherent tune spreads. Recent works [3, 4] propose approaches to treat head-tail modes with space charge. However, numerical simulations appear to be indispensable for a comprehensive stability analysis in different beam parameter regimes and with various collective effects taken into account.

A study for head-tail modes with space charge requires extensive parametrical scans of long time scale (tens of thousands of turns) runs. In the case of the weak head-tail instability it is not possible to scale up e.g. the impedance and hence reduce the run time, since mode coupling excites strong head-tail modes above the associated threshold. For reliable stability predictions it is thus necessary to use code modules, in this case primarily space-charge and wake-field implementations, which have been validated accurately, and are applicable for long time scale runs. In the present work we use two different particle tracking codes, PATRIC [5] and HEADTAIL [6], in order to take the advantage of different kinds of implementations and thus treat the task. The PATRIC code, a part of numerical Transverse and Longitudinal Instabilities and Cures

cal development effort at GSI, was optimised for relatively short-term effects with well-resolved betatron oscillations and exact self-consistent space charge solvers, while the HEADTAIL code, created at CERN, was designed historically for longer-term phenomena, including electron-cloud effects, with an option of very fast once-per-turn modulus. For the code validation in the range of moderate and strong space charge we suggest to use the model of an airbag bunch in barrier potential. This bunch model, being rather artificial, has a simple analytical solution [7], can be easily implemented in a simulation and, as we demonstrate here, gives very useful insight into physics of head-tail modes with incoherent interactions.

PHYSICAL MODEL

An analytical solution for head-tail modes in bunches with space charge has been derived in Ref. [7]. The model assumes the airbag distribution in phase space and the square-well (barrier) potential and thus a constant line density. The longitudinal momentum distribution has then two opposing flows of particle $[\delta(v_0 - v_b) + \delta(v_0 + v_b)]$, the synchrotron tune in this bunch is $Q_s = v_b/2\tau_b R f_0$, where τ_b is the full bunch length in radian, R is the ring radius and f_0 is the revolution frequency. The model considers “rigid slices”, i.e. only dipole oscillations without transverse emittance variation are included. It is also assumed that all betatron tune shifts are small compared to the bare tune $|\Delta Q| \ll Q_0$. The resulting tune shift due to space

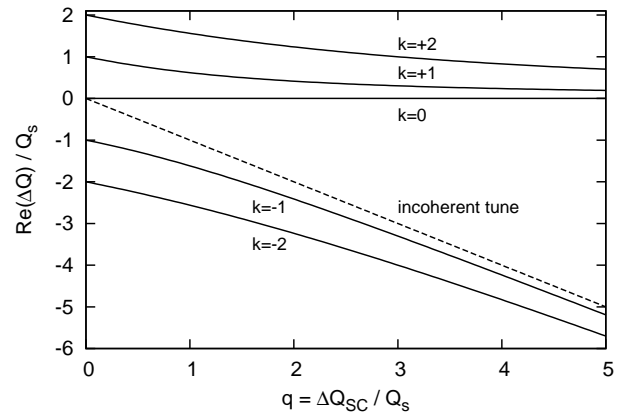


Figure 1: Tune shifts of five head-tail modes versus space charge parameter q as given by the airbag theory Eq. (1), the dashed line is the incoherent betatron tune $Q_0 - \Delta Q_{sc}$.

charge (without wake) is given by

$$\Delta Q = -\frac{\Delta Q_{sc}}{2} \pm \sqrt{\frac{\Delta Q_{sc}^2}{4} + k^2 Q_s^2}, \quad (1)$$

where "+" is for modes $k \geq 0$, ΔQ_{sc} is defined as a positive value. These space-charge tune shifts are illustrated in Fig. 1, where we introduce a space charge parameter $q = \Delta Q_{sc}/Q_s$.

The solution for weak head-tail instability is obtained in Ref. [7] assuming the wake potential $W(\tau) = W_0 \exp(-\alpha\tau)$ with a sufficiently short range compared to the bunch length, $\alpha\tau_b \gg 1$. For the mode $k = 0$, which is not affected by space charge, the tune shift is given by

$$\Delta Q = \Delta Q_0(\alpha/\zeta + i), \quad (2)$$

where we define ΔQ_0 as the growth rate for the mode $k = 0$,

$$\Delta Q_0 = -\frac{\zeta}{\alpha^2} \frac{\kappa W_0}{2Q_0}, \quad (3)$$

here $\zeta = \xi Q_0/\eta$ is the normalized chromaticity, $\Delta Q_\xi/Q = \xi \Delta p/p$, and

$$\kappa = \frac{I_0 q_{ion}}{2\pi\gamma m\beta c\omega_0^2}, \quad (4)$$

where β, γ are relativistic parameters, I_0 is the beam current, q_{ion} is the ion charge and ω_0 is the angular revolution frequency.

For modes $k \neq 0$ this theory predicts

$$\Delta Q = -\Delta Q_{sc} + \frac{\Lambda_0 \pm \sqrt{\Lambda_0^2 + 4k^2 Q_s^2 \Lambda_s}}{2\Lambda_s}, \quad (5)$$

where

$$\begin{aligned} \Lambda_0 &= \Delta Q_0(\alpha/\zeta + i) + \Delta Q_{sc}, \\ \Lambda_s &= 1 - \left[\frac{\Delta Q_0 \pi}{2\zeta Q_s \tau_b} \right]^2. \end{aligned}$$

SPACE CHARGE SIMULATIONS

In order to verify the space-charge implementation for long-time simulations with a particle tracking code, we have introduced the barrier-airbag bunch distribution in both PATRIC and HEADTAIL codes. For space charge, the frozen electric field model was used, i.e. a fixed potential configuration which follows the mass center for each single slice. This approach is justified for the "rigid-slice" regime, which is also a basis assumption for the airbag theory [Eq. (1)] we compare with. A homogeneous and round transverse bunch profile (K-V distribution) was used in simulations in this work, a linear rf force was assumed.

A good agreement between theory and simulations with space charge has been achieved with both codes. Figures 2 and 3 show the coherent bunch spectrum, which is the Fourier transform of the total transverse bunch offset, for Transverse and Longitudinal Instabilities and Cures

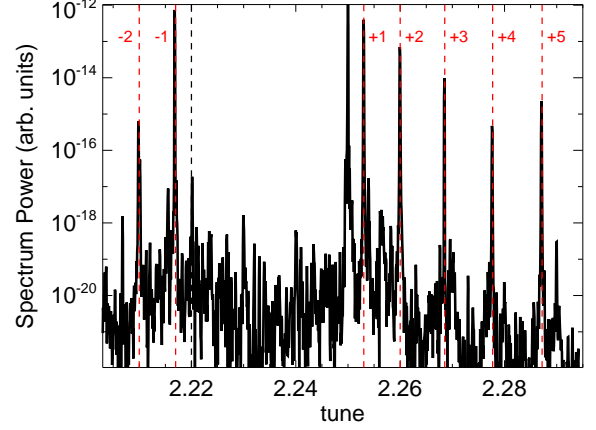


Figure 2: Dipole bunch spectrum from simulations (the HEADTAIL code, airbag bunch) with moderate space charge $q = 3$. The bare tune is $Q_0 = 2.25$, $Q_s = 0.01$, the black dashed line is the incoherent tune ($Q_0 - \Delta Q_{sc}$), red dashed lines are head-tail modes from Eq. (1)

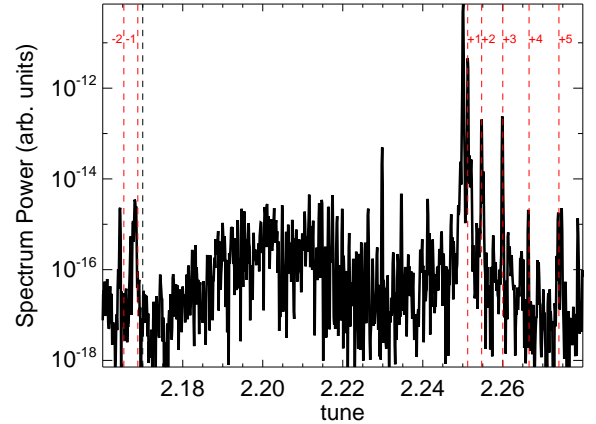


Figure 3: Dipole bunch spectrum from simulations (the PATRIC code, airbag bunch) with strong space charge $q = 8$. The legend and the tunes correspond to Fig. 2.

moderate $q = 3$ and stronger $q = 8$ space charge. Prediction of the airbag theory is laid over in red.

In our parametric studies we have found that even for moderate and weak space charge it is not enough to execute only one space-charge kick per turn. At least one kick per approximately 1 rad phase advance is necessary for an accurate description of space-charge effect. In the example presented here, $Q_0 = 2.25$, this means at least 16 kicks per turn, and for the SIS-100 tune $Q_0 = 18.7$ this means 120 kicks per turn.

In order to test the relevance of the airbag theory for realistic bunches we have performed simulations with a Gaussian bunch, i.e. Gaussian line density profile and Gaussian momentum distribution. For comparisons between

bunch profiles, the line density of the airbag-bunch is taken to be equal to the density in the middle of the Gaussian bunch. Especially for strong space charge a surprisingly good agreement between the airbag theory and simulation results for the Gaussian bunch is found, see Fig. 4. The frequency shift of head-tail eigenmodes due to space charge seems to be a very robust effect, which does not depend on the details of the bunch form and distribution, especially for strong space charge.

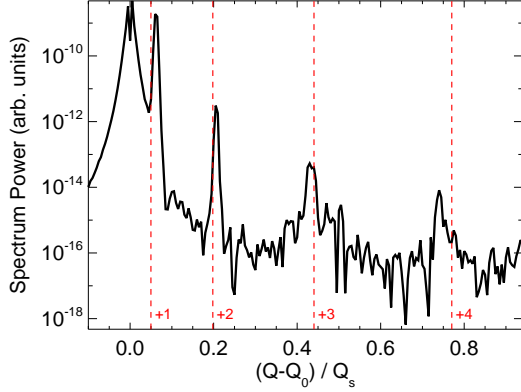


Figure 4: Dipole bunch spectrum from simulations for a Gaussian bunch with strong space charge $q = 20$. Red dashed lines are head-tail modes for the airbag bunch, Eq.(1). To clarify the notation we note that the tune shifts without space charge are $\Delta Q/Q_s = 1$ for $k = 1$, $\Delta Q/Q_s = 2$ for $k = 2$, etc.

INSTABILITY SIMULATIONS

For wake-field verifications the exponential wake potential $W(\tau) = W_0 \exp(-\alpha\tau)$ has been implemented in the codes. The wake momentum kick for a bunch slice τ_0 (or z_j) is calculated [8] as

$$\begin{aligned} \Delta x' &= \frac{2\pi\kappa}{R} \int_{-\tau_b}^{\tau_0} \frac{\lambda(t, \tau)}{\lambda_0} \bar{x}(t, \tau) W(\tau_0 - \tau) d\tau = \\ &= \frac{2\pi q_{\text{ion}} \kappa}{\lambda_0 R^2} \sum_{i>j} N_i \bar{x}_i W(z_j - z_i), \end{aligned} \quad (6)$$

where $\lambda(t, \tau)$ is the bunch line density for the time t in the bunch position τ , λ_0 is the peak line density, \bar{x} is the transverse slice offset, $x' = dx/ds$, N_i is the particle number in the i th slice, and κ is given in Eq. (4).

The analytic solution for the exponential wake potential with the airbag bunch [Eqs. (2)-(5)] predicts the mode $k = 0$ to be the most unstable head-tail mode. As we illustrate in Fig. 5, $k \neq 0$ modes always have smaller growth rates than ΔQ_0 . Here we normalize $\text{Im}(\Delta Q)$ by the growth rate of the $k = 0$ mode. In this example the $k = -1$ mode becomes more unstable than the $k = 0$ mode at $\Delta Q_0 \approx 0.003$, which is due to the fact that the model breaks down. As we can see in Fig. 6, after $\Delta Q_0 \approx 0.002$

Transverse and Longitudinal Instabilities and Cures

(where $\Delta Q_0/Q_s \approx 0.2$) the mode $k = -1$ does not approach the $k = 0$ mode anymore, although in reality they should cross and produce mode coupling, but the latter is not included in this theory. The observation that space charge suppresses the largest growth rate among $k \neq 0$ is not general and depends on parameters $\alpha\tau_b$, Q_s , etc. Here we take $\alpha\tau_b = 40$ (in the simulations presented below as well), for other parameters space charge can cause an increase in the largest growth rate among $k \neq 0$. Finally, it should be noted that in a real bunch with space charge the $k < 0$ modes are strongly damped by any betatron- and synchrotron tune spread, additionally these modes are damped by the non-resonant dipole behaviour [9].

In the simulations we can observe only the dominant $k = 0$ mode, but it is possible to resolve real tune shifts also for $k \neq 0$, at least for small wake amplitudes. Our first comparison between simulations and the theory is for the case without space charge, see Fig. 6. Particle tracking simulations reproduces fairly well weak head-tail instability from the theory, both growth rates and real tune shifts. Only close to $\Delta Q_0 \approx 0.002$ discrepancies appear, because the airbag approach breaks down, as we discuss above.

The next step is to combine the wake-field effect with space charge. Results of this parametric scan are presented in Fig. 7, where a constant space charge tune shift $\Delta Q_{\text{sc}} = 3Q_s$ was assumed. This is related to an increasing impedance, not increasing beam intensity. Also here we successfully reproduce the theory results for the head-tail instability with space charge. In agreement with the theory the simulations demonstrate that space charge does not affect the $k = 0$ mode and strongly changes other head-tail modes.

Another test for the relevance of the airbag theory for realistic bunches is to perform simulations with the exponential wake for a Gaussian bunch. Space charge intensity $q = 5$ in the middle of the bunch is assumed, $\alpha\tau_b = 10$ is

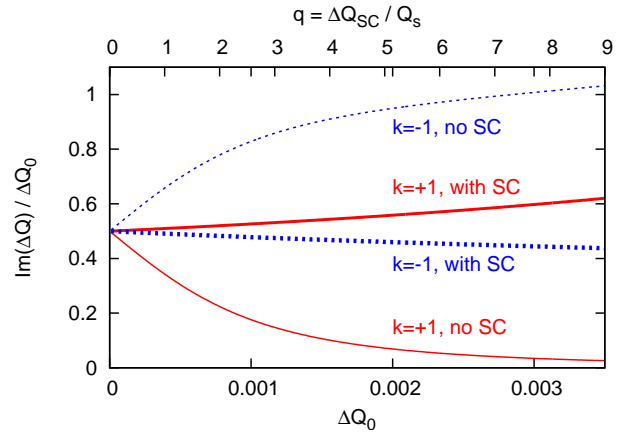


Figure 5: Growth rate given by the airbag theory Eqs. (2)-(5). The beam intensity increases along the horizontal axis, thus the wake amplitude (lower axis range) and space charge (upper axis range) grow proportionally.

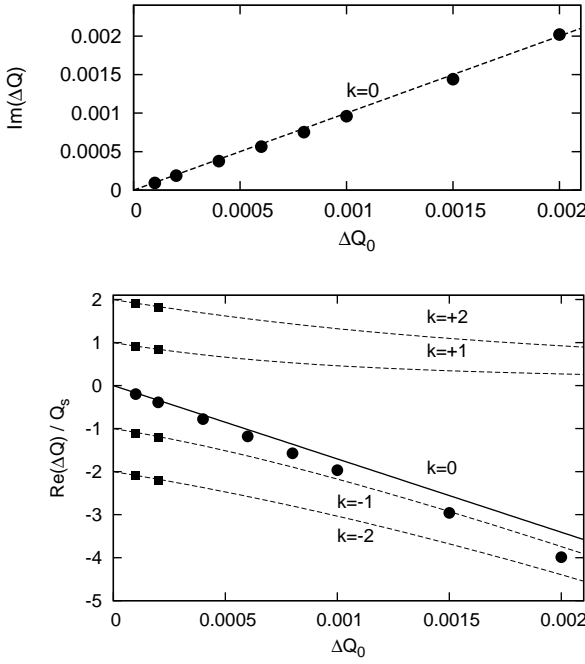


Figure 6: Growth rates (upper plot) and real tune shifts (bottom plot) of head-tail modes without space charge. Symbols: simulation output (the PATRIC code, airbag bunch); lines: airbag theory Eqs. (2)-(5). $\text{Re}(\Delta Q)$ for $k \neq 0$ is shown only for small ΔQ_0 .

taken. Simulations with and without space charge showed a dominating $k = 0$ mode which is not affected by space charge. Figure 8 demonstrates that if growth rates and real tune shifts from simulations are multiplied by a factor 1.89, a good agreement with the airbag theory Eqs. (2) is achieved. This means that head-tail modes in a Gaussian bunch correspond, accurate to a constant factor, to the airbag theory, although the Gaussian bunch has totally different linear density and momentum distribution. Obviously, this scaling factor depends on bunch parameters and wake-field properties. Frequencies of $k \neq 0$ modes were difficult to resolve in this case, which may be related to the distribution type. However, one would suspect that this scaling factor is not effective for $k \neq 0$ modes due to effect of space charge. The appropriate way to clarify this issue may be simulations for parameters where $k \neq 0$ modes are the most unstable. This should be investigated in further studies.

Finally it is worth to mention that in our verification study we have proved that it is enough to apply the wake-field kick once per turn. Still, in the case with space charge numerous kicks are necessary for the space-charge module.

WAKE FIELD IN LONG BUNCHES

The validation of the wake-field module with the exponential wake potential $W(\tau) = W_0 \exp(-\alpha\tau)$ is not complete, since the most important case in e.g. SIS-

Transverse and Longitudinal Instabilities and Cures

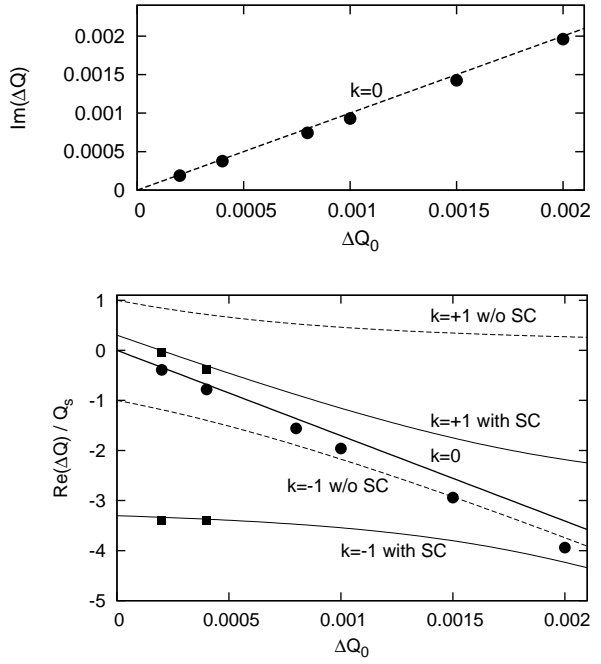


Figure 7: Growth rates (upper plot) and real tune shifts (bottom plot) of head-tail modes in the airbag bunch with space charge $q = 3$. The legend corresponds to Fig. 6. $\text{Re}(\Delta Q)$ for $k = \pm 1$ is shown only for small ΔQ_0 . The theory lines for cases with/without space charge are shown for comparison.

synchrotrons — long bunches and long-range wake fields — is not included. In order to do this we again consider a problem which has an analytical solution. Take a coasting beam without effects of chromatic tune spread (zero chromaticity, or a space-charge induces large gap between coherent frequency and incoherent spectrum) and the resistive-wall impedance in the thick-wall regime. The latter implies that the skin depth $\delta_{sk} = \sqrt{2c/Z_0\sigma_{rw}\Omega_{cb}}$ is small compared to the pipe wall thickness. This impedance is given by

$$\text{Re}(Z_{rw}^\perp) = \frac{L_{rw}}{2\pi b^3} Z_0 \delta_{sk}, \quad (7)$$

where $Z_0 = 376.7 \text{ Ohm}$, b is the pipe radius and $\Omega_{cb} = Q_{cb}\omega_0$ is the coherent frequency of the coasting beam. Since Z_{rw}^\perp is large for small frequencies as $1/\sqrt{\Omega}$, the eigenmode with the smallest frequency is the most unstable. As the coasting-beam unstable spectrum is represented by slow waves $\Omega_{cb} = (n - Q_0)\omega_0$, the resistive-wall impedance mostly excites the eigenmode with

$$Q_{cb} = 1 - Q_f \quad (8)$$

and with the mode index n , which is the closest integer above the bare tune Q_0 . Here Q_f is the fractional part of Q_0 . The growth rate of a coasting-beam mode is

$$\text{Im}(\Delta Q_{cb}) = \frac{I_0 q_{ion}}{4\pi c \gamma m Q_0 \omega_0} \text{Re}(Z^\perp). \quad (9)$$

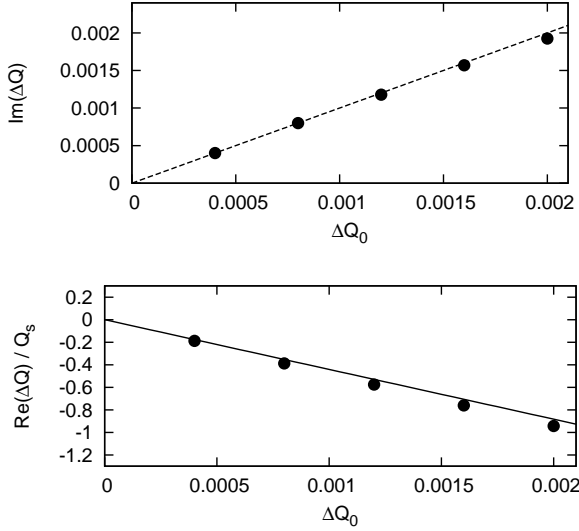


Figure 8: Growth rates and real tune shifts of the $k = 0$ head-tail mode with space charge $q = 5$. Circles are simulation results for a Gaussian bunch, both $\text{Im}(\Delta Q)$ and $\text{Re}(\Delta Q)$ are scaled by 1.89. Lines are given by the airbag theory.

In our verification simulations we consider a coasting beam with Gaussian momentum distribution, wake-field kicks are implemented according to Eq. (6). The resistive-wall wake function which corresponds to the impedance Eq. (7) is given by

$$W_{\text{rw}}(z) = -\frac{cL_{\text{rw}}}{b^3} \left(\frac{\beta}{\pi} \right)^{3/2} \sqrt{\frac{Z_0}{z \sigma_{\text{rw}}}}. \quad (10)$$

In our parametric study for coasting beams we compare the theory of the impedance Eq. (7) and the slow-wave Eqs. (8) and (9), on the one hand, with simulation results for the wake field Eq. (10), on the other hand. Figure 9 demonstrates this comparison, where the growth rate is normalized by the parameter $\Delta Q_{\text{rw}0}$, which is the value of $\text{Im}(\Delta Q_{\text{cb}})$ calculated for $\Omega_{\text{cb}} = \omega_0$, or formally $Q_f = 0$. For this good agreement it was essential to include previous turns in the wake module Eq. (6), which seems to be obvious, since the “head” (talking in bunch terms) is strongly kicked by the “tail” from the previous turn in a coasting beam. Furthermore, the wake function Eq. (10) is a long-range potential, basically non-saturating. However, even for small $(1 - Q_f)$ it was enough to include three previous turns to achieve an adequate description.

The coasting-beam mode type we could verify with the resulting beam-offset structure. Figure 10 shows an example of the dipole moment pattern (which is equal to the beam-offset for the dc beam) of an unstable beam from a HEADTAIL simulation. The dashed line is the fixed-location signal $\Delta_x(z)$, which is the natural output of our simulations, and the solid line is the snapshot signal $\Delta_x^{\text{snap}}(z) = e^{-iQ_0 z/R} \Delta_x(z)$, which we would see at a certain time. The

Transverse and Longitudinal Instabilities and Cures

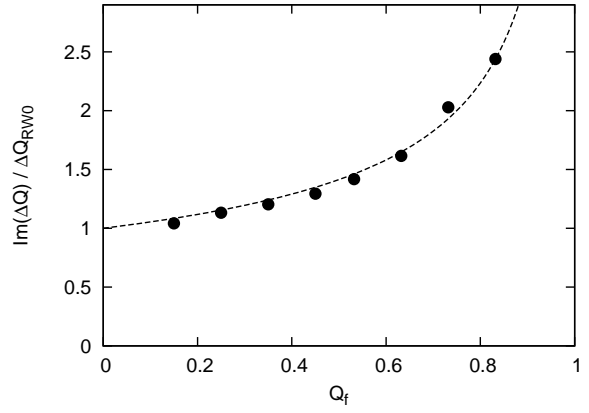


Figure 9: Growth rate of the resistive-wall instability in a coasting beam versus fractional part of the betatron tune. Dashed line is the analytic result Eq. (9) for the mode Eq. (8) and for the impedance Eq. (7). Circles: simulations with the wake field Eq. (10).

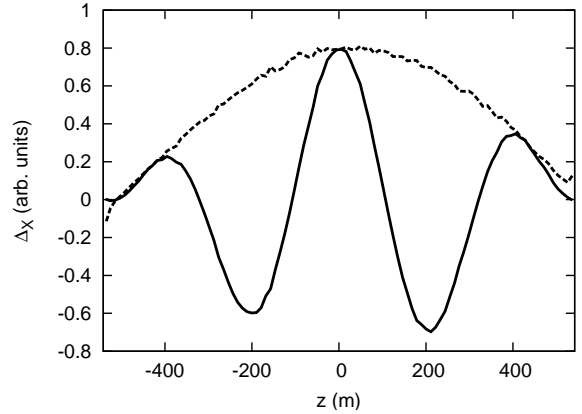


Figure 10: Transverse offset structure over the circumference of an unstable coasting beam from an exemplary simulation. Dashed line is the fixed-location signal, the solid line is the corresponding snapshot signal.

betatron tune is $Q_0 = 2.532$ here, accordingly, we observe the $n = 3$ mode.

CONCLUSIONS

Space-charge and wake-field modules, suitable for long time scale simulations of the weak head-tail instability, have been verified with the airbag-barrier model [7], which has an analytical solution for the exponential wake potential.

Simulations for moderate and strong space charge demonstrated a good agreement between PATRIC and HEADTAIL simulations and the theory. It was found that betatron oscillations should be well resolved in the particle tracking for a correct description of space-charge effects. Simulations with a Gaussian bunch showed that the airbag theory Eq. (1), although considering a simple bunch

model, describes rather accurately effects of space charge, especially for strong space charge.

Instability simulations reproduced fairly well the theory predictions concerning growth rates and real tune shifts, both with and without space charge. Simulations with a Gaussian bunch demonstrated that the airbag theory Eqs. (2)-(5) can be very helpful even in understanding effects of space charge for the head-tail instability in realistic bunches. Our wake-field validation with a coasting beam has proved that the method can be surely used for long bunches and long-range wake potentials.

ACKNOWLEDGEMENTS

We thank Giovanni Rumolo (CERN) for fruitful collaboration, and Alexey Burov (FNAL) for useful discussions.

REFERENCES

- [1] FAIR Baseline Tech. Report 2006:
<http://www.gsi.de/fair/reports/btr.html>
- [2] F. Sacherer, Proc. First Int. School of Particle Accelerators, Erice, p. 198 (1976)
- [3] A. Burov, Phys. Rev. ST Accel. Beams **12**, 044202 (2009)
- [4] V. Balbekov, FERMILAB-PUB-09-322-APC (2009)
- [5] O. Boine-Frankenheim, V. Kornilov, Proc. of ICAP2006, 2-6 Oct., Chamonix Mont-Blanc, (2006)
- [6] G. Rumolo and F. Zimmermann, Phys. Rev. ST Accel. Beams **5**, 121002 (2002)
- [7] M. Blaskiewicz, Phys. Rev. ST Accel. Beams **1**, 044201 (1998)
- [8] A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators*, John Wiley & Sons, Inc. (1993)
- [9] O. Boine-Frankenheim, V. Kornilov, to be published (2009)

PROGRESS WITH UNDERSTANDING AND CONTROL OF NONLINEAR BEAM DYNAMICS AT THE DIAMOND STORAGE RING

R. Bartolini, Diamond Light Source Ltd, Oxfordshire, OX11 0DE, UK and John Adams Institute, University of Oxford, OX1 3RH, UK.

Abstract

The Diamond light source started operation for users in January 2007. With the successful commissioning of the nominal optics, delivering a 2.75 nm emittance beam at 3 GeV, we now routinely provide the users with a 250 mA beam with a lifetime of 20 h, exceeding the minimum specified current-lifetime product of 3000 mAh.

Driven by the necessity to guarantee a correct implementation of the nonlinear optics, a significant experimental and theoretical effort is ongoing to understand and improve the nonlinear beam dynamics in the storage ring. The necessity to control the nonlinear beam dynamics is even more urgent with the installation of a large number of small gap (5 mm) in-vacuum insertion device and the need to control the injection efficiency with Top-Up operation. We report here the present status of the analysis of the nonlinear beam dynamics and the main experimental results.

INTRODUCTION

Diamond is a third generation light source which entered in operation in January 2007 [1-2]. The storage ring lattice is a Double Bend Achromat (DBA) where the achromatic condition is broken and dispersion leaks in the straight sections in order to reduce the natural emittance of the machine. The storage ring consists of 24 DBA cells, with ten quadrupoles and seven sextupoles per cell, making a total of 48 dipoles, 240 quadrupoles and 168 sextupoles. The sextupoles are combined function magnets which also have skew quadrupole and dipole correctors in the horizontal and vertical plane. These magnets all have independent power supplies, allowing a large degree of freedom in the choice of both the optimisation and the correction of the linear and nonlinear optics. The ring is also equipped with a set of 7 BPMs per cell providing a total of 168 BPMs, each with turn-by-turn capabilities.

The sextupoles were carefully optimised in order to provide sufficient dynamic aperture and momentum aperture for injection and a Touschek lifetime of at least 10 h for the nominal operating current of 300 mA in a 2/3 fill. Extensive numerical simulations were performed to ensure that the injection efficiency and the Touschek lifetime were maintained even with the complement of IDs planned for Phase-I and Phase-II. Currently this includes ten in-vacuum IDs at 5 mm minimum gap, two superconducting multipole wigglers and an APPLE II device. Two customised optics are also planned in two long straight sections, providing two vertical mini-beta sections with two virtual horizontal focuses.

Striving for the lowest emittance achieved so far in a medium energy machine (2.75 nm), the correct operation of the ring requires a very strict control of the optics of the storage ring. During the commissioning the correct implementation of the linear optics was achieved with the use of the LOCO package [3] and its implementation in the MATLAB Middlelayer [4]. The residual β -beating was reduced to less than 1% peak-to-peak and the nominal emittance of 2.75 nm was measured with very good correction of the linear coupling. The correction is achieved with LOCO by fitting the quadrupoles to match the model and measured orbit response matrix. The required quadrupole gradient corrections are below 2% peak-to-peak and are compatible with the measurements of the quadrupole gradient performed prior to the installation of these magnets.

While the correct implementation of the linear optics is nowadays not unusual in modern third generation light sources, the analysis and correction of the nonlinear model of the storage ring of most modern light sources still stops short of an equivalently good solution [5]. In this context, Diamond has put in place a significant experimental and theoretical effort to provide tools and strategies that allow a correct implementation of the nonlinear model of the storage ring. In this paper we report the current status of the investigation and the main experimental results.

CHARACTERISATION OF THE NONLINEAR BEAM DYNAMICS

The nonlinear dynamics of the storage ring is optimised in order to provide sufficient aperture for injection and adequate Touschek lifetime for the electron beam in the various operating conditions. This is achieved by extending the dynamic aperture and the momentum aperture of the ring. Numerical tracking is ultimately used to validate the optimisation and currently available codes such as Tracy-II [6] or elegant [7] have automated numerical computation of the ring apertures.

It is desirable however to provide dynamical quantities that characterise the nonlinear behaviour of the ring that can be used at the design stage, that can give insight on the beam dynamics and provide further guidelines to the optimisation. When these quantities can be accessed experimentally in the machine, they provide a valuable tool to compare the nonlinear model of the ring with the real nonlinear behaviour of the beam in the storage ring. The dynamical quantities typically used are the detuning with amplitude and the nonlinear resonance driving terms, which can be computed to the desired perturbative order with codes such as Tracy-II and Mad-X/PTC [8]. A

crucial problem in nonlinear beam dynamics is related to the fact that these quantities are not necessarily well correlated with the dynamic and momentum apertures of the ring which is the ultimate goal of the optimisation. It is well known that, in general, the dynamic aperture and the momentum aperture cannot be simply improved by targeting detuning with amplitude, one or even a few resonant driving terms. Therefore the optimisation has to be validated numerically. Nevertheless these quantities allow the comparison of the nonlinear model to the ring.

Another very interesting numerical tool that allows a quick analysis of the resonance net around the working point in the tune diagram is the Frequency Maps (FMs) introduced in accelerator physics by J. Laskar [9]. This allows drawing the region occupied by the beam in frequency space. The diffusion coefficients allow a clear visual understanding of the strength of a given resonance crossed by the beam. FMs can also be measured experimentally and the diffusion strength can be somehow substituted with the measured losses thus allowing a valuable comparison of the real machine with the model.

Finally, the measurement of the spectral lines of the betatron oscillations and their connection to the resonance driving terms [10-11] has been proposed to compare the behaviour of the real machine with the nonlinear model. The spectral lines can be associated with the amplitude and strength of a given resonance driving term and give a complementary handle to control the implementation of the model to the real machine.

FREQUENCY MAPS AND DYNAMIC APERTURE STUDIES AT DIAMOND

The nonlinear beam dynamics activity at Diamond has aimed at achieving a good implementation of the nonlinear model of the storage ring in order to guarantee sufficient injection efficiency and Touschek lifetime. Crucial to this work was the installation of two pinger magnets in Sept. 2007 and the possibility of acquiring turn-by-turn data from all BPMs in the ring. The pingers can independently kick the beam in the horizontal and in the vertical plane to large amplitudes, scanning the whole dynamic aperture available. They were used to determine the dynamic aperture (DA) and to measure the frequency map (FM) of the storage ring.

The comparison between the measured DA and the prediction from the model is reported in Fig. 1. Considering that the yellowish regions in the model DA are likely to be lost over long term tracking, we can only claim a qualitative agreement in the shape of the DA, especially in the horizontal plane. The measurements show a significant disagreement which reaches a factor two in the vertical plane. This degree of disagreement is not unusual, even for modern third generation light sources [5]: it points to the fact that our present knowledge of the nonlinear model does not fully capture the complexity of the nonlinear motion of the beam in the storage ring. In fig. 2 we report a measurement of the

momentum aperture obtained from a scan of the lifetime as a function of the RF voltage, in a condition where the beam lifetime is Touschek dominated. Again the agreement is qualitatively good but the simulations fail to reproduce the exact voltage for which the maximum lifetime is attained. The model gives maximum lifetime for a voltage which is about 15% smaller than the measurements corresponding to 0.5% underestimate of the momentum aperture.

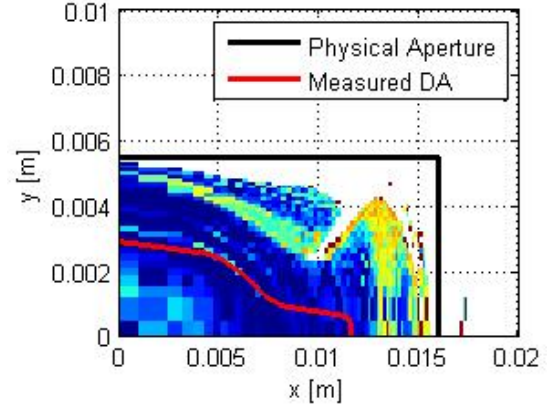


Fig. 1: Comparison of the measured dynamic aperture with the prediction from the model. The colour code illustrates the diffusion strength. Tracking is performed with Tracy-II.

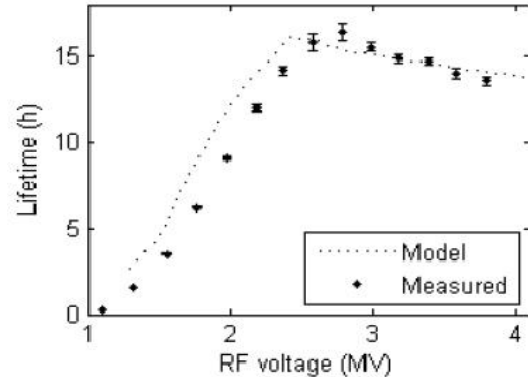


Fig. 2: Comparison of the measured and model lifetime as a function of the RF voltage.

The reason for this disagreement is still the object of investigation. On the one hand we are trying to include in the model all possible known error sources, on the other hand we are investigating the appropriate modelisation of the magnetic elements trying to include thick lense multipoles, fringe fields, edge focussing effects and their momentum dependence. In particular we have included a better description of the magnetic length of the magnets by using the individual data coming from magnetic measurements and we have introduced octupolar components in the quadrupoles which was considered to be the main source of error. While these additional ingredients have improved the match between model and real machine, as shown later for the FMs, it is worth

pointing out that even when including the best knowledge of the nonlinearities in the ring we still obtain slightly different values for the chromaticities which are (3, 1.5) in the model while we measure (2, 2) in the machine. Errors in the calibration tables for the sextupole magnets are also a possible source of disagreement between the machine and the model.

In Fig. 3 we report the measurements of the frequency map obtained by scanning the aperture with the pinger magnets and measuring the corresponding betatron tunes. In Figs. 4-6 we report the corresponding FMs obtained from numerical tracking in the model. Different refinements of the nonlinear model were used. Fig. 4 corresponds to the bare lattice with the nominal sextupole lengths. In Fig. 5 we used the correct magnetic length for the sextupoles obtained from magnetic measurements. Fig. 6 includes the octupolar errors in the quadrupoles. It is clear that a better description of the model improves the agreement but there are still significant deviations which are not explained. Work is ongoing to include all the remaining multipolar errors in the quadrupoles, the main errors in the dipoles and sextupoles and a better calibration table of the sextupole gradient.

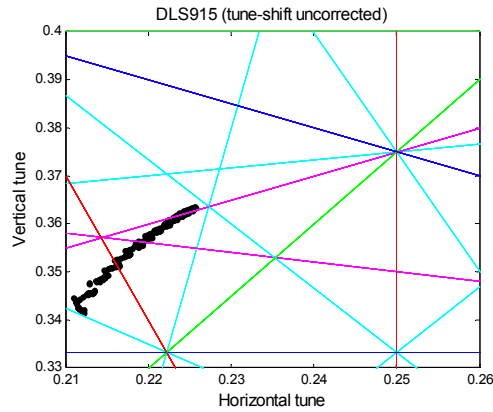


Fig. 3: Measured Frequency Map on the bare lattice without Insertion Devices.

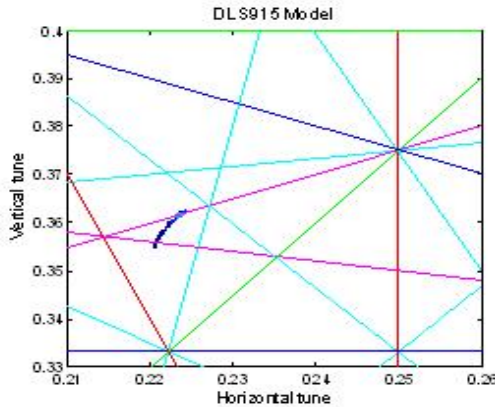


Fig. 4 Model Frequency map. Bare lattice without Insertion Devices.

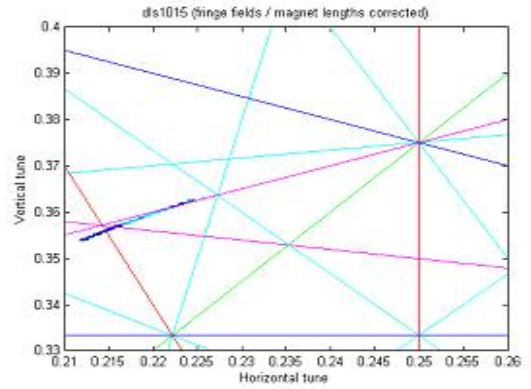


Fig. 5: Model Frequency Map as in Fig. 4, adding to the model the correct magnetic length of the sextupoles as per magnetic measurements.

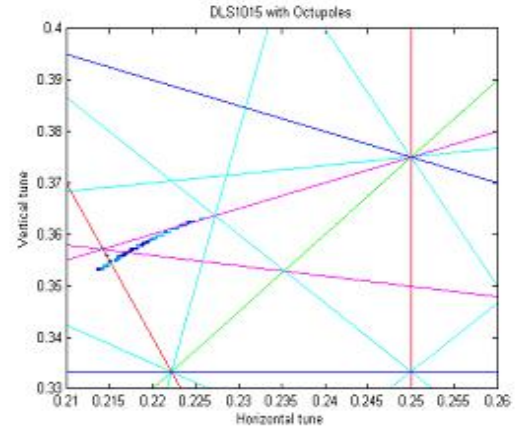


Fig. 6 Model Frequency map as in Fig. 5 adding to the model the octupolar errors in the quadrupoles as per magnetic measurements.

SPECTRAL LINES MEASUREMENTS

It has been shown in the past that Fourier analysis of the betatron oscillations can provide a wealth of information about the nonlinear dynamics of the beam in the storage ring [10]. The basic idea is to connect the amplitude and phase of the Fourier coefficients of the spectral lines with the amplitude and phase of the driving terms of a given resonance. A more complete investigation in the framework of the map pointed out that a full reconstruction of the nonlinear model is possible **at least** if turn-by-turn data with sufficiently high precision are available [11]. More recently a new algorithm has been proposed for the reconstruction of the nonlinear machine model based entirely on the comparison of the amplitude and phase of the spectral lines [12]. The model reconstruction has been demonstrated in tracking data and a first experimental investigation has shown that it is indeed possible to correct simultaneously several resonance driving terms [13]. The procedure introduced in [12] mimics closely the approach that LOCO takes to correct the linear optics of the ring where the role of the orbit response matrix is taken by the Fourier coefficients

of the spectral lines excited by nonlinear resonance driving terms, measured at all BPMs, and the role of the quadrupoles is now taken by the sextupoles. A fit procedure aims at defining the sextupole values which match the spectral lines measured on the machine with the one obtained from numerical tracking from the model. We have recently substantially improved the fit algorithm by taking into account the phase information by considering the real and the imaginary part of the spectral lines. When applied to tracking data this information allows a faster and more robust reconstruction. Its application to experimental data will be carried out in the near future.

Several experiments with pinged beams were performed at Diamond. They showed that the amplitude of the spectral lines related to nonlinear resonances can be measured with very good precision and corrected to restore the original pattern of the amplitude along the ring. In the experiment we targeted the amplitude of the $Q_x - Q_y$ spectral line measured in the vertical plane and the amplitude of the $-2Q_x$ spectral line measured in the horizontal plane. These are related to the driving terms of the resonances $Q_x \pm 2Q_y$ and $3Q_x$ respectively. We have verified experimentally that targeting a single spectral line can produce a good correction of the driving term as shown in Fig. 7. However this does not necessarily improve the DA and momentum aperture of the ring and can produce unrealistic sextupole gradients (black line in Fig. 9) If two spectral lines are taken into account the correction can have beneficial effects on the performance of the ring and an improvement of the Touschek lifetime of 10% was measured.

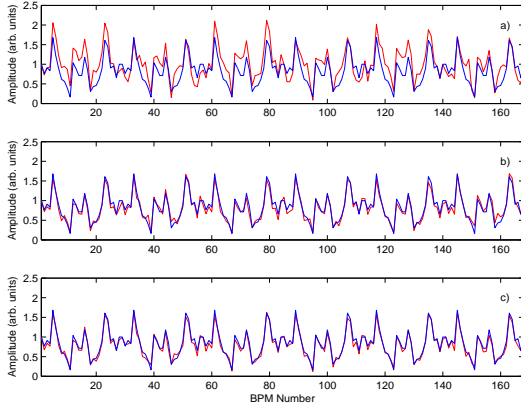


Fig. 7: Measurement and correction of the spectral line $Q_x - Q_y$ measured in the vertical plane: red – measured, blue – model, before correction (top) after one iteration (middle) after two iterations (bottom).

This technique has some limits: firstly it is based on the assumption that the first order perturbative theory adequately describes the nonlinear beam dynamics, secondly it relies on very precise measurements of the turn-by-turn data. Decoherence of the excited oscillations reduces the number of turns available and the machine tune stability has also to be controlled carefully if meaningful results are to be extracted. Nevertheless the indication provided by the experiment shows that this

technique holds great potential for the characterisation of the nonlinear beam dynamics in storage rings.

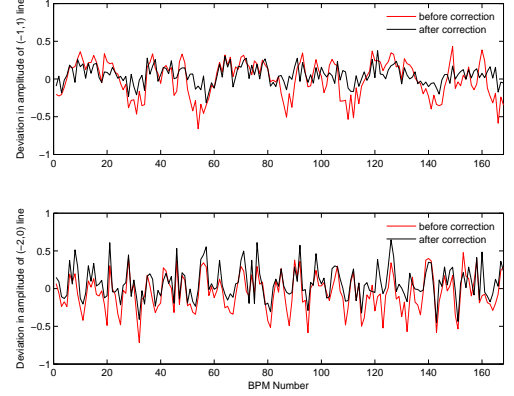


Fig. 8: Measurement and correction of the spectral line $Q_x - Q_y$ measured in the vertical plane (top) and of the $-2Q_x$ measured in the horizontal plane (bottom). Before correction (red), after correction (black).

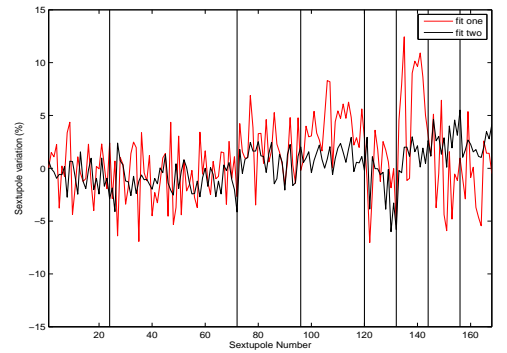


Fig. 9: Sextupoles gradient variation required by the fit procedure outlined in the text; black for the case where one resonance was targeted; red for the case where two resonances were targeted.

ACKNOWLEDGEMENTS

I would like to thank the Diamond Accelerator Physics Group, G. Rehm at Diamond., F. Schmidt at CERN and P. Kuske at Bessy-II for their contributions to this work.

REFERENCES

- [1] R.P. Walker, pg. 2718, EPAC06.
- [2] R. Bartolini, pg. 1109, PAC07.
- [3] J. Safranek, NIM **A338**, pg. 27, (1997).
- [4] G. Portmann, pg. 620, EPAC98.
- [5] R. Bartolini, pg. 993, EPAC08.
- [6] L. Nadolski, Tracy-II–Soleil version.
- [7] M. Borland, elegant, APS LS-287, (2000).
- [8] P.K. Skowronski et al., pg. 3381, PAC07.
- [9] H. S. Dumas et al., PRL **70**, pg. 2975, (1993)
- [10] J. Bengtsson, CERN 88-05, (1988).
- [11] R. Bartolini, Part. Acc. **59**, pg. 93, (1998).
- [12] R. Bartolini et al., pg. 1452, PAC05.
- [13] R. Bartolini PRSTAB **11**, 104002, (2008).

DESIGN AND CONTROL OF ULTRA LOW EMITTANCE LIGHT SOURCES*

Johan Bengtsson[#]

BNL, Upton, NY 11973, U.S.A.

Abstract

In the quest for brightness, the horizontal emittance remains one of the main performance parameters for modern synchrotron light sources. A control theory approach that takes the nonlinear dynamics aspects into account, using a few simple (linear) optics guidelines, at an early stage generates robust designs. Modern analytic- and computational techniques enable the optics designer to avoid the fallacy of the traditional approach guided by the Theoretical Minimum Emittance (TME) cell: the "chromaticity wall". In particular, by using an interleaved computational approach with the nonlinear dynamics analyst/model. We also outline how to implement the correction algorithms for a realistic model so that they can be re-used as part of an on-line model/control server for commissioning- and operations of the real system.

TRADE-OFFS: GLOBAL OPTIMIZATION

The (natural) horizontal emittance ε_x originates from the equilibrium:

damping \leftrightarrow diffusion

of three different processes: radiation damping, quantum fluctuations, and IntraBeam Scattering (IBS). One can show that (fundamental limit is IBS):

$$\varepsilon_x \sim \frac{1}{R^2 \cdot P}$$

where R is the bend radius, and P the radiated power.

The design of a synchrotron light source is essentially a matter of balancing the conflicting entities schematized in Fig. 1 (optimized for Insertion Device (ID) beam lines) [1].

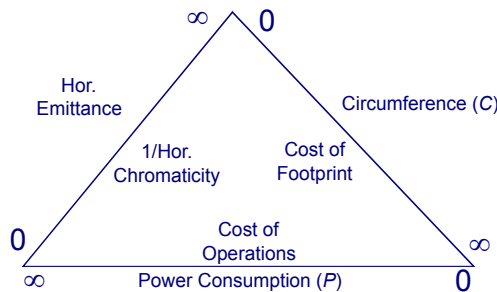


Figure 1: Synchrotron Light Source Optimization.

Traditionally, the approach has been driven by the so-called Theoretical Minimum Emittance (TME) cell [2-3].

[#]bengtsson@bnl.gov

*Work supported by U.S. DOE, Contract No.DE-AC02-98CH10886.

However, the approach is misguided (reductionalist), since it only considers the linear optics, i.e., ignores how to control the resulting (linear) chromaticity, and hence does not lead to realistic/robust designs. In particular, it creates an artificial "chromaticity wall" [4]. It also leads to lattices with dispersion at the cavity; which potentially increases the effective transverse beam size due to synchro-betatron coupling (by i.e. operating with finite (linear) chromaticity).

To capture the control aspects of the nonlinear dynamics from the start of the NSLS-II, we have provided the following (linear) optics guidelines [5]:

- max chromaticity per cell,
- min peak dispersion,
- max values for the beta functions.

For an intuitive (systems) approach, see e.g. the MAX-IV conceptual design [6].

WHAT'S KNOWN

The first dedicated third generation light sources were commissioned in the early 80s, i.e., they have been optimized for over 20 years. Basically:

- The horizontal emittance (isomagnetic lattice) is given by

$$\varepsilon_x [\text{nm}\cdot\text{rad}] = 7.84 \times 10^3 \cdot \frac{(E[\text{GeV}])^2 F}{J_x N_b^3}$$

where N_b is the number of dipoles, $J_x + J_z = 3$,

and $F \geq 1$. No dipole gradients $\Rightarrow J_x \sim 1$.

- Generalized Chasman-Green lattices: DBA, TBA, QBA, 7-BA [6].
- Effective emittance \Rightarrow chromatic cells.
- Increasing N_b reduces ε_x but also reduces the peak dispersion, which makes the chromatic correction less effective \Rightarrow "chromaticity wall".
- Damping wigglers (DWs): damping rings and conversion of HEP accelerators [7-8].
- Mini-Gap Undulators (MGUs), Three-Pole-Wigglers (TPWs) inside the DBA [9].

WHAT'S NEW

The NSLS-II design is conservative, i.e., it is based on well known techniques, but the approach is also novel because it combines these in a unique way:

- Use of damping wigglers to reduce horizontal emittance and as high flux X-ray sources \Rightarrow achromatic cells and weak dipoles.
- Medium energy ring (3 GeV) with ~ 30 DBA cells.
- Vertical orbit stability requirements.
- Generalized higher order achromat.

CHALLENGES

Given the design goals and approach, challenges related to non-linear dynamics issues are:

- Medium energy: control of Touschek lifetime and momentum aperture.
- 30 DBA cells: control of tune footprint.
- Control of impact of DWs and IDs -> include leading order nonlinear effects from DWs in the Dynamic Aperture (DA) optimizations.
- Optics requirements for IDs and top-up injection are contradictory: introduce alternating straights with high- and low horizontal beta functions => reduced symmetry (30 => 15).
- DBA: momentum dependence of optics functions => sufficient number of chromatic sextupole families.

There are also technical challenges:

- Weak dipoles: introduce TPWs (adjacent to the dipoles) => control of peak beta functions and horizontal dispersion.
- Vertical orbit stability: sub micron => pushing the state-of-the-art [10-11].

LATTICE PARAMETERS

The main lattice parameters are summarized in Tab. 1, where values specific to the NSLS-II are in bold type.

Table 1: NSLS-II Lattice Parameters

Energy (E_0)	3 GeV
Circumference (C)	791.5 m
Beam Current (I_b)	500 mA
Bending Radius (R)	25.0 m
Dipole Energy Loss (U_0)	286.5 keV
Emittance: (ϵ_x, ϵ_y) bare/w. 8 DWs	(2.1, 0.01)/(0.6, 0.01) nm·rad
Momentum Compaction	0.00037
RMS Energy Spread: bare/w. 8 DW	0.05/ 0.1%
Working Point (ν_x, ν_y)	(32.4, 16.3)
Chromaticity (ξ_x, ξ_y)	(-100, -42)
Peak Dispersion ($\hat{\eta}_x$)	0.45 m
Beta Function (β_x, β_y): long/short straight	(18, 3)/(3, 3) m

ROBUST DESIGN AND CONTROL

Typically, the approach has been to first design the linear optics, and then attempt to control (fix) the DA, aka perturbative point-of-view. In other words, a “top-up” (reductionist) rather than “top-down” (systems) approach, see e.g. refs [12-14]. Clearly, a prerequisite for a robust design and effective commissioning is a realistic model, see Fig. 2.

Challenge: for a streamlined approach, how to re-use the design model for model based (on-line) control?

- See section MODEL BASED CONTROL.

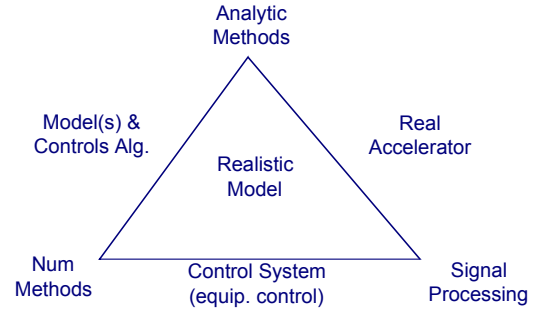


Figure 2: Prerequisite for Robust Design and Control: a Realistic Model.

For a systematic approach one may view the design process as “Closed-Loop” Control, see Fig. 3 applied to:

- lattice design,
- control of DA,
- guidelines for engineering tolerances, ring magnets, and insertion devices,
- correction algorithms,
- aka TQM (Total Quality Management) in industry.

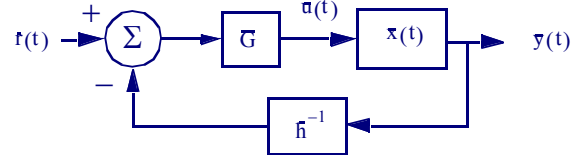


Figure 3: Closed-Loop Control Paradigm.

Similarly, a Use Case approach is a rational method to capture and refine the often elusive requirements for:

- model based control [15].

By treating the control system as an abstraction and analyzing how abstract “actors” (e.g. individuals, groups, other sub-systems, etc.) interact with the system, one avoids the typical gridlock between different stake holders (“What’s the requirement?” vs. “What’s the best you can do?”). Instead, by focusing the effort on “what” rather than “how”, the process provides for a sequence of successive refinements that generates a set of quantitative, measurable requirements: aka a spiral approach.

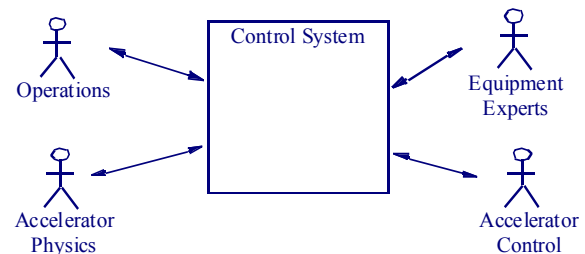


Figure 4: Use Case Approach.

MODELING CONSIDERATIONS

Of course, these methods assume that a realistic model has been provided. In particular, the following aspects must be addressed:

- A confinement problem governed by the Lorentz force: $d\vec{p}/dt = q(\vec{E} + \vec{v} \times \vec{B})$.
- The single particle dynamics is described by the relativistic Hamiltonian for a charged particle in an external electro-magnetic field (aka volume preserving flow) \Rightarrow Symplectic integrators.
- The residual beam size is in dynamic equilibrium between “cooling” from radiation damping (described by classical radiation), and “heating” due to diffusion from quantum fluctuations (i.e. recoil from the emitted photons) \Rightarrow Modified symplectic integrator.
- Need to model a realistic magnetic lattice, i.e., that includes mechanical misalignment- and magnetic field errors, and related correction algorithms.
- Must be able to compute- and optimize the global properties of a realistic lattice: the optics, diffusion coefficients, driving terms, tune foot print, etc.
- No theory of stability (for the general nonlinear case) \Rightarrow Perturbation theory. Hence, “analytic” results must be validated by numerical simulations.
- Control nonlinear effects by the: lattice symmetry, driving terms/resonances (Lie generators), and tune foot print; obtained either from Taylor maps, Lie series, and map normal form (analytically) or frequency maps (numerically).

Challenge: How to combine the numerical methods for modeling of a realistic lattice with the analytical techniques for analysis of its properties?

- Introduce a polymorphic number class for transparent floating point- and TPSA (Truncated Power Series Algebra) [16] calculations with object-oriented programming [17] \Rightarrow a Lagrangian object, aka PTC (Polymorphic Tracking Code) [18].

Challenge: How to re-use the beam dynamics model and related correction algorithms developed during the design phase as an on-line model for the commissioning?

- Implement a well designed software library that can be re-used by for instance the Controls Group.

MODEL

The Hamiltonian is (equations of motion for a medium size ring)

$$H = (1 + h_{\text{ref}}(s)) \left\{ -\frac{q}{p_0} A_s(s) + \frac{1}{2(1+\delta)} \left[\left(p_x - \frac{q}{p_0} A_x(s) \right)^2 + \left(p_y - \frac{q}{p_0} A_y(s) \right)^2 \right] - \delta + O(p_{x,y})^4 \right\}$$

with the multipole expansion

Synchrotron Radiation Facilities

$$\frac{q}{p_0} A_s(s) = -\text{Re} \left\{ \sum_{n=1}^{\infty} \frac{1}{n} (ia_n(s) + b_n(s)) (x + iy)^n \right\}$$

The map is obtained by splitting the Hamiltonian into two integrable parts ($: f(\vec{x}) : g(\vec{x}) \equiv [f(\vec{x}), g(\vec{x})]$)

$$H = H_{\text{drift}} + H_{\text{kick}}$$

which leads to

$$\begin{aligned} \mathcal{M}(\Delta s) &= \exp \left(: - \int_0^{\Delta s} H ds : \right) = e^{-H\Delta s} \\ &= e^{-H_{\text{drift}}\Delta s/2} e^{-H_{\text{kick}}\Delta s} e^{-H_{\text{drift}}\Delta s/2} + O(\Delta s^3) \\ &= \mathcal{M}_{\text{drift}} \left(\frac{\Delta s}{2} \right) \mathcal{M}_{\text{kick}}(\Delta s) \mathcal{M}_{\text{drift}} \left(\frac{\Delta s}{2} \right) + O(\Delta s^3) \end{aligned}$$

aka a 2nd order symplectic integrator. In particular, it can be generalized to 4th order.

For insertion devices, the vector potential can be obtained from the magnetic field (numeric model or measurements) by

$$A_x(s) = - \int_s B_y(z) dz, \quad A_y(s) = \int_s B_x(z) dz, \quad A_z(s) = 0$$

The corresponding kick $\mathcal{M}_{\text{kick}}$ map is provided for instance by RADIA [19].

For an analytic model (to leading order)

$$\begin{aligned} \frac{q}{p_0} A_x(s) &= \frac{B_u}{k_z(B\rho)} \cos(k_x x) \cosh(k_y y) \sin(k_z s), \\ \frac{q}{p_0} A_y(s) &= \frac{k_x B_u}{k_y k_z (B\rho)} \sin(k_x x) \sinh(k_y y) \sin(k_z s) \end{aligned}$$

$$\text{with } k_z^2 = k_y^2 - k_x^2 = (2\pi/\lambda_u)^2.$$

Assuming that the corresponding Taylor map has been obtained (to an arbitrary order) from the beam dynamics model, the map can be factored (Lie series)

$$\mathcal{M} = \mathcal{A}^{-1} \dots e^{f^{(4)}} e^{f^{(3)}} \mathcal{R} \mathcal{A}$$

The (Lie) generators (driving terms) provide a means to control the DA [20]. They can also be measured from turn-by-turn data and Fourier analysis [21] \Rightarrow “closing-the-loop” between model and the real lattice [22].

The map can also be (recursively) transformed into normal form [23]

$$\mathcal{M} =^{-1} e^{-g(\vec{J}, \vec{\phi})} e^{K(\vec{J})} e^{g(\vec{J}, \vec{\phi})} \mathcal{R} \mathcal{A}$$

from which we obtain the global properties of the lattice, e.g. the tune shift

$$\bar{\nu}(\vec{J}) = -\frac{1}{2\pi} \frac{\partial K(\vec{J})}{\partial(\vec{J})}$$

The information flow for the corresponding computer model is shown in Fig. 5.

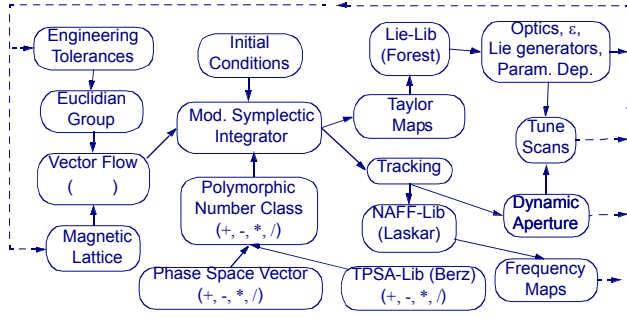


Figure 5: The NSLS-II “Wind Tunnel”, i.e., a Virtual Accelerator (aka Polymorphic Tracking Code).

DYNAMIC APERTURE

The requirements are:

- on-momentum Dynamic Aperture (DA): 11 mm (robust top-up injection),
- off-momentum DA: 2.5% (Touschek life time),
- Tune footprint for the bare lattice (w/ DWs): ~0.05 (to accommodate engineering tolerances, IDs, etc.).

The last requirement is based on a (conservative) estimate of the tune footprint for stable beam in existing medium energy light sources, i.e., about ~0.1.

Note, due to the high number of DBA cells (30), as compared to existing medium energy synchrotron light sources, the control of the amplitude dependent tune shift per cell needs to be about 3 times better for a similar nonlinear performance. Hence, tight engineering tolerances are required.

The DA is essentially determined by the tune footprint and the sextupolar resonances (to 2nd order in the sextupole strength, aka 4th order resonances).

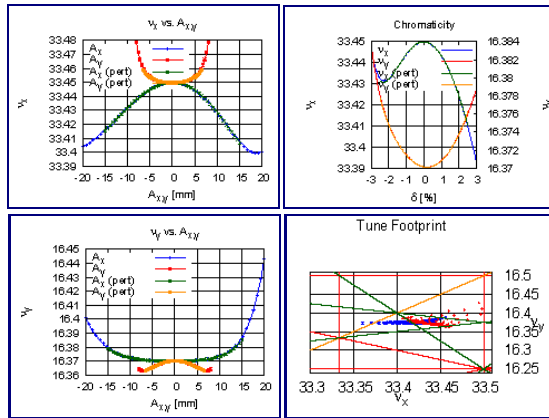


Figure 6: Tune Footprint (w/ DWs).

The resulting tune footprint and frequency maps after optimization are summarized in Figs 6 and 7. In particular, the introduction of DWs requires [24]:

- optics correction (local/global control of symmetry and working point),
- and sextupole re-optimization (due to the residual local optics perturbations from the DWs).

The analytic model for the amplitude dependent tune shift and residual nonlinear chromaticity need to include terms to 6th order in the sextupole strength:

$$\bar{\nu} = \bar{\nu}(\bar{J}) + O(\bar{J}, \delta)^4$$

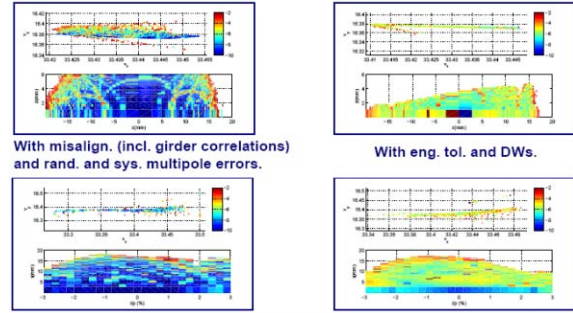


Figure 7: Impact of Engineering Tolerances and DWs.

MODEL-BASED CONTROL

Assuming that a realistic model of the system has been provided, and that a robust design has been delivered, which will be implemented, the question arises:

- How to control the real system?

Since a model-based approach is required, ideally, the model and controls algorithms developed during the design work would be re-used on-line. In particular, by pursuing a client/server approach [25].

Challenge:

- How to migrate from high level application prototypes developed for beam studies into thin (simple) clients suitable for day-to-day operations.

By providing a software architecture that provides both [26-27]:

- a flexible environment for rapid prototyping with a scripting language,
- and a model server with thin clients,

see Fig 8.

High level controls applications are ideally implemented and tested before commissioning, but one problem is that there is typically quite some lag time until hardware, etc. become available. It is thus desirable to have a simulator for the entire accelerator. A transparent approach is summarized in Fig. 9 [27-28].

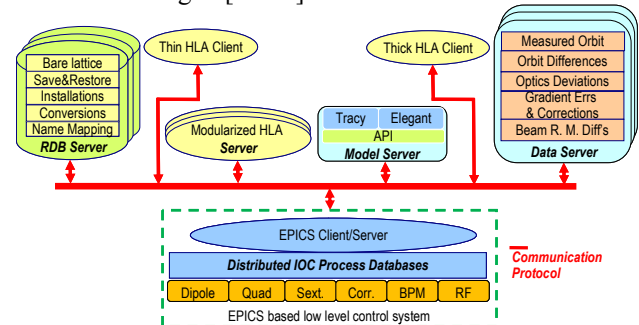


Figure 8: A Client/Server Architecture.

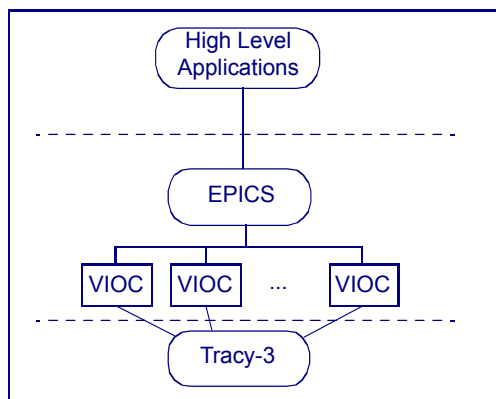


Figure 9: A Transparent Accelerator Simulator.

CONCLUSIONS

- By using modern methods, a self-consistent, realistic computer model has been implemented, i.e., where the same model is used for numerical simulations and by analytic techniques.
- The model has been used to guide the NSLS-II design. In particular, it has provided an effective framework to control the dynamic aperture, to provide guidelines for engineering tolerances, and magnet- and insertion device design. In other words, “closing-the-loop” between conceptual design and the performance of the final hardware.
- The model is also being used by the Controls Group, as a simulator for the accelerator, by interfacing with the control system, for e.g. testing of high level applications & controls algorithms. Furthermore, it also provides a transparent implementation of a model server with thin clients for the commissioning of the accelerator.

REFERENCES

- [1] S. Ozaki, J. Bengtsson, S. Kramer, S. Krinsky, V. Litvinenko “Philosophy for NSLS-II Design with Sub-Nanometer Horizontal Emittance” p 77-79 PAC07 (2007).
- [2] S.Y. Lee, L. Teng “Theoretical Minimum Emittance Lattice for an Electron Storage Ring” p 2679-2681 PAC91 (1991).
- [3] S.Y. Lee “Emittance Optimization in Three- and Multiple-Bend Achromats” Phys. Rev. E, vol 54, no 2, p 1940-1944 (1996).
- [4] R. Talman “Accelerator X-Ray Sources” (Wiley-VCH, Berlin, 2006).
- [5] NSLS-II Preliminary Design Report (2007).
- [6] M. Eriksson et al “The MAX-IV Design: Pushing the Envelope” p 74-76 PAC07 (2007).
- [7] H. Wiedemann “An Ultra-Low Emittance Mode for PEP Using Damping Wigglers” Proc. Synchr. Rad. Instr. Conf. p 24-32 (1987).
- [8] K. Balewski et al “Progress Report on PETRA III” p 3317-3319 EPAC06 (2006).
- [9] A. Nadji et al “A Modified Lattice for SOLEIL with Large Number of Straight Sections” SSILS (2001).
- [10] C. Steier et al “Operational Experience Integrating Slow and Fast Orbit Feedbacks at the ALS” p 2786-2788 EPAC04 (2004).
- [11] M. Böge et al “Commissioning of the Fast Orbit Feedback at SLS” p 3386-3388 PAC03 (2003).
- [12] J. Bengtsson, W. Joho, P. Marchand, L. Rivkin, A. Streun “Status of the Swiss Light Source Project SLS” p 685-687 EPAC96 (1996).
- [13] J. Bengtsson, W. Joho, P. Marchand, G. Mülhaupt, L. Rivkin, A. Streun “Increasing the Energy Acceptance of High Brithness Synchrotron Light Storage Rings” Nucl. Instr. Meth., A404, p 237-247 (1997).
- [14] J. Bengtsson “A Control Theory Approach for Dynamic Aperture” p 3478-3480 EPAC06 (2006).
- [15] J. Bengtsson, Bob Dalesio, Timur Shaftan, Toshi Tanabe “NSLS-II: Model Based Control – A Use Case Approach” NSLS-II Tech Note 51 (2008).
- [16] M. Berz “The Method of Power Series Tracking for the Mathematical Description of Beam Dynamics” Nucl. Instr. Meth., A258, p 431-436 (1987).
- [17] J. Bengtsson, E. Forest, CBP, LBL unpubl. (1994).
- [18] E. Forest, F. Schmidt, E. McIntosh “Introduction to the Polymorphic Tracking Code” CERN-SL-2002-044 (AP) KEK-Report 2002-3 (2002).
- [19] P. Elleaume, O. Chubar, J. Chavanne “Computing 3D Magnetic Fields from Insertion Devices” p 3509-3511 PAC97 (1997).
- [20] J. Bengtsson “The Sextupole Scheme for the Swiss Light Source” SLS Note 9/97 (1997).
- [21] J. Bengtsson “Non-Linear Transverse Dynamics for Storage Rings with Applications to the Low-Energy Antiproton Ring (LEAR) at CERN” CERN 88-05 (1988).
- [22] R. Bartolini, I.P.S. Martin, J.H. Rowland, P. Kuske, F. Schmidt “Correction of Multiple Nonlinear Resonances in Storage Rings” Phys. Rev. ST 11, 104002 (2009).
- [23] E. Forest “A Hamiltonian-Free Description of Single Particle Dynamics for Hopelessly Complex Periodic Systems” J. Math. Phys. p 1133-1144 (1990).
- [24] J. Bengtsson “NSLS-II: Control of Dynamic Aperture” BNL-81770-2008-IR (2008).
- [25] M. Böge “Commissioning of the SLS Using CORBA Based Beam Dynamics Applications” p 292-294 PAC01 (2001).
- [26] B. Dalesio “NSLS-II Control System” ICALEPCS07 (2007).
- [27] G. Shen “A Software Architecture for High Level Applications” PAC09 (2009).
- [28] M.T. Heron et al “Progress of the Implementation of the DIAMOND Control System” ICALEPCS05 (2005).

NOVEL METHODS FOR SIMULATING RELATIVISTIC SYSTEMS USING AN OPTIMAL BOOSTED FRAME*

J.-L. Vay[†], W. M. Fawley, C. G. Geddes, E. Cormier-Michel, LBNL, Berkeley, CA, USA
D. P. Grote, LLNL, CA, USA

Abstract

It can be computationally advantageous to perform computer simulations in a Lorentz boosted frame for a certain class of systems. However, even if the computer model relies on a covariant set of equations, it has been pointed out that algorithmic difficulties related to discretization errors may have to be overcome in order to take full advantage of the potential speedup. We summarize the findings, the difficulties and their solutions, and show that the technique enables simulations important to several areas of accelerator physics that are otherwise problematic, including self-consistent modeling in three-dimensions of laser wakefield accelerator stages at energies of 10 GeV and above.

INTRODUCTION

In [1], we have shown that the ratio of longest to shortest space and time scales of a system of two or more components crossing at relativistic velocities is not invariant under a Lorentz transformation. This implies the existence of an “optimum” frame of reference minimizing a measure of the ratio of space and time scales. Since the number of computer operations (e.g., time steps), for simulations based on formulations from first principles, is proportional to the ratio of the longest to shortest time scale of interest, it follows that such simulations will eventually have different computer runtimes, yet equivalent accuracy, depending solely upon the choice of frame of reference. The scaling of theoretical speedup was derived for a generic case of two crossing identical rigid particle beams, and for three particle acceleration related problems: particle beams interacting with electron clouds [2], free electron lasers (FEL) [3], and laser-plasma accelerators (LWFA) [4]. For all the cases considered, it was found that the ratio of space and time scales varied as γ^2 for a range of γ , the relativistic factor of the frame of reference relative to the optimum frame. For systems involving phenomena (e.g., particle beams, plasma waves, laser light in plasmas) propagating at large γ , demonstrated speedup of simulations being performed in an optimum boosted frame can reach several orders of magnitude, as compared to the same simulation being performed in the laboratory frame.

We summarize the difficulties and limitations of the method, the solutions that were developed, and the simu-

lations that we have performed to date. We show that the technique enables simulations important to several areas of accelerator physics that are otherwise problematic. For the first time, it allows for direct self-consistent simulations of laser wakefield accelerator stages at 10 GeV and beyond using current supercomputers in a few hours, while the same calculations in the laboratory frame would take years using similar resources and are thus impractical. It also allows simulations of electron cloud effects in high energy physics accelerators (modeled so far with codes based on quasistatic approximations) using more standard Particle-In-Cell methods. This renders these types of simulations accessible to a wider range of existing computer codes, alleviates the added complication due to pipelining when parallelizing a quasistatic code, and removes the approximations of the quasistatic method which may not be applicable in some situations. For free electron lasers, the new technique offers the possibility of calculating self-consistently configurations that are not accessible with standard FEL codes due to the limitations of the approximations that they are based on. Finally, the method may offer a unique way of calculating self-consistently, and in three-dimensions, coherent synchrotron radiation effects which are of great importance in several current and future accelerators.

DIFFICULTIES

Even if the fundamental electrodynamics and particles equations are written in a covariant form, the numerical algorithms that are derived from them may not retain this property and special techniques have been developed to allow simulations in boosted frames. As an example, we considered in [5] an isolated beam propagating in the laboratory frame at relativistic velocity. When applying the effect of the beam field on itself using the Newton-Lorentz equation of motion, the contribution from the radial electric field is largely canceled by the contribution from the azimuthal magnetic field. However, we showed that the so-called ‘Boris particle pusher’ [6] (which is widely used in PIC codes), does make an approximation in the calculation of the Lorentz force which leads to an inexact cancellation of the electric component by the magnetic component. The magnitude of the error grows with the beam relativistic factor and in practice, it is unacceptably large for simulations of ultra-relativistic charged beams, where the cancellation needs to be nearly complete. The issue was resolved by changing the form of the Lorentz force term in the Boris pusher, and solving analytically the resulting implicit system of equations (see [5] for details).

* Work supported by US-DOE Contracts DE-AC02-05CH11231 and DE-AC52-07NA27344, US-LHC program LARP, and US-DOE SciDAC program ComPASS. Used resources of NERSC, supported by US-DOE Contract DE-AC02-05CH11231.

[†] jlvey@lbl.gov

An additional practical complication of numerical simulation in a boosted frame is that inputs and outputs are often specified (or desired) in the laboratory frame. For example, in LWFA simulations, laser and plasma parameters have to be transformed from the laboratory to the new relativistic boosted frame, so that the electromagnetic waves will be Doppler-shifted, and the background plasma, with higher density, is now drifting. In the PIC code Warp [9], the laser is injected at a plane that is fixed in the laboratory frame and drifting in the boosted frame. Likewise, the initial phase-space distribution of a particle beam is generally known in the laboratory. For calculations in boosted frames of large γ , deriving the initial beam conditions at a given time can be easy if the initial conditions are simple (e.g., initial Gaussian beam in vacuum), or more difficult and/or computationally costly if injecting the beam in a particle accelerator for example, where its longitudinal extent in the boosted frame can cover several lattice periods. In order to circumvent this difficulty, a procedure was implemented in Warp which injects the beam through a transverse plane that is fixed in the laboratory, but drifting in the boosted frame, similarly to the laser injection method. Due to long range space charge forces, it is still necessary to provide a reasonable estimate of the beam distribution near the injection plane; this is accomplished by the use of “frozen” drifting macroparticles.

After the relativistic PIC algorithm evolves the system in the boosted frame, the results must be transformed back to the laboratory frame. We have found it convenient in Warp to record quantities at a number of regularly spaced “stations”, immobile in the laboratory frame, at a succession of discrete times, for both detailed time histories and laboratory time-averages. Since the space-time locations of the diagnostic grids in the laboratory frame generally do not coincide with the space-time positions of the macroparticles and grid nodes used for the calculation in a boosted frame, some interpolation is performed during the data gathering process.

Finally, in simulations of laser-plasma acceleration stages (see below), we observed a short wavelength instability with a growth rate that rises with the velocity of the boosted frame and the inverse of the grid resolution, which we have controlled through the use of low dispersion electromagnetic solvers [7] and low-pass digital filtering. The details of the instability and its cures will be detailed in a future paper [8].

Together with mitigation of numerical artifacts as just described, these techniques allow simulations using boosted frames, with orders of magnitude speedup over the same simulations performed using a laboratory frame, as shown below. Additional details of the input and output procedures can be found in [10].

Computer Codes (Design, Simulation, Field Calculation)

EXAMPLES OF APPLICATION

Laser wakefield acceleration

Laser driven plasma waves offer orders of magnitude increases in accelerating gradient over standard accelerating structures (which are limited by electrical breakdown), thus holding the promise of much shorter particle accelerators. Yet, computer modeling of the wake formation and beam acceleration requires fully kinetic methods and large computational resources due to the wide range of space and time scales involved [13]. For example, modeling 10 GeV stages for the LOASIS (LBNL) BELLA proposal [14] in one-dimension demanded as many as 5,000 processor hours on a NERSC supercomputer [15]. As discussed in [1], the range of scales can be greatly reduced if one adopts the common assumption that the backward-emitted radiation can be neglected, enabling, for the first time, the full-PIC simulation of the next generation of laser systems.

Warp simulations at plasma density $n_e = 10^{19} \text{ cm}^{-3}$ were performed in 2-1/2D and 3D using reference frames moving anywhere between $\gamma_f = 1$ (laboratory frame) and 10. These simulations are scaled replicas of 10 GeV stages that would operate at actual densities of 10^{17} cm^{-3} [16, 17] and allow short run times to permit effective benchmarking between the algorithms. Agreement within a few percent was observed on the beam peak energy and average energy between calculations in all frames, showing that the boosted frame simulations gain speed without sacrificing accuracy. A speedup of 100 was measured between the calculation in the frame at $\gamma = 10$ and the calculation in the laboratory frame.

The boosted frame model was then used to conduct full scale simulations of 10 GeV stages at plasma densities of 10^{17} cm^{-3} in 2-1/2D and 3D simulations. Simulations at 10^{18} cm^{-3} were also conducted to establish scaling. Relativistic factors of the boosted frame were 130 and 40 respectively, i.e. close to the relativistic factor associated with the wake velocity in the laboratory frame. The 3D run at full scale took almost 4 hours with $\gamma_f = 130$ using 512 cores on the cluster Lawrence Livermore at LBNL. This provided direct simulation of next generation experiments and possible laser-plasma collider stages. Good agreement with the scaled energy gain was obtained. Such simulations are impractical in the laboratory frame, with projected time of 15 years on the same resources using the $2\gamma_f^2$ formula for the estimated speedup, and scaling from standard PIC runs.

Electron cloud driven instabilities

Several existing and planned future particle accelerators have limitations due to the electron cloud instability that may negatively impact the beam quality and in some cases even lead to severe beam loss. A calculation of electron cloud driven instability [2] for an ultra-relativistic beam was performed with the Warp code framework in (a) standard PIC mode using the new particle pusher in a Lorentz boosted frame; (b) in quasistatic mode [11] using linear

maps to push beam particles into the accelerator lattice. The two runs were in good agreement and completed using similar computer resources and runtimes. The speedup factor of the PIC boosted frame calculation compared to a PIC calculation in the laboratory frame was estimated at 500. For many calculations of electron cloud instability, the boosted frame approach may not resolve any additional physics not included in the quasistatic approach. We note, however, that the quasistatic method requires significant special coding to take into account eventual longitudinal motion of electrons [11], as well as a special parallelization scheme [12] for parallelization along the axis of beam propagation, which are not standard to PIC codes. By contrast, the boosted frame method includes naturally the longitudinal dynamics and requires more modest modifications to an existing standard PIC code or framework (none for parallelization, if the PIC code is already parallel).

Free electron lasers

In a short wavelength free-electron laser, a high energy electron beam interacts with a static magnetic undulator. In the optimal boost frame with Lorentz factor γ , the down-shifted FEL radiation and up-shifted undulator have identical wavelengths and the number of required time-steps (presuming the Courant condition applies) decreases by a factor of $2\gamma^2$ for fully electromagnetic simulations. Examples of boosted-frame simulations have been compared [18] to results obtained with the eikonal (i.e., SVEA) and wiggler-period averaged code Ginger [19]. It was concluded that if the necessary FEL physics can be studied with an eikonal code, it will run much faster than a full electromagnetic code in whatever frame. However, if there are important physical phenomena that cannot be resolved properly by an eikonal code, a boosted-frame electromagnetic code is a very attractive alternative to a brute force full electromagnetic calculation in the laboratory frame.

Coherent synchrotron radiation

Another application for which the Lorentz-boosted frame method might be useful is that of modeling coherent synchrotron radiation (CSR) [20] emitted by high current, high brightness relativistic electron beams. Because full scale electromagnetic simulation of CSR in the laboratory frame is difficult due to the wide range of scales (chicane lengths of order meters, radiation wavelengths of orders microns), in order to make the calculation tractable most CSR simulation codes apply simplifications such as ignoring transverse variation of CSR across the electron beam. We have begun preliminary work of simulating CSR emission with the boosted frame method with Warp, examining the behavior of a high current, short electron beam transiting a simple dipole magnet. Our early results show that upon exit from the undulator the electron beam shows the characteristic energy loss variation with longitudinal position that one expects from previous theoretical analyses of CSR. Further studies are currently underway.

Computer Codes (Design, Simulation, Field Calculation)

CONCLUSION

The non-invariance of the range of scales of a physical system implies that the computational cost of a certain class of computer simulations depends strongly on the choice of the simulation frame of reference. Algorithmic difficulties arise due to the loss of covariance upon discretization of the Maxwell-Vlasov system of equations, and the need to transform input/output data between the laboratory frame and the Lorentz boosted frame. So far, the difficulties that have arisen have been overcome and no “show-stopper” has been identified at this time. First principles simulations in boosted frames have been performed successfully with the code Warp in application to laser wakefield acceleration, electron cloud driven instabilities and free electron lasers, with speedups ranging between a few and several orders of magnitude. Our recent progress show that first principles modeling in a Lorentz boosted frame is a viable alternative or complement to using reduced descriptions like the quasistatic [11] or eikonal [19] approximations, or performing simulations with scaled parameters [17], and in many cases includes physics that is not accessible to the other descriptions. This includes direct three-dimensional simulations of laser wakefield accelerator stages at 10 GeV and beyond, electron cloud effects in high energy physics accelerators, physics that is inaccessible to standard free electron lasers codes, and coherent synchrotron radiation.

ACKNOWLEDGMENTS

We are thankful to D. L. Bruhwiler, J. R. Cary, B. Cowan, E. Esarey, M. A. Furman, C. Huang, S. F. Martins, W. B. Mori, B. A. Shadwick, C. B. Schroeder and M. Venturini for insightful discussions.

REFERENCES

- [1] J.-L. Vay, *Phys. Rev. Lett.* **98** (2007) 130405.
- [2] *Proc. International Workshop on Electron-Cloud Effects*, Daegu, S. Korea (2007).
- [3] N. Kroll, P. Morton, M. Rosenbluth, *IEEE J. Quantum Electron* **QE-17** (1981) 1436.
- [4] T. Tajima and J. M. Dawson, *Phys. Rev. Lett.* **43** (1979) 267.
- [5] J.-L. Vay, *Phys. Plasmas*, **15** (2008) 056701.
- [6] J. P. Boris, *Proc. Fourth Conf. Num. Sim. Plasmas*, Naval Res. Lab., Wash., D. C. (1970) 3-67.
- [7] M. Karkkainen, E. Gjonaj, T. Lau, T. Weiland, *Proc. International Computational Accelerator Physics Conference*, Chamonix, France (2006).
- [8] J.-L. Vay *in preparation*.
- [9] D. P. Grote, A. Friedman, J.-L. Vay, I. Haber, *AIP Conf. Proc.* **749** (2005) 55.
- [10] J.-L. Vay *et al.*, *Proc. Particle Accelerator Conference*, Vancouver, Canada (2009) TU1PB104.
- [11] C. Huang *et al.*, *J. of Comput. Phys.* **217** (2006) 658-679.
- [12] B. Feng, C. Huang, V. Decyk, W.B. Mori, P. Muggli, T. Katsouleas, *J. Comput. Phys.* **228** (2009) 5340.

- [13] C. G. R. Geddes *et al.*, *J. Phys. Conf. Series V* **125** (2008) 12002/1-11.
- [14] <http://loasis.lbl.gov>
- [15] D. L. Bruhwiler *et al.*, *Proc. 13th Advanced Accelerator Concepts Workshop*, Santa Cruz, CA (2008) 29.
- [16] E. Cormier-Michel, *et al. Proc. 13th Advanced Accelerator Concepts Workshop*, Santa Cruz, CA (2008) 297.
- [17] C. G. R. Geddes *et al.*, *Proc. Particle Accelerator Conference*, Vancouver, Canada (2009) WE6RFP075.
- [18] W. M. Fawley, J.-L. Vay, *Proc. Particle Accelerator Conference*, Vancouver, Canada (2009) WE5RFP029.
- [19] W. M. Fawley, LBNL Tech. Rpt. LBNL-49625-Rev. 1 (2004); also SLAC Rpt. LCLS-TN-04-3.
- [20] M. Venturini, *et al.*, *Phys. Rev. ST Accel. Beams* **8** (2005) 014202.

MODELING TECHNIQUES FOR DESIGN AND ANALYSIS OF SUPERCONDUCTING ACCELERATOR MAGNETS*

P. Ferracin, LBNL, Berkeley, CA 94720, USA

Abstract

Superconducting magnets for particle accelerators are complex devices requiring the use of sophisticated modelling techniques to predict their performance. A complete description of the magnet behaviour can only be obtained through a multi-physics approach which combines magnetic models, to compute magnetic fields and electro-magnetic forces, mechanical models, to study stresses arising during assembly, cool-down and excitation, and electrical-thermal models, to investigate temperature margins and quench phenomena. This approach is essential in particular for the next generation of superconducting accelerator magnets, which will likely implement strain sensitive conductors like Nb_3Sn and will handle forces significantly larger than in the present LHC dipoles. The design of high field superconducting magnets has benefited from the integration between CAD, magnetic, and structural analysis tools allowing a precise reproduction of the magnet 3D geometry and a detailed analysis of the three-dimensional strain in the superconductor. In addition, electrical and thermal models have made possible investigating the quench initiation process and the thermal and stress conditions of the superconducting coil during the propagation of a quench. We present in this paper an overview of the integrated design approach and we report on simulation techniques aimed to predict and improve magnet behaviour.

INTRODUCTION

The R&D on the next generation of superconducting magnets for particle accelerators is currently focused on quadrupoles and dipoles for future luminosity and energy upgrades of the LHC [1]. Other possible applications include neutrino factories and cable test facilities [2]. These magnets will operate at field approaching 15 T, i.e. beyond the limits of NbTi superconductor, and with stored energies and electro-magnetic (e.m.) forces significantly larger than in the magnets presently used in the LHC. At the moment, Nb_3Sn , the only practical superconductor capable of generating fields higher than 10 T, appears as the best candidate for this future generation of superconducting magnets. However, Nb_3Sn is a brittle and strain-sensitive superconductor whose current carrying capability depends on its strain status [3,4]. As a result, the performance of Nb_3Sn magnets can be strongly affected by the mechanical stresses in the windings during magnet operation. It is therefore mandatory to understand and predict the strain in the superconductor, and devise a support structure capable of minimizing the stresses in the coils from magnet assembly to excitation.

*Supported by the Office of Science, Office of Basic Energy Sciences, U. S. Department of Energy under Contract No. DE-AC02-05CH1123.
#pferracin@lbl.gov

The computation of the mechanical status of the superconducting material is a very complex task, considering all the stages involved in the fabrication of Nb_3Sn coils, like cabling, winding, heat treatment to 650 °C, and epoxy impregnation (Fig. 1). These steps are then followed by magnet assembly, pre-loading, cool-down and powering, which further contribute to the final strain conditions of the Nb_3Sn superconductor.

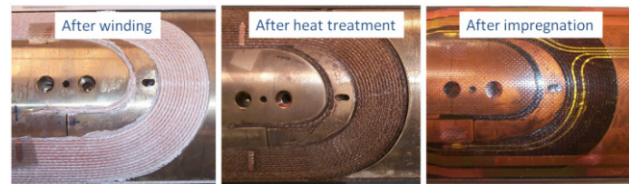


Figure 1: Nb_3Sn coil after winding (left), after reaction (center), and after impregnation (left).

We present in this paper an overview on modelling works performed in the LBNL Superconducting Magnet Program and aimed at design Nb_3Sn superconducting magnets, predict their behaviour, and analyze and improve quench performance. We start with a description of the tools and techniques adopted, and we then discuss how the models can be used to optimize coil and magnet lay-outs, improve fabrication process, and predict and minimize coil stress from assembly to quench.

INTEGRATED MODELING: TOOLS AND TECHNIQUES

The design and analysis of superconducting magnets can be seen as one single process that integrates different tools to provide a full characterization of the magnet components during assembly, cool-down, magnet excitation and quench. We present in this section an overview of codes and techniques utilized for superconducting magnet design, starting from simplified scaling laws to full 3D magnet models. A complete description of the integrated design approach applied to accelerator magnets can be found in [5].

Coil and Magnet 2D Design

The first step of magnet design consists in a preliminary estimate, through analytical tools or scaling laws, of the amount of conductor required for a given field and aperture [6-8]. Then, the definition of a 2D cross-section of superconducting cable, coil and support structure constitutes the second design step. In this phase, a 2D analysis of the magnetic and mechanical behaviour of the magnet can be performed with programs like Poisson [9], Roxie [10], Opera 2D [11] and ANSYS [12] (Fig. 3). The output of such programs gives field, harmonics and short-sample predictions for the magnet performance, as well as stress in all magnet components.

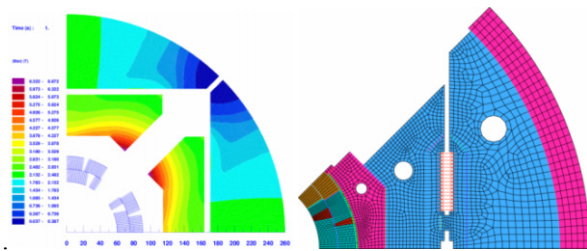


Figure 3: 2D finite element model for magnetic (Roxie, left) and mechanical (ANSYS, right) analysis.

Coil 3D Design

With the optimized coil cross-section coordinates, it is then possible to proceed and complete the cable windings through return and lead ends, using computer programs such as Bend [13] and Roxie. The resulting full 3D model of the coil, including each individual turn (Fig. 4), can be then uploaded in CAD programs, like ProE [14], for the coil parts fabrication, or in Opera 3D and ANSYS for magnetic and mechanical analysis (Fig. 5).

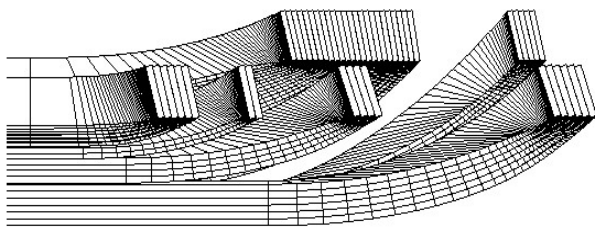


Figure 4: 3D coil model generated by the program Bend.

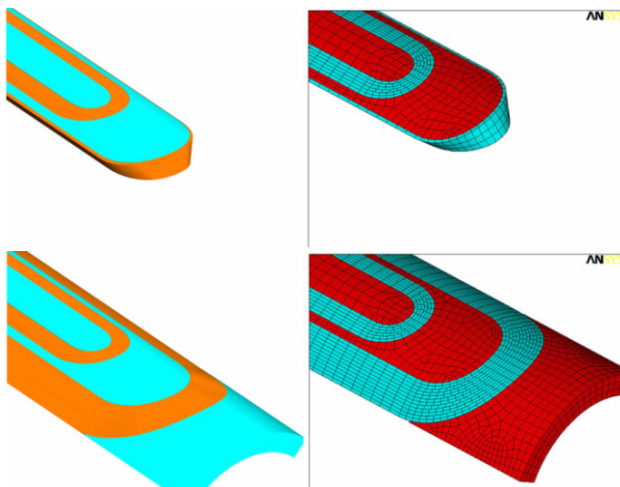


Figure 5: 3D coil model imported in ProE (left) and ANSYS (right).

Magnet 3D Design

At this point, the final step consists in the implementation of the full 3D magnet geometry. The CAD models are used to define assembly and loading procedures, and to generate drawings for part fabrication. Opera 3D and Roxie models compute conductor peak fields in the coil ends, and optimize iron geometry and field quality in the end regions (Fig. 6). Finally, an ANSYS 3D model can focus on the design of the support structure, with particular emphasis on the coil axial

support system. The integration of all these programs will have as an output a full description of the magnetic, mechanical and thermal condition of the superconductor, and will provide fundamental information to analyze magnet performance, identify quench triggering mechanisms, and define corrective strategies to improve training and minimize stress in the superconductor.

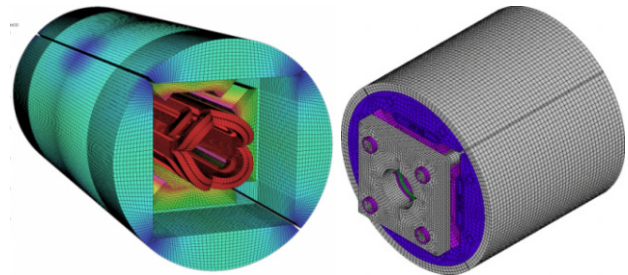


Figure 6: 3D finite element model for magnetic (Vector Field, left) and mechanical (ANSYS, right) analysis.

COIL AND SUPPORT STRUCTURE DESIGN

In Nb_3Sn magnets, the risk of conductor degradation due to high stress requires that already from the initial phases of the design, aimed at defining the 2D coil and magnet cross-sections, one of the main objectives must be minimizing the accumulated stress on the conductor. At the same time, the support structure must provide enough coil pre-load after cool-down so that turns do not separate from coil parts when the e.m. forces start to act on the windings. We present in this section the results of a conceptual design study focused at the optimization of the coil and support structure cross-sections of quadrupoles.

Coil Cross-Section Optimization

In a shell-type coil, the e.m. azimuthal forces are directed towards the mid-plane, where the accumulated stress reaches its peak at high field. In a multi-layer coil configuration, the choice of the number of turns per layer can play a significant role in reducing the accumulated stresses.

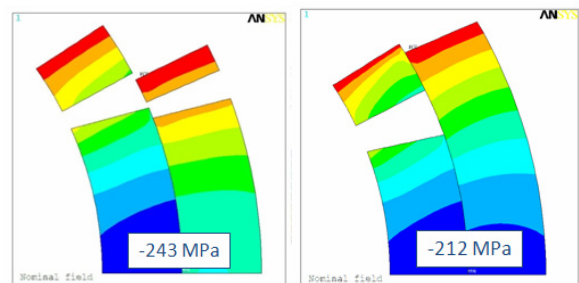


Figure 7: 2D computations of azimuthal stress with e.m. forces. Red (blue) contours indicate area of low (high) compressive stress.

In Fig. 7 we plotted the stress at maximum field of two different coil designs considered for the LHC Accelerator Research Program (LARP) quadrupole magnet HQ

[15,16]. In the first design (left cross-section), the first layer is characterized by a larger number of turns than the second layer. The consequent larger accumulated e.m. force determines a peak stress of -243 MPa. By redistributing the number of turns between the two layers (right cross-section) the total e.m. force is shared and the accumulated stress reduced by 30 MPa.

It is important to notice that both designs reach a similar maximum gradient, but the first one is clearly more efficient in term of amount of conductor. In this case, conductor efficiency was sacrificed to improve stress profiles.

Magnet Cross-Section Optimization

Once the coil design is chosen, the following step regards the optimization of the support structure. By providing pre-load to the coil during assembly and cool-down, and counteracting the e.m. forces during excitation, the structure may have a significant impact on the conductor peak stresses. An example is depicted in Fig. 8 and Fig. 9 where the accumulated stress in the coil at maximum field in the LARP quadrupole magnets TQ and LQ are shown. Although both quadrupoles implemented the same coil and operated at the same gradient, a modification of the geometry of the components surrounding the coil implemented in LQ contributed to a 20 MPa reduction of accumulated stress [17].

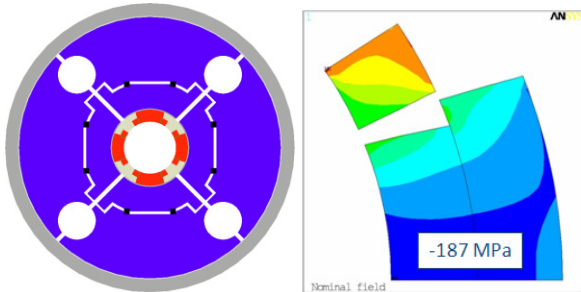


Figure 8: Cross-section of the LARP quadrupole magnet TQ (left) and 2D computations of azimuthal stress with e.m forces (right).

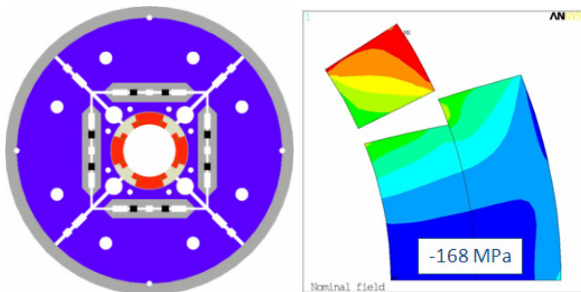


Figure 9: Cross-section of the LARP quadrupole magnet LQ (left) and 2D computations of azimuthal stress with e.m forces (right).

OPTIMIZATION OF COIL FABRICATION PROCESS

The use of finite element magnetic-mechanical models can play an important role also in the optimization of the Magnets

coil fabrication process. The models can be focused on investigating the effect of different coil part materials, which must be compatible with the high-temperature heat treatment, at the same time contributing to minimize coil strain during magnet operation. As an example, we present in this section the results of the study, reported in [18-20], regarding the winding poles of the LARP quadrupole magnet TQ.

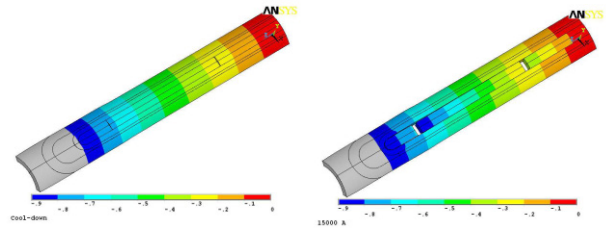


Figure 10: 3D mechanical computations of the deformed shapes (with displacements enhanced by a factor of 50) of the TQ coils after cool-down (left) and at 15 kA (right).

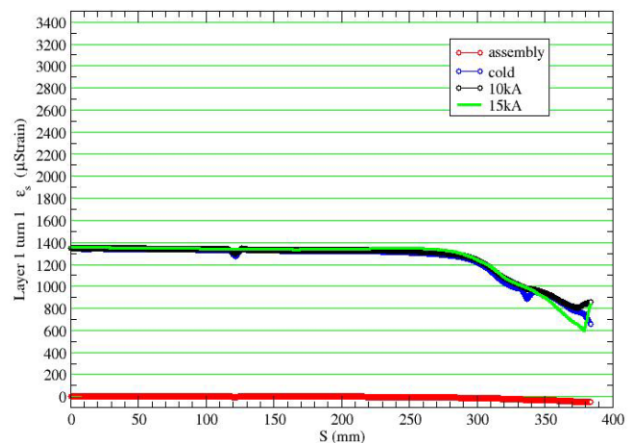
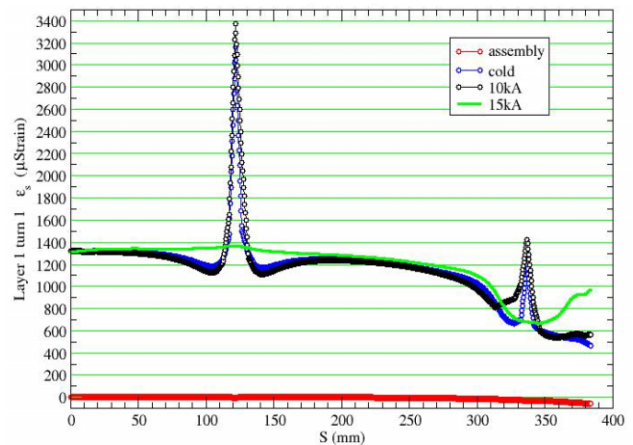


Figure 11: 3D mechanical computations of the pole turn axial strain as a function of the axial position with bronze pole (top) and titanium alloy pole (bottom).

Nb_3Sn coils are usually wound around solid pieces, called poles, which then participate to the full reaction and impregnation process. For fabrication purposes, the winding poles are segmented in several pieces with a length of approximately 200-300 mm. The first set of

$\cos(\theta)$ coils for the TQ quadrupole (TQ01 series) were wound around segmented aluminium bronze poles. During the TQS01 magnet tests, it was observed that quench-origins clustered around the area in between adjacent pole pieces. The numerical simulations indicated that, after cool-down, the segments tend to separate, thus creating gaps which increase during excitation (see deformed shape in Fig. 10). As a result, peaks of axial strain in the pole turn arise in correspondence of the gaps, indicating that the cables around the poles experience high tensions during magnet powering (see Fig 11, top).

This phenomenon was attributed to the friction between the iron components surrounding the coils, characterized by a low thermal contraction coefficient, and the bronze pole pieces, which feature a high thermal contraction coefficient. The friction induces high tension in the bronze poles, which then separate when the e.m. forces were applied. For the second set of coils (TQ02 series) it was therefore decided to change the pole material from bronze to titanium alloy (Ti6Al4V), characterized by a low thermal contraction coefficient: the titanium pole segments remain in compression during all magnet operations, thus eliminating gaps and high strain in the pole turns (see Fig. 11, bottom).

STRAIN AND DISPLACEMENTS IN COIL END REGIONS

Nb_3Sn magnets operating at fields approaching 15 T are subjected to very high axial e.m. forces which stretch the coils along the longitudinal direction. If not counterbalanced, these forces may generate mechanical motions and tensional strain resulting in potential degradation of magnet performance. For these reasons, the LBNL Superconducting Magnet Program has started implementing in the magnet design a coil axial support system based on end plates and aluminium rods (see Fig. 12) with the goal of preventing separation between turns and pole pieces in the end regions. In parallel, 3D models have been focused on mechanical solutions to mitigate the effect of the axial forces.

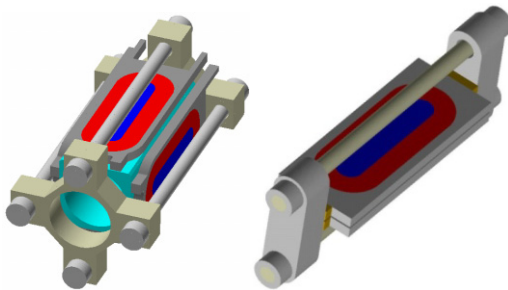


Figure 12: Coil axial supports implemented in the quadrupole SQ (left) and the dipole SD (right).

The importance of the axial support can be seen in Fig. 13, where the model results indicate how the deformation of the turns in the end regions of the TQ quadrupole coils can be significantly reduced by providing sufficient pre-load after cool-down.

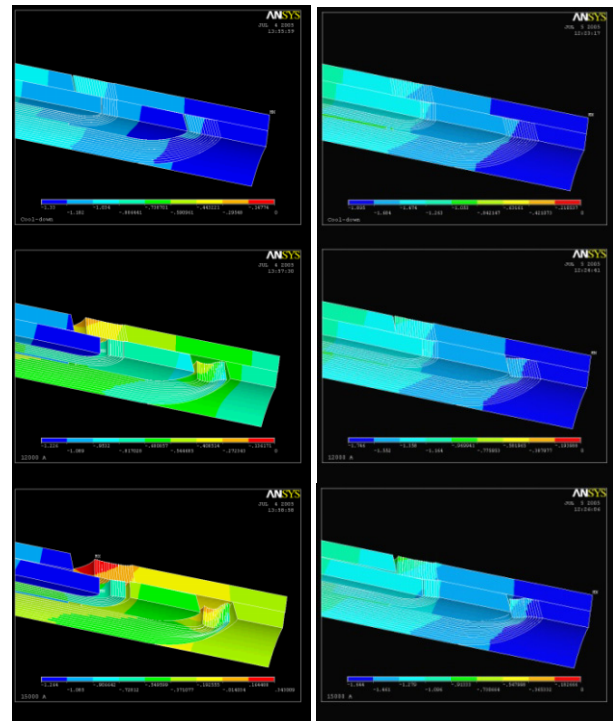


Figure 13: 3D mechanical computations of the deformed shapes (with displ. enhanced by a factor of 50) of the TQ coils without (left) and with (right) axial support after cool-down (top), at 12 kA (centre), and at 15 kA (bottom).

In order to investigate the impact of axial support and prove its effectiveness, in 2007 the LARP program launched a series of tests with the subscale quadrupole magnet SQ02 [21]. The magnet was tested with (SQ02b) and without (SQ02c) axial end load. The training performance is plotted in Fig. 14: the magnet reached its expected current limits when fully supported, but it showed a clear degradation in quench performance (in the order of 5 to 10%) when no axial load was applied: the degradation was progressive over consecutive quenches, i.e., the quench current gradually decreased from 10.2 kA to 9.7 kA in seven quenches.

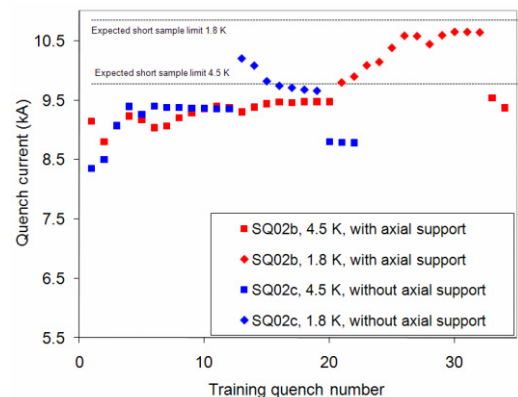


Figure 14: Training performance of SQ02b (at 4.5 K and 1.8 K, with axial support), and SQ02c (at 4.5 K and 1.8 K, without axial support). The dashed lines represent the expected current limits based on strand measurements.

A 3D analysis, reported in [22], pointed out a high tensile strain (about 4500 μstrain) in the cable at the end of the straight section as a probable cause of the performance degradation (Fig. 15).

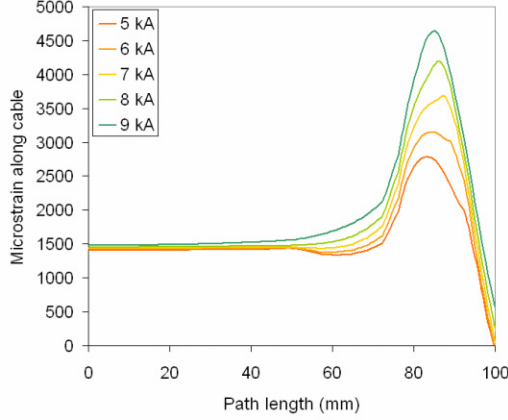


Figure 15: Computed strain during excitation along the cable on a path moving from the centre of the straight section to the end, in the SQ02c conditions.

MODELING OF QUENCH INITIATION AND PROPAGATION

In addition to the increase of strain analyzed in the previous section, the effect of axial e.m. force may include a relative sliding between the pole turns and the winding poles (see deformed shapes in Fig. 16, left). In the presence of friction, any sliding between two surfaces results in energy dissipation.

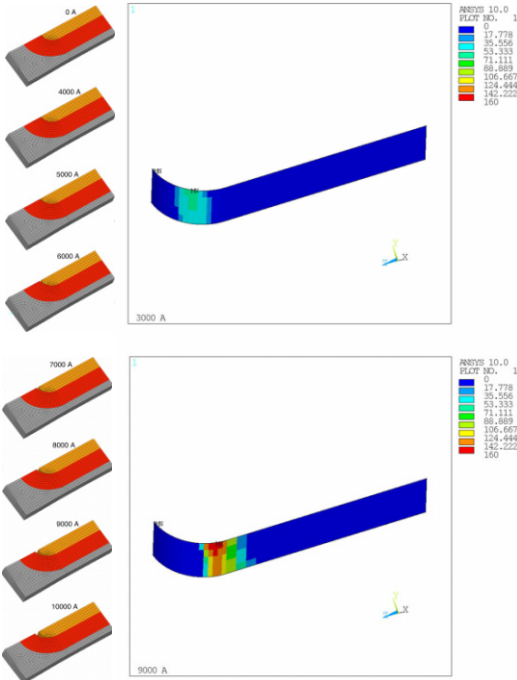


Figure 16: 3D mechanical computations of the deformed shapes (with displ. enhanced by a factor of 50) of the SQ02 coils (left) and frictional energy dissipation [J/m^2] (right) during excitation.

With a 3D finite element model, it is possible to compute the energy dissipation during excitation, and gain useful insight on the areas subjected to premature quenching. This study, reported in [23,24], was performed to analyze the performance of the SQ02 magnet. The results, plotted in Fig. 16, show that, from 0 to 3 kA the release of frictional energy near the end peaks at about 70 J/m^2 (top figure). During the following current steps, the dissipated energy progressively increases to a maximum of 160 J/m^2 (bottom figure) and its location gradually moves towards the straight section. The quench locations recorded during the SQ02a were consistent with the energy dissipation pattern predicted by the model.

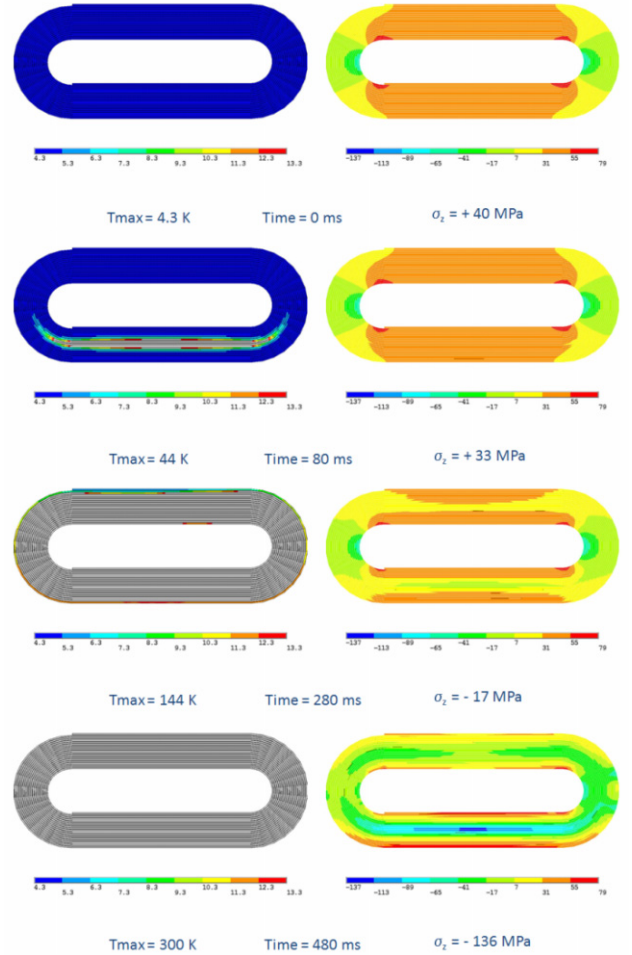


Figure 17: 3D computations of the thermal (K, left) and mechanical (MPa, right) status of the SM coil during quench propagation as a function of time.

If the energy dissipated by a conductor motion is enough to determine an increase of the local temperature of the superconductor beyond its critical level, a quench is originated and starts propagating. After the analysis of quench-triggering mechanisms, the computations can be focused towards a detailed representation of the thermal and mechanical status of a superconducting coil during the propagation of a quench. The goal is to determine the peak temperatures reached by the superconductor and to

extend the analysis to the mechanical response to a quench: the temperature profile evaluated by a thermo-electrical model can be transferred to a 3D mechanical model and the evolution of the coil stresses during the quench propagation investigated.

This analysis, reported in [25] was applied to the SM magnet and is summarized in Fig. 17. The plots on the left of the figure show the propagation of the normal zone (gray area) from quench initiation to 480 ms, when the hot spot temperature reaches a maximum of 300 K. The plots on the right of the figure represent the corresponding stress distribution in the coil. In the hot spot area the coil starts with a longitudinal tension of about 40 MPa before quench propagation. After the quench is initiated, the hotter regions attempt to expand and push against the colder surroundings. The hot spot experiences compression in every direction after the quench, with a maximum compression of about -140 MPa. The study can be used to determine that maximum hot spot temperature allowed in Nb₃Sn superconducting coils before the induced thermal strain degrade the superconductor.

CONCLUSIONS AND NEXT STEPS

The use of the strain-sensitive Nb₃Sn superconductor in the next generation of high-field magnets for particle accelerators will require an in-depth knowledge of the coil mechanical status, which can be obtained only with integrated magnetic, mechanical and thermal design and analysis. We described how a combined use of different tools and techniques allows a full representation of magnet geometry and operation conditions. We then pointed out the importance of focusing the analysis tools on optimizing coil and support structure to minimize coil stress, investigate the strain status of the superconductor from assembly to quench aftermath, and identify and correct quench initiation and training mechanism.

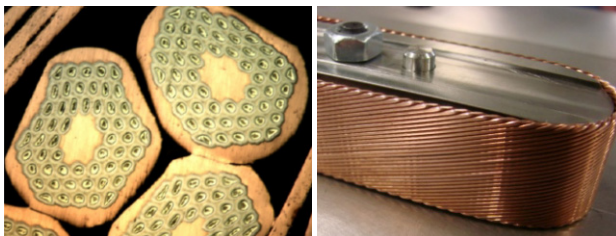


Figure 17: Deformation of strands in the cable edge (left) and cable bending around winding pole (right).

As a next step, we believe that a finer modelling of the mechanical status of the superconductor down to the filament level will constitute a fundamental step to improve our understanding of the behaviour of Nb₃Sn superconducting magnets. To reach this goal, it will be necessary to include in the simulations strand geometries resulting from the coil fabrication process, in particular cabling, winding, reaction and potting (Fig. 17) and implement 2D and 3D models of the Rutherford cable (Fig. 18) capable of reproducing the deformed shape of filaments and strands [26].

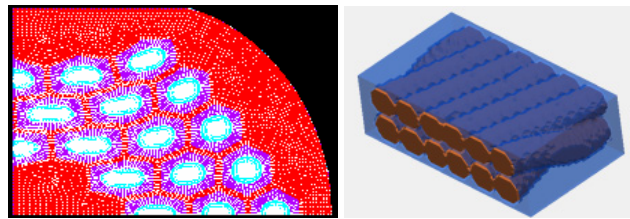


Figure 17: 2D model of a deformed strand (left), and 3D models of a Rutherford cable (right).

REFERENCES

- [1] P. Wanderer *et al.*, IEEE Trans. Appl. Supercond., 19 (3), 1122 (2009).
- [2] G. de Rijk, CERN 2009-001, 114 (2009).
- [3] Godeke *et al.*, Supercond. Sci. Techn. 19, R100 (2006).
- [4] Dietderich and Godeke, Cryogenics 48, 331 (2008).
- [5] S. Caspi and P. Ferracin, IEEE Trans. Appl. Supercond. 16 (2), 1298 (2006).
- [6] S. Caspi, P. Ferracin, and S. Gourlay, IEEE Trans. Appl. Supercond. 16 (2), 354 (2006).
- [7] L. Rossi and Ezio Todesco Phys. Rev. ST Accel. Beams 9, 102401 (2006).
- [8] L. Rossi and Ezio Todesco Phys. Rev. ST Accel. Beams 10, 112401 (2007).
- [9] K. Halbach, LLNL report UCRL-17436, 1967.
- [10] www.cern.ch/roxie.
- [11] Vector Fields Limited, 24 Bankside, Kidlington, Oxford OX5 1JE, England.
- [12] ANSYS, Inc., Southpointe, 275 Technology Drive, Canonsburg, PA 15317, USA.
- [13] J. Cook, IEEE Trans. Magn. 27 (4), 1976 (1991).
- [14] PTC, 140 Kendrick Street, Needham, MA 02494, USA.
- [15] H. Felice *et al.*, IEEE Trans. Appl. Supercond. 18 (2), 281 (2008).
- [16] H. Felice *et al.*, IEEE Trans. Appl. Supercond. 19 (3), 1235 (2009).
- [17] Paolo Ferracin *et al.*, IEEE Trans. Appl. Supercond. 19 (3), 1106 (2009).
- [18] S. Caspi *et al.*, IEEE Trans. Appl. Supercond. 17 (2), 1122 (2007).
- [19] S. Caspi *et al.*, IEEE Trans. Appl. Supercond. 18 (2), 179 (2008).
- [20] S. Caspi *et al.*, IEEE Trans. Appl. Supercond. 19 (3), 1221 (2009).
- [21] P. Ferracin *et al.*, IEEE Trans. Appl. Supercond. 17 (2), 1019 (2007).
- [22] P. Ferracin *et al.*, IEEE Trans. Appl. Supercond. 18 (2), 285 (2008).
- [23] P. Ferracin *et al.*, IEEE Trans. Appl. Supercond. 17 (2), 2373 (2007).
- [24] P. Ferracin and S. Caspi, Cryogenics 47 (11-12), 595 (2007).
- [25] P. Ferracin *et al.*, IEEE Trans. Appl. Supercond. 14 (2), 361 (2004).
- [26] D. Arbelaez *et al.*, "Cable Deformation Simulation and a Hierarchical Framework for Nb₃Sn Rutherford Cables", presented at EUCAS 2009.

TRANSIENT, LARGE-SCALE 3D FINITE ELEMENT SIMULATIONS OF THE SIS100 MAGNET*

S. Koch[§], T. Weiland[‡], Technische Universitaet Darmstadt, D-64289 Darmstadt, Germany
H. De Gersem[¶], Katholieke Universiteit Leuven - Campus Kortrijk, B-8500 Kortrijk, Belgium

Abstract

Numerical simulations are frequently used in the design, optimization and commissioning phase of accelerator components. Strict requirements on the accuracy as well as the complex structure of such devices lead to challenges regarding the numerical simulations in 3D. In order to capture all relevant details of the geometry and possibly strongly localized electromagnetic effects, large numerical models are often unavoidable. The use of parallelization strategies in combination with higher-order finite-element methods offers a possibility to account for the large numerical models while maintaining moderate simulation times as well as high accuracy. Using this approach, the magnetic properties of the SIS100 magnets designated to operate within the Facility of Antiproton and Ion Research (FAIR) at the GSI Helmholtzzentrum für Schwerionenforschung GmbH in Darmstadt, are calculated. Results for eddy-current losses under time-varying operating conditions are reported.

INTRODUCTION

For the operation of the heavy-ion synchrotron SIS100 as a part of the FAIR project at GSI, the magnetic flux density in the aperture of the dipole magnets is required to be ramped at rates as high as 4 T/s in order to keep up with the acceleration of the particles. As a consequence, eddy-current effects arising at the end regions of the dipole magnets with respect to the beam orbit, become an important issue in the magnet design and optimization process.

Despite the laminated structure of the iron yoke, eddy currents and, in turn, resistive losses appear in the conductive iron sheets. In the actual design, not only the superconductive current coils are operated at the appropriate temperature of 4.5 K, but also the ferromagnetic yoke and the mechanical support. As a consequence, the eddy-current losses induced by fast ramping appear in the cold mass of the system. These losses increase the load of the cryogenic system and therefore the power consumption of the facility significantly. Hence, one of the design goals is to reduce the Joule losses inside the magnet to an acceptable level. Several design optimizations aimed at this issue have already been proposed, e.g. in [1], [2], whereas the Nuclotron magnet [3], [4] served as a starting point for the

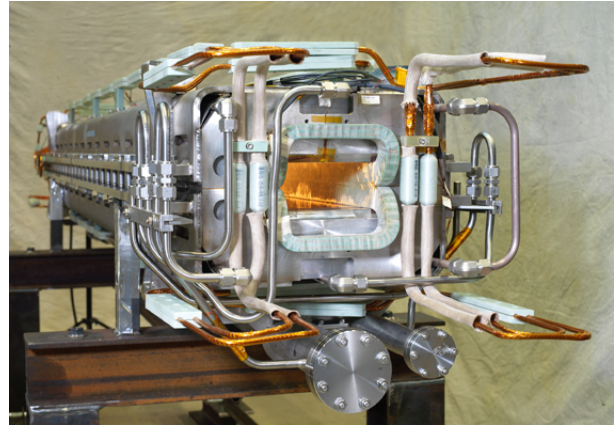


Figure 1: Full-length prototype dipole magnet for the SIS100 including cooling tubes and mechanical assembly (photograph: J. Guse, GSI (www.gsi.de)).

design. These optimizations are based on experiments as well as numerical simulations in 2D and 3D.

While the original Nuclotron magnet is 1.4 m long [4], the length of the current prototype of the SIS100 dipole shown in Fig. 1 is increased to 2.8 m [5]. Therefore, the number of dipole magnets required to cover the circumference of the synchrotron is, in turn, lowered by a factor close to two when compared to a virtual installation of the short magnet. As the major fraction of the eddy currents arises at the end regions of the iron yoke, the resulting overall losses are reduced accordingly. The increased length, however, provides additional challenges regarding numerical simulations when using volume-based discretization methods such as the finite element (FE) method. It leads to larger numerical models which in turn require a longer simulation time and more computational resources. One way to deal with the large numerical models is the use of parallelization techniques. Simulations related to the full length prototype shown in Fig. 1 are carried out using the finite element method in combination with higher-order shape functions. The eddy-current losses in the different parts of the yoke assembly are calculated.

NUMERICAL MODELING

Transient Magnetoquasistatic Formulation

Even though the desired ramping of the aperture field is fast when considering the amount of energy dissipated in the electrically conductive iron yoke, the time variation of

*Work supported by GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt under contract F&E, DA-WE11

[§] koch@temf.tu-darmstadt.de

[‡] thomas.weiland@temf.tu-darmstadt.de

[¶] Herbert.DeGersem@kuleuven-kortrijk.be

the resulting fields still justifies the application of the magnetoquasistatic approximation the Maxwell equations. Different formulations based on potentials or field quantities can be derived as reviewed, e.g., in [6], [7]. Among these, the formulation in terms of the magnetic vector potential \vec{A} is well suited for the problem under consideration. It is introduced in order to represent the magnetic flux density \vec{B} according to

$$\vec{B} = \nabla \times \vec{A}. \quad (1)$$

This leads to the second order partial differential equation

$$\nabla \times (\nu \nabla \times \vec{A}) + \sigma \frac{\partial}{\partial t} \vec{A} = \vec{J}_s, \quad (2)$$

where $\nu = 1/\mu$ denotes the reluctivity, μ the permeability, σ the electric conductivity and \vec{J}_s the source current density. Note that \vec{A} is interpreted as the modified magnetic vector potential and therefore features an implicit gauge in the conductive regions [8]. The quantities of interest within this application, the electric field strength \vec{E} and the magnetic flux density \vec{B} , are conveniently obtained from the solution for \vec{A} along with its temporal derivative by

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} \quad (3)$$

and Eq. 1. In contrast to the very similar formulation in terms of the \vec{E} itself, the source current density \vec{J}_s in Eq. 2 is not required to be differentiated in time. This is especially important as non-smooth excitation functions are provided for this type of magnet simulations.

Spatial Discretization

In order to solve the formulation in Eq. 2 numerically, tetrahedral meshing is applied. The computational domain is limited to the dipole magnet and the considered parts of the mechanical assembly. At a sufficient distance to the structure, closed boundary conditions are applied. By using vectorial FE shape functions \vec{w}_j ensuring tangential continuity of the approximated quantity, the magnetic vector potential is approximated in terms of the local Ritz approach

$$\vec{A} \approx \sum_j a_j \vec{w}_j. \quad (4)$$

In case of a lowest order approximation six vectorial shape functions, each of which associated with one edge of a tetrahedron, are used to represent the magnetic vector potential \vec{A} . This approximation is capable of modeling a linear variation of the vectorial quantity \vec{A} itself as well as a constant flux density \vec{B} resulting from the application of the curl operator according to Eq. 1 consistently. The subsequent higher order set of shape functions, being quadratically exact in \vec{A} and linearly exact in the curl of \vec{A} features 20 shape functions per element and, in turn, the same number of degrees of freedom a_j in Eq. 4. Choosing the same set of shape functions as test functions \vec{w}_i as part of the

Computer Codes (Design, Simulation, Field Calculation)

standard Galerkin procedure turns Eq. 2 into the weak formulation. Discretization of the latter by means of the shape functions \vec{w}_j results in the semi-discrete representation of Eq. 2:

$$\mathbf{K}_\nu \mathbf{a} + \mathbf{D}_\sigma \frac{d}{dt} \mathbf{a} = \mathbf{j}_s. \quad (5)$$

In this differential-algebraic equation, \mathbf{K}_ν denotes the stiffness matrix, \mathbf{D}_σ the damping matrix, \mathbf{a} collects the degrees of freedom a_j for the magnetic vector potential and the elements of \mathbf{j}_s describe the excitation current density weighted by the test functions \vec{w}_i . The entries of the matrices and vectors mentioned are given by

$$(\mathbf{K}_\nu)_{i,j} = \int_\Omega (\nabla \times \vec{w}_j) \cdot \nu \cdot (\nabla \times \vec{w}_i) dV, \quad (6)$$

$$(\mathbf{D}_\sigma)_{i,j} = \int_\Omega \vec{w}_j \cdot \sigma \cdot \vec{w}_i dV, \quad (7)$$

$$(\mathbf{j}_s)_i = \int_\Omega \vec{J}_s \cdot \vec{w}_i dV, \quad (8)$$

where Ω denotes the computational domain and the characteristic material coefficients ν and σ are assumed to be tensor-valued.

Temporal Discretization and Linearization

The differential-algebraic representation of the magnetoquasistatic formulation in terms of the magnetic vector potential in Eq. 5 is discretized in time using the implicit Euler scheme. For constant time steps Δt indexed by n , with $t_n = t_0 + n\Delta t$, the algebraic nonlinear system of equations reads

$$\mathbf{K}_\nu \mathbf{a}^{(n+1)} + \frac{1}{\Delta t} \mathbf{D}_\sigma \mathbf{a}^{(n+1)} = \underbrace{\frac{1}{\Delta t} \mathbf{D}_\sigma \mathbf{a}^{(n)} + \mathbf{j}_s^{(n+1)}}_{\mathbf{f}^{(n+1)}}, \quad (9)$$

whereas $\mathbf{a}^{(n+1)} = \mathbf{a}(t_{n+1})$, $\mathbf{a}^{(n)} = \mathbf{a}(t_n)$ and $\mathbf{j}_s^{(n+1)} = \mathbf{j}_s(t_{n+1})$. The fix-point formulation of Eq. 9 is given by $\mathbf{F}(\mathbf{a}^{(n+1)}) = 0$ with

$$\mathbf{F}(\mathbf{a}^{(n+1)}) = \mathbf{K}_\nu \mathbf{a}^{(n+1)} + \frac{1}{\Delta t} \mathbf{D}_\sigma \mathbf{a}^{(n+1)} - \mathbf{f}^{(n+1)}. \quad (10)$$

It is nonlinear with respect to the entries of \mathbf{K}_ν due to ferromagnetic saturation. In order to advance from the time instance t_n to t_{n+1} the nonlinear system needs to be solved. Based on Eq. 10 linearization is carried out by means of a Newton-Krylov method. The search direction $\mathbf{d}^{(n+1,k+1)}$ within each nonlinear step k is found by solving the system

$$\mathbf{J}_F^{(n+1,k)} \mathbf{d}^{(n+1,k+1)} = -\mathbf{F}(\mathbf{a}^{(n)}) \quad (11)$$

with the Jacobian \mathbf{J}_F of \mathbf{F} using a Krylov-subspace iteration. Here, the Jacobian is given in matrix form by

$$\mathbf{J}_F^{(n+1,k)} = \mathbf{K}_{\nu_a}^{(n+1,k)} + \frac{1}{\Delta t} \mathbf{D}_\sigma \quad (12)$$

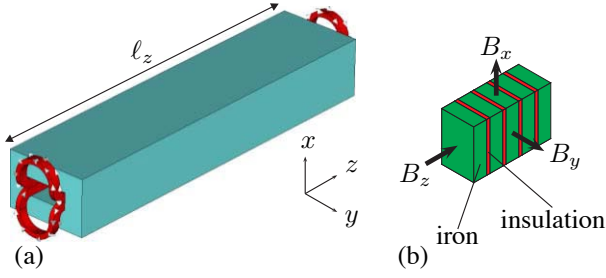


Figure 2: (a) Simplified model of the SIS100 magnet yoke and the coil; (b) Strongly enlarged view of the laminated structure of the yoke based on iron plates and insulation sheets (relation of thickness is not to scale).

while the entries of the matrix \mathbf{K}_{ν_d} follow from replacing ν in Eq. 6 by the differential reluctivity tensor ν_d , as described, e.g., in [9]. As a consequence, the sparsity pattern of the Jacobian \mathbf{J}_F is identical to the one of a standard system matrix in linear magnetoquasistatic modeling. From the search direction the new nonlinear iterate $\mathbf{a}^{(n+1,k+1)}$ is obtained by

$$\mathbf{a}^{(n+1,k+1)} = \mathbf{a}^{(n+1,k)} + \alpha \mathbf{d}^{(n+1,k+1)} \quad (13)$$

with the relaxation factor $\alpha = 1$ corresponding to full Newton step. A solution candidate $\mathbf{a}^{(n+1,k+1)}$ for $\mathbf{a}^{(n+1)}$ is accepted, if the criterion

$$r_{\text{nlm}} = \frac{\|\mathbf{F}(\mathbf{a}^{(n+1,k+1)})\|_2}{\|\mathbf{f}^{(n+1)}\|_2} < \epsilon_{\text{nlm}} \quad (14)$$

in terms of the relative nonlinear residual r_{nlm} is met for a prescribed tolerance ϵ_{nlm} . Within each nonlinear step $k+1$ the system in Eq. 11 is solved using Krylov-subspace iterations. These iterations are terminated once the appropriate residual tolerance ϵ_{k+1} for the current nonlinear step $k+1$ is reached. Using an adjustable stopping criterion ϵ_{k+1} as described in [10] can significantly reduce the total number of Krylov iterations per time step. An implementation according to [11] based on the NOX package of the TRILINOS framework [12] is used for the numerical simulations reported in this paper.

Homogenization of the Laminated Yoke Material

A schematic view of the SIS100 dipole reduced to the ferromagnetic yoke and the current coil is shown in Fig. 2(a). Due to the high ramp rate required for the application in the synchrotron, the yoke is built from laminated steel plates as indicated in Fig. 2(b). It is, however, cumbersome if not impossible to resolve the very thin insulation layers between adjacent iron sheet within a 3D numerical simulation. Therefore, the packing factor γ_p , commonly employed to classify laminated steels, is used to construct a homogeneous, anisotropic material to be considered in the simulation. For a straight yoke the co-ordinates can be

aligned with the main axes of the anisotropy. As a consequence, diagonal tensors

$$\nu = \text{diag}(\nu_{xx}, \nu_{yy}, \nu_{zz}); \sigma = \text{diag}(\sigma_{xx}, \sigma_{yy}, \sigma_{zz}) \quad (15)$$

for the reluctivity ν and the conductivity σ are sufficient to model the homogenized yoke material as described in [13].

While in general (case A) all entries of the reluctivity tensor in Eq. 15 are dependent on the value of the magnetic flux density as well as on the packing factor γ_p , a simplification can be introduced at moderate saturation levels up to 1.6 T [14]. Namely, the reluctivity ν_{zz} is observed to be constant with respect to the saturation level and it is only related to the packing factor by $\nu_{zz} = \nu_0(1 - \gamma_p)$. This type of modeling the anisotropic yoke material is referred to as case B in the following. Details and a comparison of both material models with respect to the eddy-current losses in the yoke can be found in [13].

BENCHMARK MODEL

The results reported in the following sections are obtained by using an in-house simulation tool based on the FEMSTER library [15] for the FE layer and the TRILINOS framework [12] for the linear algebra infrastructure as well as parallelized matrix and vector classes in terms of MPI [16]. Pre- and postprocessing as well as automatic mesh generation is carried out in CST STUDIO SUITE™ [17].

Simulation Setup

In order to validate the simulation tool in terms of a comparison of the results for the eddy-current losses to the ones obtained by other codes, a simplified benchmark model of the SIS100 dipole was defined [14], [18]. The according geometry is shown in Fig. 2(a). Further parameters for the magnetoquasistatic simulations of the simplified model can be summarized as follows:

- Length of the iron yoke $\ell_z = 1200$ mm;
- Outer dimensions of the rectangular cross-section of the yoke: 276 mm \times 186 mm;
- Aperture size 146 mm \times 56.4 mm;
- Conductivity tensor of the laminated steel: $\sigma_{xx} = \sigma_{yy} = 3.2 \cdot 10^6$ S/m, $\sigma_{zz} = 0$;
- Maximum excitation current: $I_{\text{max}} = 48$ kA;
- Coil excitation: triangular cycle 0...48 kA...0 of 1 s duration ($\partial B / \partial t = 4$ T/s);
- Homogeneous, anisotropic, nonlinear yoke material for different values of the packing factor γ_p (case A, case B);
- Constant time step $\Delta t = 0.01$ s.

Convergence of the Eddy-Current Losses

On the basis of the discretized model several simulations are carried out. In a first set, the eddy-current losses are calculated for different values of the packing factor γ_p in the range of 0.93 to 0.98. The losses are evaluated in each

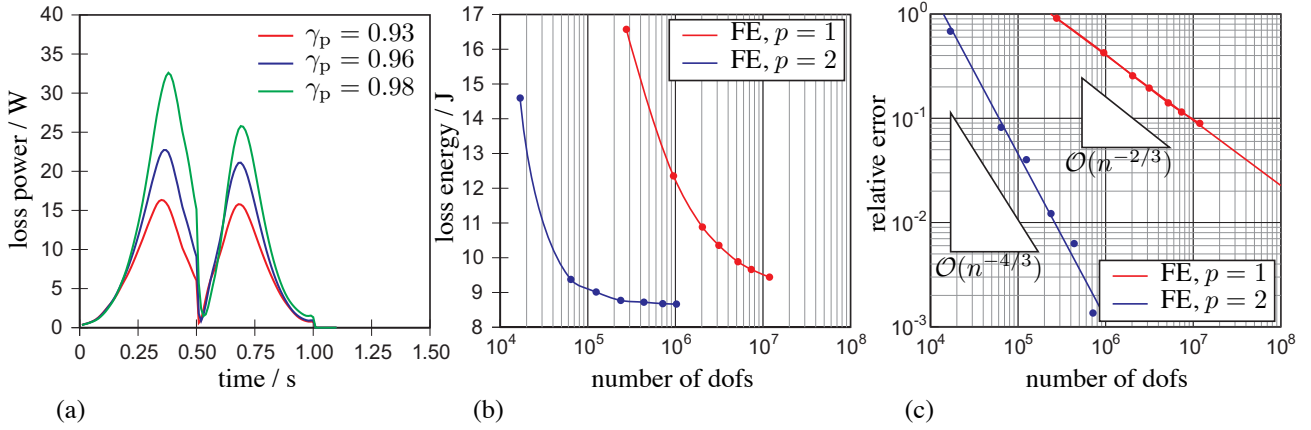


Figure 3: (a) Time dependence of the volume-integrated eddy-current losses in the magnet yoke in a triangular excitation cycle for different values of the packing factor γ_p . (b) Convergence of the loss energy integrated over the length of the excitation cycle for different levels of discretization: For lowest order shape functions ($p = 1$) and for shape functions allowing for a linear variation ($p = 2$) of the curl of the primary unknown. (c) Convergence of the relative discretization error of the power losses with respect to a reference solution obtained using 10^6 degrees of freedom (dofs) in combination with linear shape functions ($p = 2$).

time step and integrated spatially over the entire yoke volume. Fig. 3(a) shows the results on the time-axis. While the curves agree at the beginning of the cycle, the losses increase differently due to beginning ferromagnetic saturation starting approximately at $t = 0.2$ s. For higher packing factors the resulting eddy-current losses become larger. However, a high value is required in order to avoid adversely affecting the magnetic length of the dipole.

A second set of simulations addresses the discretization error of the numerical simulation which is quantified in terms of the loss energy over one excitation cycle, here. In order to compare the results to the ones reported in [18], the simplified homogenization strategy (case B) is implemented in the simulation tool. Using automatic mesh generation, several discrete representations of the magnet model are created. These are equipped with either lowest ($p = 1$) or second order ($p = 2$) FE shape functions. For each selected combination of a FE mesh and one of the sets of shape functions, a transient nonlinear simulation is carried out. The results of the integrated, time-averaged losses are shown in Fig. 3(b) with respect to the according number of degrees of freedom (dofs). Very large numerical models are required in order to obtain reliable results from the simulation, in particular for the case of lowest order shape functions. This becomes even more evident when determining the relative error which is illustrated in Fig. 3(c). Here, a simulation involving approximately 10^6 dofs based on second order FE shape functions provides the reference solution for the eddy-current losses. The higher approximation order of the linear shape functions leads to an improved convergence in terms of the discretization error. As a consequence, a fixed accuracy can be achieved using fewer degrees of freedom when compared to the curve related to the lowest order shape functions. Furthermore, the high accuracy required for the application can only be

reached using the higher order shape functions as, e.g., a discretization involving ten million lowest order degrees of freedom only achieves an accuracy of around 1%. The slope of the curves agrees with the theoretically expected values for the respective order of approximation.

Comparison to Different Simulation Code

Using the simplified homogenization model (case B) for the laminated iron yoke, the results for the eddy-current losses can directly be compared to results from the literature. The first two results columns of Table 1 show the

Table 1: Time-integrated eddy-current losses for different values of the packing factor γ_p using a nonlinear (case A) as well as a constant reluctivity (case B) in z -direction.

packing factor γ_p	case A Fig. 3(a)	case B cf. [14], [18]	
0.93	7.68 J	8.77 J	8.66 J
0.96	9.86 J	11.58 J	11.30 J
0.98	12.51 J	15.05 J	13.90 J

time-integrated losses in the magnet yoke for the two different material models obtained by using the in-house simulation tool. For comparison, the third column contains the values reported in [14], [18]. The difference in the results for the two material models increases for higher packing factors and reaches up to 20% for $\gamma_p = 0.98$. However, when comparing the last two columns of Table 1 for case B, a good agreement within 2.5% is observed for lower packing factors. The remaining differences can be assigned to the boundary conditions, which were not exactly specified in advance. Further contributions might be found in the temporal discretization as well as in the coil modeling.

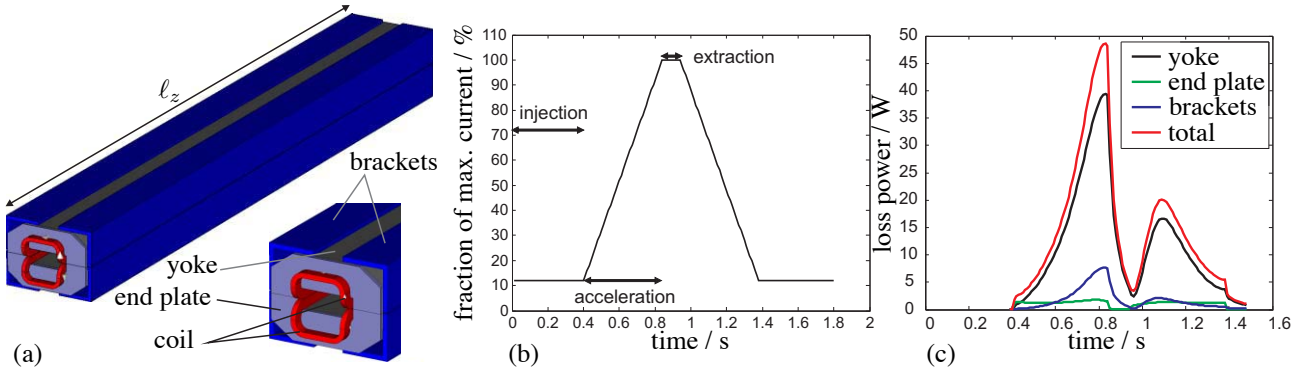


Figure 4: (a) CAD model of the full-length SIS100 dipole ($\ell_z = 2.8$ m) containing yoke, end plates and brackets ensuring mechanical stability. The new model uses a modified, lower coil design in order to reduce the eddy-current losses in the end regions. (b) Realistic excitation cycle resulting in a maximum aperture field of 2.1 T. (c) Eddy-current losses with respect to time for the separate parts of the yoke assembly.

SIMULATION RESULTS FOR THE FULL-LENGTH DIPOLE MAGNET

Using the in-house simulation software described above, a new set of simulations is carried out on the basis of more realistic parameters for the yoke geometry and for the excitation signal. A design featuring a long yoke offers advantages regarding the magnitude of the eddy-current losses as well as the homogeneity of the magnetic flux density in the aperture region. The geometry used for the simulations reported in the following is outlined in Fig. 4(a). End plates and brackets made from stainless steel, included for mechanical stability, are considered in the simulations as these parts are supposed to contribute to the overall losses in the magnet. In contrast to prior simulations, a more realistic operation cycle, as shown in Fig. 4(b), is considered. Further parameters, as far as differing from the previous setup, are as follows:

- Length of the iron yoke $\ell_z = 2800$ mm;
- Thickness of endplates and brackets; $d = 15$ mm;
- Isotropic conductivity of stainless steel used for end plates and brackets: $\sigma_{\text{steel}} = 2.0 \cdot 10^6$ S/m;
- Isotropic permeability of stainless steel: $\mu_{\text{steel},r} = 1.01$;
- Coil excitation: as indicated in Fig. 4(b);
- Homogeneous, anisotropic, nonlinear yoke material, packing factor $\gamma_p = 0.98$ (case A).

Due to the increased length of the magnet yoke, a larger number of tetrahedra is required for the numerical model when compared to the benchmark model. The superior convergence of higher-order shape functions ($p = 2$) is exploited in order to achieve the desired accuracy in an acceptable time. As the parallelization of the simulation tool is implemented on the basis of MPI, shared-memory as well as distributed-memory computing environments can be used. The results for the eddy-current losses shown in Fig. 4(c) are obtained on a workstation featuring two quad-core processors. A typical transient nonlinear simulation

Computer Codes (Design, Simulation, Field Calculation)

of this type takes approximately 36 hours, the major part of which is spent on the solution of the linear systems of equations. In Fig. 4(c) the time-characteristic of the eddy-current losses in the different parts of the yoke assembly are illustrated. The losses in the ferromagnetic yoke form the major contribution to the overall losses in this setup. Due to the high packing factor and the resulting anisotropic properties, relaxation effects arise and in turn lead to different magnitudes of the two maxima [19]. Also for the brackets, relaxation effects typical for the isotropic case equivalent to a packing factor $\gamma_{p,\text{steel}} = 1$ are observed in terms of an almost vanishing second maximum in the loss characteristic. As the end plates are thin and feature a smaller electrical conductivity as the laminated steel, the magnitude of the resulting losses is very small when compared to the remaining contributions. Furthermore, the respective relaxation time is short with respect to the considered simulation interval. Therefore, the loss characteristic is similar to the one expected for an equivalent linear RL-network related to the respective part of the model. Table 2 shows

Table 2: Eddy-current losses in the separate parts of the yoke assembly for two different levels of discretization both using FE shape functions of second order ($p = 2$).

Discretization	433 246 dofs	791 072 dofs
yoke	12.97 J	12.73 J
end plates	1.19 J	1.18 J
bracket	1.95 J	1.95 J
total	16.11 J	15.85 J

the integrated losses over the excitation cycle depicted in Fig. 4(b) for two different discretizations. The second column corresponds to the results of Fig. 4(c). In Fig. 5 the magnitude of the z -component of the magnetic flux density, which is responsible for the eddy currents, is shown in the end-region of the magnet. The sub-figures correspond to time-instances at the beginning (Fig. 5(a)), in the middle (Fig. 5(b)) and at the end of the acceleration phase.

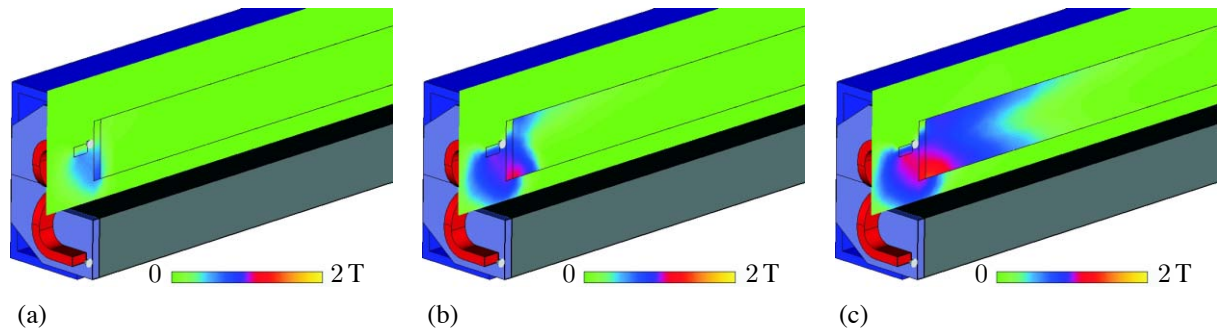


Figure 5: Magnitude of the z -component of the magnetic flux density responsible for the eddy-currents at the end regions of the yoke assembly in logarithmic color scale at different time instances: (a) $t = 0.42$ s, (b) $t = 0.68$ s, (c) $t = 0.82$ s.

CONCLUSION

Transient, nonlinear simulations required to predict the eddy-current losses in the SIS100 dipole are carried out by means of a 3D finite element simulation tool. The results are compared to a different, independent simulation code on the basis of a benchmark model. A good agreement in terms of the calculated eddy-current losses is observed. The developed simulation tool is well suited for the simulation of large and geometrically complicated structures. Thanks to the parallelized simulation framework, it is further scalable to even larger numerical models while maintaining moderate computational times as well as high accuracy by using higher-order finite element shape functions.

REFERENCES

- [1] A. Kalimov, E. Fischer, G. Hess, G. Moritz, and C. Mühle. Investigation of the power losses in a laminated dipole magnet with superconductive coils. *IEEE Trans. Appl. Supercond.*, 14(2):267–270, June 2004.
- [2] A. Kovalenko, N. Agapov, E. Fischer, H. Khodzhbagiyani, G. Kuznetsov, G. Moritz, and A. Smirnov. New results on minimizing ac power losses in a fast cycling 2 T superferic dipole with a cold yoke. *IEEE Trans. Appl. Supercond.*, 16(2):338–341, June 2006.
- [3] A.M. Baldin, N.N. Agapov, S.A. Averichev, A.M. Donyagin, E.I. D'yachkov, H.G. Khodzhbagiyani, A.D. Kovalenko, L.G. Makarov, E.A. Matyushevsky, and A.A. Smirnov. Superconducting fast cycling magnet of the Nuclotron. *IEEE Trans. Appl. Supercond.*, 5(2):875–877, June 1995.
- [4] A. Kovalenko. Status of the Nuclotron. In *Proceedings of the 4th European Particle Accelerator Conference (EPAC94)*, pages 161–164, London, UK, June 1994.
- [5] E. Fischer, H.G. Khodzhbagiyani, and A.D. Kovalenko. Full size model magnets for the fair sis100 synchrotron. *IEEE Trans. Appl. Supercond.*, 18(2):260–263, June 2008.
- [6] O. Bíró and K. Preis. On the use of the magnetic vector potential in the finite element analysis of three-dimensional eddy currents. *IEEE Trans. Magn.*, 25(4):3145–3159, July 1989.
- [7] O. Bíró and K. Preis. Finite Element Analysis of 3-D Eddy Currents. *IEEE Trans. Magn.*, 26(2):418–423, March 1990.
- [8] C.R.I. Emson and J. Simkin. An Optimal Method for 3-D Eddy Currents. *IEEE Trans. Magn.*, MAG-19(6):2450–2452, November 1983.
- [9] S. Koch, H. De Gersem, and T. Weiland. Transient 3D finite element simulations of the field quality in the aperture of the SIS-100 dipole magnet. *IEEE Trans. Appl. Supercond.*, 19(3):1162–1166, June 2009.
- [10] R.S. Dembo, S.C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19(2):400–408, April 1982.
- [11] M. Pernice and H.F. Walker. NITSOL: A Newton iterative solver for nonlinear systems. *SIAM J. Sci. Comput.*, 19(1):302–318, January 1998.
- [12] M. Heroux, R. Bartlett, V. Howle, R. Hoekstra, J. Hu, T. Kolda, R. Lehoucq, K. Long, R. Pawlowski, E. Phipps, A. Salinger, H. Thornquist, R. Tuminaro, J. Willenbring, and A. Williams. An overview of Trilinos. *Sandia National Laboratories, SAND2003-2927*, 2003.
- [13] S. Koch, H. De Gersem, T. Weiland, E. Fischer, and G. Moritz. Transient 3D finite element simulations of the SIS100 magnet considering anisotropic, nonlinear material models for the ferromagnetic yoke. *IEEE Trans. Appl. Supercond.*, 18(2):1601–1604, June 2008.
- [14] E. Fischer, R. Kurnyshov, G. Moritz, and P. Shcherbakov. 3-D transient process calculations for fast cycling superferic accelerator magnets. *IEEE Trans. Appl. Supercond.*, 16(2):407–410, June 2006.
- [15] D. White, P. Castillo, B. Fassenfest, R. Rieben, and M. Stowell. FEMSTER: A C++ class library of higher-order discrete differential forms. *Lawrence Livermore National Laboratory*, 2006.
- [16] MPI. Argonne National Laboratory, Argonne, IL, USA. www.mcs.anl.gov.
- [17] CST STUDIO SUITE™. Computer Simulation Technology AG, Darmstadt, Germany. www.cst.com.
- [18] P. Shcherbakov, E. Fischer, and R. Kurnyshov. 3D magnetic field and eddy loss calculations for iron dominated accelerator magnets using ansys compared with results of non-commercial codes. In *Proceedings of the EPAC 2006*, pages 2595–2597, Edinburgh, Scotland, UK, 2006.
- [19] E. Fischer, R. Kurnyshov, and P. Shcherbakov. Analysis of the eddy current relaxation time effects in the FAIR SIS 100 main magnets. *IEEE Trans. Appl. Supercond.*, 17(2):1173–1176, June 2007.

A PARALLEL HYBRID LINEAR SOLVER FOR ACCELERATOR CAVITY DESIGN

I. Yamazaki*, X. S. Li†, and E. G. Ng‡

Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

Abstract

Numerical simulations to design high-energy particle accelerators give rise to large-scale ill-conditioned highly-indefinite linear systems of equations that are becoming increasingly difficult to solve using either a direct solver or a preconditioned iterative solver alone. In this paper, we describe our current effort to develop a parallel hybrid linear solver that balances the robustness of a direct solver with the efficiency of a preconditioned iterative solver. We demonstrate that our hybrid solver is more robust and efficient than the existing state-of-the-art software to solve these linear systems on a large number of processors.

INTRODUCTION

Numerical simulations to design high-energy particle accelerators [10] give rise to large sparse linear systems of equations that are becoming increasingly difficult to solve using standard techniques [9]. Although significant progress has been made in the development of high-performance direct solvers [2, 4, 11], the dimension of the systems that can be directly factored is limited due to large memory requirements. Preconditioned iterative solvers [3, 14, 16] can reduce the memory requirements, but they often suffer from slow convergence.

To overcome these challenges, a number of parallel hybrid solvers have been developed based on a domain decomposition idea called the Schur complement method [5, 6]. In this method, the unknowns in interior domains are first eliminated using a direct method, and the remaining Schur complement system is solved using a preconditioned iterative method. These hybrid solvers often exhibit great parallel performance because the interior domains can be factored in parallel, and the direct solver is effective to factor the relatively-small interior domain. In addition, the preconditioned iterative solver is shown to be robust to solve the Schur complement systems, where most of the fill occurs, in a number of applications [5, 6]. In particular, for a symmetric positive definite system, the Schur complement has a smaller condition number than the original matrix [15, Section 4.2], and fewer iterations are often needed to solve the Schur complement system. Hence, these hybrid solvers have the potential to balance the robustness of the direct solver with the efficiency of the iterative solver.

Unfortunately, for a highly-indefinite linear system from the accelerator simulation, these existing hybrid solvers often suffer from slow convergence when solving the Schur

complement system. This is true especially on a large number of processors because these solvers are designed to achieve good scalability of time to compute the preconditioners, but the quality of the preconditioner often degrades as more processors are used.

To overcome these drawbacks, we have been developing a new implementation of the Schur complement method which provides the robustness and flexibility to solve large highly-indefinite linear systems on a large number of processors [12, 17]. In this paper, we demonstrate the effectiveness of our hybrid solver to solve these linear systems on hundreds of processors using a linear system whose dimension is greater than those used in our previous papers. We also point out how our implementation has been modified since the last publication in order to solve such large linear systems with millions of unknowns.

SCHUR COMPLEMENT METHOD

The Schur complement method is a non-overlapping domain decomposition method, which is also referred to as iterative substructuring. In this method, the original linear system is first reordered into a 2×2 block system of the following form:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad (1)$$

where A_{11} and A_{22} respectively represent *interior domains* and *separators*, and A_{12} and A_{21} are the *interfaces* between A_{11} and A_{22} . By eliminating the unknowns associated with the interior domains A_{11} in the bottom part of (1), we obtain the block-triangular system

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & S \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ \hat{b}_2 \end{pmatrix}, \quad (2)$$

where S is the Schur complement defined as

$$S = A_{22} - A_{21}A_{11}^{-1}A_{12}, \quad (3)$$

and $\hat{b}_2 = b_2 - A_{21}A_{11}^{-1}b_1$. Hence, the solution of the linear system (1) can be computed by first solving the Schur complement system

$$Sx_2 = \hat{b}_2, \quad (4)$$

then solving the interior system

$$A_{11}x_1 = b_1 - A_{12}x_2. \quad (5)$$

Note that interior domains are independent of each other, and A_{11} is a block-diagonal matrix. Hence, the relatively

* ic.yamazaki@gmail.com

† xsli@lbl.gov

‡ egng@lbl.gov

small interior domains can be efficiently factored in parallel using a direct solver. On the other hand, a large amount of fill can be introduced in S . In order to reduce the memory requirement, the Schur complement system (4) is solved using a preconditioned iterative method. We note that within the iterative method, the matrix-vector product with the Schur complement S can be computed by applying the sequence of the sparse matrix operations (3) on the vector, and hence, S does not have to be stored explicitly for this phase of the solver.

PARALLEL IMPLEMENTATION

To obtain high-performance, our implementation takes full advantage of the state-of-the-art software. Specifically, the 2×2 block system (1) is computed using PT-SCOTCH [8], which implements a parallel nested bisection algorithm.¹ Then, to compute the LU factors of the interior domains in parallel, our hybrid solver uses either a “1-level” or “2-level” configuration, where each interior domain is factored using either a single processor or multiple processors, respectively. Currently, we use a serial direct solver SuperLU [4] or a parallel direct solver SuperLU_DIST [11] for the 1-level or 2-level configuration, respectively. To preserve the sparsity of the LU factors, each interior domain is permuted using METIS [7] or PT-SCOTCH. Then, the Schur complement system (4) is solved using a Krylov subspace method from PETSc [13] combined with an ILU preconditioner. Finally, the solution of the interior system (5) is computed using the previously computed LU factors of the interior domains.

Since computing the ILU preconditioner of the Schur complement is often the computational and memory bottleneck, we give a brief description of how the preconditioner is computed. In our current implementation, the preconditioners are the exact LU factors of a sparsified Schur complement. Specifically, let us denote the ℓ -th interior domain and corresponding interfaces by $A_{11}^{(\ell)}$, and $A_{12}^{(\ell)}$ and $A_{21}^{(\ell)}$, respectively, such that the coefficient matrix of the 2×2 block system (1) with k interior domains can be written as

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right) = \left(\begin{array}{cccc|c} A_{11}^{(1)} & & & & A_{12}^{(1)} \\ & A_{11}^{(2)} & & & A_{12}^{(2)} \\ & & \ddots & & \vdots \\ & & & A_{11}^{(k)} & A_{12}^{(k)} \\ \hline A_{21}^{(1)} & A_{21}^{(2)} & \dots & A_{21}^{(k)} & A_{22} \end{array} \right). \quad (6)$$

In the 1-level configuration, the ℓ -th processor stores the nonzeros of $A_{11}^{(\ell)}$ and $A_{21}^{(\ell)}$ in row-wise order, and the nonzeros of $A_{12}^{(\ell)}$ in column-wise order. If the 2-level configuration is used, then the rows of $A_{11}^{(\ell)}$ and $A_{12}^{(\ell)}$, and the columns of $A_{21}^{(\ell)}$ are evenly distributed among the processors assigned to the ℓ -th interior domain. Furthermore, the

rows of A_{22} are evenly distributed among the processors that solve the Schur complement system (4).

With the block structure (6) and the LU factorization of the interior domain $A_{11}^{(\ell)}$, which is denoted by $A_{11}^{(\ell)} = L_{11}^{(\ell)} U_{11}^{(\ell)}$,² the Schur complement is computed as

$$\begin{aligned} S &= A_{22} - \sum_{\ell=1}^k A_{21}^{(\ell)} (A_{11}^{(\ell)})^{-1} A_{12}^{(\ell)} \\ &= A_{22} - \sum_{\ell=1}^k ((U_{11}^{(\ell)})^{-T} (A_{21}^{(\ell)})^T)^T ((L_{11}^{(\ell)})^{-1} A_{12}^{(\ell)}) \\ &= A_{22} - \sum_{p=1}^{p_A} E^{(\ell)}(:, j_1^{(p)} : j_2^{(p)}) F^{(\ell)}(j_1^{(p)} : j_2^{(p)}, :), \end{aligned}$$

where p_A is the number of processors used to solve the whole system, $E^{(\ell)} = ((U_{11}^{(\ell)})^{-T} (A_{21}^{(\ell)})^T)^T$, $F^{(\ell)} = (L_{11}^{(\ell)})^{-1} A_{12}^{(\ell)}$, the p -th processor owns the $j_1^{(p)}$ -th through $j_2^{(p)}$ -th columns of $A_{21}^{(\ell)}$ and the corresponding rows of $A_{12}^{(\ell)}$, and $E(:, j_1 : j_2)$ and $F(j_1 : j_2, :)$ are the matrices consisting of the j_1 -th through j_2 -th columns of a matrix E and the corresponding rows of F , respectively. In other words, after $E^{(\ell)}$ and $F^{(\ell)}$ are computed by the processors assigned to the ℓ -th interior domain, the $j_1^{(p)}$ -th through $j_2^{(p)}$ -th columns of $E^{(\ell)}$ and the corresponding rows of $F^{(\ell)}$ are sent to the p -th processors. Then, the p -th processor computes the corresponding outer-product updates of the Schur complement S , and sends the rows of the updates to the processor that owns the corresponding rows of A_{22} . Subsequently, the rows of S are evenly distributed among the processors that solve the Schur complement system. We note that the matrices $E^{(\ell)}$ and $F^{(\ell)}$ are computed with regards to the sparsity of interface $A_{21}^{(\ell)}$ and $A_{12}^{(\ell)}$.

In the above expression, the columns of $E^{(\ell)}$ and rows of $F^{(\ell)}$ are distributed in the same way as those of $A_{21}^{(\ell)}$ and $A_{12}^{(\ell)}$, respectively. This is not necessary. In our new implementation, the matrices $E^{(\ell)}$ and $F^{(\ell)}$ are distributed to balance the computational cost to compute the outer-product updates $E^{(\ell)}(:, j_1^{(p)} : j_2^{(p)}) F^{(\ell)}(j_1^{(p)} : j_2^{(p)}, :)$ among the processors assigned to the ℓ -th interior domain. Furthermore, to reduce the costs, the sparsity of the matrices $E^{(\ell)}$ and $F^{(\ell)}$ are enforced by discarding nonzeros with magnitudes less than a prescribed drop tolerance σ_1 . Hence, an approximation \tilde{S} to the Schur complement S is computed.

After the approximate Schur complement \tilde{S} is computed, it is preprocessed to enhance numerical stability, and then its sparsity is enforced using a drop tolerance σ_2 . Finally, SuperLU_DIST is used to compute the LU factor of \tilde{S} , which is used as the preconditioner. To preserve the sparsity of the LU factor, \tilde{S} is permuted based on a parallel nested bisection of the supernodal graph of \tilde{S} . See [17], for the preprocessing techniques used on \tilde{S} .

¹In the previous implementation [12], HID of HIPS [5] was used to compute the block system using a single processor.

²The matrix $A_{11}^{(\ell)}$ is scaled and permuted to enhance numerical stability and preserve the sparsity of $L_{11}^{(\ell)}$ and $U_{11}^{(\ell)}$. For clarity, the scaling and permutation are not shown in the expression.

NUMERICAL EXPERIMENTS

We present preliminary results of our hybrid solver to solve a highly-indefinite linear system on hundreds of processors. Our test matrix **tdr455k** is from a simulation of an international linear collider [1, 10], and its dimension is 2,738,556. To solve the Schur complement system, we used the unrestarted GMRES from PETSc [13]. The GMRES iteration was started with the zero vector, and the computed solution was considered to have converged when the ℓ_2 -norm of the initial residual was reduced by at least twelve order of magnitude. This is the solution accuracy enforced in the actual simulations. All the experiments were conducted on the Cray XT4 machine at the National Energy Research Scientific Computing Center.

To demonstrate the effectiveness of our hybrid solver, we compare its performance with that of a direct solver SuperLU_DIST [11], and that of a state-of-the-art hybrid solver HIPS [5]. The primary difference between our hybrid solver and HIPS is the way the preconditioner is computed for solving the Schur complement system. HIPS computes the preconditioner based on an ILU factorization of \tilde{S} , where the sparsity of the preconditioner is enforced based on both numerical values and locations of nonzeros. Specifically, the fill is allowed only between separators adjacent to the same domain. Furthermore, HIPS factors each interior domain using a single processor, and the number of interior domains needs to be at least as large as the number of processors used to solve the whole linear system. This allows HIPS to achieve good parallel scalability of time to compute the preconditioner. However, for a highly-indefinite system, we found that HIPS often suffers from slow convergence. This was especially true on a large number of processors since a large number of interior domains must be generated, which increases the size of the Schur complement.

To compare the performance of our hybrid solver with that of SuperLU_DIST, Figure 1 shows the total solution times as functions of the number of processors used to solve the linear system. In the figure, “1-level” and “2-level” are the two configurations of our hybrid solver, which use a single processor and multiple processors to factor each interior domain, respectively. In our numerical experiments with the 1-level configuration, we set the number of interior domains to be equal to the number of processors used to solve the linear system. On the other hand, for the 2-level configuration, the number of interior domains was fixed to be 16, and the processors were evenly distributed among the interior domains.

By looking at the left plot of Figure 1, we see that the solution time with our hybrid solver scaled better than that with SuperLU_DIST, (i.e., the hybrid solver could reduce the solution time using up to 512 processors, while SuperLU_DIST did not scale beyond 128 processors). Furthermore, we see that the solution times with the 1-level configuration scaled similarly to that with the 2-level configuration. This is because with the small drop toler-

ances $(\sigma_1, \sigma_2) = (10^{-6}, 10^{-5})$, the number of GMRES iterations was nearly independent of the number of interior domains, and GMRES converged within 20 iterations even when the number of interior domains needed to be increased for the 1-level configuration to run on more processors (see Table 1). In comparison, HIPS required 151 iterations on 16 processors, and it failed to converge within 1,000 iterations on 32 processors even though the drop tolerances were set to be zero. Furthermore, even when HIPS converged, our hybrid solver solved the linear system faster since it takes full advantage of the state-of-the-art software.

(σ_1, σ_2)	Number of domains				
	16	32	64	128	256
$(10^{-6}, 10^{-5})$	11	15	15	17	16
$(10^{-5}, 10^{-4})$	32	60	116	205	290

Table 1: Number of iterations with our hybrid solver.

We note that larger drop tolerances reduce the memory requirement of our hybrid solver. For example, in Figure 1, less memory was needed in the right plot since the drop tolerances were increased by an order of magnitude from those in the left plot. Specifically, in the left plot, about 15% of the nonzeros were discarded from the matrices E and F , and about 50% of the nonzeros were discarded from the approximate Schur complement \tilde{S} . On the other hand, in the right plot, about 30% of the nonzeros were discarded from E and F , and about 75% of the nonzeros were discarded from \tilde{S} . Unfortunately, with large drop tolerances, the number of GMRES iterations may increase as more interior domains are generated. For example with the drop tolerance $(\sigma_1, \sigma_2) = (10^{-5}, 10^{-4})$, the number of iterations increased from 32 to 290 when the number of interior domains increased from 16 to 256 (see Table 1). As a result, with the 1-level configuration, the solution time did not scale beyond 64 processors (see Figure 1). On the other hand, with the 2-level configuration, we can increase the number of processors while keeping the size of the Schur complement the same. Subsequently, with the 2-level configuration, the solution time was reduced using up to 256 processors in Figure 1.

Unfortunately, with the drop tolerance $(\sigma_1, \sigma_2) = (10^{-5}, 10^{-4})$, the 2-level configuration could not reduce the solution time using 512 processors from that using 256 processors (see Figure 1). This is because with the 2-level configuration, the size of the Schur complement is fixed, and we used only 16 processors to solve the Schur complement system even when more processors were available. As a result, the time to solve the Schur complement system became the dominant part of the total solution time as the number of processors increased. However, the total solution time was still reduced from that of SuperLU_DIST by a factor of 2.3 on 512 processors.

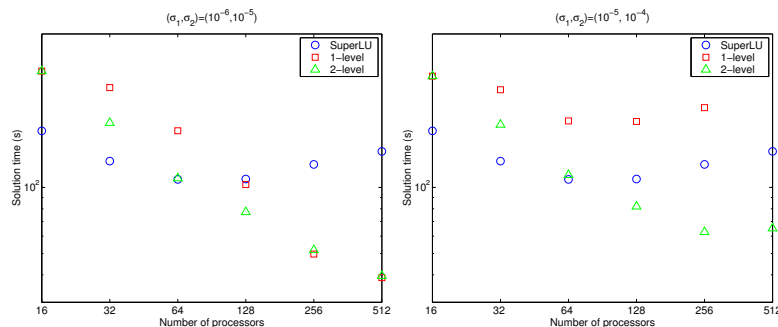


Figure 1: Solution times required by SuperLU_DIST and our hybrid solver.

CONCLUSIONS

For a large-scale particle accelerator simulation, solving large highly-indefinite linear systems becomes the memory bottleneck. We have described our current effort to address this challenge with a new parallel hybrid solver. The preliminary results have demonstrated that our hybrid solver improves the parallel scalability of a state-of-the-art direct solver. Furthermore, in comparison to a state-of-the-art hybrid solver, our hybrid solver is more robust and efficient to solve these linear systems on a large number of processors. As a result, our hybrid solver has the potential to enable a large-scale particle accelerator simulation by using a large number of processors and reducing the memory required by a processor. We are working to improve the parallel performance of our solver and conducting further experiments to solve larger systems using thousands of processors.

ACKNOWLEDGMENTS

The authors thank Lie-Quan Lee at the SLAC National Accelerator Laboratory for providing the test matrix. This research was supported in part by the Director, Office of Science, Office of Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We used the resources at the National Energy Research Scientific Computing Center.

REFERENCES

- [1] Community Petascale Project for Accelerator Science and Simulation (ComPASS). <https://compass.fnal.gov>.
- [2] P. Amestoy, I. Duff, J. Koster, and J.-Y. L'Excellent. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41, 2001.
- [3] O. Axelsson. *Iterative solution methods*. Cambridge University Press, New York, 1994.
- [4] J. Demmel, S. Eisenstat, J. Gilbert, X. Li, and J. Liu. A supernodal approach to sparse partial pivoting. *SIAM J. Matrix Analysis and Applications*, 20:720–755, 1999.
- [5] J. Gaidamour and P. Henon. HIPS: a parallel hybrid direct/iterative solver based on a schur complement. In *Proc. PMAA*, 2008.
- [6] L. Giraud, A. Haidar, and L. T. Watson. Parallel scalability study of hybrid preconditioners in three dimensions. *Parallel Computing*, 34:363–379, 2008.
- [7] Karypis Lab, Digital Technology Center, Department of Computer Science and Engineering, University of Minnesota. METIS - Serial Graph Partitioning and Fill-reducing Matrix Ordering. <http://glaros.dtc.umn.edu/gkhome/metis/metis>.
- [8] Laboratoire Bordelais de Recherche en Informatique (LaBRI). SCOTCH - Software package and libraries for graph, mesh and hypergraph partitioning, static mapping, and parallel and sequential sparse matrix block ordering. <http://www.labri.fr/perso/pelegrin/scotch/>.
- [9] L.-Q. Lee, L. Ge, M. Kowalski, Z. Li, C.-K. Ng, G. Schussman, M. Wolf, and K. Ko. Solving large sparse linear systems in end-to-end accelerator structure simulations. *Parallel and Distributed Processing Symposium, International*, 1:8a, 2004.
- [10] L.-Q. Lee, Z. Li, C.-K. Ng, and K. Ko. Omega3P: A parallel finite-element eigenmode analysis code for accelerator cavities. Technical Report SLAC-PUB-13529, Stanford Linear Accelerator Center, 2009.
- [11] X. Li and J. Demmel. SuperLU_DIST: A scalable distributed-memory sparse direct solver for unsymmetric linear systems. *ACM Trans. Mathematical Software*, 29(2):110–140, 2003.
- [12] X. Li, M. Shao, I. Yamazaki, and E. Ng. Factorization-based sparse solvers and preconditioners. In *Journal of Physics: Conference Series, SciDAC 2009*.
- [13] Mathematics and Computer Science Division, Argonne National Laboratory. The portable, extensible, toolkit for scientific computation (PETSc). www.mcs.anl.gov/petsc.
- [14] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, Philadelphia, 2004.
- [15] B. Smith, P. Bjorstad, and W. Gropp. *Domain Decomposition. Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, New York, 1996.
- [16] H. van der Vorst. *Iterative Krylov methods for large linear systems*. Cambridge University Press, New York, 2003.
- [17] I. Yamazaki, X. Li, and E. Ng. Preconditioning schur complement systems of highly-indefinite linear systems for a parallel hybrid solver. In *Submitted to the proceedings of the international conference on preconditioning techniques for scientific and industrial applications*, 2009.

HARD- AND SOFTWARE-BASED ACCELERATION TECHNIQUES FOR FIELD COMPUTATION

Martin Schauer[#], Peter Thoma^{*}

[#] CST of America, San Mateo, CA, United States

^{*} CST AG, Darmstadt, Germany

Abstract

Due to a high demand in more realistic graphics rendering for computer games and professional applications, commercial, off-the-shelf graphics processing units (GPU) increased their functionality over time. Recently special application programming interfaces (API) allow programming these devices for general purpose computing. This paper will discuss the advantages of this hardware platform for time domain simulations using the Finite-Integration-Technique (FIT). Examples will demonstrate typical accelerations over conventional central processing units (CPU).

Next to this hardware based accelerations for simulations also software based accelerations are discussed. A distributed computing scheme can be used to accelerate multiple independent simulation runs. For memory intense simulations the established Message Passing Interface (MPI) protocol enables distribution of one simulation to a compute cluster with distributed memory access. Finally, the FIT framework also allows special algorithmic improvements for the treatment of curved shapes using the perfect boundary approximation (PBA), which speeds up simulations.

INTRODUCTION

Simulation performance is a frequently discussed topic since users of simulation software want to achieve faster time-to-market in order to gain a competitive advantage. As a prerequisite for the following studies, “performance” needs to be defined first and includes the full design process from the idea to the realization.

1. Pre-processing
 - CAD Modeling / Workflow integration
 - Parameter definition
2. Solver
 - Advanced numerical algorithms
 - High performance computing
3. Post-processing
 - Derive secondary quantities
 - Optimize parameter

It is important to note that on the solver side not only the speed, but also the accuracy of this algorithm needs to be taken into account.

$$\text{Performance} = \text{speed} * \text{accuracy}$$

Most of the performance studies in this paper are discussed within the FIT framework [1]. We would also like to emphasize that the right solver choice can speed up the simulation significantly and should therefore be preferred, before hard- and software based acceleration techniques are chosen.

The paper is subdivided into two parts: hardware and software based acceleration techniques. Figure 1 illustrated how these techniques play together in [2].

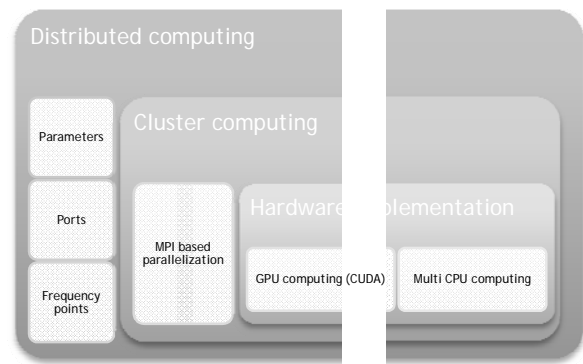


Figure 1: Overview of acceleration techniques and their hierarchy

HARDWARE ACCELERATION TECHNIQUES

Multi-CPU/multi-core is a classical hardware acceleration technique. In the recent years GPU computing emerged as a competing technique. In terms of performance we need to distinguish between algorithms with a high count of operations per memory access and algorithms with a low count.

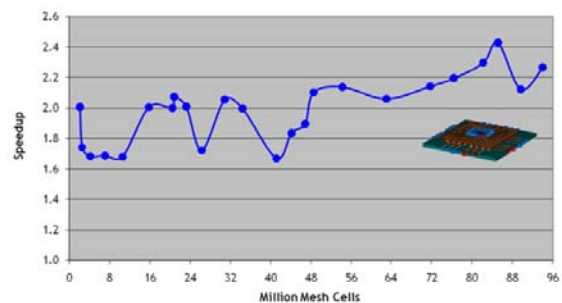


Figure 2: Simulation speed-up of the latest Intel generation CPU [3] vs. the previous generation

Figure 2 demonstrates the performance of a time domain FIT simulation using the latest hardware. This algorithm falls into the “memory-bandwidth” limited category and therefore benefits directly from recent CPU developments [3]. GPU computing even further accelerates this example [4]. Figure 3 highlights the speed-up when one, two or four GPUs are used over a reference system with the latest generation of CPUs.

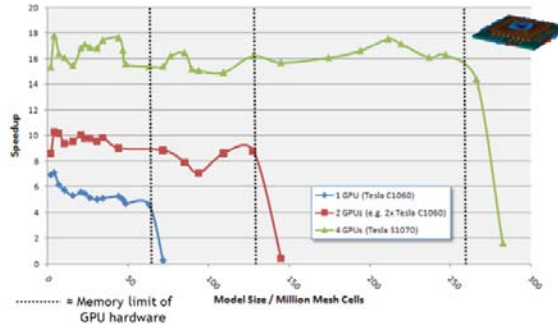


Figure 3: Speed-up of GPU computing. Reference is the new CPU generation [3].

Again, the speed-up is possible due to the higher memory bandwidth on the GPU hardware platform, which allows a greater throughput of the memory bandwidth limited algorithm.

Frequency domain techniques on the other hand require a matrix inversion within the solution process. This tends to require much more floating point operations per memory access. Therefore they benefit from multiple cores/CPU's even on systems with lower memory bandwidth.

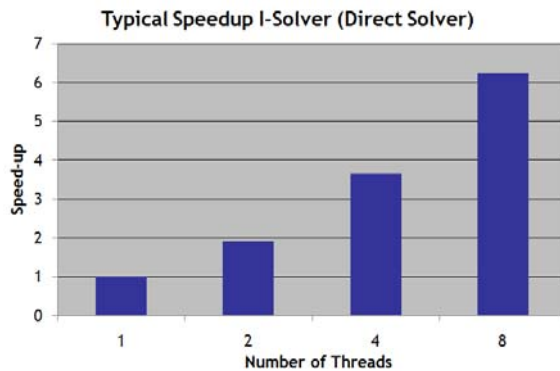


Figure 4: Typical speed-up of a direct methods-of-moments (MoM) solver vs. number of threads used on 8 core machine

Figure 4 shows the speed-up of multi core usage for a direct methods-of-moments (MoM) solver. It can be observed that an excellent speed-up of more than 6 on a 8 core machine can be achieved.

SOFTWARE ACCELERATION TECHNIQUES

The message passing interface protocol (MPI) [5] is a standard library used for parallel processing across multiple compute nodes via a standard interconnect protocol, e.g. Internet Protocol (IP). It is used in [2] to implement a Domain Decomposition scheme. Unlike the distributed computing scheme, see figure 1, MPI introduces some overhead in the simulation. In its time domain implementation the nodes need to exchange information about field components across the domain boundaries in each time step. Therefore the cluster computing becomes most efficient if the overhead - the time to synchronize the nodes - is small compared to the time to update the fields in the volume of one domain.

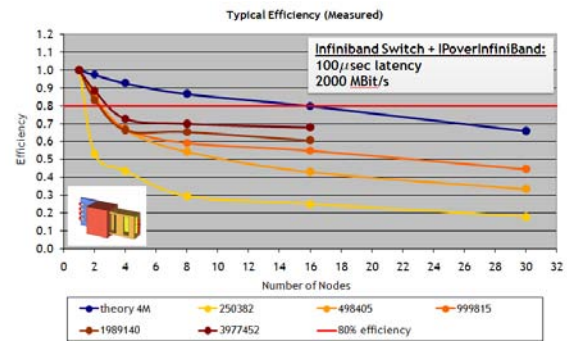


Figure 5: Efficiency of MPI based cluster computing for different discretizations vs. number of used cluster nodes

Figure 5 visualizes the efficiency of cluster computing, when multiple compute nodes are used. In addition it includes multiple discretizations in terms of mesh cell count for the same example. In theory the efficiency drops as the overhead gets larger with multiple nodes, highlighted in blue. If the model is large enough, one can get close to this limit, making it an effective scheme for simulations with distributed memory.

For independent simulations, multiple computers in the network can be used to simulate different

- Port excitations (time-domain solver)
- Frequency samples (frequency domain solver)
- Parameter combinations during parameter sweep or optimization

in parallel. Figure 6 illustrates the process of submitting a simulation from the frontend machine to the main controller. Since all the setup is done on the frontend it detects N independent simulations and asks the main controller for N individual jobs. The main controller checks the status of the connected solvers and distributes the simulations. After the solver is finished it reports back to the main controller. The frontend at the end automatically merges the results of the individual simulations, in order to make this process seamless for the user. The overall simulation performance therefore is an almost perfect linear speed-up with the number of

computing nodes for independent simulations. The fact that there is no user interaction required makes this scheme very efficient and scalable.

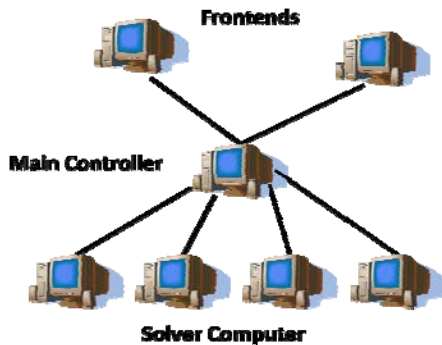


Figure 6: Computers involved and nomenclature of distributed computing

For the software/algorithmic acceleration, [2] offers an extension to FIT, the perfect boundary approximation (PBA) technique [6]. This conformal discretization method circumvents the geometry approximation error in classical finite-difference (FD) type of simulations. It allows using a much coarser mesh in order to achieve the same accuracy level and therefore speeds up the simulation, while the advantages of a structured meshing algorithm are maintained*. In addition the convergence during mesh refinement behaves more consistent, an important feature for any a posteriori error estimation.

SUMMARY

This paper introduced several hardware and software acceleration techniques, in particular for electromagnetic field computations. Figure 1 summarizes these and also illustrates how they interact with each other for maximum performance, e.g. cluster of GPU accelerated nodes.

As mentioned in the definition of performance the whole process from the idea to the final simulation design needs to be taken into account, including simulation setup and post processing. Another important factor for the performance is the right solver choice, e.g. time domain or frequency domain and mesh choice, e.g. hexahedral or tetrahedral.

Finally, the discretization method needs to be accurate (e.g. Perfect Boundary Approximation) otherwise the acceleration technique will only make a slow method less slow, but never fast.

* In [2] the user is not limited to structured hexahedral meshes, but can also use an unstructured mesh in time domain FIT simulations. This multilevel subgridding scheme also includes the PBA extension, giving the user full meshing flexibility and highest accuracy.

REFERENCES

- [1] Weiland, T.: A discretization method for the solution of Maxwell's equations for six-component fields: Electronics and Communication, (AEÜ), Vol. 31, pp. 116-120, 1977.
- [2] CST STUDIO SUITE™ 2009, www.cst.com
- [3] Intel® Xeon® processor 5500 series, <http://www.intel.com/technology/architecture-silicon/next-gen/>
- [4] NVIDIA® Tesla™ GPU http://www.nvidia.com/object/tesla_computing_solutions.html
- [5] Message Passing Interface (MPI), <http://www.mcs.anl.gov/research/projects/mpi/>
- [6] Krietenstein, B.; Schuhmann, R.; Thoma, P.; Weiland, T.: The Perfect Boundary Approximation technique facing the challenge of high precision field computation: Proc. of the XIX International Linear Accelerator Conference (LINAC'98), Chicago, USA, pp. 860-862, 1998.

GRAPHICAL PROCESSING UNIT-BASED PARTICLE-IN-CELL SIMULATIONS*

Viktor K. Decyk, Department of Physics and Astronomy, Tajendra V. Singh and Scott A. Friedman, Institute for Digital Research and Education, UCLA, Los Angeles, CA 90095, U. S. A.

Abstract

New emerging multi-core technologies can achieve high performance, but algorithms often need to be redesigned to make effective use of these processors. We will describe a new approach to Particle-in-Cell (PIC) codes and discuss its application to Graphical Processing Units.

INTRODUCTION

High Performance Computing (HPC) has been dominated for the last 15 years by distributed memory parallel computers and the Message-Passing Interface (MPI) programming paradigm. The computational nodes have been relatively simple, with only a few processing cores. This computational model appears to be reaching a limit, with several hundred thousand simple cores in the IBM Blue Gene. The future computational paradigm will likely consist of much more complex nodes, such as Graphical Processing Units (GPUs) or Cell Processors, which can have hundreds of processing cores, with different and still evolving programming paradigms, such as NVIDIA's CUDA. One anticipates that the next generation HPC computers, unlike Blue Gene, will consist of a relatively small number (<1,000) nodes, each of which will contain hundreds of cores. High performance on the node will in most cases require new algorithms. Between nodes, however, it is likely that MPI will continue to be effective.

Particle-in-Cell (PIC) codes [1-2] are one of the most important codes in plasma physics and other sciences, and use substantial computer time at some of the largest supercomputer centers in the world. Such codes integrate the trajectories of many charged particles, each interacting via electromagnetic fields they themselves produce. In anticipation of future requirements, we have been developing algorithms for PIC codes on this new class of multi-core nodes. As much as possible, we would like these new algorithms to be general enough that they would run well on most of the new emerging architectures. We decided to start with NVIDIA GPUs, because they are powerful, inexpensive, and widely available.

These GPUs consist of 12-30 multiprocessors, each of which has 8 processor cores. The control logic performs the same operation on 32 cores at a time. There is a large (up to 4 GBytes) global memory, which has very high aggregate bandwidth (up to 140 GBytes/sec), far higher than the memory bandwidth of a traditional processor. The memory latency (400-600 clocks) is quite

high, however. To hide this latency, the NVIDIA GPUs support thousands of threads simultaneously, and can switch threads in one clock period. To use this architecture, there are two challenges to any algorithm. The first is that the high global memory bandwidth is achieved only when adjacent threads read adjacent locations in memory (stride 1 access, or in the vocabulary of NVIDIA, data coalescing). This is due to the fact that memory is read 64 bytes at a time, and if all 64 bytes are used, memory bandwidth is maximized. The second is that there is no cache. However, each multiprocessor has a small (16 KB), fast (4 clocks) memory which can be shared by threads running on that multiprocessor. It is best to read and write global memory only once (with stride 1 access), storing the data that has to be read more than once or does not have stride 1 access, in small pieces. From this we concluded that ordered, streaming algorithms are optimal for this and similar architectures.

PIC codes have 3 major components. The first is a deposit step, where particles contribute charge or current field elements to grid points located near the particle's position. The deposit generally involves a scatter operation. The second is a field solver, where some subset of Maxwell's equation is solved to obtain values of electric and/or magnetic field points on a grid from the charge or current grid points. The third is a particle push step, where particles interpolate electric or magnetic fields at a particle's position by interpolating from nearby field elements. The push generally involves a gather operation. Normally, most of the time is spent in the deposit and push steps, since there are usually many more particles than grids. PIC codes typically have low computational intensity. That is, the number of floating point operations (FLOPs) compared to the number of memory accesses is small, around 2 or 3, so that optimizing memory operations is very important. Parallel algorithms for distributed memory parallel computers have been available for many years [3], and such codes have effectively used 1,000-100,000 processors.

PIC codes can implement a streaming algorithm by keeping particles constantly sorted by grid. This minimizes global memory access, since all the particles at the same grid point read the same field elements: the field elements need to be read only once for the entire group (and can be stored in registers). Cache is not needed, since gather/scatter operations are no longer required. Most importantly, it is possible to store particles so that the deposit and push procedures all have optimal stride 1 memory access. The challenge is whether one can sort the particles in an optimal way.

In this paper, we will discuss an implementation of a streaming algorithm for a simple 2D electrostatic

*Work supported by Northrop Grumman, UCLA IDRE, and USDOE (SciDAC)

particle code on the NVIDIA GPUs. This code involves depositing charge, solving a Poisson equation with a spectral method, and implementing a particle push with electric forces only. It is based on one of the codes from the UPIC Framework [4]. An electromagnetic code would differ only in the local operations (depositing current in addition to charge, including magnetic forces in the push), but not in the structure of the algorithm or its parallelization. The entire code runs on the GPU, in contrast to an earlier work which implemented only a deposit algorithm [5].

ORDERED CHARGE DEPOSIT

In a traditional PIC code, the particle coordinates are stored as (or ultimately translated to) grid units, where the integer part of a coordinate refers to the nearest grid point, and the deviation from the grid point is used as a weight in the interpolation. In the charge deposit, one would first extract the integer part of the coordinate, and then add the weights to the nearest grid points. The most commonly used interpolation is linear, which in 2D would involve the four nearest grid points. Two arrays are used as the data structures, a particle array *part* and a charge array *q*. In Fortran, they would be declared as follows:

```
dimension part(idimp,nop), q(nx+1,ny+1)
```

where *idimp* is the number of coordinates describing a particle. In this case there are 4 coordinates, corresponding to two positions, *x* and *y*, and two velocities, *vx* and *vy*, respectively. The size of the grid is given by *nx* and *ny*, and *nop* is the number of particles. The charge on a particle is given by *qm*. The traditional deposit loop is:

```
do j = 1, nop
  n = part(1,j)      ! extract x grid point
  m = part(2,j)      ! extract y grid point
  dxp = qm*(part(1,j) - real(n)) ! find weights
  dyp = part(2,j) - real(m)
  n = n + 1; m = m + 1 ! add 1 for Fortran
  amx = qm - dxp
  amy = 1.0 - dyp
  q(n+1,m+1) = q(n+1,m+1) + dxp*dyp ! deposit
  q(n,m+1) = q(n,m+1) + amx*dyp
  q(n+1,m) = q(n+1,m) + dyp*amy
  q(n,m) = q(n,m) + amx*amy
enddo
```

When particles are sorted, a new data structure is needed. Particles can still be stored in a 2D array as before, but they are now grouped together, and there could be gaps between groups, since the number of particles per grid can vary. The location of where a group of particles at a grid starts and the number of particles at that grid are stored in a separate array. The new data structures are declared in Fortran as follows:

```
dimension part(idimp,npmax) ! npmax > nop
dimension npic(2,nx*ny)
```

The element *npic*(1,*k*) contains the number of particles at grid *k*, and the element *npic*(2,*k*) contains the location in the array *part* where this group starts. The loop over particles now becomes a double loop as follows:

```
do k = 1, nx*ny      ! new outer loop over grids
  joff = npic(2,k)    ! memory offset
  do j = 1, npic(1,k) ! inner loop over particles at grid
    x = part(1,j+joff) ! obtain the x coordinate
    ...
  enddo
enddo
```

A charge deposit loop for ordered particles can be written:

```
k2 = 0
do k = 1, nx*ny      ! outer loop over grids
  k2 = k2 + 1        ! increment cell address
  sql = 0.0; squ = 0.0 ! zero out local accumulators
  sqlu = 0.0; squu = 0.0
  joff = npic(2,k)
  do j = 1, npic(1,k) ! loop over particles at grid
    dxp = qm*(part(1,j+joff)) ! find weights
    dyp = part(2,j+joff)
    amx = qm - dxp
    amy = 1.0 - dyp
    squu = squu + dxp*dyp ! first sum charges locally
    sqlu = sqlu + amx*dyp
    sql = sql + dyp*amy
    sqll = sqll + amx*amy
  enddo
  q(k2) = q(k2) + sqll ! then deposit sum in array
  q(k2+1) = q(k2+1) + sqlu
  q(k2+nx+1) = q(k2+nx+1) + sqlu
  q(k2+nx+2) = q(k2+nx+2) + squu
enddo
```

Note that the integer part of a coordinate no longer needs to be stored, since it is known from the grid location. This improves accuracy with 32 bit arithmetic, since all bits are used to store weights. The contribution of all the particles at a grid are first summed locally into register variables, then added to the grid. This reduces the number of memory references needed and improves the computation intensity of this subroutine from less than 2 to around 5.

PARALLEL CHARGE DEPOSITS

This ordered algorithm does not run safely in parallel, however, since a particle at one grid writes to other grids, and two threads cannot safely update the same grid at the same time. There are two possible approaches. A traditional approach is to implement an atomic update, where the sum $s = s + x$ is performed as a single,

uninterruptible operation by locking or protecting the memory in some fashion. CUDA supports atomic updates for integers, but not for floating point numbers. Protecting memory, however, is slow and not very portable in most computer languages. An alternative approach is to have each thread write to its own memory locations, which includes additional guard cells that are added up later. Such techniques are common in distributed memory algorithms, but require additional memory. We shall adopt the latter approach, which is known as domain decomposition.

The parallel algorithm will assign each thread `ngrid` grid points, which are defined in an array `kcell`. The charge density array `q` now needs to include guard cells for an extra row and column. If `ngrid < nx`, the number of guard cells needed is `ngrid+2`. The worst case is `ngrid = 1`, and `nthreads = nx*ny`, when 3 guard cells are needed for each grid. These new data structures are declared in Fortran as follows:

```
dimension q(2*ngrid+2,nthreads)
dimension kcell(2,nthreads)
```

The element `kcell(1,kth)` contains the initial grid index and `kcell(2,kth)` contains the final grid index for thread `kth`.

For a conventional processor, the parallelization can be expressed by adding an OpenMP style outer loop:

```
!$OMP PARALLEL
!$OMP DO
do kth = 1, nthreads ! parallel loop over threads kth
  kmin = kcell(1,kth) ! minimum cell number for thread
  kmax = kcell(2,kth) ! maximum cell number for thread
  ngrid = kmax - kmin + 1 ! number of cells for thread
  k2 = 0
  do k = kmin, kmax
    ! charge deposit loop for ordered particles as previously
    ! shown
    q(k2,kth) = q(k2,kth) + sqll ! deposit sum in array
    q(k2+1,kth) = q(k2+1,kth) + squl
    q(k2+ngrid+1,kth) = q(k2+ngrid+1,kth) + sqllu
    q(k2+ngrid+2,kth) = q(k2+ngrid+2,kth) + squu
  enddo
enddo
!$OMP END DO
!$OMP END PARALLEL
```

This algorithm will run correctly in parallel. However, it will not run optimally on the GPU. The reason is that adjacent threads do not read adjacent locations in memory (stride 1 access is not maintained). To achieve this, we must declare the arrays so that the thread index is the first dimension in Fortran arrays (in C, the last dimension):

```
dimension q(nthreads,2*ngrid+2), kcell(nthreads,2)
```

More importantly, we also need to partition the particle array and its associated data descriptor by thread index as well. We shall also assign the same number of grids to each thread:

```
dimension part(nthreads,idimp,npmax/nthreads)
dimension npic(nthreads,2,ngrid)
```

Other than reorganizing the data with the new partition, the algorithm remains the same.

FIELD SOLVER

The field solver used in this test code solved Poisson's equation, using spectral methods and making use of CUDA's CUFFT library. The algorithm has three steps. First perform a real to complex 2D FFT on the charge density `q`. Next, multiply the complex charge density `qk` by the quantity $-i\mathbf{k}/k^2$ to obtain the complex electric field `fk`. Finally, perform a complex to real 2D FFT to obtain the electric field `f` in real space.

We decided to use the `cufftExecC2C` function which performs multiple 1D complex to complex FFTs, and build our own 2D real to complex FFT using a well known algorithm [6]. The CUDA function requires the data to be packed with no gaps between elements. Since the input charge density array has some of the data in guard cells, we add the guard cells as we copy to a contiguous array. If we choose the parallel loop index to correspond to the index of the output array, this operation can be safely run in parallel. The output of this operation is the form:

```
complex, dimension q(nx/2,ny)
```

Once the data is copied, we perform multiple FFTs in `x` for each `y`. We then transpose the data, while modifying it as required by the algorithm[6]. This transpose has stride 1 only on the input. NVIDIA has examples of how to improve this, but so little time was used here, we did not do so. Finally, we perform multiple FFTs in `y` for each `x`. The result is an complex array `qk` of the form:

```
complex, dimension qk(ny,nx/2+1)
```

The field solver calculates the two component electric field `fk` in fourier space from `qk`, and the operation is reversed to obtain the 2 component electric field in real space:

```
complex, dimension fk(ny,2,nx/2+1)
complex, dimension f(nx/2,2,ny)
```

The final step is to create an electric field array with guard cells, described next.

PARALLEL PARTICLE PUSH

The particle push integrates Newton's equation of motion using a leap-frog scheme:

$$\begin{aligned} v(t+dt/2) &= v(t-dt/2) + f(x(t))*dt \\ x(t+dt) &= x(t) + v(t+dt/2)*dt \end{aligned}$$

where dt is the time step, and $f(x(t))$ is the force at the particle's position, found by interpolation. The push subroutine is structured the same as the deposit. To maintain stride 1 memory access, the electric field is partitioned just like the charge density:

$$f(nthreads, 2, 2*ngrid+2)$$

The partitioning is done by the field solver as its final step. The 8 components needed to interpolate the electric field for the group of particles at a grid are read once and stored in register variables, then reused by all the particles in the group. The inner loop of the push subroutine is as follows:

```
do k = 1, ngrid
  k2 = k2 + 1
  fxll = f(kth, 1, k2)           ! read forces
  fyll = f(kth, 2, k2)
  fxul = f(kth, 1, k2+1)
  fyul = f(kth, 2, k2+1)
  fxlu = f(kth, 1, k2+ngrid+1)
  fyly = f(kth, 2, k2+ngrid+1)
  fxuu = f(kth, 1, k2+ngrid+2)
  fyuu = f(kth, 2, k2+ngrid+2)
  joff = npic(kth, 2, k)
  do j = 1, npic(kth, 1, k) )    ! loop over particles at grid
    dxp = part(kth, 1, j+joff)   ! obtain coordinates
    dyp = part(kth, 2, j+joff)
    vx = part(kth, 3, j+joff)
    vy = part(kth, 4, j+joff)
    amx = 1.0 - dyp
    amy = 1.0 - dyp              ! find acceleration
    dx = dyp*(dyp*fxuu + amx*fxlu)
      + amy*(dyp*fxul + amx*fxll)
    dy = dyp*(dyp*fyuu + amx*fyly)
      + amy*(dyp*fyul + amx*fyll)
    vx = vx + qtm*dx             ! update coordinates
    vy = vy + qtm*dy
    dx = dyp + vx*dt
    dy = dyp + vy*dt
    part(kth, 1, j+joff) = dx     ! write coordinates
    part(kth, 2, j+joff) = dy
    part(kth, 3, j+joff) = vx
    part(kth, 4, j+joff) = vy
  enddo
enddo
```

PARALLEL PARTICLE SORTING

After the particles have been pushed, they may need to be placed in a new grid group and location in memory.

We have tried about a dozen algorithms, and finally selected one which appeared to be best. This algorithm assumes that most particles remain in the group, in which case they are written to the same location they had originally. This maintains stride 1 memory access as much as possible.

Within a thread, the groups are processed left to right. If a particle is going to a group outside the thread, the particle coordinates and destination group number are written to a message buffer owned by the thread. In addition, the location of the hole created by the departing particle in the original group is recorded in a hole array. If a particle is going to a group within the thread, there are two possibilities. For a particle going to a group to the right (which has not yet been processed), it is temporarily buffered at the end of the particle array in the destination group, and the location of the hole in the original group is recorded. Once a group has been processed, any holes in the group are filled from the temporarily buffered particles, starting from the last written particle and hole. Finally, if a particle is going to a group to the left (which has already been processed), it is placed either in a hole, if there is one, or added to the end of the group of particles.

Once all the particles are processed (there is an implicit synchronization point here), each thread examines the message buffers created by the other threads to see if any particles belong in this thread. To optimize this search, an array *icell* is created, which contains for each thread, the index of other possible threads to search. For linear interpolation and a uniform partition, this number is normally 8. The array *icell* changes whenever the partition described by the array *kcell* changes. The incoming particles are either placed in a hole, if there is one, or added to the end of the appropriate group. Finally, if any holes in a group are left, they are filled with particles from the end of the group.

For particles leaving a thread group, this algorithm is very similar to the message-passing schemes used by distributed memory PIC codes [3,7].

PERFORMANCE RESULTS

Porting this code to the GPU required first translating six (kernel) subroutines into C, and replacing the loop over threads

```
for (kth = 0; kth < nthreads; kth++)
```

with a special CUDA construct:

```
kth = blockIdx.x*blockDim.x+threadIdx.x;
```

In addition, memory had to be allocated on the GPU, and initial data copied from the host. Finally, wrapper functions were written to enable the kernel subroutines to be called from the main Fortran code. At the end of the simulation, the final charge density array was copied to be host to check for correctness.

The following benchmarks were run on a Macintosh Pro, with a 2.66 GHz Intel Xeon W3520(Nehalem) host and a C1060 Tesla card. Because Mac OS does not support the Tesla card, we installed the Fedora 11 Linux operating system on this hardware. The benchmark application had a 256x512 grid and 4,718,592 particles, with a timestep of $\omega_p dt = 0.025$. It was run in single precision, and the time reported is the time per particle per time step. For the benchmark, up to 131,072 threads were possible. However, it turned out that the optimal result was obtained for 8,192 threads (so that each thread had 16 grids), with 128 threads/multiprocessor.

	Intel Nehalem	Nvidia Tesla	Speedup
Deposit:	8.2 nsec.	0.16 nsec.	51
Push:	19.9 nsec.	0.56 nsec.	36
Sort:	-	1.30 nsec.	-
Total:	30.0 nsec.	2.27 nsec.	13

The overall speedup for the entire code was about 13. Most of the time was spent in the sorting step, particularly handling particles moving from one thread group to another. The field solver consumed only a small part of the total time in both cases. It should be noted that the new algorithm is more accurate in single precision than the original algorithm, as explained in the Ordered Charge Deposit section.

DISCUSSION

This version of the PIC code made use of global memory only. Access to global memory is the slowest part of the hardware, and is important to optimize that first. It is a very general algorithm and should run on any processor. The charge and push subroutines improved extremely well, with performance within a factor of 2 of the memory bandwidth limit. Clearly, the sorting step needs the most attention. We expect to improve the algorithm in the future by making use of faster local memories. We were somewhat surprised that we could effectively use distributed memory algorithms on such a device, in avoiding data conflicts and maintaining stride 1 memory access. We were impressed that using CUDA was so simple.

The 2D code used for development here is challenging because there are few operations and the overall computational intensity with the new algorithm improves only from 2 to 4 times. Our target application, however, is a 3D electromagnetic code, and our estimate is that the computational intensity with the new algorithm should improve from 2 to 30 times, so we expect much better results there. The sorting, even if not improved, should become relatively less important.

REFERENCES

- [1] Charles K. Birdsall and A. Bruce Langdon, *Plasma Physics via Computer Simulation* [McGraw-Hill, New York, 1885].
- [2] Roger W. Hockney and James W. Eastwood, *Computer Simulation Using Particles* [McGraw-Hill, New York, 1981].
- [3] P. C. Liewer and V. K. Decyk, "A General Concurrent Algorithm for Plasma Particle-in-Cell Codes," *J. Computational Phys.* 85, 302 (1989).
- [4] V. K. Decyk, "UPIC: A framework for massively parallel particle-in-cell codes," *Computer Phys. Comm.* 177, 95 (2007).
- [5] P. G. Stanchev, W. Dorland, and N. Gumerov, "Fast parallel Particle-to-Grid interpolation for plasma PIC simulations on the GPU," *J. Parallel Distrib. Comput.* 68, 1339 (2008).
- [6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* [Cambridge University Press, 1986], p. 504.
- [7] P. M. Lyster, P. C. Liewer, R. D. Ferraro, and V. K. Decyk, "Implementation and Characterization of Three-Dimensional Particle-in-Cell Codes on Multiple-Instruction-Multiple-Data Parallel Supercomputers," *Computers in Physics* 9, 420 (1995).

VIZSCHEMA – A STANDARD APPROACH FOR VISUALIZATION OF COMPUTATIONAL ACCELERATOR PHYSICS DATA*

S. Shasharina[#], J. Cary, M. Durant, S. Kruger, S. Veitzer, Tech-X Corporation, Boulder, CO, 80303, U.S.A.

Abstract

Even if common, self-described data formats are used, data organization (e.g. the structure and names of groups, datasets and attributes) differs between applications. This makes development of uniform visualization tools problematic and comparison of simulation results difficult. VizSchema is an effort to standardize metadata of HDF5 format so that the entities needed to visualize the data can be identified and interpreted by visualization tools. This approach allowed us to develop a standard powerful visualization tool, based on VisIt, for visualization of large data of various kinds (fields, particles, meshes) allowing 3D visualization of large-scale data from the COMPASS suite for SRF cavities and laser-plasma acceleration.

INTRODUCTION

Visualization is extremely valuable in providing better understanding of scientific data generated by simulations and guiding researchers in designing more meaningful experiments. Scientific models need to be compared with each other and validated against experiments. Consequently, most computational scientists rely on visualization tools. However, visualization and data comparison is often made difficult by the fact that various simulations use very different data formats and visualization tools.

Self-describing data formats are increasingly being used for storage of data generated by simulations. Such formats allow the code to store and access data within a file by name. The file storage system then takes care of developing an index for the data. In addition, the data can be decorated with attributes describing the units, dimensions, and other metadata for a particular variable. The self-describing formats now in use also help to deal with binary incompatibilities. Because different machine architectures use different binary representations for numbers, a binary file written by one processor may not be readable by another processor. Self-describing data file formats and interfaces ensure that the data is written in a universal binary format on all processors, and that software reading the data translates it to the appropriate architecture-specific format.

The Hierarchical Data Format (current version is HDF5) [1] and the NetCDF [2] format are in common use in the fusion, accelerator and climate modeling communities. HDF5 allows one to create a multi-tiered data structure inside of a file, so that one can create nested structures of groups and datasets.

Examples of HDF5 use include plasma physics codes such as VORPAL [3], a 3D plasma simulation code developed under development and Tech-X, and SYNERGIA [4], a multi-particle accelerator simulation tool developed at Fermilab. Both codes are actively used in the COMPASS SciDAC project [5]. Many other communities (earth sciences, fusion simulations) also use HDF5.

In spite of the fact that all these codes use self-describing data format, their files are organized very differently. They often do not share the node structure, do not agree on attributes, use different names for physically similar variables and store data in different structures. In other words, self-describing formats, though powerful, do not impose universally interpretable data structures.

For example, VORPAL put particles data in one dataset with all spatial information coming first: $x = \text{data}[0,:]$, $y = \text{data}[1,:]$, $z = \text{data}[2,:]$, followed by momenta: $p_x = \text{data}[3,:]$, $p_y = \text{data}[4,:]$, $p_z = \text{data}[5,:]$, while SYNERGIA intermixes momenta and spatial information: $p_x = \text{data}[0,:]$, $x = \text{data}[1,:]$ etc.

How one can guess from looking at the data what is what? How does one recognize that a particular dataset represents a mesh and what kind of mesh is it? How does one indicate that a dataset is mapped to a particular mesh? Which data ordering is used (is it grouped by components or position indices)? Using some standards and common metadata within these formats could resolve this problem.

Visualization tools used by different teams are also very non-uniform. For a long time, scientific community used IDL [6] and AVS/Express [7]. Lately, many teams are moving towards the freely available, open source, high-quality visualization tools such as VisIt [8].

In this paper we present our efforts to develop such a standard for computational applications dealing with field and particles data. Our approach is based on first identifying the entities of interest to visualization, relationships between these entities and then defining intuitive and minimalistic ways to express them using metadata and common constructs used in self-described data formats: *groups*, *datasets*, and *attributes*. We call this data model VizSchema.

It is then used to implement a VisIt plugin (called Vs) which reads visualization entities from HDF5 files into memory and creates VisIt data structures thus providing a data importing mechanism from VizSchema compliant HDF5 files into VisIt.

In what follows we describe the VizSchema data model, Vs plugin, give examples of visualization and discuss future directions.

*Work supported by DOE grant DE-FC02-07ER54907.

[#]sveta@txcorp.com

VIZSCHEMA DATA MODEL

Principles

In this section we describe the elements of the VizSchema. These elements identify the data structures that one needs to expose in order to do visualization. They are not about HOW the visualization is performed (i.e. the type of light or position of the camera); instead, they are WHAT is being visualized (data and geometry) and WHAT needs to be exposed for minimal default visualization.

In designing the schema we use the following guiding principles:

- VizSchema assumes that data comes as one of three types: *variables* (data which lives on a mesh described outside of the HDF5 node containing the data), *variables with meshes* (data which mixes physical values with the spatial information which is contained within the same HDF5 nodes) and *meshes*.
- These entities are identified by HDF5 markup and have particular attributes specific to their types.
- All the markup for the schema should be contained in the attributes so that users could choose the names of the data itself (typically contained in groups and datasets) as they please. The markup can be generated during I/O or added in a post-processing step. We expect these attributes to start with “vs”.
- VizSchema attributes can refer to other entities using their short or fully-qualified names. If a short name is used, the reader will first search in the same space and then enlarge the search until the matching name is found.
- Each vs entity has an attribute vsType, which describes its category (variable or mesh, for example).
- Some entities have different kinds (i.e. subtypes), in which case a vsKind attribute specifies the kind.

Although, the schema entities described below use HDF5 lingo, mapping to the NetCDF lingo is straightforward; one needs just to substitute the term “variable” in place of “dataset.” In the remainder of this section we give some details of the VizSchema elements.

Variables and Variables With Mesh

We assume that data comes as one of two kinds: a *variable* or a *variable with mesh*. A variable represents data, which lives on a mesh described outside of the variable array, while a variable with mesh contains spatial information within itself. In Particle-in-Cell simulations, all fields share the same mesh, so this mesh is described once and the values of the electric and magnetic fields do not contain the spatial information but rather depend on the tool to determine the mesh that they live on. Such fields are typically “variables.” For particle data, one typically outputs their momentum and position in one dataset, so here the tool is supposed to generate a point mesh from within this dataset. So, particle data is a “variable with mesh.” The suggested markup gives the information to the visualization tool to interpret the data.

In the following pseudo-code snippet we show the variable markup in HDF5:

```
Dataset "phi" {
  Att vsType = "variable"
  Att vsMesh = "mycartgrid"
  Att vsCentering = "zonal"
}
```

The vsType attribute in this example indicates that this dataset needs to be visualized and needs a mesh called *mycartgrid* to be defined elsewhere in the file. The optional attribute vsCentering instructs that the data should be interpolated to a zone (with the default being nodal). The dimensions of the variable can be derived from querying the dataspace and are not needed in the explicit metadata.

Since variables with mesh mix spatial and other data in one dataset, there should be a way to specify the data structure. If the dataset’s first N indices specify the coordinates (like in VORPAL), one could use the following markup:

```
Dataset "vorpalelectrons" {
  Att vsType = "variableWithMesh"
  Att vsNumSpatialDims = N
}
```

If the layout of data is different from this order (for example, like in SYNERGIA), one needs to use vsSpatialIndices, which would indicate which indices of the dataset contain spatial information:

```
Dataset "synergiaElectrons" {
  Att vsType = "variableWithMesh"
  Att vsSpatialIndices = [1, 3, 5]
}
```

Since the data can be ordered in many various ways, one also needs to describe the ordering of the data or the order of indices starting from the fastest-varying. For example, for the 3D case:

```
compMinorC = (i0, i1, i2, ic)
compMinorF = (ic, i2, i1, i0)
compMajorC = (ic, i0, i1, i2)
               (same as compMinorF for 1D)
compMajorF = (i2, i1, i0, ic)
               (same as compMinorC for 1D)
```

In component minor order, the indices (i0, i1, i2, ic) are such that the component index, ic, appears last. The C reference would be array[i0][i1][i2][ic], while the Fortran reference would be array(i0,i1,i2, ic). In component major, the indices (ic, i0, i1, i2) are such that the component index, ic, appear first. The C reference would be array[ic][i0][i1][i2], while the Fortran reference would be array(ic,i0,i1,i2).

When addressing the array in memory, two adjacent memory locations can differ by incrementing either the first index (Fortran) or the last index (C). Since the data is generally written to HDF5 files without changing the

order, the component index must be specified. The default value of this attribute is compMinorC. This attribute is needed to reorder data as expected by a visualization tool.

Derived Variable

It is often useful to define additional variables, which are not being dumped by a simulation but present an interesting thing to see as well. That is why, in addition to the prime variable described above, we allow defining expressions using regular mathematical symbols. For example, one could define a density of electric energy as follows:

```
Group anygroupname {
  Att vsType = "variableDefinition"
  Att vsDefinition = "elecEnergyDensity =
(E_0+E_0+E_1+E_1+E_2+E_2)"
}
```

In defining this, we assume that the visualization tool can parse and evaluate such expressions. These assumptions are valid for our VisIt plugin implementation, which uses Python as its expression language.

Meshes

There is no uniform classification of meshes across tools and experiments. Based on our experience with several codes, we determined that the following mesh type categorizations are fairly general:

- Structured grid, which is defined by a list of points defined by their coordinates.
- Rectilinear grid, which is defined by the lists of increasing coordinate values for each axis and is a specialization of a structured grid
- Uniform grid (sometimes also called uniform Cartesian), which has constant distances between nodes in all directions and is a specialization of a rectilinear mesh
- Unstructured grid, which are defined by points and cells of various types.

The VizSchema markup for these mesh types is shown by the following examples. The first example describes a structured mesh with component-minor ordering. The dataset contains the mesh's points as an array ordered in X, Y, and Z, with 3 values (x,y,z) at each mesh point, for a total of 4 array dimensions:

```
Dataset "mystructmesh" {
  Att vsType = "mesh"
  Att vsKind = "structured"
  Att vsIndexOrder = "compMinorC"
  Att vsStartCell = [0, 0, 0]
}
```

The second example describes a 2D rectilinear mesh. It is a group containing 2 datasets, each of which contains the mesh points along one axis (X, Y). The optional vsAxis* attributes provide a name for each axis.

```
Group "myrectgrid" {
  Att vsType = "mesh"
```

```
  Att vsKind = "rectilinear"
  Att vsAxis0 = "axis0"
  Att vsAxis1 = "axis1"
  Dataset axis0[n0]
  Dataset axis1[n1]
}
```

The third example describes a 3D uniform mesh. Since all the mesh points are uniformly distributed, the coordinates of each point do not have to be provided. Instead, the VS attributes give the start and end position and number of cells along each axis, permitting a visualization tool to generate the mesh.

```
Group "myunigrid" {
  Att vsType = "mesh"
  Att vsKind = "uniform"
  Att vsStartCell = [0, 0, 0]
  Att vsNumCells = [200, 200, 104]
  Att vsLowerBounds = [-2.5, -2.5, -1.3]
  Att vsUpperBounds = [2.5, 2.5, 1.3]
}
```

The final example describes a 3D unstructured mesh. Such a mesh is generated from two arrays, one containing the coordinates of the mesh points, and the other containing entries giving the set of points that compose each cell in the mesh. By default, the coordinate array is named "points" and the cell array is named "polygons". The optional attributes vsPoints and vsPolygons permit arrays with non-default names to contain this information.

```
Group "mypolymesh" {
  Att vsType = "mesh"
  Att vsKind = "unstructured"
  Att vsPoints = "points"
  Att vsPolygons = "polygons"
}
```

The list of supported kinds of meshes will be growing as we encounter more kinds of simulation data. Some of them will need to have mappings to already existing types with the data translations implemented in the Vs plugin.

Multi-Domain Data

Quite often simulation data comes from multiple domains and uses different names in these domains, while it would be natural to treat it as one variable in a continuous domain. For such cases, we use vsMD attribute, which instructs visualization tools to connect data having the same value of this attribute.

Here is an example of two domain blocks that will be treated as a single multi-domain mesh named "edgeMesh" and the two variables psiPriv and psiSol are declared to be an md variable named psi:

```
Dataset "privMesh" {
  Att vsType = "mesh"
  Att vsKind = "structured"
  Att vsMD = "edgeMesh"
}
Dataset "solMesh" {
  Att vsType = "mesh"
  Att vsKind = "structured"
  Att vsMD = "edgeMesh"
}
```



```

Dataset "psiPriv" {
  Att vsType = "variable"
  Att vsMesh = "privMesh"
  Att vsMD = "psi"
}
Dataset "psiSol" {
  Att vsType = "variable"
  Att vsMesh = "solMesh"
  Att vsMD = "psi"
}

```

Summary Of The Data model

To summarize, the visualization data model consists of variables, variables with mesh and meshes and their metadata. Variables metadata includes their names, their meshes, data ordering and centering. Variables with mesh have metadata for their name, data ordering, centering and separation of values from the spatial information. Meshes metadata depends on the mesh kind and fully describes each kind.

There are also variables defined as expressions and links that allow creating multi-domain variables.

In addition to the metadata described above, visualization needs additional metadata needed for correct allocation of the memory. For example, each dataset has its internal type (int, for example) and dimensions. This metadata should also be extracted before the visualization is possible but does not have to be present in the data markup.

VS PLUGIN

Based on the data model described above, we implemented a C++ data reader class, which reads all the needed metadata from HDF5 files into the memory. This reader creates an object that reflects the structure of an HDF5 file as it is seen by visualization – lists of variables with the meshes that they live on, variables with meshes, derived variables and meshes and all their metadata. Once such object is created, one uses the reader's methods for reading these entities by their name. All the data is returned as a void* array (consistent with HDF5 model) for which memory should be allocated based on the metadata of this entity. The interface of the reader class is independent of the type of the visualization tool and is implemented for HDF5 data.

Next we created a VisIt plugin using the reader's API. This plugin is available for the download at <https://ice.txcorp.com/trac/vizschema/wiki/WikiStart>. We are in the process of adding it to VisIt repository so it will be available upon VisIt installations.

EXAMPLES

Several codes adopted VizSchema and now provide the compliant output during I/O. One can also change the files after they have been generated using PyTables [9] (we have successfully using to change data as the schema evolved and also to annotate SYNERGIA files in accordance with the schema).

The plugin code was tested on Linux and OS X and is installed on such supercomputers as franklin.nersc.gov. Figs. 1-4 show some examples of visualizations done using the VizSchema plugin for VisIt. Fig. 1 is a screen capture of OASCR Award for Scientific Visualization at the 2008 Scientific Discovery through Advanced Computation Conference (Seattle) for the video, "Visual Inspection of a VORPAL Modeled Crab Cavity."

Fig. 2 has been used as a cover for one of the issues of SciDAC review magazine [10]. Fig. 3 shows visualization for SYNERGIA data. Fig. 4 shows an example of multi-domain visualization and demonstrates that VizSchema is general enough to accommodate applications outside of computational accelerator physics: data from FACETS (Framework Application for Core-Edge Transport Simulations) [11].

CONCLUSIONS AND FUTURE DIRECTIONS

Standardization of the HDF5 output using consistent markup for visualization proved to be useful in accelerator physics applications as well as other domains having notions of fields and particles. The developed VisIt plugin is available for all interested parties.

In the nearest future we intend to extend the schema and the plugin with more detailed metadata for unstructured meshes and bring more applications into the VizSchema realm.

It will be interesting to develop a means to automatically annotate data with the markup. One could have a text or XML input for mapping internal data to the data elements of the schema and then use PyTables to add the expected attributes.

ACKNOWLEDGMENT

We would like to thank the VisIt, FACETS and VORPAL teams for helping us in this work. We also thank C. Geddes and G. Weber for providing us with images and J. Amundsen for providing us with SYNERGIA data.

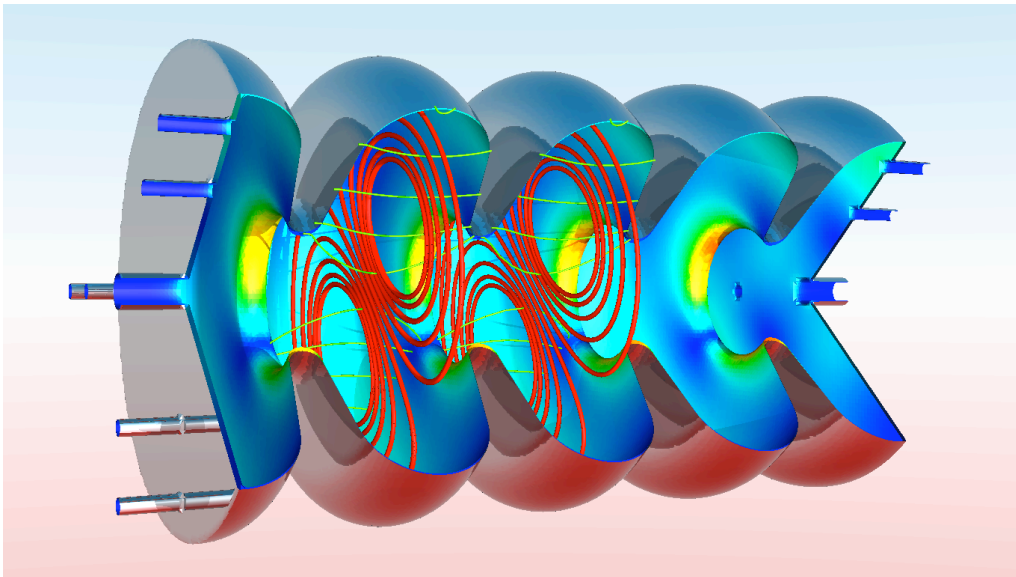


Figure 1: Examples of a visualization of VORPAL data: electromagnetic fields (red and green) and magnetic stress on the cavity (on the walls).

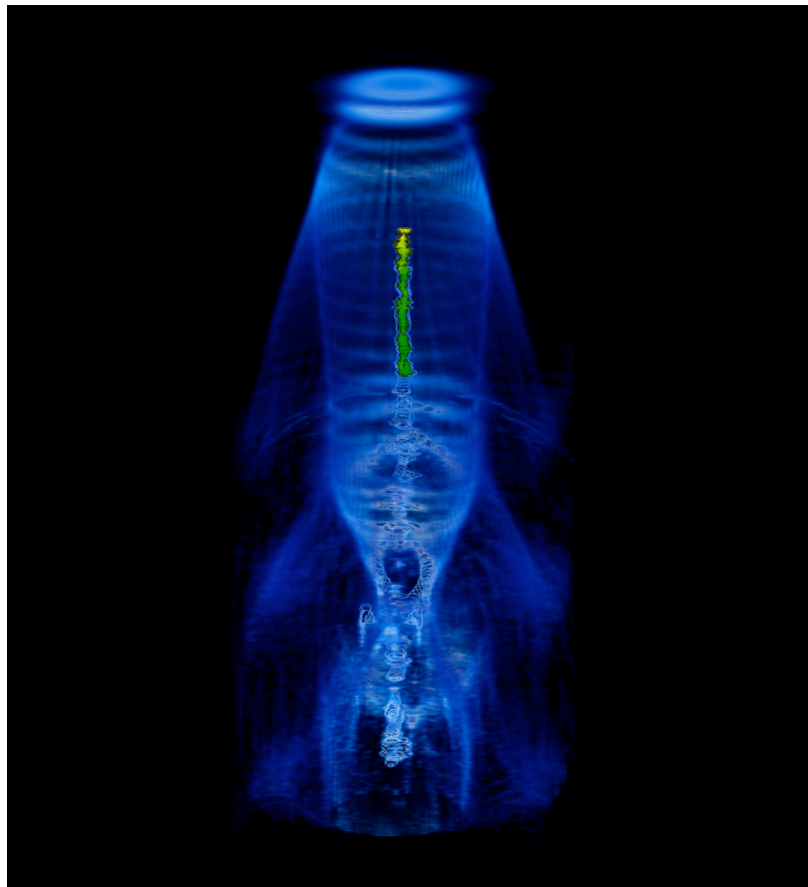


Figure 2: A three-dimensional VORPAL simulation models the self-consistent evolution of the wake resulting from a laser pulse and the acceleration of particles in a laser-plasma particle accelerator. Shown in volume rendering are the wake (blue) and a particle bunch (green and yellow). Courtesy of *G.H. Weber and C. Geddes*.

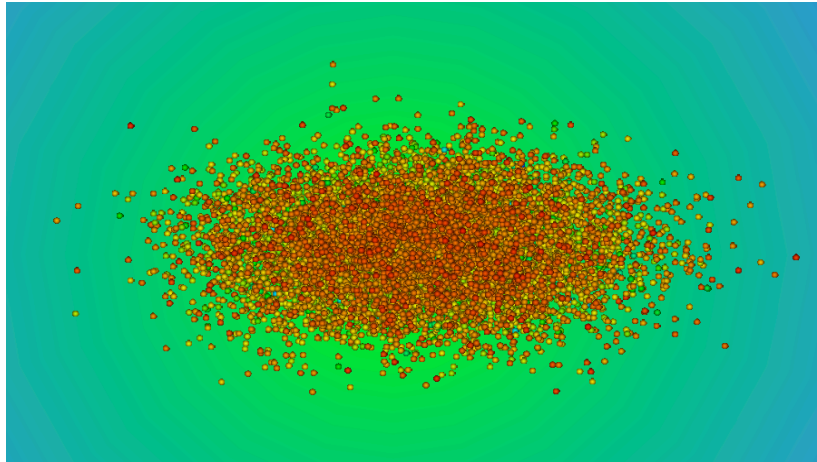


Figure 3: Examples of a visualization of SYNERGIA data: beam colored by the energy of the particles.

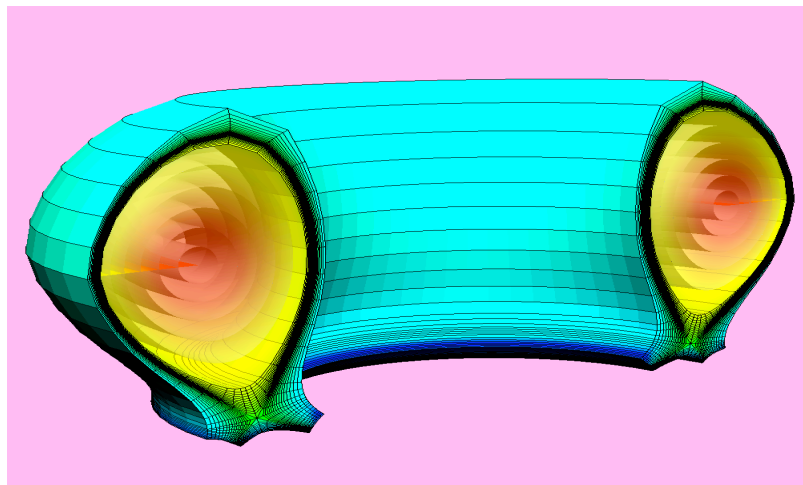


Figure 4: Examples of a visualization of FACETS data: electron temperature defined in multiple domains.

REFERENCES

- [1] <http://hdf.ncsa.uiuc.edu/HDF5/>.
- [2] <http://www.unidata.ucar.edu/packages/netcdf>
- [3] C. Nieter and J/ R. Cary, "VORPAL: a versatile plasma simulation code," *J. Comp. Phys.* 196, 448-472 (2004).
- [4] J. Amundson and P. Spentzouris, J. Qiang and R. Ryne, Synergia: A 3D Accelerator Modelling Tool with 3D Space Charge, *Journal of Computational Physics*, Volume 211, Issue 1, 1 January 2006, Pages 229-248.
- [5] <http://www.scidac.gov/physics/COMPASS.html>.
- [6] <http://www.rsinc.com/idl/index.asp>.
- [7] <http://www.avsc.com/>.
- [8] H. Childs, E. S. Brugger, K. S. Bonnell, J. S. Meredith, M. Miller, B. J. Whitlock and N. Max, A Contract-Based System for Large Data Visualization, *Proceedings of IEEE Visualization 2005*, pp 190-198, Minneapolis, Minnesota, October 23--25, 2005.
- [9] <http://www.pytables.org>.
- [10] <http://www.scidacreview.org/0903/index.html>.
- [11] J. R. Cary, J. Candy, J. Cobb, R. H. Cohen, T. Epperly, D. J. Estep, S. Krashennikov, A. D. Malony, D. C. McCune, L. McInnes, A. Pankin, S. Balay, J. A. Carlsson, M. R. Fahey, R. J. Groebner, A. H. Hakim, S. E. Kruger, M. Miah, A. Pletzer, S. Shasharina, S. Vadlamani, D. Wade-Stein, T. D. Rognlien, A. Morris, S. Shende, G. W. Hammett, K. Indireskumar, A. Yu. Pigarov and H. Zhang, Concurrent, parallel, multiphysics coupling in the FACETS project, *SciDAC 2009, J. Physics: Conf. Series* 180, 012056 (2009).

THE OBJECT ORIENTED PARALLEL ACCELERATOR LIBRARY (OPAL), DESIGN, IMPLEMENTATION AND APPLICATION

A. Adelman*, Ch. Kraus, Y. Ineichen, PSI, Villigen Switzerland
S. Russell, LANL, Los Alamos, USA,
Y. Bi, J.J Yang, CIAE, Beijing, China

Abstract

OPAL (Object Oriented Parallel Accelerator Library) is a tool for charged-particle optic calculations in accelerator structures and beam lines including 3D space charge, short range wake-fields and 1D coherent synchrotron radiation and particle matter interaction. Built from first principles as a parallel application, OPAL admits simulations of any scale, from the laptop to the largest High Performance Computing (HPC) clusters available today. Simulations, in particular HPC simulations, form the third pillar of science, complementing theory and experiment. OPAL has a fast FFT based direct solver and an iterative solver, able to handle efficiently exact boundary conditions on complex geometries. We present timings of OPAL-T using the FFT based space charge solver with up to several thousands of cores.

OPAL IN A NUTSHELL

OPAL is a tool for charged-particle optics in accelerator structures and beam lines. Using the MAD language with extensions, OPAL is derived from MAD9P and is based on the CLASSIC class library, which was started in 1995 by an international collaboration. The Independent Parallel Particle Layer (IP^2L) is the framework which provides parallel particles and fields using data parallel ansatz, together with Trilinos for linear solvers and preconditioners. Parallel input/output is provided by H5Part/Block a special purpose API on top of HDF5. For some special numerical algorithms we use the Gnu Scientific Library (GSL).

OPAL is built from the ground up as a parallel application exemplifying the fact that HPC (High Performance Computing) is the third leg of science, complementing theory and experiment. HPC is now made possible through the increasingly sophisticated mathematical models and evolving computer power available on the desktop and in super computer centres. OPAL runs on your laptop as well as on the largest HPC clusters available today.

The state-of-the-art software design philosophy based on design patterns, makes it easy to add new features into OPAL, in the form of new C++ classes. Figure 1 presents a more detailed view into the complex architecture of OPAL.

OPAL comes in the following flavors:

- OPAL-T
- OPAL-CYCL

* andreas.adelman@psi.ch

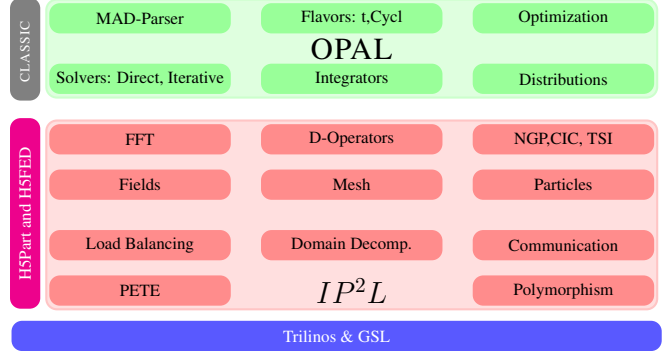


Figure 1: The OPAL software structure

- OPAL-MAP (not yet fully released)
- OPAL-ENVELOPE (not yet fully released)

OPAL-T tracks particles with time as the independent variable and can be used to model beam lines, dc guns, photo guns and complete XFEL's excluding the undulator. Collective effects such as space charge (3D solver), coherent synchrotron radiation (1D solver) and longitudinal and transverse wake fields are considered. When comparing simulation results to measured data, collimators (at the moment without secondary effects) and pepper pot elements are important devices. OPAL-CYCL is another flavor which tracks particles with 3D space charge including neighboring turns in cyclotrons, with time as the independent variable. Both flavors can be used in sequence, hence full start-to-end cyclotron simulations are possible. OPAL-MAP tracks particles with 3D space charge using split operator techniques. OPAL-ENVELOPE is based on the 3D-envelope equation (à la HOMDYN) and can be used to design XFEL's

Documentation and quality assurance are given our highest attention since we are convinced that adequate documentation is a key factor in the usefulness of a code like OPAL to study present and future particle accelerators. Using tools such as a source code version control system (subversion), and source code documentation (Doxygen) together with an extensive user manual we are committed to provide users as well as co-developers with state-of-the-art documentation for OPAL. Rigorous quality control is realized by means of daily build and regression tests.

In the sequel we will only discuss features of OPAL-T based on the current production version 1.1.5.

MODELS

In recent years, precise beam dynamics simulations in the design of high-current low-energy hadron machines as well as of 4th generation light sources have become a very important research topic. Hadron machines are characterized by high currents and hence require excellent control of beam losses and/or keeping the emittance of the beam in narrow ranges. This is a challenging problem which requires the accurate modeling of the dynamics of a large ensemble of macro or real particles subject to complicated external focusing, accelerating fields and wake fields, particle-matter interaction, as well as the self-fields caused by Coulomb interaction of the particles. In general the geometries of particle accelerators are large and complicated which has a direct impact on the numerical solution method.

Some of the effects can be studied by using a low dimensional model, i.e., envelope equations [1, 2, 3, 4]. These are a set of ordinary differential equations for the second-order moments of a time-dependent particle distribution. They can be calculated fast, however the level of detail is mostly not sufficient for quantitative studies. Furthermore, a priori knowledge of critical beam parameters such as the emittance is required with the consequence that the envelope equations cannot be used as a self-consistent method.

One way to overcome these limitations is by considering the Vlasov-Poisson description of the phase space, including external fields and self-fields and, if needed, other effects such as wakes. To that end let $f(\mathbf{x}, \mathbf{v}, t)$ be the density of the particles in the phase space, i.e., the position-velocity (\mathbf{x}, \mathbf{v}) space. Its evolution is determined by the collisionless *Vlasov equation*,

$$\frac{df}{dt} = \partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \frac{q}{m_0} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \nabla_{\mathbf{v}} f = 0, \quad (1)$$

where m_0 , q denote particle mass and charge, respectively. The electric and magnetic fields \mathbf{E} and \mathbf{B} are superpositions of external fields and self-fields (space charge),

$$\mathbf{E} = \mathbf{E}_{\text{ext}} + \mathbf{E}_{\text{self}} + \mathbf{E}_{\text{wake}}, \quad \mathbf{B} = \mathbf{B}_{\text{ext}} + \mathbf{B}_{\text{self}}. \quad (2)$$

If \mathbf{E} and \mathbf{B} are known, then each particle can be propagated according to the equation of motion for charged particles in an electromagnetic field,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}(t)}{dt} = \frac{q}{m_0} (\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

After the movement of the particles \mathbf{E}_{self} and \mathbf{B}_{self} have to be updated. To that end we change the coordinate system into one moving with the particles. By means of the appropriate *Lorentz transformation* [5] we arrive at a (quasi-) static approximation of the system in which the transformed magnetic field becomes negligible, $\hat{\mathbf{B}} \approx 0$. The Computer Codes (Design, Simulation, Field Calculation)

transformed electric field is obtained from

$$\hat{\mathbf{E}} = \hat{\mathbf{E}}_{\text{self}} = -\nabla \hat{\phi}, \quad (3)$$

where the electrostatic potential $\hat{\phi}$ is the solution of the *Poisson problem*

$$-\Delta \hat{\phi}(\mathbf{x}) = \frac{\hat{\rho}(\mathbf{x})}{\varepsilon_0}, \quad (4)$$

equipped with appropriate boundary conditions. Here, $\hat{\rho}$ denotes the spatial charge density and ε_0 is the dielectric constant. By means of the inverse Lorentz transformation the electric field $\hat{\mathbf{E}}$ can then be transformed back to yield both the electric and the magnetic fields in (2).

In OPAL the discretized Poisson equation is either solved by a combination of a Green function and FFT or by a conjugate gradient algorithm, preconditioned with algebraic multi-grid using smoothed aggregation (SA-AMG PCG). This 3D solver has the unique capability to include the exact boundary geometry. The right hand side in (4) is discretized by sampling the particles at the grid points. In (3), $\hat{\phi}$ is interpolated at the particle positions from its values at the grid points. We also note that the FFT-based Poisson solvers and similar approaches [6, 7] are usually restricted to box-shaped or open domains in order to obtain good performance.

Field Solver

A Direct FFT Based Poisson Solver In our implementation of the PIC method, firstly a rectangular 3D grid containing all particles is constructed. Subsequently, the charges are interpolated onto the grid points. Then the discretized Poisson equation is solved on the grid to obtain the scalar field at the grid points. The electric field is calculated on the grid and interpolated back on to the positions of the particles.

In 3D Cartesian coordinates, the solution of the Poisson equation at point \mathbf{x} can be expressed by

$$\phi(\mathbf{x}) = \frac{1}{4\pi\varepsilon_0} \int G(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \quad (5)$$

with G the 3D Green function

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{(\mathbf{x} - \mathbf{x}')^2}} \quad (6)$$

assuming open boundary conditions. The typical steps of calculating space charge fields using Hockney's FFT algorithm is sketched in Algorithm 1, where the quantities with superscript D (discrete) refer to grid quantities.

The image charge of a beam near a cathode is not negligible, hence open boundary conditions are not justified in such a case. To find the space-charge forces on the beam from the image charge by the standard Green function method, we need to solve the Poisson equation with a computational domain containing both the image charge and the beam. We are using a shifted-Green function [8]

Algorithm 1 3D Space Charge Calculation

- 1: **procedure** 3DSpaceCharge(In: ρ , G , Out: \mathbf{E}_{sc} , \mathbf{B}_{sc})
- 2: Create 3D rectangular grid which contains all particles,
- 3: Interpolate the charge q of each macro-particle to nearby mesh points to obtain ρ^D ,
- 4: Lorentz transformation to obtain ρ^D in the beam rest frame \mathbf{S}_{beam} ,
- 5: FFT ρ^D and G^D to obtain $\hat{\rho}^D$ and \hat{G}^D ,
- 6: Determine $\hat{\phi}^D$ on the grid using $\hat{\phi}^D = \hat{\rho}^D \cdot \hat{G}^D$,
- 7: Use FFT⁻¹ of $\hat{\phi}^D$ to obtain ϕ^D ,
- 8: Compute $\mathbf{E}^D = -\nabla\phi^D$,
- 9: Interpolate \mathbf{E} at the particle positions \mathbf{x} from \mathbf{E}^D ,
- 10: Perform Lorentz back transform to obtain \mathbf{E}_{sc} and \mathbf{B}_{sc} in frame \mathbf{S}_{local} and transform back to \mathbf{S}_{lab} .
- 11: **end procedure**

technique in order to efficiently compute the correct potential at the cathode. With this technique, the FFT is used to calculate the cyclic convolution and the previous algorithm can be used to calculate the potential in the shifted field domain.

At emission from a dc gun, or when calculating neighboring turns in a cyclotron, the electrostatic approximation is not valid anymore. To overcome this problem we divide the beam into n energy bins. The space charge solver uses now n separate Lorentz transformations.

To show the parallel performance of OPAL-T we consider two problems, the first one has $5 \cdot 10^6$ particles on a $64 \times 64 \times 128$ mesh and 200 time steps are considered. The used CPU time as a function of cores is shown in Figure 2. We obtain in the order of $5 \cdot 10^6$ particle pushes per second on a 16 nodes (HP BL460c blades) cluster, each node having a dual-socket quad core Intel Xeon E5450 3.0 GHz with 16GB ECC RAM. For a high-bandwidth low-latency communication the InfiniBand interconnect is used.

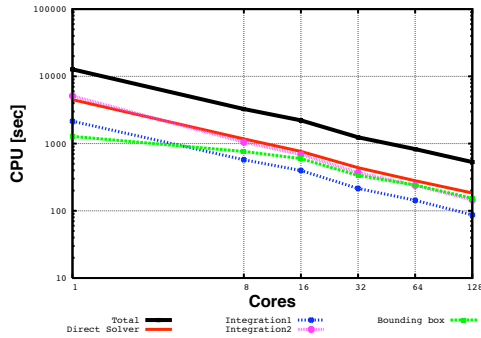


Figure 2: CPU time of a production run showing the scaling of the most important parts of OPAL-T on a 128 core HP Cluster.

The second problem consists of $1 \cdot 10^8$ particles on a 512^3 mesh and timings for 3 integration steps are shown in Figure 3. The timings were obtained on the Cray XT5 cluster of the Swiss Supercomputing Center (CSCS) in Manno. Each of the 1844 compute nodes consists of 2 quad-core AMD Opteron 2.4 GHz Shanghai processors giving 8 cores in total per node with 16 GBytes of memory. The high-speed network based on a SeaStar 2.2 communications processor which is able to provide 2 GBytes/s of injection bandwidth for the node, with a theoretical peak of 9.6 GBytes/s of bandwidth in each direction for the through-flow of packets out on the network.

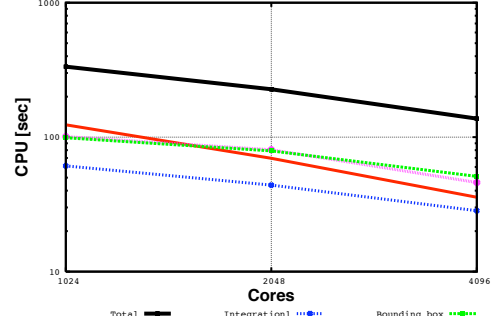


Figure 3: CPU time of a test run showing the scaling of the most important parts of OPAL-T on a Cray XT5.

A Fast Iterative Parallel Poisson Solver on Irregular Domains The problem is discretized by finite differences. Depending on the treatment of the Dirichlet boundary the resulting system of equations is symmetric or ‘mildly’ nonsymmetric positive definite. In all cases, the system is solved by the preconditioned conjugate gradient algorithm with smoothed aggregation (SA) based algebraic multigrid (AMG) preconditioning. Additionally we investigated variants of the implementation of SA-AMG that lead to considerable improvements in the execution times. We demonstrate good scalability of the solver on distributed memory parallel processor with up to 2048 processors in [9]. In this paper we also compare our SAAMG-PCG solver with the FFT-based solver described in the preceding paragraph.

Particle Matter Interaction

The physics models describing particle matter interaction includes energy loss and Coulomb scattering. The nuclear scattering is not yet included for particles in the order of hundreds of MeVs. Their contribution is negligible compared to Coulomb scattering. The energy loss model is based on the Bethe-Bloch equation. Comparing the stopping power with the PSTAR program of National Institute of Standards and Technology (NIST), we find errors in the order of 10% for copper, from several MeV to 10 GeV. Important for our immediate application at PSI, the error is within 3% in the region from 50 MeV to 1 GeV. In general, there is energy straggling when a beam passes through

the material. For relatively thick absorbers such that the number of collisions is large, the energy loss distribution is Gaussian [10]. The Coulomb scattering is treated as two independent events: the multiple Coulomb scattering and the large angle Rutherford scattering, using the distribution given in [11].

Validation A 72 MeV cold Gaussian beam with $\sigma_x = \sigma_y = 5$ mm is send through a copper slit with the half aperture of 3 mm from 0.01 to 0.1 m. Figure 4 shows some trajectories of particles which are either absorbed or deflected by the collimator. Most of the particles were absorbed

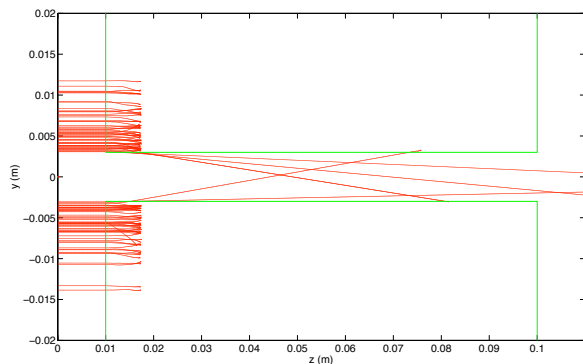


Figure 4: Trajectories of particles which are either absorbed or deflected by the collimator.

within a range of about 7.4 mm, except for a few which were deflected by the collimator. As a benchmark of the elliptic collimator models in OPAL, the energy spectrum and angle deviation is compared against two general-purpose Monte Carlo codes, MCNPX [12] and FLUKA [13, 14], as shown in Fig. 5. The deflected particles contribute to the energy spectrum and angle deviation after a collimator. These particles may be lost downstream.

ACKNOWLEDGMENTS

The majority of computations have been performed on the Cray XT5 in the framework of the PSI CSCS “Horizon” collaboration. We acknowledge the help of the XT5 support team at CSCS, most notably Timothy Stitt and D. Kiselev for the MCNPX simulations and fruitful discussions w.r.t. the particle matter interaction models.

REFERENCES

- [1] Frank James Sacherer. *Transverse Space-Charge Effects in Circular Accelerators*. PhD thesis, University of California, Berkeley, 1968.
- [2] Frank J. Sacherer. RMS envelope equations with space charge. *IEEE Trans. Nucl. Sci.*, 18(3):1105–1107, 1971.
- [3] J. Struckmeier and M. Reiser. Theoretical studies of envelope oscillations and instabilities of mismatched intense

Computer Codes (Design, Simulation, Field Calculation)

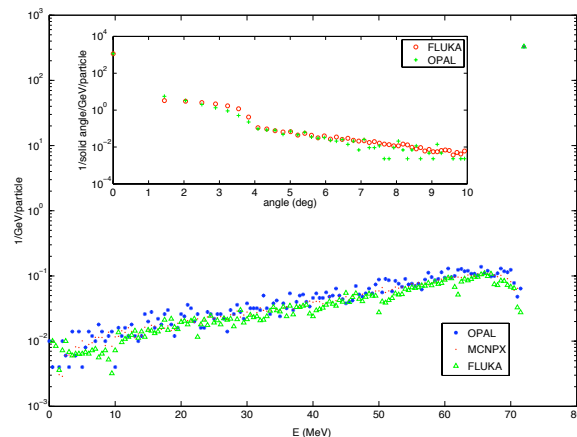


Figure 5: Energy spectrum and angle deviation (small plot)

charged-particle beams in periodic focusing channels. *Part. Accel.*, 14(2-3):227–260, 1984.

- [4] R. L. Gluckstern. Analytic model for halo formation in high current ion linacs. *Phys. Rev. Lett.*, 73(9):1247–1250, 1994.
- [5] L. D. Landau and E. M. Lifshitz. *Electrodynamics of Continuous Media*. Pergamon, Oxford, 2nd edition, 1984.
- [6] Ji Qiang and Robert D. Ryne. Parallel 3D Poisson solver for a charged beam in a conducting pipe. *Comp. Phys. Comm.*, 138(1):18–28, 2001.
- [7] Ji Qiang and Robert L. Gluckstern. Three-dimensional Poisson solver for a charged beam with large aspect ratio in a conducting pipe. *Comp. Phys. Comm.*, 160(2):120–128, 2004.
- [8] Ji Qiang, Steve Lidia, and Robert D Ryne. Three-dimensional quasistatic model for high brightness beam dynamics simulation. *Phys. Rev. ST Accel. Beams*, 9(4):1–10, Apr 2006.
- [9] A Adelmann, P Arbenz, and Y Ineichen. A fast parallel poisson solver on irregular domains applied to beam dynamic simulations. *arXiv, physics.comp-ph*, July 2009.
- [10] William R. Leo. *Techniques for nuclear and particle physics experiments*. Springer-Verlag, Berlin Heidelberg New York, 2nd edition, 1994.
- [11] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, 3rd edition, 1998.
- [12] D. Pelowitz ed. MCNPX Users Manual. Technical Report Version 2.5.0 LA-CP-05-0369.
- [13] G. Battistoni, S. Muraro, P.R. Sala, F. Cerutti, A. Ferrari, S. Roesler, A. Fasso, and J. Ranft. The fluka code: Description and benchmarking. In *Proceedings of the Hadronic Shower Simulation Workshop 2006*, Fermilab, 2006. M. Albrow, R. Raja eds., AIP Conference Proceeding 896, 31-49, (2007).
- [14] A. Fasso, A. Ferrari, J. Ranft, and P.R. Sala. Fluka: a multi-particle transport code. Tech. Report CERN-2005-10 (2005), INFN/TC.05/11, SLAC-R-773, 2005.

RECENT PROGRESS AND PLANS FOR THE CODE *elegant* *

M. Borland [†], V. Sajaev, H. Shang, R. Soliday, Y. Wang, A. Xiao, ANL, Argonne, IL 60439, USA
W. Guo, BNL, Upton, NY, 11973, USA

Abstract

elegant is an open-source accelerator code that has been in use and development for approximately two decades. In that time, it has evolved from a graduate student project with a narrow purpose to a general code for the design and modeling of linacs and storage rings. *elegant* continues to evolve, thanks in no small part to suggestions from users. *elegant* has seen extensive application to modeling of linacs, particularly for applications related to free-electron lasers and energy recovery linacs. Recent developments have emphasized both linac and storage-ring-related enhancements, along with parallelization. In this paper, we briefly review the features of *elegant* and its program suite. We then describe some of the recent progress made in the ongoing development of *elegant*. We also discuss several noteworthy applications and directions for future work.

INTRODUCTION

The program *elegant* [1] is now widely used in the accelerator community and is available as source code or in binary form for many operating systems. It started more than two decades ago as a graduate student project when the lead author concluded that it was easier to write a new code than to modify existing codes to include needed features. Since then, it has undergone almost continuous incremental improvement, with releases at approximately six-month intervals. The original structure and philosophy of the code are well suited to this process.

A basic *elegant* run has two inputs: a command input file and a lattice definition file. The command input file contains a series of namelist-like structures defining a series of commands to set up and execute a run. The lattice input file defines the lattice using a format that is very similar to that popularized by the program MAD [2].

One of the design goals of *elegant* was to make adding a new element no harder than writing code to implement the physics of the element. Toward this end, a set of data structures was defined that allows the developer to describe the properties and parameters of any new element, as well as the properties of those parameters. This element dictionary has made incremental improvement of the code relatively painless. (It is also used to automatically generate the manual pages for all elements.)

* Work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[†] borland@aps.anl.gov

elegant attempts to implement as many features as possible using a lumped-element concept. For example, one may impart charge to a beam or change the Twiss parameters of a beam using a lumped element. This has the advantage of allowing *elegant* to vary or optimize such properties just as it could for a property of a quadrupole or any property of another traditional beamline element. Similarly, many local diagnostic outputs are obtained by inserting one of several diagnostic elements into the beamline.

elegant was the first accelerator code to make thorough use of self-describing data for input and output, starting originally with the Access With Ease (AWE) protocol [3] and transitioning in 1993 to the Self-Describing Data Sets (SDDS) protocol [4]. This feature is as important as the element dictionary in allowing incremental improvement and delivering new results to users in a consistent, usable fashion. With SDDS we can add new data to the output without disrupting users and applications that use the output files. We can also make use of general-purpose pre- and post-processing tools that are not *elegant*-specific.

In what follows, we discuss recent improvements in *elegant* and some of the programs distributed with it. We'll begin by discussing improvements of a general nature, followed by a discussion of new features that are specific to ring modeling. Next, we'll summarize the status of on-going parallelization of the code, then turn to a discussion of recent changes to related programs. Finally, we will briefly review some recent applications of *elegant* and plans for future development. This paper covers changes starting with version 16.0 and ending with version 22.1.

GENERAL IMPROVEMENTS

Although *elegant* ("ELEctron Generation ANd Tracking") was written for electron tracking, repeated requests were made to allow tracking of other particles. The new `change_particle` command allows the user to choose different particles by name or specify the charge and mass of the particle of interest.

Optimization is an important feature of *elegant* and perhaps one of its strengths, compared to other codes. *elegant*'s optimizer uses a single penalty function that is the sum of many terms, each of which is specified as an expression by the user. Essentially anything the program computes, including intermediate and final results of tracking, can be used in an optimization term. New features in optimization include the ability to define optimization terms from templates, so that many similar optimization terms may be added without much effort. We've also added

the ability to read optimization terms from an SDDS file, so that they can be generated by external programs or scripts.

`elegant` features several elements for modeling wigglers and undulators, including the new UKICKMAP element, which implements undulator kick maps [5]. A script was written and distributed with `elegant` that translates RADIA [6] kick map output into SDDS for input to `elegant`. The CWIGGLER element (a canonically integrated wiggler using code from Y. Wu [7]) was improved to use more general field expansions, such as those resulting from certain helical or vertically polarized devices. The CWIGGLER element now also includes classical and quantum synchrotron radiation effects when tracking.

Another element that involves an undulator is LSRMDLTR, which simulates a laser/undulator beam heater. This element was upgraded to include a time-dependent laser profile as well as synchrotron radiation.

A number of methods of simulating synchrotron radiation effects are provided in `elegant`. One of these is element-by-element simulation during tracking. By default, modeling of quantum effects uses Gaussian energy scattering of simulation particles [8], which might not be accurate for beams with very small emittance or energy spread. To allow examination of these, the CSBEND element now allows modeling synchrotron radiation using the energy and angle distributions for the emitted photons.

Of course, one can use tracking with element-by-element synchrotron radiation modeling to compute the beam properties along a beamline or even at equilibrium in a storage ring. However, a more efficient method [9] is propagation of the beam envelope using matrix techniques, including the damping and diffusion effects of radiation. This is now available in `elegant`, both for storage rings and transport lines, using the `moments_output` command. The results of these computations at any number of points in the lattice may also be subjected to optimization. One application of this is coupling minimization.

Another source of beam size and energy spread is intra-beam scattering (IBS), which can be modeled in `elegant` using the IBSCATTER element. The algorithm behind IBSCATTER has been improved to include the effect of vertical dispersion [10], to allow multipole scattering locations, and to allow modeling of IBS with acceleration [11]. The program `ibsEmittance`, which is distributed with `elegant`, includes the same changes and can be used for computing equilibrium properties in storage rings.

In order to model IBS along a beamline using IBSCATTER, it is necessary to insert many IBSCATTER elements in the lattice. To make this easy, we added the `insert_elements` command, which allows inserting a new element at multiple locations in a lattice without editing the lattice definition file. A companion command, `replace_elements`, is also new. It allows replacing existing elements with new ones. In both cases, the result can be saved as a new lattice file.

Another example of using `insert_elements` would be to insert many WATCH elements in a lattice in order to get Computer Codes (Design, Simulation, Field Calculation)

phase-space dumps at many locations. In the past, this wouldn't work as expected because all the output files would have the same name, so that only the last occurrence would be retained. In addition, in modeling a large beamline one could easily attempt to open more files than the operating system allows. Hence, we improved the WATCH element to allow versioning of the filenames and to make use of the SDDS library's file disconnection feature to allow an essentially unlimited number of output files.

`elegant` models lumped-element beam pipe apertures of various types. Using `insert_elements` provides a new way to add apertures to an existing lattice at specific locations; e.g., one could insert the same aperture downstream of all quadrupoles with the same name. To further improve aperture specification, we've added the `aperture_data` command, which allows providing the beam aperture as a function of position along a beamline using an SDDS file. This file can, of course, be plotted together with loss distribution data or beam size data, using `sddspplot`.

When modeling errors in long transport lines, the simulated beam may be completely lost on the apertures before making a full pass. This makes trajectory correction using traditional methods problematical. To address this, we added two new trajectory correction methods. The `thread` method attempts to thread the beam through the system looking only at transmission. The `one-to-best` method pairs each corrector with the downstream BPM showing the largest response. Both methods attempt to imitate what might be done in early-stage commissioning.

Singular-value decomposition (SVD) is a standard technique for orbit and trajectory correction in situations where the beam is fully transmitted. `elegant`'s correction algorithm has been updated to use SVD, including various methods of downselecting the singular values that are used.

`elegant` is used frequently as part of a procedure for routine correction of lattice functions in the APS [12]. We know that many of our gradient errors result from orbit offsets in sextupoles. Under such circumstances, the sextupoles act like combined function sextupoles and quadrupoles. For convenience in the correction procedure, we added the KQUSE element, which is a combined canonically integrated quadrupole and sextupole magnet. This magnet can also be used in advanced storage ring designs that posit combined function magnets of this type.

STORAGE RING MODELING

A topic of considerable recent interest in the storage ring community has been the use of pulsed sextupole magnets to perform injection [13]. To support modeling of such concepts, `elegant` now includes the MBUMPER element, which simulates a time-dependent multipole kicker. The waveform for the kicker is supplied using an SDDS file.

No matter what injection method is used, having sufficient dynamic aperture (DA) is an important consideration in obtaining high injection efficiency. We've improved the aperture search algorithms in `elegant` to include multi-

line scans from the origin as well as a “smart” calculation of the aperture area that ignores lobes that may indicate an unreliable result. In addition, the DA area may now be optimized using `elegant`’s built-in optimization methods. Another method of improving dynamic aperture is minimization of resonance driving terms [14]. The computation and optimization of these quantities is now included in `elegant`, so that they can be optimized along with other linear and nonlinear properties of the lattice.

Particularly in light source rings, coupling has an important effect on DA. `elegant` now supports computation of coupled lattice functions, based on Ripken’s method [15], and allows these to be optimized. This could be used, for example, to correct coupling in the presence of errors.

DA is just one aspect of storage ring optimization. Equally important is the position-dependent momentum aperture [16], which determines the Touschek lifetime. This computation is now included in `elegant`. It can be used for storage rings, of course, but has also been applied to single-pass systems like energy recovery linacs (ERLs) [17].

One of `elegant`’s strengths in ring simulations is modeling collective effects. This includes transverse and longitudinal short-range wakes and resonant impedances. The former are computed turn-by-turn, while the latter persist over many turns. An effective way to avoid spurious transients in such simulations is to ramp the impedance gradually from zero [18]. This feature has been added for transverse and longitudinal wakes, impedances, and rf modes.

Typically when modeling a ring with impedances, one needs to track a large number of particles for many turns to get reliable results. Hence, one cannot afford element-by-element tracking. At the same time, one needs to include higher-order transport effects, e.g., chromaticity or tune shift with amplitude, as these may provide damping. This can be done with the new `ILMATRIX` element, which stands for Individualized Linear `MATRIX`. This element can stand in for an entire storage ring or a superperiod, for example. The user specifies the periodic lattice functions, the tunes, and the momentum- and amplitude-dependent tune shifts. Synchrotron radiation effects can also be included with the (pre-existing) `SREFFECTS` element, providing a very fast simulation with all the essential features.

In addition to impedance elements, `elegant` can now simulate transverse space-charge kicks in a storage ring [19]. This is accomplished using the `insert_sceffects` command, which inserts a number of `SCMULT` elements. Each of these elements imparts an effective space-charge kick that simulates the effect of weak space-charge forces over the intervening distance from the previous element. Using `insert_sceffects`, it is trivial to vary the number of elements used in order to verify convergence.

PARALLELIZATION

With the increasing emphasis on multicore processors in laptops and desktops, parallelization is essential to the `Computer Codes (Design, Simulation, Field Calculation)`

ture of any simulation code. `Pelegant`, the parallel version of `elegant`, has been successfully run on dual-core laptops and 1000-core supercomputers. The status of parallelization is detailed elsewhere in this conference [20, 21].

Parallelization of `elegant` is being performed gradually, concurrent with on-going improvements to the serial version, while maintaining a single set of source code files. The initial approach was to parallelize only those elements that involve “embarrassingly parallel” operations, then gradually parallelize the elements that involve inter-process communication. The code is capable of switching between parallel and serial mode automatically as required, based on information in the element dictionary. This approach resulted in a very useful parallel version in about 6 months, which was put to immediate use. At present, just under 90% of the elements have been parallelized for multi-particle tracking. Optimization that involves tracking also makes use of parallel computation.

Originally, the master node handled all input/output (I/O) and performed particle scatter/gather operations as needed. As a result of I/O and memory bottlenecks, this approach was limited to about 60M particles (for 16 GB of RAM on the master node). A significant recent improvement was the addition of parallel I/O using the parallel `SDDS` library [21] and subsequent elimination of the central role of the master processor. This has allowed simulation with hundreds of millions of particles with significantly improved performance.

In addition to basic tracking of multi-particle beams, which involves particle-based domain decomposition, several other operations were recently parallelized. These include frequency map analysis, dynamic aperture searching, and momentum aperture searching [20]. We believe `Pelegant` is the first parallel code to offer these features, although the `elegantRingAnalysis` script [22] provides equivalent functionality.

RELATED PROGRAMS

The consistent use of `SDDS` files makes it easy to deploy `elegant` as a component of a larger application. Examples of this abound in the use of `elegant` along with other accelerator codes and free-electron laser (FEL) codes to perform start-to-end modeling for FEL light sources. One necessary component of such simulations is the ability to translate phase-space conventions among codes. Recent additions of this type include a pair of programs to translate between `elegant` and `ASTRA` [23] conventions, and another pair to translate between `elegant` and `TRACK` [24] conventions. Several scripts are available to translate `IMPACT` [25] output, including phase-space output, into `SDDS`.

Modeling the effects of coherent synchrotron radiation (CSR) is of course important in linac-based light sources, but it is also of interest in storage ring design. A convenient way to model CSR in a storage ring is to use the steady state CSR impedance with shielding [26]. This impedance can be computed and placed in an `SDDS` file ready for use with

elegant using the new program `csrImpedance`.

Different aspects of synchrotron radiation are handled by the new programs `sddsurgent` and `sddsfluxcurve`, which augment `sdsdbrightness` in providing computations of synchrotron radiation properties. `sddsurgent` provides computation of flux distributions and spectra using code from the programs URGENT [27] and US [28]. `sddsfluxcurve` provides computation of flux tuning curves using code from US. All of these programs take beam distribution data from `elegant` output files.

Another recent addition is `touschekLifetime` [17], which allows computation of Touschek lifetime using optics and momentum aperture data generated by `elegant`.

`elegant`'s commandline interface and use of SDDS strongly supports script-based automatation of simulations. Along these lines, we have written the graphical user interface script `elegantRingAnalysis` [22], to provide a convenient interface to many storage ring computations. `elegantRingAnalysis` is designed to take advantage of a computing cluster, but can run on a single processor. `elegantRingAnalysis` uses `sdsfindresonances`, another recent tool, to find resonances in frequency map data.

SOME RECENT APPLICATIONS

In this section we briefly highlight a few recent applications, some of which use the new features discussed above.

Short-pulse x-rays in rings: APS has investigated the application of Zholents' scheme [29] for crab-cavity-based short pulse x-ray production from a storage ring. Originally [30], we used serial `elegant`, but the studies later benefited immensely from the parallel version. In particular, the optimization of sextupole to reduce vertical emittance dilution [31] and the exploration of the use of pulsed cavities [32] both benefited from rapid turn-around with the parallel version. Figure 1 shows an example of using `Pelegant` to optimize the results of one-pass tracking to minimize the vertical emittance growth while maintaining the desired chromaticity. `elegant` was also used to investigate this scheme for the Diamond Light Source [33]. Several other short-pulse schemes, one based on a vertical kicker [34], another based on rf phase modulation [35], and a third based on circulation of a short injected pulse [36], were also investigated with `elegant`.

Storage ring optimization: Optimization of dynamic and momentum aperture is a challenging aspect of storage ring design. Using `elegant` and `Pelegant`, we implemented several highly successful direct methods [37] of optimizing dynamic and momentum aperture based on tracking and genetic optimization. Figure 2 shows an example of frequency map analysis performed with `elegantRingAnalysis` on 100 processors for an optimized APS lattice with 10 long straight sections. Earlier, `elegant` was used to develop potential replacement rings for the APS [38, 39] as well as ultimate storage ring light sources [40]. `elegant` is also being used to optimize the NSLS-II lattice in the presence of strong damping wig-

glers using both direct optimization and minimization of resonant driving terms [14, 41], to model injection into the ring [42], and to investigate instabilities [43].

ERL design and modeling: `elegant` has been used extensively for design and simulation of ERLs, both at APS and elsewhere (e.g., [44, 45, 46, 47]). The APS group has produced several designs [48, 49] for possible upgrades. This included simultaneous optics matching for beams of multiple energies in the same beamline, as in Figure 3, which shows an optics solution for a two-pass 7-GeV linac. Tracking with `Pelegant` was essential in evaluating concerns about the microbunching instability [50]. Other `elegant`-based investigations at APS are Touschek scattering simulation and loss minimization [19], intrabeam-scattering simulation [11], optics correction [51], and x-ray compression [52].

FEL design and simulation: The recent success of the Linac Coherent Light Source (LCLS) at SLAC [53] has demonstrated the power of modern simulation tools, including `elegant`, to accurately predict the performance of future accelerators [54, 55]. `elegant` was used for development of the LCLS design and is part of on-going work to develop new operating modes for LCLS [56, 57].

The FERMI project has made use of `elegant` for a wide variety of design and simulation problems. This includes study of the laser heater [58] and related diagnostics [59], trajectory correction [60], jitter [61], beam instabilities [62], bunch length diagnostics [63], and the microbunching instability [64]. Figure 4 shows the evolution of a density modulation in FERMI, modeled with `Pelegant`.

Many other FEL projects are using `elegant`, including efforts in Korea [65], the United Kingdom [66, 67], Italy [68], Sweden [69], Switzerland [70], the United States [71, 72], Germany [73, 74], Japan [75], and China [76].

International Linear Collider (ILC): The ILC is a proposed next-generation electron-positron collider based on superconducting technology. `elegant` has been used in several aspects of ILC design, including the electron source [77], positron source [78, 79], bunch compression [80], and damping ring [19, 81].

FUTURE DEVELOPMENT

While there is no formal plan for future work on `elegant`, we anticipate that in the not-too-distant future the following enhancements will be made available. Users are encouraged to send suggestions for additional features.

1. Simultaneous parallel optimization of dynamic and momentum aperture, to allow use of the built-in simplex optimizer to perform storage ring nonlinear optimization. This should be more convenient than the existing method using a genetic optimization script.

2. Upgrading of the CSR algorithm to include shielding, non-relativistic beams, and multiple magnet effects, using the method of Sagan *et al.* [82]. Although the existing algorithm seems to correspond very well to experiments on LCLS, this upgrade will extend the validity to longer

bunches and lower energies.

3. Addition of higher-order wakes and long-range resistive wall wakes is desirable.

4. Inclusion of IBS in moments computations is desirable in order to have a self-consistent result including radiation effects and IBS.

5. Improved coupling correction, using cross-plane response matrices and vertical dispersion correction. At present, coupling correction can only be done using somewhat artificial methods, such as correction of the coupled lattice functions or moments, or using an external script.

6. Built-in lattice correction using LOCO [83]. At present, this is performed by an external script. With parallel resources, it should be possible to quickly perform the simulated response matrix measurement and correction.

CONCLUSION

elegant and related tools are under continuous, incremental development for linac and storage ring simulation. We have reviewed some of the many new features added in the last three to four years, as well as highlighted some applications. Details of the features discussed in this paper may be found in the *elegant* manual, which is available on-line [84]. An additional source of information and assistance is the on-line forum [85].

ACKNOWLEDGEMENTS

We appreciate assistance from Y. Wu in understanding his symplectic integrator code, which is used by the CWIGGLER element. We also appreciate the many suggestions and questions from *elegant*'s users, which contribute significantly to the on-going improvement of the program.

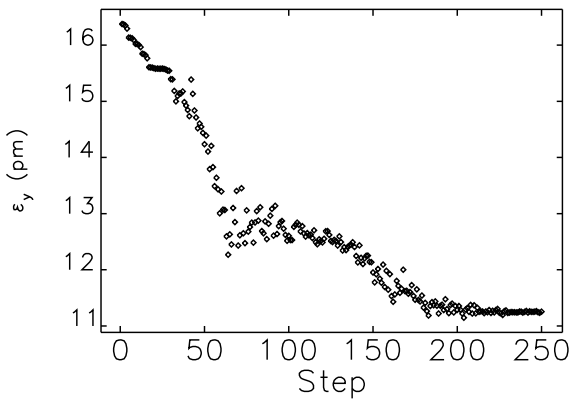


Figure 1: Vertical emittance optimization in the presence of crab cavities performed on 60 processors.

REFERENCES

[1] M. Borland, APS LS-287, September 2000.
 [2] H. Grote and F. C. Iselin, CERN/SL/90-13(AP) (1991).
 Computer Codes (Design, Simulation, Field Calculation)

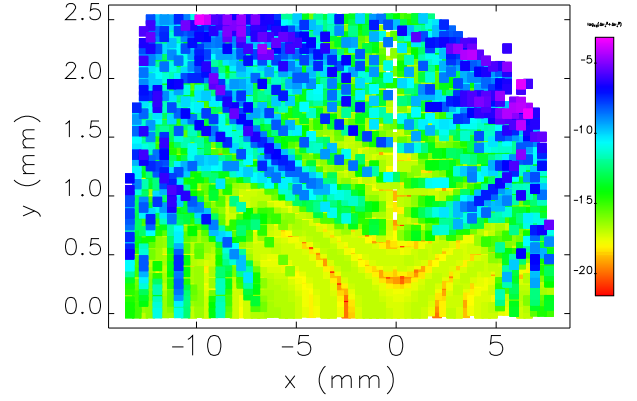


Figure 2: Frequency map analysis for a possible future APS lattice, using *elegantRingAnalysis* and 100 processors.

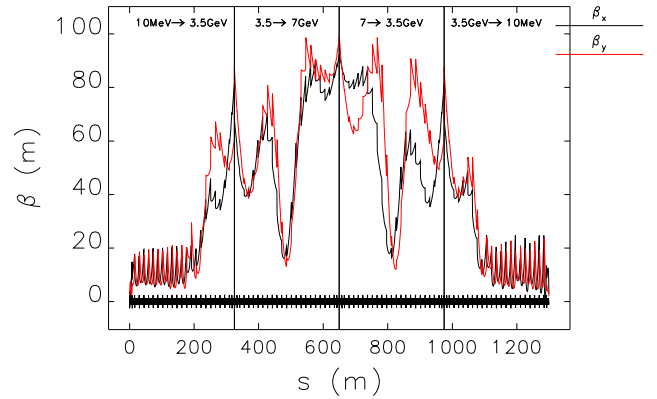


Figure 3: Two-pass energy recovery linac optics determined by simultaneous matching of four beams [49].

[3] M. Borland, Appendix A.2, SLAC-402 (1991).
 [4] M. Borland, Proc. PAC95, 2184-2186 (2006).
 [5] P. Elleaume, Proc. EPAC92, 661-663 (1993).
 [6] P. Elleaume *et al.*, Proc. PAC97, 3509-3511 (1998).
 [7] Y. Wu *et al.*, Phys. Rev. E **68**, 046502 (2003).
 [8] M. Sands, SLAC-R-121, 1970.
 [9] K. Ohmi *et al.*, Phys. Rev. E **49**, 751-765 (1994).
 [10] F. Zimmerman, CERN-AB-2006-002 (2006).
 [11] A. Xiao, Proc. Linac 2008, 296-298 (2009).
 [12] V. Sajaev and L. Emery, Proc. EPAC02, 742-744 (2002).
 [13] H. Takaki *et al.*, Proc. EPAC08, 2204-2206 (2008).
 [14] J. Bengtsson, SLS Note 9/97 (1997).
 [15] G. Ripken, DESY Report No. R1-70/04, 1970.
 [16] M. Belgroune *et al.*, Proc. PAC03, 896-898 (2003).
 [17] A. Xiao and M. Borland, Proc. PAC07, 3453-3455 (2007).
 [18] M. Blaskiewicz, private communication.
 [19] A. Xiao *et al.*, Proc. PAC07, 3456-3458 (2007).
 [20] Y. Wang *et al.*, these proceedings.

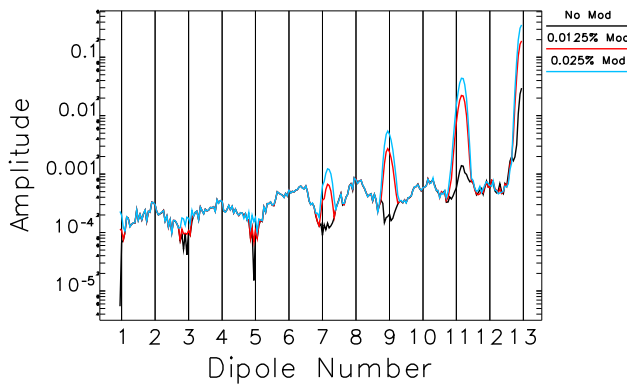


Figure 4: Modulation amplitude evolution in FERMI, for three initial levels of 25- μ m modulation [64].

- [21] H. Shang *et al.*, these proceedings.
- [22] M. Borland, Proc. PAC05, 4200-4202 (2005).
- [23] K. Flöttmann, www.desy.de/~mpyflo/Astra_dokumentation/
- [24] V. N. Aseev *et al.*, Proc. PAC05, 2053-2055 (2005).
- [25] J. Qiang *et al.*, J. Comp. Phys. **163**, 434 (2000).
- [26] W. Warnock, SLAC-PUB-5375, 1990.
- [27] R. Walker, private communication.
- [28] M. Sánchez del Río and R. Dejus, SPIE Proc. **3448** (1998).
- [29] A. Zholents *et al.*, NIM A **425** 385-389 (1999).
- [30] M. Borland, Phys. Rev. ST Accel. Beams **8**, 074001 (2005).
- [31] M. Borland and V. Sajaev, Proc. PAC05, 3886-3888 (2005).
- [32] M. Borland *et al.*, Proc. PAC07, 3429-3431 (2007).
- [33] R. Bartolini *et al.*, Proc. EPAC06, 160-162 (2006).
- [34] W. Guo *et al.*, Phys. Rev. ST Accel. Beams **10**, 020701 (2007).
- [35] G. Decker *et al.*, Phys. Rev. ST Accel. Beams **9**, 120702 (2006).
- [36] X. Huang, Proc. PAC07, 1305-1307 (2007).
- [37] M. Borland *et al.*, Proc. PAC09, TH6PFP062, to be published.
- [38] A. Xiao *et al.*, Proc. PAC07, 3447-3449 (2007).
- [39] V. Sajaev *et al.*, Proc. PAC07, 1139-1141 (2007).
- [40] M. Borland, www.aps.anl.gov/asd/oag/presentations/borland-2008-06-03.pdf
- [41] W. Guo *et al.*, Proc. PAC09, TU5RFP008, to be published.
- [42] I. Pinayev *et al.*, Proc. PAC07, 1350-1352 (2007).
- [43] S. Krinsky *et al.*, Proc. PAC07, 1344-1346 (2007).
- [44] R. Hajima and R. Nagai, Proc. ERL07, 133-138 (2007).
- [45] R. A. Bosch *et al.*, Proc. PAC07, 1287-1289 (2007).
- [46] D. Angal-Kalanin, ERL09 Workshop, WG2, 138 (2009).
- [47] N. Nakamura *et al.*, Proc. PAC09, TU5RFP084, to be published.
- [48] M. Borland *et al.*, Proc. PAC07, 1121-1123.
- [49] M. Borland *et al.*, Proc. PAC09, MO3PBI01, to be published.
- [50] M. Borland, Proc. Linac08, 444-446 (2009).
- [51] V. Sajaev, Proc. PAC09, TH6PFP007, to be published.
- [52] V. Sajaev and M. Borland, Proc. Linac08, 447-449 (2009).
- [53] P. Emma *et al.*, Proc. PAC09, TH3PBI01, to be published.
- [54] H. Loos *et al.*, Proc. Linac08, 1095-1099 (2009).
- [55] K. L. Bane *et al.*, Phys. Rev. ST Accel. Beams **12**, 030704 (2009).
- [56] Y. Ding *et al.*, Proc. PAC09, WE5RFP040, to be published.
- [57] D. Xiang and G. Stupakov, Phys. Rev. ST Accel. Beams **12**, 030702 (2009).
- [58] S. Spampinati *et al.*, Proc. FEL07, 362-365 (2007).
- [59] S. Spampinati *et al.*, Proc. FEL08, 342-344 (2008).
- [60] S. Di Mitri *et al.*, Proc. EPAC08, 2993-2995 (2008).
- [61] P. Craievich *et al.*, NIM A **604** (2009) 457-465.
- [62] P. Craievich *et al.*, Proc. EPAC08, 1607-1609 (2008).
- [63] P. Craievich *et al.*, Proc. DIPAC07, 168-170 (2007).
- [64] M. Borland, Phys. Rev. ST Accel. Beams **11**, 030701 (2008).
- [65] E. Kim and M. Yoon, Proc. FEL08, 125-127 (2008).
- [66] P. Williams *et al.*, Proc. EPAC08, 145-147 (2008).
- [67] C. Christou *et al.*, Proc. Linac08, 462-464 (2009).
- [68] C. Vaccarezza *et al.*, Proc. EPAC06, 107-109 (2006).
- [69] S. Thorin *et al.*, Proc. EPAC08, 139-141 (2008).
- [70] Y. Kim *et al.*, Proc. Linac08, 483-485 (2009).
- [71] M. Dunning *et al.*, Proc. PAC07, 1257-1259 (2007).
- [72] A. Zholents *et al.*, Proc. Linac08, 501-503 (2009).
- [73] E. Prat *et al.*, Proc. EPAC06, 1951-1953 (2006).
- [74] M. Abo-Bakr *et al.*, Proc. FEL07, 220-223 (2007).
- [75] H. Ohgaki *et al.*, Proc. FEL07, 390-393 (2007).
- [76] Z. Liu, Proc. FEL06, 277-280 (2006).
- [77] F. Zhou *et al.*, Proc. PAC07, 3127-3129 (2007).
- [78] A. Drozhdin *et al.*, Proc. EPAC08, 2871-2873 (2008).
- [79] W. Gai *et al.*, Proc. PAC09, WE6PFP068, to be published.
- [80] E. Kim, Proc. PAC07, 2814-281 (2007).
- [81] A. Xiao and L. Emery, Proc. PAC07, 3450-3452 (2007).
- [82] D. Sagan *et al.*, Phys. Rev. ST Accel. Beams **12**, 040703 (2009).
- [83] J. Safranek, NIM A **388**, 27 (1997).
- [84] www.aps.anl.gov/asd/oag/manuals/elegant_latest/elegant.html
- [85] https://www.aps.anl.gov/Accelerator_Systems_Division/Operations_Analysis/phpBB3/

UPDATE ON MAD-X AND FUTURE PLANS

F. Schmidt (MAD-X custodian*), CERN, Geneva, Switzerland

Abstract

After a intense and hectic code development during the LHC design phase the MAD-X [1] program (Methodical Accelerator Design – Version X) is going through a period of code consolidation. To this end the development on the core has been frozen and most efforts are concerned with a solid debugging in view of a trustworthy production version for the LHC commissioning. On the other hand, the demand on further code development from the LHC pre-accelerators and CLIC are dealt with PTC [2] related parts of the code where the implementation is in full swing. Having reached a mature state of the code the question arises what kind of future can be envisaged for MAD-X.

INTRODUCTION

During the year 2000 the MAD-X project had been started and a first version had been released in June 2002 [1]. Since then all parts of the code have been adapted frequently to deal with constantly changing requirements needed for the LHC design. Since some time now we have shifted our focus to the upcoming LHC commissioning. Since MAD-X will be playing a relevant role in the LHC operation.

Module	Description	Keeper
MAD-X C Core	Maintenance & Debug	H. Grote
APERATURE [3]	Modeling LHC Aperture	J.B. Jeanneret
C6T [4]	SixTrack Converter	F. Schmidt
CORORBIT	Orbit Correction	W. Herr
DYNAP	Tracking Postproc.	F. Zimmermann
EMIT	Emittance, Radiation	R. Tomas
ERROR	Error Assignment	W. Herr
IBS	Intra-Beam Scattering	F. Zimmermann
MAKETHIN	Thin lens Converter	H. Burkhardt
MATCH	Matching Procedures	E. Laface
PLOT	Mad-X Plotting	R. de Maria (interim)
PTC [2]	PTC proper	É Forest KEK
PTC_NORMAL	Normal Form Coeff.	F. Schmidt
PTC_TRACK	Thick lens Lattice Track	V. Kapin ITP (RU)
PTC_TWISS	Ripken Optics Para.	J.L. Nougaret
SODD [5]	Resonance Comp.	F. Schmidt
SURVEY	Machine Survey	F. Tekker
SXF [6]	Stand. eXchange Format	N. Malitsky BNL
TOUSCHEK	Touschek Effect	C. Milardi IFNL/LNF F. Zimmermann
TWISS	Classical Optics Para.	–
THINTRACK	Thin lens Lattice Track	Y. Sun

Table 1: Module Keepers, People in RED are collaborators from outside CERN, corresponding laboratories in BLUE.

*MAD-X Module keepers see Tab.1

During the years 2008/2009 we have been working on a consolidation of the code such that we can provide a trustworthy MAD-X production version for the LHC commissioning. In particular, since we are presently developing an on-line model [7] of LHC that is based on MAD-X.

On the other hand, there are new demands from other applications notably from the LHC upgrade program, the studies on the LHC pre-accelerators and the CLIC project. Presently, these demands concentrate on better performing modules based on PTC. These developments have been continued and they are not interfering with the consolidation effort for the LHC commissioning.

In this report the key features of the MAD-X are reviewed followed by the most relevant recent MAD-X highlights. The consolidation phase was used to further professionalize the code maintenance which is dealt with in some detail in the next chapter. Besides several outdated features of MAD-X, due to the fact that it is a successor to MAD8, we had to use shortcuts to allow for a gluing of MAD-X with PTC. These shortcuts cannot easily be undone and pose limitations for a further development of the code as a whole. In particular since we will need MAD-X proper in its present state for the LHC commissioning effort. This report closes with a discussion of how one could modernize the code and achieve a better integration of PTC with MAD-X.

KEY FEATURES

The status of MAD-X can best be understood by looking at the design goals of MAD-X followed by a description of what PTC is and how it is connected to MAD-X. From the start we have decided to distribute the work of the code development and maintenance to a large group of module keepers (see Tab. 1) with a code custodian orchestrating their efforts. The job of the custodian is to direct the development of the code and to guarantee its integrity with the goal to provide a code that can handle all requirements for the LHC design and the commissioning alike.

Design Goals

The task at the time was to provide a code in a very short period to allow the design work for the LHC. No grand scheme could be attempted instead we made use of most of the well debugged MAD8 source code (Fortran77) including the traditional MAD8 strengths like sequence editing, matching, plotting, closed-orbit and error routines. However, MAD8 also included out-dated techniques and some feature were flawed if not plain wrong. Those features had to be eliminated: e.g. thicklens non-symplectic tracking

was replaced by a symplectic thintrack module. Another major step has been the replacement of Lie-Algebra by the more flexible Differential Algebra Map [8] and Normal Form techniques [9]. Memory management via C code was used to replace CERN-made Zebra (Fortran77) and the concept of link-lists for the elements in an accelerator have been adopted. A better modular structure of the code was attempted with C interfaces to isolated modules in Fortran or C. The input language (parser in C) was chosen close to MAD9 language and includes FLOW statements like (while, if, macro etc). Deferred expressions are invoked by the “:=” symbol rather than the traditional one “=”. Moreover, a complete pattern matching has been implemented.

Through out all the MAD-X modules output data are written to self-described TFS Format with a two-way TFS-SDDS [10] converter, both formats are self-described data sets. For a complex code like MAD-X the version control system CVS has been an essential tool mandatory for a sound program evolution in particular when dealing with a large number of contributors.

The documentation is presented via the MAD-X web page (google search: madx cern). The user guide can be found via a HTML page. A keyword search and a PDF manual is supplied as well. On the MAD-X web page one can find supporting reports and minutes of the MAD-X meetings. One can download the source code and executables. Moreover, all examples can be found collected in the madx-examples CVS repository.

There have been 2 major objectives: a usable MAD-X code needed for the LHC design which has been delivered by July 2002. After years of development MAD-X has now been prepared for the LHC commissioning and the LHC on-line model for the end of 2009. It should also be mentioned that MAD-X also has to serve all LHC pre-accelerators. To this end a CVS repository CAMIF (Cern Accelerator Mad-X Input Files) has been worked out by CERN staff for all Cern machines as of the PS booster till the LHC and including the transfer lines. Last but not least MAD-X has become a major design code for CLIC in particular with the PTC extension.

PTC

The single most important change to MAD8 is the inclusion of PTC of É Forest [2] as an integral part of MAD-X. The code PTC is composed of two distinct parts (in Etienne’s words):

- A) The independent polymorphic library FPP [8, 11] which handles the operations on polymorphs. Polymorphs are capable of transforming real numbers into Taylor at execution. FPP also contains all the critical operation on Taylor maps: Normal Forms, various factorizations, etc
- B) The code PTC proper is at heart a symplectic integrator with classical radiation added on top. It also moves around quadratic stochastic beam envelopes.

Computer Codes (Design, Simulation, Field Calculation)

However PTC is a physics engine only. It offers hardly any tool of convenience in itself, except when absolutely mandatory like a closed orbit searcher. PTC remains É Forest’s independent (Fortran90) program and is linked to MAD-X via a converter. All modules based on PTC (also in Fortran90) have to be provided by the MAD-X team, e.g. ptc_track (symplectic thicklens), ptc_twiss and ptc_normal (responsibilities see Tab. 1). The design goal for the integration of the two codes has been: all accelerator elements of MAD-X should be treated in PTC and agree with MAD-X as long as the physics is correct with the added value that the elements in PTC are correctly described for any momentum deviation. Special emphasis is on a proper description of RBEND/SBEND, only the later is truly an element of MAD-X. This concept may best be illustrated by an example. At the PAC03 [12] it was demonstrated that MAD (green dots) cannot handle delta-p dependence very well. For this cyclotron I have prepared a MAD-X run both executing a TWISS and PTC_TWISS command. The latter (blue dots) agrees perfectly with the theoretical prediction (red line).

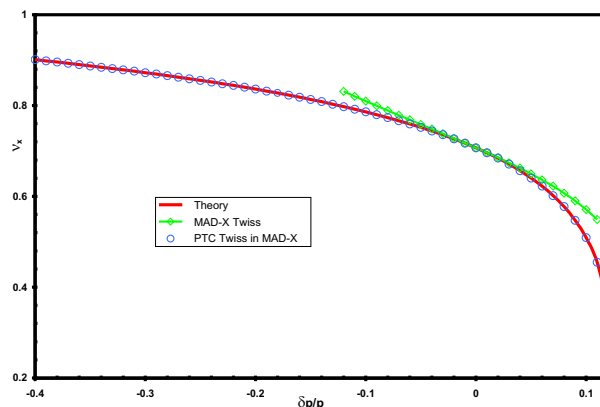


Figure 1: Simple cyclotron for MAD-X and PTC in comparison with theory.

This result is not a surprise to us but the very reason why we have complemented MAD-X with PTC. Please note, that the delta-p in this example is very large contrary to the LHC where it is so small that TWISS is quite sufficient for our purposes.

HIGHLIGHTS

MAD-X General Status

In 2008/2009 the MAD-X team has used the time for a consolidation of the MAD-X code: except for cosmetic changes the development of the core code has been stopped. Most efforts have been concentrated on a rigorous debugging (presently, September 2009, still on-going). In this mature state a major effort has been launched to maximize performance (culprits have been: table access and

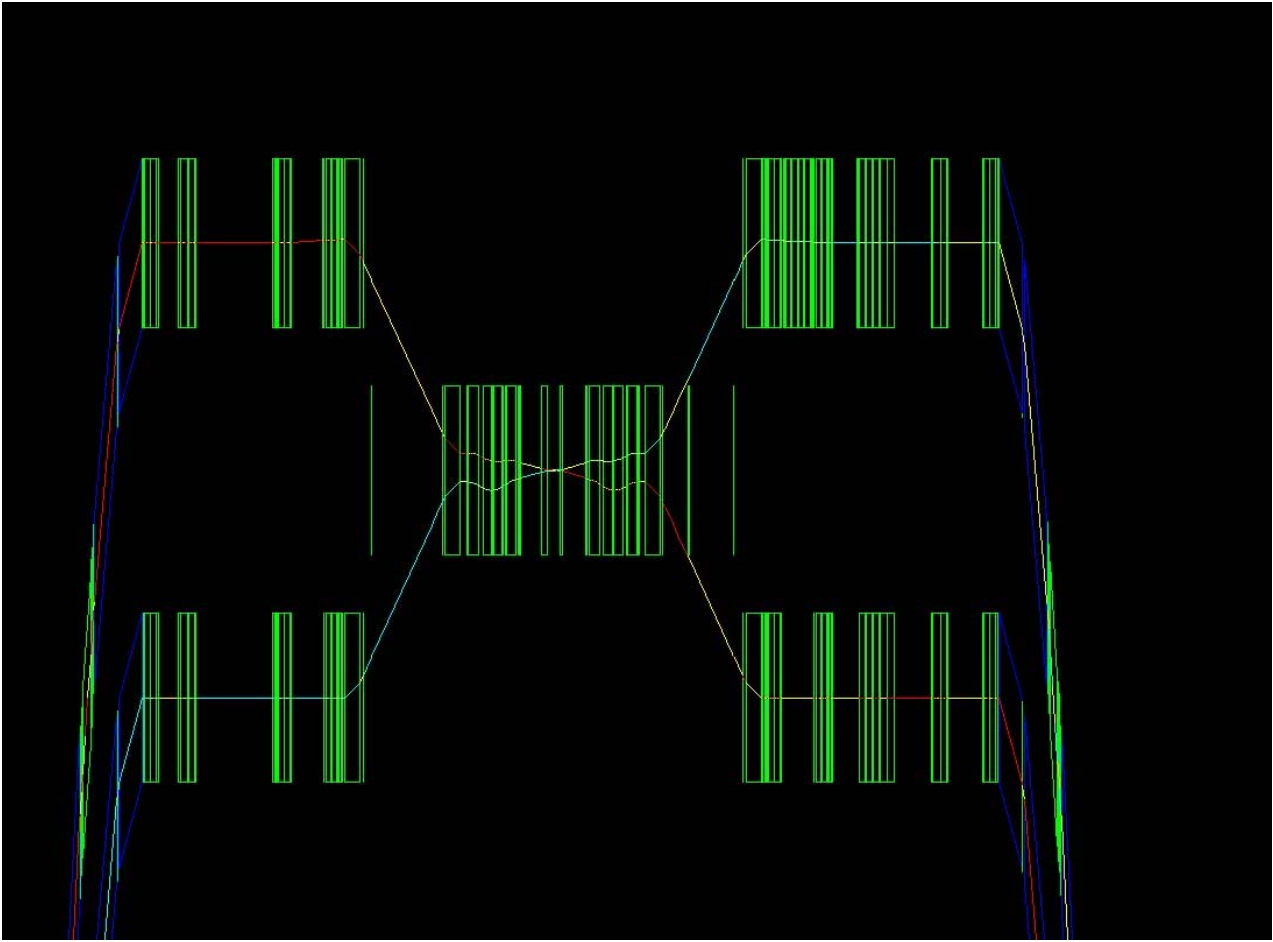


Figure 2: High Beta LHC Interaction Zone in the stand-alone PTC model.

database searches). As a result speed-up factors of 3-5 have been reported in LHC matching. One major change has been the suppressing of the highly controversial individual BV flag for separation dipoles in the common area used to distinguish between beams going in either direction. Instead, we did change the LHC sequence file for the counter-clockwise beam (BEAM2). After this special effort we are confident that our goal has been reached to provide a trustworthy MAD-X program with all needed features for the upcoming LHC commissioning.

It should also be mentioned that MAD-X has become one of the codes that can communicate accelerator input between those codes via AML and the Universal Accelerator Parser [13].

Module Development

During the MAD-X consolidation phase the status of the original modules have been reviewed. The goal was to have all modules ready for the LHC commissioning.

Aperture considerations will be very relevant in the early stages of the LHC commissioning. To this end the aperture module was modified various times, both to add new fea-

Computer Codes (Design, Simulation, Field Calculation)

tures and to clean-up the coding.

Studies for CLIC has revealed that the standard TWISS module suffered from various shortcomings with respect to chromatic functions and delta-p dependence in general. A full solution of all these problems would require a re-write of the code for which we simply lack the manpower. Therefore, we decided to fix the worst problems, in particular some improvement for the chromatic functions and a numerical calculation of the chromaticity which was incorrect in presence of linear coupling. Unfortunately, we cannot simply replace this module with `ptc_twiss` since computing speed is relevant for large machines like LHC in particular in matching.

For CLIC studies at very large delta-p dependence there is no choice but using `ptc_twiss` which delivers the desired results under any condition. We have therefore assigned a module keeper to this module. The goals are:

- a) Provide the same elaborate information as the standard TWISS module.
- b) Adding additional parameters, in particular dispersion and momentum compaction in second or higher order both for rings and lines alike.

Unfortunately, the thintrack modules had relied on the results of the TWISS module. This had to be changed since these 2 modules are based on different Hamiltonian descriptions. The thintrack module is now fully self-contained.

Presently, we are just missing a few more module features: the noise element will be added and the treatment of radiation will be improved both in the thintrack module.

PTC Advancements

Up to now PTC was based on one DA package by M. Berz [14]. Recently, Etienne got in contact with Lingyun Yang [15]. His DA package is written in C++ requiring an elaborate interface in PTC. Now both DA packages co-exist in PTC. Results are identical and the C++ package is even slightly better in performance. It should also be noted that this new package is optimized for use with many variables which still needs to be fully exploited.

The combination of PTC with ORBIT [16], has been recently achieved and successfully applied for the J-PARC Main Ring to fight emittance growth by a skew resonance compensation [17].

Another major development has been the addition of spin routines to PTC. Applying it to MIT-Bates South Hall Ring [18] it could be demonstrated that spin resonances as calculated by PTC agree perfectly with the prediction of the linear response theory [19].

MAD-X in the LHC Control System

The LHC control system LSA [20] uses TWISS TFS output files from MAD-X for their controls during all the LHC cycle.

Similarly relevant is the LHC on-line model [7] to safeguard LHC operation. This on-line model is based almost exclusively on the MAD-X code. This system is being developed in close collaboration of the operation and accelerator physics groups at CERN, the OP and ABP group respectively.

Lastly, we are in the process of constructing a full 3D model of the LHC with the PTC code in stand-alone mode. A MAD-X run provides pieces of the machine that are placed in the virtual tunnel according to SURVEY information. The harmonics to high order are added to all thick elements, magnet assemblies are misaligned, and separately the magnets within them. A complete aperture model around the ring is also taken into account. In Fig. 2 one finds a blow up of one of the high beta interaction zones. By constructions this is really a single structure and the light blue and red tracks are particles being tracked in both directions. In the common area these particles are steered via crossing and separation schemes respectively and are brought into collision at the interaction point quite like in the real world.

Computer Codes (Design, Simulation, Field Calculation)

CODE MAINTENANCE

Integrity

The code development set-up with a sizable number of module keepers requires a more rigorous testing to ensure code integrity. To this end we have set up an automatic system to test several examples for every MAD-X module. These tests are started automatically once a MAD-X CVS tag has been set. The tests are performed 3 times with the standard executable and 2 specially compiled executables. This is needed to find hidden run-time errors, un-initialized variables, out-of-bound arrays etc. In case of problems the module keepers are alarmed and asked to correct code and/or the examples, which are kept in a separate madx-examples CVS repository. We have adopted the concept to provide a production and a debug version, the latter being the latest and greatest but at the risk of the user. A production version for LHC commissioning requires a successful passing of the complete example testing.

Portability

The next relevant design goal is portability across the 3 relevant platforms LINUX, Windows and MAC-OSX. We have one responsible for each platform. Particular worrisome is the fact that we are dealing with a multi-language code which now also includes a C++ part. With the advent of free GNU Fortran compilers (g95 and gfortran) and several commercial ones (Intel ifort, Lahey If95, Nag f95) we could transform the Fortran77 to Fortran90 and provide just one executable which combines MAD-X and PTC. Presently, the ifort compiler has become our workhorse since it has worked effortlessly on Linux and Windows (available for MAC) and it offers a significant performance boost compared to the other compilers.

Special Issues

Separate output buffers for combined Fortran/C code may lead to garbled-up output. This problem is taken care of by automatic python scripts that modify the code prior to compilation such that the corresponding buffer is flushed whenever C calls F or the reverse. To allow for 64 bit compilation the C part of the code needed some fixing. For portability reasons we provide just the 32 bit executables. Tools have been provided to facilitate memory leak detection. In regular intervals we have to check both Fortran and C parts and take action if too many leaks have been found. For portability on the various Linux distributions we need statically linked executables. Unfortunately, the rapid Linux evolution requires changes to the Makefile for every new Linux distribution. The link step will therefore require fiddling in most cases that have not yet been covered.

LIMITATIONS

- The MAD-X parser in C is by now very complex and difficult to modify.
- The TWISS module has reached its end-of-life status. It exists for performance reasons only.
- The PLOT module is also difficult to maintain.

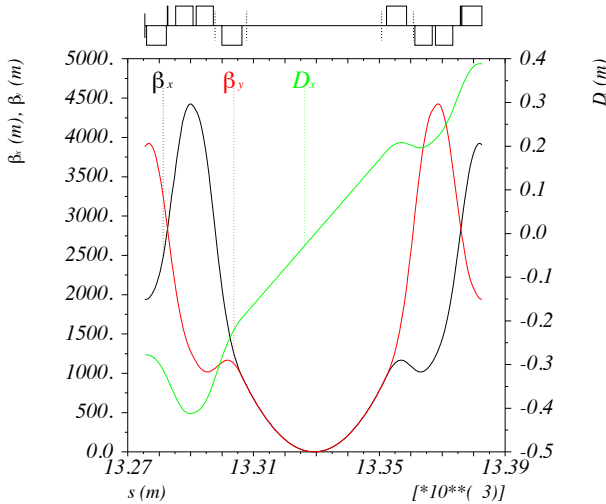


Figure 3: Plot example a high-beta LHC interaction zone.

This can best be explained with an plot example of a high-beta LHC interaction zone. It shows 3 Twiss parameters with 2 vertical axis, one at either side of the graph. Moreover, on the top of the graph the machine elements are symbolically placed at their proper s location in this case of the LHC. These kinds of complex plots are well suited for the design of accelerators. Any replacement of this module would require the exact same functionality. Trouble with the present module is that it is highly complex and difficult to fix, e.g. the invisible subscripts at the vertical axis should be modified. The interactive plot mode on Linux requires the libX11.a library to produce a statically linked executable which requires special installations for each new Linux distribution.

- The documentation is pretty old-fashioned and should be replaced by a professional tool.
- The main problem: MAD-X and PTC have separate structures, e.g. matching (MACROs needed) are extremely inefficient since in each step the PTC structure has to be recreated. In the present multi-language environment there is no clear path of how to overcome this problem.

PLANS

MAD-X has reached a mature status and can be used as is for the LHC commissioning. Minor modifications and bug fixes will continue.

PTC development can happily continue or speed-up to serve smaller machines and CLIC.

A truly integrated code would require a new start, i.e. something like MAD-11.

The design goals of MAD-11 as I see them are:

- MAD-X should be a subset of MAD-11 with exactly the same input language facilities, modules and features except for bugs.
- Parser, plot module and documentation should be build with external, well established and long lifetime tools.
- A single accelerator structure in memory both used for MAD and PTC is clearly needed.
- The big question is if such a goal can be reached in the present situation of a multi-language set-up rather than a monolith in a single computer language.
- Following the experience of the MAD-X project a reasonable estimate of the project duration is 5 years given appropriate resources and manpower. The goal would be to reach the same functionality as MAD-X but with a much better structure of the code.

REFERENCES

- [1] H. Grote and F. Schmidt, PAC03, p. 3497, Portland, USA.
- [2] É Forest, E. McIntosh and F. Schmidt, KEK Report 2002-3.
- [3] I.K. Waarum, Sør-Trøndelag University College, Trondheim, Norway, 2004.
- [4] F. Schmidt, CERN SL/94-56 (AP), (1994) (Rev. July 2008).
- [5] F. Schmidt, CERN SL/Note 99-099 (AP).
- [6] H. Grote, et al., RHIC/AP/155.
- [7] I.V. Agapov et al., PAC07, p. 3384, Albuquerque, USA.
- [8] É Forest, Y. Nogiwa and F. Schmidt, ICAP06, p. 191, Chamonix, France.
- [9] M. Berz, É. Forest and J. Irwin, Part. Accel. 1989, Vol. 24, pp. 91-107.
- [10] M. Borland, ICAP98, pp. 23-27, Monterey, USA.
- [11] based on an idea of J. Bengtsson and on the DA packages of M. Berz, Lingyun Yang or others.
- [12] D. Trbojevic et al., PAC03, p. 3405, Portland, USA.
- [13] D. Sagan et al., EPAC06, p. 2278, Edinburgh, Scotland.
- [14] M. Berz, Part. Accel. 1989, Vol. 24, pp. 109-124.
- [15] Lingyun Yang, this conference.
- [16] É Forest et al., discussed at the 2006 ICFA HB06 workshop.
- [17] É Forest et al., to be published.
- [18] É Forest and S.R. Mane, private communication.
- [19] S.R. Mane, Nucl. Inst. Meth A 601 (2009), pp. 256-263.
- [20] M. Lamont, et al, ICALEPCS07, p. 307, Knoxville, USA.

HIGH-FIDELITY INJECTOR MODELING WITH PARALLEL FINITE ELEMENT 3D ELECTROMAGNETIC PIC CODE PIC3P *

A. Candel[†], A. Kabel, L. Lee, Z. Li, C. Ng, G. Schussman and K. Ko,
SLAC, Menlo Park, CA 94025, U.S.A.

Abstract

SLAC's Advanced Computations Department (ACD) has developed the parallel Finite Element 3D electromagnetic code suite ACE3P for modeling of complex accelerator structures. The Particle-In-Cell module Pic3P was designed for simulations of beam-cavity interactions dominated by space charge effects. Pic3P solves the complete set of Maxwell-Lorentz equations self-consistently and includes space-charge, retardation and boundary effects from first principles. In addition to using conformal, unstructured meshes in combination with higher-order Finite Element methods, Pic3P also uses causal moving window techniques and dynamic load balancing for highly efficient use of computational resources. Operating on workstations and on leadership-class supercomputing facilities, Pic3P allows large-scale modeling of photoinjectors with unprecedented accuracy, aiding the design and operation of next-generation accelerator facilities. Applications include the LCLS RF gun.

THE PARALLEL CODE PIC3P

In Pic3P, the full set of Maxwell's equations is solved numerically in time domain using parallel higher-order Finite Element methods. Electron macro-particles are pushed self-consistently in space charge, wake- and external drive fields.

Finite Element Time-Domain Field Solver

Ampère's and Faraday's laws are combined and integrated over time to yield the inhomogeneous vector wave equation for the time integral of the electric field \mathbf{E} :

$$\left(\varepsilon \frac{\partial^2}{\partial t^2} + \sigma \frac{\partial}{\partial t} + \nabla \times \mu^{-1} \nabla \times \right) \int^t \mathbf{E}(\mathbf{x}, \tau) d\tau = -\mathbf{J}(\mathbf{x}, t), \quad (1)$$

with permittivity ε and permeability μ . The effective conductivity σ provides a simple model for Ohmic losses.

The computational domain is discretized into curved tetrahedral elements and $\int^t \mathbf{E} d\tau$ in Equation (1) is expanded into a set of hierarchical Whitney vector basis func-

tions $\mathbf{N}_i(\mathbf{x})$ up to order p within each element:

$$\int^t \mathbf{E}(\mathbf{x}, \tau) d\tau = \sum_{i=1}^{N_p} e_i(t) \cdot \mathbf{N}_i(\mathbf{x}). \quad (2)$$

For typical simulation runs with second-order elements (curved and using second-order basis functions), $N_2 = 20$. Up to $N_6 = 216$ different basis functions can be used in each element. Tangential continuity between neighboring elements reduces the global number of degrees of freedom, in contrast to discontinuous Galerkin methods.

Substituting Equation (2) into Equation (1), multiplying by a test function and integrating over the computational domain results in a system of linear equations (second-order in time) for the coefficients e_i . Numerical integration is performed with the unconditionally stable implicit Newmark-Beta scheme [1]. More detailed information about the employed methods has been published earlier [2].

Higher-Order Particle-Field Coupling

Electron macro particles are specified by position \mathbf{x} , momentum \mathbf{p} , rest mass m and charge q . The total current density \mathbf{J} in Equation (1) is then approximated as

$$\mathbf{J}(\mathbf{x}, t) = \sum_i q_i \cdot \delta(\mathbf{x} - \mathbf{x}_i(t)) \cdot \mathbf{v}_i(t), \quad (3)$$

for delta-particles with $\mathbf{v} = \frac{\mathbf{p}}{\gamma m}$, $\gamma^2 = 1 + |\frac{\mathbf{p}}{mc}|^2$. The classical relativistic collision-less Newton-Lorentz equations of motion are integrated using the standard Boris pusher [3].

Starting with consistent initial conditions and fulfilling the discrete versions of Equation (1) and the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (4)$$

simultaneously during time integration leads to numerical charge conservation.

The use of higher-order Finite Elements not only significantly improves field accuracy and dispersive properties [4], but also leads to intrinsic higher-order accurate particle-field coupling. For delta-particles, the numerical current deposition involves the (exact) evaluation of line integrals over vector basis functions along the elemental particle trajectory segments for a given time step, using Gaussian quadrature [5].

* Work supported by the U. S. DOE ASCR, BES, and HEP Divisions under contract No. DE-AC002-76SF00515.

[†] candel@slac.stanford.edu

Code Performance Optimization

Pic3P has been optimized for high efficiency both on workstations as well as leadership-class supercomputers. In the following, two specific methods to optimize the parallel performance of Pic3P are introduced, i.e. the causal moving window technique and dynamic parallel load balancing.

Causal Moving Window Field computations can be restricted to the causality region around the particle bunch without any loss of accuracy, assuming the (approximate) trajectory of the particles through the gun is known. The “moving window” feature of Pic3P makes use of adaptive p -refinement where the Finite Element basis function order can be specified for every element, including the omission of certain elements. This can save orders of magnitude in memory and CPU time requirements and allows full EM 3D PIC simulations on desktop computers [6].

Dynamic Load Balancing Pic3P is designed to scale to large problem sizes, and both the fields and the particles are partitioned using a specialized load balancing scheme. As in ACE3P’s time-domain code T3P, the computational Finite Element mesh and the corresponding field degrees of freedom are typically partitioned using graph-based methods (e.g., with ParMETIS) [7]. The macro-particles are typically partitioned using a geometric “recursive coordinate bisection” method (e.g., with Zoltan). Figure 1 shows a typical example of the field and particle partitioning. The absence of explicit locality of particle and field data

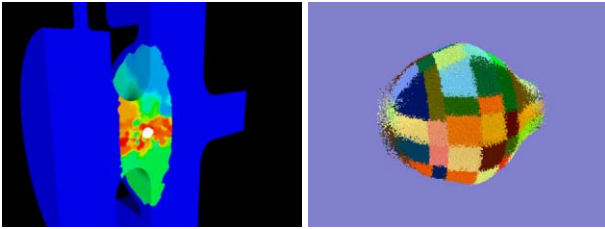


Figure 1: Parallel dynamic load balancing in Pic3P for a typical LCLS RF gun PIC simulation. Partitioning of fields (left) and particles (right) onto different processors is indicated by colors. On the left, the particle bunch is shown in white while the omitted field region from the causal moving window scheme is shown in blue.

necessitates a sophisticated parallel particle-field coupling scheme, which is outlined in the following. First, the geometric domain (such as a collection of bounding boxes) containing all particles is collectively obtained by all processors. Second, every process determines whether it owns or needs fields for the particle region for the next particle pushing step. Then, MPI communicators are constructed accordingly, connecting sub-groups of involved processes, and the “PIC Mesh”, which consists of tetrahedral elements and corresponding field coefficients, is collected with collective MPI operations such as Allgatherv. Finally, the Computer Codes (Design, Simulation, Field Calculation)

particles are pushed using the (now locally available) PIC Mesh information. The current deposition “scatter” step is done similarly.

By segmenting the particle region into fewer or more pieces, the memory and time requirements during the communication steps can be traded, such that both strong and weak optimal scalability can be achieved. The execution order of PIC Mesh updates for different communicators is optimized and re-ordered such that disjoint sub-groups of processes can communicate simultaneously. By slightly enlarging the bounding box around the particles (by extrapolating the trajectories), a new PIC Mesh only needs to be created from time to time (e.g., every 50 time steps), and re-using the same objects and communicators leads to significant savings in computational resources, as only the changing degrees of freedom for electromagnetic fields and particle currents need to be communicated.

This dynamic load balancing scheme allows the solution of large problems with hundreds of millions of field degrees of freedom (DOFs) and billions of particles and enables unprecedented accuracy in self-consistent state-of-the-art simulations of beam-cavity interactions on modern supercomputers. Table 1 shows typical runtime parameters for LCLS RF gun PIC simulations performed with Pic3P on the Jaguar Cray XT5 machine at NCCS. For small prob-

Table 1: Typical runtime parameters on a Cray XT5 for LCLS RF gun emittance calculations with Pic3P. For the first case, causal moving windowing and a direct solver were used. For the other cases, full-domain calculations were performed with an iterative field solver (i.e., CG with incomplete Cholesky preconditioner).

CPU's	Particles	Elements	DOFs	Walltime/step
12	100k	305k	53k	0.23 secs
120	1M	305k	2M	1.3 secs
1200	10M	2.8M	17.7M	2.4 secs
12,000	100M	23M	142M	24 secs

lem sizes (i.e., up to a few million DOFs, and less than 100 CPUs), the use of direct factorization methods for the field solution is most efficient. For larger problems, iterative solvers are used, since they display much better scalability. For typical problem sizes, the performance of the memory subsystem is most important. Due to the small relative size of the particle bunch compared to the overall field domain in this LCLS RF gun application, the use of more than a few thousand CPUs is not efficient in the current implementation, as the global PIC Mesh construction and communication begins to strain the network subsystem. However, the current scheme could be easily modified to only use a subset of the processors for particle pushing. This should lead to better scalability for extremely large problems with highly localized particle distributions, similar to the good scalability experienced for small to mid-size problems.

VALIDATION: LCLS RF GUN

Benchmark PIC simulations of the 1.6-cell S-band LCLS RF gun are presented in the following [8]. Simulation parameters are: π -mode, 120 MV/m, 1 nC, 10 ps, 1 mm beer-can initial bunch distribution, centroid injection phase -58° and no solenoid. These parameters allow comparisons between the 3D results from simulations with Pic3P and PARMELA and the 2D results from simulations with Pic2P and MAFIA.

Figure 2 shows a comparison of transverse emittance results by the different codes. For Pic3P simulations, a

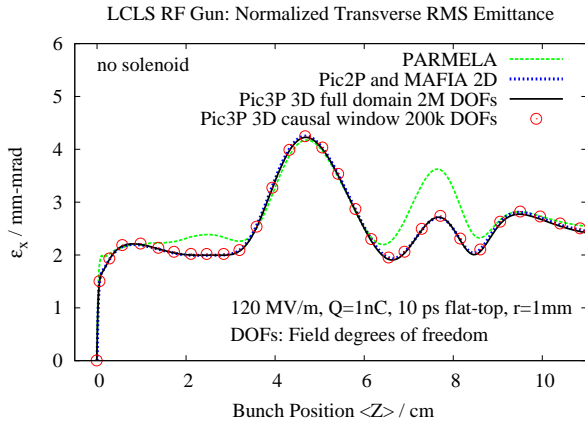


Figure 2: Comparison of normalized transverse RMS emittance as a function of beam position in the LCLS RF gun as calculated with PARMELA, Pic2P and MAFIA 2D (both agree), and Pic3P, where the causal moving window technique reduces the problem size by one order of magnitude.

conformal, unstructured 3D (1/4) mesh model with only 305k tetrahedral elements is sufficient to reach convergence, with mesh refinement along the center of the beam pipe. High fidelity cavity mode fields are obtained with the parallel Finite Element frequency domain code Omega3P and directly loaded into Pic3P as drive fields.

Excellent agreement between 3D results from Pic3P and the 2D results from Pic2P is found, as expected from the high cylindrical symmetry in the fields, as well as perfect agreement with MAFIA 2D as expected from the convergence behavior of the codes. PARMELA results differ as space-charge effects are significant, presumably because wakefield and retardation effects are ignored, as detailed in a previous study [2]. Simulation results starting from a measured initial bunch distribution have been published earlier [6].

SUMMARY

SLAC has developed the first parallel higher-order Finite Element 3D PIC code Pic3P, for realistic modeling of space-charge dominated beam-cavity interactions. As a part of the ACE3P code suite, Pic3P uses state-of-the-art parallel Finite Element methods on conformal, unstructured meshes. Computer Codes (Design, Simulation, Field Calculation)

Pic3P has been optimized for large-scale modeling on modern supercomputers and causal moving window techniques and dynamic load balancing enable high-fidelity simulations of electron injectors with unprecedented efficiency and accuracy. Applications include emittance calculations for the LCLS RF gun. Benchmark simulations show excellent agreement between Pic3P and MAFIA TS2 results, while results obtained with the electrostatic code PARMELA show some differences, presumably because wakefields and retardation effects are omitted.

ACKNOWLEDGMENTS

This work was supported by the US DOE ASCR, BES, and HEP Divisions under contract No. DE-AC002-76SF00515. This research used resources of the National Energy Research Scientific Computing Center, and of the National Center for Computational Sciences at Oak Ridge National Laboratory, which are supported by the Office of Science of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231 and No. DE-AC05-00OR22725. – We also acknowledge the contributions from our SciDAC collaborators in numerous areas of computational science.

REFERENCES

- [1] N. M. Newmark, “A method of computation for structural dynamics”, Journal of Eng. Mech. Div., ASCE, vol. 85, pp. 67-94, July 1959.
- [2] A. Candel et al., “Parallel Higher-order Finite Element Method for Accurate Field Computations in Wakefield and PIC Simulations”, Proc. ICAP 2006, Chamonix Mont-Blanc, France, Oct. 2-6, 2006.
- [3] J. P. Boris, “Relativistic plasma simulation-optimization of a hybrid code”, Proc. Fourth Conf. Num. Sim. Plasmas, Naval Res. Lab, Wash. D.C., pp. 3-67, Nov. 2-3, 1970.
- [4] M. Ainsworth, “Dispersive properties of high-order Nedelec/edge element approximation of the time-harmonic Maxwell equations”, Philos. trans.-Royal Soc., Math. phys. eng. sci., vol. 362, no. 1816, pp. 471-492, 2004.
- [5] J. W. Eastwood, W. Arter, N. J. Brealey, R. W. Hockney, “Body-fitted electromagnetic PIC software for use on parallel computers”, Computer Physics Communications, vol. 87, Issues 1-2, Particle Simulation Methods, pp. 155-178, May 2, 1995.
- [6] A. Candel et al, “High-Fidelity RF Gun Simulations with the Parallel 3D Finite Element Particle-In-Cell Code Pic3P”, AIP CP1149, Proc. Workshop on Polarized Electron Sources, pp. 1114-1118, Oct 1-3, 2008
- [7] A. Candel et al., “State of the art in electromagnetic modeling for the Compact Linear Collider”, Proc. J. Phys.: Conf. Ser., vol. 180, pp. 012004 (10pp), 2009.
- [8] L. Xiao et al., “Dual Feed RF Gun Design for the LCLS”, Proc. PAC 2005, Knoxville, Tennessee, May 15-20, 2005.

BEAM DYNAMICS IN THE LOW ENERGY PART OF THE LOW EMITTANCE GUN (LEG)

M. Dehler, S. G. Wipf, Paul Scherrer Institut, Switzerland

Abstract

One option for the electron source of the SwissFEL is the Low Emittance Gun (LEG), which is currently under development at PSI. It consists of a pulsed DC gun operating at 500 keV and has the option of using either a photo cathode or a field emitter array. The gun is followed by a pulsed in-vacuum solenoid and a two-frequency cavity, not only used to accelerate the beam but also to create a highly linear energy correlation required for ballistic bunching. All components are rotationally symmetric, so a full particle-in-cell simulation of the setup using 2 1/2 D MAFIA, including space charge, wake fields and beam loading effects, shows the base line performance. Given the relatively low beam energy and high brightness of the beam, there were concerns with respect to the sensitivity to mechanical misalignments in the structure. So we investigated these using the 3D in-house code CAPONE and calculated tolerances, which are well within acceptable limits.

INTRODUCTION

To realize compact X-ray free electron lasers, electron sources with a high brilliance and ultra low emittance are required. The SwissFEL project at PSI is based on the development of such concepts, allowing a substantial reduction in size and cost of such a facility. In order to reach the Angstrom wavelength range, peak currents of 1.5 kA, a relative energy spread of 10^{-4} and normalized transverse slice emittances in the order of 300 nm rad are crucial in the standard operation mode.

Several options are under discussion for the electron source. One consists of an S-band RF gun with a photo cathode[1] running at gradients of 100 MV/m, which will be tested in the 250 MeV injector facility currently under construction. The other, examined here, employs a pulsed diode at an accelerating gradient of 125 MV/m gradient over a four millimeter gap[2], where a prototype is under commissioning at PSI. The baseline scenario assumes a photo-cathode, but cathodes using field emitter arrays may be promising candidates[4, 5].

Even after the high gradient acceleration in the diode, the electron beam is still fragile at energies of 500 keV. A low initial beam current of 5.5 A with an overall charge of 200 pC is beneficial in that respect, but requires a large bunch compression ratio of 270 to obtain a peak current sufficient for lasing. Therefore after the diode, the electron beam is accelerated off-crest in a two-frequency cavity. The fundamental mode at 1.5 GHz is combined with a higher har-

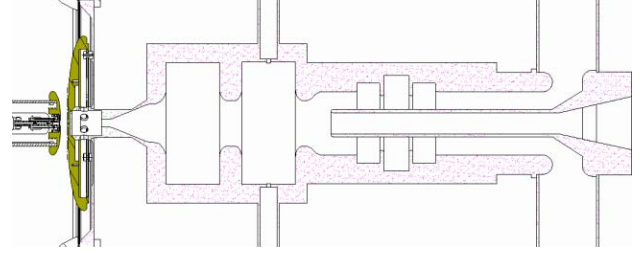


Figure 1: Pulsed DC gun with dual frequency cavity (solenoid not visible)

monic at 4.5 GHz to introduce a highly linear energy chirp to do ballistic bunching in the following drift delivering a peak current of 20 A to the linac. The beam current and phase space at the end of the drift are compatible with those expected from the S-band RF gun option, so that we can use the same linac design for both options. Fig. 1 shows the layout. A pulsed solenoid (not shown in the figure) between diode and two-frequency cavity corrects the residual divergence of the beam after the gun.

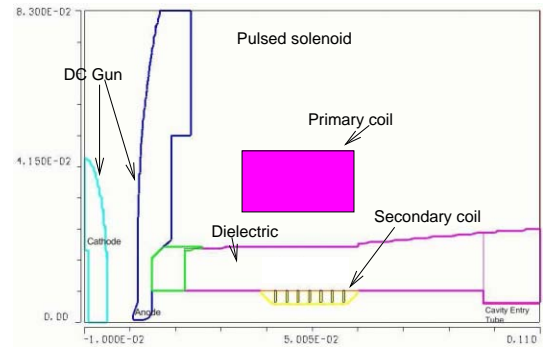


Figure 2: Diode and pulsed solenoid

The simulations have been performed in two steps. Since all elements are rotationally symmetric, the base line performance assuming perfect alignment and ideal beam properties is obtained from a 2½D particle in cell simulation using MAFIA TS2[6]. The influence of misalignment, beam offsets etc. was obtained with the in-house code CAPONE[7] and is described in a separate section.

BASE LINE PERFORMANCE

As was mentioned above, the assembly is rotational symmetric, so it was modeled in two dimensional cylindrical coordinates. The setup consists of the following, the cath-

ode and anode disks, followed by a pulsed solenoid and the dual-frequency cavity. To be able to resolve the space charge of the particles within the thin bunch, the mesh is very fine especially close to the z-axis with a minimum resolution of $33 \mu\text{m}$ and an overall grid size of 250,000 points, at which the full numerical convergence of the results was obtained. The nominal gap between anode and cathode is 4 millimeters. Both were modeled as precise as possible in order to obtain sufficient resolution for the fine detail of the surface shape. The diode voltage is 500 kV giving an average accelerating gradient of 125 MV/m.

The solenoid (Fig. 2) is a corrugated steel insert sitting inside the ceramic beam pipe. The azimuthal magnet current is excited in this insert by magnetic induction via a secondary coil indicated in the figure. While the beam is passing through the magnet, the resulting field can be assumed as constant, though with important differences to the field distribution of a truly static magnet. Where the magneto-static field is determined by the distribution of the driving currents and the permeability, we have the additional influence of induced mirror current on the various conducting parts as e.g. the anode plate or the RF cavity resulting in a shielding effect.

The PIC module in MAFIA can use only static focusing fields, so we employed the following strategy to arrive at a realistic field distribution. The metallic insert corresponding to the winding of the solenoid carries a homogeneous current distribution, something which we can safely assume from the material properties and the time domain behavior of the driving pulse. All metallic parts were specified with a very low permeability of 0.01 to push out the magnetic fields. Furthermore, a tiny shielding current (also homogeneously distributed) was assumed in the anode plate and set to a value minimizing the magnetic field between anode and cathode¹. Figure 3 shows the on-axis distribution of the longitudinal field.

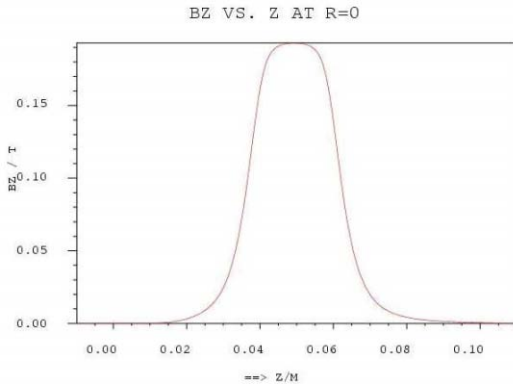


Figure 3: On axis field of pulsed solenoid

With the given parameters (0.5 kV and a 4 mm gap at the pulser), the solenoid needs a focusing field strength of

¹A fully realistic computation could only be done in 3D with the help of a time domain computation, but we still would need special tricks to load the field into the PIC simulation.

$\int B_z^2 dz = 0.92 \cdot 10^{-3} T^2 m$ to correct for the residual divergence of the beam after the DC gun.

The cavity was modeled including the subsequent filter section. The entry of the RF-Cavity fits into the end of the solenoid. The cavity is designed so that the third harmonic RF mode, TM_{012} can be superimposed on the fundamental one, the TM_{010} . Apart from acceleration, the fundamental frequency provides the energy chirp for compression. The higher harmonic serves to linearize the longitudinal phase space. Initial beam and field parameters used in the simulation are listed in table 1.

Table 1: Initial parameters used in simulation

Emittance	70 nm rad
Beam diameter (homogeneous)	600 μm
Energy	$\gamma=1.0003$
Pulse length (flat top)	40 ps
Beam current	5.5 A
Gun voltage	500 kV
Slice length for sliced emittance	1 ps
Solenoid strength	$0.92 \cdot 10^{-3} T^2 m$
Fundamental mode	
Amplitude	35.5 MV/m
Phase/deg.	-15
Harmonic mode	
Amplitude	17 MV/m
Phase/deg.	10

The field values for the solenoid and the RF cavity differ from the official parameter set in the CDR[2] for the following reasons. First, we used the latest structure geometries, which differ from those used in generating the reference case. When trying to find a good beam behavior in the vicinity of the reference value, we saw that the beam behavior is very sensitive to small variations in the input parameters. This is due to the fact that the beam at the entry of the RF structure is simultaneously focused in the transverse plane and sees a deceleration in the order of 30%, so that space charge effects become quite pronounced. We went to a new working point and reduced the phase of the fundamental mode by roughly 20 degrees to $\phi = -15$ degrees and increased the amplitude of the harmonic mode from 11 to 17 MV/m. The pronounced drop in beam energy at the entrance of the cavity is strongly reduced leading to a more stable behavior. The resulting longitudinal phase space obtained by either accelerating only with the fundamental mode or with both is shown in Fig. 4. The combination of both modes results in a nicely correlated energy spread suitable for ballistic compression in the following drift.

The big bump in emittance, as the bunch is traversing the solenoid (Fig. 5), may be irritating, but it is due to the fact that the code (as well as all other known codes) only uses the kinetic transverse momenta to compute emittances. A computation using the generalized momentum $p = mv + qA$ would not show this behavior. Only the

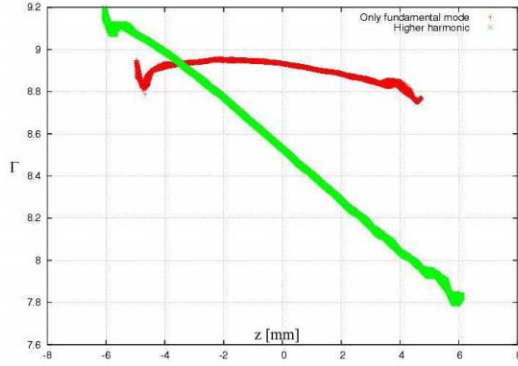


Figure 4: Longitudinal phase space at exit of the dual-frequency cavity ($z = 370\text{mm}$) with/without harmonic mode

emittance values outside the solenoid region are relevant.

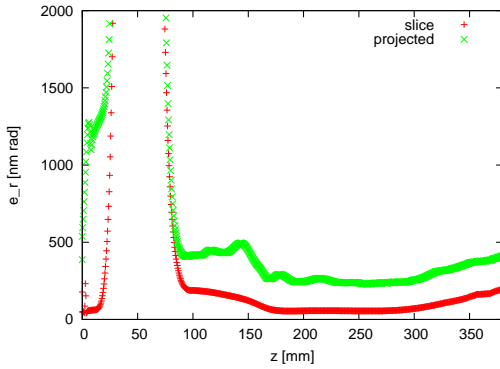


Figure 5: Evolution of slice and projected transverse emittance during flight.

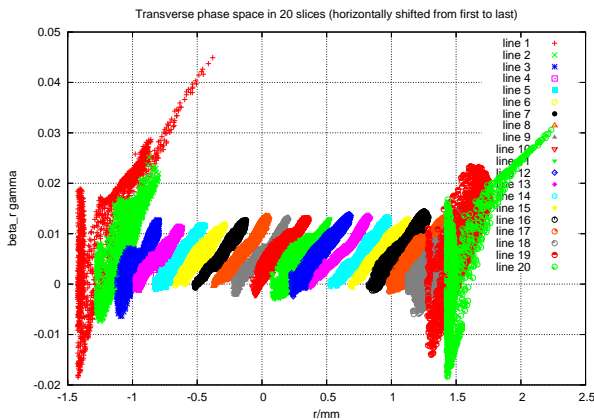


Figure 6: Phase space distribution of individual bunch slices at the exit of the dual-frequency cavity ($z = 370\text{mm}$)

The best view of the transverse phase space distribution is obtained by looking at individual slices in the bunch. In Fig. 6, the distribution of several slices is shown with the leftmost corresponding to the tail slice and the rightmost to the head slice. The distribution shows a clear separation between the slices, indicating a well-defined bunch structure. The plot also shows the evolution of the beam parameters seen after the RF structure.

that of the head of the bunch. We have a very good emittance of the center slices corresponding to values slightly above 180 nm rad. Only the outermost slices see a deterioration due to nonlinear space charge, RF forces as well as wake fields. The angle of the phase space ellipses of the center slices are very well aligned with respect to each other, so that there is no intrinsic emittance increase by subsequent bunch compression stages merging adjacent slices. Table 2 lists the beam parameters seen after the RF structure.

Table 2: Beam parameters at the exit of the RF structure ($z = 370\text{mm}$)

Parameter	Center slice	Projected
$\epsilon_t/\text{nm rad}$	176	378
$\epsilon_z/\text{eV s}$	$1.2 \cdot 10^{-9}$	$3.9 \cdot 10^{-1}$
$\sigma_r/\mu\text{m}$	297	272
$\sigma_{r'}/\text{mrad}$	1.0	1.0
γ/MeV	3.9	3.9
σ_t/ps	0.25	10.8
$\sigma_\gamma/\%$	0.12	4.7

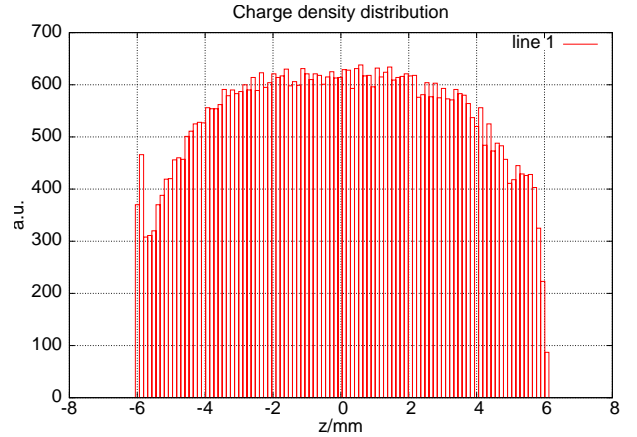


Figure 7: Intra bunch charge density at exit of the dual-frequency cavity ($z = 370\text{mm}$)

Fig. 7 shows the relatively flat charge distribution inside the bunch. The small scale fluctuations seen are due to numerical noise. The peak at the bunch tail ($z = -6\text{mm}$) is a numerical artifact due to the code.

MISALIGNMENTS

Including misalignments as structure offsets or tilts into the simulations destroys the symmetry of the problem, so a full three dimensional simulation has to be performed. To do that, we used the in-house particle-in-cell code CAPONE, which was originally developed to do ultra high resolution of electron sources based on field emission arrays[7]. The original version was restricted to have only static accelerating fields, whereas here we also need to take into account time harmonic fields.

Running an eigenvalue solver on the large grid sizes of an order of a billion mesh cells required to have a consistent field solution for the particle-in-cell algorithm is a rather tedious task – the more, since we need not only the fundamental mode in the RF structure but also the third harmonic. With the typical three dimensional solver, this means computing all modes of all azimuthal orders up to that frequency, so that approximately 100 modes need to be determined.

A rather elegant alternative was to introduce 2D field maps into CAPONE allowing the use of high resolution two dimensional solution. These are not distributed into the three dimensional calculation grid but kept in 2D format and read out there by the force calculation routine. Apart from the easier generation of these fields, this has other significant advantages. For one, the domain spanned by the grid for the particle-in-cell solver does not have to extend over the full volume of e.g. the RF cavities, but can be truncated to the size relevant for a correct representation of space charge fields and wakes. This feature can reduce the size of the particle-in-cell grid already considerably. As a second feature, we can easily introduce misalignments as offsets or tilts by simply moving the 2D field map with respect to the PIC grid.

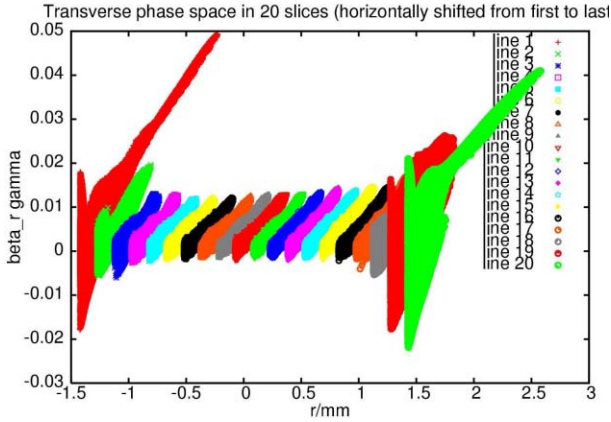


Figure 8: Phase space distribution of bunch slices as calculated with CAPONE ($z = 370\text{mm}$)

The initial conditions for the beam correspond to that of the two dimensional case (Tab. 1) with the exception of the longitudinal profile. In the three dimensional calculation, the flat top profile had a non zero rise and fall time of 0.5 picoseconds. 2D and 3D results agree well, we have normalized emittance values of 388 nm-rad in 2D versus 383 nm-rad in 3D. The one significant difference is seen in the center slice emittance, where we obtain 170 nm-rad in 2D compared to 120 nm-rad in 3D, which is probably due to artifacts in the stochastic distributions used in MAFIA as shown in Fig. 7. Fig. 8 shows the 3D slice emittances at the exit of the cavity.

For the various misalignments, the following cases were computed: The laser spot on the cathode was assumed to be misaligned by $300\text{ }\mu\text{m}$, the solenoid was calculated with RF Guns and Linac Injectors

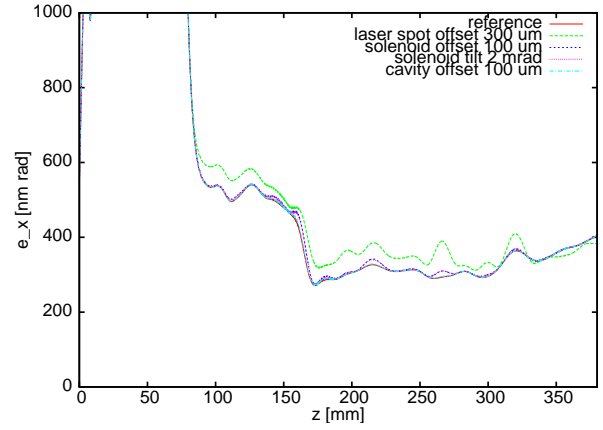


Figure 9: Envelope of the projected transverse emittance for various misalignments.

offset ($100\text{ }\mu\text{m}$) and tilted (2 mrad) and the effect of an offset of the cavity of $100\text{ }\mu\text{m}$ was simulated.

The variation of the projected emittance is relatively minor as can be seen in figure 9. Only the rather extreme offset of the laser spot leads to a visible difference. Similar is the influence on the center slice emittance. The only significant deterioration is visible, if we start with an offset beam. Fig. 11 shows the phase space distribution for the case of a shifted laser spot.

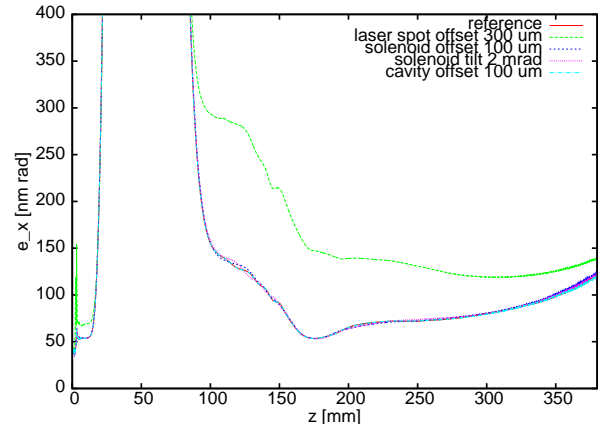


Figure 10: Envelope of the slice emittance in the bunch center for various misalignments.

Does this mean, that misalignment is a non-issue for the setup? In addition to any widening of the phase space volume occupied by the bunch, the kicks introduced by the finite tolerance lead to a displaced of the bunch centroid in the transverse phase space as can be seen in Fig. 12 and Fig. 13 for the mean position and flight angle respectively, which lead to mismatching in the following beam optics.

In principle, one can correct this effect by additional corrector magnets, but one could also try to find tolerances rendering these correctors unnecessary. The required criterion is, that the displacement of the bunch centroid still fits into the maximum allowable phase space volume (aka

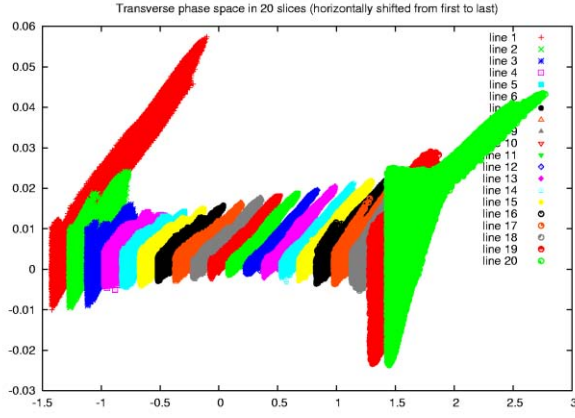


Figure 11: Phase space of bunch slices assuming a shifted laser spot.

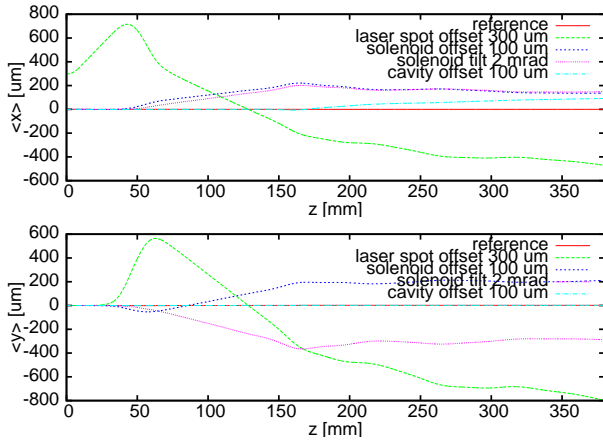


Figure 12: Variation of beam offset along beam axis

projected emittance) of the bunch of roughly 400 nm-rad. Given the geometrical offsets $\langle x \rangle$, $\langle y \rangle$ and momentum offsets $\langle u_x \rangle$ and $\langle u_y \rangle$, we can define a kind of a pseudo emittance

$$\epsilon_p = \sqrt{(\langle x \rangle \langle u_x \rangle)^2 + (\langle y \rangle \langle u_y \rangle)^2},$$

which by definition should be lower than the limit of the projected emittance. Fig. 14 shows envelopes of this pseudo emittance for the different cases of misalignment. ϵ_p scales roughly quadratically with offset or tilt, table 3 uses this relationship to derive alignment tolerances.

Table 3: Alignment tolerances to keep pseudo emittance below 400 nm

Parameter	Tolerance
laser spot offset	45 μm
solenoid offset	70 μm
solenoid tilt	1.3 mrad
cavity offset	140 μm

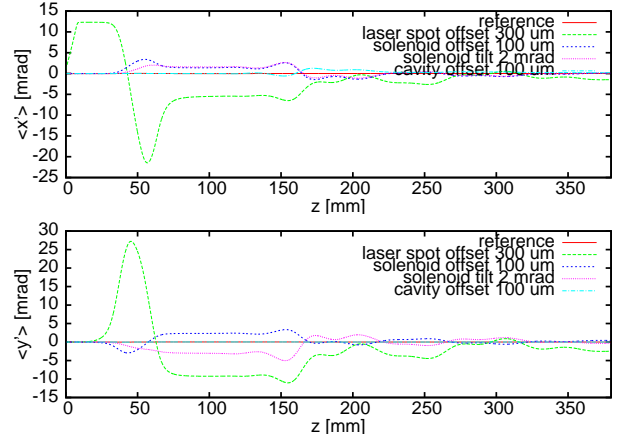
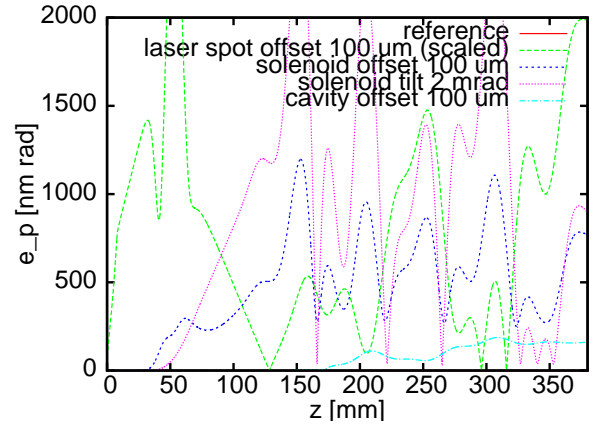


Figure 13: Variation of beam angle along beam axis


 Figure 14: Envelope of the pseudo emittance ϵ_p (definition see main text) for various misalignments.

CONCLUSIONS

A simulation of the Low Emittance Gun (LEG), comprising the pulsed DC source at 500 kV, the pulsed solenoid and the two frequency cavity, has been completed. If one compares the settings and results with those in the official specification[2], the following remarks have to be made. Using the official parameter set, we see (as in up-to-date simulations by Anne Oppelt with tracking codes[8]), that the setup is extremely sensitive to minor variations in phase and amplitude of the cavity modes. At the entrance of the cavity the beam energy drops considerably, this is due to the fact that it is strongly focused, so that space charge forces are quite dominant. Given a run time of the simulation of 6-8 CPU hours and a corresponding limited number of possible iterations, it was not possible to find a satisfactory solution in the vicinity of the official parameters. In the end, a different working point was found, where the beam is more on the crest of the fundamental mode and which has the following advantages: The drop in beam energy at the entrance of the cavity is visibly reduced and the radius and projected emittance at the exit are lowered. If one compares emittance values computed by tracking codes such as

BET (beam envelope tracker)[9] with those of a particle in cell simulation, one sees the following: Slice parameters in the bunch center agree relatively well, but the tracking codes do underestimate the deterioration seen by the head and tail of the bunch, so that the PIC code computes a larger value for the projected emittance.

Using the in-house code CAPONE, the influence of misalignments on the performance has been calculated. Other than the setting of amplitudes and phases for the RF cavity mentioned above, tolerances for the geometrical alignment of the setup are relatively relaxed for the operation.

REFERENCES

- [1] Y. Kim et al., “Linear Accelerator for the PSI-XFEL FEL3 Beamline”, Proc. of the LINAC 2008, Vancouver, Canada (2008)
- [2] A. Oppelt et al., “Towards a low emittance X-ray FEL at PSI”, Proc. of the 2007 Free-Electron Laser Conference (FEL2007), Novosibirsk (Russia)
- [3] J.-Y. Raguin et al., “A two-frequency RF cavity for the PSI-XFEL: Design and beam dynamics simulations”, Nuclear Instruments and Methods in Physics Research A 593 (2008) 125 128
- [4] E. Kirk et al., “Development of All-metal Field Emitter Arrays for high Current Applications”, proc. International Vacuum Electronics Conference, May 15-17, Kitakyushu, Japan (2007)
- [5] M. Dehler et al., “Full scale simulation of a field-emitter arrays based electron source for free electron lasers”, J. Vac. Sci. Technol. B 24 (2), pp. 892-897 (2006)
- [6] <http://www.cst.de>
- [7] A. Candel, M. Dehler, “Massive parallel particle in cell code for the simulation of field emitter based electron sources”, proc. 8th International Computational Accelerator Physics Conference, Jun. 29- Jul. 02, St. Petersburg, Russia (2004), Nucl.Instrum.Meth.A558:154,2006
- [8] A. Oppelt, personal communication
- [9] R. Bakker, PSI in-house development.

THE XAL INFRASTRUCTURE FOR HIGH LEVEL CONTROL ROOM APPLICATIONS

A. Shishlo[#], C. K. Allen, J. Galambos, T. Pelaia, ORNL, Oak Ridge, TN 37831, U.S.A.
C. P. Chu, SLAC, Menlo Park CA.

Abstract

XAL is a Java programming framework for building high-level control applications related to accelerator physics. The structure, details of implementation, and interaction between components, auxiliary XAL packages, and the latest modifications are discussed. A general overview of XAL applications created for the SNS project is presented.

INTRODUCTION

The development of XAL [1] was started in 2001 at the SNS project as a framework for high level accelerator physics applications. The Java programming language was chosen because it addresses the need for a GUI interface, database services, plotting, and numerical simulations. When XAL development first began there was a lack of free mathematical and plotting packages, but the situation has since improved. EPICS has been chosen as a communication protocol. Today the XAL framework consists of the following parts:

- A hardware representation of the machine for connectivity and control.
- A beam simulation model termed the "online model" for model reference and comparison to the hardware operation.
- An application framework to provide a common "look and feel" and functionality for all XAL applications.
- Services that run continuously in the background (24/7), and which can communicate with several XAL applications simultaneously.
- A set of auxiliary mathematics, graphics, and plotting packages.
- The channel access communication library.

In this paper we present descriptions of these parts of XAL and an overview of applications implemented on the base of this framework for the SNS project.

ACCELERATOR MODEL

An accelerator model represents a structural view of an accelerator. According to this model the accelerator consists of ordered accelerator sequences which usually represent accelerator beam lines, and they can have other ordered sub-sequences or nodes corresponding to physical devices. An instance of such a structure is shown on Fig. 1. The lowest level of the accelerator model hierarchy is represented by such components as magnets, BPMs, wire scanners, RF gaps, position markers etc. Usually accelerator nodes correspond to real physical devices, but

it is not necessarily a one-to-one mapping.. For instance, at SNS there are single devices consisting of a quadrupole + dipole windings + BPM strip-lines. We consider these functionalities as three separate accelerator nodes (quad + dipole corrector + BPM), all at the same position.

XAL uses an XML file called an "optics_source" as a natural way to initialize this accelerator hierarchy. This XML file includes all information about sequences, components, positions, parameters, and necessary EPICS's PV names for device signals.

There are two ways to prepare such files. First, it can be done manually from scratch or by modification of an existing file if you are interested in only a relatively small accelerator model for testing XAL features. Second, you can prepare an application that will generate the file for you by using a relational database. Of course, this application will be specific for each accelerator, because accelerator database structures are usually different.

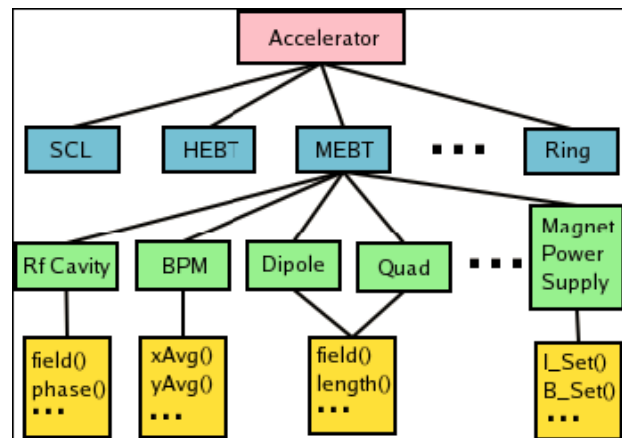


Figure 1: An example of the XAL accelerator model structure.

In the beginning of XAL development, the optics XML file was the only source for the accelerator model initialization, but later several new XML files were added to provide the model with the necessary information. First, there was an XML file with hardware node status information. This is a small file describing availability of certain diagnostics nodes, because they frequently go from the "online" to the "offline" state, and the model should know about a validity of the diagnostic signals. The second new XML file maps an accelerator node type with a particular implementation of this type in the model. This file was introduced to generalize XAL and to use it for different accelerators where the similar devices (i.e. BPM in SNS or J-PARC) can have different functionality. The third one includes information about signals from

[#]shishlo@ornl.gov

systems like a timing system that does not belong to any particular beam line. Finally, there is a XML file with the online model parameters. Today the combination of all these XML files is used to initialize the XAL accelerator model.

After initialization, the model is ready for usage. Examples of common tasks for the model and nodes are:

- Creating a new combo-sequence from existing sequences.
- Getting a position of a node or a sequence inside a parent sequence.
- Selecting nodes of a certain type and properties from a sequence.
- Getting a list of possible EPICS signals from a node.
- Getting default values of parameters of a particular node.

None of the actions listed above need live EPICS connections and therefore they do not require a real machine or virtual accelerator. There is another group of actions which require live EPICS communications. Some of these actions include getting or setting live hardware parameters such as magnetic fields or RF cavity phase and amplitude or reading diagnostic data such as BPM signals. This group provides live interfacing with the accelerator. The details of such EPICS connections and communications are hidden from the user by the simple interface of the model.

The XAL accelerator model is a very useful and convenient control system tool, but by itself it is has nothing to do with accelerator physics. To perform meaningful operations with the accelerator or beam lines we need a physical model. We call this model the “online model”.

XAL ONLINE MODEL

The XAL online accelerator model [2] performs on-the-fly calculations of beam parameters based on machine settings. These settings can be extracted from a live accelerator, from design values, from a combination of these two sources, or they can be modified by the user.

The three main components of the model are an accelerator lattice which is constructed from the accelerator nodes, a probe which describes the beam and how it is to be modeled, and a set of algorithms for probe tracking through different elements of the lattice. The online model implements the Element-Algorithm-Probe design pattern introduced by Malitsky and Talman [3]. This design strategy separates the machine representation from the beam model and the dynamics calculations.

A lattice can be generated for any sequence of the accelerator model described in the previous section. In the transformation to the online model lattice view, devices may be split into more than one piece, and drift spaces are added (note – the XAL initialization database does not have drifts as accelerator nodes, only actual device information).

The different probes in the online model represent different physics aspects of charged particles beam. There are three probe types commonly used:

- The envelope probe: This is a correlation matrix of moments in a 6D phase space up to second order. By using this probe we can simulate the beam emittance transformation along the lattice.
- The transfer map probe: This represents the transformation matrix of 6D coordinates from the beginning of the sequence to a particular point at the lattice. This probe is usually used for ring modeling.
- The particle probe: This represents the center of the beam. It is frequently used for phase scan analysis and orbit predictions.

Each probe has one or several corresponding algorithms describing the tracking of the probe through the lattice.

The XAL online model went through a series of verifications and benchmarks [2]. The most important feature of the online model is the speed of calculations. The simulation time for SNS sequences is usually much less than one second, and that enables use of the online model in the control room even if a problem includes multiple runs of the model (optimization procedures).

XAL APPLICATIONS FRAMEWORK

In the early days of XAL development, each application was created with its own JFrame menus, toolbars, and standard functionalities like open, save etc. This approach meant duplicated efforts, a different look and feel for each application, and maintenance difficulties. To avoid this the XAL Application Framework was created [4].

An application framework is a set of classes that actual applications extend, and it is used as a common starting point for all XAL applications. There are several advantages of using this framework.

The framework provides all applications with the same look and feel, which helps operators and users to more easily get acquainted with new applications. An example of the application framework template is shown in Fig. 2. A standard “windows application” menu bar, toolbar and empty panel is provided as a starting point, modeled after the familiar windows application format. All menu and toolbar items and actions within them can be customized by changing a simple configuration file which is unique for each application.

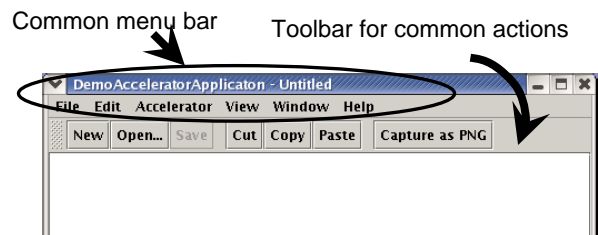


Figure 2: The accelerator application framework template.

The framework uses a document-view architecture, i.e. a single application can have multiple documents

associated with it, and each with its own window view. The application document can be stored (restored) in (from) a file. Of course, this functionality should be provided by the document.

An Accelerator extension of the XAL application framework document class provides an accelerator model with a specific set of menu items. This extension enables the application to read an accelerator file, and to choose and create accelerator sequences.

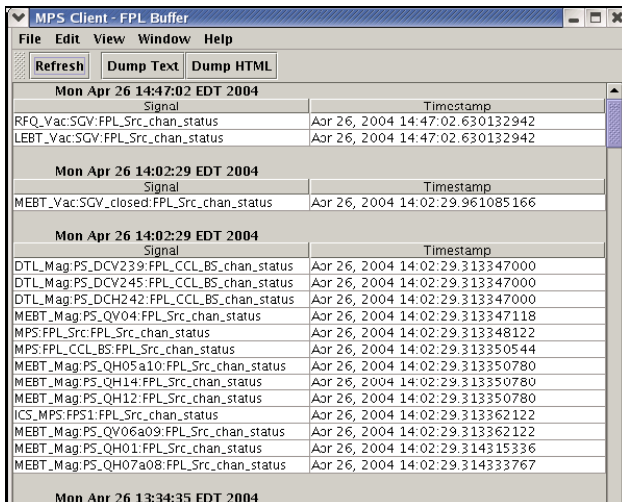
The application framework usage is not mandatory for XAL applications, but it speeds up and facilitates their development and modifications.

XAL SERVICES

XAL services are a special type of XAL applications. They run 24/7 in the background; they can communicate with any number of standard XAL applications; and they do not have a GUI interface. A Service-Application communication uses XML-RPC for inter-process data exchange and uses multicast DNS for discovering subscribers and publishers. Knowledge of the details of multicast DNS and XML-RPC are hidden from the user.

There are several advantages of using the services. First, users need not worry about starting or restarting them. They will restart automatically in the case of shutdown or crash. Second, they reduce the amount of network traffic by avoiding duplicated EPICS and database requests from different XAL applications running simultaneously. And finally, they are prototypes of a future XAL distributed agent systems that is currently under development.

In this paper we present as examples two services that are useful for each accelerator facility.



Mon Apr 26 14:47:02 EDT 2004	
Signal	Timestamp
RFQ_Vac:SGV:FPL_Src_chan_status	Aor 26, 2004 14:47:02.630132942
LEBT_Vac:SGV:FPL_Src_chan_status	Aor 26, 2004 14:47:02.630132942
Mon Apr 26 14:02:29 EDT 2004	
Signal	Timestamp
MEBT_Vac:SGV_Closed:FPL_Src_chan_status	Aor 26, 2004 14:02:29.961085166
Mon Apr 26 14:02:29 EDT 2004	
Signal	Timestamp
DTL_Mag_PS_DCV239:FPL_CCL_BS_chan_status	Aor 26, 2004 14:02:29.313347000
DTL_Mag_PS_DCV245:FPL_CCL_BS_chan_status	Aor 26, 2004 14:02:29.313347000
DTL_Mag_PS_DCH242:FPL_CCL_BS_chan_status	Aor 26, 2004 14:02:29.313347000
MEBT_Mag_PS_QV04:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313347118
MPS_FPL_Src:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313348122
MPS_FPL_CCL_BS:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313350544
MEBT_Mag_PS_QH05a10:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313350780
MEBT_Mag_PS_QH14:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313350780
MEBT_Mag_PS_QH12:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313350780
ICS_MPS_FPS1:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313362122
MEBT_Mag_PS_QV06a09:FPL_Src_chan_status	Aor 26, 2004 14:02:29.313362122
MEBT_Mag_PS_QH01:FPL_Src_chan_status	Aor 26, 2004 14:02:29.314315336
MEBT_Mag_PS_QH07a08:FPL_Src_chan_status	Aor 26, 2004 14:02:29.314333767
Mon Apr 26 13:34:35 EDT 2004	

Figure 3: An example of the XAL MPS client application.

MPS Service

MPS service is our Machine Protection System (MPS) post-mortem application. Originally it was a standalone client application. Later it was migrated to the service framework. This service is always running in the background monitoring MPS events – capturing the

Accelerator/Storage Ring Control Systems

stream of signals that emanate from each trip and sorting them to determine the root cause of the trip. It also provides statistics and views of the MPS trip history. Any number of client applications can view this data. Figure 3 shows a client view of the MPS trip.

PV Logger

Another service application is the XAL “PV Logger”. This application logs predefined sets of control system signal values to a database, at specified intervals and upon requests from any XAL application. One example of use is to grab machine settings directly used by accelerator physics such as magnet, RF and BPM values. This provides complete sets of information needed to configure the online model, taken by a background process. At SNS we have several PV Loggers covering the needs of different systems like accelerator physics, SNS cooling system, beam loss monitors, etc.

XAL AUXILIARY PACKAGES

XAL has numerous general purpose packages. They were developed at different times, and the list is still growing. Usually these packages are independent from the rest of XAL, and they can be easily ripped off and used elsewhere.. Below we discuss a few of the most interesting and useful of these packages .

XAL Plotting Package

Development of the XAL plotting package [5] started at the early stages of development as a research project to study how fast data plotting can be updated. Later, interactive features of the package were found useful in several applications. This package is not intended to completely replace existing and freely available powerful plotting packages, but it is sufficient for instances where simple charts and color surface plotting is required.

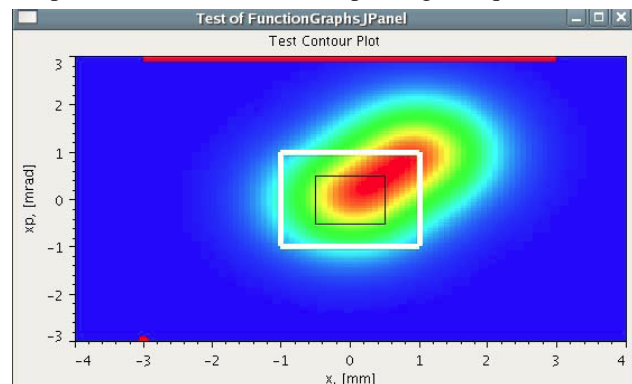


Figure 4: A color surface plot with XAL plotting package.

In terms of the Model-view-controller (MVC) pattern, the XAL internal plotting package is simplified, and it has only two components. The view and controller are combined, and they are implemented in the FunctionGraphsJPanel Java class, which does not have any subclasses. There are four major types of data (Model components). Two are related to 2D chart plotting. The third can be used for bar-charts, and the last one for color

surface 3D plotting. An example of the color surface data plotting with 100x100 point graphics area resolution is shown in Fig. 4.

XAL Channel Access Package

To communicate with accelerator hardware XAL uses the EPICS Channel Access protocol. EPICS communication uses a single “Process Variable” (PV) as the fundamental unit for communication via an EPICS protocol called Channel Access. XAL has a Channel class that encapsulates the communication with a process variable.

The Channel class is an abstract class that has the same interfaces that most control system would provide. This abstract layer insulates the rest of XAL from possible changes in the existing implementation of the EPICS protocol. For EPICS PV communication, we extend it to a concrete class that wraps Java Channel Access (JCA) [6] or Channel Access for Java (CAJ) [7] packages. JCA has interfaces to native C routines, and until recently it was the only non-Java library we used. Now there is CAJ - a 100% pure Java implementation of the EPICS Channel Access library.

The Channel class conveniently hides from users the underlying actions required to make connections to PVs. It also has member functions to provide Process Variable parameters other than the value (e.g. times stamps, units, display limits, etc.). Additionally, it has the capability to switch between synchronous and asynchronous communication and PV monitoring. Another useful feature of the Channel class is the ability to apply a specified “transform” to a Channel value. For example, one may apply a scaling transformation on a power supply current, in order to get a magnetic field level.

Typically applications dealing with the accelerator classes never actually use Channel objects directly, but rather they use methods that provide the information of interest. For example, with a magnet the user may call a getField() or setField() method to get or set the magnetic field. The actual Channel and control system connection details are hidden.

BRICS Package

XAL Bricks package (gov.sns.tools.bricks) is a tool to facilitate writing of GUI interfaces. The main idea of Bricks is to keep information about the GUI window elements and their appearances in a XML file which can be used later to restore the GUI window. The developer can get the reference to the GUI window simply by specifying the name of this XML file. Then, all effort can be concentrated on the functionality of an application instead of its appearance.

The Bricks package also has been integrated with the XAL application framework. During the runtime the bricks definition file can be loaded to create the main XAL application window. After that, all child views are available for usage inside the XAL application document.

To create XML Bricks files XAL includes a GUI builder based on the Bricks package. This builder is a

XAL application and works like others GUI builders for many existing Integrated Development Environments (IDEs). The key difference between existing GUI builders and XAL Bricks is that Bricks does not create any Java code. This design for Bricks is inspired by OpenStep’s™ Interface Builder™ [8]. The snapshot of the Bricks GUI builder application is shown in Fig. 5.

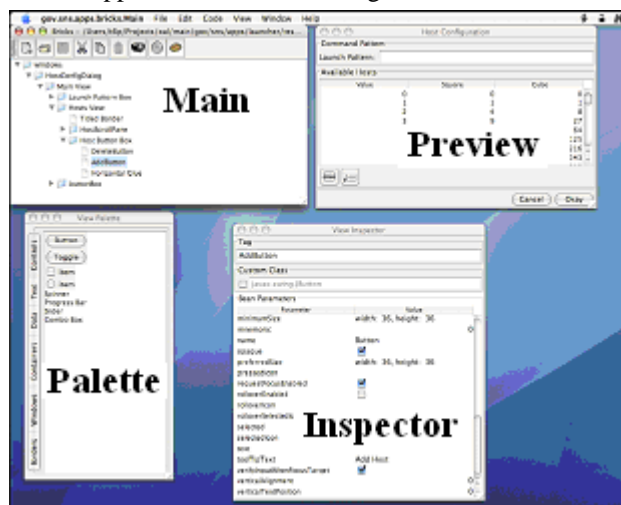


Figure 5: XAL Bricks GUI builder.

The main window for the Bricks document shown in Fig. 5 displays a hierarchy of windows and their views for a future GUI interface. A palette of views shows possible views such as buttons, tables, text fields etc. so the developer can simply select and drop views into the main window. A preview window reveals how the GUI window will look. An inspector is used for editing the properties of a selected view. Copy, Cut and Paste support along with drag and drop support make it easy to rearrange views. The built in code assistant helps developers avoid errors by generating references to views to pass directly into their code.

General Optimization Package

The XAL optimization package (gov.sns.tools.solver) is a “home grown” product of a long evolution of such packages during the XAL development. Currently it is the third generation of optimization packages in XAL. The base components of the package are:

- The Solver component: This is a primary class for setting up and running an optimization process. To create an instance of the Solver we need instances of the Algorithm Market, Stopper, and Solution Judge classes. The Problem class is needed to start the optimization process in the Solver.
- The Algorithm Market: This is a collection of algorithms that will compete during the optimization process. The user can choose arbitrary combination of algorithms from the existing ones in the package.
- The Stopper Class and Subclasses: This decides if the process should stop. The decision is made on the basis of time, number of iterations, a level of

satisfaction, etc. Users can choose which criterion or which combination of criteria will be used.

- The Solution Judge component: The implementation of the Solution Judge interface compares the new solution and the best known one at a particular moment.
- The Problem Class: This class consists of the list of variables, restriction rules, possible hints for different algorithms, and a scorer that estimates the model function.

At the moment there are four available algorithms: a random search, a random search with a shrinking range of search, a direct gradient method, and a simplex algorithm.

The XAL optimization package has no upward connection to other XAL classes except the XAL message center, and it can be freely used outside of XAL.

Least Square Method Package

There is a special linear squares method (LSM) package (gov.sns.tools.fit.lsm) in XAL to solve optimization problems that can be reduced locally to a linear problem. Usually, we want to find an approximation of measured points (x,y) by a known function with unknown parameters. This type of problems has exact solutions, so there is no need to use the general XAL optimization. The XAL LSM package implements two algorithms: a classic LSM and a Levenberg-Marquardt Method (LMM) [9]. Both methods assume that the user will provide methods to calculate partial derivatives of the model function with respect to the model parameters.

The two types of functions most frequently used in XAL applications are a Gaussian distribution and a damped sinusoidal oscillation,

$$y_{gauss}(x) = a_0 + a_1 \cdot x + \exp\{a_2(x - a_3)^2\},$$

$$y_{damp}(x) = a_0 + a_1 \cdot \exp(a_2 x) \cdot \sin(a_3(x - a_4)),$$

where a_i are the unknown parameters. The Gaussian distribution function is used in wire scanner data analysis, and the damped oscillations are used in SNS ring tune calculations. To provide initial values for the model parameters, both fitting classes have a “guess” method which analyzes the initial data and suggests reasonable values. There is also a polynomial class that calculates coefficients of arbitrary power series.

Another advantage of the LSM package with respect to the general XAL optimization is that errors for model parameters can be estimated if the user specifies the errors in the measured data.

Formulas Evaluation Package

The formulas evaluation package (gov.sns.tools.formula) is used to evaluate a formula with a given set of variables provided by the user. The formulas are presented as text. A Formula Interpreter class compiles the text and evaluates the formula. The compilation process is performed only once, so the efficiency of following evaluations is very high. This allows use of the package for fitting problems inside the

optimization package. The formulas evaluating package has a variety of operators including arithmetic operators, function operators, and logical operators. The set of internal functions can be extended.

XAL APPLICATIONS

At this moment, only the SNS project has more than 50 XAL applications. They were developed at different times by different developers, and many of them do not comply completely with all standards for correctly implementing XAL applications. In this paper we consider only few of them which could be useful everywhere.

To write a correctly implemented general XAL application the developer should follow several rules:

- The application cannot use java classes from another application. If a Java class is useful for more than one application, it should be moved into the core XAL packages.
- The source code of the application should not contain any specific information related to a particular accelerator.
- The application should use the accelerator model for hardware interaction. The model should be initialized from the default XML accelerator model file.
- The number of external packages should be minimized.

Virtual Accelerator

A virtual accelerator (VA) is a simulation program running permanently and generating sensible diagnostics signals in response to changing parameters of the model. From the point of view of a client that communicates (through EPICS or another protocol) with this VA there is no difference between the VA and a real machine. The VA is a very useful tool for developing and testing control room applications, especially during the early stages when the real accelerator is not functioning yet.

In the beginning of XAL development, the virtual accelerator was based on PARMILA or TRACE3D codes and an EPICS Portable Channel Access Server (PCAS) [10]. The structure of this VA is shown in Fig. 6. Later, the XAL online model was used as a simulation program.

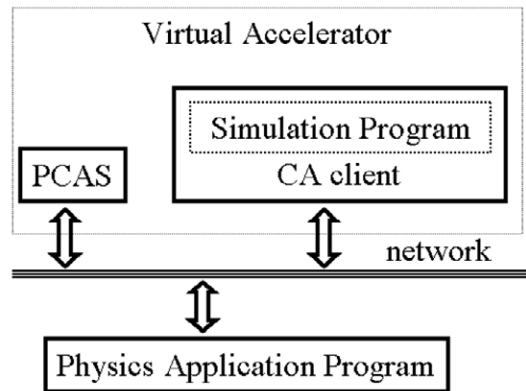


Figure 6: The structure of a virtual accelerator.

Until recently, PCAS was the only option to provide communications between an application and the simulation program. It was inconvenient because each time the user had to generate a special initialization file with a list of EPICS PV names that will be used in the VA. After replacing the JCA wrapper around the native EPICS C-library by the CAJ (pure Java EPICS communication package [7]), the PCAS was dropped from XAL. Now the XAL Virtual Accelerator Application does not need any external executables and produces all necessary EPICS signals by itself. The user needs only the XML Accelerator Model file to start VA.

Orbit Correction Application

An orbit correction application was developed to correct transverse orbit errors with dipole correctors and bend magnets in the linac and the ring. Among the features of this application, there is the ability to save and restore the existing orbit, to correct orbit errors to zero or a reference orbit, to switch on and off any particular dipole correctors from the process, and to initialize from the XML accelerator model file. This application utilizes the XAL optimization package by specifying a goal of optimization as a combination of zeroing orbit errors, by providing a smooth orbit, and by keeping the currents in the correctors within their control limits. The satisfaction function of this optimization is a nonlinear combination of all goals. The use of nonlinearity eliminates the need to introduce artificial weights, as is often necessary with many linear optimization approaches.

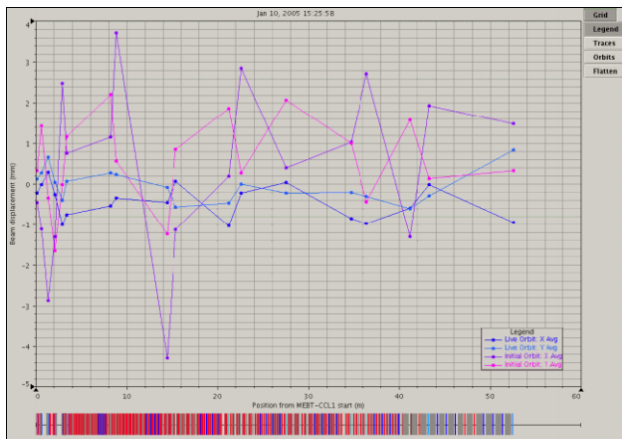


Figure 7: Orbit correction for the MEBT-DTL-CCL1 SNS's linac sections.

When correcting an orbit, the user can choose whether to correct the orbit based on the XAL online model or through empirical measurement. For the first case, the orbit response coefficients of the dipole correctors for the downstream beam position monitors (BPMs) come from the online model. If the model is incorrect the orbit correction will fail. In this situation the empirically measured coefficients can be used. The empirical approach will work even in a case where there are dipole correctors and BPM polarity errors. The disadvantage of

this method is the long period of time needed to measure the orbit response to each dipole corrector.

The result of the orbit correction for part of the SNS linac is shown in Fig. 7. The four trajectories are the initial horizontal and vertical trajectories (purple and pink, respectively), and the horizontal and vertical trajectories after correction (blue and light blue, respectively). The initial trajectory excursion is about 4 mm horizontally and 2 mm vertically. After correction, trajectory oscillation in both planes is within 1 mm.

The XAL orbit correction application can also serve as a live orbit display.

CONCLUSION

The XAL framework has three basic interconnected components: the hierarchical accelerator model, the online model, and the XAL application framework. Additionally, XAL includes a collection of independent Java packages that can be used anywhere.

Correctly implemented XAL applications can be easily ported to other accelerators, and there are persistent efforts to make the XAL more portable.

ORNL/SNS is managed by UT-Battelle, LLC, for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

- [1] J Galambos et al. "XAL Application Programming Framework", ICALEPCS'03, Gyeongju, Korea, October 2003, WE204, <http://www.JACoW.org>.
- [2] C.K. Allen et al. "A novel online simulator for applications requiring a model reference", ICALEPCS'03, Gyeongju, Korea, October 2003, WE116, <http://www.JACoW.org>.
- [3] N. Malitsky and R. Talman, "The Framework of the Unified Accelerator Libraries", Monterey, CA, USA, September, 1998, ICAP 1998.
- [4] T. Pelaia "XAL Application Framework and Bricks GUI Builder", ICALEPCS'07, Knoxville, TN USA, October 15-19, 2007, TPPA09, <http://www.JACoW.org>.
- [5] A. Shishlo et al. "Java Swing-Based Plotting Package Residing Within XAL", ICALEPCS'07, Knoxville, TN USA, October 15-19, 2007, TPPA08, <http://www.JACoW.org>.
- [6] <http://www.aps.anl.gov/xfd/SoftDist/swBCDA/jca2/index.html>.
- [7] <http://caj.cosylab.com>.
- [8] Nancy Craighill, OpenStep for Enterprises, John Wiley & Sons, Inc., New York, NY, 1997.
- [9] Numerical recipes in C: The art of scientific computing. Cambridge University Press., 1992.
- [10] A. Shishlo et al. "The EPICS Based Virtual Accelerator - Concept and Implementation", PAC'03, May 12-16, Portland, OR, 2003, WPPE017, <http://www.JACoW.org>.

IMPROVEMENT PLANS FOR THE RHIC/AGS ON-LINE MODEL ENVIRONMENTS *

K.A. Brown[†], L. Ahrens, J. Beebe-Wang, J. Morris, S. Nemesure,
G. Robert-Demolaize, T. Satogata, V. Schoefer, S. Tepikian
C-AD Dept., BNL, Upton, NY

Abstract

The on-line models for Relativistic Heavy Ion Collider (RHIC) and the RHIC pre-injectors (the AGS and the AGS Booster) can be thought of as containing our best collective knowledge of these accelerators. As we improve these on-line models we are building the framework to have a sophisticated model-based controls system. Currently the RHIC on-line model is an integral part of the controls system, providing the interface for tune control, chromaticity control, and non-linear chromaticity control. What we discuss in this paper is our vision of the future of the on-line model environment for RHIC and the RHIC pre-injectors. Although these on-line models are primarily used as Courant-Snyder parameter calculators using live machine settings, we envision expanding these environments to encompass many other problem domains.

INTRODUCTION

The Collider Accelerator Department (C-AD) at Brookhaven National Laboratory (BNL) operates a series of accelerators that serve the purpose of providing beams to the RHIC experiments. These accelerators include the AGS, which first operated with beams in 1960 and the two RHIC rings, that began beam operations in 2000.

RHIC consists of two super-conducting accelerators, 2.4 miles in circumference, with counter-rotating beams. It has six interaction regions where the two beams can be put into collisions with zero crossing angle. We currently operate with collisions in two of these regions. RHIC can be operated in many different modes and with many different types of beams [1]. For example, RHIC is able to run with two different ion beams in the two rings simultaneously (e.g., gold and deuteron beams in collision) [2]. RHIC can accelerate gold ions up to 100 GeV/nucleon and polarized proton beams up to 250 GeV/c (for more on RHIC performance see [3, 4]).

To deliver polarized protons to RHIC, the beam accelerates as H^- ions through the 200 MeV LINAC is stripped to H^+ and brought up to 2.16 GeV/c in the Booster synchrotron. The beam is then transferred into the AGS and accelerates to 23.8 GeV/c. Finally the beam is transferred to the two RHIC rings, ending with polarized protons up to 250 GeV. For ion operations the process starts at the Tandem Van de Graff. A gold beam, for example, is stripped

of some of the outer shell electrons at the Tandem and is brought through a long transport line to the Booster. From the Booster the gold ions are transferred to the AGS and stripped to Au^{77+} in the transfer line. The final two electrons are stripped off in the AGS to RHIC transfer line.

ACCELERATOR CONTROLS SYSTEM

We have two ways to view accelerator controls interfaces. One can take an engineering view in which we think in terms of power supply configurations and in physical units of current and voltage. This paradigm had worked well for decades, before large scale computing was able to take over the more computationally involved process of working in terms of beam parameters, such as betatron tune, chromaticity, and other Courant-Snyder parameters [5].

What is important is that we develop controls that allow the best mapping between how we think of the accelerator and how we control it. This also allows more of the information that describes the various subsystems to be captured into the controls systems. For example, if you have a transport model of a beam line in the controls, the system will contain not only the transfer functions for control units (e.g., some 0 to 10 volt reference to a power supply, derived from a 16 bit digital to analog conversion module) but also transfer functions from current to field, gradient, and even normalized strengths. This then captures not only the power supply information, but also the magnet information. The controls system now begins to hold the best collective knowledge of the accelerators. It could even contain the best collective knowledge of the beam dynamics.

The controls systems at C-AD span multiple generations of technologies. The controls for RHIC represent the largest systems, in terms of total number of control points (over 220,000 settings and over 160,000 measurements) [6]. From the point of view of the online models, there are then multiple interfaces that need to be defined to collect live parameters of the accelerators.

Generally speaking, all of the controls systems are hierarchical with multiple physical and software layers. At the lowest level we speak of a front end computer (FEC) that directly interfaces to some piece of hardware (a power supply or an instrumentation module). The front end systems interface to the console layers of the system through high speed Ethernet employing fiber-optic network connections. In this respect one can think of the controls system as a widely distributed computer system where computational work is performed in parallel. This is somewhat

* Work performed under Contract Number DE-AC02-98CH10886 with the auspices of the US Department of Energy.

[†] kbrown@bnl.gov

naïve, since each FEC performs a very specific task (they cannot share the work of some other FEC). Our reason for making this analogy is to highlight two things. First is the concept of distributing the accelerator state for the controls to the lowest levels of the hierarchy. The other is to allow us to make the point that this form of a controls system implements an engineering paradigm of controls. The work is distributed according to the power supply configurations (which often don't parallel magnet configurations), not according to the magnet configurations. So the work that the online model performs cannot be distributed down to the lowest level of the controls hierarchy, but must remain at a higher abstraction in the software layers.

HOW MODELS FIT IN

The software layers of the controls can be divided up into three parts. There is the software on the FEC's, then there is the middleware manager (server) layer, and finally there is the console layer. The online models reside in the middleware manager layer, that being the lowest level of the software hierarchy that can perform that work. Each manager acts as a domain-specific arbiter between the console applications and the FEC's. Multiple hardware systems can be employed to distribute managers, but a single manager, unless it threads itself off into multiple instances, remains fixed on a single computer server as a single process.

The basic design of the RHIC online model server allows for multiple instances using different computational model engines [7]. The main computational engine used is an in-house system we call Opticalc. The communication interface for the online model uses CDEV [8]. The main client interfaces that connect the online model to the controls system are the *RampManager* server and *RampEditor* console application [9]. Through this system we achieve control of the accelerator in units of betatron tune and chromaticity. The division of work, between the online model server, the *RampManager*, and the FEC's permits each to focus on a specific function. The model server works purely in normalized strengths, with the *RampManager* handling the management of transfer functions, along with the *WaveformGenerator* manager that builds the actual references that are sent to the FEC and eventually into the various power supplies. In this system the *RampManager* acts as the central authority, managing the interfaces between the hardware, the model, and the operator.

The AGS/Booster online model system is not yet an integral component of the controls system. It acts more as an advisor, providing basically what is a fast offline interface to the operator. The basic design follows very closely the design of the RHIC online model system [10, 11], except *madx* [21] is used as the computational engine. The communication interface uses CDEV, and the design allows for multiple instances using different computational model engines. The main client is the *AGSModelViewer*, which obtains controls data (e.g., the tune control functions, the main magnet ramp, and the RF functions) as

well as data logged from instrumentation (i.e., to compare tune and chromaticity measurements with the model). The *AGSModelViewer* also works with a longitudinal model of the accelerators, encapsulated in a library interface, and so presents to the operator both longitudinal and transverse beam parameters.

The kind of online model system we are working towards is a manager, advisor cooperative, in which the *RampManager* is the central authority and the model server acts as an advisor to all other applications. This is shown in figure 1. The RHIC system is very close to this kind of system, lacking only a model viewer that is as extensive as the AGS/Booster system. The model viewer, when built to encapsulate not just the simple linear model, but to allow comparison to other models and measurements, grows in functionality to a controls viewer. That is, the conceptual view of the accelerator and the mapping to the controls begin to converge where the *RampEditor* is the controls interface and the model viewer is the visualization interface. Figure 2 shows a snapshot of the *AGSModelViewer* application interface.

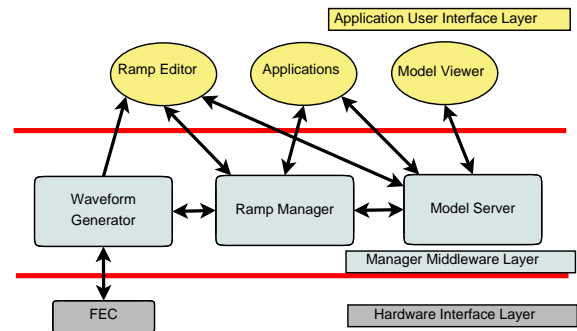


Figure 1: The manager, advisor online model cooperative in the controls system.

OFFLINE SIMULATION ACTIVITIES

Offline simulation activities tend to be focused on understanding phenomena that limit beam performance. This can include relatively simple simulations, such as matching at injection, to highly specialized simulations, such as predicting the performance of transverse stochastic cooling in RHIC. It is interesting to review the current set of activities and consider what offline work could prove useful as either online simulations that become part of the controls or as activities that can be improved by linking to the online system.

In the Booster and the AGS a significant amount of the offline simulation activities are associated with improving the performance of polarized protons operations. To preserve polarization in the AGS the vertical betatron tune needs to be kept very near the integer [12]. A significant amount of work goes into comparing beam based measurements to predictions, including momentum offset and other effects. This work will eventually lead to precise

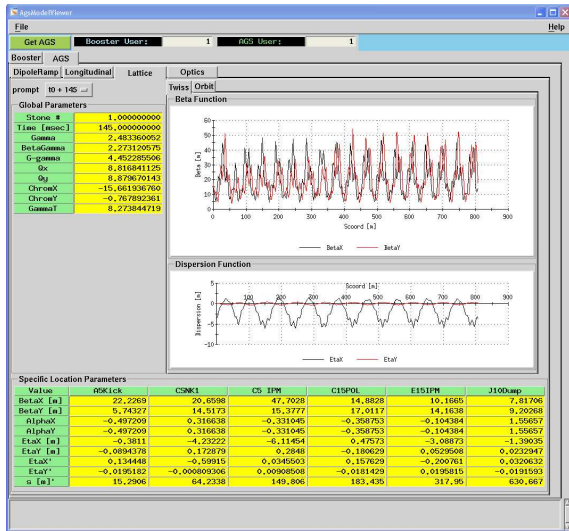


Figure 2: AGSModelViewer showing polarized proton lattice at injection. β -functions are distorted by the presence of partial snakes in the lattice.

tune control in the AGS, through the online model interface. Another activity is learning how to match the Booster to AGS transfer line optics to the AGS lattice distorted by the partial snakes [13]. The main area of activity here is in trying to understand how to model the snakes and bring that model into the online model [14]. Keeping transverse emittance as small as possible is important for polarization preservation, simulations of emittance growth on the injection foil during Booster injection are another area of activity [15]. Having these simulations online could benefit machine performance, especially since the dynamics are complicated enough to not be entirely intuitive.

Other areas of offline work include space charge calculations at AGS injection, spin tracking, and understanding horizontal resonances. A new project in the AGS is the horizontal tune jump system. Two fast quadrupoles were added to the AGS lattice [16]. They will be programmed to perform tune jumps across 82 horizontal spin resonances. Including this system into the online model has already proved desirable, where some pieces of the online system were used to provide predictions.

In RHIC, the focus of offline simulations is on polarized proton performance and ions operation performance.

Even though RHIC has full snakes (each ring has two 180 degree snakes), there are still spin resonances that can cause polarization loss. Areas of study related to spin include spin tracking and polarization preservation above 100 GeV, where there are strong resonances [17]. There can also be polarization losses during stores. Polarized protons are also susceptible to beam-beam interactions [18].

For ions operations, the dominant intensity limitation is instabilities that occur after the gamma transition jump. Various areas of exploration are being pursued, including electron cloud effects and nonlinear chromaticity. During store intrabeam scattering (IBS) limits luminosity. New lat-

Controls and Computing

tices have been developed to suppress IBS [19]. The development of a new lattice is an offline activity. This is because great care must be taken to build functions that are realistic for the power supplies to follow.

Other areas of offline activity include orbit feedback, orbit response matrix measurements (ORM) [20], and injection steering and matching. This last activity is actually a mostly online activity, but developing optical matches between the AGS to RHIC transfer line and the RHIC lattice remains an offline activity. Significant work also goes into offline analysis of collimation.

Finally a significant amount of effort goes into simulation of individual elements or systems. These include improving our understanding of the RHIC injection kickers and for the development of a superconducting 56 MHz RF cavity.

INFRASTRUCTURE IMPROVEMENTS

All online software systems are version controlled and under a backup system.

The lattice descriptions are managed outside of the main controls infrastructure, although they are under some form of version control. All lattices are under a backup system. The lattices used in the online models are provided by the responsible physicist for the given system. For these lattices the plan is to encapsulate them into a true version controlled system. The responsible physicists will then employ this system to make changes.

There are a number of tools, either obtained from other accelerator laboratories, or built in-house, that are not managed within the main controls infrastructure. Some of these tools are used by the online system, such as *madx*, and others could be useful for the online systems, such as certain in-house built spin tracking modules. This software should, but may not, be under source control and backups. Since some of this software resides on individuals computers. Documentation of this software is usually minimal and mechanisms for bug tracking and reporting are not implemented.

A particular problem we are focusing on is the linking of the online models to offline analysis. One example of this is the ORM analysis. In this case clearly the online and offline descriptions of the accelerator need to be identical. But there is no mechanism, besides individual initiative, that ensures they are identical. So we are moving to systems where everyone uses the same lattice descriptions, there is a responsible contact for each lattice and lattice tool, and there is user guide documentation.

We have formed an online model working group to work through these infrastructure improvements, consisting of members of the physics group, controls group, and operations. A significant amount of the work is currently focused on model development and verification for the online models. But the infrastructure is being built to permit offline and online simulations to use the same descriptions of the accelerators and beam lines.

MOVING OFFLINE WORK ONLINE

We often discuss the process of moving offline work to online [22]. Some offline work needs only better access to real time machine parameters. Some is associated with beam experiments and is only useful during those periods. What is important, from our perspective, is that there is a need for mechanisms that allow faster offline work to be done using real time machine parameters.

Members of the physics group do most of the offline work. Since they have a diverse set of tools they employ for simulations, we focus on providing a framework with which they can easily access the information they need to perform offline studies. This can satisfy the difficulty of getting real time parameters to offline work, but it doesn't bring the offline work into the online infrastructure. For this we need to define what physics is needed in the online environment. This is a continual process, since as models become more sophisticated and realistic, we can seriously consider bringing them into the online system. We also need to consider that much of the offline work is computationally demanding. Bringing it online may mean purchasing better servers or bringing cluster computing into the controls system.

Finally, but most importantly, bringing offline work into the online environment necessarily brings it to a larger audience. This provides the ability to use these offline tools in the online system as teaching aides.

LONG TERM PLANS

Our goal is to support a model based accelerator controls system and to provide an accelerator physics based approach to accelerator operations. We intend to provide support to accelerator physicists and operators assisting physicists. To do these things we need to have well understood models of the accelerators and transfer lines that have been tested against beam based experiments. At C-AD there is a strong collaboration among physics, operations, and controls. We intend to build on this collaboration to achieve our goals.

We are actively investigating improved simulation engines for the online models, investigating ways to make *madx* faster and learning about other systems such as *ptc* [23]. We are also looking into ways of standardizing and improving the data formats with tools such as *SXF/ADXF* [24] and *HDF* [25].

REFERENCES

- [1] T. Roser, *Proceedings of the 2007 Asian Particle Accelerator Conference (APAC 07)*, Indore, India, pp.74-78, (2007)
- [2] C. Gardner, et al, *Proceedings of 11th European Particle Accelerator Conference (EPAC 08)*, Magazzini del Cotone, Genoa, Italy, 23-27 Jun 2008, pp 2548
- [3] A. Drees, et al, *Proceedings of Particle Accelerator Conference (PAC 07)*, Albuquerque, New Mexico, 25-29 Jun 2007, pp 722
- [4] C. Montag, et al, *Proceedings of 11th European Particle Accelerator Conference (EPAC 08)*, Magazzini del Cotone, Genoa, Italy, 23-27 Jun 2008, pp 2566
- [5] E. D. Courant and H. S. Snyder, *Annals Phys.* **3**, 1 (1958) [*Annals Phys.* **281**, 360 (2000)]
- [6] D. S. Barton et al., *Nucl. Instrum. Meth. A* **499**, 356 (2003).
- [7] T. Satogata, K. Brown, F. Pilat, A. Alai Tafti, S. Tepikian and J. van Zeijts, *Prepared for IEEE Particle Accelerator Conference (PAC 99)*, New York, New York, 29 Mar - 2 Apr 1999
- [8] J. Chen, et al. *Proceedings of ICALEPCS 1995*
- [9] J. van Zeijts *Proceedings of ICALEPCS 2001*
- [10] K. Brown, J. Niederer, T. Satogata, A. Alai Tafti, N. Tsoupas and J. van Zeijts, *Proceedings of IEEE Particle Accelerator Conference (PAC 99)*, New York, New York, 29 Mar - 2 Apr 1999
- [11] K.A. Brown et al., *Workshop on Automated Beam Steering and Shaping (ABS)*, CERN, Geneva, Switzerland, 14-16 Dec 1998
- [12] H. Huang et al., *Proceedings of Particle Accelerator Conference (PAC 07)*, Albuquerque, New Mexico, 25-29 Jun 2007, pp 748.
- [13] K.A. Brown et al., *Proceedings of IEEE Particle Accelerator Conference (PAC 09)*, Vancouver, B.C., May 4-8, 2009
- [14] V. Schoefer et al., *These Proceedings*
- [15] K.A. Brown et al., *Proceedings of IEEE Particle Accelerator Conference (PAC 09)*, Vancouver, B.C., May 4-8, 2009
- [16] J.W. Glenn et al., *Proceedings of IEEE Particle Accelerator Conference (PAC 09)*, Vancouver, B.C., May 4-8, 2009
- [17] M. Bai et al., *Proceedings of IEEE Particle Accelerator Conference (PAC 09)*, Vancouver, B.C., May 4-8, 2009
- [18] Y. Luo et al., *Proceedings of 11th European Particle Accelerator Conference (EPAC 08)*, Magazzini del Cotone, Genoa, Italy, 23-27 Jun 2008.
- [19] A. V. Fedotov et al., *Presented at 42nd ICFA Advanced Beam Dynamics Workshop on High-Intensity, High-Brightness Hadron Beams (HB 2008)*, Nashville, Tennessee, 25-29 Aug 2008.
- [20] J. Safranek, *NIM A*388 (1997) 27
- [21] H. Grote and F. Schmidt, *Proceedings of the 2003 Particle Accelerator Conference*, Portland, OR, pp. 3497-3499 (2003)
- [22] N. Malitsky et al., *Proceedings of Particle Accelerator Conference (PAC 05)*, Knoxville, Tennessee, 16-20 May 2005, pp 880.
- [23] E. Forest, F. Schmidt, and E. McIntosh, *Technical Report CERN-SL- 2002-044, KEK-Report 2002-3*, 2002
- [24] N. Malitsky and R. Talman, *Proceedings of ICAP 2006*, Chamonix, France, 2-5 October 2006, pp 297-302
- [25] The HDF Group. <http://www.hdfgroup.org/HDF5/>

BEAM-BEAM SIMULATIONS FOR KEKB AND SUPER-B FACTORIES

K. Ohmi, KEK, Tsukuba, Japan

Abstract

Recent progress of KEKB and nano beam scheme adopted in KEKB upgrade are discussed. For the present KEKB, chromatic x-y coupling, which was the key parameter to improve luminosity, is focussed. Beam-beam simulations with weak-strong and strong-strong models for nano beam scheme are presented. A weak-strong simulation was done in the presence of the longitudinal microwave instability. Finally status of beam simulations in KEK supercomputers is presented.

INTRODUCTION

Crab cavity has been installed into KEKB to boostup the luminosity performance. Basically the crab cavity should give us potential to increase the beam-beam parameter more than 0.1. Actually various errors disturb to achieve the high beam-beam performance. For example linear x-y coupling at IP induces an emittance growth with couple to the beam-beam nonlinear interaction. Fast turn by turn fluctuation of the beam position also induced an emittance growth with couple to the interaction. To achieve the high beam-beam parameter, errors should be removed as possible as we could. Tuning of colliders is just the work to remove errors. Tolerance for errors are estimated in simulations, but it is hard to know how much errors exist, how to correct the errors and how the errors were corrected in an accelerator.

Recently KEKB achieved the new luminosity record. The luminosity record increases 20%, from 1.76 to $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in June 2009. It is twice of the design luminosity, $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Tuning of chromatic x-y coupling improved the luminosity remarkably.

For KEKB upgrade, we turn to the strategy to boost-up the luminosity. Higher beam-beam parameter is hard to achieve against various errors. Increasing current is also problem for the operation cost. Nano-beam scheme, in which low emittance and low beta beams collide with a large crossing angle, is alternative way.

We discuss simulations of the crab crossing of the present KEKB and nano-beam scheme of the KEKB upgrade in Sec II and III, respectively. In Sec. IV, the computer environment of KEK is reviewed.

Beam-Beam Interaction

RECENT PROGRESS OF KEKB

Chromatic x-y coupling

The existence of the chromatic x-y coupling was known by a measurement of the synchro-beta sideband in the beam size on the x-y tune space [1]. Simulations including the chromatic coupling has been performed using a symplectic integration method of the chromaticity [2]. Hamiltonian which expresses generalized chromaticity is given by

$$H_I(x, \bar{p}_x, y, \bar{p}_y, \delta) \quad (1)$$

$$= \sum_{n=1} (a_n x^2 + 2b_n x \bar{p}_x + c_n \bar{p}_x^2 + 2d_n xy + 2e_n x \bar{p}_y + 2f_n y \bar{p}_x + 2g_n \bar{p}_x \bar{p}_y + u_n y^2 + 2v_n y \bar{p}_y + w_n \bar{p}_y^2) \bar{\delta}^n / 2.$$

The coefficients $10 \times n$ are related to n -th order chromaticity of 10 Twiss parameters, $\alpha_{x,y}, \beta_{x,y}, \nu_{x,y}$ and $r_i, i = 1, 4$. Transfer map using H as a generating function guarantees the 6D symplectic condition.

Alternative way is the direct map for the betatron variables $\mathbf{x} = (x, p_x, y, p_y)^t$ and z as

$$\mathbf{x}(s+L) = M_4(\delta) \mathbf{x}(s). \quad (2)$$

$$z(s+L) = z(s) + \mathbf{x}^t M_4^t(\delta) S_4 \partial_\delta M_4(\delta) \mathbf{x} / 2 \quad (3)$$

where $M_4(\delta)$, which is the revolution matrix at the interaction point, which contains 10 Twiss parameters and their chromaticity. The transformation for z guarantees the 6-D symplectic condition.

Twiss parameters at the interaction point is measured by turn by turn position monitors located at the both side of the interaction point [3, 4]. Their chromaticity is given by scanning RF frequency in the range of $\pm 200 \sim 300$ Hz.

Figure 1 shows the measured x-y coupling parameters as functions the momentum deviation. The parameters are fitted by polynomial of the momentum deviation, as follows,

$$\begin{aligned} r_1(\delta(\%)) &= 0.00848 - 0.00435\delta + 0.00909\delta^2 + 0.151\delta^3 \\ r_2(\delta(\%)) &= 0.0137 + 0.00696\delta + 0.0222\delta^2 - 0.320\delta^3 \\ r_3(\delta(\%)) &= 0.189 - 0.304\delta + 2.45\delta^2 - 1.24\delta^3 \\ r_4(\delta(\%)) &= 0.0277 - 0.942\delta + -0.512\delta^2 - 0.301\delta^3 \end{aligned} \quad (4)$$

The coefficients, which are chromaticity, varies run by run, and differ from prediction of the optics design code like SAD. Therefore the accelerator model based on the measured chromaticity is important.

Using these transformation, synchro-beta resonances and their effects on the beam-beam interaction have been

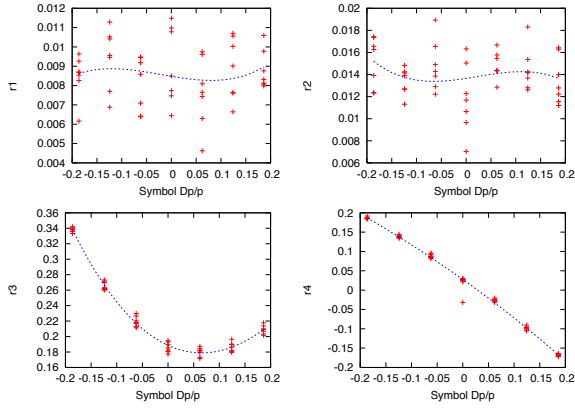


Figure 1: Measurement of the chromaticity for x-y coupling in KEKB-LER.

studied [5]. The transformation is implemented into both of the weak-strong (BBWS) and strong-strong (BBSS) codes for beam-beam interaction. Figure 2 shows the luminosity as a function of the chromaticity given by the weak-strong simulation. The chromaticity for r_4 , $dr_4/d\delta$, can affect the luminosity: the luminosity degradation is 10-15 % for $dr_4/d\delta \sim 200$. Skew sextupole magnets are installed to correct the chromaticity. Luminosity increases 20% due to the chromaticity tuning as shown in Figure 3.

KEKB UPGRADE - NANO BEAM SCHEME

KEKB upgrade is progressing the design with the nano-beam scheme. In the nano-beam scheme, low emittance and low beta beams collide with a large crossing angle. Parameters as candidates of KEKB upgrade are summarized in Table 1.

Crab waist technique to fit the waist of the beam to the axis of colliding beam is taken into account in KEKB upgrade [6].

The ratio of the horizontal projection of the bunchlength and horizontal size, $\phi\sigma_z/\sigma_x$, which is called Piwinski angle, indicates the overlap area of two beam at collision. In the low emittance approach, the large ratio $\phi\sigma_z/\sigma_x \approx 20$ means a large number of slice in the simulation.

Weak-strong simulation

Weak-strong simulation is convenient to survey feasibility of the design. Macro-particles, which represents weak beam, collide with a fixed charge distribution as the strong beam. The strong beam is a fixed tri-Gaussian distribution for x, y, z with the size $\sigma_{x,y,z}$. The strong beam is sliced along z . The number of slices are 200. Since the overlap area is $\Delta z = \sigma_x/\phi \sim 20 \times \sigma_z$, the area, which mostly contribute the luminosity, is sliced into 10 pieces in the average. Hourglass effect in the area is serious when the beta function is smaller than the overlap area, $\beta_y < \sigma_x/\phi$. The luminosity obtain by the weak-strong simulation is filled in

Beam-Beam Interaction

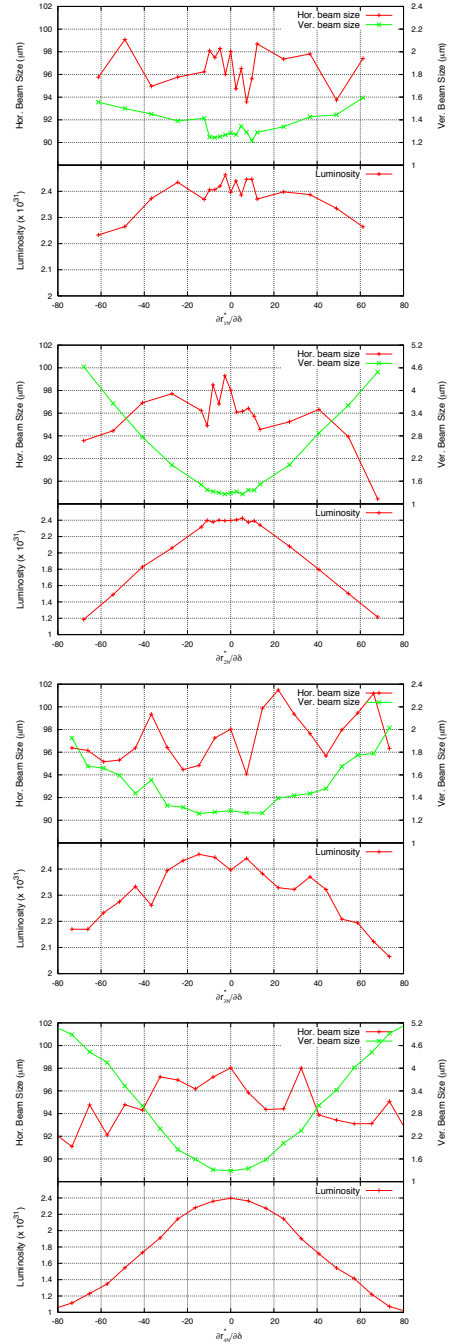


Figure 2: Simulated luminosity for the chromatic x-y coupling.

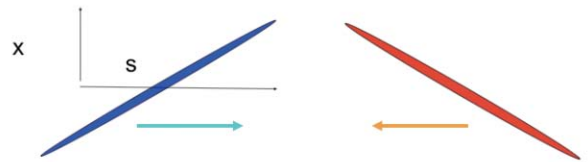


Figure 4: Schematic view of the low emittance approach for KEKB upgrade.

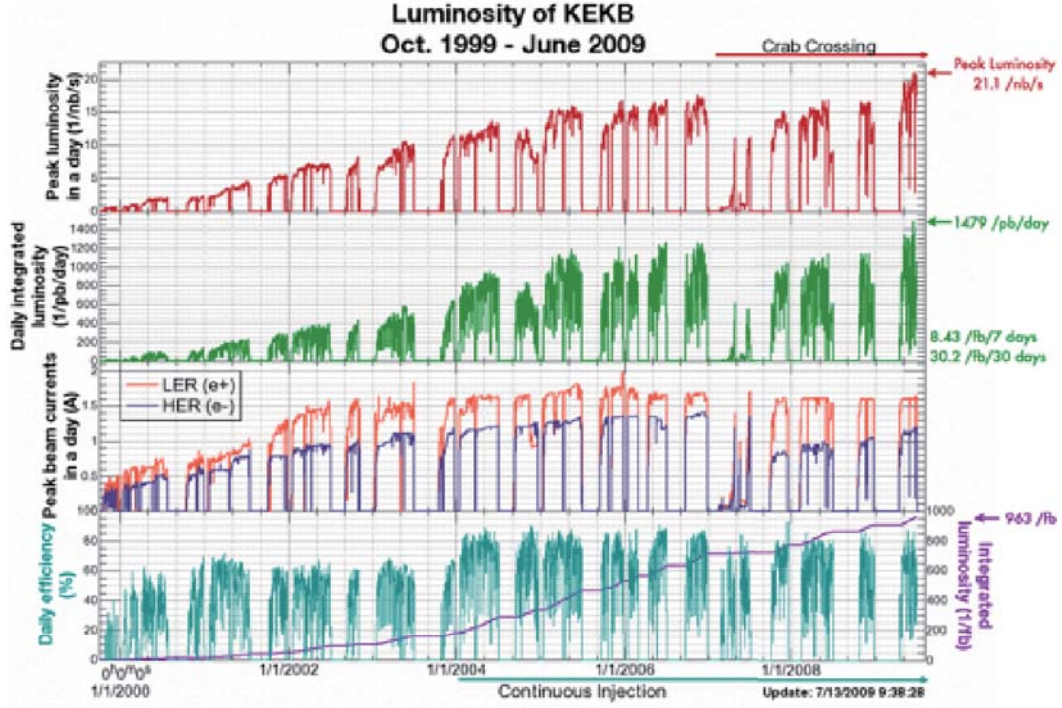


Figure 3: Luminosity trend of KEKB.

Table 1: Parameters for KEKB upgrade and Frascati Super B factories.

variable	HighCurrent	NanoBeam-1	NanoBeam-2	Frascati
E_p/E_e (GeV)	3.5/8	3.5/8	3.5/8	4/7
ε_x (nm)(L/H)	24/18	2.8/2.0	2.8/2.0	2.8/1.6
ε_y (pm)	240/90	33.6/10.7	20.7/36.0	7/4
β_x (mm)	200/200	44/25	17.8/25	35/20
β_y (μ m)	3/6	0.21/0.37	0.26/0.26	0.22/0.39
σ_x (μ m)	69/60	11/7.07	7.06/7.07	9.9/5.66
σ_y (μ m)	0.85/0.73	0.084/0.063	0.073/0.097	0.039/0.039
σ_z (μ m)	5/3	5/5	5/5	5/5
ϕ (mrad)	0	30	30	24
$\phi\sigma_z/\sigma_x$	0	14/21	21/21	14/25
σ_x/ϕ (μ m)	-	0.37/0.24	0.24/0.24	0.35/0.20
N_p/N_e (10^{11})	12/5.25	10.7/6.17	10.7/6.17	0.55/0.55
$N_{bunch}/\text{Cir(m)}$	5000/3016	2230/3016	2252/3016	1251/1800
ξ_y	0.3/0.5	0.081/0.081	0.079/0.079	0.147/0.150
L ($\text{cm}^{-2}\text{s}^{-1}$)	5.3×10^{35}	$8(2.9) \times 10^{35}$	$8(8.5) \times 10^{35}$	$(11) \times 10^{35}$

table 1, where bracket and non bracket is those with and without the crab waist. In NanoBeam-1, luminosity without crab waist degrade 1/3 of that with crab waist, while no big difference for NanoBeam-2. Essential point whether such difference arises or not is the overlap area σ_x/ϕ is larger than β_y or not: that is, the hourglass effect is strong or not. If hourglass effect becomes serious ($\sigma_x/\phi > \beta_y$), luminosity degrades without crab waist. In $\sigma_x/\phi < \beta_y$, luminosity without crab waist does not degrade. Figure 5 shows luminosity and beam-beam parameter as a function of current in this condition. The gain of the crab waist is not remarkable for $\sigma_x/\phi < \beta_y$ in this current range. The gain is higher at higher current, and at further large current, corresponds to the beam-beam parameter > 0.1 , the gain of the crab waist is remarkable even $\sigma_x/\phi < \beta_y$. Figure 6 shows luminosity in transverse tune space. Clear synchro-beta resonance line $2\nu_x + \nu_s = \text{integer}$ is seen in the both case of crab waist on and off.

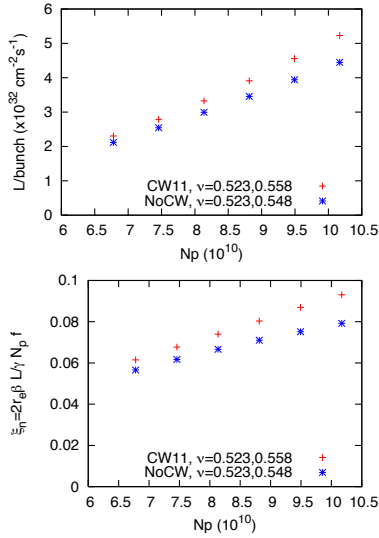


Figure 5: Luminosity and beam-beam parameter for NanoBeam-2 as function of the beam current.

Strong-strong simulation

For strong-strong simulation, both beams are represented by macro-particles. Both beams are sliced into many pieces, and collide slice by slice with solving Poisson equation during the interaction. Poisson equation has to be solved many times, square of the number of slice. Since the number of slice is 100-200, Poisson equation is solved 10^4 times per collision. The radiation damping time is 4000 turns for KEKB, therefore the collision has to be repeated 10^4 times, with the result that the total number is $10^8 - 10^9$ times. Note that the potential is calculated for two beams, and twice per slice to interpolate potential along z [7]. KEK super computer HITACHI SR11000 computes one Potential solution in 10 ms. Simulation of the present KEKB in which Poisson equation is solve $10^2 \times 10^4 \text{ turn} = 10^6$ times, takes a few hours.

Beam-Beam Interaction

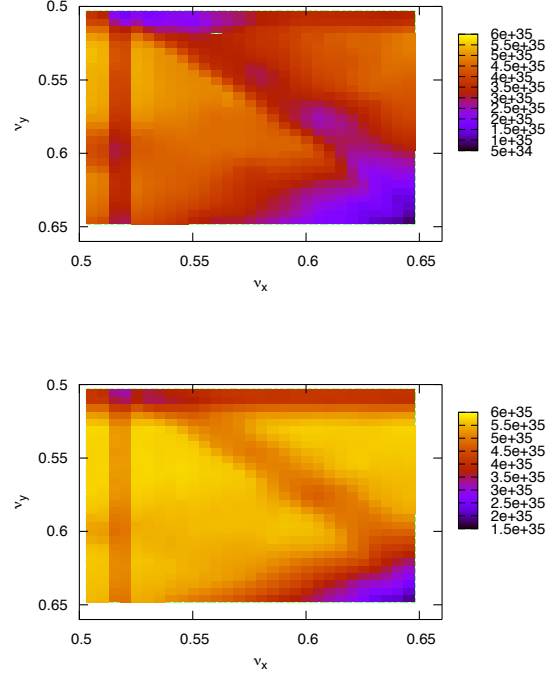


Figure 6: Luminosity on the transverse tune space. Top and bottom are without and with crab waist, respectively.

Soft Gaussian approximation reduces the simulation time extremely. Therefore we adopt a mixed method of the PIC solver and soft Gaussian approximation. When colliding two beam (slice) separation is closer than $5\sigma_x$, PIC solver is used, otherwise soft Gaussian approximation is used. Figure 7 shows the luminosity given by the strong-strong simulation. The luminosity somewhat degrade from the design value. The beam-beam parameter obtained by the luminosity is 0.08. No coherent motion was seen. The horizontal size did not change, while vertical size was enlarged. This result means that there is no serious problem for beam-beam dynamics point of view.

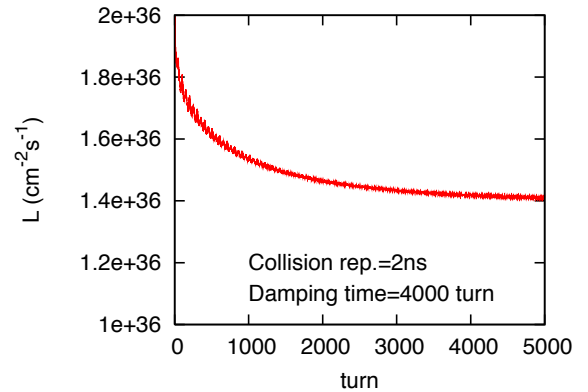


Figure 7: Luminosity evolution given by the strong-strong simulation.

Effect of the micro-wave instability

Bunch current in KEKB upgrade is higher than the present current of KEKB. Bunch lengthening and micro wave instability have been observed even in KEKB. Though the chamber is more carefully designed, residual impedance or CSR contribution may cause micro-wave instability. Actually missing impedance exist in the present KEKB [8]. The horizontal beam-beam force integrated along the bunch length is Bassetti-Erskine type for tri-Gaussian distribution in x-y-z plane. When the micro-wave instability arises, the transverse beam-beam force is distorted and fluctuated in the nano-beam scheme as shown in Figure 8.

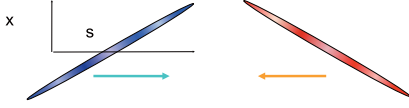


Figure 8: Collision of beams with a longitudinal density modulation.

The combined effect of the beam-beam and micro-wave instability is studied with the weak-strong simulation. The weak-beam is represented by macro-particles on 6-dimensional phase space as is done generally. The strong-beam is represented by macro-particles on the longitudinal phase space, where the transverse distribution is fixed and determined by the design σ_x and σ_y . Longitudinal wake field is introduced for the strong beam. When the strong beam is unstable in longitudinal, the weak beam experiences fluctuating beam-beam force from the strong beam. Figure 9 shows the longitudinal profile of the strong beam, and luminosity and weak beam size evolution. The impedance (wake) is the resonator model used by Y. Cai [9],

$$Z(\omega) = \frac{R_S}{1 + iQ \left(\frac{\omega_R}{\omega} - \frac{\omega}{\omega_R} \right)} \quad (5)$$

where $R_S = 2.5 \times 10^{-6}$ s/m, $\omega_R = 2\pi \times 31.3$ GHz and $Q = 1$. The luminosity is smaller than the design value in the simulation, $8 \rightarrow 6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. Bunch lengthening is the reason of the luminosity degradation, but no other complex effects is seen. There is not blow-up in the transverse beam size. Since the longitudinal profile is lengthening but is stable, the beam-beam force does not fluctuate. The resonator impedance may be mild for the micro wave instability. Actual impedance is more complex, thus the beam profile may fluctuate. The simulation should be performed with a realistic impedance.

COMPUTING AT KEK

Common memory or distributed memory

Two super computers are used in KEK. One is a common memory type of parallel computer, HITACHI SR11000. It consists of 16 nodes, where each node equips 16 CPU's Beam-Beam Interaction

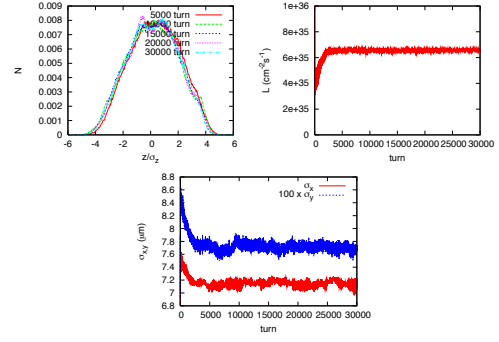


Figure 9: Longitudinal profile for the strong beam. Luminosity and transverse beam size of the weak beam.

(POWER 5) with a common memory (24 GB). The total power is 2 TFlops.

Another is the distributed memory type of computer, Blue Gene. It consist 10,000 nodes, which contain 2 PowerPC440 in each, and the total power is 57 TFlops.

Simulations for accelerators have been carried out mainly using SR11000, even though the total CPU power is lower.

The potential solver is called 10^2 times per collision for the present KEKB. It is called 10^4 or more for the upgrade of KEKB. The total numbers of Poisson solve is 10^6 and $10^8 - 10^9$ for the present KEKB and KEKB upgrade, respectively. It is similar number for JPARC-MR. A potential calculation including the distribution to all CPU's should be finished less than 1 ms to complete these simulations in a reasonable computation time, < 100 hours.

HITACHI SR11000m in which communication between CPU is via memory, solves potential in 10 ms. For Blue Gene, the overhead of network communication should be cared.

PIC simulation is carried out as follows,

1. Particle loops are paralleled.
2. Count distribution $\rho(x_i, x_j)$ and take summation for all nodes.
3. Potential $\varphi(x_i, y_j)$ is calculated by solving 2D Poisson equation with parallel for the mesh.
4. $\varphi(x_i, y_j)$ is distributed to all nodes, and is used to track particles.

When the mesh is 128×128 with 8 Byte memory, 128 kB data has to be communicated in the process of (2) and (4) per one potential calculation. This communication is performed via memory for SR11000, but via network for Blue Gene. The network overhead is estimated by the help of Dr. Doi (IBM-Japan). MPI_Allreduce for 32 node spend 15 sec for 10^4 times: that is 1.5 ms. Since the data communication is performed along "tree" structure, it is twice for 1024 node: i.e., $\log_2 1024 / \log_2 32 = 2$. Anyway the communication time has already over the required time 1 ms per one potential calculation.

The common memory type of supercomputer SR11000 is still useful for particle in cell simulations of circular accelerators. However computing power of super parallel computers based on the distributed memory increases more and more. Super parallel computers are trend of the computing. KEK super computers are replaced by new one with 1 PFlops in 2011. RIKEN computer center with 10 PFlops starts at 2012. We have to keep up the trend in spite of the network overhead in our simulation scheme.

ACKNOWLEDGMENT

The author thanks KEKB commissioning team, and KEKB upgrade and superB design teams. This work is supported by the Large Scale Simulation Program No.09-04 (FY2009) of High Energy Accelerator Research Organization (KEK).

REFERENCES

- [1] K. Ohmi, proceeding of EPAC08.
- [2] Y. Seimiya and K. Ohmi, proceedings of PAC09. To be published.
- [3] Y. Ohnishi et al., to be published in Phys. Rev. ST-AB.
- [4] K. Ohmi et al., to be published.
- [5] D. Zhou et al., KEK Preprint 2009-10.
- [6] P. Raimondi et al., proceedings of PAC07.
- [7] K. Ohmi et al., Phys. Rev. ST-AB 7, 104401 (2004).
- [8] D. Zhou, in this proceedings.
- [9] Y. Cai et al., Phys. Rev. ST-AB 12, 061002 (2009).

RECENT ADVANCES OF BEAM-BEAM SIMULATION IN BEPCII*

Y. Zhang[†], IHEP, Beijing, China

Abstract

The luminosity of BEPCII (the upgrade project of Beijing electron-positron collider) have reached $3.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ @ 1.89 GeV in May 2009. In this paper we'll compare the beam-beam simulation results with the real machine. In the case the single bunch current is lower than 8mA, the simulation coincides well with the real. Some phenomenon related to synchro-betatron resonances during machine tuning and simulation is shown. The tune is close to half integer help us increase luminosity, however the detector background increases at the same time. It is believed that the beam-beam dynamic effect result in the drop of the dynamic aperture. We also study the possible luminosity contribution from the crab waist scheme in BEPCII.

INTRODUCTION

BEPCII is an upgrade project from BEPC. It is a double ring machine. Following the success of KEKB, the crossing scheme was adopted in BEPCII, where two beams collide with a horizontal crossing angle $2 \times 11 \text{ mrad}$. The design luminosity of BEPCII is $1.0 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ at 1.89 GeV, about 100 times higher than BEPC. The construction started in January 2004 and completed in June 2008 when the detector is positioned. The luminosity was only achieved $1.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, since the two profile monitors in the positron ring excite very strong longitudinal instability [1]. When the two monitors were removed, $2.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ was achieved. In May 2009, we decided to move the horizontal tune more closer to half integer, which help us achieve $3.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, and the project was reviewed by the government in July 2009. Table 1 shows the main design and achieved parameters in BEPCII.

The beam-beam code used in our simulation is a Particle-in-Cell code [2]: (1) the transport map in the arc is linear approximation which is same as Hirata's BBC code where the synchrotron radiation and quantum excitation is included, (2) the beam-beam force is calculated by solving Poisson equation using FFT, (3) finite bunch length effect is included by longitudinal slices, and the interpolation scheme is used to improve the convergence of slice number [3], (4) the finite horizontal crossing angle is included by Lorentz Boost [4].

In the following, we'll compare the simulated beam-beam limit with the achieved in the real machine. According to the simulation, the synchro-betatron resonances

Table 1: Parameters of BEPCII (Design and Achieved)

	Design	Achieved
E [GeV]	1.89	1.89
C [m]	237.53	
N_b	93	70
I_b [mA]	9.8	8
\mathcal{L} [$\times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$]	10	3.0
ξ_y	0.04	0.025
θ_c [mrad]	2×11	
β_x^* / β_y^* [m]	1 / 0.015	
ϵ_x / ϵ_y [nm]	144 / 2.2	
σ_z [cm]	1.5	
σ_e	5.16×10^{-4}	
ν_x / ν_y	6.53 / 7.58	6.51 / 5.58
ν_s	0.034	0.032
τ_x / τ_y [turn]	31553 / 31553	
τ_s [turn]	15777	

would be excited in some tune region, and similar phenomenon appears in the tune scan of real machine. We'll also show that the dynamic effect reduce the aperture in the near half-integer region. We also study the possible luminosity contribution of the crab-waist scheme in our machine. At last a summary and discussion is presented.

BEAM-BEAM LIMIT

The beam-beam parameter is defined as

$$\xi_u = \frac{Nr_e}{2\pi\gamma} \frac{\beta_u^0}{\sigma_u(\sigma_x + \sigma_y)} \quad (1)$$

where N is the particle number per bunch, r_e the classical electron radius, γ the relativistic factor and it should be noted that β^0 is unperturbed beta function and σ is perturbed beam size. If we do not consider the luminosity loss caused by finite bunch length and crossing angle, the bunch luminosity can be expressed as

$$L = \frac{N^2 f_0}{4\pi\sigma_x\sigma_y} \quad (2)$$

where f_0 is the revolution frequency, and it should be noted that σ is perturbed beam size. For flat beams $\sigma_y \ll \sigma_x$, the achieved beam-beam parameter can be expressed with bunch luminosity as

$$\xi_y = \frac{2r_e\beta_y^0}{N\gamma} \frac{L}{f_0} \quad (3)$$

* Work supported by National Natural Sciences Foundation of China (10725525 and 10805051)

[†] zhangy@ihep.ac.cn

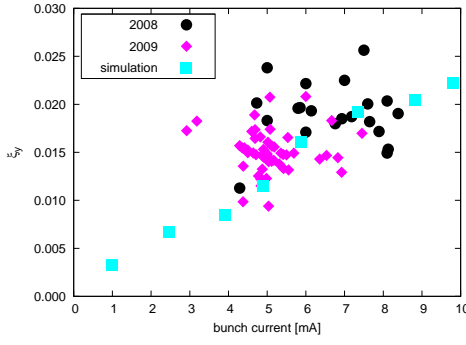


Figure 1: The achieved and simulated beam-beam parameter for $\nu_x \approx 0.53$. 2008 and 2009 refers to that before and after the removal of the profile monitors respectively.

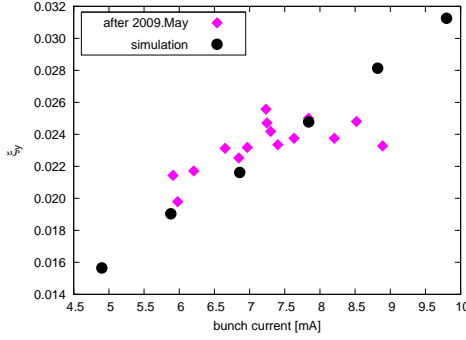


Figure 2: The achieved and simulated beam-beam parameter for $\nu_x \approx 0.51$.

Figure 1 shows the beam-beam limit for $\nu_x \approx 0.53$. There exists clear difference between 2008 and 2009, which is due to the data error. Only one bunch is used to tune the machine in 2008, and we would like to inject more bunches (5 or 10) in 2009. The latter is more credible. The maximum beam-beam parameter is close to 0.025 by simulation, and in the real machine 0.015-0.020 can be achieved stably.

Figure 2 shows the beam-beam limit for $\nu_x \approx 0.51$. The maximum beam-beam parameter is about 0.035 by simulation, and in the real machine 0.022-0.025 can be achieved stably.

In both cases, the achieved beam-beam parameter can be greater than simulation by optimization for lower bunch current. However the real is less than the simulation when $I_b > 8\text{mA}$. One possible explanation is the crosstalk between the beam-beam kick and the nonlinear map in the arc, since there is no cells in the arc and only 4-groups sextupoles are used to correct the chromaticity.

SYNCHRO-BETATRON RESONANCES

There are two kind of synchro-betatron resonances in our machine: one is caused by beam-beam kick, which can be seen in Figure 3 during machine tuning and in Figure 4 by simulation, the corresponding resonance line is $2\nu_x + \nu_s = n$.

Beam-Beam Interaction

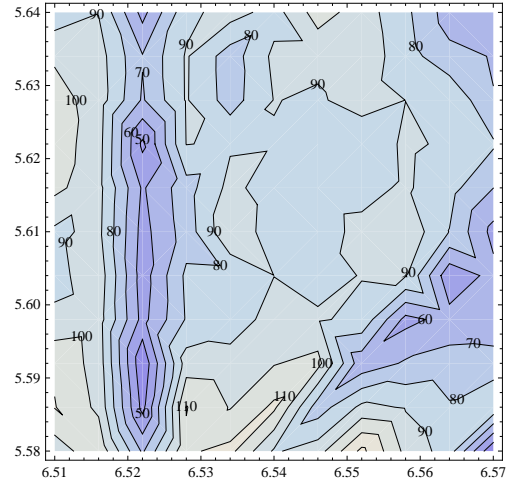


Figure 3: Luminosity versus tune of the electron ring.

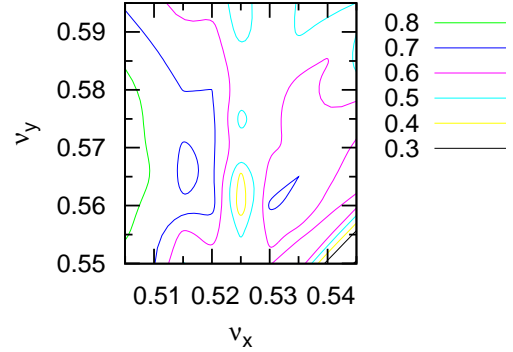


Figure 4: Luminosity versus tune by simulation.

The other is $2\nu_x + \nu_s = n$ which is more important during the lattice design near $\nu_x \sim 0.51$, since it may limit the dynamic aperture. In fact the beam-beam effect is also very effective to the resonance. In the first stage when we move the horizontal tune to 0.51, the luminosity is very sensitive to the knobs: RF voltage, tune and orbit etc. It is found that the sextupole configuration is not good, see Figure 5. After the problem is fixed, the tuning knob is normal.

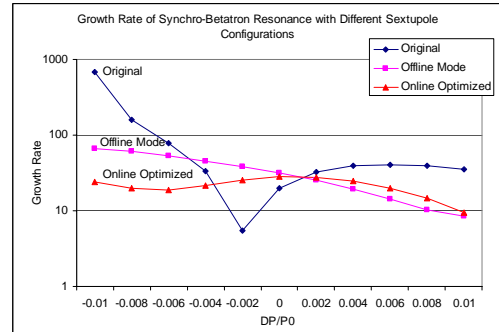


Figure 5: Growth rate of synchro-betatron resonance for different sextupole configurations.

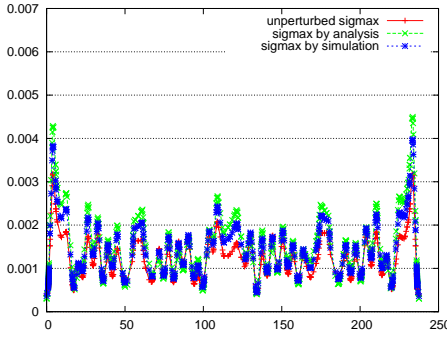


Figure 6: The perturbed horizontal beam size along the ring for $\nu_x = 0.53$ and $I_b = 8\text{mA}$

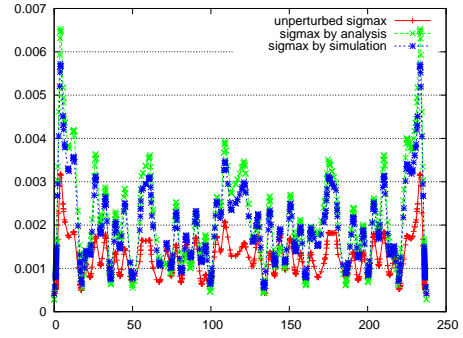


Figure 7: The perturbed horizontal beam size along the ring for $\nu_x = 0.51$ and $I_b = 8\text{mA}$

DYNAMIC EFFECT

The beam-beam interaction perturbs the twiss parameter along the ring, which is the so-called dynamic beta effect. With linear approximation, the beam-beam force is treated as linear focusing force in both transverse directions, and the perturbed beta function at IP is

$$\beta = \frac{\beta_0}{\sqrt{1 + 4\pi\xi \cot \mu_0 - 4\pi^2\xi^2}} \quad (4)$$

where β_0 is unperturbed and ξ is the achieved beam-beam parameter. The transverse emittance is also perturbed, which is the so-called dynamic emittance,

$$\epsilon = \frac{1 + 2\pi\xi \cot \mu_0}{\sqrt{1 + 4\pi\xi \cot \mu_0 - 4\pi^2\xi^2}} \epsilon_0 \quad (5)$$

The perturbed parameters can be calculated by iteration.

When our machine is running near $\nu_x \sim 0.53$, the detector's background is low enough to take data. And we can reduce the background by tuning the horizontal orbit in the IR region. When the horizontal tune is more closer to half integer, the luminosity achieves $3.0 \times 10^{32} \text{cm}^{-2} \text{s}^{-1}$ but the detector cannot work due to high background. Tuning the horizontal orbit cannot improve the background.

We use two methods to study the dynamic effect. One is the linear theory analysis, the other is making use of the simulation code: the macroparticles after multi-turn beam-beam kick is transported along the ring, and the RMS bunch size is calculated by statistics. Figure 6 and Figure 7 shows the horizontal beam size along the ring for $\nu_x = 0.53/0.51$ respectively. It is very clear that the RMS size enlarges near the horizontal final focus magnet along with the tune's close to half integer. It should be noted that the aperture is only about 12σ without collision. It seems that we could not make full use of the high luminosity region near half integer without modification of IR.

CRAB WAIST SCHEME

There is still long way for us to achieve the design luminosity. The crab waist scheme proposed by Raimondi [6] Beam-Beam Interaction

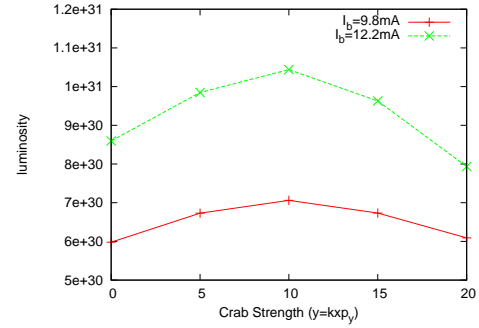


Figure 8: Luminosity versus the crab transformation strength with design parameters.

may help us, so it's necessary to study the feasibility in our machine. There is 3 steps to implement the scheme:

1. large Piwinski angle $\phi = \sigma_z \tan \theta / \sigma_x$, however it's only 0.43 in BEPCII
2. reduce vertical beta, which is comparable to the overlap area $\beta_y \approx \sigma_x / \theta$, and $\beta_y = 0.015 < \sigma_x / \theta = 0.034$ in BEPCII
3. crab waist transformation $H = \frac{1}{4\theta} x p_y^2$, which means $H = 22 x p_y^2$ in BEPCII

It seems that we're not very lucky to use this scheme.

First we try to determine the optimum strength of crab waist transformation, which is shown in Figure 8. The optimum is only ~ 0.2 of the full crab rotation. The beam-beam limit with crab on is shown in Figure 9 for $\nu_x = 0.53$. The achieved beam-beam parameters is increased from 0.025 to 0.030, which means the maximum luminosity is $\sim 11 \times 10^{30} \text{cm}^{-2} \text{s}^{-1}$. When ν_x is more closer to half integer, the luminosity contribution is not as good as that 0.53. It can be concluded that the scheme could help us to some extent, however we still could not find a lattice solution due to dynamic aperture limitation coming from the strong sextupoles.

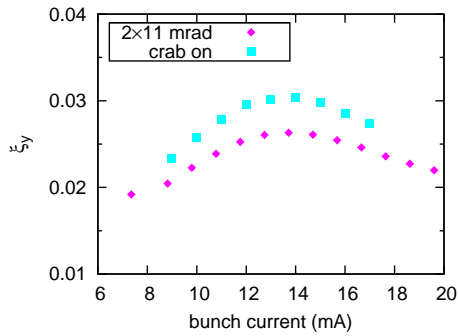


Figure 9: Beam-beam parameter versus bunch current with crab on/off.

SUMMARY

The machine has been tuning with solenoid on since June 2008. In the real machine, $\xi_y = 0.015 \sim 0.020$ is achievable near $\nu_x = 0.53$, however the simulated beam-beam limit is 0.025. Near $\nu_x = 0.51$, $\xi_y = 0.020 \sim 0.025$ is achievable, and the simulated beam-beam limit is 0.035. Not only the achieved but also the simulated is not satisfying. It is the crossing angle which reduces the beam-beam limit. We still don't know what cause the difference between the simulated and the real. It is suspected that the nonlinear map in the arc contributes to the difference. The element-by-element tracking in the arc instead of 6×6 linear map is scheduled.

By the simulation study, we notice that $2\nu_{x,\pi} + 2\nu_s = n$ could excite synchro-betatron resonances and lead lumi-

nosity loss. During tune scan of the real machine, we find similar resonance line, however it seems not very strict to conclude they prove each other. Both the dynamic aperture and beam-beam effect is sensitive to another resonance line $2\nu_x + \nu_s = n$.

The dynamic effect reduces the aperture near half integer, which makes the high luminosity region cannot be used to take data till now. On the other hand, the crab waist scheme would not increase luminosity very much. In fact lower beta at IP will help us increase luminosity with or without crab. That is to say it's a good choice to enlarge the aperture near the horizontal final focus magnet. In one words there is still long way to achieve the design luminosity.

REFERENCES

- [1] C. Zhang, L. Ma and J.Q. Wang, "Construction and Commissioning of BEPCII", in Proceedings of PAC 2009.
- [2] Y. Zhang, K. Ohmi and L. Chen, "Simulation study of beam-beam effects", Phys. Rev. ST Accel. Beams 8, 074402 (2005).
- [3] K. Ohmi, M. Tawada, Y. Cai, S. Kamada, K. Oide, and J. Qiang, "Beam-Beam Limit in e^+e^- Circular Colliders", Phys. Rev. ST Accel. Beams 7, 104401 (2004).
- [4] K. Hirata, "Analysis of Beam-Beam Interactions with a Large Crossing Angle", Phys. Rev. Lett. 74, 2228 (1995).
- [5] K. Hirata, "BBC: Program for Beam-Beam Interaction with Crossing Angle", <http://wwwslap.cern.ch/collective/hirata>
- [6] P. Raimondi, 2nd SuperB Workshop, Frascati, 2006

MODELING OF ULTRA-COLD AND CRYSTALLINE ION BEAMS*

H. Okamoto[#], H. Sugimoto, Hiroshima University, Hiroshima, Japan
 Y. Yuri, Takasaki Advanced Radiation Research Institute, JAEA, Gumma, Japan
 M. Ikegami, Keihanna Research Laboratory, Shimadzu Corporation, Kyoto, Japan
 J. Wei, Tsinghua University, Beijing, China

Abstract

An ultimate goal in accelerator physics is to produce a “zero-emittance” beam, which is equivalent to making the beam temperature the absolute zero in the center-of-mass frame. At this limit, if somehow reached, the beam is Coulomb crystallized. Schiffer and co-workers first applied the molecular dynamics (MD) technique to study the fundamental features of various Coulomb crystals. Their pioneering work was later generalized by Wei et al. who explicitly incorporated discrete alternating-gradient (AG) lattice structures into MD simulations. This paper summarizes recent numerical efforts made to clarify the dynamic behavior of ultra-cold and crystalline ion beams. The MD modeling of beam crystallization in a storage ring is reviewed, including how one can approach the ultra-low emittance limit. Several possible methods are described of cooling an ion beam three-dimensionally with radiation pressure (the Doppler laser cooling).

INTRODUCTION

Mutual Coulomb interactions among stored particles play a substantial role in beam dynamics especially when those particles are densely distributed in phase space [1,2]. The volume occupied by the particles in six-dimensional phase space is called “emittance” that can directly be linked to the beam “temperature” measured in the center-of-mass frame. In theory, the emittance of a beam converges to zero (except for quantum noises) at the ultra-low temperature limit [3]. It can thus be said that space-charge-induced phenomena become more prominent as the emittance or temperature goes down.

The emittance is approximately conserved if the rate of Coulomb collisions between individual particles is low [4]. That is basically due to the Hamiltonian nature of lattice elements (magnets, cavities, etc.) that only produce conservative forces. In practice, however, we almost always prefer a beam with a lower emittance. To meet this general requirement, we must introduce dissipative interactions into the system to “cool” the beam. Needless to say, the ultimate goal of cooling is to make the beam temperature the absolute zero.

Many questions arise, however: is it really possible in principle to establish a zero-emittance state? Can such an ultimate state, if it exists, be stable? How does the beam look like at that limit? These questions have been answered since the mid 1980’s [5-13]. Schiffer and co-workers first carried out systematic theoretical researches on strongly-coupled non-neutral plasmas by employing

the MD technique [5-9], but their work was based on the smooth approximation that may eliminate possible realistic effects in cooler storage rings. This fact motivated the later, more sophisticated MD work by Wei et al. who took discrete lattice structures into account [10-13]. Their MD simulations actually revealed essential differences between ultra-cold states in a uniform channel and those in an AG channel. Through all these continuous efforts, it is now strongly believed that stable zero-emittance beams can exist, at least, in theory.

The purpose of this paper is to give a brief review of computer modelling of ion beams in the ultra-low temperature regime. After showing the primary conditions to form and maintain a crystalline ion beam in a storage ring, we outline the MD method employed generally for crystalline-beam studies. We then proceed to the description of several cooling models including the Doppler laser cooling [14,15] that is currently the only solution toward beam crystallization. Although the Doppler limit is actually very close to the absolute zero, the powerful laser cooling force only operates in the longitudinal direction of beam motion [16,17]. It is thus necessary to somehow make it work three-dimensionally. For this purpose, we here consider the *resonant coupling method* (RCM) that can easily be implemented in a real storage ring [18,19]. Finally, a unique storage-ring lattice free from momentum dispersion [20,21] is described which can resolve the problem of “tapered cooling” [12,22].

CRYSTALLINE BEAMS

Schiffer, Hasse and others numerically demonstrated that a system of many identical charged particles confined by a time-independent harmonic potential exhibits a spatially ordered configuration at the low-temperature limit [5-9]. This phenomenon is referred to as “Coulomb crystallization”. In this unique state of matter, the Coulomb repulsion among particles just balances with the external focusing potential. Suppose a coasting ion beam, for instance. If the line density is sufficiently low, all ions are aligned along the design beam orbit at equal intervals (string crystal). By increasing the line density, we can convert this one-dimensional (1D) configuration into a two-dimensional (2D) (zigzag crystal). The zigzag crystal is eventually transformed to a three-dimensional (3D) figure (shell crystal) if we put more ions in the beam. The threshold line density from a particular crystalline structure to another can be estimated from the Hasse-Schiffer theory [8]. Similar structural transitions occur even for bunched beams. Figure 1 shows a typical multi-shell Coulomb crystal predicted by a MD simulation.

*Work supported in part by Grants-in-Aid for Scientific Research.
[#]okamoto@sci.hiroshima-u.ac.jp

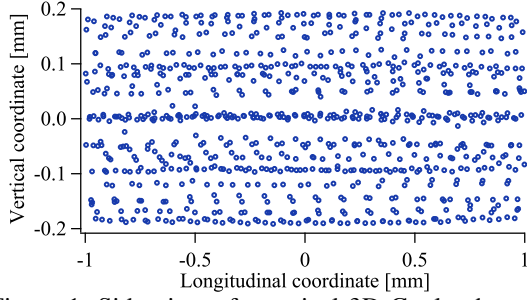


Figure 1: Side view of a typical 3D Coulomb crystal numerically obtained with a MD simulation code.

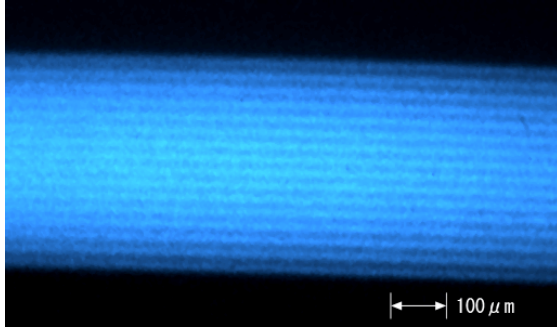


Figure 2: Laser-induced fluorescence from a multi-shell Coulomb crystal formed in a linear Paul trap at Hiroshima University.

When the external force is uniform and time-independent as assumed in early papers on Coulomb crystals [5-9], each particle is frozen at a certain fixed point in the rest frame once the beam reaches a crystalline ground state. The situation is definitely different in a storage ring where the beam experiences periodic driving forces all the time. The periodic change of the focusing potential excites resonant instability under a specific condition satisfied. It is particularly important to avoid crossing *linear* (second order) resonance stopbands throughout a cooling process toward the absolute zero. To eliminate the possibility of dangerous resonance crossing, the bare betatron phase advance per lattice period must be less than 90 degrees [12,23]:

$$v_{x(y)} < N_{sp} / 4, \quad (1)$$

where (v_x, v_y) are the horizontal and vertical betatron tunes, and N_{sp} denotes the lattice superperiodicity of the ring. The time-dependency of the external potential also enhances heating from *intrabeam scattering* [24,25]; we thus need to provide sufficiently strong transverse cooling force to overcome this effect. Another primary condition pointed out by Wei et al. is the following:

$$\gamma < \gamma_T, \quad (2)$$

which means that the beam energy γ (the Lorentz factor) must be below the transition energy γ_T of the ring.

Various Coulomb crystals have now been realized *experimentally* in ion traps [26-28]. It is actually straightforward to produce an ultra-cold ion plasma in a

trap by using the Doppler cooling technique. Figure 2 is a laser-induced fluorescence image of a $^{40}\text{Ca}^+$ plasma laser-cooled in a linear Paul trap. There is no doubt that a multi-shell crystalline structure has been formed. It is indeed possible to make string and zigzag crystals as well. This experimental evidence probably convinces many people that it must be possible to crystallize an ion beam in a similar way. In reality, however, nobody has succeeded in producing a crystalline beam in spite of serious attempts by European groups [16,17,29,30]. Although “moving” Coulomb crystals were generated in a ring-shaped Paul trap system [31,32], crystalline beams in a real storage-ring accelerator have the nature more complex than Coulomb crystals in such a compact low-energy device. In addition, the lattice parameters are not so flexible which often prevents us from approaching an ultra-low temperature state [33]. How to achieve efficient 3D laser cooling in a storage ring is also a big issue.

MD APPROACH

Periodic Boundary Condition

The Particle-In-Cell (PIC) algorithm has often been adopted to study the dynamic behavior of space-charge-dominated beams. Beam crystallization is, however, clearly beyond the scope of the PIC method that relies on spatial meshes and macro-particles. Since collective interactions over the whole beam and Coulomb collisions among individual particles both play an important role at low temperature, we have to compute the space-charge potential as precisely as possible. The best way is to simply sum up the Coulomb potentials of all particles, but that is indeed impractical even with a modern high-performance computer when the beam consists of a large number of particles.

In MD simulations, the so-called periodic boundary condition is employed to save computing time. We first slice the beam in the longitudinal direction and load some number of “real” particles (not “macro” particles) in the reference cell we are looking at. Within this particular cell, interparticle Coulomb interactions are calculated from the potential of the form

$$\phi_{short}^{(j)} = \frac{1}{\sqrt{(x-x_j)^2 + (y-y_j)^2 + (z-z_j)^2}}, \quad (3)$$

where (x, y, z) is the spatial coordinates of a particle, and (x_j, y_j, z_j) the coordinates of one of the other particles. When the cell contains n particles, the short-range Coulomb forces acting upon the particle at (x, y, z) are evaluated by summing up Eq. (3) over all $n-1$ partners.

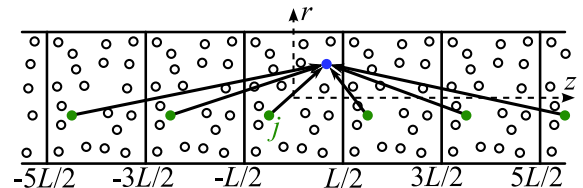


Figure 3: Periodic boundary condition

If n is not too large, the CPU time required for this process should be reasonable. Now, the question is how to compute the long-range Coulomb forces coming from other cells. The MD algorithm assumes that all longitudinal cells have an identical particle distribution at each integration step, as illustrated in Fig. 3. Then, the long-range Coulomb potential generated by the j th particle's images in all cells can be reduced to the Ewald-type integral [34]:

$$\phi_{\text{long}}^{(j)} = \frac{2}{L} \int_0^\infty \frac{\cosh(kz^{(j)}/L) J_0(kr^{(j)}/L) - 1}{e^k - 1} dk, \quad (4)$$

where L is the cell length, J_0 is the Bessel function of zero-th order, $z^{(j)} = |z - z_j|$ ($-L/2 < z^{(j)} < L/2$), and $r^{(j)} = \sqrt{(x - x_j)^2 + (y - y_j)^2}$. Although the beam is not perfectly uniform in the longitudinal direction, this should be a good approximation because the total long-range potential is probably insensitive to the details of the image charge distribution. The cell length L should be neither too short nor too long, so that we have a proper number of real particles in the reference cell. It has been confirmed that, as long as the ratio n/L is fixed, we always reach the same crystalline structure.

Beam-Frame Hamiltonian

In order to apply the periodic boundary condition, we have to observe the beam motion in the center-of-mass frame. Strictly speaking, the general relativity formalism is required to derive the correct equations of motion because the beam is not in an inertial system. Let us here suppose an ion beam travelling at the speed βc with c being the light velocity. For simplicity, consider only dipole and quadrupole magnetic fields. Then, the beam-frame Hamiltonian is approximately given by [35]

$$H = \frac{p_x^2 + p_y^2 + p_z^2}{2} - \frac{\gamma}{\rho} x p_z + \frac{x^2}{2\rho^2} - \frac{K(s)}{2} (x^2 - y^2) + \frac{r_p}{\beta^2 \gamma^2} \phi, \quad (5)$$

where the scaled canonical variables are $(x, y, z; p_x, p_y, p_z)$, ρ is the local curvature of the design beam orbit, $K(s)$ corresponds to the focusing gradients of quadrupole magnets, r_p is the classical radius of the particle, and the independent variable is the path length $s = \beta \gamma c \cdot t$ with t being proper time. The scalar Coulomb potential is calculated in a MD code from $\phi = \sum_j (\phi_{\text{short}}^{(j)} + \phi_{\text{long}}^{(j)})$. Not surprisingly, H has the form identical to the well-known Hamiltonian for standard beam-orbit theories. We have developed a MD code “CRYSTAL” that integrates this Hamiltonian motion in a symplectic manner. Solenoid magnets, radio-frequency (rf) cavities, and other insertion elements can also be incorporated in the code, if necessary.

When the beam is bunched, the CRYSTAL code automatically set the MD cell length equal to the rf bucket size, assuming that all bunches have an identical particle

distribution. We have confirmed that, at low temperature, the Coulomb potentials from other bunches are generally quite weak.

COOLING MODELS

Cooling interactions must be introduced separately from the Hamiltonian framework because they are not conservative. Several cooling models can be considered in the CRYSTAL code to study the dynamics of cold beams.

Linear Friction

The simplest cooling force is the linear 1D friction defined by

$$\Delta p_q = -f_q p_q^{\text{in}} \quad (q = x, y, z), \quad (6)$$

where f_q is the constant friction coefficient, and Δp_q stands for the momentum change in q -direction before and after the cooling section; namely, $\Delta p_q = p_q^{\text{out}} - p_q^{\text{in}}$. This cooling force just tries to equalize the velocities of all particles. The linear friction can certainly cool regular *hot* beams, but at very low temperature, it works as a heating source. For example, unlike in a uniform focusing channel, the betatron oscillations of particles never vanish in an AG lattice as long as the beam has finite transverse extent.

Tapered Cooling

As mentioned above, too strong a transverse linear friction eventually starts to heat up the beam due to the oscillatory nature of the stationary state. The same argument also applies to the longitudinal motion because of the existence of dipole fields in a storage ring. Once a crystalline ground state is reached, particles with different horizontal positions follow slightly different closed orbit every turn. On the other hand, their revolution frequencies must be identical to maintain the ordered structure. This means that the average longitudinal speeds of those particles are different depending on the horizontal coordinates. Therefore, the linear friction as in Eq. (6) again operates as a heating source in the longitudinal direction. To compensate the *dispersive* heating at ultra-low temperature, the *tapered* cooling force is necessary [12,22]:

$$\Delta p_z = -f_z (p_z^{\text{in}} - C_{xz} x^{\text{in}}), \quad (7)$$

where C_{xz} is the tapering factor that depends on the lattice design [36]. The tapered force yields not only longitudinal but also horizontal cooling effects when it is applied to the beam at a position with finite momentum dispersion [22].

Laser Cooling

The longitudinal dissipative force $F_{+(-)}$ generated by a laser light co-propagating (counter-propagating) with an ion beam can be expressed as [14,15]

$$F_{\pm} = \pm \frac{1}{2} \hbar k_L \Gamma \frac{S_L}{1 + S_L + (2\delta_{\pm} / \Gamma)^2}, \quad (8)$$

where Γ is the natural linewidth of the cooling transition, k_L is the wave number of the laser, S_L is the saturation parameter, and $\delta_{+(-)}$ is the detuning of the co-propagating (counter-propagating) laser frequency from the natural resonant frequency ω_0 of the ion. When the co-propagating (counter-propagating) laser frequency in the laboratory frame is $\omega_{+(-)}$, we have the Doppler-shifted detuning $\delta_{\pm} \approx \omega_{\pm} \gamma [1 \mp \beta(1 + p_z/\gamma)] - \omega_0$. Assuming a Gaussian laser, the saturation parameter is given by $S_L = S_0 \exp[-2(x^2 + y^2)/w^2]$, where S_0 corresponds to the peak saturation parameter on the axis of laser propagation, and w is the laser spot size that depends on the Rayleigh length z_R . When the center of a laser cooling section is located at the longitudinal coordinate s_0 , we can write $w(s) = w_0 \sqrt{1 + \gamma^2(s - s_0)^2 / z_R^2}$ with w_0 being the waist size of the laser.

The Doppler limit T_D of laser cooling is determined by the balance between the dissipative force in Eq. (8) and diffusive heating originating from the random nature of photon emission and absorption. In the simple 1D case, T_D can be evaluated from the formula $k_B T_D / 2 = \hbar \Gamma / 4$, where k_B is the Boltzmann constant. In the CRYSTAL code, a proper amount of random kick is applied to the beam in every integration step to include the diffusive heating effect. To check the reliability of our laser-cooling algorithm, we performed test simulations changing some fundamental parameters. An example is shown in Fig. 4 where the final equilibrium temperature reached in our laser-cooling simulation is plotted as a function of integration time step. The horizontal straight line corresponds to the theoretical Doppler limit that agrees fairly well with the CRYSTAL simulation results unless the integration step is too large.

RESONANT COUPLING SCHEME

Principle

In order for the Doppler cooling mechanism to be effective, each ion must absorb many photons while it passes through a cooling section. This requirement can readily be met in the longitudinal direction by introducing the laser light along the beam orbit in a straight section. As to the transverse directions, it is practically impossible to achieve efficient, direct laser cooling because we cannot ensure a large overlap between the beam and laser. Unlike an ion plasma in a compact trap, a typical ion beam in a storage ring is much hotter and thinner, which makes it extremely difficult to accomplish 3D laser cooling.

A possible solution to extend the powerful longitudinal laser-cooling force to the transverse dimensions is the use of dynamic coupling that correlates one dimension to the others [18]. Mathematically, what we must do is to create additional linear potentials proportional to $x \cdot z$ and $y \cdot z$

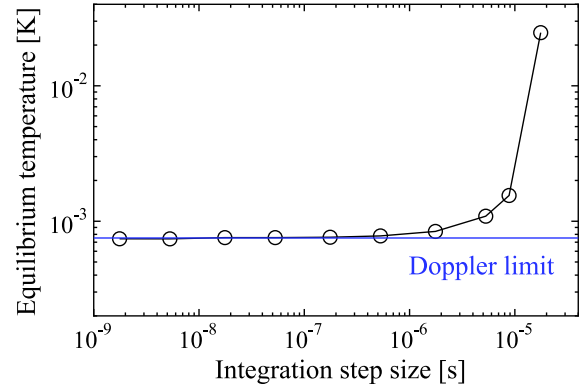


Figure 4: Results of Monte-Carlo simulations to test the laser-cooling model. Low-energy $^{24}\text{Mg}^+$ ions have been assumed in this example.

(or $x \cdot y$). We then move the operating point of the storage ring to excite linear coupling resonances:

$$\nu_x - \nu_y \approx \text{integer}, \quad \nu_x - \nu_z \approx \text{integer}, \quad (9)$$

where ν_z is the synchrotron tune. As theoretically demonstrated in previous papers [18,19], the transverse indirect cooling rate is considerably enhanced under these conditions. The effectiveness of RCM has been partially confirmed in a storage-ring experiment [37] where the method was actually employed to improve the vertical cooling efficiency. In this experiment, a solenoid magnet was turned on to produce the linear $x \cdot y$ coupling.

Coupling Sources

There are several practical ways to provide the linear coupling potential required for indirect transverse cooling. The excitation of horizontal-vertical coupling is particularly easy; we simply put either a skew quadrupole magnet or a solenoid. It is also straightforward to couple the longitudinal motion of a stored particle with the horizontal motion; all we have to do is to place a regular rf cavity at a dispersive position [19]. An alternative solution for longitudinal-transverse coupling is the use of coupling rf cavities operating in a deflective mode [18]. The coupling-cavity scheme is more flexible in controlling the transverse cooling efficiency because it does not rely on the dispersion function of the lattice.

Another interesting option for indirect transverse cooling is the Wien filter [38]. The single-particle motion within this static electromagnetic device approximately obeys the Hamiltonian

$$H_w = \frac{p_x^2 + p_y^2 + p_z^2}{2} + \frac{1}{2} \kappa_x^2 x^2 - \kappa_x x p_z, \quad (10)$$

where κ_x is a constant parameter proportional to the filter voltage. As pointed out in Ref. [38], a sort of tapered force defined by Eq. (7) is naturally developed when we apply a cooling laser to stored ions within the filter. The momentum dispersion must, therefore, be finite in the straight section where this device is located. It is worthy to recall that tapered cooling requires no synchro-betatron resonance to enhance the horizontal cooling rate [22]. We

can thus strongly cool even coasting beams. Efficient vertical cooling can very easily be carried out with RCM by equalizing two betatron tunes (ν_x, ν_y) and then switching on either a solenoid or a skew quadrupole.

MD RESULTS

It has been demonstrated in many previous papers that beam crystallization is theoretically feasible under several conditions satisfied. The necessary conditions include not only Eqs. (1) and (2) but also efficient transverse cooling plus tapering. In practice, however, it is difficult to meet all these requirements simultaneously. For example, the condition (1) cannot strictly be fulfilled unless we execute beam cooling in all lattice periods. If we bunch the beam with a single rf cavity, that also weakly breaks the lattice symmetry, reducing N_{sp} to unity. The bare betatron tune $\nu_{x(y)}$ then has to be less than 0.25 per turn, which is clearly unacceptable in a regular storage ring.

We have performed a number of systematic MD simulations to see how close we can come to a crystalline state with existing technologies. First of all, laser cooling has to be chosen for our final goal because of its ultra-low limiting temperature. Although several other cooling methods are available, none of them can reach the temperature range close to the absolute zero. Then, we need to enhance the transverse cooling rate in some way. The best option for this purpose should be the application of RCM, considering its simplicity; in fact, all we need is to adjust the betatron tunes to proper resonant values if linear coupling sources are present in the ring. Figure 5 shows a MD result in which we have assumed low-energy $^{24}\text{Mg}^+$ ions circulating in the cooler storage ring “S-LSR” at Kyoto University [21]. Clearly, a coasting string crystal has been formed. In this example, a *horizontal* Wien filter is used to activate linear synchro-betatron coupling, while a weak solenoid field is switched on for $x \cdot y$ coupling. S-LSR actually has a solenoid magnet (originally for electron cooling) in one of six straight sections. The bare tunes have been set at $(\nu_x, \nu_y) = (1.46, 1.46)$ to improve the vertical cooling rate with RCM. The ordered configuration in Fig. 5, whose

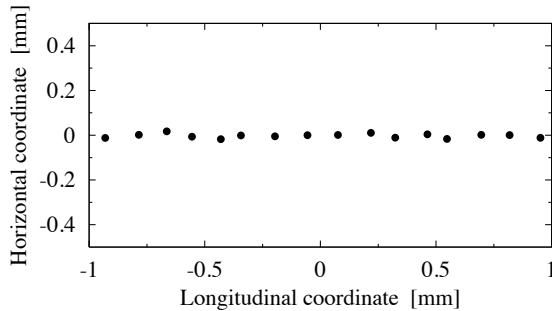


Figure 5: Spatial configuration of an ultra-cold ion beam predicted by a MD simulation in which the realistic laser-cooling model and the lattice of S-LSR have been considered.

normalized root-mean-squared (rms) emittance is below $10^{-12} \text{ m} \cdot \text{rad}$, lasts many turns even without the cooling force. It is also possible to form a zigzag crystalline beam while its stability is not guaranteed. Note that, by switching on an rf cavity, we can establish string and zigzag-like configurations of finite lengths (bunched crystals). In that case, the Wien filter is no longer necessary because, as explained in the last section, a regular rf cavity sitting in a dispersive position naturally induces synchro-betatron coupling.

In contrast to 1D and 2D crystals, none of numerical attempts to produce a stable shell crystal with realistic methods have been successful. There are two primary obstacles that prevent the formation of stable 3D crystalline structures:

- Transverse collective instability (linear coherent resonance) due to the lattice symmetry breakdown originating from local cooling forces and coupling sources.
- Lack of an optimal tapered force.

We immediately recognize that it is not easy to evade the first obstacle in practice. The second one is also quite troublesome. Although a Wien filter provides a tapered force, it is difficult to adjust the tapering coefficient to the optimum value [38]. Even if the optimum tapering is realized somehow, the filters have to be placed in all superperiods to keep the high lattice symmetry, which is not realistic in general.

DISPERSION-FREE LATTICE

One of the two obstacles mentioned above (i.e. lack of an optimal tapered force) can be overcome by introducing special bending elements [20]. Since the necessity of the tapered force comes from the existence of momentum dispersion in a usual storage ring, what we should try is to minimize dispersive effects. It is actually possible to eliminate linear dispersion all around the storage ring by combining an electro-static dipole field with a magnetic dipole [20].

The dispersion-free bending element enables us to construct multi-shell crystalline structures, as depicted in Fig. 6, without the use of the tapered force. The bunched

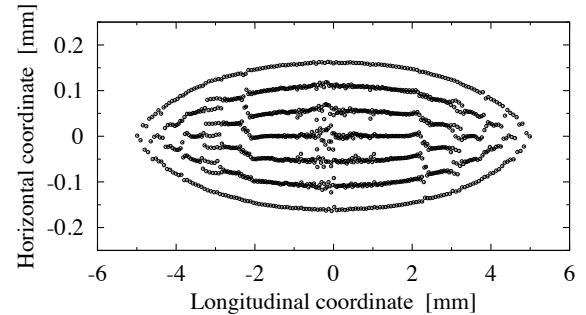


Figure 6: Ordered structure formed in a ten-fold symmetric test storage ring operating in a dispersion-free mode. The linear friction model in Eq. (6) has been employed to cool a bunched $^{24}\text{Mg}^+$ beam.

3D crystalline state was actually reached with the linear friction model in Eq. (6). The ordered beam in Fig. 6 is, however, unstable without the cooling force; the crystalline structure is destroyed as soon as we stop cooling the beam. This is because in the present simulation, we put only one rf cavity in the test ring to bunch the beam. If rf cavities are placed and excited in all lattice periods, then the stability of the crystalline beam is remarkably improved. In any case, it is impossible to form such an ordered 3D configuration as shown in Fig. 6 unless the ring is operated in the dispersion-free mode.

CONCLUSIONS

We have surveyed relatively recent MD results on Coulomb crystallization of ion beams circulating in a storage ring. Emphasis is placed upon the importance of including the actual lattice structure of the ring into MD simulations. In fact, the periodic nature of AG focusing and momentum dispersion peculiar to a circular machine make the dynamic behavior of crystalline beams much more complex than that of Coulomb crystals in a harmonic potential. To form various crystalline beams, we need a storage ring that has a high superperiodicity and is equipped with a laser cooler. It is particularly important to preserve the lattice symmetry as strictly as possible, so that the destructive effect from transverse coherent instability is minimized. In a regular ring, it is also strongly required to develop an optimal tapered force in order to avoid dispersive heating at ultra-low temperature.

MD simulations indicate that 1D and 2D crystalline beams can be generated with advanced beam cooling techniques, e.g. 3D laser cooling based on RCM. The use of the dispersion-free bending element may enable us to form even 3D crystalline structures (while it depends on how well we can maintain the lattice symmetry). In any case, a careful combination of state-of-the-art accelerator technologies will make it possible to produce an ultra-cold beam that has a normalized rms emittance of the order of 10^{-10} m·rad or even lower.

REFERENCES

- [1] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York, 1994).
- [2] T. Katayama and T. Koseki (eds.), *Proc. of Int. Workshop on Beam Cooling and Related Topics (COOL 03)* (Yamanashi, Japan, 2003).
- [3] H. Okamoto, H. Sugimoto, and Y. Yuri, to be published in *J. Plasma Fusion Res.*
- [4] This is not the case for high-energy electron and positron beams that naturally radiate photons.
- [5] J. P. Schiffer and P. Kienle, *Z. Phys. A* **321**, 181 (1985).
- [6] A. Rahman and J. P. Schiffer, *Phys. Rev. Lett.* **57**, 1133 (1986).
- [7] J. P. Schiffer, *Phys. Rev. Lett.* **61**, 1843 (1988).
- [8] R. W. Hasse and J. P. Schiffer, *Ann. Phys. (N.Y.)* **203**, 419 (1990).
- [9] R. W. Hasse and V. V. Avilov, *Phys. Rev. A* **44**, 4506 (1991).
- [10] J. Wei, X.-P. Li, and A. M. Sessler, *Phys. Rev. Lett.* **73**, 3089 (1994).
- [11] J. Wei, X.-P. Li, and A. M. Sessler, "Crystalline Beams", *AIP Conf. Proc.* **335**, p.224 (1995).
- [12] J. Wei, H. Okamoto, and A. M. Sessler, *Phys. Rev. Lett.* **80**, 2606 (1998).
- [13] Y. Yuri and H. Okamoto, *Phys. Rev. ST Accel. Beams* **8**, 114201 (2005).
- [14] T. Hänsch and A. Schawlow, *Opt. Commun.* **13**, 68 (1975).
- [15] D. J. Wineland and H. Dehmelt, *Bull. Am. Phys. Soc.* **20**, 637 (1975).
- [16] S. Schröder et al., *Phys. Rev. Lett.* **64**, 2901 (1990).
- [17] J. S. Hangst et al., *Phys. Rev. Lett.* **67**, 1238 (1991).
- [18] H. Okamoto, A. M. Sessler, and D. Möhl, *Phys. Rev. Lett.* **72**, 3977 (1994).
- [19] H. Okamoto, *Phys. Rev. E* **50**, 4982 (1994).
- [20] M. Ikegami et al., *Phys. Rev. ST Accel. Beams* **7**, 120101 (2004).
- [21] A. Noda, M. Ikegami, and T. Shirai, *New J. Phys.* **8**, 288 (2006).
- [22] H. Okamoto and J. Wei, *Phys. Rev. E* **58**, 3817 (1998).
- [23] K. Okabe and H. Okamoto, *Jpn. J. Appl. Phys.* **42**, 4584 (2003).
- [24] A. Piwinski, in *Proc. 9th Int. Conf. on High Energy Accelerators*, Stanford, CA, 1974, p.405 (SLAC, Stanford, 1974).
- [25] J. Bjorken and S. Mtingwa, *Part. Accel.* **13**, 115 (1983).
- [26] F. Diedrich et al., *Phys. Rev. Lett.* **59**, 2931 (1987).
- [27] D. J. Wineland et al., *Phys. Rev. Lett.* **59**, 2935 (1987).
- [28] M. Drewsen, C. Brodersen, L. Hornekær, and J. S. Hangst, *Phys. Rev. Lett.* **81**, 2878 (1998).
- [29] H.-J. Miesner et al., *Phys. Rev. Lett.* **77**, 623 (1996).
- [30] N. Madsen et al., *Phys. Rev. Lett.* **83**, 4301 (1999).
- [31] T. Schätz, U. Schramm, and D. Habs, *Nature (London)* **412**, 717 (2001).
- [32] U. Schramm, M. Bussmann, and D. Habs, *Nucl. Instrum. Meth. A* **532**, 348 (2004).
- [33] It is quite easy for any Paul traps to fulfill both conditions (1) and (2). More importantly, the low-energy Coulomb crystals in the circular trap are not exposed to strong momentum dispersion that causes a serious trouble in a real storage ring.
- [34] P. P. Ewald, *Ann. Phys. (Leipzig)* **64**, 253 (1921).
- [35] J. Wei, *Proc. of PAC'01*, Chicago, June 2001, TTPH011, p. 1678 (2001).
- [36] H. Okamoto, *Phys. Plasmas* **9**, 322 (2002).
- [37] I. Lauer et al., *Phys. Rev. Lett.* **81**, 2052 (1998).
- [38] M. Ikegami, H. Sugimoto, and H. Okamoto, *J. Phys. Soc. Jpn.* **77**, 074502 (2008).

DEVELOPING THE PHYSICS DESIGN FOR NDCX-II, A UNIQUE PULSE-COMPRESSING ION ACCELERATOR*

A. Friedman, J. J. Barnard, R. H. Cohen, D. P. Grote, S. M. Lund, W. M. Sharp, LLNL, USA
 A. Faltens, E. Henestroza, J.-Y. Jung, J. W. Kwan, E. P. Lee, M. A. Leitner, B. G. Logan,
 J.-L. Vay, W. L. Waldron, LBNL, USA
 R. C. Davidson, M. Dorf, E. P. Gilson, I. Kaganovich, PPPL, USA

Abstract

The Heavy Ion Fusion Science Virtual National Laboratory (a collaboration of LBNL, LLNL, and PPPL) is using intense ion beams to heat thin foils to the “warm dense matter” regime at $\lesssim 1$ eV, and is developing capabilities for studying target physics relevant to ion-driven inertial fusion energy. The need for rapid target heating led to the development of plasma-neutralized pulse compression, with current amplification factors exceeding 50 now routine on the Neutralized Drift Compression Experiment (NDCX). Construction of an improved platform, NDCX-II, has begun at LBNL with planned completion in 2012. Using refurbished induction cells from the Advanced Test Accelerator at LLNL, NDCX-II will compress a ~ 500 ns pulse of Li^+ ions to ~ 1 ns while accelerating it to 3–4 MeV over ~ 15 m. Strong space charge forces are incorporated into the machine design at a fundamental level. We are using analysis, an interactive 1D PIC code (ASP) with optimizing capabilities and centroid tracking, and multi-dimensional Warp-code PIC simulations, to develop the NDCX-II accelerator. This paper describes the computational models employed, and the resulting physics design for the accelerator.

INTRODUCTION

The Heavy Ion Fusion Science Virtual National Laboratory (HIFS-VNL) is a collaboration of Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, and the Princeton Plasma Physics Laboratory. The VNL is using intense ion beams to enable the study of matter in the poorly-understood “warm dense matter” (WDM) regime at $\lesssim 1$ eV, and is developing capabilities for experimental studies of inertial-fusion target physics relevant to ion-driven inertial fusion energy. For an overview, see [1].

The need for rapid target heating motivated the development of ion beam compression in the presence of a neutralizing plasma (which serves to minimize the beam space-charge forces that otherwise would inhibit compression to a compact volume). Bunching factors exceeding 50 have been achieved on the Neutralized Drift Compression Experiment (NDCX) at LBNL. Funding for an improved research platform, NDCX-II, has been approved (via the American Recovery and Reinvestment Act of 2009) and construction is beginning at LBNL, with planned comple-

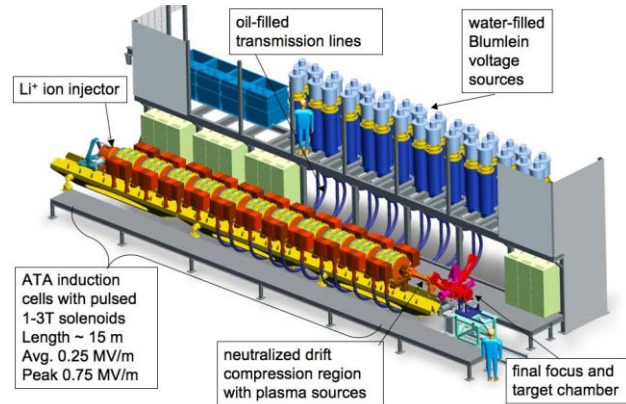


Figure 1: CAD rendering of a design concept for NDCX-II.

tion in 2012. This will be the first ion induction accelerator specifically designed to heat targets with short pulses. Using refurbished induction cells, Blumlein voltage sources, and transmission lines from the decommissioned Advanced Test Accelerator (ATA) at LLNL, NDCX-II will compress a ~ 1 m, ~ 500 ns pulse of Li^+ ions to ~ 1 cm, ~ 1 ns while accelerating it to 3–4 MeV over ~ 15 m. This is accomplished in two stages: the induction accelerator itself, which shortens the pulse to ~ 0.2 m, 20 ns ($\sim 5\times$ increased speed and $\sim 5\times$ decreased length); and a downstream neutralized drift compression line, which enables the final compression (spatial and temporal) through a factor of ~ 20 or more. The layout is shown in Fig. 1.

The ferrite in each ATA cell offers 0.014 V-s of flux swing, while the Blumleins can source as much as 250 kV with a FWHM of 70 ns. Passive pulse-shaping elements can be inserted into the “compensation boxes” attached to each cell, offering some flexibility in the accelerating waveforms. It is possible to generate longer pulses (and necessary to do so at the front end of the machine), as well as pulses with waveforms optimized to confine the beam-ends, but the need to minimize costs motivates keeping the voltage of any new power supplies to 100 kV or less. The NDCX-II lattice period is unchanged from that of the ATA, at 0.28 m; the accelerating gaps (across which the driving inductive electric field appears) are 2.8 cm long. Because ions do not rapidly reach high speeds, transverse confinement of the beam against its own space charge requires that the existing DC solenoids be replaced by much stronger pulsed solenoids with fields approaching 3 T.

For three reasons, it will be necessary to rebuild the radially innermost parts of the ATA cells. Firstly, a major

* This work was performed under the auspices of the USDOE by LLNL under Contract DE-AC52-07NA27344, by LBNL under Contract DE-AC02-05CH11231, and by PPPL under Contract DE-AC02-76CH03073.

loss of Volt-seconds available for acceleration results when the return flux of a pulsed solenoid passes through a ferromagnetic core. Thus, we plan to introduce extra magnetic elements (a copper layer, magnetic iron, or a secondary winding) outboard of each solenoid, to force flux to return inboard of the ferrite. This requires reducing the original 6.7 cm ATA beam-pipe radius to 4.5 cm. Secondly, an accelerating gap must be “on” while any of the beam overlaps its fringe field; the transit time for a single particle can be comparable to the time for the beam to pass a fixed axial position z . To shorten this fringe, it is again necessary to reduce the pipe radius (or at least the radius of the conductor that sets the axial scale of the fringe field) to of order 4 or 4.5 cm. Finally, the ATA cells contain end plates that allow large eddy currents; these plates will be made thinner.

Non-relativistic ions exhibit complex dynamics, more so when the beam is space-charge-dominated longitudinally and transversely. The beam manipulations in NDCX-II are actually enabled by longitudinal space charge forces. We are using analysis, an interactive 1D PIC code (ASP) with optimizing capabilities and a centroid-offset model, and both (r, z) and 3D Warp [2] simulations, to develop the NDCX-II accelerator. This paper describes the computational models used for these studies, and the resulting physics design. A companion paper [3], and earlier papers on this work [4, 5, 6], provide complementary information.

1-D ASP CODE MODEL

The ASP code (an acronym for “Acceleration Schedule Program”) follows the beam’s evolving longitudinal phase space (z, v_z) , and was developed to facilitate the synthesis and optimization of acceleration schedules. It uses well-established particle-in-cell methods, but solves for the space-charge field with a modified 1-D Poisson formulation [7]. In order to capture the fall-off with axial distance of a beam slice’s influence (associated with image charges in the beam pipe wall), we use the following prescription for the electrostatic potential $\phi(z)$:

$$\nabla^2 \phi \simeq \frac{d^2 \phi}{dz^2} - k_{\perp}^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (1)$$

where the transverse inverse scale length k_{\perp} is defined by:

$$k_{\perp}^2 = 4/(g_0 r_b^2); \quad g_0 = 2 \ln(r_w/r_b). \quad (2)$$

Here, r_b is the nominal beam radius, r_w is the pipe (“wall”) radius, and the so-called “g-factor” g_0 has been introduced to make correspondence with the long-wavelength limit. For very short wavelength variations the effect of the k_{\perp}^2 term is properly small, while at long wavelengths the electric field is given by

$$E_z = -\frac{\partial \phi}{\partial z} = -\frac{g_0}{4\pi\epsilon_0} \frac{d}{dz} (\pi r_b^2 \rho) = -\frac{g_0}{4\pi\epsilon_0} \frac{d\lambda(z)}{dz}, \quad (3)$$

where $\lambda(z)$ is the line charge density (C/m) [7, 8]. Our prescription for g_0 differs slightly from that of [7] because we

Linear Accelerators

seek to model a space-charge-dominated beam of roughly constant density, that is, $\lambda(z) \propto r_b^2(z)$. In practice, comparisons with (r, z) Warp calculations allow us to set g_0 .

ASP represents the accelerating-gap fields $E_{z,i}(z, t)$ due to each time-varying gap voltage $V_i(t)$ via the “Lee” model [9]; we use the single-term approximation (Eq. A7 of the reference) because the gap is narrow.

The general “type” of each accelerating waveform (including ideal ramps or flat-tops, circuit-models of various kinds, and self-adjusting “ear” waveforms that counteract beam end expansion) is set by the user. The code automatically adjusts each waveform using estimated times of beam entry into and exit from the gap’s fringe field, constrained by user-specified limits to peak voltage and Volt-seconds.

In addition to the longitudinal phase-space coordinates (z_k, p_{zk}) of each simulation particle (beam slice) k , ASP also tracks the transverse coordinates $(x_k, y_k, p_{xk}, p_{yk})$ of its centroid. In a perfectly aligned system, all of these would remain zero. However, misalignments drive the beam off axis. Furthermore, because of the head-to-tail energy variation along the beam, once a displacement occurs the different beam slices gyrate in the solenoids’ field at their own peculiar rates. The result is a “corkscrew” distortion [10]. Fortunately, methods for minimizing its amplitude have been developed for electron linacs [11], and our studies (described below) show that they remain effective for NDCX-II with its large energy variation.

ASP was designed to be interactive; it is written in the Python scripting language, with a few computationally-intensive routines in Fortran. A few hundred simulation particles (beam slices) are used, and a run without iterations takes a few minutes on a single processor. Iterations of two kinds are routinely carried out: variations of the applied voltage waveforms $V_i(t)$ to establish the acceleration schedule, and iterative tuning of the “steering” dipole magnet strengths to minimize the beam’s corkscrew and off-axis shift. On a 4-processor computational node of a 2.2 GHz AMD Linux cluster, each “steering” run takes between two and three hours. Using a script which launches four jobs at a time, an ensemble of 20 cases with differing random alignment errors runs overnight.

In order to initialize the ASP beam, we begin by carrying out a Warp simulation of the source diode and matching section, and record the beam parameters as it passes through a plane upstream of the first gap; we then construct a smooth beam in ASP that resembles the Warp beam, using a simple parameterization. This method requires manual intervention and is crude, but works well enough. We have most recently implemented a method that uses the Warp data to construct a profile which is then used as the basis in detail for the ASP initial beam [3].

1-D PHYSICS DESIGN

We identified two principles which served as guides to development of an effective acceleration schedule. The first of these is to “shorten the beam first” (compress it longitudinally).

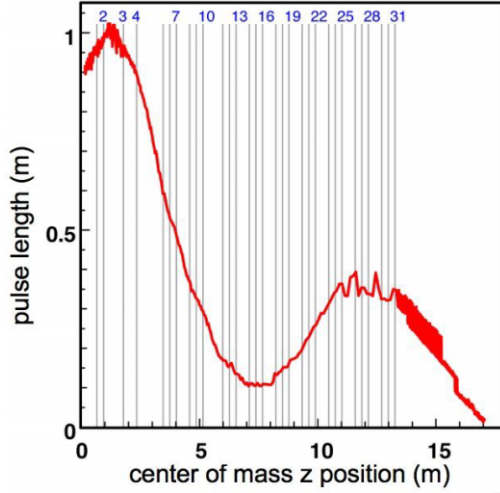


Figure 2: Pulse length vs. axial coordinate z . Numbers label the accelerating gaps.

dinally before the main acceleration) via a process of non-neutral drift compression. We use the initial cells primarily to impose a velocity tilt, while others provide longitudinal control; space is left for the drift compression. The goal is to achieve a sub-70 ns beam transit time through the acceleration gaps (including their fringe fields) as early as possible, so that we can use as many Blumlein-driven 200-kV pulses as possible. It is necessary, at this stage, to compress the beam carefully so as to minimize the effects of space charge which might lead to beam non-uniformity. Thus we seek to achieve (ideally) a large linear velocity “tilt” with $v_z(z) \propto z$ and a smooth density profile. We obtain the desired waveforms using a least-squares optimization that penalizes both nonlinearity and nonuniformity.

The second guiding principle is to “let the beam bounce.” Rapid inward motion in the beam frame is required to reduce the pulse duration below 70 ns. Space charge ultimately inhibits this compression. At that point, the beam is shorter than the fringe field of a gap, and is not sustainable because confining fields are ineffective. Thus, the beam “bounces” — that is, it starts to lengthen. Nonetheless, the duration remains below 70 ns because it is now moving faster and is confined by additional ramped pulses. We allow it to lengthen while applying additional acceleration via flat-top pulses, and longitudinal confinement via ramped (“triangular”) pulses. Alternatively, trapezoidal pulses may be used instead of a combination of flat-tops and triangles. The final few gaps apply the velocity tilt for neutralized drift compression onto the target.

Figure 2 shows the evolution of the beam length, while Fig. 3 shows the evolution of the pulse duration. While the design has yet to be finalized, the system described here is representative, with the final ~ 20 induction cells driven by the ATA Blumleins and the rest by lower-voltage sources. Most of the required waveforms (shown in Fig. 4) are simple enough to be formed with passive circuits in the “compensation boxes” that are attached to the ATA cells. ASP solves simplified circuit equations for the 200 kV triangular

Linear Accelerators

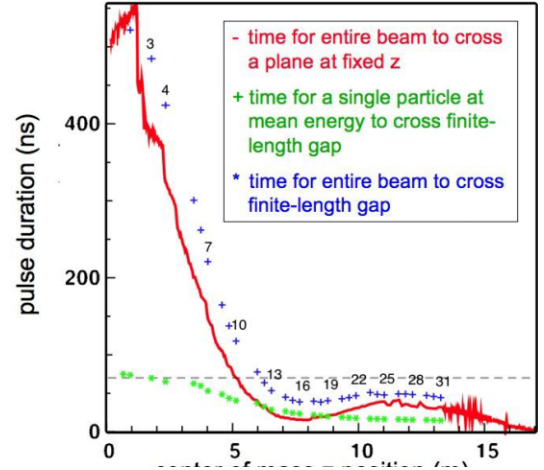


Figure 3: Pulse duration vs. z ; the key time is that for the entire beam to traverse a gap’s fringe field from entrance to exit.

pulses; laboratory tests of such pulse shaping are underway, and to date a high-quality ~ 60 ns ramp has been achieved (our latest ASP runs use this experimental data). Similar circuits may be able to form the initial tilt-generating waveforms. The low-voltage highly shaped “ear” waveforms that provide beam-end control (shown in black in Fig. 4) will be driven by programmable circuits.

Figure 5 shows the evolution of the phase space and current. The final panel shows the beam when its centroid is at the plane of best longitudinal focus. This plane is estimated by an RMS measure [12] and refined by searching for that plane through which the most current flows in a 1-ns window. In ASP, the beam is assumed perfectly neutralized after it exits the accelerator (the space-charge field is ramped to zero over a user-specified length).

(R, Z) SIMULATIONS

As mentioned earlier, we are carrying out both (r, z) and full 3-D simulations of NDCX-II using the Warp code.

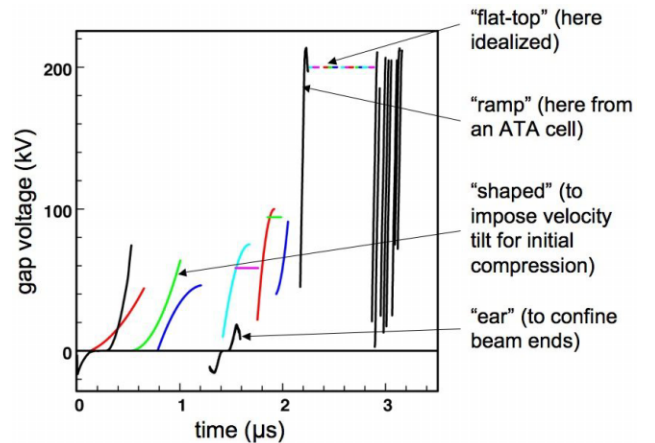


Figure 4: Waveforms for acceleration as developed via ASP.

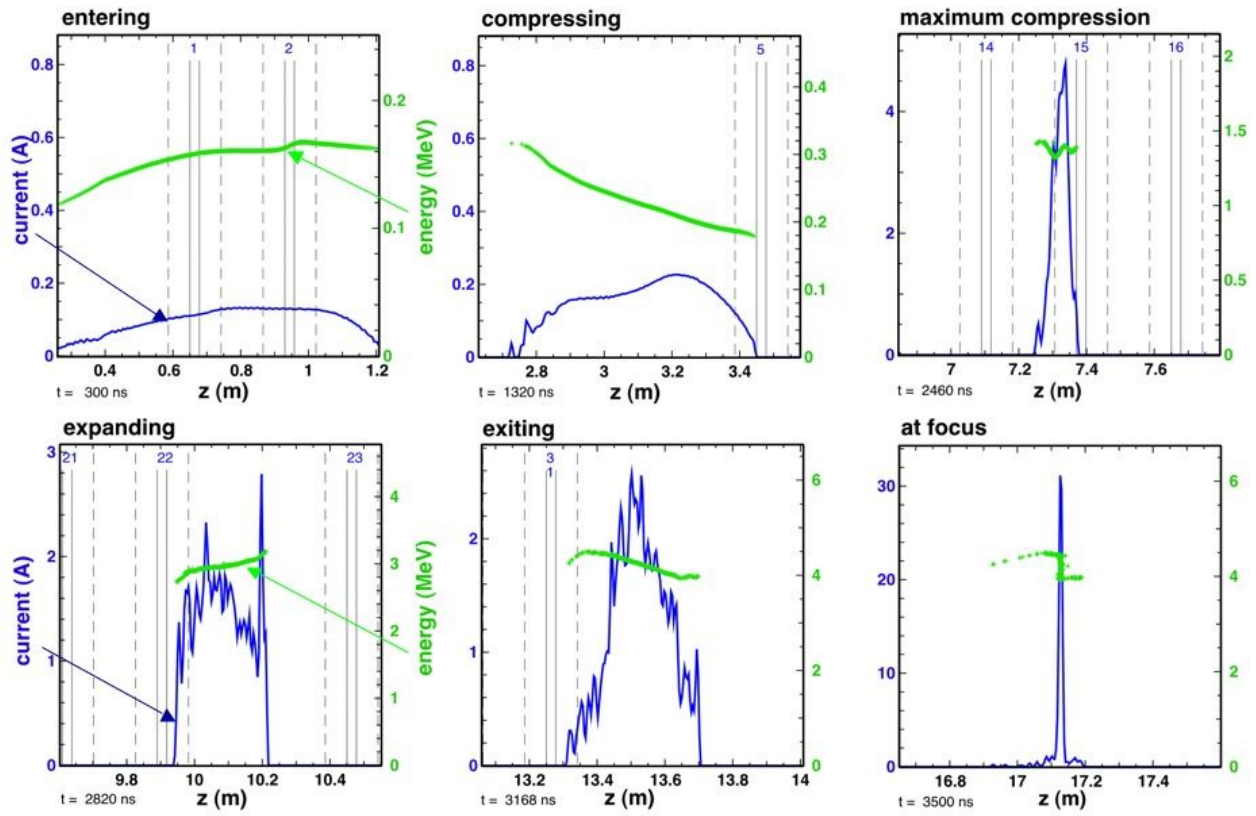


Figure 5: ASP snapshots.

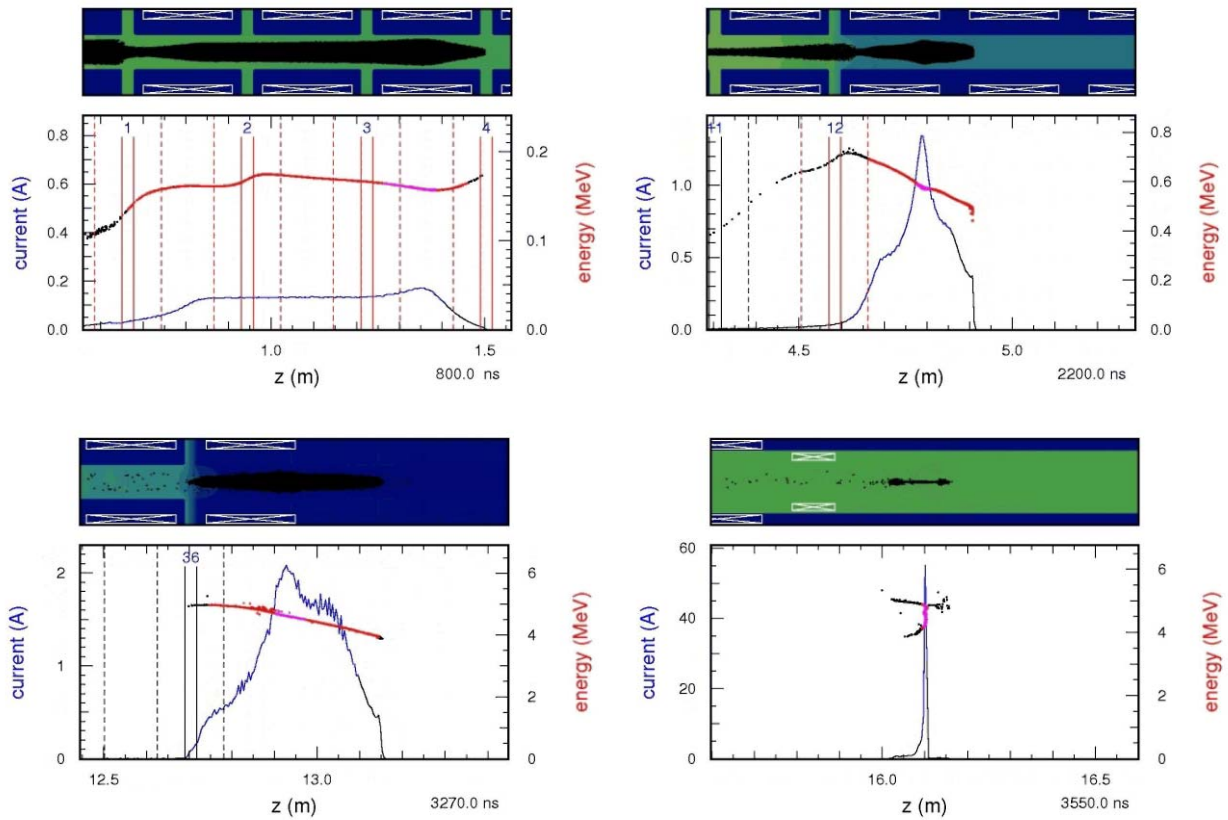


Figure 6: Warp RZ snapshots.

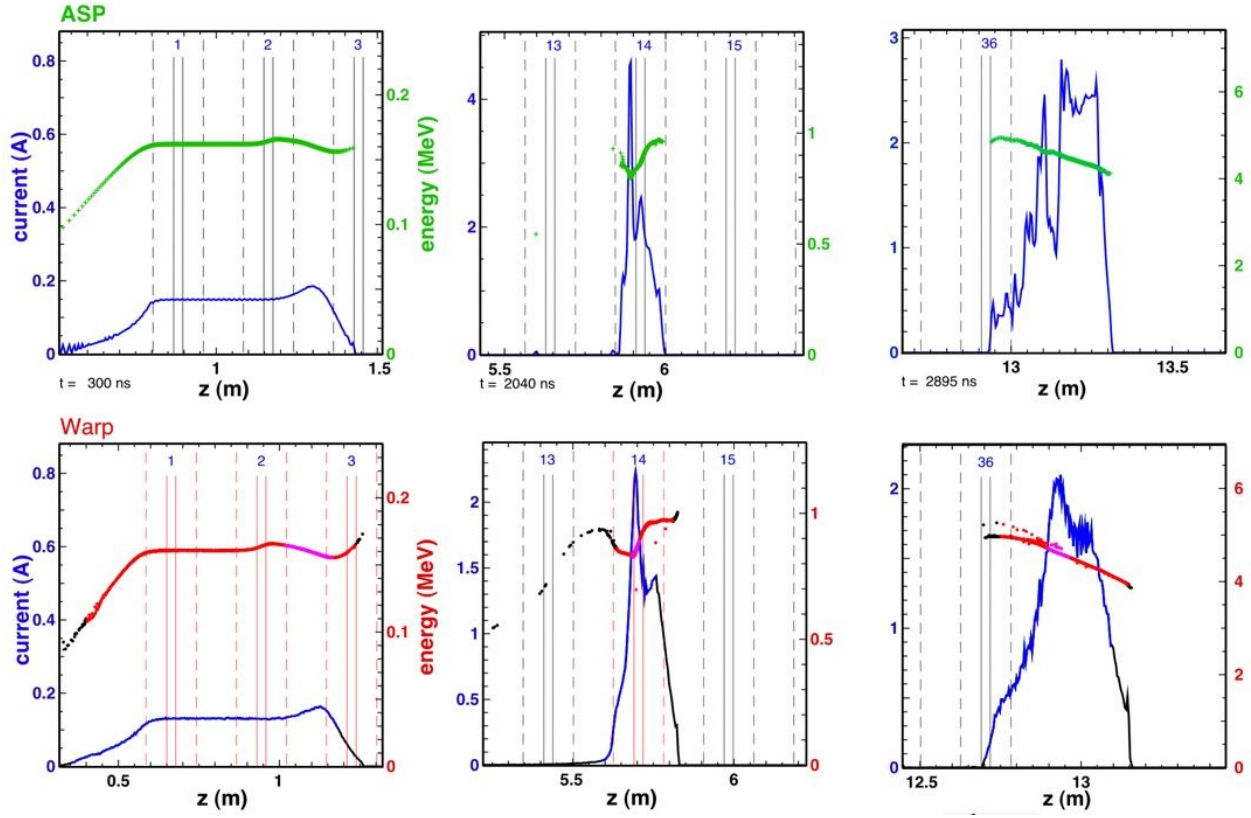


Figure 7: ASP vs. Warp comparison.

These simulations include detailed acceleration gaps and focusing magnet fields, and solve for the beam self-field in a conducting pipe. The gap fields have a two-dimensional variation (that is, the fringe field has reduced axial extent at larger radius); we derive these fields by solving Laplace's equation with differing potential on the walls at each side of the gap; the true field is inductive, but for these slowly varying fields the approximation is excellent. As an example, Fig. 6 shows a few snapshots from a Warp movie; the elevation view does not clearly convey the compactness of the beam, which has a long, low-density tail. (A number of such movies may be found at [13].) In view of the radial variation of the gap field and the approximate nature

of the 1-D self-field, it was not clear at the outset that a set of waveforms developed in ASP would be effective in accelerating, confining, and compressing the Warp beam. We were pleased to see that indeed our design procedure works reliably provided the initial beam is well described in ASP (see Fig. 7), and efficiently relative to iterative Warp runs.

The (r, z) mode of Warp is also used to design the injector (Fig. 8). For scoping we are using one refinement patch around the emitter; to obtain converged voltage waveforms it will be important to use fully adaptive multi-level AMR.

3-D EFFECTS: MISALIGNMENTS, CORKSCREW, AND THEIR MITIGATION

With nominal random solenoid-end misalignments of up to 0.5 mm (thus allowing for transverse offsets as well as pitch and yaw tilts), both ASP and Warp simulations indicate good beam transmission through the accelerator. See the 3-D Warp movie with random 1-mm shifts of solenoid ends at [13]. 3-D Warp simulations show that the degradation of the focal intensity is smooth, and modest for small errors. Nonetheless, it is highly desirable to use steering dipole magnets to minimize both the corkscrew amplitude and the off-axis displacement. In Fig. 9 we show how iterative dipole tuning can be effective in such a system. The most recent designs for NDCX-II are built around blocks of three cells, followed by a non-accelerating lattice period with a solenoid, steering dipoles for x and y , and diagnostics to measure the beam centroid displacement.

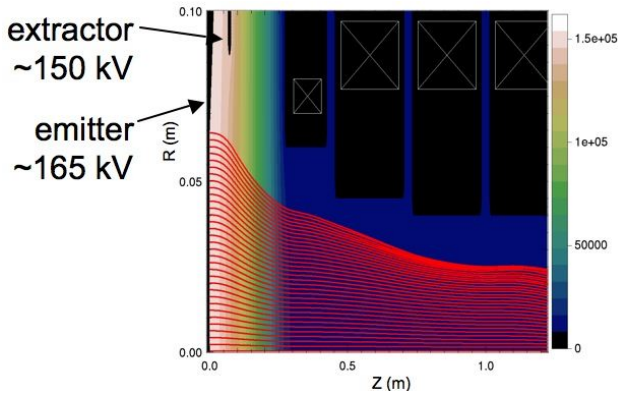


Figure 8: Warp simulation of a possible injector configuration. Linear Accelerators

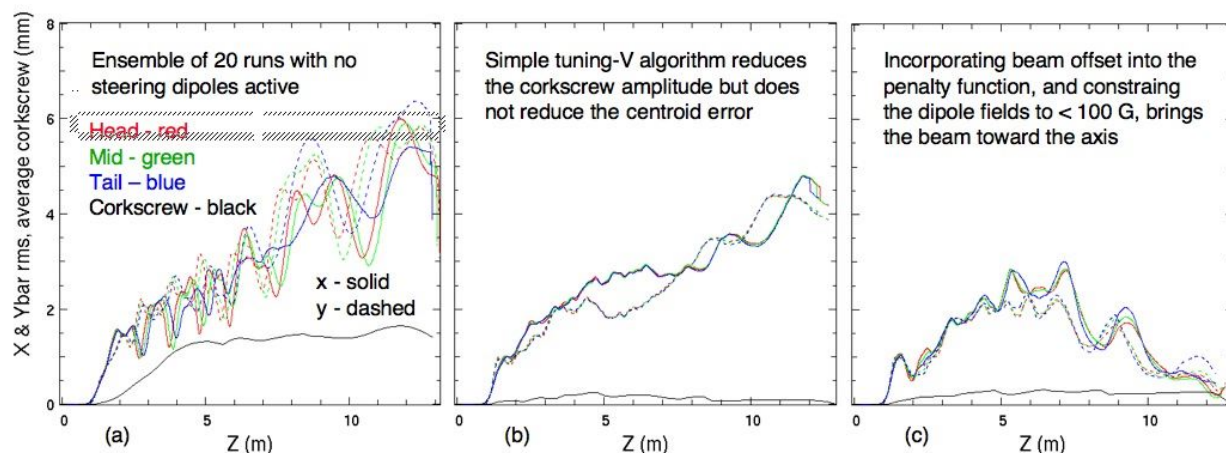


Figure 9: ASP studies of steering, showing RMS of x and y centroid coordinates vs. z for head, middle, and tail particles, and the corkscrew amplitude; the results are averages over 20 simulations with differing random offsets of solenoid ends up to 1 mm. The penalty function was evaluated at the next sensor downstream from the dipole being varied.

DISCUSSION

We have developed the physics design for a novel accelerator, using a new computational tool (ASP) in tandem with an existing tool (Warp). Computational aspects of this work, including the 1-D field solution with tuned axial falloff, self-adjusting waveforms that make maximal use of available volt-seconds and voltage, iterations for waveform design and for beam steering, and the interactive very-high-level code framework based on Python, have all worked out well in this application.

The baseline design for the NDCX-II project will be established later this year; it will resemble the designs shown here, but should take the beam to a lower final kinetic energy (the goal for the WDM application is a mean kinetic energy of 2.8 MeV, as appropriate for uniform heating by a beam that slows through the Bragg peak in a foil target).

This paper has described the physics design of the NDCX-II accelerator, and has not attempted to cover many other important aspects of the full machine, which include the ion source and the pulsed power, as well as the neutralized drift line, final focusing solenoid, and target chamber (into all of which which plasma must be injected at sufficient density that the beam space charge is uniformly cancelled). For discussion of these topics the reader may refer to the Proceedings of the recent Heavy Ion Fusion Symposium [14] and presentations at PAC09 [15, 16].

REFERENCES

- [1] B. G. Logan, *et al.*, *Nucl. Instr. and Meth.* **A 577**, 1 (2007).
- [2] D.P. Grote, A. Friedman, I. Haber, W. Fawley, J.-L. Vay, *Nucl. Instr. and Meth.* **A 415**, 428 (1998).
- [3] W. M. Sharp, A. Friedman, D. P. Grote, R. H. Cohen, S. M. Lund, M. Leitner, J.-L. Vay, and W. L. Waldron, "Modeling the NDCX-II Physics Design," ICAP'09, San Francisco, Sept. 2009, <http://www.JACoW.org> (these *Proceedings*).
- [4] A. Friedman, *et al.*, *Nucl. Instr. and Meth.* **A 606**, 6 (2009).
- [5] W. M. Sharp, A. Friedman, D.P. Grote, E. Henestroza, M. A. Leitner, and W. L. Waldron, *Nucl. Instr. and Meth.* **A 606**, 97 (2009).
- [6] W. M. Sharp, A. Friedman, D.P. Grote, E. Henestroza, M. A. Leitner, and W. L. Waldron, "Simulating an Acceleration Schedule for NDCX-II," PAC'09, Vancouver, May 2009, TH5FPF072, <http://www.JACoW.org>.
- [7] J. J. Barnard, G. J. Caporaso, S. S. Yu, and S. Eylon, "One Dimensional Simulations of Transients in Heavy Ion Injectors," PAC'93, Washington, DC, May 1993, pp. 712-714, <http://www.JACoW.org>.
- [8] J. J. Barnard and S. M. Lund, Course notes, *Beam Physics with Intense Space-Charge*, U.S. Particle Accelerator School, Univ. of Maryland, June 2008, Pp. 13-14 of Section on *Longitudinal Physics Part II*, <http://uspas.fnal.gov/materials/08UMD/BeamPhysics.html>
- [9] E. P. Lee (private communication); see: D. R. Welch, *et al.*, *Phys. Rev. ST Accel. Beams* **11**, 064701 (2008), Appen. A.
- [10] Y.-J. Chen, *Nucl. Instr. and Meth.* **A 292**, 455 (1990).
- [11] Y.-J. Chen, *Nucl. Instr. and Meth.* **A 398**, 139 (1997).
- [12] A. Friedman, "Some kinematic aspects of neutralized drift compression," LLNL Note LLNL-TR-402447, March 2008.
- [13] <http://hifweb.lbl.gov/public/movies/ICAP09>
- [14] *Proc. 17th Int. Sympos. on Heavy Ion Inertial Fusion (HIF2008)*, Tokyo, Aug. 4-9, 2008, published as *Nucl. Instr. and Meth.* **A 606**, (2009).
- [15] I. D. Kaganovich, *et al.*, "Designing Neutralized Drift Compression for Focusing of Intense Ion Beam Pulses in Background Plasma" PAC'09, Vancouver, May 2009, <http://www.JACoW.org>.
- [16] R. C. Davidson, E. A. Startsev, M. Dorf, I. D. Kaganovich and H. Qin, "Collective Instabilities and Beam-Plasma Interactions for an Intense Ion Beam Propagating through Background Plasma," PAC'09, Vancouver, May 2009, <http://www.JACoW.org>.

SELF FIELD OF SHEET BUNCH: A SEARCH FOR IMPROVED METHODS

G. Bassi, University of Liverpool and Cockcroft Institute, Liverpool, UK *
J. A. Ellison, K. Heinemann, University of New Mexico, Albuquerque, NM, USA †

ABSTRACT

We consider a sheet bunch represented by a random sample of \mathcal{N} simulation particles moving in a 4D phase space. The mean field (=‘self field’) of the bunch is computed from Maxwell’s equations in the lab frame with a smoothed charge/current density. The particles are tracked in the beam frame, thus requiring a transformation of densities from lab to beam frame. We seek improvements in speed and practicality in two directions: (a) choice of integration variables and quadrature rules for the field calculation; and (b) finding smooth densities from scattered data. For item (a) we compare our singularity-free formula with the retarded time as integration variable, which we currently use, with a formula based on Frenet-Serret coordinates. The latter suggests good approximations in different regions of the retardation distance which could save both time and storage. For item (b) we discuss Fourier vs. kernel density estimation and mention quasi vs. pseudo-random sampling.

INTRODUCTION

In this paper we discuss current and future approaches to numerically integrating the Vlasov-Maxwell system for a sheet bunch. More information on our current work can be found in [1]-[3]. We first present the mathematical problem in the lab frame. We write the field as an integral of the time history of the source. Then the initial value problem (IVP) for the Vlasov equation defines the $u = ct$ evolution of the phase space density, f_L . The coefficients of the Vlasov equation depend on the Maxwell self field and thus contain integrals over the time history of f_L .

It is both physically and computationally advantageous to determine the so-called beam frame phase space density, f_B . We define the beam frame phase space variables in terms of the lab frame. The independent variable in the lab frame (LF) is u and the independent variable in the beam frame (BF) is arc length s along a suitably defined reference orbit. The lab to beam phase space variable transformation gives the relation between f_B and f_L and f_B satisfies a BF Vlasov equation. Our goal is an efficient computation of the s -evolution of f_B given its value f_{B0} at say $s = 0$. However, this problem is not well posed; solutions are not unique. The root of this is a causality issue; at s , certain coefficients of the BF Vlasov equation need information about f_B outside the interval $[0, s]$. This problem, which is pertinent to the BF and absent in the LF, is easily resolved to what we believe is a good approximation.

We want to numerically integrate the 4D BF Vlasov

equation and we do this in terms of a random sample of \mathcal{N} points which simulate the 4D phase space density. We work in a high performance computing (HPC) environment. Even so, we do not have a fast enough algorithm to take \mathcal{N} large enough to obtain an accurate estimate of the 4D density. Furthermore, there are probably more efficient ways to obtain the 4D density, e.g., the method of local characteristics. However, the self field calculation only needs the BF spatial density, ρ_B , and a 2D current density type function, which we denote by τ_B . We believe our sample of 4D points is large enough to accurately estimate these 2D quantities and this makes a simulation approach feasible. We randomly generate an initial sample of BF phase space points from f_{B0} , and move this sample according to the BF equations of motion. Having resolved the causality issue, the self field can be computed at arc length s from the history of $g_B = (\rho_B, \tau_B)$. The calculation of g_B requires a density estimation procedure from our scattered data which we discuss. To move the points from s to $s + \delta s$ we freeze the self field at s and move the points according to the equations of motion. Important to our approach is the discovery of an s -independent grid on which to represent the spatial density and a parallel implementation of our algorithm.

STATEMENT OF PROBLEM FOR SHEET BUNCH IN LAB FRAME

We consider particle motion in the $Y = 0$ plane in a right handed coordinate system, (Z, X, Y) , under an external magnetic field $\mathbf{B}_{ext}(Z, X, Y) = B_{ext}(Z)\mathbf{e}_Y$. The equations of motion without self field are

$$\dot{\mathbf{R}} = \frac{\mathbf{P}}{m\gamma(\mathbf{P})c}, \quad \dot{\mathbf{P}} = qB_{ext}(Z)\frac{1}{m\gamma(\mathbf{P})c} \begin{pmatrix} P_X \\ -P_Z \end{pmatrix}, \quad (1)$$

where $\mathbf{R} = (Z, X)^T$, $\mathbf{P} = (P_Z, P_X)^T$, $\dot{} = d/du$, m is the electron rest mass, q is the electron charge and γ is the Lorentz factor. The associated 4D phase space density, $f_L(\mathbf{R}, \mathbf{P}; u)$, evolves according to the Liouville equation $\partial_u f_L + \dot{\mathbf{R}} \cdot \partial_{\mathbf{R}} f_L + \dot{\mathbf{P}} \cdot \partial_{\mathbf{P}} f_L = 0$, where f_L is normalized so that its integral over a phase space region represents the fraction of the beam in that region. All densities in this paper are normalized in this way.

We are interested in the evolution of f_L when coupled to the self field and we begin with the coupled Vlasov-Maxwell initial boundary problem in 3D with a shielding boundary condition and initial data at $u = u_i$ where u_i will be specified further below. In general, the self field will push the particles out of the $Y = 0$ plane unless the bunch is a ‘sheet bunch’ and the self

* gabriele.bassi@stfc.ac.uk

† Work supported by US DOE grant DE-FG02-99ER41104

field is ‘symmetric’. We call the bunch a ‘sheet bunch at u ’ if its phase space density f has, at time u , the form $f(Z, X, Y, P_Z, P_X, P_Y; u) = \delta(Y)\delta(P_Y)f_L(\mathbf{R}, \mathbf{P}; u)$. Denoting the components of the self field by $E_Z, E_X, E_Y, B_Z, B_X, B_Y$ we call the self field “symmetric at u ” if $E_Z(\mathbf{R}, Y; u), E_X(\mathbf{R}, Y; u), B_Y(\mathbf{R}, Y; u)$ are even in Y and if $E_Y(\mathbf{R}, Y; u), B_Z(\mathbf{R}, Y; u), B_X(\mathbf{R}, Y; u)$ are odd in Y . The point here is that if the initial bunch is a sheet bunch and if the initial self field is symmetric then, thanks to the reflection symmetry of Maxwell’s equations with respect to the $Y = 0$ plane, the bunch is a sheet bunch for all $u \geq u_i$ and the self field is symmetric for all $u \geq u_i$. We assume that these initial conditions on f and the self field are fulfilled whence, for all $u \geq u_i$, $E_Y(\mathbf{R}, 0; u) = B_Z(\mathbf{R}, 0; u) = B_X(\mathbf{R}, 0; u) = 0$. Thus no particle is pushed out of the $Y = 0$ plane.

We now focus on the self field components $\mathcal{F}_L(\mathbf{R}; u) = (E_Z(\mathbf{R}, 0; u), E_X(\mathbf{R}, 0; u), B_Y(\mathbf{R}, 0; u)) \equiv (E_Z(\mathbf{R}; u), E_X(\mathbf{R}; u), B_Y(\mathbf{R}; u))$ which together with $B_{ext}(Z)$ are responsible for the in plane forces. The solution of the initial boundary problem for \mathcal{F}_L , with $\mathcal{F}_L(\mathbf{R}; u) = \partial_u \mathcal{F}_L(\mathbf{R}; u) = 0$ at $u = u_i$ and the shielding boundary condition $\mathcal{F}_L = 0$ for $Y = \pm h/2$, is

$$\mathcal{F}_L(\mathbf{R}; u) = -\frac{1}{4\pi} \sum_{k=-\infty}^{\infty} (-1)^k \times \int_{\mathbb{R}^2} d\mathbf{R}' \frac{\mathbf{S}(\mathbf{R}'; u - [|\mathbf{R}' - \mathbf{R}|^2 + (kh)^2]^{1/2})}{[|\mathbf{R}' - \mathbf{R}|^2 + (kh)^2]^{1/2}}. \quad (2)$$

The source is

$$\mathbf{S}(\mathbf{R}; u) = Z_0 Q H(u - u_i) \begin{pmatrix} c \partial_Z \rho_L + \partial_u J_{L,Z} \\ c \partial_X \rho_L + \partial_u J_{L,X} \\ \partial_X J_{L,Z} - \partial_Z J_{L,X} \end{pmatrix}, \quad (3)$$

where H is the indicator function on $[0, \infty)$ and $\mathbf{J}_L = (J_{L,Z}, J_{L,X})^T$. This solution is obtained by writing the Maxwell equations in the wave equation form and using the retarded Green function and the method of images. If the initial condition is not zero a homogeneous solution must be added. Without the boundary condition only the $k = 0$ term remains.

The Vlasov IVP for the LF phase space density is

$$\partial_u f_L + \dot{\mathbf{R}} \cdot \partial_{\mathbf{R}} f_L + \dot{\mathbf{P}} \cdot \partial_{\mathbf{P}} f_L = 0, \quad f_L(\mathbf{R}, \mathbf{P}; u_i) = f_{L0}(\mathbf{R}, \mathbf{P}), \quad (4)$$

where

$$\dot{\mathbf{R}} = \frac{\mathbf{P}}{m\gamma(P)c}, \quad (5)$$

$$\dot{\mathbf{P}} = \frac{q}{c} \left(\begin{pmatrix} E_Z(\mathbf{R}; u) \\ E_X(\mathbf{R}; u) \end{pmatrix} + \frac{[B_{ext}(Z) + B_Y(\mathbf{R}; u)]}{m\gamma(\mathbf{P})} \begin{pmatrix} P_X \\ -P_Z \end{pmatrix} \right).$$

The Vlasov equation and the self field are coupled by the 2D charge and current densities, $Q\rho_L$ and $Q\mathbf{J}_L$, where

$$\rho_L(\mathbf{R}; u) = \int_{\mathbb{R}^2} d\mathbf{P} f_L(\mathbf{R}, \mathbf{P}; u), \quad (6)$$

$$\mathbf{J}_L(\mathbf{R}; u) = \int_{\mathbb{R}^2} d\mathbf{P} (\mathbf{P}/m\gamma(\mathbf{P})) f_L(\mathbf{R}, \mathbf{P}; u). \quad (7)$$

Computer Codes (Design, Simulation, Field Calculation)

We believe the IVP (4) is well posed, that is, there exists a unique solution depending continuously on the initial data. Furthermore, given $f_L(\cdot; u)$, $\mathcal{F}_L(\cdot; u)$ and a small positive δu the solution at $u + \delta u$ can be determined approximately by freezing the field at u and moving forward along characteristics defined by the Vlasov equation in (4). Note that $\mathcal{F}_L(\mathbf{R}, u)$, requires knowing $\rho_L(\cdot; v)$ and $\mathbf{J}_L(\cdot; v)$ for $u_i \leq v \leq u$.

In this paper we focus on the numerical solution of the IVP (4). It is computationally intensive even in a HPC environment and so a fast algorithm is utmost on our mind. Actually, for several reasons, the Vlasov equation is integrated in the beam frame and this will be discussed in the next section. Several approximations will be involved. Ultimately the approximations must be judged by how accurately they give an approximation to $f_L(\cdot; u)$ as defined by (4).

The physical problem we have in mind is a single pass four magnet chicane, and this determines $B_{ext}(Y)$ (e.g., see [1]). We take u_i to be the time at which the head of the bunch reaches the chicane. To have a well defined “head”, as well as for other reasons, we consider a bunch of compact spatial support, and we assume this in the following (see [2]). In the regime we have studied we believe the self field at $u = u_i$ is negligible and so (2) is appropriate. In applications such as this, it is important to determine the evolution of the so-called BF phase space density, f_B , with the arc length s along a reference orbit as the independent variable, and with f_B given at $s = 0$, which we take to be the entrance to the chicane. This frame is also convenient as the phase space variables are small and so linearizations are possible. We now turn our attention to the beam frame.

BEAM FRAME FOR SHEET BUNCH

The beam frame is defined in terms of the reference orbit $\mathbf{r}_r(s) = (Z_r(s), X_r(s))^T$ in the $Y = 0$ plane, which in turn is defined by the Lorentz equations without the self field given in (1). Here s is the arc length along the orbit and $\mathbf{r}_r(0) = 0$ is the entry point of the reference orbit into the chicane. The unit tangent vector, \mathbf{t} , to the reference orbit is just $\mathbf{t}(s) = \mathbf{r}'_r(s)$ and we define the unit normal vector, \mathbf{n} , by $\mathbf{n}(s) = (-X'_r(s), Z'_r(s))^T$ so that \mathbf{n} is a $\pi/2$ counterclockwise rotation from \mathbf{t} . It follows from (1) that $\mathbf{t}'(s) = -qB_{ext}(Z_r(s))\mathbf{n}(s)/P_r$ where $P_r = m\gamma_r\beta_r c$ is the momentum of the reference particle. This determines the curvature κ up to a sign and we choose $\kappa(s) = qB_{ext}(Z_r(s))/P_r$. Thus $\mathbf{t}'(s) = -\kappa(s)\mathbf{n}(s)$ and $\mathbf{n}'(s) = \kappa(s)\mathbf{t}(s)$.

The BF Frenet-Serret coordinates are $\xi = (s, x)$, where x is the perpendicular distance along \mathbf{n} . Let $\mathbf{T} : U := \mathbb{R} \times (-x_M, x_M) \rightarrow \mathbf{T}(U) \subset \mathbb{R}^2$, where $\mathbf{T}(\xi) = \mathbf{r}_r(s) + x\mathbf{n}(s)$ and $x_M > 0$ is chosen sufficiently small so that \mathbf{T} is a bijection. This leads to the phase space variable transformation $(\mathbf{R}, \mathbf{P}) \leftrightarrow (s, x, p_s, p_x)$ defined by

$$\mathbf{R} = \mathbf{T}(\xi), \quad \mathbf{P} = \mathbf{P}(s, p_s, p_x), \quad (8)$$

where $\mathbf{P}(s, p_s, p_x) := P_r(p_s \mathbf{t}(s) + p_x \mathbf{n}(s))$. Our lab to beam transformation has two more transformations so that we have $(\mathbf{R}, \mathbf{P}; u) \rightarrow (s, x, p_s, p_x; u) \rightarrow (u, x, p_s, p_x; s) \rightarrow (z, x, p_z, p_x; s)$. We have included the independent variables and in the second transformation the variables s and u are interchanged making s the new independent variable. In the final transformation $z := s - \beta_r u$ replaces u as a dependent variable and $p_z := (\gamma - \gamma_r)/\gamma_r$ replaces p_s . Note that the variables $\mathbf{r} := (z, x)^T$ and $\mathbf{p} := (p_z, p_x)^T$ are small near the reference orbit which corresponds to $\mathbf{r} = \mathbf{p} = 0$.

The main objects in this paper are the phase space densities. The exact relation between the LF phase space density, f_L , and the BF phase space density, f_B , is

$$f_B(\mathbf{r}, \mathbf{p}; s) = \frac{P_r^2}{\beta_r^2} f_L(\mathbf{T}(\xi), P_r[p_s(\mathbf{p})\mathbf{t}(s) + p_x\mathbf{n}(s)]; (s - z)/\beta_r), \quad (9)$$

where

$$p_s(\mathbf{p})^2 = \left(\frac{1}{\beta_r}\right)^2 (1 + p_z)^2 - p_x^2 - \frac{1}{\gamma_r^2 \beta_r^2}. \quad (10)$$

Recall that we consider $f_L(\cdot; u)$ to be well defined by (4) for $u \geq u_i$ and note that $f_B(\cdot; s)$ requires $f_L(\cdot; u)$ for a range of u values. This density transformation is unusual in that there is an interchange of u and s in their roles as independent and dependent variables and this requires that in the dynamics s must be an increasing function of u . There are subtleties in the derivation of the transformation (9) and details will be given in [2].

Introducing the inverse of (8), we write

$$f_L(\mathbf{R}, \mathbf{P}; u) = \frac{\beta_r^2}{P_r^2} f_B\{s(\mathbf{R}) - \beta_r u, x(\mathbf{R}), p_z(\mathbf{P}, s(\mathbf{R})), \mathbf{P} \cdot \mathbf{n}(s(\mathbf{R}))/P_r; s(\mathbf{R})\}, \quad (11)$$

where $(s(\mathbf{R}), x(\mathbf{R})) = \mathbf{T}^{-1}(\mathbf{R})$ for $\mathbf{R} \in \mathbf{T}(U)$, $p_s(\mathbf{p}) = \mathbf{P} \cdot \mathbf{t}(s)/P_r$ can be solved for $p_z(\mathbf{P}, s)$ using (10) and $p_x = \mathbf{P} \cdot \mathbf{n}(s(\mathbf{R}))/P_r$. Because $\mathbf{R} = \mathbf{T}(\xi)$ in (8) is only locally invertible we require the coordinate densities $\rho_L(\mathbf{R}; u)$ and $\rho_B(\mathbf{r}; s)$ to have compact support.

To define the BF Vlasov equation we need the BF equations of motion. Using (5) and (8) we obtain

$$\begin{aligned} z' &= 1 - \frac{\mathcal{K}(\xi)(1 + p_z)}{p_s(\mathbf{p})}, & x' &= \frac{\mathcal{K}(\xi)p_x}{p_s(\mathbf{p})}, \\ p_z' &= \frac{q\mathcal{K}(\xi)}{m\gamma_r c^2} \left(\mathbf{E}_{\parallel}(\mathbf{T}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{t}(s) \right. \\ &\quad \left. + \frac{p_x}{p_s(\mathbf{p})} \mathbf{E}_{\parallel}(\mathbf{T}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{n}(s) \right) \\ p_x' &= \frac{1}{P_r} \left(P_r p_s(\mathbf{p}) \kappa(s) - q\mathcal{K}(\xi) B_{ext}(Z_r(s) - xX_r'(s)) \right) \\ &\quad + \frac{q(1 + p_z)}{P_r p_s(\mathbf{p}) \beta_r c} \mathcal{K}(\xi) \mathbf{E}_{\parallel}(\mathbf{T}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{n}(s) \\ &\quad - \frac{q}{P_r} \mathcal{K}(\xi) B_Y(\mathbf{T}(\xi); \frac{s-z}{\beta_r}). \end{aligned} \quad (12)$$

Computer Codes (Design, Simulation, Field Calculation)

Here $' = d/ds$, $\mathbf{E}_{\parallel} := (E_Z, E_X)^T$ and $\mathcal{K}(\xi) := 1 + \kappa(s)x$.

The BF phase space density f_B , given in (9), satisfies the Vlasov equation

$$\partial_s f_B + \mathbf{r}' \cdot \nabla_{\mathbf{r}} f_B + \mathbf{p}' \cdot \nabla_{\mathbf{p}} f_B = 0, \quad (13)$$

where the coefficients are given by (12). This can be verified by simply plugging (9) into (13) and using (4). It is also consistent with the BF equations of motion themselves, as it must be, since the Vlasov equation for (12) is (13). In this context it is worth mentioning that the vector field in (12) is divergence free (see [2]).

Since there exists a unique f_L defined by the IVP (4), Eq. (9) gives us a unique f_B which we want to compute. However, we want to compute it based on the $s = 0$ IVP for (13), where

$$f_B(\mathbf{r}, \mathbf{p}; 0) =: f_{B0}(\mathbf{r}, \mathbf{p}), \quad (14)$$

by using an algorithm which marches forward in s . The initial condition can be determined from (9) given f_{L0} and the solution of (4) for a small forward time interval starting at u_i . However, in our applications, e.g. [1], we have been given f_{B0} , not f_{L0} .

Clearly, (13) is a nonlinear partial differential integral equation where the \mathbf{p}' coefficient depends on the self field $\mathcal{F}_L(\mathbf{R}(s, x), (s - z)/\beta_r)$ and thus on f_B through f_L . However, there is a causality issue. The quantity $\mathcal{F}_L(\mathbf{R}(s, x), (s - z)/\beta_r)$ requires knowledge of $f_B(\cdot; \tau)$, not only for $\tau \in [0, s]$, but also for some τ outside this interval. Nevertheless, since the main contribution to \mathbf{p}' comes from $[0, s]$, we obtain a feasible s-stepping algorithm for f_B given the initial condition (14). In the next section we discuss what we do when knowledge of $f_B(\cdot; \tau)$ is needed for $\tau \notin [0, s]$.

The basic computational issue is to solve the IVP (13-14) in such a way that to good approximation (9) is satisfied, with f_L defined by (4). This requires determining f_{L0} from f_{B0} , which we discuss in [2].

Before leaving this section we note that in our computations we use the approximate BF equations of motion,

$$\begin{aligned} z' &= -\kappa(s)x, & p_z' &= F_{z1}(z, x; s) + p_z F_{z2}(z, x; s), \\ x' &= p_x, & p_x' &= \kappa(s)p_z + F_x(z, x; s), \end{aligned} \quad (15)$$

where

$$\begin{aligned} F_{z1} &= \frac{q}{P_r c} \mathbf{E}_{\parallel}(\mathbf{R}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{t}(s), \\ F_{z2} &= \frac{q}{P_r c} \mathbf{E}_{\parallel}(\mathbf{R}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{n}(s), \\ F_x &= \frac{q}{P_r c} [\mathbf{E}_{\parallel}(\mathbf{R}(\xi); \frac{s-z}{\beta_r}) \cdot \mathbf{n}(s) - cB_Y(\mathbf{R}(\xi); \frac{s-z}{\beta_r})]. \end{aligned} \quad (16)$$

These equations were obtained from the exact BF equations (12) by linearizing the terms without the self field and by approximating the coefficients of the self field terms (see [2]). Without the self field, the general solution of (15) can be written in terms of the dispersion function, $D(s)$, and the

momentum compaction function, $R_{56}(s)$. We have found that it is numerically more efficient to integrate (15) in the interaction picture based on this zero self field solution (see [1]).

SELF FIELD FROM BF PHASE SPACE DENSITY AND CAUSALITY DISCUSSION

An s -stepping algorithm to evolve an approximation to $f_B(\cdot; s)$ according to (13-14) needs to compute

$$\mathcal{F}_L(\mathbf{T}(\xi); (s-z)/\beta_r) = - \int_{\mathbb{R}^2} d\mathbf{R}' \frac{\mathbf{S}[\mathbf{R}'; (s-z)/\beta_r - |\mathbf{R}' - \mathbf{T}(\xi)|]}{4\pi|\mathbf{R}' - \mathbf{T}(\xi)|}, \quad (17)$$

approximately, where $\mathcal{F}_L(\mathbf{R}; u)$ was defined in (2). Here we ignore the shielding, as it adds little computational complexity. To compute (17) we need $\rho_L(\mathbf{R}'; v(\mathbf{R}'))$ and $\mathbf{J}_L(\mathbf{R}'; v(\mathbf{R}'))$ in terms of f_B as \mathbf{R}' varies over the support of $\rho_L(\mathbf{R}'; v(\mathbf{R}'))$, where $v(\mathbf{R}') = (s-z)/\beta_r - |\mathbf{R}' - \mathbf{T}(\xi)|$. Using (11) we obtain

$$\rho_L(\mathbf{R}; u) = \beta_r^2 \int d\mathbf{p} f_B\{s(\mathbf{R}) - \beta_r u, x(\mathbf{R}), \mathbf{p}; s(\mathbf{R})\} |\partial_{p_z} p_s(\mathbf{p})|, \quad (18)$$

and a similar formula for $\mathbf{J}_L(\mathbf{R}; u)$. Thus given $f_B(\cdot; \tau)$ over a suitable τ domain, (17) can be computed. Furthermore, using the fact that f_B is nonnegative and continuous, it follows that the support of $\rho_L(\mathbf{R}; u)$ is the same as the support of $\rho_B\{s(\mathbf{R}) - \beta_r u, x(\mathbf{R}); s(\mathbf{R})\}$.

There are two issues here. First, it would be inefficient to integrate (17) over all of \mathbb{R}^2 . Thus it is important to determine a good superset for the region where the integrand is nonzero. However, this \mathbf{R}' region appears to be difficult to determine efficiently. In the next section we explore two sets of integration variables better suited to the integration in (17). Second, we must deal with the issue that (17) requires knowledge of $f_B(\cdot; \tau)$ for $\tau \notin [0, s]$. to good approximation becomes

We begin by considering $s = 0$. We assume that the support of the initial z -density, $\int dx dp_z dp_x f_{B0}$, is $[-a, a]$. For $s = 0$ we have $z = -\beta_r u$, so the arrival time at $s = 0$ of the particle with z -coordinate ξ is $u = -\xi/\beta_r$. Thus the head of the bunch corresponds to $z = a$ and the tail to $z = -a$. Recall that we defined u_i to be the time at which the head of the bunch enters the chicane, thus $u_i = -a/\beta_r$. To evaluate (17) we need $f_L(\mathbf{R}', \mathbf{P}; -z/\beta_r - |\mathbf{R}' - \mathbf{T}([0, x])|)$ and so we need $f_L(\mathbf{R}', \mathbf{P}; v)$ for $v \in [u_i = -a/\beta_r, a/\beta_r]$. This will be discussed in [2].

We claim that at s we need partial information on $f_B(\cdot; \tau)$ for some $\tau > s$, i.e., there is a causality issue. To demonstrate this we choose $\mathbf{r} = 0$ and $\mathbf{R}' = \mathbf{R}_r(s) + \Delta \mathbf{t}(s)$ in the integrand of (17), where Δ is positive and small relative to the bunch size. Thus the integrand requires the value of $\rho_L(\mathbf{R}_r(s) + \Delta \mathbf{t}(s); s/\beta_r - \Delta)$ and by (18) $f_B\{s(\mathbf{R}_r(s) + \Delta \mathbf{t}(s)) - s + \beta_r \Delta, x(\mathbf{R}_r(s) + \Delta \mathbf{t}(s)), \mathbf{p}; s(\mathbf{R}_r(s) + \Delta \mathbf{t}(s))\}$ is needed. Now applying Computer Codes (Design, Simulation, Field Calculation)

Taylor's theorem $s(\mathbf{R}_r(s) + \Delta \mathbf{t}(s)) = s + \Delta + O(\Delta^2)$ and $x(\mathbf{R}_r(s) + \Delta \mathbf{t}(s)) = O(\Delta^2)$ and we obtain $f_B\{(1 + \beta_r)\Delta + O(\Delta^2), O(\Delta^2), \mathbf{p}; s + \Delta\}$. Thus we need partial information on $f_B(\cdot; s + \Delta)$ and this completes the claim.

A geometrical argument may be more insightful. Take $\mathbf{r} = 0$, then the LF observation point is at $(\mathbf{R}_r(s), s/\beta_r)$. The 'backward lightcone' of this event is $C_L = \{(\mathbf{R}', u') : |\mathbf{R}' - \mathbf{R}_r(s)| = s/\beta_r - u', \mathbf{R}' \in \mathbf{T}(U)\}$ with BF image $C_B = \{(z', x', s') : |\mathbf{R}_r(s') + x' \mathbf{n}(s') - \mathbf{R}_r(s)| = (z' - s' + s)/\beta_r, \xi' \in U\}$. Solving $|\mathbf{R}_r(s') + x' \mathbf{n}(s') - \mathbf{R}_r(s)| = (z' - s' + s)/\beta_r$ for small $\Delta := s' - s$ gives $z' = \Delta + \beta_r \sqrt{\Delta^2 + x'^2}$. Thus there are small positive Δ with (z', x') in the beam, i.e., C_B contains points in the beam $s' > s$. In the drift, this argument even works without the small Δ assumption.

We suspect that whenever causality is violated, within the support of f_B , it is only violated in a small τ interval for $\tau > s$. However, in our applications $f_B(\cdot; s)$ is slowly varying, that is, $f_B(\cdot; s + \Delta) \approx f_B(\cdot; s)$ for $|\Delta|$ less than the z -size of the beam. Thus we believe that to good approximation (17) can be determined from $f_B(\cdot; \tau)$ for $\tau \leq s$, if we take $f_B(\cdot; \tau) = f_B(\cdot; s)$ when $\tau > s$, and this resolves the causality issue. Recall that the IVP (4) defines $f_L(\mathbf{R}, \mathbf{P}; u)$ for $u \geq u_i$, given $f_{L0}(\mathbf{R}, \mathbf{P})$. So the true test of any approximation is, does the $f_B(\mathbf{r}, \mathbf{p}; s)$ we calculate approximately satisfy (9)? We believe our approximations do, but have no proof.

For each s , the computation of \mathcal{F}_L in (17) must be done for a large number of \mathbf{r} values, thus both arguments of \mathcal{F}_L vary. However, since $f_B(\cdot; s)$ is slowly varying we believe that $\mathcal{F}_L(\mathbf{T}([s, x]), (s-z)/\beta_r)$ is also slowly varying in s , for fixed \mathbf{r} . Replacing s by $s + z$ and expanding gives $\mathcal{F}_L(\mathbf{R}_r(s) + x \mathbf{n}(s), (s-z)/\beta_r) \approx \mathcal{F}_L(\mathbf{R}_r(s) + M(s) \mathbf{r}, s/\beta_r)$, where $M(s) = [\mathbf{t}(s), \mathbf{n}(s)]$. Thus to move points, the second argument is independent of z , and this increases the efficiency. This is what we do in the code. There is still a causality issue but it can be resolved as above.

OUTLINE OF CALCULATION OF THE BEAM FRAME SPATIAL DENSITY

Our ultimate goal is to compute the evolution of $f_B(\cdot; s)$ for $s > 0$ given f_{B0} , i.e., the solution of the IVP (13-14), but this is beyond our current capability as mentioned in the Introduction. However the computation of the evolution of the associated spatial density $\rho_B(\cdot; s)$ is possible. From (10) $\partial_{p_z} p_s(\mathbf{p})$ is $1/\beta_r^2$ to good approximation, thus by (11) we can take $\rho_L(\mathbf{R}; u) = \rho_B\{s(\mathbf{R}) - \beta_r u, x(\mathbf{R}); s(\mathbf{R})\}$. Also, to good approximation, \mathbf{J}_L can be determined from ρ_B and τ_B where $\tau_B(\mathbf{r}, s) = \beta_r \int_{\mathbb{R}^2} p_x f_B(\mathbf{r}, \mathbf{p}; s) d\mathbf{p}$ (see [1]-[3]). Thus, from the history of $g_B := (\rho_B, \tau_B)$, the self field (17) can be computed. Our current goal is a fast method to calculate $g_B(\cdot; s)$, for $s > 0$ given f_{B0} , consistent with the IVP (13-14).

We begin by generating an initial, $s = 0$, set of phase space points from f_{B0} using a random number genera-

tor. We currently use pseudo random points and are in the process of investigating quasi random points, which are known to have a significant advantage in numerical integration. From the scattered data we calculate $g_B(\cdot; 0)$ using a density estimation procedure. We follow the phase space points using (15) in the interaction picture. At s we have the scattered set of phase space points and the history of $g_B(\mathbf{r}; \tau)$ for $-2a \leq \tau < s$ on a 3D grid in (\mathbf{r}, τ) . To go from $s \rightarrow s + \delta s$ we first calculate $g_B(\cdot; s)$ from scattered data using well known density estimation procedures from Statistics. Then from the knowledge of $g_B(\cdot; \tau)$ for $-2a \leq \tau \leq s$ we calculate the self field from (17) and use this and (15) to move the points from $s \rightarrow s + \delta s$. The bulk of the calculations are done in parallel.

The main issues for a fast algorithm are the field calculation, the density estimation and the parallel implementation. The latter is elementary, the first two are discussed in the next two sections.

FIELD CALCULATION

The bunch is small and moving. Thus, for each s , it is important to have a good estimate of the location and spatial extent of the bunch. This allows for an s -independent grid on which to represent $g_B(\mathbf{r}; s)$. We first discuss a model we currently use in our bunch compressor studies, which gives a compact, s -independent grid. Then we discuss two sets of integration variables for the integration in (17).

Compact Support Model of Bunch

The location and spatial extent of the bunch depends on the initial phase space density. The one we use in our microbunching studies, [1], is

$$f_{B0}(\mathbf{r}, \mathbf{p}) = (1 + \varepsilon(z))\mu(z)\rho_c(p_z - hz)\rho_t(x, p_x), \quad (19)$$

where $\varepsilon(z) = A \cos(k_0 z)$, μ is a flat top with support $[-a, a]$, ρ_c is $N(0, \sigma_u)$ and ρ_t is $N(0, \text{diag}(\sigma_{x_0}, \sigma_{p_{x_0}}))$, where σ_u and $\sigma_{p_{x_0}}$ are small. Taking $\sigma_u = 0$ and $\sigma_{p_{x_0}} = 0$, the initial conditions for (15) have the form $z = \tilde{z}$, $p_z = h\tilde{z}$, $x = \tilde{x}$, and $p_x = 0$ and without the self field the solution is

$$z = (1 + hR_{56}(s))\tilde{z} - D'(s)\tilde{x}, \quad x = hD(s)\tilde{z} + \tilde{x}. \quad (20)$$

Here h is the chirp parameter and $D(s)$ and $R_{56}(s)$ were introduced in the context of (15). Using (\tilde{z}, \tilde{x}) as coordinates we find that the support of $g_B(\cdot; s)$ is essentially independent of s .

To obtain our compact support model we scale the tilde variables and our final transformation is $\mathbf{r} = A(s)\tilde{\mathbf{r}} = A(s)\Sigma\hat{\mathbf{r}} =: B(s)\hat{\mathbf{r}}$. Here Σ is a diagonal matrix chosen so that the support of the beam is just inside the circle $\hat{\mathbf{r}}^T \hat{\mathbf{r}} = 1$. Thus the support of $g_B(\mathbf{r}; s)$ is given by

$$\mathbf{r}^T E(s) \mathbf{r} \leq 1, \quad E(s) = B(s)^{-T} B(s)^{-1}, \quad (21)$$

Computer Codes (Design, Simulation, Field Calculation)

that is, it is the interior of an ellipse. Using (21) to determine the \mathbf{R}' region of integration seems complicated. We now introduce two sets of integration variables which simplify the determination of the support and the task of integration.

Polar Coordinates

We transform to polar coordinates (χ, θ) in (2) and then take the temporal argument v in place of the radial coordinate χ . That is, we make the transformation $\mathbf{R}' \rightarrow (\theta, v)$ via $\mathbf{R}' = \mathbf{R} + \chi \mathbf{e}(\theta)$, $\mathbf{e}(\theta) = (\cos \theta, \sin \theta)^T$, $v = u - \sqrt{\chi^2 + (kh)^2}$.

This removes the integrable singularity giving the field simply as an integral over the source,

$$\mathcal{F}_L(\mathbf{R}, u) = -\frac{1}{4\pi} \int_{u_i}^u dv \int_{-\pi}^{\pi} d\theta \mathbf{S}[\mathbf{R} + (u-v)\mathbf{e}(\theta), v], \quad (22)$$

where we have ignored shielding. Using the slowly varying assumption, as discussed at the end of Section 4, we obtain

$$\begin{aligned} \mathcal{F}(\cdot, (s-z)/\beta_r) &\approx \mathcal{F}(\mathbf{R}_r(s) + M(s)\mathbf{r}, s/\beta_r) \\ &= -\frac{1}{4\pi} \int_{u_i}^{s/\beta_r} dv \int_{-\pi}^{\pi} d\theta \mathbf{S}[\tilde{\mathbf{R}}(\theta, v; \mathbf{r}, s), v], \end{aligned} \quad (23)$$

where $\tilde{\mathbf{R}}(\theta, v; \mathbf{r}, s) = \mathbf{R}_r(s) + M(s)\mathbf{r} + (s/\beta_r - v)\mathbf{e}(\theta)$. This is what we must calculate, at arc length s , as \mathbf{r} varies over the bunch. Note that for $\mathbf{r} = 0$ the approximation is exact. Also taking $\mathbf{r} = 0$, $v = s/\beta_r - \Delta$ and $\mathbf{e}(\theta) = \mathbf{t}(s)$, we see the same causality issue as before and resolve it in the same way.

Except for v close to s/β_r the θ support in (23) is tiny and it is important for a fast algorithm to compute this accurately. Using the slowly varying approximation and (18), it can be shown that the support of $\mathbf{S}(\tilde{\mathbf{R}}, v)$ is, to good approximation, the same as $\rho_B(M^T(\beta_r v)(\tilde{\mathbf{R}} - \mathbf{R}_r(\beta_r v)), \beta_r v)$. The support of the latter is given by (21) with \mathbf{r} replaced by $M^T(\beta_r v)(\tilde{\mathbf{R}} - \mathbf{R}_r(\beta_r v))$ and s replaced by $\beta_r v$. This gives a quartic in $\exp(i\theta)$ and solving this gives reasonable θ limits. This is discussed in some detail in [1] and [3].

Currently the θ integration is done with the superconvergent trapezoidal rule. The remaining v -integrand varies with v , \mathbf{R} and u in ways we have not yet quantified and so we use an adaptive integrator based on the Gauss-Kronrod algorithm. We are investigating two improvements. The adaptive integrator is slow and we are studying the v dependence of the integrand after the θ integration with the hope of using a non-adaptive algorithm. Solving the quartic to determine the θ support may not be the best approach. The ellipse in (21) is quite elongated and we are investigating replacing it by a parallelogram which should simplify the calculation.

The computational effort for the calculation of one component of the self field is $O(N_z N_x N_v N_\theta)$, where N_z and N_x are the number of grid points in \hat{z} and \hat{x} respectively, N_v is the number of evaluations for the v integrand.

tion, and N_θ is the number of evaluations for the θ integration. Typical values for our simulations in [1] are $N_z = 1000$, $N_x = 128$, $N_v = N_\theta = 1000$, therefore $O(N_z N_x N_v N_\theta) = O(10^{12})$.

Beam Frame Coordinates

The transformation $\mathbf{R}' \rightarrow \xi'$ via $\mathbf{R}' = \mathbf{T}(\xi')$ in (2) gives

$$\mathcal{F}_L(\mathbf{T}(\xi); (s - z)/\beta_r) = -\frac{1}{4\pi} \int_{\mathbb{R}^2} d\xi' \frac{\mathcal{K}(\xi')}{|\mathbf{T}(\xi') - \mathbf{T}(\xi)|} \times \mathbf{S}[\mathbf{T}(\xi'); \frac{s - z}{\beta_r} - |\mathbf{T}(\xi') - \mathbf{T}(\xi)|]. \quad (24)$$

The nonsingular part of the integrand for $\mathbf{E}_{||}$ is given by

$$\begin{aligned} & \mathcal{K}(\xi') \mathbf{S}_{||}[\mathbf{T}(\xi'); \frac{s - z}{\beta_r} - |\mathbf{T}(\xi') - \mathbf{T}(\xi)|] / Q Z_0 c \\ &= [(\frac{1}{\gamma^2} - \beta_r^2 x' \kappa(s')) D_1 \rho_B() + D_3 \rho_B()] \mathbf{t}(s') \\ &+ \mathcal{K}(\xi') [D_2 \rho_B() + D_1 \tau_B()] \mathbf{n}(s'), \end{aligned} \quad (25)$$

where $() = (z', x'; s')$, $z' = s' - s + z - \beta_r |\mathbf{T}(\xi') - \mathbf{T}(\xi)|$ and $\mathbf{S}_{||}$ is defined in the obvious way. Thus the support of the integrand is the support of $g_B(z', x'; s')$. For each s' , the x' region of integration can now be determined from (21) by solving $\mathbf{r}(\xi')^T E(s') \mathbf{r}(\xi') = 1$ where $\mathbf{r}(\xi')^T = (s' - s + z - \beta_r |\mathbf{T}(\xi') - \mathbf{T}(\xi)|, x')$. This is a quartic equation in x' , which is not so surprising as x' is analogous to θ in the previous subsection. We can now write (24) as an iterated integral as in (23).

We are in the process of comparing this with the polar coordinate approach. The integrand is given naturally in terms of g_B which is nice. For example, one sees the factor, $D_3 \rho_B$, which gives rise to the compression and is not there for a rigid bunch. Also, studies by Warnock suggest (i) this is a natural framework for the study of 1D collective force approximations, (ii) it could lead to a good approximation for s' near s , which includes the singularity, and (iii) a multipole expansion for s' a few bunch lengths from s is possible and may save time and storage. We are pursuing these ideas.

DENSITY ESTIMATION

Density estimation is a significant and active area of Statistics, see for example [4] and [5]. We compare the computational effort of three methods for the same level of accuracy.

One approach to density estimation is based on orthogonal series and we have studied the Fourier series case in some detail following [4]. Here ρ_L and \mathbf{J}_L are obtained at every s step by computing the Fourier coefficients of the truncated Fourier series via Monte Carlo integration of the random sample of phase space points. Details are given in [1] and [3]. The computational effort is $O(\mathcal{N} J_z J_x) + O(N_z N_x J_z J_x)$, where \mathcal{N} is the number of simulated points, and J_z and J_x are the number of Fourier Computer Codes (Design, Simulation, Field Calculation)

coefficients in \hat{z} and \hat{x} respectively. Typical values in our microbunching simulations are $\mathcal{N} = 5 \times 10^8$, $J_z = 150$ and $J_x = 50$. Therefore the computational effort is $O(10^{12})$ and is of the same order as the computational effort for the polar coordinate field calculation discussed previously.

A second approach employs cloud in cell charge deposition where at every s step the random sample is placed on our fixed grid (See Section 3.5 of [5] and [6]). Here ρ_L and \mathbf{J}_L are obtained by computing the Fourier coefficients of the truncated Fourier series by a simple quadrature. The computational effort in this case is $O(\mathcal{N}) + O(N_z N_x J_z J_x)$. We have found that using \mathcal{N} , J_z , J_x as above, $N_z = 1000$ and $N_x = 128$, gives the same approximation as the Monte Carlo approach of the previous paragraph. This computational effort of $O(10^9)$ is much smaller than for the orthogonal series method and negligible with respect to the computational effort for the polar coordinate field calculation. This is the present method implemented in our code.

A third approach applies kernel density estimator techniques to the random sample. This approach is still in the testing phase where we are investigating standard kernels like bivariate Gaussians or bivariate Epanechnikov kernels (all with a uniform bandwidth, h). The computational effort for the bivariate Epanechnikov kernel is $O(\mathcal{N} \tilde{N}_z \tilde{N}_x)$, where \mathcal{N} is as before but now $\tilde{N}_z \tilde{N}_x$ is the number of grid points inside the square which encloses the circle of radius h centered at the scattered particle position \hat{z}, \hat{x} . For $\mathcal{N} = 5 \times 10^8$, $N_z = 1000$, $N_x = 128$ we approximately get $\tilde{N}_z = 24$, $\tilde{N}_x = 3$, $O(\mathcal{N} \tilde{N}_z \tilde{N}_x) = O(10^{10})$. Thus this method is comparable in speed to the second method and is worthy of further investigation.

So far, for all three approaches, the initial random sample is generated from pseudo-random numbers but quasi-random numbers will be tested soon.

ACKNOWLEDGMENTS

We gratefully acknowledge Bob Warnock for useful discussions.

REFERENCES

- [1] G. Bassi, J. A. Ellison, K. Heinemann, and R. Warnock, Phys. Rev. ST Accel. Beams, **12**, 080704 (2009).
- [2] G. Bassi, J.A. Ellison, K. Heinemann and R. Warnock, "Relation of phase space densities in laboratory and beam centered coordinates", in preparation.
- [3] J. A. Ellison, G. Bassi, K. Heinemann, M. Venturini, and R. Warnock, Proc. PAC07, Albuquerque, USA.
- [4] S. Efromovich, "Nonparametric Curve Estimation: Methods, Theory, and Applications", (Springer, Berlin, 1999).
- [5] B.W. Silverman, "Density Estimation", (Chapman and Hall, 1986).
- [6] B. Terzic, I.V. Pogoderov, C. Bohn, Phys. Rev. ST Accel. Beams, **10**, 034201 (2007).

SIMULATION OF MICROWAVE INSTABILITY IN LER OF KEKB AND SUPERKEKB

D. Zhou*, K. Ohmi, K. Oide, Y. Suetsugu, K. Shibata, KEK, 1-1 Oho, Tsukuba 305-0801, Japan

Abstract

Microwave instability in the LER of KEKB may be one obstacle to achieving high luminosity as expected by beam-beam simulations. To understand the single-bunch beam dynamics of KEKB LER, we constructed a numerical impedance models by calculating ultra-short wake potentials of various vacuum components, resistive wall and coherent synchrotron radiation. The geometrical wakes were calculated by 3D electromagnetic code GdfidL. And CSR impedance were estimated by a dedicated code. Similar work was also done for LER of SuperKEKB. Using these impedance models we simulated the microwave instability at LER of KEKB and SuperKEKB by solving Vlasov-Fokker-Planck (VFP) equation in the longitudinal phase space. The results of impedance calculation and simulations were presented in this paper.

INTRODUCTION

KEKB [1] has been operated for more than 10 years since its first commissioning from Dec. 1, 1998. In June 2009, the peak luminosity reached $2.11 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ with stored beam currents of 1.64/1.12A (LER/HER) due to crab crossing and off-momentum optics corrections. One of the merits of KEKB [2] which contributed to such high luminosity is squeezing the vertical beta function at interaction point (IP) to 0.59 cm. Correspondingly, the natural bunch length is around 4.6 mm. And at normal operating bunch current of 1.0 mA at LER, the measured bunch length is around 7 mm.

Since the beam-beam simulations showed that the crab crossing should boost the luminosity by a factor of 2 [3], the present achieved luminosity is still far from expectations. Besides chromatic coupling induced by lattice non-linearity [4], microwave instability in the LER may be another potential obstacle for KEKB to achieving higher luminosity by way of increasing beam currents.

Recently, Y. Cai et al. studied the microwave instability in the LER of KEKB using a broadband resonator impedance model [5]. In that work, it was demonstrated that the model described the longitudinal beam dynamics very well when comparing with experimental observations. As predicted by Cai's model, the threshold of microwave instability at LER of KEKB is 0.5 mA, which is well lower than the present operating current of 1.0 mA. In this paper, we introduce the studies on microwave instability in the LER of KEKB and SuperKEKB using numerically calculated impedance models.

QUASI GREEN'S FUNCTION OF WAKE POTENTIAL

To study the longitudinal single-bunch instabilities, we first calculate the ultra-short wake potentials of various vacuum components. GdfidL installed on a cluster with 256 GB memory is available at KEK. As trade-off between the capability of the cluster and the interested frequency range, 0.5 mm bunch length was chosen for most vacuum components of KEKB LER.

Fig. 1 shows the total geometrical wake potentials of LER of KEKB and SuperKEKB. The length of driving gaussian bunch used in GdfidL is 0.5 mm. Due to significant improvements in the vacuum components, the impedance of SuperKEKB rings will be well suppressed. Coherent synchrotron radiation (CSR) is another important impedance source at LER of KEKB and SuperKEKB. The bending radius of normal dipoles at KEKB LER and wigglers are 15.87 m and 16.3 m, respectively. For SuperKEKB LER, only half of the wigglers will remain. Such magnets will produce CSR as bunch length get short to a few minimeter. Thus a dedicated code was developed by K. Oide in 2008 to calculate the CSR impedance in LER of SuperKEKB. In this code, the paraxial approximation was adopted [6]. Electronic fields due to CSR were calculated in the frequency domain and then wake potential was obtained by Fourier transformation. The calculated CSR wake potentials of 0.5 mm bunch are shown in Fig. 2. Interference between adjacent magnets caused modulations at the tail parts of the CSR wake potentials.

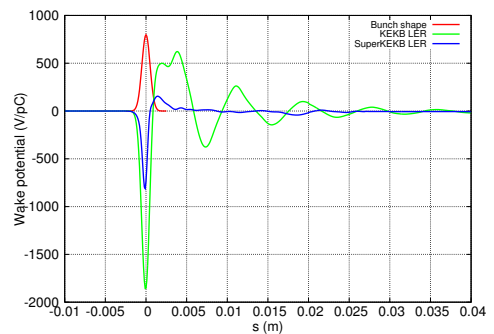


Figure 1: Calculated geometrical wake potentials of 0.5 mm bunch for LER of KEKB and SuperKEKB.

* dmzhou@post.kek.jp

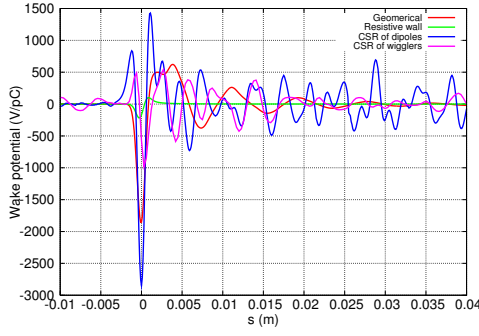


Figure 2: Calculated geometrical, resistive wall and CSR wake potentials of 0.5 mm bunch for LER of KEKB.

SOLVING VFP EQUATION

Basically we follow R.L. Warnock and J.A. Elissson's work [7] to solve Vlasov-Fokker-Planck (VFP) equation numerically. VFP equation including collective wake force is written as

$$\frac{\partial \psi}{\partial s} + \frac{\partial q}{\partial s} \cdot \frac{\partial \psi}{\partial q} + \frac{\partial p}{\partial s} \cdot \frac{\partial \psi}{\partial p} = \frac{2\beta}{c} \frac{\partial}{\partial p} [p\psi + \sigma_p^2 \frac{\partial \psi}{\partial p}] \quad (1)$$

$$\psi = \psi(q, p, s) \quad (2)$$

where $\psi(q, p, s)$ is the probability density in the longitudinal phase space and is normalized as $\iint \psi(q, p, s) dp dq = 1$. $q = z$ is the longitudinal coordinate and $p = \Delta p/p_0$ is the relative momentum deviation. The corresponding longitudinal distribution is calculated from $\psi(q, p, s)$ as

$$\lambda(q, s) = \int \psi(q, p, s) dp \quad (3)$$

and will be used in calculating wake forces. The Hamiltonian's equations are

$$\frac{\partial q}{\partial s} = \frac{\omega_s \sigma_z}{c \sigma_p} \cdot p \quad (4)$$

$$\frac{\partial p}{\partial s} = -\frac{\omega_s \sigma_p}{c \sigma_z} \cdot q - I_n \cdot F(q, s) \quad (5)$$

where $I_n = \frac{Ne^2}{E_0 C}$ and wake force is

$$F(q, s) = \int_{q'=-\infty}^{\infty} W_0(q' - q) \lambda(q') dq' \quad (6)$$

The unit of W_0 is V/pC. N is bunch population. C is circumference of the ring. E_0 is the design energy of the ring.

Operator splitting

The technique of operator splitting [8, 9], or called time splitting, is widely used in solving partial differential equations (PDEs). To solve the VFP equation, we rewrite the VFP Eq. 1 and split the operators into three parts:

$$\frac{\partial \psi}{\partial s} = \mathcal{L} \psi = \left(\sum_{i=1}^3 \mathcal{L}_i \right) \psi \quad (7)$$

The solution of Eq. 7 is given by

$$\psi^{n+1} = e^{\Delta s \mathcal{L}} \psi^n \quad (8)$$

where

$$\mathcal{L}_1 = -\frac{\omega_s \sigma_z}{c \sigma_p} \cdot p \cdot \frac{\partial}{\partial q} + \frac{\omega_s \sigma_p}{c \sigma_z} \cdot q \cdot \frac{\partial}{\partial p} \quad (9)$$

$$\mathcal{L}_2 = I_n \cdot F(q, s) \cdot \frac{\partial}{\partial p} \quad (10)$$

$$\mathcal{L}_3 = \frac{2\beta}{c} \frac{\partial}{\partial p} [p + \sigma_p^2 \frac{\partial}{\partial p}] \quad (11)$$

The \mathcal{L}_1 and \mathcal{L}_2 represents Liouville operator, and the \mathcal{L}_3 is called Fokker-Planck operator. The Liouville operator is reversible and the Fokker-Planck operator is irreversible. A simple first-order splitting is formulated as

$$e^{\Delta s \mathcal{L}} \approx e^{\Delta s \mathcal{L}_1} e^{\Delta s \mathcal{L}_2} e^{\Delta s \mathcal{L}_3} \quad (12)$$

And high-order splitting instead of Eq. 12 can be applied to achieve better approximation. For example, second-order symmetric splitting scheme can be adopted

$$e^{\Delta s (\mathcal{L}_1 + \mathcal{L}_2)} \approx e^{\Delta s/2 \mathcal{L}_1} e^{\Delta s \mathcal{L}_2} e^{\Delta s/2 \mathcal{L}_1} \quad (13)$$

To get good approximation, usually we split the one-turn map in the ring to k integration steps. This scheme can be written as

$$e^{C \mathcal{L}} \approx [e^{C/k \mathcal{L}_1} e^{C/k \mathcal{L}_2} e^{C/k \mathcal{L}_3}]^k \quad (14)$$

Discrete operator

The discrete version of Liouville operator is Frobenius-Perron operator. Let \mathcal{F}_1 and \mathcal{F}_2 are the Frobenius-Perron operators corresponding to \mathcal{L}_1 and \mathcal{L}_2 , then the evolution of probability density corresponding to reversible operators can be evaluated as

$$\psi^*(q, p) = \mathcal{F}_1 \psi(q, p, n \Delta s) = \psi(R^{-1}(q, p), n \Delta s) \quad (15)$$

$$\psi^{**}(q, p) = \mathcal{F}_2 \psi^*(q, p) = \psi^*(K^{-1}(q, p)) \quad (16)$$

where the rotation mapping R is

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} \cos(\mu_s \Delta s / C) & \beta_z \sin(\mu_s \Delta s / C) \\ -\sin(\mu_s \Delta s / C) / \beta_z & \cos(\mu_s \Delta s / C) \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix} \quad (17)$$

and the kick mapping K is

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} q \\ p - I_n F(q, s) \Delta s / C \end{bmatrix} \quad (18)$$

For Fokker-Planck operator, we propose exponentially fitting scheme (EFS) [10] for discretization

$$\frac{\psi_i^{n+1} - \psi_i^n}{\Delta s} = \frac{2\beta}{c} \psi_i^{n+1} + \frac{\beta p_i (\psi_{i+1}^{n+1} - \psi_{i-1}^{n+1})}{c \Delta p} + \rho_i^{n+1} \frac{2\beta \sigma_p^2 \psi_{i+1}^{n+1} - 2\psi_i^{n+1} + \psi_{i-1}^{n+1}}{c \Delta p^2} \quad (19)$$

Table 1: Main parameters of KEKB LER

Parameter	Value	Unit
Circumference	3016.25	m
Beam energy	3.5	GeV
Bunch population	6.6	10^{10}
Natural bunch length	4.58	mm
Synchrotron tune	0.024	
Longitudinal damping time	2000	turn
Energy spread	7.27	10^{-4}

where

$$\rho_i^{n+1} = \frac{p_i \Delta p}{2\sigma_p^2} \coth \frac{p_i \Delta p}{2\sigma_p^2} \quad (20)$$

As proved in [10], the EFS has the properties of 1) being uniformly stable for all values of integration step Δs , damping coefficient β and mesh size Δp ; 2) being oscillation free.

SIMULATION RESULTS

Based on the algorithms described in the previous section, a code of VFP solver was developed and used in simulations of microwave instability at LER of KEK and SuperKEKB. The main parameters of KEKB LER are listed in Table 1. For SuperKEKB LER, we choose bunch length as 5 mm and other parameters as the same as KEKB LER.

The numerical impedance model predicts much weaker bunch lengthening against measurements [11, 12] as shown in Fig. 3. A pure inductance of around 90 nH should be added in this impedance model to get similar bunch lengthening. But when the pure inductance was added, the threshold of MWI get much higher as shown in Fig. 4. This disagreement indicates that there are unknown impedance sources in the KEKB LER. According to Fig. 4, threshold of MWI with CSR impedance is around 0.7 mA. Without CSR, the threshold is around 1.1 mA. It can be concluded that CSR is important source to drive microwave instability.

For bunch length of 5 mm, no serious bunch lengthening and energy spread growth are seen with bunch current up to 1.6 mA at SuperKEKB LER, as shown in Fig. 5 and Fig. 6.

SUMMARY AND DISCUSSIONS

Accurate impedance model is essential for studying microwave instability. When comparing with beam observations, the numerical impedance model for KEKB LER gave insufficient bunch lengthening and higher threshold for MWI. The discrepancy between numerical model and measurements are around 90nH in case of bunch lengthening.

CSR in storage rings like KEKB LER was not well understood yet. Interference between adjacent bending magnet instabilities and feedback

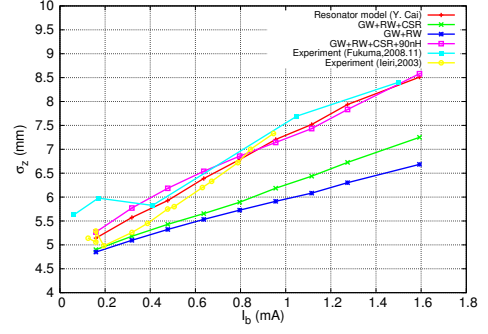


Figure 3: Bunch length as function of bunch current at KEKB LER. The resonator model is given in [5]

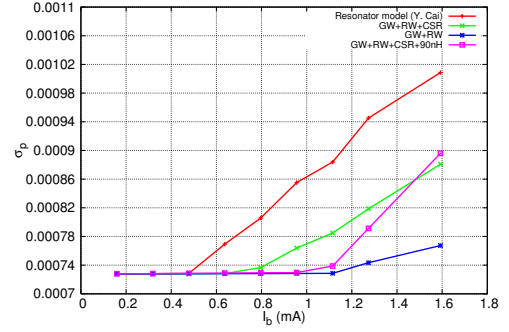


Figure 4: Energy spread as function of bunch current at KEKB LER.

nets seems to be strong. Thus more benchmarking on the CSR code is needed.

ACKNOWLEDGEMENTS

The author D. Zhou would like to acknowledge the members of KEKB group for careful beam measurements. Sincere thanks are due to M. Zobov (INFN) and Y. Cai (SLAC) for valuable discussions.

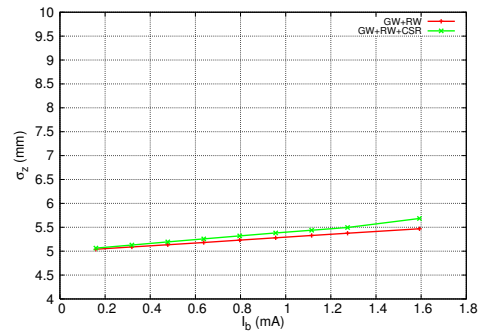


Figure 5: Bunch length as function of bunch current at SuperKEKB LER.

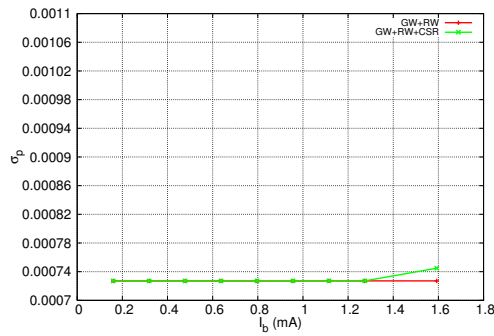


Figure 6: Energy spread as function of bunch current at SuperKEKB LER.

REFERENCES

- [1] KEKB B-Factory Design Report, KEK-Report-95-7, (1995).
- [2] K. Oide, "KEKB B-Factory, The Luminosity Frontier", Progress of Theoretical Physics, Vol. 122, No. 1, July 2009.
- [3] K. Ohmi, et al., "Luminosity Limit Due to the Beam-Beam Interactions With or Without Crossing Angle, Phys. Rev. ST Accel. Beams 7, 104401 (2004).
- [4] D. Zhou, et al., "Simulations on Beam-Beam Effect Under the Presence of General Chromaticity", KEK Preprint 2009-10, May 2009.
- [5] Y. Cai, et al., "Potential-Well Distortion, Microwave Instability, and Their Effects with Colliding Beams at KEKB", Phys. Rev. ST Accel. Beams 12, 061002 (2009).
- [6] T. Agoh and K. Yokoya, "Calculation of Coherent Synchrotron Radiation Using Mesh", Phys. Rev. ST Accel. Beams 7, 054403 (2004)
- [7] R. Warnock and J. Ellison, "A General Method for Propagation of Phase Space Distribution, with Application to the Sawtooth Instability", SLAC-PUB-8404, 2000.
- [8] S. I. Tzenov, "Contemporary Accelerator Physics", Singapore, World Scientific, p. 217 (2004).
- [9] W. Hundsdorfer and J. G. Verwer, "Numerical Solution of Time-Dependent Advection-Diffusion-Reaction equations", Springer series in Computational Mathematics, vol. 33, Springer, Berlin, p. 325 (2003).
- [10] D.J. Duffy, "A Critique of the Crank Nicolson Scheme Strengths and Weaknesses for Financial Instrument Pricing", Wilmott magazine, p. 68-76, July 2004.
- [11] T. Ieiri and H. Koiso, "Measurement of Longitudinal Impedance at KEKB", Proc. of 14th Symposium on Accelerator Science and Technology, KEK, Tsukuba, p. 443 (2003).
- [12] H. Ikeda, et al., "Negative Momentum Compaction at KEKB", Proc. of Workshop on e^+e^- in the 1-2 GeV range: Physics and Accelerator Prospects, Alghero (SS), Italy, September 2003.

STUDY OF BEAM-SCATTERING EFFECTS FOR A PROPOSED APS ERL UPGRADE *

A. Xiao [†], M. Borland, X. Dong, ANL, Argonne, IL 60439, USA

Abstract

Beam-scattering effects, including intra-beam scattering (IBS) and Touschek scattering, may become an issue for linac-based 4th-generation light sources, such as X-ray free-electron lasers (FELs) and energy recovery linacs (ERLs), as the electron density inside the bunch is very high. In this paper, we describe simulation tools for modeling beam-scattering effects that were recently developed at the Advanced Photon Source (APS). We also demonstrate their application to a possible ERL-based APS upgrade. The beam loss issue due to the Touschek scattering effect is addressed through momentum aperture optimization. The consequences of IBS for brightness, FEL gain, and other figures of merit are also discussed. Calculations are performed using a particle distribution generated by an optimized high-brightness injector simulation.

INTRODUCTION

The Coulomb scattering between particles inside a beam has been widely studied for circular accelerators. They were largely ignored for linacs in the past, since significant effects are not expected for one-pass, low-repetition-rate systems with relatively large beam size. The scattering rate is quite low, and there is not enough time for the beam to develop any noticeable diffusion. The situation has dramatically changed since linac-based 4th-generation light sources are on the horizon. To provide users with synchrotron radiation with unprecedented high brightness, the required linac beam must have extremely low emittance with significant charge and a high repetition rate. To ensure that the machine can be run safely with acceptable beam losses and that the beam quality will be not harmed by IBS, we developed a series of simulation capabilities in elegant [1]. They provide the ability to simulate beam-scattering effects for an arbitrarily distributed linac beam with energy variation.

Beam-scattering effects are traditionally separated into two categories, Touschek effect and IBS, based on whether the scattered particles are lost immediately after the scattering event or not, respectively. In the case of IBS, we only see diffusion that leads to increased emittance in 6-D phase space; whereas in Touschek, a single scattering event may result in loss of the scattered particles. Different theoret-

ical approaches are used to calculate the beam size diffusion rate and beam loss rate. In developing our simulation tools, we followed the same path: the widely used Bjorken-Mtingwa's [2] formula is chosen for calculating the emittance growth rate due to the IBS effect, while a combination of Piwinski's formula and Monte Carlo simulation is used for determination beam loss rates and positions.

Both the Bjorken-Mtingwa formula and Piwinski's formula were developed for stored beam, which has constant energy, and both assume a Gaussian bunch. These assumptions are generally invalid for a linac beam. In previous papers [3, 4, 5, 6], we discussed the beam loss issue for a one-pass transport system (Gaussian beam, constant beam energy), and the IBS for a arbitrarily distributed accelerating beam. In this paper, we describe newly developed methods that give us the ability to simulate the beam loss for an arbitrarily distributed linac beam, and summarize the already existing IBS tools. We also give an example application to a possible ERL-based APS upgrade design [7] using a particle distribution generated by an optimized high-brightness injector simulation [8].

A PROPOSED APS ERL UPGRADE

The APS has an eye on building an ERL for a future upgrade. Figure 1 shows the layout of one proposed design. The existing APS ring is used as part of the new machine. Since the radiation shielding of the APS already exists, there is concern about beam loss rate from the high-average-current ERL beam. Also, because of energy recovery, we will find that a small energy deviation generated at high energy may exceed the energy aperture at the end of deceleration, resulting in beam loss. Therefore, a detailed simulation tool that can determine the beam loss rate and the beam loss position precisely is needed.

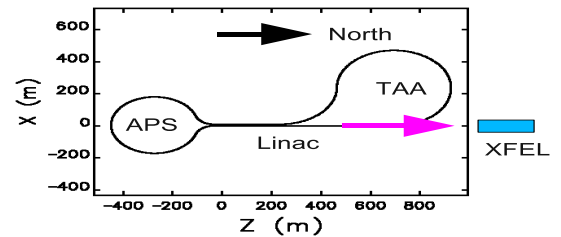


Figure 1: Layout of a proposed APS ERL upgrade.

In general, a linac beam departs from the normally assumed Gaussian distribution that holds for a stored beam,

* Work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[†] xiaoam@aps.anl.gov

especially in the longitudinal dimension. Figure 2 shows the particle distribution from our optimized high-brightness injector simulation. This distribution is used in the later examples of our simulation tools. The major beam parameters of the simulated bunch are: $\varepsilon_{x,n} = \varepsilon_{y,n} = 0.35\mu m$, $\sigma_p = 2.63 \cdot 10^{-3}$, $\sigma_l = 0.6mm$, bunch charge = $77pC$, and repetition frequency = $1.3GHz$.

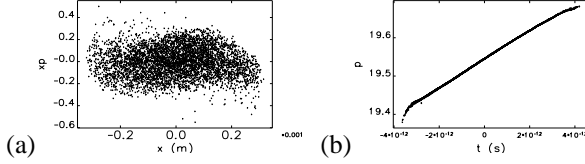


Figure 2: Particle distribution from optimized high-brightness injector simulation: (a) horizontal, (b) longitudinal ($p=\beta\gamma$).

TOUSCHEK EFFECT

Simulation of beam loss due to Touschek effect is performed in several steps. The procedure is illustrated in Figure 3. To start the simulation, the beamline under study is first divided into many small sections by inserting a special element TSCATTER into it. (This is easily done using the elegant command `insert_elements`.) The total number and locations where one should insert TSCATTER elements depends on the rapidity with which the energy and optical functions vary. To ensure reliable results, these variations should be small between successive scattering elements.

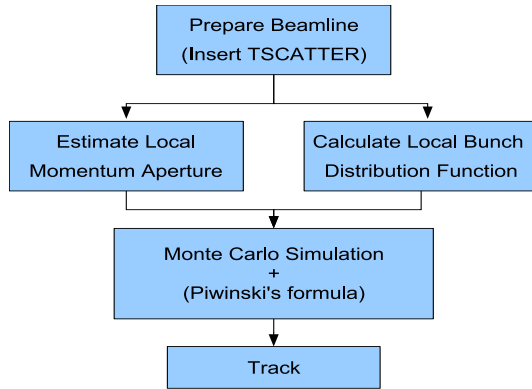


Figure 3: Procedure of the Touschek-caused beam loss simulation.

Estimate local momentum aperture

Unlike in storage rings, the fractional momentum aperture varies over a large range in linacs, as illustrated in Figure 4. In order to efficiently study beam loss from Touschek scattering, we need to know the approximate local momentum aperture and use these results for later simulation [9].

`elegant` provides the `momentum_aperture` command to determine local momentum aperture. We also added a Multi-Particle Beam Dynamics

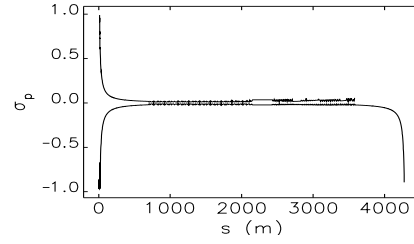


Figure 4: Local momentum aperture for example APS ERL lattice (tracking stopped before last rf module ($E=189$ MV)).

switch in the Touschek simulation module that can generate a momentum aperture bunch at each TSCATTER position and then track the bunch to the end of the beamline. The lost particles are collected and an approximate local momentum aperture is obtained. Figure 5 shows an example of the aperture bunch and lost particles.

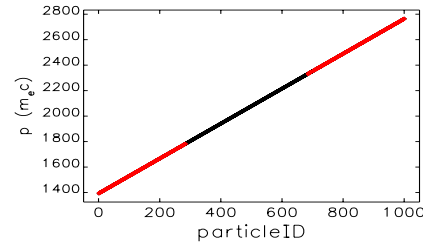


Figure 5: A momentum aperture bunch and lost particles. All particles have $x(x', y, y', t) = 0$.

The results obtained from this step are very important for performing an economical yet detailed Monte Carlo simulation in subsequent steps.

Calculate local bunch distribution function

The electron bunch from a high-brightness injector is typically not Gaussian distributed, especially in the longitudinal direction. As shown in Figure 2, the energy spread of the entire bunch is more than two orders of magnitude larger than the “intrinsic” energy spread. The traditional formulae, which calculate the beam-scattering effect based on the assumption of a Gaussian beam, are therefore invalid. To match our simulation result more closely to the real machine, we track the simulated electron bunch from the gun through the beamline with inserted TSCATTER elements. At each TSCATTER position, the tracked particles’ coordinates are saved so that a corresponding distribution histogram can be made. This distribution histogram (table) is read back later by the Monte Carlo simulation module in order to obtain the particle density by interpolation of the table.

To accomplish this task, we wrote a general tool to generate an n -dimension histogram for a collection of n -tuples, where n can be any integer. Of course, it’s natural to build a 6D histogram from a particle distribution. One concern is that, in order to have a meaningful 6D histogram, we need a

huge number of particles from the injector simulation. For example, to have 11 bins in each dimension, the total number of bins is about $11^6 \approx 2 \cdot 10^6$, and the total number of particles need to be larger than this value. Another concern is to interpolate in 6D, any point's value is determined by the nearby $2^6 = 64$ grid point values. This calculation is in principle possible using our software. However, to demonstrate our method in this paper, we separated the beam distribution into two parts: transverse (4D) and longitudinal (2D), see Figure 6. The simulated gun bunch has 500,000 particles.

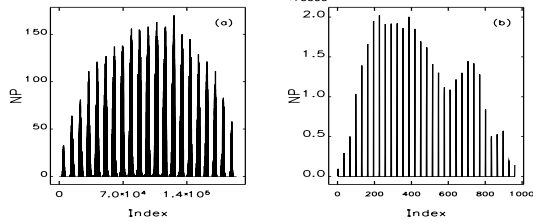


Figure 6: Histogram of the simulated bunch at the beginning of the APS ELR beamline: (a) transverse (4D), (b) longitudinal (2D). (The index is an n-bit counter where each bit has the size of the number of bins of the corresponding dimension.)

Monte Carlo simulation

In the center-of-mass (CM) system¹, the probability of one of the two interacting particles being scattered into a solid angle $d\Omega^*$ is given by the differential Møller cross section [10]

$$\frac{d\sigma^*}{d\Omega^*} = \frac{r_e^2}{4\gamma^{*2}} \left[\left(1 + \frac{1}{\beta^{*2}} \right)^2 \frac{4 - 3\sin^2 \Theta^*}{\sin^4 \Theta^*} + \frac{4}{\sin^2 \Theta^*} + 1 \right], \quad (1)$$

where r_e is the classical electron radius; γ^* and β^* are the relative energy and velocity of scattered electrons in the CM system, respectively; Θ^* is the angle between the momenta before and after scattering; and $d\Omega^* = \sin \Theta^* d\Theta^* d\Psi^*$.

The Touschek scattered particles' distribution is modeled by Monte Carlo simulation. To use Equation (1), a pair of particles with same position (x, y, t) are generated randomly. Their momenta are transformed into the CM system using the Lorentz transformation. In the CM system, the scattering angles (Θ^*, Ψ^*) are selected randomly. The scattered particles' momenta, together with the associated Møller cross section, are then transformed back to the laboratory coordinate system. Therefore, a single random scattering event includes 11 random numbers (3 positions, 6 momenta, and 2 scattering angles).

The total scattering rate R is given by integral over all possible scattering angles and over all electrons in the

bunch. In the CM system,

$$R^* = 2 \int |v^*| \sigma^* \rho(\vec{x}_1^*)^* \rho(\vec{x}_2^*)^* dV^*, \quad (2)$$

where v^* is the scattered electrons' velocity, σ^* is the total Møller cross section, $\vec{x}^* = (x^*, y^*, z^*, p_x^*, p_y^*, p_z^*)$, $\rho(\vec{x}_i^*)^*$ is the electron phase-space density, and $dV = dx^* dy^* dz^* dp_{x1}^* dp_{y1}^* dp_{z1}^* dp_{x2}^* dp_{y2}^* dp_{z2}^*$. σ^* is integrated over the solid angle $d\Omega^*$ with $\Theta^* \in (0, \frac{\pi}{2}]$, $\Psi^* \in [0, 2\pi]$:

$$\sigma^* = \int_0^{2\pi} \int_0^{\pi/2} \frac{d\sigma^*}{d\Omega^*} \sin \Theta^* d\Theta^* d\Psi^*. \quad (3)$$

The reason for $\Theta^* \in (0, \frac{\pi}{2}]$ is that, if one electron is scattered into the region $0 < \Theta^* \leq \frac{\pi}{2}$, then the other is scattered into the region $\frac{\pi}{2} \leq \Theta^* < \pi$. The factor "2" in Equation (2) includes both regions.

For the problem we are interested in, we assume that $p_x \ll p_z$ and $p_y \ll p_z$, which means that the Lorentz transformation is mainly taking place along the z direction, and σ^* is parallel to the z^* -axis. Transforming to the laboratory coordinate system gives

$$|v|\sigma = \frac{|v^*|}{\gamma} \frac{\sigma^*}{\gamma} \quad (4)$$

and

$$R = 2 \int |v| \sigma \rho(\vec{x}_1) \rho(\vec{x}_2) dV, \quad (5)$$

with

$$dV = dx_\beta dy_\beta d\Delta z dx'_\beta dy'_\beta d\Delta p_1 d\Delta p_2. \quad (6)$$

Equation (5) can be computed using the Monte Carlo integration with N uniform distributed samples in the n-dimensional volume V , e.g.,

$$\int_V f(\vec{x}) d\vec{x} \approx \frac{V}{2N} \sum_{i=1}^M f(\vec{x}_i), \quad (7)$$

where "2" represents two particles involved in a sampled scattering event, and the integration is calculated for each scattered particle respectively. For the problem of interest (beam loss calculation), M ($M < 2N$) is the total number of particles with $\delta > \delta_m$, where δ_m is an input value and should be chosen slightly smaller than the local momentum aperture for an economical simulation.

Figure 7 shows the Monte Carlo integration convergence vs. the total number of valid simulated scattered particles M . Based on this, we use $5 \cdot 10^6$ as the default value of M in *elegant*. Figure 8 shows the comparison of the local scattering rates calculated from Piwinski's formula [11] and our Monte Carlo simulation for a Gaussian-distributed beam. We can see that the agreement is excellent.

For a non-Gaussian-distributed beam, *elegant* has the ability to read the real beam distribution function from a

¹For clarity, we use (*) to denote all quantities in the CM system, as opposed to quantities in the laboratory coordinate system.

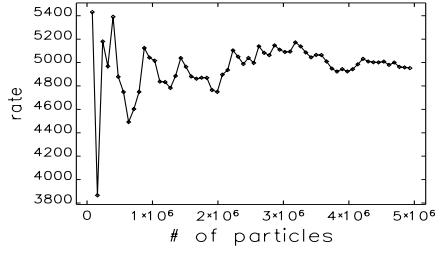


Figure 7: Scattering rate (in an arbitrary scale) vs. number of valid simulated scattered particles.

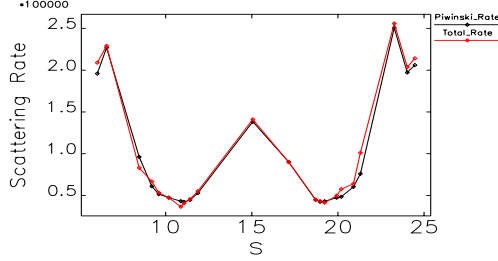


Figure 8: Local Touschek scattering rate (1/s): Piwinski formula (black) and Monte Carlo simulation (red).

histogram table. The table is interpolated to get the values of $\rho(\vec{x}_1)$ and $\rho(\vec{x}_2)$ in Equation (5). Figure 9 shows the comparison of simulated scattering rate for the assumed Gaussian-distributed beam and the realistic beam distribution. In the dispersion-free regions ($\eta = 0$), we obtained similar results for both cases, which is expected since the transverse beam distribution is very close to the Gaussian distribution. At a location with $\eta \neq 0$, the simulated rate depends on energy spread and the local value of the dispersion. Since the energy distribution is not Gaussian, the results from the Gaussian bunch are unreliable.

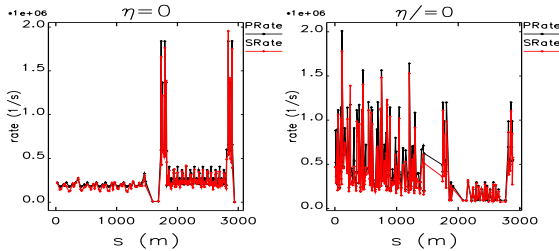


Figure 9: Simulated local Touschek scattering rate (1/s) for Gaussian beam (PRate) and realistic beam distribution (SRate).

Simulation of loss rate and position

Beam-scattering is a random process and can happen at any place along the beamline. The Monte Carlo simulation at one location is already very time consuming. To simulate the Touschek effect at every location of the beamline and obtain a stable statistical result is almost a non-realistic task. From Figure 8, we see clearly that the Monte Multi-Particle Beam Dynamics

Carlo simulation results are very close to Piwinski's formula for a Gaussian-distributed bunch. This fact inspired us to consider using Piwinski's formula to calculate the integrated scattering rate over a section of beamline and using the Monte Carlo simulation to generate random scattered particles. This allows obtaining accurate results with far few particles. Each scattered particle represents a scattering rate of

$$R_i = \frac{r_i}{\sum r_i} \int R_{Piwinski}, \quad (8)$$

where r_i is the associated local scattering rate $\frac{V}{2N} f(\vec{x}_i)$ in Equation (7), $\sum r_i$ is the value of Equation (7), and $\int R_{Piwinski}$ is the integrated Piwinski rate over the section of beamline. For a non-Gaussian-distributed bunch the local scattering rate can not be given by Piwinski's formula. In this case, Equation (8) is modified by multiplying by a factor $\frac{R_{MonteCarlo}}{R_{Piwinski}}$, where $R_{MonteCarlo}$ and $R_{Piwinski}$ are both local rates calculated at the same place.

The scattered particles are then tracked from the scattering location to the end of the beamline. The lost particles R_i and locations are collected, and the total beam loss rate and loss position are given by adding results from all the small beamline sections together.

As shown in Figure 7, to obtain a stable statistical result, the total number of valid events M (not the total number of samples $2N$) needs to be large enough. (Recall that M is the number of particles for which $\delta > \delta_m$.) In the case of calculating beam loss rate, it implies that the input value of δ_m should be close to the real momentum aperture for an economical calculation. A value of $0.8\delta_0$, where δ_0 is the estimated momentum aperture, is used in our example simulation.

We examined the scattering rate that each simulated particle represents and, not surprisingly, found a large variation. Some simulated particles represent very likely scattering events, while some represent very low probability events. We sorted all simulated particles by the associated scattering rate. Figure 10 illustrates the sum of the scattering rate ($\sum r_i$) vs. the number of simulated particles (\sum_i).

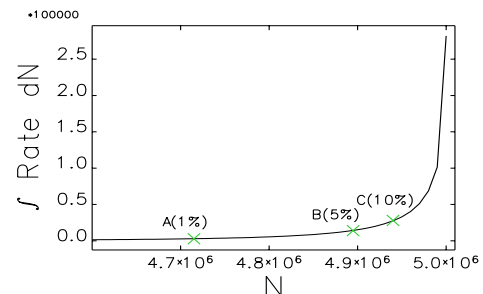


Figure 10: Integrated scattering rate vs. number of simulated particles. Particles are sorted with increasing associated scattering rate.

From this plot we can see that about 5% of simulated particles represent about 99% of the scattering rate. If we

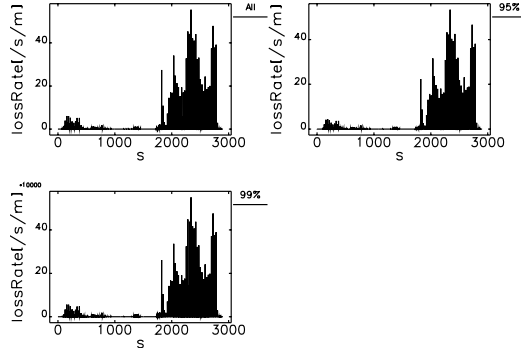


Figure 11: Simulated loss rate vs. position for various values of the scattering rate cut-off.

make use of an estimate of the local momentum aperture, which we do, then a large portion of the simulated particles will be lost somewhere along the beamline. Of those, we need only track that 5% of the particles, which represents 99% of the scattering events. The resulting error will be negligible. Figure 11 compares the computed loss rate for tracking scattered particles with 95%, 99%, and 100% of the total scattering rate, respectively. It's clear that the differences are small. In practice, the user can determine what percentage of scattering they would like to simulate, and elegant will choose the corresponding high-probability scattering events automatically for beam loss study. This strategy makes the calculation even more economical.

An application to the APS ERL upgrade is shown in Figure 12. Without optimized sextupoles installed in the turn-around-arc (TAA) section, the beam loss rate is too high. After optimization of the sextupoles configuration, the beam loss rate in the APS ring portion (from about 2600m to 3600m) is reduced significantly, to a level that is safe for operation.

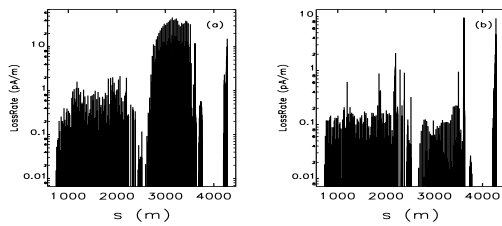


Figure 12: Simulated loss rate vs. position for APS ERL: (a) without sextupole optimization; (b) with optimized sextupoles

IBS EFFECT

The IBS effect is another widely studied beam-scattering effect in storage rings. The emittance growth rate τ_d in the direction d (x , y , or z) is given by the Bjorken-Mtingwa [2] formula for a Gaussian-distributed beam:

$$\frac{1}{\tau_d} = \frac{\pi^2 c r_0^2 m^3 N \ln \Lambda}{\gamma \Gamma} f, \quad (9)$$

Multi-Particle Beam Dynamics

where c is the speed of light, r_0 is the classical particle radius, m is the particle mass, N is the number of particles per bunch (or in the beam for the unbunched case), $\ln \Lambda$ is a Coulomb logarithm, γ is the Lorentz factor, Γ is the 6-dimensional invariant phase-space volume of the beam

$$\Gamma = (2\pi)^3 (\beta\gamma)^3 m^3 \varepsilon_x \varepsilon_y \sigma_p \sigma_z, \quad (10)$$

and f is a complicated function of beam size.

As for the Touschek effect, for a non-Gaussian beam Equation (9) is no longer valid, and we have to search for a new method. Due to the different natures of IBS and Touschek scattering, we care more about beam size evaluation than the real particle distribution, so we choose to continue to use the Bjorken-Mtingwa formula with some modifications. Figure 2 shows that the major difference between a linac beam and a Gaussian beam is in the longitudinal direction. The “intrinsic” energy spread σ_p is much smaller than the bunch's energy spread, and

$$\frac{1}{\tau_d} \propto \frac{1}{\gamma \varepsilon_x N \varepsilon_y N \sigma_p \sigma_z} f. \quad (11)$$

This difference could result in many orders of magnitude error in the computation of τ_d and must be taken into account. In elegant we provide a slice method to overcome the problem.

First, the beamline is divided into several sections by inserting a special element ISCATTER in the beamline, similar to what we did in the Touschek simulation. Unlike the Touschek simulation, it is not necessary to put as many ISCATTER elements as TSCATTER elements along the beamline, due to the fact that IBS effects need time to develop. An IBSCATTER element is only needed when beam size has a noticeable change due to the IBS effect.

For each section of beamline, the bunch is sliced longitudinally at the beginning of the section. The beam parameters and optical functions are calculated for each slice and propagated to the end of the section. To deal with a bunch traveling through a linac with energy variation, normalized beam parameters are used and are assumed to be unchanged for each section. Because there are no synchrotron oscillations for a linac bunch, the longitudinal growth rate is increased by a factor of 2 based on Piwinski's [12] formula

$$\frac{1}{\tau_z} [\text{linac-bunch}] = 2 \frac{1}{\tau_z} [\text{circulating-bunch}], \quad (12)$$

and the effective bunch length is $\sigma_z = \frac{1}{\sqrt{2\pi}} c \Delta t$, where Δt is the slice duration.

Each slice is assumed to be Gaussian distributed in transverse coordinates and energy spread, and to be uniformly distributed in the longitudinal direction. The Bjorken-Mtingwa formula is used to calculate the growth rate τ_d locally and is integrated over the entire section for each slice. At the end of the section (just prior to the location of the next IBSCATTER element), particles in each slice are scattered smoothly or randomly based on the calculated τ_d .

Particles are then put back together as a whole bunch, and all is ready for simulation of the next section of beamline.

We applied this method to the same APS ERL lattice used for our Touschek studies. Figure 13 shows the IBS growth rate with and without slicing beam. It's clear that the IBS growth rate of each slice is higher than if calculated for the whole bunch, especially in the longitudinal direction.

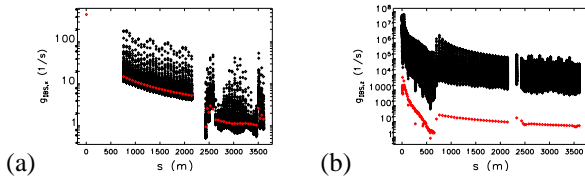


Figure 13: IBS growth rate for sliced bunch (black, each dot represents a slice) and unsliced bunch (red): (a) τ_x , (b) τ_z .

Figure 13 also shows that the longitudinal IBS growth rate is much higher than the transverse growth rate; this is expected due to the fact that the beam is much cooler in the longitudinal dimension. Figure 14 shows the beam dimensions at the end of the linac with (77 pC) and without (0 pC) IBS effect. There is no noticeable change in the transverse dimension. In the longitudinal, the energy spread at the center of the bunch increases due to the IBS effect. The change of energy spread is small and mainly happens at the center of the bunch. There should be no significant effect on brightness or FEL gain. It is possible that it may reduce CSR effects. Figure 15 shows the evaluation of energy spread of the entire bunch. There is no noticeable difference with and without the IBS effect.

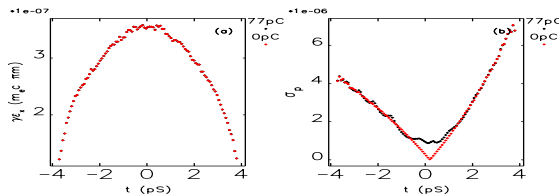


Figure 14: Particle distribution vs. longitudinal position (t) at the end of beamline with/without IBS: (a) normalized emittance; (b) energy spread.

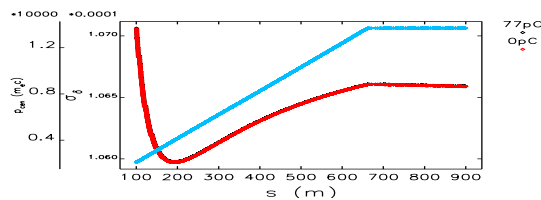


Figure 15: Bunch energy spread evolution vs. s .

CONCLUSION

We developed a method based on *elegant* to simulate beam-scattering effects for a linac beam with energy variation. The beam loss rate and location can be obtained by Multi-Particle Beam Dynamics

tracking scattered particles from Monte Carlo simulation, using realistic beam distributions. Beam-size evaluation is obtained by applying the Bjorken-Mtingwa formula to a sliced bunch. After applying the tools to an example APS ERL lattice design, we found that the Touschek scattering effect is significant. The momentum aperture of the lattice needs to be optimized carefully, and a beam collimation system can be designed based on the simulation results. The IBS growth rate is also very high for such a beam, but due to the fact that the time to travel through the linac is very short, the IBS effect has not enough time to develop. Hence there is no obvious effect on the machine's performance.

ACKNOWLEDGEMENTS

We would like to express our special thank to S. Khan for his original work [13, 14] on the Monte Carlo simulation, and F. Zimmermann for his work [15] to include vertical dispersion to the IBS calculation. We also want to thank L. Emery for many useful discussions.

REFERENCES

- [1] M. Borland, "elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation," Advanced Photon Source LS-287, April 2000.
- [2] J.D. Bjorken, S.K. Mtingwa, "Intrabeam Scattering," Part. Acc. 13 (1983) 115. 2000.
- [3] M. Borland and A. Xiao, Proc. of PAC'07, 3453 (2007); www.JACoW.org.
- [4] A. Xiao, Proc. of LINAC08, Victoria, Canada, to be published at www.JACoW.org.
- [5] M. Borland and A. Xiao, Proc. of LINAC08, Victoria, Canada, to be published at www.JACoW.org.
- [6] M. Borland and A. Xiao, Proc. of PAC'09, Vancouver, Canada, to be published at www.JACoW.org.
- [7] M. Borland *et al.*, Proc. of AccApp'07, 196 (2007).
- [8] X. Dong, private communication.
- [9] G. Hoffstaetter, M. P. Ehrlichman and A. Temnykh, Proc. of EPAC08, 1631 (2008); www.JACoW.org.
- [10] W. T. Ford, A. K. Mann and T. Y. Ling, "Beam Polarization Effects in High Energy Electron-Positron Storage Rings," SLAC-158, 1972.
- [11] A. Piwinski, "The Touschek effect in strong focusing storage rings," DESY 98-179, Nov. 1998.
- [12] A. Piwinski, "Intra-beam Scattering," CERN Accelerator School, p. 226 (1991).
- [13] S. Khan, "Simulation of the Touschek Effect for BESSY II: A Monte Carlo Approach," Proc. of EPAC94, 1192 (1994); www.JACoW.org.
- [14] S. Khan, private communication.
- [15] F. Zimmermann, "Intrabeam scattering with Non-Ultrarelativistic Corrections and Vertical Dispersion for MAD-X," CERN-AB-2006-002, Jan. 2006.

THE SIMULATION OF THE ELECTRON CLOUD INSTABILITY IN BEPCII AND CSNS/RCS*

Y. D. Liu[#], N. Wang

Institute of High Energy Physics, CAS, P.O. Box 918, 100049, Beijing, China

Abstract

Electron Cloud Instability (ECI) may take place in any positively charged particle circular accelerator especially in positron and proton storage rings. This instability has been confirmed to be a serious restriction to the beam stabilities. The physical model on the formation of electron cloud in various kinds of magnetic fields was introduced in the first section of the paper. The transverse and longitudinal wake field model to present the interaction between electron cloud and beam were introduced in another section of the paper. As an example, in positron storage in BEPCII and RCS of CSNS, the densities of electron cloud and beam instabilities caused by the accumulated electrons were simulated.

INTRODUCTION

The electron cloud accumulated in the vacuum chamber is usually associated with the transverse coupled bunch instability, bunch blow up and bunch lengthening. Experimental studies and numerical simulation have been developed for these phenomena [1]. Now BEPC has been upgraded to a two-ring collider, namely BEPCII, with electron and positron beams circulating in each separate ring. In its commissioning operation, ECI is much weaker because of many restraining methods used in positron ring. The effects of these restraining methods have been validated. In this paper, the simulation to electron cloud in different restraining conditions was introduced.

CSNS is a proton accelerator facility with consists of a linac and a rapid cycling synchrotron (RCS). Two bunches with a population of 1.88×10^{13} will be accumulated and accelerated in the RCS ring, and the electron-proton instabilities might happen in such high intensity proton ring. The ECI in CSNS/RCS is investigated in the last section. The main parameters of the BEPCII and CSNS/RCS ring are summarized in Table 1 and Table 2, respectively [2].

FORMATION OF ELECTRON CLOUD IN BEPCII AND CSNS/RCS

Electrons sourced from the (1) photoelectrons arising from the synchrotron radiation hitting the wall of the vacuum chamber, and (2) secondary emission from

antechamber by the photons hitting the wall with yield rate $Y \sim 0.1$ and reflectivity $R \sim 0.1$. If there is photon absorber, the Y and R become as small as $Y \sim 0.02$, $R \sim 0.1$.

Table 1: Parameters of the BEPCII

Parameters	Value
Beam energy $E(\text{GeV})$	1.89
Bunch population $N_b(10^{10})$	4.84
Bunch spacing $L_{sep}(\text{m})$	2.4
Bunch number n	93
Average bunch length $\sigma_z(\text{m})$	0.015
Average bunch sizes $\sigma_{x,y}(\text{mm})$	1.18, 0.15
Chamber half dimensions $h_{x,y}(\text{mm})$	60, 27
Synchrotron tune Q_s	0.033
Tune $Q_{x,y}$	6.53, 7.58
Circumference $C(\text{km})$	0.237
Average beta function $\langle \beta \rangle(\text{m})$	10

Table 2: Parameters of the CSNS/RCS

Parameters	Symbol, unit	Value
Inj./Ext. Energy	E_{in}/E_{ext} , GeV	0.08/1.6
Circumference	C , m	248
Bunch population	N_p , $\times 10^{12}$	9.4
Harmonic number	H	2
Repetition freq.	f_0 , H_z	25
Betatron tune	ν_x/ν_y	5.86/5.78
Beam pipe radii	a/b , cm	10
Proton loss rate	P_{loss} , turn^{-1}	1.33×10^{-4}
Proton e- yield	Y_p , e-/p/loss	100
Ionization e-	Y_i , e-/p/loss	1.31×10^{-5}

The percentage of photoelectron escaping out of the antechamber depends on the width of antechamber. In the simulation the beam field is presented by B-E formula and the numerical solver of Poisson-Superfish in the central region of $(10\sigma_x, 10\sigma_y)$ and out of this region, respectively. In the simulation we assume that secondary electrons yield (SEY) with and without TiN coating in the chamber is 1.06 and 1.8, respectively.

Simulation results show that the EC density can be reduced by about: 5x if the antechamber is adopted, 6x if the TiN is coated only, 3x if the photon absorber is made in the wall of the chamber only, and 5x if the electrode is installed in the beam chamber. In BEPCII, the antechamber, the photon absorber, and the TiN coating approaches have been adopted. With these three effects taken into account in the simulation, the electron density will be decreased about 80 times, i.e., from $1.1 \times 10^{13} \text{ m}^{-3}$ in the case without any restraining method to $1.3 \times 10^{11} \text{ m}^{-3}$,

*Work supported by National Natural Science Foundation of China (10605032)

[#]liuyd@mail.ihep.ac.cn

electrons hitting the walls, are attracted by the beam electric field and accumulate around the positron beam. Photoelectrons are produced in the chamber and

which is lower than the threshold causing the strong head-tail instability as described in later sections.

Figure 1 shows the electron cloud distribution in the vacuum chamber with different transverse shapes with or without electrodes.

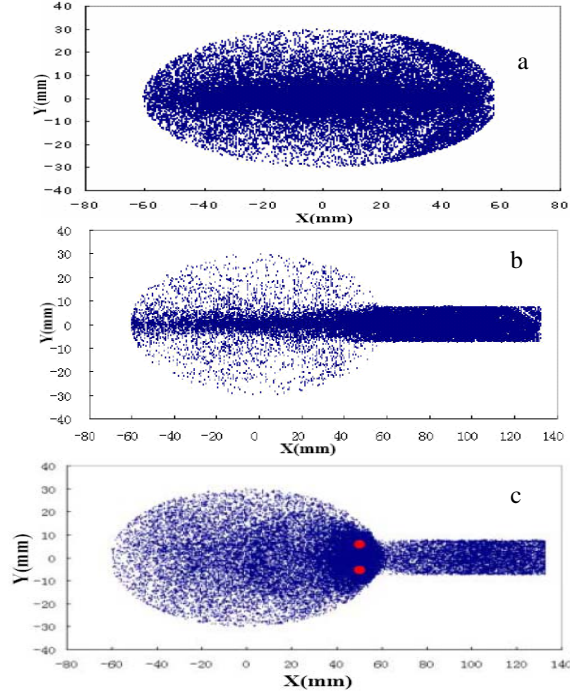


Figure 1: The EC distribution in vacuum chamber (a; elliptic pipe; b: antechamber pipe; c: antechamber with electrodes).

In the above simulation, we don't consider the effect of magnetic fields including dipole, quadrupole, sextupole and solenoid fields. In the region of dipole magnetic field without considering the fringe field, the magnetic field is only in vertical direction, $B=B_y$. For a quadrupole magnetic field, \mathbf{B} can be expressed by

$$\begin{aligned} B_x &= k_1 y, \\ B_y &= k_1 x, \\ k_1 &= \frac{1}{B\rho} \frac{\partial B_y}{\partial x}, \end{aligned} \quad (1)$$

where $B\rho$ is magnetic rigidity. For a sextupole magnetic field, \mathbf{B} can be expressed by

$$\begin{aligned} B_x &= k_2 xy, \\ B_y &= \frac{1}{2} k_2 (x^2 - y^2), \\ k_2 &= \frac{1}{B\rho} \frac{\partial^2 B_y}{\partial x^2}, \end{aligned} \quad (2)$$

In uniform solenoid field, the magnetic field is only in longitudinal direction, $B=B_z$.

According to the simulation in Figure 2, it is clear that in the magnetic fields, the electron cloud density is much lower than the density in drift region. The uniform solenoid field is the most effective way to confine the photoelectrons. All of the photoelectrons are confined to the vicinity of the vacuum chamber wall. So in the positron ring of BEPCII, the solenoid has been wound on the vacuum chamber of straight section and the magnetic field designed to be 30 gauss, which will be enough to clear the electron cloud in the central region.

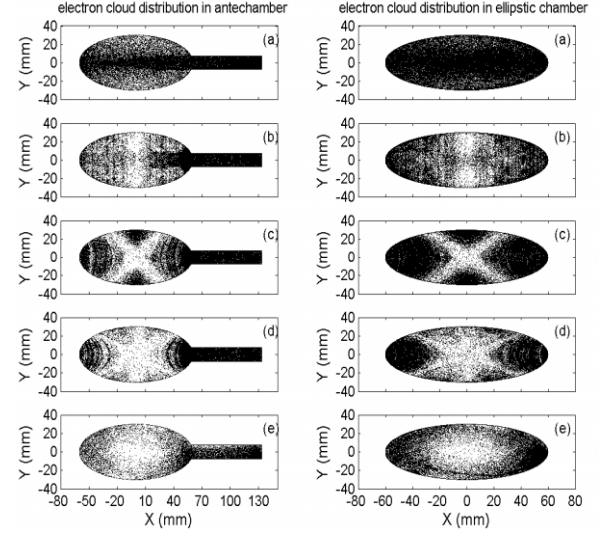


Figure 2. Distribution of electron cloud in various kinds of magnetic field (left: antechamber chamber; right: elliptic chamber; a: free field region; b: dipole field; c: quadrupole field; d: sextupole field; e: solenoid field $B_z=10$ G).

In CSNS/RCS, the photoelectron effects are much smaller. So there are three candidate mechanisms of electron production considered in this article, including: lost protons hitting the chamber wall, electrons produced by residual gas ionization, and secondary electron emission.

The electron yield due to residual gas ionization is determined by the ionization cross section and the vacuum pressure in the beam chamber. Residual gases of CO and H_2 are considered, whose ionization cross sections are $\sigma(CO)=1.3 \times 10^{-22} \text{ m}^2$ and $\sigma(H_2)=0.3 \times 10^{-22} \text{ m}^2$. The corresponding electron yield at vacuum pressure $p=10$ nTorr and room temperature ($T=294$ K) is 1.22×10^5 e/p/turn. The electrons are produced along the beam trajectory.

The mechanism of electron yield due to proton loss is not yet well known. In the simulation, we use the simple model proposed by Furman et al [3], that the number of electrons generated by lost protons hitting the vacuum chamber wall is $N_{px} Y_{xp_{loss}}$, per turn for the whole ring, where Y is the effective electron yield per lost proton, and p_{loss} is the proton loss rate per turn per beam particle. According to the beam loss tracking simulation of the

RCS ring, a total beam loss of 6% mostly occurs in the collimation region during the first 1 ms. by using the assumption of 100 e/p/loss, we obtain an electron production rate of 1.16×10^2 per turn, which is 3 orders higher than that of gas ionization. We assume the lost proton time distribution to be proportional to the longitudinal bunch intensity.

In the simulation, the electrons are simulated by macro particles. We use 1000 macro-particles to represent primary electrons generated when each bunch slice passes through the electron region. The secondary electron emission occurs when the particles hit on the beam chamber wall. The macro-electrons are tracked dynamically in the transverse plane. The space charge force is computed by the PIC method and applied to particles at each slice in the bunch and each step in the gap. The motion of macro-electrons and macro-protons are tracked during the EC region. After that, the bunch is transformed according to the six-dimensional linear equation. Figure 3 shows the building progress of the electron cloud in RCS with different proton loss. The results show maximum electron density at the bunch tail, and the electron density keep almost unchanged at the bunch head.

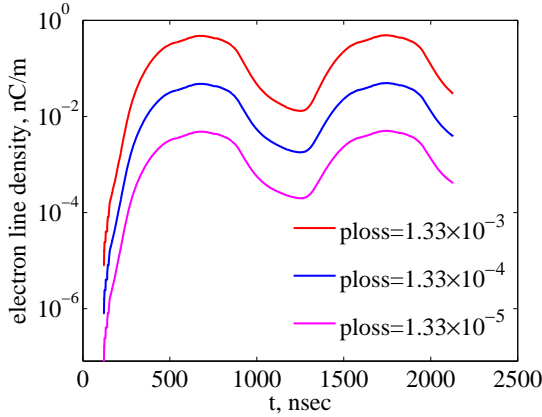


Figure 3: The density of EC for different proton losses.

THE TRANSVERSE AND LONGITUDINAL INSTABILITY CAUSED BY THE ELECTRON CLOUD

In the simulation to the coupled bunch instability, a bunch is expressed as a macro-particle and the EC can be represented by macro-particles. The force between the bunch and electron is represented by the Bassetti-Erskine formula and the solver of Poisson-Superfish in the central region of $(10\sigma_x, 10\sigma_y)$ and out of the region, respectively. By tracking the motion of the bunch and the formation of the EC in the same time, the oscillation amplitudes of 93 bunches and the EC density are recorded. The growth time can be obtained by fitting the amplitude of the oscillation. From the previous results, without any restraining methods the EC density is $1.03 \times 10^{13} \text{ m}^{-3}$, but when using the antechamber, photon absorber and TiN coating, the density will be decrease to $1.35 \times 10^{11} \text{ m}^{-3}$. In these two conditions we track the coupled bunches

oscillation in vertical direction, obtaining the growth time $\tau_{ly} \sim 0.08 \text{ ms}$ and $\tau_{2y} \sim 4.3 \text{ ms}$. The Growth behavior of the coupled bunch oscillation and the sideband spectra are shown in Figure 4.

Based on the head-tail model, a code was developed to simulate the beam size blow up. In the model, concentrating electron cloud at one location s of the ring, the EC and the bunch are represented by N_e and N_p macro-particles with transverse uniform and Gaussian distributions, respectively. We use vectors (x_e, x'_e) and (y_e, y'_e) to describe the transverse motion of electron respectively, without considering the longitudinal force imposed by the EC. The particle's synchrotron oscillation in a bunch being included, the motion of bunch macro-particles are described by the 3D vector, $(x_p, x'_p, y_p, y'_p, z, \frac{\Delta p}{p})$. The bunch is divided into N_s slices, which interact with the EC one another and cause the distortion of the EC distribution. The macro-particles in different slices can change their positions as the synchrotron oscillation occurs.

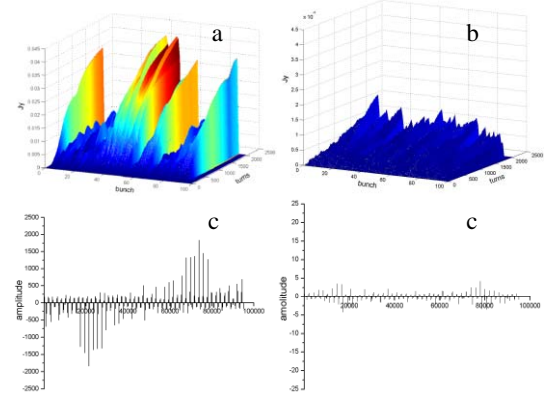


Figure 4: Growth behavior of coupled-bunch oscillation (Tracking result: a and b; sideband spectra: c and d).

After tracking the motions of bunch macro-particles for 4096 turns in the different EC densities, we find that the threshold by simulation is comparable to the analysis result. The tracking results are shown in Figure 5. It is clear that the threshold of the blow up of electron cloud is about $1.0 \times 10^{12} \text{ m}^{-3}$.

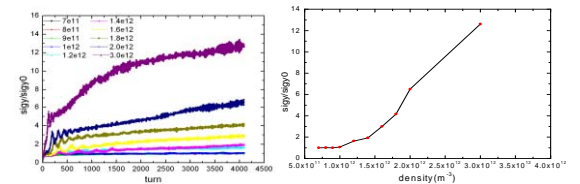


Figure 5: Beam vertical size in the different EC densities.

The same method was used to calculate the transverse bunch oscillation in RCS. There we include the RF acceleration progress.

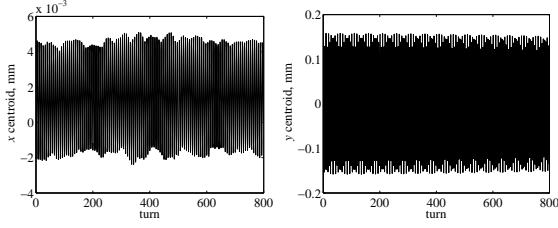


Figure 6: Bunch transverse RMS in RCS.

On the other hand, the positron bunches have to lose some amount of their kinetic energy to build the electron cloud during the interaction with the electrons. The energy variation inside the bunch can be seen as a longitudinal wake. The bunch particles have an additional energy spread due to the longitudinal wake from the electron cloud. the longitudinal electric field of the electron cloud is expressed as,[4]

$$E_z = Z_0 \int_r^a j_r dr, \quad (3)$$

where Z_0 the impedance in free space and j_r is transverse current density of electron cloud.

As an example of BEPCII, assuming the bunch current is 9.8mA, the electric field caused by different electron cloud density is shown in Figure 7.

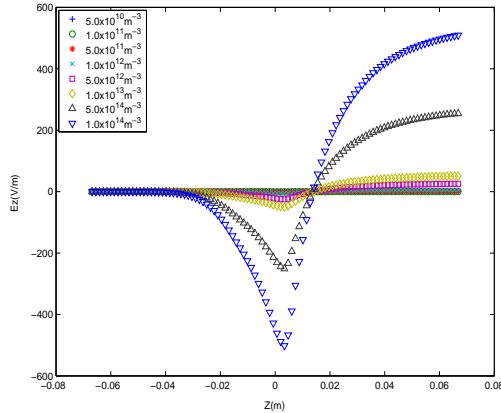


Figure 7: Longitudinal electric fields for different electron cloud densities.

The bunch length and energy spread varied due to the longitudinal wake caused by the electron cloud. The longitudinal interaction between bunch and electron cloud depends on the bunch current and density of electron cloud. A normal method to simulate the process of bunch lengthening is to track the motions of many macro-particles presenting the bunch. The motion of macro-particles is described in the longitudinal phase by [5]

$$z_i(n) = z_i(n-1) + \frac{\alpha c T_0}{E} \varepsilon_i(n) \quad (4)$$

$$\begin{aligned} \varepsilon_i(n) = & \varepsilon_i(n-1) - \frac{2T_0}{\tau_\varepsilon} \varepsilon_i(n-1) + \\ & 2\sigma_{\varepsilon 0} \sqrt{\frac{T_0}{\tau_\varepsilon}} R_i(n) - U_0 + \\ & V \cos(\phi_s + \frac{2\pi h}{C} z_i(n)) + V_i[z_i(n)]. \end{aligned} \quad (5)$$

where $\varepsilon_i(n)$ and $z_i(n)$ are the energy and position coordinates of the i^{th} particle after n revolutions in the storage ring. T_0 is the revolution period; τ_ε the damping time; U_0 the energy lost per turn; ϕ_s the synchronous phase; h the harmonic number; C the ring circumference; E the bunch energy; α the momentum compaction factor; $\sigma_{\varepsilon 0}$ the natural energy spread; R_i a random number obtained from a normal distribution with mean 0 and RMS 1. The wake potential V_i caused by the electron cloud depends on the longitudinal electric field of E_z .

In the simulations 10^6 macro-particles are tracked over 6 longitudinal damping times and the bunch length are calculated by averaging particle positions in the last damping time. For BEPCII, positron bunch current 9.8mA, bunch natural length 13.53mm, the longitudinal electric field and the tracking result for bunch length in different electron cloud density are shown in Figure 8.

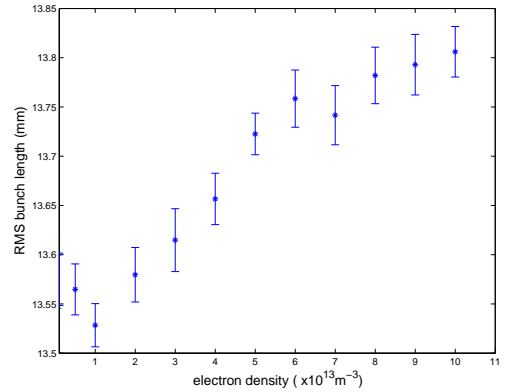


Figure 8: Bunch length for different electron densities.

Synchrotron tune shift is reduced about 5% of the undisturbed tune with the electron cloud density of $1.0 \times 10^{14} \text{ m}^{-3}$, as shown in Figure 9. The longitudinal action between electron cloud and bunch can be seen as an electron cloud potential well, which causes the possible bunch distortion.

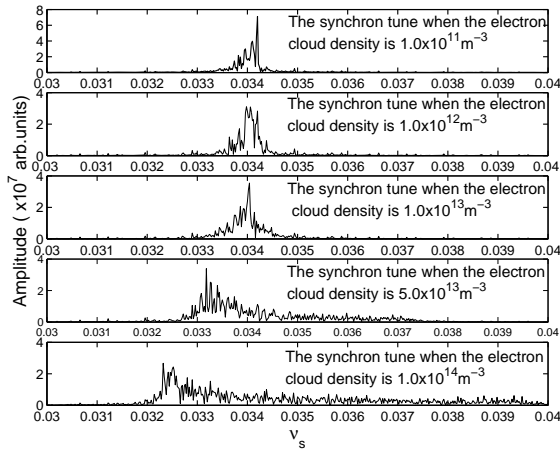


Figure 9: Synchrotron tune shift by the electron cloud.

According to the simulation results for bunch length in different electron cloud density, the electric field due to electron cloud can lead to the bunch lengthening. The bunch initial Gaussian distribution is shifted slightly in the forward direction for compensating the additional energy loss to the electron cloud. The longitudinal action between electron cloud and bunch can be seen as an electron cloud potential well which cause the possible bunch distortion.

CONCLUSION

With simulation, the efficiency for antechamber with photon absorber, TiN coating and clearing electrode to reduce the EC density is explored. All of these results are very meaningful for understanding the mechanism, as needed for the design and operation of storage rings for factory-like colliders. Particularly we have decided to adopt antechamber with photon absorber and TiN coating

in the BEPCII to cure the ECI. The EC density can be suppressed to below the threshold of strong head-tail like instability, while the coupled bunch instability can be damped with feedback system. The longitudinal effect of electron cloud serves as a potential well to interact the dynamics of bunch particles. Tracking methods to simulate the bunch length in different electron cloud density show that the bunch lengthening caused by the electron cloud density $1.0 \times 10^{14} \text{ m}^{-3}$, is just 2.0% of the natural bunch length. Its effect appears to be negligible for BEPCII.

ACKNOWLEDGEMENTS

The author is grateful for the suggestions from the members of accelerator physics group in IHEP and KEK.

REFERENCES

- [1] K. Ohmi, F. Zimmermann, "Head-tail Instability caused by Electron Cloud in Positron Storing Rings", *Phys. Rev. Lett.* 85, 3821(2000).
- [2] WEI Jie, FU Shi-Nian, FANG Shou-Xian et al. China Spallation Neutron Source Accelerators: Design, Research, and Development. In: C. Biscari et al. *Proceedings of EPAC'06*. Edinburgh, Scotland: EPS-AG, 2006. 366-368.
- [3] Furman M A, Pivi M. Simulation Results for the Electron Cloud at the PSR. In: Lucas P, Webber S. *Proceedings of PAC'01*. Chicago, U.S.A.: IEEE, Inc, 2001. 707-709.
- [4] A. Novokhatski, J. Seeman, "Simulation of electron cloud multipacting in solenoidal magnetic field", SLAC-PUB-10327.
- [5] R. Siemann, "Computer Simulation of Bunch Lengthening in SPEAR", *NIM* 203 (1982).

MODELING LASER STRIPPING WITH THE PYTHON ORBIT CODE

T. Gorlov, A. Shishlo, ORNL, Oak Ridge, Tennessee, 37831, U.S.A.

Abstract

Laser assisted hydrogen stripping has become a widely discussed alternative to the existing stripper foil approach. A simulation tool for this new approach is presented. The application is implemented in the form of an extension module to the Python ORBIT parallel code that is under development at the SNS. The physical model in the application utilizes quantum theory to calculate the evolution and ionization of hydrogen atoms and ions affected by the superposition of electromagnetic and laser fields. The algorithm, structure, benchmark cases, and results of simulations are discussed for several existing and future accelerators.

INTRODUCTION

One of the serious problems with operating the SNS facility in Oak Ridge involves the injection system of the accumulator ring. The current system uses a thin carbon foil to convert H⁻ beam from the linac to protons at the ring injection point. The planned upgrade of SNS involves a power increase of the injected beam that will lead to excessive heating and to rapid failure of the stripper foil. For this reason SNS is developing alternative injection processes for higher powers.

There are two such investigations at SNS. The first involves the development of better stripper foils [1] and the second is the replacement of the stripper foil by a laser-assisted stripping (LS) process [2]. Moreover, LS is an attractive method for other projects using conversion-injection of H⁻ beam. This paper presents a computational model for the three step LS developed at the SNS [3-5], and that can also be applied to other projects.

Basically, the theoretical description of LS requires the self-consistent application of quantum mechanics, laser physics, and accelerator physics. The central problem is the excitation-ionization of a hydrogen beam in a superposition of electromagnetic and laser fields $H^0 + \gamma \rightarrow H^{0*} \rightarrow p + e^-$. Success of the ionization process for each particle of the beam can be predicted with probability P using quantum mechanics. The problem of LS injection is conditioned by the requirements of the total LS efficiency \bar{P} and output emittance parameters of the proton beam. By solving the problem one can determine the requirements on the input hydrogen beam and the laser beam parameters for successful injection.

The LS is a new scientific field [3-5] with no established computational component. The proof-of-principle (POP) of LS has been successfully demonstrated at the SNS [3]. To computationally support the experiment a simple quantum model [4] of adiabatic rapid passage (ARP) was applied. The model considers a two level hydrogen atom and linear frequency growth in time of the laser field in the atom's rest frame. The model consists of a system of two linear differential equations

that can be solved by any Math package. Many physical phenomena taking place in a real experiment and significantly affecting the final LS efficiency are not included in the model. Nevertheless the model yields a good estimation of LS efficiency and can be used both as an initial stage for LS calculation and for benchmarking more detailed models. It should be noted that the purpose of the POP experiments was to demonstrate feasibility of the LS idea founded on the basic principles of quantum mechanics. The expectation of the model was successfully met experimentally.

For the next experiments planned in the SNS project it is necessary to demonstrate the feasibility of LS injection for the detailed SNS requirements. Experimental LS involves many different phenomena that should be included in the calculation. These include: the Stark effect and splitting of the hydrogen atom energy levels; spontaneous decay; electric field ionization; and possible circulation of the external electromagnetic field. Computing the LS for the next experiments is necessary for determining the simplest technical equipment and for optimizing the LS efficiency over the numerous parameters in the LS scheme.

A short description of the LS physical model, taking into account all the listed phenomena, can be found in [3]. A computer model of LS presented in this paper has been realized in form of an extension module in the PyORBIT parallel code developed at the SNS [6]. The choice of the implementation is conditioned by the general direction of development of accelerator codes at the SNS. The main advantages of the chosen direction are: rapid and pure object oriented prototyping of applications at the Python level; the widespread use and detailed documentation of Python; high performance execution of the classes at the C++ level; simple writing of extension modules; and parallel computing with PyORBIT based on MPI library. Moreover the present PyORBIT already contains templates for developing extension modules. If the developer has a unique physical problem requiring different classes than those in PyORBIT, he can create new extension modules to solve the problem.

This paper is organized as follows. Section 2 gives a short overview of the physical model of the LS and formulates the mathematical problem for computing the LS. The purpose of the section is to show the amount and kinds of computations required for solving the problem of LS. Section 3 describes the organization of the LS code and classes for getting the most efficient computations. Section 4 presents benchmarks and tests of the LS code. Section 5 outlines the scope of problems that can be solved by the code. Section 6 summarizes the paper summarizing and suggests some problems to be treated in the future.

OVERVIEW OF PHYSICAL MODEL

A brief description of the complete physical model for the three step LS process can be found in [5]. The first step of LS deals with the Lorentz stripping of the H^- ion: $H^- \rightarrow H^0 + e^-$. This is a probabilistic process that can be calculated with a simple semi-empirical formula [7, 8]. This overview of physical model is devoted mainly to the second and third steps of the LS, namely the excitation and ionization of the H^0 beam. In the absence of interaction between particles, each particle of the beam can be treated independently and then the total effects of the LS can be calculated statistically.

The model assumes that after the first step of the LS process there is a hydrogen atom in the ground state with initial position and momentum $\{\mathbf{r}_0, \mathbf{p}_0\}_i$ in the laboratory frame (LF) (Fig. 1a). The atom moves under the influence of a strong high frequency laser field and a slowly oscillating external electromagnetic field.

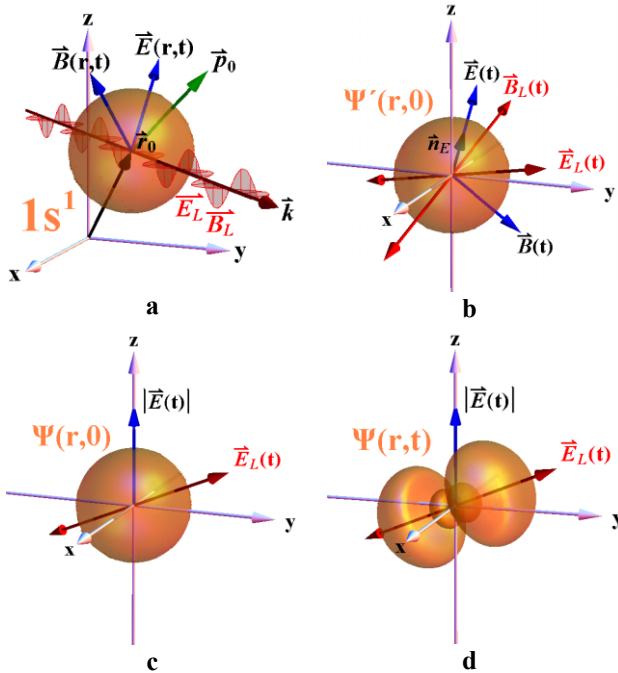


Figure 1. Schematics of a hydrogen atom in different frames considered for solving the excitation-ionization problem: **a**-laboratory frame, **b**-particle rest frame (inertial), **c**-particle rest frame (non-inertial) with z axes directed along the E field and $t=0$, **d**-the same as **c** but with $t>0$. The laser and external field vectors are shown as red and blue arrows, respectively. The orange figure represents the electron cloud of the atom.

The state of the atom travelling in the superposed electromagnetic and laser fields evolves. The problem is to find the probability of ionization of the hydrogen atom as a function of time. The state functions calculated for each H^0 particle will determine the further evolution of the beam. As a result it will be possible to calculate emittance parameters and currents of the H^0 , p , and e^- beams. The evolution of the hydrogen atom implies evolution of the electron wave function in the atom

$\Psi(\mathbf{r}, t)$. The probability of ionization of the H^0 atom can be calculated via the wave function by the following expression: $p(t) = 1 - \int \Psi^*(\mathbf{r}, t) \Psi(\mathbf{r}, t) d^3r$.

The quantum problem for the hydrogen atom is solved in the particle rest frame (PRF). A first step is to Lorentz transform the fields and time from the LF (Fig. 1a) to the PRF (Fig. 1b). Theoretically, it is possible to solve the quantum problem in the new frame using either of two approaches: application of the Schrödinger equation (SE) for the wave function or application of the master equation (ME) for the density matrix (DM). The first method does not take into account the phenomenon of spontaneous decay but the second one does. In other respects, both methods equally well solve the problem of the hydrogen atom in the superposition of laser and external electromagnetic fields taking into account the Stark effect and field ionization.

The SE in the inertial frame (Fig. 1 b) can be written in the following form:

$$i\hbar \frac{\partial \Psi}{\partial t} = (\hat{H}_0 + \hat{H}_{E,B} + \hat{V}(t))\Psi \quad (1)$$

Here \hat{H}_0 - operator of the unperturbed hydrogen atom, $\hat{H}_{E,B}$ - interaction between electron and the electromagnetic field, and $\hat{V}(t)$ - interaction between electron and the laser field. In principle, equation (1) can be solved directly. However, this is a 4D partial differential equation of the second order that is difficult to solve without simplifying assumptions. All magnetic fields in the PRF can be omitted in comparison with the electric fields in determining the evolution of the atom. Next, for solving the problem it will be convenient to transform equation (1) and the electric fields into a frame where the external electric field (non laser field) is directed along the z axis (Fig. 1c). The new frame is non-inertial and equation (1) is transformed into the following:

$$i\hbar \left(\frac{\partial \Psi}{\partial t} + \vec{\nabla} \Psi \cdot \left\{ \frac{\partial \mathbf{M}}{\partial \vec{n}_E} \cdot [\vec{\omega}_E \times \vec{n}_E] \right\} \mathbf{M}^{-1} \vec{r} \right) = (\hat{H}_0 + \hat{H}_{E,B} + \hat{V}(t))\Psi \quad (2)$$

Here $\mathbf{M} = \mathbf{M}(\vec{n}_E)$ is a transformation matrix from the inertial frame \mathbf{r} (Fig.1 b) to the non inertial frame \mathbf{r}' (Fig.1 c) $\mathbf{r}' = \mathbf{M}\mathbf{r}$. If the circulation of the field in the inertial PRF is zero $\vec{\omega}_E = 0$ then the new frame (Fig. 1c) is inertial and equation (2) will look like the usual SE (1). Equation (2) can be solved by the well-known method [5], [9] in which the solution $\Psi(\mathbf{r}, t)$ is represented in the form of an eigenfunction expansion:

$$\Psi(\vec{r}, t) = \sum_{n=1}^N c_n(t) \psi_n(\vec{r}) \quad (3)$$

where the $\psi_n(\mathbf{r})$ are stationary parabolic wave functions of the hydrogen atom in an electric field satisfying the equation: $(\hat{H}_0 + \hat{H}_E) \psi_n = (E_{0n} - i\Gamma_n/2) \psi_n$. Here $\hat{H}_E = -\mu_z E_z(t)$

is the interaction operator between the electron and the external quasi-static electric field which leads to the Stark splitting of the hydrogen atom into a multi level system. In this model we assume that $E_z(t)$ is a quasi-static field not leading to atomic excitation or transitions between levels. It follows that it is necessary to solve the problem of the Stark effect prior to equation (2) and to find $E_{0n}(E_z)$, $\Gamma_n(E_z)$ and $\psi_n(E_z)$ in parabolic coordinates as functions of the field E_z .

The dependences $E_{0n}(E_z)$ and $\Gamma_n(E_z)$ were calculated numerically in the form of tables by the method described in [10] for E_z in the range between 0 and the threshold fields at which the level disappears. The functions $\psi_n(E_z)$ were calculated using second-order perturbation theory [11].

Substituting (3) into (2), after some manipulation we obtain a system of linear differential equations for complex parameters $c_n(t)$ that can be solved numerically by the Runge-Kutta method. Finally, the system of equations has the following form:

$$\dot{c}_m(t) = \sum_{n=1}^N c_n(t) f_{nm}(t) \quad m=1...N \quad (4)$$

The initial condition of the equation is the ground state of the atom at the initial instant: $c_n(t_0) = \delta_{1n}$. The functions $f_{nm}(t)$ contain parameters of the hydrogen atom (Stark parameters and wave functions) and the laser field. The number of equations, N , is defined by the number of hydrogen levels involved in the problem. Each level with the principal quantum number n has n^2 Stark sublevels. If we consider the excitation of n^{th} level then we should include $N=1^2+2^2+\dots+n^2 = n(n+1)(1+2n)/6$ levels into equations (3) and (4).

The DM formalism is another approach for solving the problem that takes spontaneous decay into account. DM elements for the wave function (3) $\rho_{nm}(t) = c_n(t)c_m^*(t)$, $n, m=1...N$ define the state of the quantum system and the wave function (3). The ME for evolution of the density matrix can be obtained similar to the SE. Finally, the system of differential equations for the DM elements has the following form:

$$\dot{\rho}_{nm}(t) = \sum_{k=1}^N [\rho_{km}(t)a_{knm}(t) + \rho_{nk}(t)b_{knm}(t)] \quad (5)$$

$$n, m=1...N$$

with the initial conditions defined by the ground state of the atom at the initial instant: $\rho_{nm}(t_0) = \delta_{1n}\delta_{1m}$. The DM elements are complex and satisfy the relation $\rho_{nm}(t_0) = \rho_{mn}^*(t_0)$. Functions $a_{knm}(t)$ and $b_{knm}(t)$, similar to (4), contain the parameters of the hydrogen atom and laser field. The ME is bigger then the SE by a factor of N and consists of $N^2 = (n(n+1)(1+2n)/6)^2$ differential equations, as can be seen by comparing Eqs. (4) and (5). If we consider the excitation of the $n=3$ level, then it will be necessary to solve a system of 196 differential equations for complex elements of DM.

Beam Injection/Extraction, Transport and Targetry

After solving the equations (4) or (5) it is easy to calculate the function of ionization probability. Substituting (3) into the definition of the ionization probability we obtain:

$$p(t) = 1 - \sum_{n=1}^N c_n(t)c_n^*(t) = 1 - \sum_{n=1}^N \rho_{nn}(t) \quad (6)$$

The quantum mechanical model of LS described above covers many phenomena, excluding spin-orbital interaction and fine structure splitting. However the question of the applicability of the model and code for the calculation of LS in extremely strong external electromagnetic fields requires special consideration. In this case, perturbation theory can yield incorrect wave functions. For correct solution of the SE (1) or (2) it is necessary to take into account continuum spectra in the quantum problem without using perturbation theory.

COMPUTER CODE

The code consists of many classes responsible for different functions. This section reviews the basic classes.

Laser field

This class provides the components of the electric and magnetic laser field as a function of particle coordinates and time in the laboratory frame (Fig. 1a). Most laser beams can be well approximated by the elliptical Gaussian mode [12]. In the LS problem the components of the laser field are represented in a complex form for simplification of mathematical equations and the code:

$$E(x, y, z, t) = \frac{2\sqrt{cPw_xw_y}}{\pi\sqrt{h_xh_y}} e^{ik(ct-z) - \frac{x^2}{h_x} - \frac{y^2}{h_y}} \quad (7)$$

$$h_{x,y} = w_{x,y}^2 - \frac{2i(z - f_{x,y})}{k}$$

where $k=2\pi/\lambda$, P is the power of the beam and c is the speed of light. The magnetic field B can be written as $B=E/c$. The elliptical Gaussian laser beam (7) is specified by four geometrical parameters: w_x, w_y – horizontal and vertical waists, and f_x, f_y – positions of the waists.

In general, the result of the interaction of a particle beam with the laser beam depends on the local characteristics of the laser beam at the point of interaction. For this reason, it is convenient to operate with r_x, r_y – sizes of the beam and α_x, α_y – angles of divergence of the beam at this point. The local parameters can be expressed mathematically using the first four parameters.

The expression (7) defines a continuum laser beam. For the next LS experiments at the SNS a pulsed laser source ($\tau \sim 50$ ps, $\lambda=355$ nm) will be used. Because this is not an ultrashort laser pulse, the temporal shape of the laser power P in (7) can be described by a Gaussian function: $\exp[-4 \cdot \ln(2) \cdot (t-z/c)^2/\tau^2]$. Here τ is the full width at half-

maximum FWHM. The electric field (7) should be multiplied by the square root of this function.

Stark effect

This class provides Stark parameters for equations (3) and (4) as functions of the electric field. These functions have been calculated previously by another code and stored as data files. The energy E_{0n} and lifetime⁻¹ Γ_n are stored in the form of numerical functions and the dipole transitions $\mathbf{d}_{nm} = \int \psi_n^*(\mathbf{r}) \mathbf{e} \mathbf{r} \psi_m(\mathbf{r}) d^3r$ are stored as coefficients of the power series expansion $\mathbf{d}_{nm}(E_z) = \mathbf{d}_0 + \mathbf{d}_1 E_z + \mathbf{d}_2 E_z^2$.

A resonant phenomenon occurring at the interaction of the laser field and the atom significantly affects the efficiency of LS. This phenomenon, mathematically contained in equations (3) and (4), strongly depends on the precision of the energies $E_{0n}(E_z)$. For this reason, the more exact method [10] was used for computation of the energies. For computation of the dipole transitions playing a minor role, the perturbation theory has been applied.

Two level atom

This is a class for computation of laser stripping based on the SE or ME when only two levels $N=2$ are included in the series expansion (3). The two level atom model for the interaction of light with matter leads to the system (4) consisting of two equations. Practically, the approach is valid only when: no external static field is applied to the atom; the electric field is directed along the z axis in the PRF (Fig 1.c); and the eigenfunctions (3) are expressed in spherical coordinates. In other words, the class can be applied for computation of LS without external static electric fields in the PRF and for linearly polarized laser fields. From the point of view of computing, this is a high performance class that can be used for initial estimations of LS.

Schrödinger equation

This class performs the calculation using the SE approach described in the previous section. It takes into account everything except spontaneous decay from upper to lower levels. Nevertheless the class can be used for the calculation of LS in the presence of external quasi-static fields for atoms having small time of interaction with laser field. The validity of the computation requires a negligible spontaneous decay rate, or small evolution times in contrast with lifetime of an exited atom. Practically, this is the best class when considering the performance-potentialities relationship, and it is useful for optimization of the laser and external field parameters.

Density matrix

This class is based on the DM approach described in the previous section. Although the method includes many phenomena, it is computationally too slow to be used for optimization requiring multiple computations. Practically, the method can be used for the final calculation of LS efficiency after optimization by the SE method.

Diagonal density matrix

In the absence of laser fields the master equation (5) is modified such that the equations for the diagonal elements $\rho_{nn}(t)$ are separated from those for nondiagonal elements. As a result we obtain system of N linear differential equations for the diagonal elements of DM:

$$\dot{\rho}_{nn}(t) = \sum_{k=1}^N \rho_{kk}(t) f_{kn}(t) \quad n=1 \dots N \quad (8)$$

This class, as well as the previous class, can be applied for computation of LS after calculation by the SE method. However in contrast to the previous class for calculating the 2nd and the 3rd steps of the LS this class calculates only the 3rd step and has much faster performance then the previous one. This class, in combination with (6), is very useful for the computation of emittance growth of stripped atoms, which can be very important for some projects. The derivative of (6) $\partial p(t)/\partial t$ gives the spatial distribution of the current density of stripped atoms. The density is nonzero only in the presence of transverse magnetic field in the LF. The space distribution leads to the spread of trajectories of the stripped particles in the magnetic field and to emittance growth.

Another important application of the class is for computation of beam losses during foil stripping. In this case the beam loss is due to the few percent of partially stripped beam that remains neutral H^0 after passing through the foil. The empirical distribution of the initial exited states $\rho_{nn}(t_0) = f(n)$ of the neutral beam can be found in [13]. After applying this class it is possible to calculate the evolution of the beam and its losses.

A similar thing can be done for LS. The initial distribution of exited states of the H^0 beam $\rho_{nn}(t_0) = f(n)$ can be written analytically without numerical simulations and the class can be applied to compute losses and emittance growth due to LS.

Ionization of H^- beam

This class calculates the stripping of H^- beam travelling in external electromagnetic fields. This is the first step of the LS. In this case only one state of the ion is possible. The system of equations (8) is transformed into the single equation:

$$\dot{\rho}_{11}(t) = -\Gamma(E(t)) \rho_{11}(t) \quad (9)$$

with the condition $\rho_{nn}(t_0) = 1$. The empirical dependence $\Gamma(E)$ for the H^- ion can be found in [7, 8].

BENCHMARKS

This section presents benchmarks of the LS code based on: fulfilment of quantum mechanical laws; comparison of the classes with each other; and checking the results of computation with a simple analytical case. Let us begin with a particular case that can be calculated analytically. This method was applied for checking the two-level

model [4]. Let the electric component of the laser field in the PRF have the following form:

$$\vec{E}(t) = \vec{E}_0 e^{i\left(\omega_0 t + \frac{\Gamma t^2}{2}\right)} \quad (10)$$

where ω_0 is a resonant frequency between the 1st and the 2nd level of hydrogen. Assume that the frequency of the laser field grows linearly with time, passing through resonance at $t=0$: $\omega(t) = \omega_0 + \Gamma t$. We direct the field \vec{E}_0 along z axes and after calculation we obtain the following evolution of the 1s and 2p states in the spherical basis (Fig 2). The population of the upper state (2p-state) calculated in our case as $|c_{2p}|^2$ can be predicted analytically [4] at $t \rightarrow \infty$. The numerical results are in good agreement with the analytical formula.

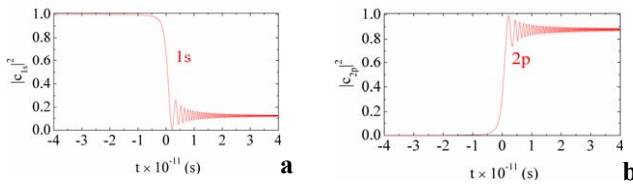


Figure 2. Evolution of populations of 1s-2p states of a hydrogen atom in a benchmark laser field.

The result should not depend on the coordinates we use to solve the problem, spherical or parabolic. Let us direct vector \vec{E}_0 in (6) along $\mathbf{n} = \{1, 1, 1\}$ direction for generality and solve the problem in parabolic coordinates by the SE. Then we obtain the following picture (Fig 3.) that shows that the evolution in the parabolic basis matches the evolution in the spherical basis (Fig. 2a).

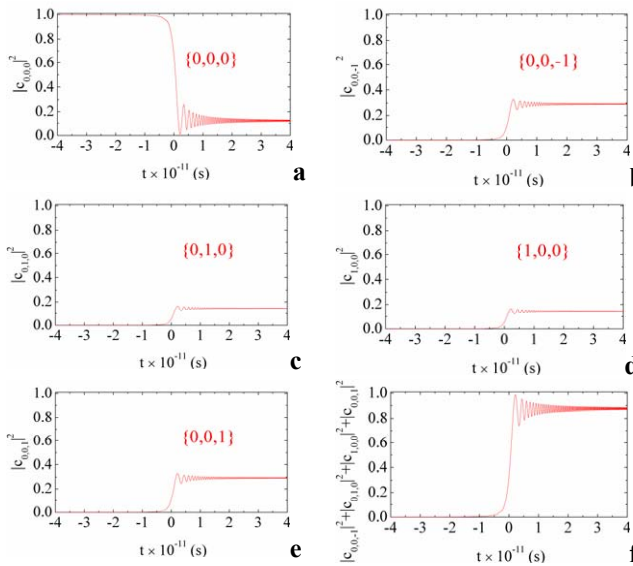


Figure 3. Evolution of populations of parabolic states with principal quantum numbers $n=1, 2$ for hydrogen atom.

The sum of the populations of the 2nd states in the parabolic basis shown in Fig. 3f matches the population

of the 2p state in the spherical basis (Fig. 2b). If we exclude spontaneous decay from ME and solve Eq. (5), then we obtain exactly the same results shown in Fig. 3 for the diagonal elements of DM. As expected, the results are matching because equation (5) follows directly from (4) when spontaneous decay in (5) is omitted. The sum of all populations is 1 for all 3 methods.

If we set $\Gamma=0$ in (10), then the atom will be affected by the resonant frequency ω_0 and we will observe Rabi oscillations. These can be also described analytically and shown to match for all the methods. In this way, we tested different classes and found them to be in agreement with the basic principles of quantum mechanics, with the analytical formulas, and with each other.

Another instructive example is a computation of the atom evolution in an external static electric field while applying the same laser field (10). The result of the calculation is shown in Fig. 4, both for the SE and the ME. One can see that the different parabolic states with different energies will be excited in different moments of time. This happens due to resonance physics of the excitation process and the linearly changing frequency of the laser field.

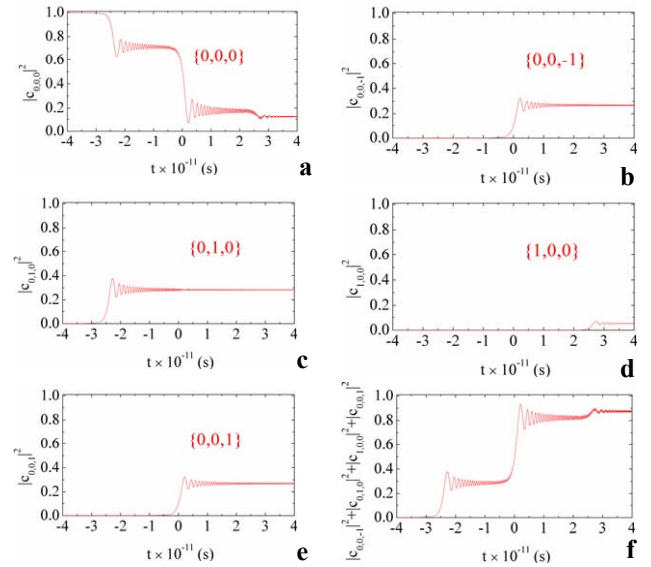


Figure 4. Evolution of populations of parabolic states with principal quantum numbers $n=1, 2$ in the presence of an electric field directed along the z axis.

This picture corresponds to the realistic LS excitation by the laser field (7) showing behaviour similarly (10) in the PRF and realizing the idea of the ARP.

APPLICATIONS OF THE CODE

The purpose of the code is the optimization of laser stripping systems. Below we list only a few of the questions that can be resolved using these models. One of the main technological problems of LS is getting a high efficiency while using a relatively low power laser. Let us consider the LS scheme with a single pass laser beam (7). The parameters $r_x, r_y, \alpha_x, \alpha_y$ of the beam can be optically

adjusted. The optimization of these parameters can be easily realized using the code together with an optimization package.

Another problem is the possible modification of the temporal shape of the laser pulse, which can become very important for LS. A laser micropulse with a given peak power and FWHM can be extremely non-optimized for stripping of a hydrogen beam with given FWHM. In this case it is reasonable to try to modify the laser micropulse shape while conserving the pulse energy. The LS code can provide parameters of the laser pulse optimized for best LS efficiency.

In most cases, σ – the polarization of the laser field in the PRF, provides better LS efficiency in contrast with π - polarization because of less Stark splitting of energies ΔE_n of excited states. In any particular case, questions about polarization can be resolved with the code.

The three step LS scheme for intermediate experiments at the SNS is shown in Fig. 5. Two dipole magnets of opposite polarities provide stripping of the first and the second electrons. The magnetic field in the region of interaction between the magnets is minimal and can be described by formula $\mathbf{B} = C(\mathbf{x}\mathbf{e}_z + \mathbf{z}\mathbf{e}_x)$ where C (T/m) is obtained from the distance between the magnets and their strength.

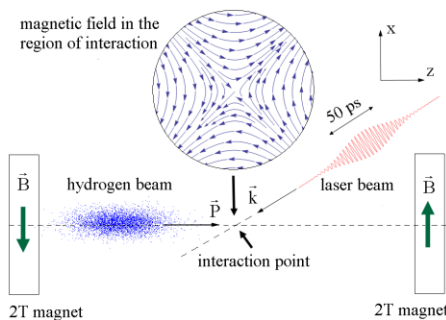


Figure 5. Schematics of laser stripping for the SNS.

Using the code it is possible to optimize the distance between the magnets. On the one hand, increasing the distance between the magnets decreases the field strength in the interaction region. This leads to a decrease of perturbation of the atom, less Stark effect, and improvement of the resonant excitation and LS efficiency. On the other hand, greater distance between the magnets leads to longer travel times of excited atoms from the interaction point to the second magnet. This circumstance increases the effect of spontaneous decay and decreases total LS efficiency. From the point of view of the code the distance between the magnets can be considered as a parameter that can be optimized.

Computations for the next LS experiments at the SNS conclude that it is possible to obtain 90% of LS excitation using a single pass laser micropulse with 2 MW peak power and 50 ps FWHM in time. The H^0 beam is considered to have the same temporal FWHM and given parameters of emittance. More details about the estimations can be found in [14].

SUMMARY

- A new code for the calculation of the three-step laser assisted stripping of H^- beams has been developed.
- The computational model was implemented as an extension module for the PyORBIT parallel code developed at the SNS.
- Benchmarks of the code demonstrate its reliability and confirm its correspondence with the mathematical model and with the basic principles of quantum mechanics.
- The code was used for optimizing the laser stripping experiments for the SNS project.
- In the future it will be necessary to perform theoretical investigations on the excitation of hydrogen atoms in a strong static field. It will then be necessary to include continuum spectra into the quantum mechanical problem.

ACKNOWLEDGMENTS

Research sponsored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy. The authors thank V. Danilov, J. Holmes, A. Menshov and the laser stripping team of SNS for their help.

REFERENCES

- [1] M. Plum, *et al.*, Nucl. Instr and Meth. A 590 (2008)
- [2] Mini-Workshop on Laser Assisted H^- Beam stripping, Oak Ridge, February 2009. <http://wiki.ornl.gov/events/lahbsa/default.aspx>
- [3] V. Danilov, *et al.*, PRSTAB, **10**, 5 (2007) 053501
- [4] V. Danilov, *et al.*, PRSTAB, **6**, (2003) 053501
- [5] T. Gorlov, V. Danilov, A. Shishlo, PAC'09, Vancouver, May 2009.
- [6] A. P. Shishlo *et al.*, "The Python shell for the ORBIT code", this conference.
- [7] L. R. Scherk, Can. J. Phys. **57**, 558 (1979).
- [8] P. B. Keating *et al.*, Phys. Rev. A **52**, 4547 (1995).
- [9] L. D. Landau and E. M. Lifshitz, Quantum Mechanics, 2nd ed. (Pergamon, New York, 1978)
- [10] R. J. Damburg and V. V. Kolosov, J. Phys. **B 9**, 3149 (1976); "Rydberg states of Atoms and molecules", edited by R. F. Stebbings, Cambridge University Press, (1983).
- [11] A. A. Kamenski and V. D. Ovsianikov, J. Phys. **B 33**, 5543 (2000).
- [12] T. Freearge, J. Coutts, J. Walz, D. Leibfried and T. W. Hänsch, J. Opt. Soc. Am. **B 14**, 2010 (1997).
- [13] V. Danilov, "Uncontrolled H^0 excited-state losses for SNS injection as function of injection energy", SNS TechNote 54, 2002. <http://it.sns.ornl.gov/cgi-bin/asd/note1LinePrint>
- [14] T. Gorlov *et al.*, "Possibility of laser stripping intermediate experiment with present equipment and beam parameters", SNS TechNote 184, 2009. <http://it.sns.ornl.gov/cgi-bin/asd/note1LinePrint>

USING GEANT4-BASED TOOLS TO SIMULATE A PROTON EXTRACTION AND TRANSFER LINE

F.W. Jones, R. Baartman, and Y.-N. Rao
TRIUMF*, 4004 Westbrook Mall, Vancouver V6T 2A3, Canada

Abstract

The simulation toolkit GEANT4 has been used to create high-level tools for specific user groups, such as SPENVIS in space physics and GATE in medical imaging. In Accelerator Physics, comparable efforts are being devoted to develop general-purpose programs for simulating beam lines and accelerators, allowing access to Geant4's facilities for 3D geometry, tracking, and interactions in matter without the need for specialised programming techniques. In this study we investigate the use of two high-level tools based on Geant4, BDSIM and G4BEAMLIN, to model a 65-meter beam line supplying protons from the TRIUMF cyclotron to the ISAC Rare Isotope Beam facility. We outline some features of the codes and comment on their different approaches to defining the beam line geometry. Due to its ability to model some important aspects such as rectangular dipoles and magnetic fringe fields, G4beamline was utilized for the simulations presented here, for validation of the model and the investigation of beam losses.

INTRODUCTION

In using simulation tools to investigate particle losses in accelerators and beam lines, the effects of particle interactions in matter, and in particular the secondary particles arising from electromagnetic and hadronic interactions, are very important for safety issues, loss monitoring and diagnostics, and radiation damage and activation of hardware.

The Geant4 simulation toolkit[1] offers a versatile way to track particles in an accelerator or beam line geometry, with realistic fields. For interactions in matter it offers a wide range of physics processes and models and a host of other facilities for studying losses with tracking of all relevant secondaries. The choice of physics models allows tuning of the simulation to the particular energy range and particles of interest.

Tapping into the power of Geant4 generally requires facility in C++, as the user must supply C++ code to define and implement the geometry, to specify the sampling of track information in sensitive detectors, and to instantiate the necessary "manager" objects to initialize and coordinate the simulation. Although C++ skills are part of the culture of high energy particle physics, they are not always as easy to find in other fields, and this has prompted the development of higher-level tools built from Geant4.

Accelerator physicists can benefit from two such tools,

BDSIM and G4Beamline. In the following we will describe some of the capabilities of these tools and our development of a prototype model of a TRIUMF beam line in each code. A limitation in BDSIM (being addressed by the code authors at the time of writing) prevented us from advancing to a full simulation, but in G4Beamline we proceeded to refine the model and to validate its tracking and optical properties against measured beam profiles. The validated model enabled us to perform simulations aimed at estimating the influence of multiple scattering in the cyclotron extraction foil on losses in the beam line, of which some first results will be presented.

GEANT4, BDSIM, AND G4BEAMLIN

GEANT4

Geant4 provides a software toolkit for tracking and simulation, in a 3D geometry, of particle interactions in matter. It is object-oriented and scalable to very large and diverse applications. A key characteristic of its design is to allow the user to plug in new or modified simulation components without the need for any modification of the Geant4 code itself.

The code is written in C++ and is implemented as a collection of class libraries in various categories. For a given application, the user provides code for a main program and auxiliary classes which instantiate the components of the simulation: geometry, particles, physics processes, data collection objects (sensitive detectors), and so on. The main program also invokes the "glue", or manager, classes from the toolkit which initialize and coordinate the simulation run. For each instance of an application, the main program and user-written classes are compiled and linked together with the Geant4 libraries to make an executable.

This approach follows the principle that for a simulation code *the most powerful and general input language is the language the code is written in*. For Geant4 any input system less complex than C++ code may limit the expression of complex problems. A somewhat gentler principle is that *a scripting language for object-oriented simulation should itself be object-oriented*.

On the other hand, simpler and easier input methods can be devised for problems with a specialized and well-defined scope, particularly if the scale of the problem is relatively small. For accelerators and beam lines, BDSIM and G4Beamline provide the needed functionality using input methods similar to basic scripting or shell languages.

* TRIUMF receives federal funding via a contribution agreement through the National Research Council of Canada

BDSIM

BDSIM[2] has been in development since c.2000 and was originally motivated by studies of backgrounds and other issues in the beam delivery systems of next-generation linear colliders (first CLIC and later ILC). The program is conceived in a sufficiently general way to support many other beam line applications. Notably, it incorporates an input parser which allows beam lines defined in MAD(X) language to be rapidly brought into BDSIM with the addition of only a few parameters such as the beam pipe radius and the outer dimensions of magnetic elements.

The other feature that puts BDSIM into a special class of beam simulation programs is that particle tracking in vacuum is done by *element transfer maps*, as in matrix-based codes such as DIMAD and TRANSPORT. The step-wise Geant4 tracking is invoked only when a particle enters a material. Hence, particles that remain in the vacuum chamber are tracked very quickly through the beam line. In studies of beam losses, this allows reduced computation time and improved statistics since the “uninteresting” particles that do not hit anything will be disposed of quickly. This feature also allows easy comparison and benchmarking against other map-based optics and tracking codes.

For the present study, BDSIM presented some limitations that revealed its origin in high-energy collider simulations. Dipole magnets are only of sector bend type (edges perpendicular to the reference path) and edge angles cannot be specified. Although a rectangular bend (RBEND) element is documented, it is in fact implemented as a sector bend. Moreover, the volumes representing dipoles and quadrupoles in the geometry always have circular cross sections. These conditions are not realistic for our application, in which the bends are of rectangular shape, and in which the edge focusing and fringe fields play a significant role in the optics of the beam line.

G4Beamline

G4Beamline[3] has been in development since 2002 and was conceived as a general and flexible interface to Geant4 for studying beam lines. It offers a wide range of beam line components, including the basic magnetic elements as well as rf cavities, solenoids, absorbers and other structures. User-defined elements are also available via a collection of basic Geant4 shapes, with flexible methods for inputting or defining electric and magnetic fields. Each element is described in detail by a set of parameters which can be specified by constants, variable substitution, and general mathematical expressions. For magnets, parameters for gap, field, and iron dimensions, as well as full control over placement and alignment, allow some realism in the geometry and layout of the beam line. A built-in optimization facility allows element parameters to be tuned to specific centering or focusing objectives, or in general to any goal expression in terms of test-particle coordinates.

G4Beamline contains extensive data-collection and processing facilities, including cut-planes, sensitive detector Computer Codes (Design, Simulation, Field Calculation)

volumes (phantom or inside real elements), powerful event and track cut mechanisms, and beam profile readouts. Data can be output as ascii files, or as Root n-tuples, which can be processed and plotted by the HISTOROOT program supplied with G4Beamline.

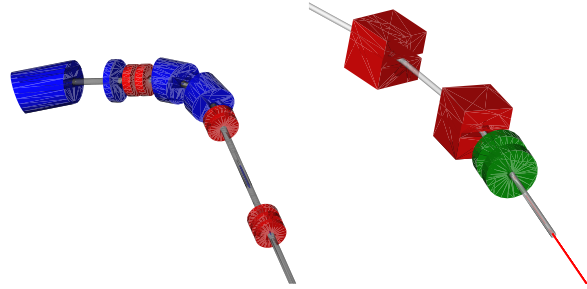


Figure 1: Geant4 visualizations with OpenInventor. The start of BL2A in BDSIM (left) and the end of BL2A in G4Beamline (right) with proton beam emerging

User Experience

Although these programs differ considerably in what is implemented, with G4Beamline generally having a more diverse set of features, the processes of setting up a simulation are quite similar, and the input usually has a 3-part structure: (1) Definition of beams, reference parameters, and other global options; (2) Definition of element prototypes; (3) Layout of the beam line.

Both BDSIM’s GMAD input language and the shell-like language used in G4Beamline support element *prototypes* defining common features, as well as element *instances* derived from the prototypes and corresponding to individual elements of the beam line. The instances can have specific names assigned to them, can set additional element parameters, and can override parameters in the prototype. This gives the input an object-oriented structure without the need for formal class definitions.

We also note that both programs can be used interactively with access to Geant4’s diverse visualization facilities, which are invaluable for debugging, diagnostic, and informational purposes. The OpenInventor visualization (via Coin3d libraries) in particular offers an efficient and functional interactive 3D viewer for the beam line geometry and particle trajectories, as exemplified in Figure 1.

The two programs differ fundamentally in the way the *reference path* (also called the “reference trajectory” or “design orbit”) of the beam line is defined. In BDSIM the reference path is *defined by the sequence of elements*, and is composed of line segments and circular arcs, the latter being defined by the effective lengths and bending angles of the dipole magnets. This makes it extremely easy to lay out the simulated beam line, and follows the convention of many optics and tracking codes. On the other hand, it can be restrictive because in real life some steering dipoles do not involve a bend in the layout of the beam line.

In contrast, in G4Beamline the reference path must be



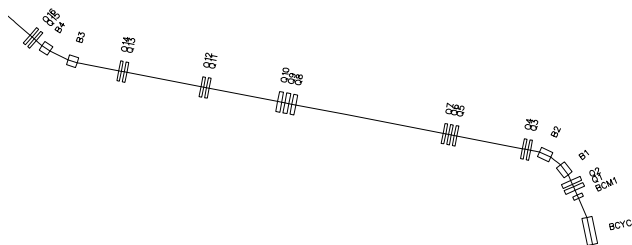
Figure 2: TRIUMF Beam Line 2A, showing final arc and quadrupole doublet (right) for delivery to ISAC target

explicitly defined by the user as a series of connected line segments. No arcs can be used, and the path through dipoles must be given in a piecewise-linear form, e.g. by bending the path by one-half the bending angle at the entrance and exit of the dipole. Once defined, this path acts as the moving origin for the local coordinate system, in which elements can be positioned and particle trajectories output. This offers greater flexibility and is closer to the way real beam lines are laid out and aligned, but it means that particle coordinates inside dipoles deviate widely from the reference path and are not directly comparable to accelerator coordinates in the conventional curvilinear system.

BEAM LINE 2A

As well as assessing Geant4-based simulation methods, the present study is motivated by their potential application to TRIUMF beam lines, and in particular to Beam Line 2A (BL2A, Figure 2) which supplies 500 MeV protons from the TRIUMF cyclotron to ISOL targets providing rare isotope beams for the ISAC and ISAC II facilities. The reliable performance of BL2A is essential to these multi-user facilities which support experiments in atomic and nuclear physics and nuclear astrophysics.

The beam line, shown schematically in Figure 3 is capable of switching between two independent targets, via the Y-magnet B3, however for present purposes we have modelled only the right-hand branch.



dinates representing the beam distribution at the foil. The COMA coordinates give a good description of the horizontal phase space and the energy distribution of particles. Due to the extreme edge angles of the cyclotron sector magnets, the vertical tracking in COMA is not very accurate. In the following section we describe a different approach used to populate the vertical phase space of the starting beam.

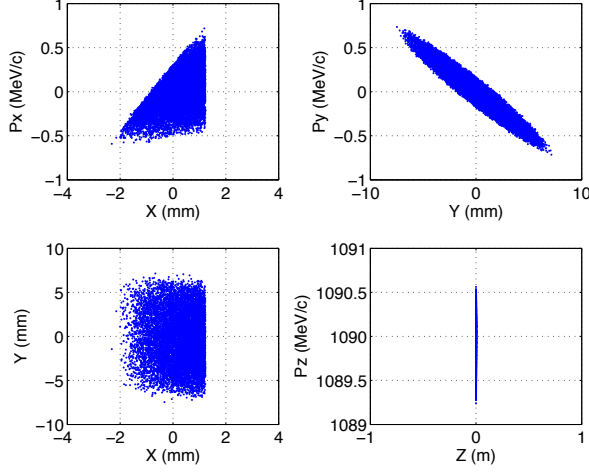


Figure 5: Scatterplot of COMA coordinates at the extraction foil

Optical model

The code TRANSOPTR has been used[6] to develop a model of BL2A which accurately describes the RMS beam characteristics. In this model, the beam sigma matrix parameters at the foil were fitted to beam profile measurements taken from monitors throughout the beam line, using the known dipole and quadrupole settings and estimates of the edge-focusing parameters at the exit from the cyclotron. Although the edge-focusing of the cyclotron magnet is not well quantified in detail, it was found that the aggregate effect could be modelled accurately by using the exit edge-angle of the optical model as an additional free parameter, thus obtaining the improved fit shown in Figure 6 (top).

We used the vertical emittance and ellipse parameters from this model in another simulation code ACCSIM[7] to generate the vertical phase space coordinates for our G4Beamline simulation. The simulation thus has two pre-processors, with the coordinates being merged (consistent with the correct total momentum vector) into an ascii input file which can be sequentially read by G4Beamline.

THE CYCLOTRON FIELD AND EXIT REGION

As seen in the optical model, the edge focusing as the proton beam exits the cyclotron, and in particular the edge angle, are important to the correct behaviour of the model throughout the length of the beam line. Although G4Beamline has an analytical treatment of the fringe fields in dipole magnets (and quadrupoles) it provides only rect-

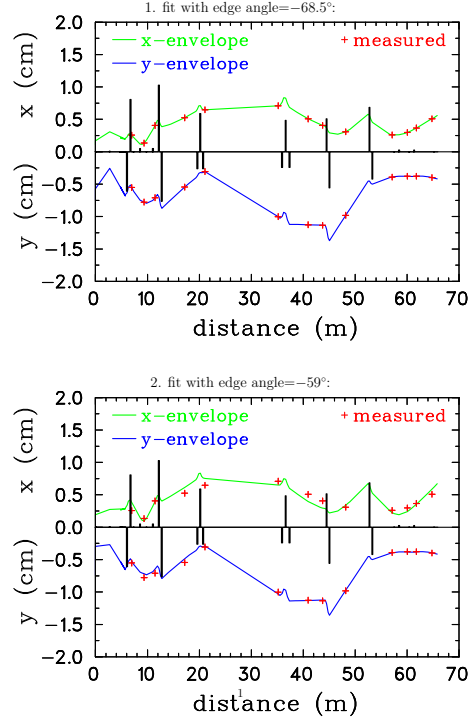


Figure 6: TRANSOPTR envelopes for different (simulated) cyclotron magnet edge angles, compared with measured beam profiles

angular and sector dipole shapes and does not allow an arbitrary edge angle to be specified.

To simulate the -68.5° exit edge angle we used a rectangular bend displaced from the reference path and rotated (Figure 7) so that protons start at the “side” of the magnet and exit at the appropriate angle from the “end” of the magnet, where the fringe field is modelled by G4Beamline.

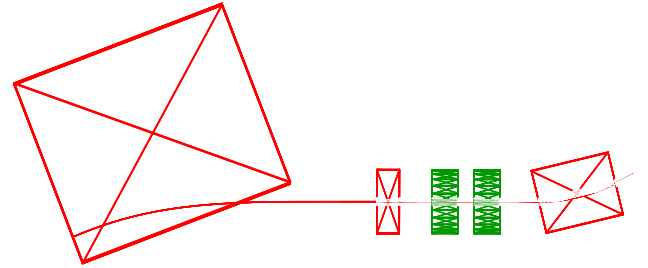


Figure 7: Rotated dipole and central proton trajectory in G4Beamline

Although this scheme gave the correct nominal bending and exit angles, we found that it could not reproduce or even approximate the horizontal focusing and vertical defocusing as seen in the beam envelopes in the optical model and supported by measurements. Although the fringe field depth is an adjustable parameter in G4Beamline we did not find that it was effective, either alone or in conjunction with the dipole edge angle, for tuning the integrated effects of the horizontal and vertical field components to obtain the correct beam envelopes at the cyclotron exit.

This issue will require further study and possibly a dif-

ferent approach to describing the dipole fringe fields. In the meantime we opted to bypass these difficulties by using ACCSIM to track the ensemble from the foil to the cyclotron exit and then export the coordinates to G4Beamline.

VALIDATION WITH ACCSIM + G4BEAMLINE

The simulation code ACCSIM combines matrix tracking with some basic interactions in matter, including Coulomb scattering, energy loss, and nuclear elastic scattering. No secondaries from inelastic interactions are tracked, however virtually all of these from the foil will be outside the acceptance of the beam line and in any case are unlikely to survive the first arc and contribute to downstream losses.

With pre-tracking by ACCSIM, in which a conventional edge-matrix represents the cyclotron edge focusing and fringe field effects, the envelopes of the optical model could be readily and precisely matched using the edge angle as a tuning parameter. Without a foil, the simulation yielded beam envelopes (RMS beam sizes) in good agreement (Figure 8, top) with those of the TRANSOPTR optical model. Introducing the foil in ACCSIM, the resulting envelopes (Figure 8, bottom) agreed very well with measurements, generally within $\sim 1\text{mm}$, and in some cases better than the TRANSOPTR envelopes which are derived using an RMS estimate of the emittance growth due to foil scattering.

In view of the sensitivity of the envelopes to initial conditions, and the fact that Geant4 is actually integrating through the dipole and quadrupole fields (with quadrupole strengths identical to those of TRANSOPTR), the close agreement of beam envelopes to the measured ones is remarkable and provides a strong validation of the model construction and of the precision of tracking in Geant4.

LOSSES DUE TO THE CYCLOTRON EXTRACTION FOIL

In a 5 mg/cm^2 extraction foil, Coulomb scattering is by far the dominant process contributing to the initial loss of protons in the cyclotron exit region and the first arc of BL2A. There are much smaller contributions from nuclear interactions, at about the 10^{-5} level, and from the very long but very sparsely populated Landau tail of the ionization energy loss distribution.

Most protons scattered to angles of more than a few milliradians will be lost within the cyclotron tank itself or in the exit horn region. Of the remainder initially accepted into BL2A, most will be lost in the first arc. A long-standing question is whether foil scattering plays any role in the distributed low level losses downstream. Initially we observed no losses at all when we introduced the extraction foil, either in a full G4Beamline simulation (with exaggerated envelopes due to the aforementioned cyclotron field problems) or in the ACCSIM-pre-tracked simulation. In both cases, multiple-scattering models are used which are not at their best in such thin layers of material.

Computer Codes (Design, Simulation, Field Calculation)

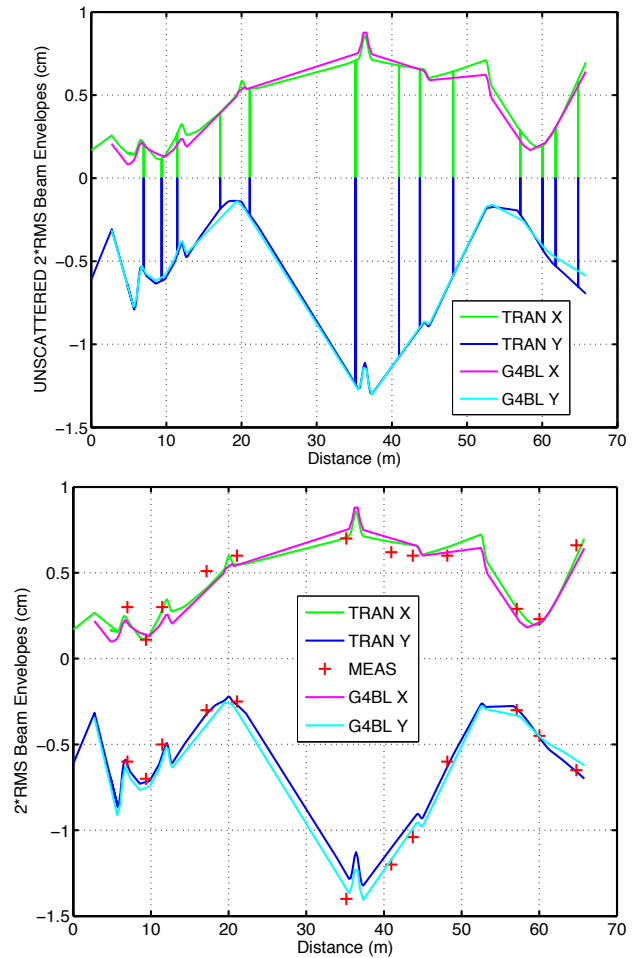


Figure 8: Beam envelopes from TRANSOPTR and G4Beamline compared for unscattered beam (top) and compared with measurements for scattered beam (bottom)

Scrutinizing the multiple scattering in ACCSIM, which is based on Molière theory with subsequent corrections and refinements, we observed that for this foil the unprojected scattering angle cuts off at around 3 milliradians, whereas the angular acceptance of the beam line with respect to the foil is roughly 10mr . Neither model includes the H^- stripping, but the preliminary analysis of the foil in ACCSIM indicated that protons traversing the entire foil thickness would undergo an average of only ~ 125 scatters.

Fortunately ACCSIM also has an iterated-single-scatter model for use in very thin foils. This model revealed the shortcomings of the multiple scattering treatment: for statistics of 10^6 protons a fraction of 4.2×10^{-4} of the beam was scattered into angles $> 3\text{mr}$ and extending out to $\sim 100\text{mr}$. This fraction is significant for a $70\mu\text{A}$ beam, since previous experiments with BL2A[8] established that mis-steering as little as 1 nA of beam at the first dipole resulted in measurable losses at the downstream spill monitors.

Accordingly we “forced” the single-scatter model in ACCSIM and used it to generate a “halo beam” consisting only of protons scattered beyond 3mr (Figure 9). This

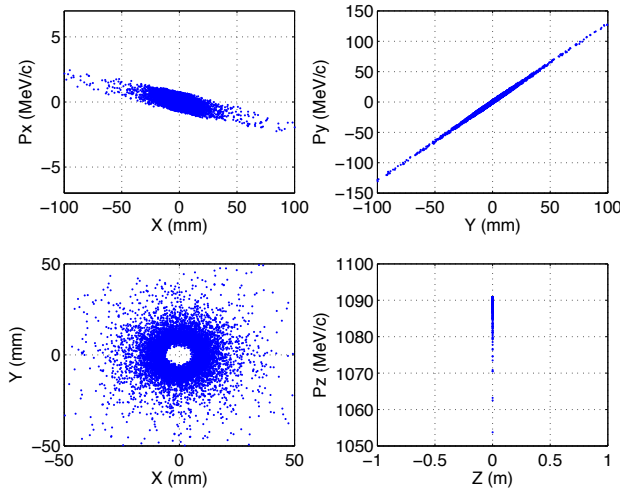


Figure 9: ACCSIM-generated halo beam around 3mr cone

ensemble was pre-tracked in ACCSIM to the cyclotron exit and then tracked in G4Beamline. Geant4 by default tracks all secondaries from the proton interactions, so we invoked G4Beamline's `kill=1` option on all the beam line components. This option stops protons as soon as they hit anything, and thus indicates the location of primary particle losses.

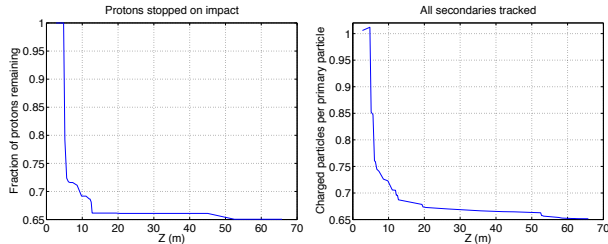


Figure 10: Left: survival of 3mr halo primary protons (all protons stopped on impact). Right: survival of charged particles (all secondaries tracked). Note lower axis limits.

As shown in Figure 10, about 65% of the halo beam is transmitted by the beam line. Of the remainder, almost all is lost within 12m, comprising the first arc and following quadrupole doublet. However, about 1% of the halo survives until ~ 45 m where it is lost in the next 7m (Q13/14 region). This region is of interest due to somewhat elevated routine spill monitor readings and radiation damage observed on a flange seal near the final arc.

In a subsequent run all secondaries were tracked and the resulting flux of charged particles (in principle detectable by spill monitors) is shown in Figure 10. Although detectable losses do propagate downstream, there is the indication of a “shadow” due to Q14 at 53m which may limit the diagnostic capacity of spill monitor BSM32 at 56m.

These results represent the rather ideal situation of a nearly perfectly-centered (within 0.5mm) beam line with perfectly-aligned elements free of field errors. In geometrical terms at least, this is the first tangible evidence that extraction foil scattering may contribute to losses far downstream from the foil location.

Computer Codes (Design, Simulation, Field Calculation)

CONCLUSIONS AND ACKNOWLEDGMENTS

Using input methods similar to those of optics and tracking codes, BDSIM and G4Beamline fulfill the promise of being able to develop a full-fledged Geant4 simulation of a beam line without any need for C++ programming and with much less effort. Although BDSIM was not completely adaptable to our application, we expect that it will continue to be developed and generalized, as it occupies a unique position with its combination of fast transfer maps in vacuum with Geant4 tracking in materials.

In G4Beamline we were able to proceed to an accurate model which was validated against measured beam envelopes. Using an iterated single-scatter treatment of Coulomb scattering in the cyclotron extraction foil, we established a baseline loss rate due to foil scattering of $\sim 1.4 \times 10^{-5}$ of the total beam intensity for a 5 mg/cm^2 foil. The G4Beamline model will be used to explore other possible loss mechanisms, such as field and alignment errors, as well as to study possible improvements in loss control, detection and monitoring.

These programs are significant achievements in the quest for realistic and predictive simulations of beam lines. We would like to thank the authors of BDSIM (Graham Blair, Ilya Agapov, Steve Malton, Lawrence Deacon) and the author of G4Beamline (Tom Roberts) for sharing their work and for answering many email queries as well as providing indispensable guidance and advice.

REFERENCES

- [1] S. Agostinelli et al., Geant4 – A Simulation Toolkit, *Nucl. Instrum. Meth. A* **506** (2003). J. Allison et al., Geant4 developments and applications, *IEEE Trans. Nuc. Sci.* **53**:1 (2006). <http://cern.ch/geant4/>
- [2] I. Agapov et al., BDSIM: a particle tracking code for accelerator beam-line simulations including particle-matter interactions, *Nucl. Instrum. Meth. A* **606** (2009). <http://www.pp.rhul.ac.uk/twiki/bin/view/JAI/BdSim>
- [3] T.J. Roberts et al., G4Beamline Particle Tracking in Matter-Dominated Beam Lines, European Particle Accelerator Conference (EPAC08) 2008. <http://g4beamline.muonsinc.com/>
- [4] J.R. Richardson and M.K. Craddock, Beam quality and expected energy resolution from the TRIUMF cyclotron, Fifth International Cyclotron Conference, Oxford 1969.
- [5] C.J. Kost and G.H. Mackenzie, COMA – A Linear Motion Code for Cyclotrons, 6th IEEE Particle Accelerator Conference, *IEEE Trans. Nuc. Sci.* **22**:3 (1975).
- [6] R. Baartman and Y.-N. Rao, TRIUMF Extraction and 500 MeV Beamline Optics, 2007 Particle Accelerator Conference (PAC07), Albuquerque NM.
- [7] F.W. Jones, Development of the ACCSIM Tracking and Simulation Code, 1997 Particle Accelerator Conference (PAC97), Vancouver.
- [8] L. Moritz et al., Investigation of Beam Trips in BL2A, TRIUMF Design Note TRI-DN-03-06 (2003).

END TO END SIMULATIONS OF THE GSI LINEAR ACCELERATOR FACILITY

G. Clemente, W. Barth, L. Groening, A. Orzhekhovskaya, S. Yaramishev, GSI, Darmstadt, Germany
 U. Ratzinger, R. Tiede, J.W. Goethe University, Frankfurt a.M., Germany.
 A. Kolomites, S. Minaev, ITEP, Moscow, Russia

Abstract

During the last year several numerical investigations have been started at GSI in order to improve the performance of the linear accelerator facility. The main activities regard the upgrade of the high current UNILAC accelerator including the severe upgrade of the HSI injector, the HITRAP decelerator and, in the frame of the future FAIR project, the development of the new dedicated proton linac. End to end beam dynamics simulations are a powerful tool concerning the machine design, commissioning and optimization. Particle distributions, generated from beam emittance measurements, are transferred through the whole chain of accelerating structures and beam transport lines. Detailed calculations of space charge effects as well as external and measured mapping of the electromagnetic fields are used to provide the most reliable results. The paper presents a general overview of all activities performed at GSI concerning the linear accelerator complex.

INTRODUCTION

The scientific program at FAIR requires a severe upgrade of the existing GSI linear accelerator complex in terms of beam brilliance and absolute beam current. To fulfill the experimental heavy ion requirements the UNILAC must provide up to 3.3×10^{11} U^{28+} particles within macropulses of 100 μ s long [1]. At the final energy of 11.4 MeV/u the beam will be injected into the SIS18 with a repetition rate up to 4 Hz.

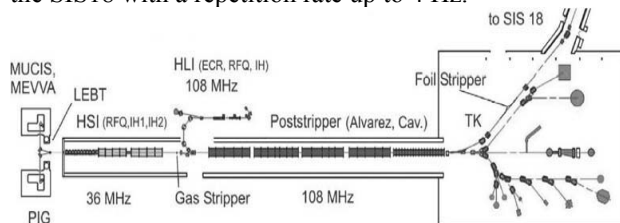


Fig.1: The present linear accelerator complex at GSI.

On the other side, FAIR will provide up to 7×10^{10} p-bar/h which, taking into account the antiproton production and cooling rate implies a primary proton beam of 2×10^{16} p/h. This intensity is far beyond the capabilities of the existing UNILAC and, for that reason, a new dedicated proton injector has to be built [2].

In parallel, activities on linear accelerators at GSI are not only focused on the FAIR project. Recently, in the frame of the atomic physics HITRAP project [3], a

4.00 MeV/u Ni^{28+} beam coming from the ESR was decelerated to 6 keV/u making available high charged and cooled beams for trapping experiments.

UNILAC UPGRADE

The next upgrade activities are mainly focused on the low energy front end which represented a bottleneck concerning operation with higher brilliance. The High Current Injector [4] consists of a 36 MHz IH-RFQ from 2.2 keV/u to 120 keV/u and a short 11 cell adapter RFQ called Super Lens. The following acceleration step is performed by two IH-DTL's which deliver a 1.4 MeV/u beam. After the stripping in a supersonic gas jet, uranium beams with charge state of 28+ are delivered to the Alvarez-DTL and accelerated to 11.4 MeV/u with minor losses.

Upgrades of the HSI

The first upgrade of the HSI-RFQ was performed in 2004 after five years of continuous operation. New electrodes were produced increasing the quality of surfaces and thus reducing the RF power consumptions from 650 kW to 380 kW. Additionally, the RFQ Input Radial Matcher (IRM) was redesigned to improve the beam transmission through the whole front-end system.

DYNAMION calculations were performed using particles distributions generated by measured emittances and predicted an intensity gain of up to 15% for high current uranium beam (15 mA). Those simulations were later on perfectly confirmed by measurements [5].

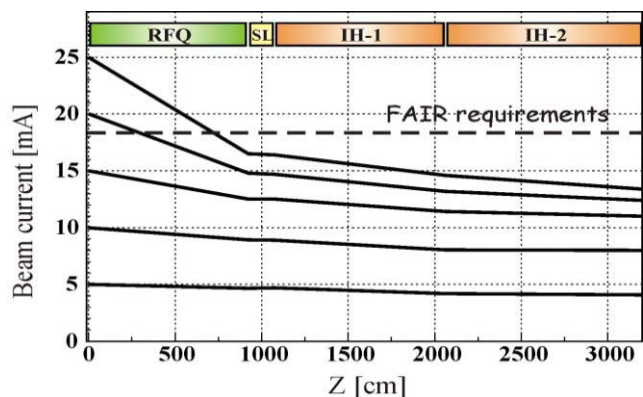


Fig.2: The HSI simulated performance before the 2009 upgrade for different input current in comparison with the requirements for FAIR.

This first upgrade increased the measured transmission up to 55 % but, as Fig 2 shows, even in simulations it was not possible to match the current requirements of FAIR. For this reason, in 2009 a new design of the RFQ was investigated in order to increase the phase advance and the acceptance of the accelerating focusing channel [6]. The new electrode profile was designed with higher voltage, keeping the maximum field at the electrode surface and with the same total length of the previous design. To compensate the change in resonance frequency the carrier ring design was consequently modified as well. The Input Radial Matcher was also redesigned in order to improve the beam matching to the RFQ by means of the LEBT quadrupole.

Beam dynamics in the beginning of the gentle buncher was optimized to provide for rapid and uniform (as possible) separatrix filling. To avoid an excessive RF-defocusing and a significant space charge influence, the modulation and the synchronous phase in the gentle buncher increase considerably slower compare to the existing RFQ-design. Dedicated code DESRFQ developed at ITEP, Moscow, was used to perform those studies.

A comparison between the new and the old design of the HSI RFQ is summarized in Tab.1.

Tab. 1: Comparison between the 2009 design and the old one of the HSI-RFQ.

Parameter	2009	1999
Voltage [kV]	155	125
Avr. Radius [cm]	0.6	0.55-0.50-0.7
Maximum Field [kV/cm]	312	318
Min. Aperture [cm]	0.41	0.38
Min Trans. σ [rad]	0.55	0.45
Norm Trans Accep. [mm rad]	0.86	0.73
Total Length [cm]	9217	9217
No. of modulated cell	394	343

Simulation Results

Particle simulations were performed with DYNAMION including integrated external 3D mapping of the electromagnetic fields. Exact topology of the Super Lens was included into DYNAMION and the intrinsic 8-terms field calculator was compared with the external 3D field mapping generated with EM Studio. The same method was applied as well to simulations of the HSI-RFQ high energy end. In particular, the use of EM Studio allowed to reproduce the exact geometry of the rods end, including curvature and shape as well as the real topology of the end flange. This combination between the intrinsic Laplace solver of DYNAMION and the external mapping generated with EMS resulted in reliable beam dynamics simulations.

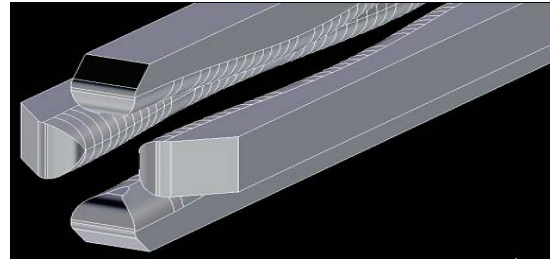


Fig.3: A detail of the modeling with EMS of the Super Lens RFQ. End flanges are not shown here but present in calculations.

Fig.4 shows the calculated the beam transmission through the new RFQ as a function of input beam current when input emittance or brilliance are kept constant. At the FAIR requirements of 18 mA transmission of around 85 % at constant emittance can be reached while 78 % of the beam reaches the end if the brilliance is preserved.

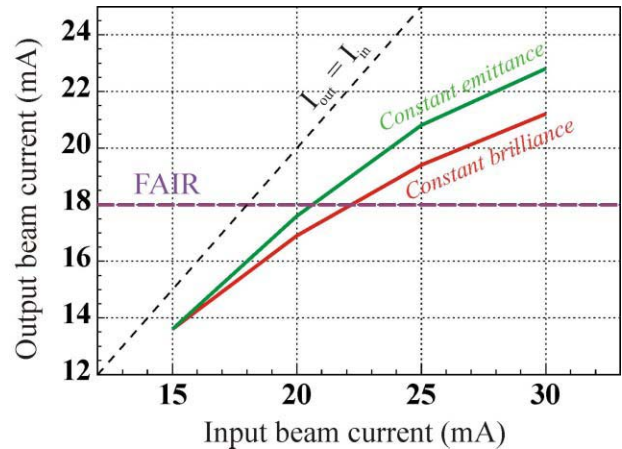


Fig.4: The transmission plot for the new HSI-RFQ when input emittance or brilliance are kept constant.

Fig.5 shows a comparison between the old and the new design of the beam current inside a given emittance at exit of the RFQ: the requirements for FAIR are fulfilled within a reasonable emittance of 20 mm mrad . Moreover, the behavior of the beam curve is a serious proof that the core of the beam remains well compact during the acceleration process.

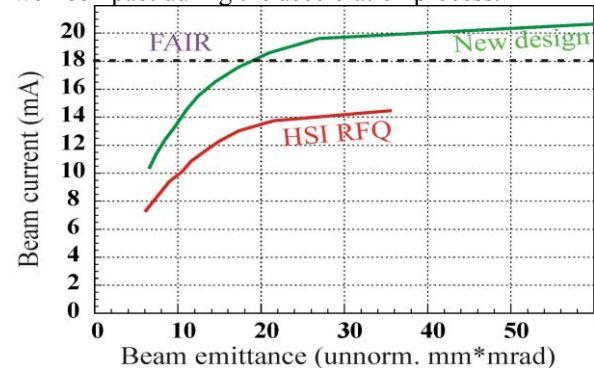


Fig.5 Beam current as function of the emittance for the old and the 2009 design.

The new HSI RFQ has been commissioned in July with high current Argon beam. Experimental results are in agreement with the calculations performed with DYNAMION and a detailed comparison is in preparation

The FAIR Proton Injector

A significant part of the scientific program at FAIR will be dedicated to antiproton physics. The acceleration chain for the production of the intense p-bar beam is shown in Fig.6 and it starts with a dedicated proton linac.

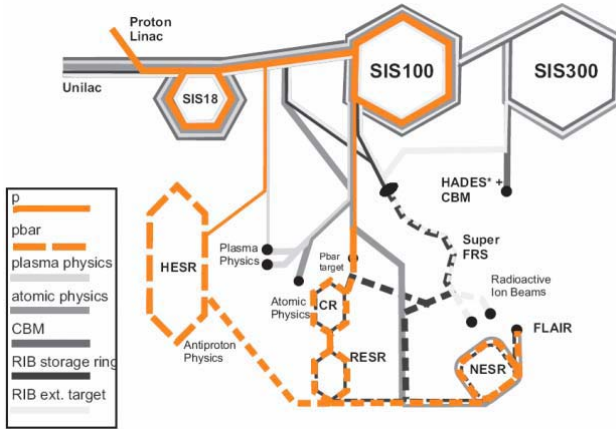


Fig.6: The FAIR accelerator chain showing the path for the production of cooled antiproton beams.

Tab.2: The main parameter of the FAIR proton injector

Source	H^+ , 95 keV, 100 mA
LEBT (2 solenoids foc.)	95 keV, 100 mA
RFQ (4-rod)	3 MeV, $\epsilon = 2 \mu m$
CH-DTL 325 MHz	70 MeV
Current [mA]	35 in operation
RF Pulse [μs]	70
Beam Pulse [μs]	36
Repetition Rate [Hz]	4
Norm. Transv. Emitt. [μm]	< 2.8
Relative Momentum Spread	$< 1 \text{ ‰}$
Length [m]	~ 23
Total Length [m]	< 40

This new injector will be operated at 325 MHz and it has to provide a 35 mA beam pulse of 36 μs . The final energy is fixed at 70 MeV in order to avoid any jump in resonance frequency and to allow an efficient and fast cooling rate. The ion source will be an ECR 2.45 GHz delivered by CEA, Saclay, and can provide 100 mA of proton beam at the extraction energy of 95 keV. Afterwards a four rod RFQ designed by Frankfurt University will accelerate the beam to 3 MeV where the main linac begins. Even if 35 mA are required for the multiturn injection into the synchrotron, the design of the linac assumes higher current, up to 100 mA from the ion source.

An example of two RFQ output distributions are shown in Fig.7 where 45 mA and 100 mA,

Linear Accelerators

respectively, are injected into an RFQ optimised for 45 mA. As one can see there are no major differences between the two distributions, showing that the RFQ design is very robust against higher current intensity.

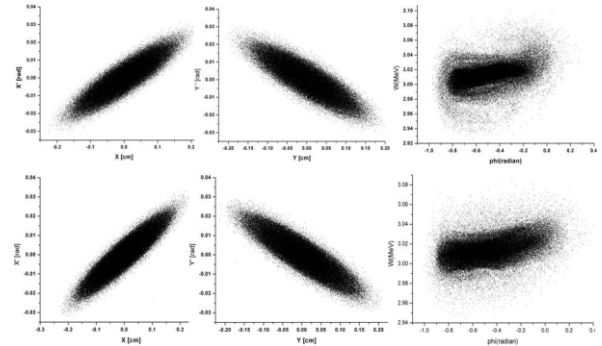


Fig.7: A 45 (top) and a 100 mA RFQ output distribution.

Tab3.: RMS emittance for two RFQ output distributions at different current

	45 mA	100 mA
RMS ϵ norm hor	0.26	0.25
RMS ϵ norm ver	0.26	0.25
RMS ϵ norm $\Delta\phi \Delta W$	1.29	1.25

After the RFQ the beam is matched into the main linac by means of an independent phase buncher; the MEFT includes an XY steerer and a current transformer placed immediately after the RFQ. Two focusing elements placed in front and beyond the buncher, respectively, ensure the full transmission while the diagnostics is completed by a compact phase probe placed just in front of the main linac.

Status of the Project

In the frame of FAIR the proton injector is well advanced. Agreement with CEA, Saclay, has been reached regarding the ion source while IAP Frankfurt has sent in production the second resonator of the DTL. The first 2.5 MW klystron from Toshiba was delivered in 2008 and, presently, the design of a high power RF test bench is in progress at GSI. Measurements on the first coupled CH are expected to be started in 2010.

Beam Dynamics through the DTL

The main linac is based on 12 CH cavities developed at Frankfurt University [7]. Those kind of cavities in combination with the KONUS beam dynamics allow an extremely efficient acceleration process in the low-medium beta range. In order to reduce the RF requirements the 12 CH cavities are grouped in six independent coupled resonators. A 1:2 scaled model built at Frankfurt University [ref] has demonstrated the validity of this concept while construction of the FAIR second resonator is at the moment in preparation.

An extended diagnostics section is integrated into the DTL after the 3rd coupled module. This section will also include scrapers to dump of outer particles and keep the beam core as compact as possible.

Fig.8 shows the general layout of the proton linac including an alternative design discussed later on

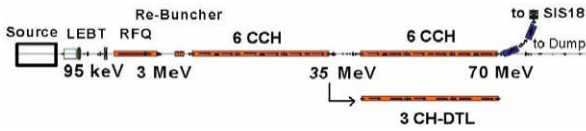


Fig.8: The general layout of the FAIR Proton injection showing the two options under discussion

Fig.9 shows a comparison of the 99% beam envelopes for 45 and 100 mA, respectively. The design reveals to be very stable against current intensity variation and, even with quadrupoles optimized for the 45 mA case, 100 mA are transported with less than 1 % beam losses.

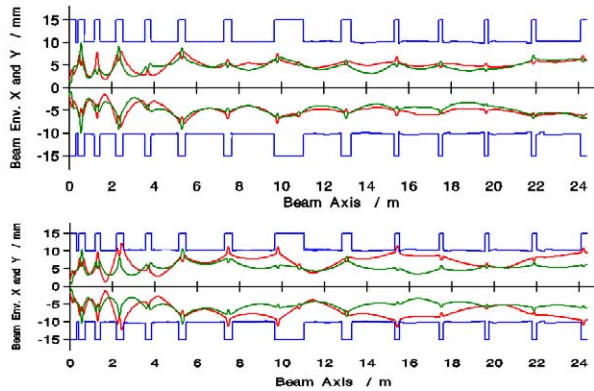


Fig.9: 99 % beam envelopes for two different RFQ output distributions. On top, 45 mA and bottom 100 mA

Examples of the output distribution for those two cases are shown in Fig.10 while Tab.4 summarizes the 100 % RMS at the output energy. In both cases the emittance requirements for the multturn injection into the SIS18 are fulfilled

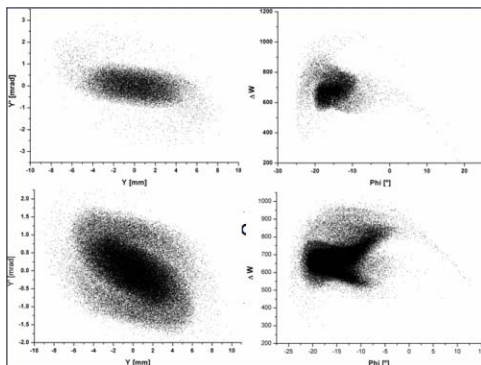


Fig.10: The transversal and longitudinal distribution for 45 mA (top) and 100 (bottom).

Tab.4: The rms parameters as dependence of the input current at the output of the proton linac

Parameter	45 mA	100 mA
RMS ϵ norm X-X' mm mrad	0.40	0.383
RMS ϵ norm Y-Y' mm mrad	0.44	0.409
RMS ϵ norm $\Delta\Phi$ - ΔW keV/ ns	2.09	2.09

Alternative Layout

The use of the KONUS allows to build long accelerating sections without any focusing element. For the proton linac, this feature could be fully exploited in the second part of the DTL after the diagnostics section where the space charge effect is already strongly reduced. The coupled cavity could be replaced by long standard CH-DTL leading to a general simplification of the mechanical design and to a reduction of the number of focusing elements. This concept is under investigation at GSI and preliminary results with a beam current of 45 mA are promising [8]. As one can see from Fig.11 and Fig.12, the beam is full transmitted through the long sections and the output distribution and RMS emittances are comparable with the original design. At present further investigations are in progress in order to test the robustness of this solution with higher current and against possible misalignments and operational errors.

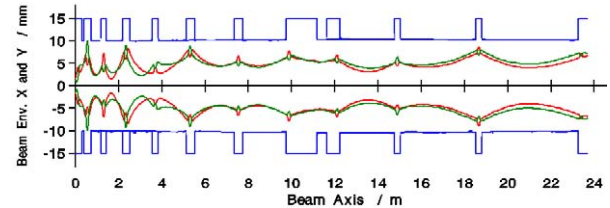


Fig.11: 99% beam envelopes for the alternative design

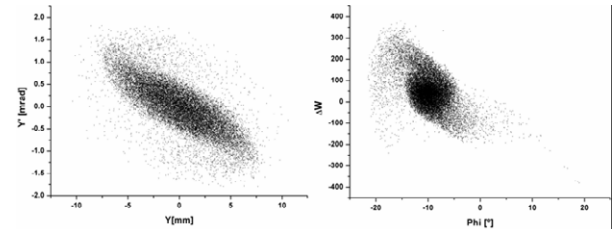


Fig.12: The output distributions for the alternative design for an RFQ output current of 45 mA.

Tab5: Rms emittance for the alternative design at 45 mA

RMS ϵ norm X-X' mm mrad	0.383
RMS ϵ norm Y-Y' mm mrad	0.409
RMS ϵ norm $\Delta\Phi$ - ΔW keV/ ns	2.09

HITRAP

In the frame of the atomic and nuclear physics program a new cooling trap called HITRAP, Heavy Ion TRAP, is under commissioning at GSI. The scientific program aims for investigations of high charged ions in the region of $Z\alpha \approx 1$ where no perturbation method can be applied. The experimental area which can be

covered by HITRAP is wide and includes, among others, investigation of the g-factor of bound electrons and alpha particles, hyperfine and X ray spectroscopy, mass measurements and surface interactions.

The general scheme of HITRAP is shown in Fig.13. A 4 MeV/u beam with $q/A = 1/3$ covering $^{238}\text{U}^{92+}$ is extracted by the ESR, bended by two dipole magnets, focused by a quadrupole duplet and decelerated to 500 keV/u by a 108 MHz IH-DTL. A 4-rod RFQ performs the last stage of deceleration to 6 keV/u. The phase and energy matching from the ESR to the IH is ensured by two independent bunchers resonating at 108 and 216 MHz, respectively.

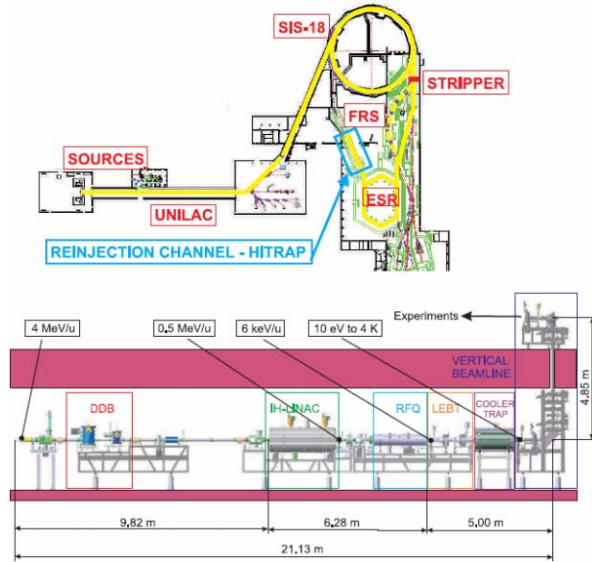


Fig.13: The location of HITRAP at GSI and a detail of the decelerator complex after the ESR.

Beam Dynamics and Ion Optic

The starting point for the optics calculations was the extracted beam from the ESR. Concerning the matching the acceptance of the IH (2.2 and 2.3 mm mrad) was used even if the emittance of the ESR is expected to be smaller (by a factor two or more). Fig. 14 shows the beam envelopes, from the exit of the ESR to the IH entrance.

The first 4 gap buncher accepts (into a single bucket) 100° out of the cw beam corresponding to 28% of intensity. The following rebuncher, operating at double frequency, reduces further the phase spread to 10° before the injection into the IH. It's important to remark that particles escaping the first RF bucket are not necessarily lost but will be transported without full deceleration through the IH.

The IH consists of 25 gaps for a total voltage of 11.15 MV. The focusing scheme is based on a KONUS lattice with an internal triplet placed behind 15 gaps.

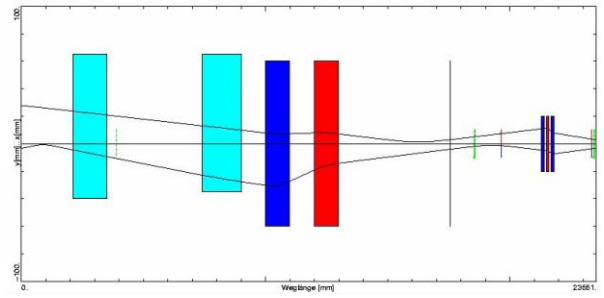


Fig.14: The beam envelopes between the ESR and the IH.

The total length is around 4 meter and the power consumption is kept lower than 180 kW. Fig.15 shows the beam distribution for a single RF bucket at the beginning of the deceleration line and at 500 keV/u after the IH.

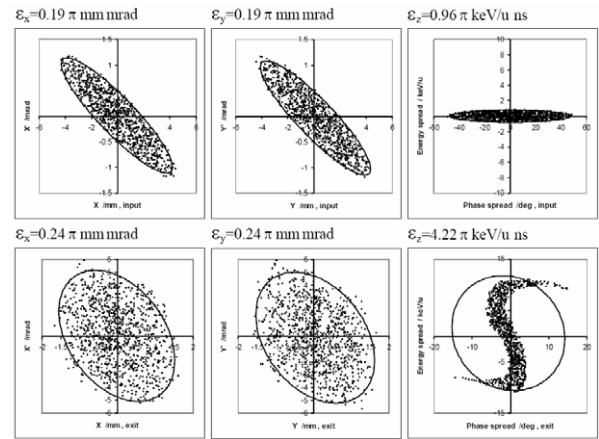


Fig.15: The beam distribution at the beginning of the decelerating line, namely in front of the first buncher (top), and after the IH (bottom) within a single RF bucket.

The design of the 4 rod-RFQ is modelled along the GSI HLI RFQ, which serves as injector for moderate charge states ($q/A > 1/9$). The HITRAP decelerator is designed for high charge states of $q/A > 1/3$, which reduces the RF-power requirements and allows for a shorter structure. The RF consumption is 80 kW and the total length is around 1.9 m with an aperture of 4 mm in radius.

The crucial RFQ-design parameters are the longitudinal emittance, the energy spread, and the phase width of the beam. The required output emittance restricts the possible input phase width and energy spread. A value of an energy spread of $\Delta W/W = 1\%$ at the RFQ-high energy input would translate to 66% at the low energy end.

The developed deceleration scheme reduces the output-energy spread to $\pm 6\%$. With an input-phase width of $\Delta\phi < 20^\circ$ and asynchronous deceleration, the beam pulse can be kept compact with reduced phase oscillations. An energy spread of 2% is the upper

useful limit while, for the radial emittance, a value of $0.24 \pi \text{ mm mrad}$ has been used for input.

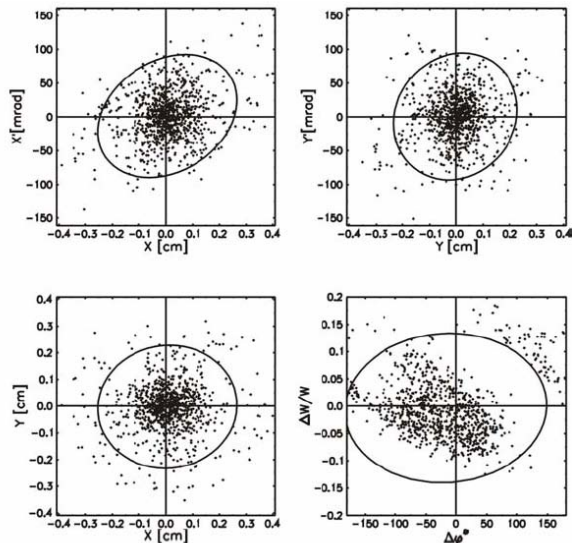


Fig. 16: Particles distribution at the exit of the RFQ assuming an energy spread of 2 % as input parameters

Presently, the HITRAP facility is under commissioning at GSI [9] and in early 2009 6 keV/u Ni^{28+} particles were detected at the end of the RFQ.

MILESTONES AND CONCLUSIONS

At GSI several activities run in parallel concerning the linear accelerator complex. The high intensity beam requirement of the FAIR project was the main motivation for a new concept design of the HSI RFQ which has been successfully commissioned in 2009. This new design was accomplished by an intense simulation campaign. Particle distributions were generated by measured emittances and external 3D mapping reproducing the exact topology of the accelerator components were integrated into the simulations code.

The R&D activities for FAIR are not limited to the existing facility since the antiproton program requires a complete new dedicated proton injector. Beam dynamics simulations are in progress applying different beam current and, concerning the structure development, the first coupled CH cavity is under fabrication. The first RF klystron has already been delivered and a high power test bench is going to be prepared at GSI.

Beside the FAIR project, recently a decelerating beam line called HITRAP has been commissioned delivering highly charged ions for trapping experiments.

REFERENCES

- [1] FAIR Technical Report. GSI, 2006
- [2] U. Ratzinger, A 70 MeV Proton Linac for the FAIR facility based on CH-DTL, LINAC2006, Knoxville, 2006
- [3] HITRAP Technical Report, GSI.
- [4] U. Ratzinger, The New GSI Prestripper Linac for High Current Heavy Ion Beams, LINAC96, Geneva, Switzerland, p. 288 (1996)
- [5] W. Barth, Development of the UNILAC Towards a Megawatt Beam Injector, Proc. of 22nd International LINAC04, Luebeck, Germany 2004.
- [6] A. Kolomiets, Upgrade of the HSI-RFQ, LINAC2008, Victoria, Canada, 2008
- [7] G. Clemente, The room temperature CH-DTL and its application for the FAIR proton injector
- [8] G. Clemente, Investigation of the Beam Dynamics Layout of the FAIR p-Linac, HB2008 Workshop, Nashville, U.S.A., 2008
- [9] L. Dahl The HITRAP Decelerator Project at GSI. Status and Commissioning Report. 2009

APERTURE AND BEAM-TUBE MODELS FOR ACCELERATOR MAGNETS*

H. De Gersem[†], Katholieke Universiteit Leuven, Belgium
S. Koch, T. Weiland, Technische Universität Darmstadt, Germany

Abstract

Standard 2D magnetodynamic finite-element models for accelerator magnets are accomplished by dedicated models for the aperture and for the beam-tube end parts. The resulting hybrid and coupled models necessitate the application of specialized algebraic solution techniques in order to preserve the computational efficiency, i.e., matrix-free iterative solvers combined with fast Fourier transforms and Schwarz-type preconditioners.

INTRODUCTION

3D finite-element (FE) models of accelerator magnets may become prohibitive when transient phenomena at small temporal and spatial scale should be resolved, i.e., eddy current effects in windings, beam tube and yoke or filamentary and coupling effects in superconductive cables. Simulation times of several hours have been reported. As a consequence, such calculations are only feasible at later stages of the design process, when geometry, materials and operating conditions are more or less fixed. At an earlier design stage, parameter variations and optimization steps are carried out, almost exclusively on the basis of semi-analytical formulae. It makes sense to support this design phase by FE models that succeed in attaining lower but acceptable accuracies within substantially smaller simulation times, compared to transient 3D simulation [3]. Efforts in the direction of this goal consider 2D FE models where extensions are implemented that deal with typical 3D effects and model some small-scale effects in a problem specific way. In this paper, two extensions for 2D and 3D FE models are proposed. The high-resolution aperture model and the beam-tube end model developed here, both significantly increase the modeling power for superconductive magnets and succeed in keeping the computation time for transient 2D FE simulation as low as a few minutes.

APERTURE MODEL

Domain Decomposition and Mixed Formulation

The magnet geometry is divided in two parts: an outer domain Ω_1 including the windings and yoke and a cylindrical inner domain Ω_2 in the magnet aperture (Fig. 1). The

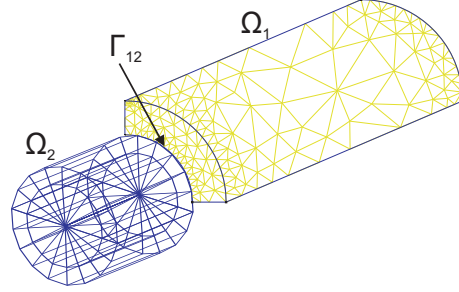


Figure 1: Magnet model: FE mesh for the outer part and tensor-product grid for the aperture.

interface is denoted by $\Gamma_{12} = \Omega_1 \cap \Omega_2$. A mixed magneto-quasistatic (MQS) formulation is applied:

$$\nabla \times \left(\nu \nabla \times \vec{A} \right) + \sigma \frac{\partial \vec{A}}{\partial t} = \vec{J}_s \quad \text{in } \Omega_1 ; \quad (1)$$

$$-\nabla \cdot (\mu \nabla \psi) = 0 \quad \text{in } \Omega_2 , \quad (2)$$

where \vec{A} is the magnetic vector potential, ψ the magnetic scalar potential, \vec{J}_s the applied current density, μ the permeability, $\nu = 1/\mu$ the reluctivity and σ the conductivity. The normal continuity of the magnetic flux density $\vec{B} = \nabla \times \vec{A} = -\mu \nabla \psi$ and the tangential continuity of the magnetic field strength $\vec{H} = \nu \nabla \times \vec{A} = -\nabla \psi$ are enforced at the interface Γ_{12} . The domains are equipped with different formulations that are dual with respect to each other. The MQS formulation in terms of \vec{A} is capable of considering the eddy current phenomena in the yoke, beam tube and windings, at the expense of being a vectorial partial differential equation (PDE). On the contrary, the MQS formulation in terms of ψ is static but is a scalar formulation. Such so-called *mixed* formulations have been frequently used for MQS simulation in the eighties, especially because of the relatively small number of degrees of freedom which was beneficial for the direct and Krylov-type solvers used at that time [11]. A drawback of a mixed formulation is the fact that the computed magnetic energy and power loss do not converge monotonically with respect to the mesh size as is the case for the non-mixed formulations. The motivation for choosing a mixed formulation will become clear below.

Discretization and System Properties

\vec{A} is discretized in Ω_1 by the standard lowest-order edge elements \vec{w}_j , whereas ψ is discretized in Ω_2 by standard

* This work was supported by the Helmholtzzentrum für Schwerionenforschung GmbH (GSI), Darmstadt and by the Katholieke Universiteit Leuven under grant STRT1/09/041.

[†] herbert.degersem@kuleuven-kortrijk.be

lowest-order nodal elements N_q :

$$\vec{A} = \sum_j u_j \vec{w}_j \quad \text{in } \Omega_1 ; \quad \psi = \sum_q v_q N_q \quad \text{in } \Omega_2 . \quad (3)$$

where the degrees of freedom (DoFs) for \vec{A} and ψ are stored in the vectors \mathbf{u} and \mathbf{v} , respectively. The application of the Ritz-Galerkin approach results in the coupled system of equations

$$\begin{bmatrix} \mathbf{K} + \mathbf{M} \frac{d}{dt} & \mathbf{B} \\ \mathbf{B}^T & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix} , \quad (4)$$

further also denoted by $\mathbf{Ax} = \mathbf{b}$ [1]. Eq. (4) is transferred in frequency domain or discretized in time by any implicit time-integration technique. For conciseness, this is not explicitly carried out in the formulae. The matrix entries are

$$\mathbf{K}_{ij} = \int_{\Omega_1} \nu \nabla \times \vec{w}_j \cdot \nabla \times \vec{w}_i d\Omega ; \quad (5)$$

$$\mathbf{M}_{ij} = \int_{\Omega_1} \sigma \vec{w}_j \cdot \vec{w}_i d\Omega ; \quad (6)$$

$$\mathbf{B}_{iq} = - \int_{\Gamma_{12}} (\nabla N_q \times \vec{w}_i) \cdot d\vec{\Gamma} ; \quad (7)$$

$$\mathbf{G}_{pq} = \int_{\Omega_2} \mu \nabla N_q \cdot \nabla N_p d\Omega ; \quad (8)$$

$$\mathbf{f}_i = \int_{\Omega_1} \vec{J}_s \cdot \vec{w}_i d\Omega . \quad (9)$$

The matrices \mathbf{K} and \mathbf{M} are both sparse, symmetric and positive semi-definite. The zero eigenvalues of \mathbf{K} are related to the undetermined gradient fields, whereas the zero eigenvalues of \mathbf{M} are related to the non-conductive model parts. The MQS formulation in terms of the magnetic scalar potential leads to a sparse, symmetric, positive definite system matrix \mathbf{G} . The mixed system (4) is sparse, symmetric and indefinite. The indefiniteness is related to the mixed character of the formulation. If the inverse of \mathbf{G} can be obtained or applied in a cheap way, the possibility exists to turn over to the Schur complement system

$$\underbrace{\left(\mathbf{K} + \mathbf{M} \frac{d}{dt} + \mathbf{B} \mathbf{G}^{-1} \mathbf{B}^T \right)}_{\mathbf{S}} \mathbf{u} = \mathbf{f} . \quad (10)$$

The Schur complement matrix \mathbf{S} is symmetric and positive semi-definite. \mathbf{S} contains large dense blocks due to \mathbf{G}^{-1} . Its application in the form of (10) is therefore substantially more expensive than the application of $\mathbf{K} + \mathbf{M} \frac{d}{dt}$. The sparsity structure of the Schur complement system is the same as for a FE, boundary-element (BE) coupled formulation where the normal derivatives are eliminated [8].

The mixed system can be solved by the Minimal Residual (MINRES) method [10]. MINRES requires a positive semi-definite preconditioner. As a first choice,

$$\tilde{\mathbf{A}}_1^{-1} = \begin{bmatrix} \tilde{\mathbf{L}}^{-1} & 0 \\ 0 & \tilde{\mathbf{G}}^{-1} \end{bmatrix} \quad (11)$$

Superconducting Magnets

is proposed. Here, $\tilde{\mathbf{L}}^{-1}$ denote any possible positive semi-definite approximation to $(\mathbf{K} + \mathbf{M} \frac{d}{dt})^{-1}$ such as, e.g., an Incomplete Cholesky (IC) preconditioner or Algebraic Multigrid (AMG) approach. As a second diagonal block, an operator $\tilde{\mathbf{G}}^{-1}$ is used as an approximation to \mathbf{G}^{-1} . It has been shown that the convergence of the MINRES method with this preconditioner is not optimal, even if no approximations are used for the diagonal blocks. Better is

$$\tilde{\mathbf{A}}_2^{-1} = \begin{bmatrix} \tilde{\mathbf{S}}^{-1} & 0 \\ 0 & \tilde{\mathbf{G}}^{-1} \end{bmatrix} \quad (12)$$

where $\tilde{\mathbf{S}}^{-1}$ is an approximation to the inverse of \mathbf{S} . If no approximations are introduced, this preconditioner causes the MINRES algorithm to converge in exactly 2 iteration steps. The construction of an approximation to \mathbf{S} is not trivial. The *additive Schwarz*-type preconditioner

$$\tilde{\mathbf{S}}_2^{-1} = \left(\mathbf{K} + \mathbf{M} \frac{d}{dt} \right)^{-1} + \mathbf{B} \mathbf{G} \mathbf{B}^{-1} \quad (13)$$

is known to be a simple and effective possibility [12]. Alternatively, it is possible to solve (10) by the Conjugate Gradient (CG) method. As preconditioners, either the approximation $\tilde{\mathbf{S}}_1^{-1} = \tilde{\mathbf{L}}^{-1}$ or the approximate inverse $\tilde{\mathbf{S}}_2^{-1}$ to the Schur complement system can be used.

Spectral Resolution and System Representation

The spatial resolution in the aperture can be improved considerably by choosing an orthogonal set of basis functions instead of the set of standard lowest order nodal shape functions $\{N_q\}$. The cylindrical aperture domain $\Omega_2 = [-R, R] \times [0, \pi] \times [-Z, Z]$ has a radius R and a length $2Z$ centered with respect to the $z = 0$ plane. In the spectral-element (SE) approach, basis functions of the form

$$M_q(r, \theta, z) = T_{q_1} \left(\frac{r}{R} \right) e^{-j\lambda\theta} T_{q_2} \left(\frac{z}{Z} \right) \quad (14)$$

are considered at a cylindrical tensor-product grid (Fig. 1). In the r - and the z -direction, the shape functions vary as Chebyshev functions, whereas in the θ -direction harmonic functions are used. To avoid a clustering of collocation points at the $r = 0$ axis, the r -coordinate is taken from $-R$ to R , the θ -coordinate is restricted to $[0, \pi]$ and an even number of collocation points is used in the r -direction. The matrices $\mathbf{B} = \mathbf{Z}^T \mathbf{Q}^T \tilde{\mathbf{D}}$ and $\mathbf{G} = \mathbf{D}^H \mu \mathbf{D}$ are available in factorized form. Here, \mathbf{Z} interpolates the fictitious interface currents between the spectral collocation points of the tensor-product grid and the edges of the FE mesh, \mathbf{Q} selects the spectral DoFs at the collocation points at the interface, $\mathbf{D} = [\mathbf{D}_r \quad \mathbf{D}_\theta \quad \mathbf{D}_z]^H$ and \mathbf{D}_r , \mathbf{D}_θ and \mathbf{D}_z are the spectral differentiation matrices along r , θ and z . The particular choice of the collocation points allow to invoke each of these differentiation matrices by means of the Fast Fourier Transform (FFT):

$$\mathbf{D}_r = \mathbf{D}_\theta = \mathbf{D}_z = \mathbf{F}^{-1} \mathbf{U} \mathbf{F} , \quad (15)$$

where \mathbf{U} is a diagonal matrix [13]. The spectral differentiation matrices are fully populated. The operator \mathbf{Q} can be represented by an index set. The interpolation operator \mathbf{Z} is stored in a sparse matrix format. The coupling matrix \mathbf{B} has a sparse row structure related to the FE DoFs and a dense column structure related to the SE DoFs. The Schur-complemented FE-SE system is comparable to a FE-BE coupled system [8] and equivalent to the FE system in which the SE domains are considered as specialized Robin-type boundary conditions applied at Γ_{12} [2].

A SE discretization is known to attain an exponential convergence of the discretization error with respect to the number of DoFs. Because a part of the model remains discretized by FEs, the convergence of the overall model is limited to the convergence of the FE part. Nevertheless, the SE discretization of the aperture allows to considerably reduce the number of DoFs required to sufficiently resolve the aperture field.

During the system-solving phase, most of the computation time is spent to the multiplication by the system matrix and the application of the preconditioner. Commonly, the matrices \mathbf{A} and \mathbf{S} are stored in a compressed way. The preconditioner steps are available by operations on the fly making use of the stored system matrices. This may become inefficient for the coupled FE-SE matrices which are substantially more dense than their pure FE counterparts. The use of properly located collocation points allows for the use of FFTs. Then, only the operations $\mathbf{K} + \mathbf{M} \frac{d}{dt}$, $\tilde{\mathbf{L}}^{-1}$ and \mathbf{Z} are carried out using a sparse storage system. \mathbf{Q} corresponds to selecting a subset of the DoFs of \mathbf{v} using an index set. Now, the matrices \mathbf{A} and \mathbf{S} are no longer given explicitly and pure algebraic preconditioning techniques such as e.g. IC, are no longer available. The preconditioners $\tilde{\mathbf{A}}_1^{-1}$, $\tilde{\mathbf{A}}_2^{-1}$, $\tilde{\mathbf{L}}^{-1}$ and $\tilde{\mathbf{S}}_2^{-1}$, however, remain available.

The expected computational effort of a pure FE model is of order $\mathcal{O}(n^3)$, where n is a measure for the number of DoFs counted along one spatial direction. This complexity is related to the application of \mathbf{K} , \mathbf{M} and $\tilde{\mathbf{L}}^{-1}$. The application of 2D FFTs scales like $\mathcal{O}(n^2 \ln n)$ which is asymptotically below $\mathcal{O}(n^3)$. Hence, the complexity of the coupled FE-SE model is expected to be comparable to the complexity of a pure FE model. This is no longer true when the coupling matrices \mathbf{B} and spectral system matrix \mathbf{G} are represented by algebraic matrices. The full population of \mathbf{B} and \mathbf{G} causes their application to scale by $\mathcal{O}(n^4)$, which would dominate the overall calculation.

Numerical Experiments

The different iterative solution techniques are compared for a 2D quarter model of a superconductive dipole magnet. The iron yoke is substantially saturated. The yoke and the windings are discretized by FEs where the aperture is discretized by the spectral basis functions $M_q(r, \theta) = T_{q1} \left(\frac{r}{R} \right) e^{-j\lambda\theta}$. The SE technique supports an easy calculation of the harmonic field coefficients that determine the

Table 1: Computation times (in seconds) for the proposed iterative solution methods for the mixed FE-SE system and the Schur complement system using the compressed row storage (CRS) form or the matrix-free solution techniques; comparison with respect to a homogeneous FE model used as a reference (indicated by the subscript "ref"); approximation to FE system matrices are created using the IC technique without fill-in.

model	system	solver	precond.	CRS	matrix-free
FE	\mathbf{L}_{ref}	CG	$\tilde{\mathbf{L}}_{\text{ref}}^{-1}$	27.23	-
FE-SE	\mathbf{A}	MINRES	$\tilde{\mathbf{A}}_1^{-1}$	9.06	11.02
FE-SE	\mathbf{A}	MINRES	$\tilde{\mathbf{A}}_2^{-1}$	8.55	7.56
FE-SE	\mathbf{S}	CG	$\tilde{\mathbf{S}}_1^{-1}$	78.34	5.58
FE-SE	\mathbf{S}	CG	$\tilde{\mathbf{S}}_2^{-1}$	55.34	3.25

quality of the accelerator magnet.

As an approximation to the FE system part, an IC factorization without fill-in is used. The performance of the proposed solution techniques is compared in Table 1, both in the case where the system is assembled into a compressed row storage (CRS) format and used in matrix form as in the case where matrix-free techniques and FFTs are applied. The experiments clearly illustrate the computational efficiency of the matrix-free techniques. The FFTs solving the SE subproblem are responsible for less than 10% of the overall computation time. The experiments show that solving the positive semi-definite Schur complement system is more efficient than solving the mixed system, as long as matrix-free techniques are used. The numerical tests indicate the advantage of a additive-Schwarz-type preconditioner over a preconditioner only accounting for the FE part. When carried out using matrix-free techniques, all hybrid FE-SE discretizations outperform the model with a pure FE discretization attaining a comparable spatial resolution.

BEAM-TUBE MODEL

Beam-Tube Modeling

The beam tube in an accelerator magnet consists of a thin pipe with a circular or elliptical cross-section. The beam tube should support the inner vacuum and is commonly made of conductive material in order to carry a blind current. During the ramping of the magnet, eddy currents are generated in the beam tube. These currents cause additional losses and may deteriorate the quality of the aperture field. The calculation of these eddy currents necessitates transient simulation, which is time consuming, particularly in combination with 3D FE models [5]. Explicitly resolving the beam-tube thickness is commonly avoided in order to restrict the model size. Moreover, as accelerator magnets exhibit a translational symmetry, relevant simulation results can already be generated on the basis of 2D FE models. A 2D model is not capable of modeling the closing paths of the beam-tube eddy currents at the front and back

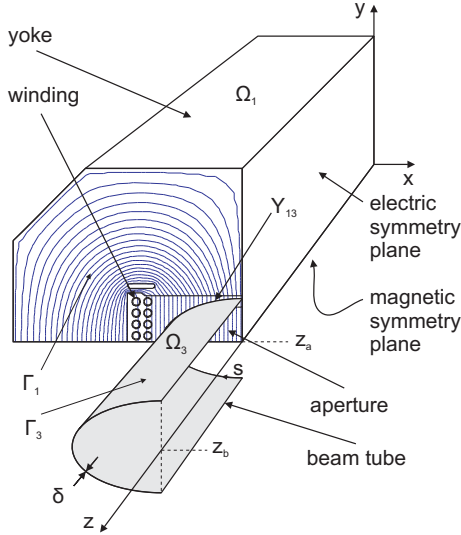


Figure 2: Accelerator magnet with beam tube: computational domain and magnetic flux lines.

side of the accelerator magnet. These closing paths are expected to have a significant influence on the distribution of eddy currents along the circumference of the beam-tube cross-section. This section is dedicated to the addition of a model part accounting for these closing paths.

2D FE Model and Thin-Sheet Approximation

The central part Ω_1 of the accelerator magnet is translationally symmetric along the z -direction: $\Omega_1 = \Gamma_1 \times \ell_z$ where Γ_1 is the cross-section and ℓ_z is the length along the z -direction (Fig. 2). The translational symmetry of Ω_1 is reflected in the choice of the edge shape functions \vec{w}_j for discretizing the magnetic vector potential \vec{A} :

$$\vec{w}_j = \frac{N_j(x, y)}{\ell_z} \vec{e}_z \quad (16)$$

where $N_j(x, y)$ are lowest-order nodal shape functions defined at a triangularization of Γ_1 and \vec{e}_z is a unit vector in the z -direction.

The thickness δ of the beam tube is much smaller than the skin depth

$$\delta_{\text{skin}} = \sqrt{\frac{2}{\omega \sigma_{\text{tube}} \mu_{\text{tube}}}} \quad (17)$$

related to the beam-tube conductivity σ_{tube} and permeability μ_{tube} and expected for the relevant angular frequencies ω . This allows to assume a current density which is homogeneously distributed along the beam-tube thickness. The cross-section of the beam tube with Γ_1 can be represented by $\Upsilon_{\text{tube}} \times [-\frac{\delta}{2}, \frac{\delta}{2}]$ where Υ_{tube} is a contour in Γ_1 (Fig. 2). The beam-tube cross-section is not explicitly resolved by the mesh. The eddy currents in the beam tube are included by augmenting the conductance matrix \mathbf{M}_σ with the line

Superconducting Magnets

integrals

$$\mathbf{M}_{\sigma, ij}^{\text{thin}} = \int_{\Upsilon_{\text{tube}}} \sigma_{\text{tube}} \vec{w}_j \cdot \vec{w}_i \delta \ell_z \, ds \quad (18)$$

where σ_{tube} is the conductivity of the beam-tube material. This approach corresponds to the most simple form of a thin-sheet model [9, 7]. In this paper, this 2D thin-sheet model is accomplished by an additional treatment for modeling the closing paths of the beam-tube current.

End Effects

End effects are the effects due to the broken translational symmetry at the front and rear sides of the model. In the center of the magnet, the beam-tube eddy currents are perpendicular to the magnet cross-section. At the end parts, some of the current lines may proceed to the next magnet and other current lines may close in a more or less circular path on the surface of the beam tube. The fact that the different closing paths feature a different impedance, disturbs the translational symmetry of the model. The constructed 2D FE model considers a perfectly conductive connection at the front and rear end. Several approaches exist to account for the additional impedances of the closing paths, while sticking to a 2D model. A common approximation is to consider the beam-tube parts exceeding the magnetically active part by decreasing the conductivity of the beam-tube material, i.e.,

$$\sigma_{\text{tube}} \rightarrow \sigma_{\text{tube}} \frac{\ell_z + \ell_{\text{end}}}{\ell_z} \quad (19)$$

where ℓ_{end} denotes the length of the outside beam-tube parts. This approach does not account for the considerably longer closing paths of the beam-tube currents at the magnetic symmetry plane compared to those at the electric symmetry plane (Fig. 2).

Beam-Tube End Model

Besides the magnetically active model part $\Omega_1 = \Gamma_1 \otimes [0, z_a]$, an additional domain $\Omega_3 = \Gamma_3 \otimes [-\frac{\delta}{2}, +\frac{\delta}{2}]$ is considered (Fig. 2). Here, $\Gamma_3 = \Upsilon_{\text{tube}} \otimes [z_a, z_b]$ is the beam-tube surface outside the active magnet part. By construction, the intersection of both domain parts is $\Omega_1 \cap \Omega_3 = \Upsilon_{\text{tube}} \otimes [-\frac{\delta}{2}, +\frac{\delta}{2}]$. The coupled formulation reads

$$\nabla \times (\nu \nabla \times \vec{A}) + \sigma \frac{\partial \vec{A}}{\partial t} + \sigma \nabla \varphi = \vec{J}_s \quad \text{in } \Omega_1 \quad (20)$$

$$-\nabla \cdot \left(\sigma \frac{\partial \vec{A}}{\partial t} \right) - \nabla \cdot (\sigma \nabla \varphi) = 0 \quad \text{in } \Omega_3 \quad (21)$$

where φ is the electric scalar potential. In Ω_1 , the electromagnetic field is computed by (20), incorporating both inductive and resistive field effects. In Ω_3 , the inductive effects are neglected with respect to the resistive effects. This allows to use the stationary-current formulation.

Discretization

The applied FE shape functions are

$$\begin{aligned} \vec{w}_j(x, y, z) &= \frac{N_j(x, y)}{z_a} \vec{e}_z & \text{in } \Omega_1; \\ P_{\tilde{q}}(s, z) &= M_{\tilde{q}}(s, z_a) \frac{z}{z_1} & \text{in } \Omega_1; \\ P_q(s, z) &= M_q(s, z) & \text{in } \Omega_2 \end{aligned} \quad (22)$$

where $N_j(x, y)$ and $M_q(s, z)$ are nodal shape functions defined on triangulations of Γ_1 and Γ_3 , respectively (Fig. 2). The indices $i, j \in \mathcal{I}_1$, $p, q \in \mathcal{I}_3$ and $\tilde{p}, \tilde{q} \in \mathcal{I}_1 \cap \mathcal{I}_3$ and the index sets \mathcal{I}_1 , \mathcal{I}_3 and $\mathcal{I}_1 \cap \mathcal{I}_3$ refer to the nodes of Γ_1 , Γ_3 and $\Upsilon_{\text{tube}} = \Gamma_1 \cap \Gamma_3$, respectively. The potentials are

$$\begin{cases} \vec{A} = \sum_j u_j \vec{w}_j \\ \varphi = \sum_{\tilde{q}} v_{\tilde{q}} P_{\tilde{q}} \end{cases} \quad \text{in } \Omega_1; \quad (23)$$

$$\begin{cases} \vec{A} = 0 \\ \varphi = \sum_q v_q P_q \end{cases} \quad \text{in } \Omega_3, \quad (24)$$

where the DoFs are gathered in the algebraic vectors \mathbf{u} and \mathbf{v} . The DoFs $v_{\tilde{q}}$ are a subset of all DoFs v_q . They share the shape functions $P_{\tilde{q}}$ defined in Ω_1 and P_q defined in Ω_3 and therefore enforce the continuity of φ at Υ_{tube} .

The FE shape functions serve for discretizing (20) and (21) at Ω_1 and Ω_3 , respectively. Eq. (20) is weighted by \vec{w}_i whereas (21) is weighted by the functions P_q . The resulting coupled system of equations reads

$$\begin{bmatrix} \mathbf{K} + \mathbf{M} \frac{d}{dt} & \mathbf{B}^T \\ \mathbf{B} \frac{d}{dt} & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \quad (25)$$

where

$$\mathbf{K}_{i,j} = \int_{\Gamma_1} \left(\frac{\nu_y}{z_a} \frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + \frac{\nu_x}{z_a} \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} \right) dA; \quad (26)$$

$$\mathbf{M}_{i,j} = \int_{\Gamma_1} \frac{\sigma}{z_a} N_i N_j dA; \quad (27)$$

$$\mathbf{f}_i = \int_{\Gamma_1} J_{s,z} N_i dA; \quad (28)$$

$$\mathbf{B}_{\tilde{p},j} = \int_{\Upsilon_{\text{tube}}} \frac{\sigma}{z_a} M_{\tilde{p}} N_j \delta ds; \quad (29)$$

$$\mathbf{G} = \int_{\Gamma_3} \sigma \nabla_{\Gamma_3} M_p \cdot \nabla_{\Gamma_3} M_q \delta dA. \quad (30)$$

Here, ∇_{Γ_3} denotes a gradient operator accounting for the curvature of Γ_3 . This system is symmetric and positive definite and can be solved by a preconditioned CG method.

Also here, it is possible to eliminate the additional degrees of freedom \mathbf{v} related to φ from the system. The resulting Schur complement system reads

$$\mathbf{K} \mathbf{u} + (\mathbf{M} - \mathbf{B}^T \mathbf{G}^{-1} \mathbf{B}) \frac{d\mathbf{u}}{dt} = \mathbf{f}. \quad (31)$$

The matrix $\mathbf{M}_{\text{schur}} = \mathbf{M} - \mathbf{B}^T \mathbf{G}^{-1} \mathbf{B}$ can be interpreted as a modified FE conductance matrix accounting for the closing paths outside of the 2D cross-sectional model. It is easily shown that $\mathbf{M}_{\text{schur}}$ is symmetric and positive semi-definite, similar to \mathbf{M} . Solving (30) can be done by a preconditioned CG method, but may be computationally expensive because of the dense blocks $\mathbf{B}^T \mathbf{G}^{-1} \mathbf{B}$.

Superconducting Magnets

SIMULATION RESULTS

The additional aperture and beam-tube end model parts are applied in combination with a transient simulation of the SIS-100 superconductive magnet which is part of the Facility for Antiproton and Ion Research (FAIR) project of the Helmholtzzentrum für Schwerionenforschung (GSI) in Darmstadt, Germany [6]. The eddy-current losses in the yoke have already been reported in e.g. [4]. 3D simulations for the beam-tube losses have been studied in [5]. This paper is concerned with 2D simulations providing a comparable accuracy at a significant lower computational costs.

The elliptical cross-section of the beam-tube domain is parametrized by (ξ, η, z) , i.e., $(x, y, z) = (\rho \cosh \xi \cos \eta, \rho \sinh \xi \sin \eta, z)$. The corresponding Laplace-Beltrami operator is

$$\nabla_{\Omega_3}^2 = \frac{1}{\rho^2 (\cosh^2 \xi - \cos^2 \eta)} \frac{\partial^2 \varphi}{\partial \eta^2} + \frac{\partial^2 \varphi}{\partial z^2}. \quad (32)$$

The discretization of the model is carried out as described above. The duty cycle of magnet operations is shown in Fig. 3a. During the ramping of the aperture field, eddy currents are induced in the beam tube. The final mesh obtained after a few adaptive mesh refinement steps as well as the magnetic flux lines are shown in Fig. 3d. Fig. 3e and Fig. 3f compare the equipotential lines for the electric scalar potential φ and the arrows for the electric field strength for the connected and unconnected cases, respectively. The figures correspond to the coordinate system (ξ, z) and show a flat projection of the beam-tube surface seen from above. The top side of the figure corresponds to the interface Υ_{tube} between both model parts. The left and right sides fall together with the magnetic symmetry plane. The vertical center line falls together with the electric symmetry plane. The bottom side of the figure corresponds to the connection of the beam tube to the next-in-line accelerator component. The closing paths of the beam-tube currents outside the magnetically active part depend on the fact whether an electric connection exists to the following component or not (compare Fig. 3e to Fig. 3f). The influence of this connection seems to be negligible. Nevertheless, the difference between the results for the model with perfectly conductive front and rear planes and the results for the model with the additional beam-tube end model, are significant, which indicates the relevance of the beam-tube end model. The double 2D model is substantially more efficient than a full 3D model. Therefore, the computational resources can be spent to achieve a finer resolution in the cross-sectional plane, e.g. to tackle the severe ferromagnetic saturation of the magnet yoke (Fig. 3d). The beam-tube losses for the duty cycle of Fig. 3a are shown in Fig. 3b. The dependence of the losses on time is mostly constant because the aperture field is ramped at a constant rate. Due to the saturation in the middle of the duty cycle, the magnetic flux density does not increase further at linear rate, which explains the slightly smaller beam-tube losses.

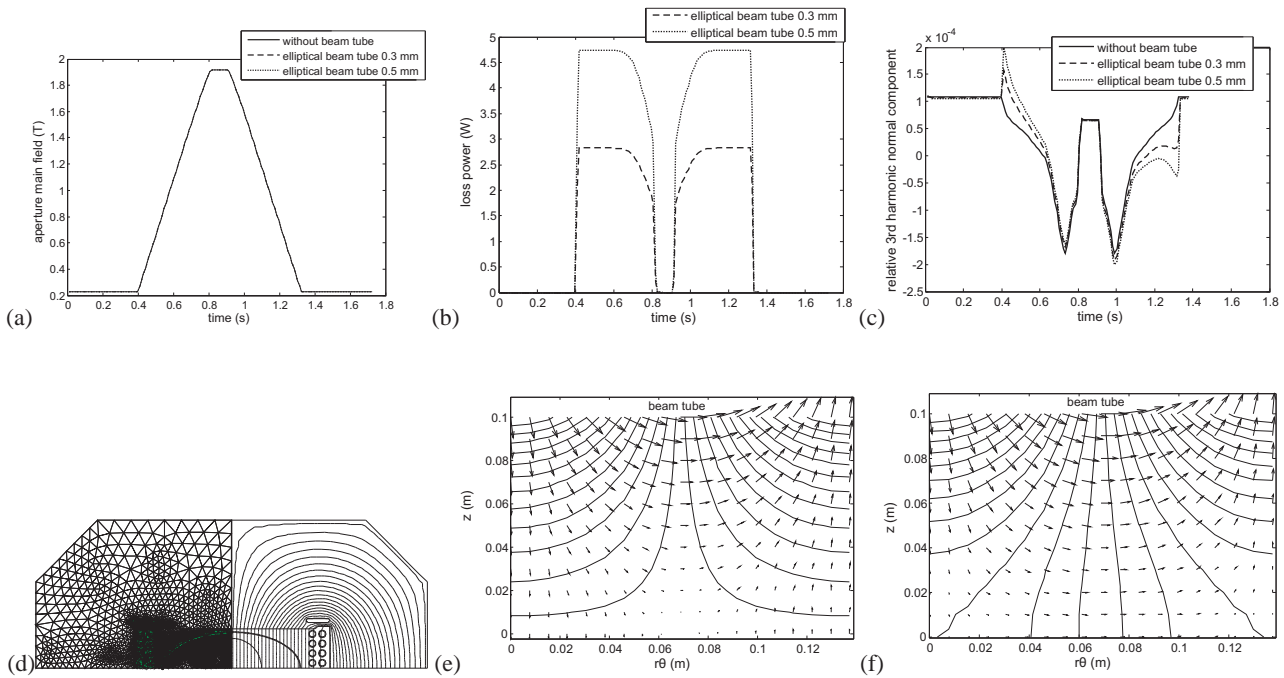


Figure 3: Transient simulation results for the SIS-100 magnet: (a) magnetic flux density in the aperture; (b) power loss; (c) sextupole field component; (d) mesh and magnetic flux lines at maximal aperture field; (e-f) distribution of the current on the beam-tube end surface with (e) and without (f) electric connection to the next magnet.

The beam-tube eddy currents have a disadvantageous effect on the quality of the aperture field as is clear for the sextupole component from Fig. 3c.

CONCLUSION

A cheap transient 2D FE model for accelerator magnets becomes valuable by increasing the resolution of the aperture by a SE discretization and by including end effects in the beam tube by an additional 2D model. The relevance of this technique is given by the fact that the overall transient simulation only takes a few minutes of calculation time and, hence, can be inserted in a flexible design process.

REFERENCES

- [1] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer Verlag, New York, 1996.
- [2] H. De Gersem and K. Hameyer. Harmonic boundary conditions for circular inclusions and their coupling to external circuits. *IEE Proc.-Sci. Meas. Technol.*, 148(6):257–262, November 2001.
- [3] T. Hemker, Oskar von Stryk, H. De Gersem, and T. Weiland. Mixed-integer nonlinear design optimization of a superconductive magnet with surrogate functions. *IEEE Trans. Magn.*, 44(6):1110–1113, June 2008.
- [4] S. Koch, H. De Gersem, T. Weiland, E. Fischer, and G. Moritz. Transient 3D finite element simulations of the SIS100 magnet considering anisotropic, nonlinear material models for the ferromagnetic yoke. *IEEE Trans. Appl. Superconduct.*, 18(2):1601–1604, June 2008.
- [5] S. Koch, J. Trommler, H. De Gersem, and T. Weiland. Modeling thin conductive sheets using shell elements in magnetoquasistatic field simulations. *IEEE Trans. Magn.*, 45(3):1292–1295, March 2009.
- [6] A.D. Kovalenko, A. Kalimov, H.G. Khodzhbagiyani, G. Moritz, and C. Mühle. Optimization of a superferic nucletron type dipole for the GSI fast pulsed synchrotron. *IEEE Trans. Appl. Superconduct.*, 12(1):161–165, 2002.
- [7] L. Krähenbühl and D. Muller. Thin layers in electrical engineering. example of shell models in analyzing eddy-currents by boundary and finite element methods. *IEEE Trans. Magn.*, 29(5):1450–1455, 1993.
- [8] S. Kurz, S. Russenschuck, and N. Siegel. Accurate calculation of fringe fields in the LHC main dipoles. *IEEE Trans. Appl. Superconduct.*, 10(1):85–88, March 2000.
- [9] T. Nakata, N. Takahashi, K. Fujiwara, and Y. Shiraki. 3-D magnetic field analysis using special elements. *IEEE Trans. Magn.*, 26(5):2379–2381, September 1990.
- [10] C.C. Paige and M.A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975.
- [11] P.P. Silvester and R.L. Ferrari. *Finite Elements for Electrical Engineers*. Cambridge University Press, Cambridge, 2nd edition, 1996.
- [12] B.F. Smith, P.E. Bjørstad, and W.D. Gropp. *Domain Decomposition: parallel multilevel methods for elliptic partial differential equations*. Cambridge University Press, Cambridge, 1996.
- [13] L.N. Trefethen. *Spectral Methods in Matlab*. SIAM, Philadelphia, 2000.

A FAST AND UNIVERSAL VLASOV SOLVER FOR BEAM DYNAMICS SIMULATIONS IN 3D*

Sylvain Franke**, Wolfgang Ackermann, Thomas Weiland,
Technische Universität Darmstadt, Institut für Theorie Elektromagnetischer Felder,
Schlossgartenstraße 8, 64289 Darmstadt, Germany

Abstract

The Vlasov equation can describe the evolution of a particle density under the effects of electromagnetic fields and thus it is possible to describe the evolution of a charged particle beam within an accelerator beam line. The Vlasov equation forms a partial differential equation in a 6D phase space which renders it very expensive if it is solved via classical methods. A more efficient approach consists in representing the particle distribution function by a discrete set of characteristic moments. For each moment a time evolution equation can be stated. These ordinary differential equations can then be integrated efficiently by means of numerical methods if all acting forces together with a proper initial condition are given. The beam dynamics simulation tool V-Code has been implemented at TEMF on the basis of the moment approach. In this paper the numerical model, main features and designated use cases of the V-Code will be presented.

INTRODUCTION

The distribution of particles in the 6-dimensional (6D) phase space can be described by a density distribution function $f(\vec{r}, \vec{p}, \tau)$ with space coordinates $\vec{r} = (x, y, z)$, normalized momentum $\vec{p} = (p_x, p_y, p_z)$ and equivalent time $\tau = c \cdot t$. Their evolution in the phase space can then be expressed by the Vlasov equation

$$\frac{\partial f}{\partial \tau} + \frac{\partial f}{\partial \vec{r}} \cdot \frac{\vec{p}}{\gamma} + \frac{\partial f}{\partial \vec{p}} \cdot \frac{\vec{F}}{m_0 c^2} = 0 \quad (1)$$

where γ represents the relativistic factor, \vec{F} the applied forces, m_0 the particles rest mass and c the speed of light in free space.

Equation (1) is applicable for any forces \vec{F} with slow variation in space [2]. Coulomb forces within an charged particle beam as well as forces from external electromagnetic fields meet this condition. Thus, the Vlasov equation is applicable for beam dynamics simulations of charged particle beams in accelerators.

It is very expensive to solve such a partial differential equation via classic numerical methods for a time varying 6D density distribution function.

MOMENT APPROACH

A more efficient approach is to consider a discrete set of characteristic moments of the particle distribution function instead of the function itself [3]. Following this approach the problem can be reduced to a set of ordinary differential equations which can be evaluated by means of standard time integration methods.

Moment Definition

The classical raw moments $\langle \mu \rangle$ are obtained from the distribution function f by a weighted integration over the whole phase space Ω

$$\langle \mu \rangle = \int_{\Omega} \mu f(\vec{r}, \vec{p}, \tau) d\Omega. \quad (2)$$

Here, the normalized density distribution function to ensure

$$1 \stackrel{!}{=} \langle 1 \rangle = \int_{\Omega} f(\vec{r}) d\Omega \quad (3)$$

has to be applied for proper algebraic relations.

A numerically advantageous choice of moments which ultimately allows the determination of the overall position and the overall momentum of a particle distribution is given by the first order raw moments

$$\mu \in \{x, y, z, p_x, p_y, p_z\} \quad (4)$$

in Cartesian coordinates.

By choosing the higher order moments in a centralized notation

$$\mu \in \{(x - \langle x \rangle)^{l_1} \cdot \dots \cdot (p_z - \langle p_z \rangle)^{l_6}, \dots\} \quad (5)$$

one automatically obtains a translatory invariant description of the shape of the particle distribution function.

For example, a subset of the second order moments

$$\begin{aligned} \sigma_x^2 &= \langle (x - \langle x \rangle)^2 \rangle \\ \sigma_y^2 &= \langle (y - \langle y \rangle)^2 \rangle \\ \sigma_z^2 &= \langle (z - \langle z \rangle)^2 \rangle \end{aligned}$$

then identify the important variances of the underlying particle distribution.

* Work supported by DFG through SFB 634.

**franke@temf.tu-darmstadt.de

Moment Evolution

The time evolution of the moment parameters can be expressed as follows:

$$\frac{\partial \langle \mu \rangle}{\partial \tau} = \frac{\partial}{\partial \tau} \int \mu f \, d\Omega = \int (f \frac{\partial \mu}{\partial \tau} + \mu \frac{\partial f}{\partial \tau}) \, d\Omega. \quad (6)$$

Here, the Vlasov equation can be used to completely eliminate the time derivative of the distribution function and substitute it with a spatial expression instead:

$$\frac{\partial f}{\partial \tau} = -(\frac{\partial f}{\partial \vec{r}} \cdot \frac{\vec{p}}{\gamma} + \frac{\partial f}{\partial \vec{p}} \cdot \frac{\vec{F}}{m_0 c^2}). \quad (7)$$

Further, the time derivative of the moment parameter can be reformulated like

$$\frac{\partial \mu}{\partial \tau} = \frac{\partial \mu}{\partial \langle \vec{r} \rangle} \frac{\partial \langle \vec{r} \rangle}{\partial \tau} + \frac{\partial \mu}{\partial \langle \vec{p} \rangle} \frac{\partial \langle \vec{p} \rangle}{\partial \tau}. \quad (8)$$

Inserting (7) and (8) in equation (6) and applying a partial integration together with the moment definition (2) allows to state the following fundamental time evolution equation

$$\begin{aligned} \frac{\partial \langle \mu \rangle}{\partial \tau} = & \langle \frac{\partial \mu}{\partial \langle \vec{r} \rangle} \rangle \langle \frac{\vec{p}}{\gamma} \rangle + \langle \frac{\partial \mu}{\partial \langle \vec{p} \rangle} \rangle \langle \frac{\vec{F}}{m_0 c^2} \rangle \\ & + \langle \frac{\partial \mu}{\partial \vec{r}} \frac{\vec{p}}{\gamma} \rangle + \langle \frac{\partial \mu}{\partial \vec{p}} \frac{\vec{F}}{m_0 c^2} \rangle \end{aligned} \quad (9)$$

In order to enable the evaluation of (9) by means of a time integration method all arguments on the right hand side have to be expressed in terms of time dependent bunch parameters i.e. by moments of the settled form. This can be achieved by Taylor expanding $\frac{1}{\gamma}$ and \vec{F} in an operation point defined by the particle distribution and utilizing a truncation according to the regarded order of moments.

If the energy spread is small compared to the mean energy of the whole bunch a linear approximation of γ is adequate

$$\gamma = \sqrt{1 + p_x^2 + p_y^2 + p_z^2} \quad (10)$$

but in general higher order forms are possible.

The series expansion for internal space charge forces and forces due to external fields can be performed independently. External fields are observed in radio frequency cavities (RF-cavities) or any kind of magnetic multipole for example.

External Field Representation

From the Maxwell equations in frequency domain and isotropic homogeneous linear media one can describe the fields within RF-cavities in cylindrical coordinates:

$$\begin{aligned} \underline{E}_r &= \frac{-1}{j\omega\mu\epsilon} \frac{\partial \underline{B}_\varphi}{\partial z}, & \underline{B}_r &= \frac{1}{j\omega} \frac{\partial \underline{E}_\varphi}{\partial z} \\ \underline{E}_\varphi &= \frac{1}{j\omega\mu\epsilon} (\frac{\partial \underline{B}_r}{\partial z} - \frac{\partial \underline{B}_z}{\partial r}), & \underline{B}_\varphi &= \frac{-1}{j\omega} (\frac{\partial \underline{E}_r}{\partial z} - \frac{\partial \underline{E}_z}{\partial r}) \\ \underline{E}_z &= \frac{1}{j\omega\mu\epsilon} \frac{1}{r} \frac{\partial (r \underline{B}_\varphi)}{\partial r}, & \underline{B}_z &= \frac{-1}{j\omega} \frac{1}{r} \frac{\partial (r \underline{E}_\varphi)}{\partial r} \end{aligned}.$$

Computer Codes (Design, Simulation, Field Calculation)

In this formulation the field components \underline{B}_r , \underline{E}_φ , \underline{B}_z are completely decoupled from \underline{E}_r , \underline{B}_φ , \underline{E}_z . Within accelerating cavities it is appropriate to consider the latter ones exclusively.

By defining $\vec{B} = \text{curl} \vec{A}$ and applying Bernoulli's separation approach

$$\vec{A}(r, z) = \vec{e}_z \underline{A}_z(r, z) = \vec{e}_z \sum_{i=0}^{\infty} \underline{A}_i(z) r^i \quad (11)$$

it is possible to describe \underline{E}_r , \underline{B}_φ , \underline{E}_z solely with derivatives of the z-component of the vector potential \vec{A}

$$\underline{B}_\varphi = -2\underline{A}_2 \cdot r - 4\underline{A}_4 \cdot r^3 - 6\underline{A}_6 \cdot r^5 - \dots \quad (12)$$

$$\underline{E}_r = \frac{1}{j\omega\mu\epsilon} (2\underline{A}'_2 \cdot r + 4\underline{A}'_4 \cdot r^3 + 6\underline{A}'_6 \cdot r^5 + \dots) \quad (13)$$

$$\underline{E}_z = \frac{-1}{j\omega\mu\epsilon} (4\underline{A}_2 + 16\underline{A}_4 \cdot r^2 + 36\underline{A}_6 \cdot r^4 + \dots). \quad (14)$$

Evaluation on the axis ($r = 0$) results in

$$\underline{E}_z|_{r=0} = -\frac{4 \cdot \underline{A}_2}{j\omega\mu\epsilon}, \quad (15)$$

which allows to determine the coefficients consecutively

$$\underline{A}_2 = -\frac{1}{4} j\omega\mu\epsilon \cdot \underline{E}_z|_{r=0} \quad (16)$$

$$\underline{A}_4 = \frac{-1}{16} \cdot (\underline{A}''_2 + \omega^2 \mu\epsilon \underline{A}_2) \quad (17)$$

$$\underline{A}_6 = \frac{-1}{36} \cdot (\underline{A}''_4 + \omega^2 \mu\epsilon \underline{A}_4) \quad (18)$$

$$\vdots$$

By successively inserting the coefficients in (12) - (14) one obtains the field components in cylindrical coordinates in terms of a series expansion

$$\underline{B}_\varphi = \frac{1}{2} j\omega\mu\epsilon \underline{E}_{z0} r - \frac{1}{16} j\omega\mu\epsilon (\underline{E}''_{z0} + \omega^2 \mu\epsilon \underline{E}_{z0}) r^3 + \dots \quad (19)$$

$$\underline{E}_r = -\frac{1}{2} \underline{E}'_{z0} r + \frac{1}{16} (\underline{E}'''_{z0} + \omega^2 \mu\epsilon \underline{E}'_{z0}) r^3 + \dots \quad (20)$$

$$\underline{E}_z = \underline{E}_{z0} - \frac{1}{4} (\underline{E}''_{z0} + \omega^2 \mu\epsilon \underline{E}_{z0}) r^2 + \dots \quad (21)$$

The function of this expansion constitute the paraxial approximation of the field distribution within a RF-cavity.

For multipole magnets a similar approach to approximate the three dimensional field distribution inside the source free vacuum tube is possible. The Maxwell equations for magnetostatic problems in source free vacuum areas can be reduced to the following equations:

$$\text{curl } \vec{B} = 0 \quad (22)$$

$$\text{div } \vec{B} = 0. \quad (23)$$

This allows to state the scalar potential equation

$$\Delta V = \text{div grad } V = 0 \quad \text{with} \quad \vec{B} = -\text{grad } V. \quad (24)$$

Applying Bernoulli's separation approach together with a series expansion for magnetic $2n$ poles leads to

$$B_r = (r^{n-1} \alpha(z) - \frac{(n+2) r^{n+1} \alpha''(z)}{4(n+1)n} + \dots) \cdot \sin(n\varphi) \quad (25)$$

$$B_\varphi = (r^{n-1} \alpha(z) - \frac{r^{n+1} \alpha''(z)}{4(n+1)} + \dots) \cdot \cos(n\varphi) \quad (26)$$

$$B_z = (\frac{r^n \alpha'(z)}{n} - \frac{r^{n+2} \alpha'''(z)}{4(n+1)n} + \dots) \cdot \sin(n\varphi) \quad (27)$$

with the multipole strength α along the longitudinal axis z .

Force Calculation

The forces resulting from the external fields can now be calculated by applying the Lorentz equation

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}). \quad (28)$$

As stated before the terms $\langle \vec{F} \rangle$ and $\langle \mu \vec{F} \rangle$ in (9) have to be expanded such that the resulting expression can be used to setup a moment description form.

A multipole expansion splits the force naturally in constant, linear, quadratic, etc. terms. By applying the moment definition (2) they can be expressed as follows:

- For constant forces $F_\nu = a$ one gets $\langle F_\nu \rangle = a$ and $\langle \mu F_\nu \rangle = 0$.
- For linear forces $F_\nu = b\mu$ one gets $\langle F_\nu \rangle = 0$ and $\langle \mu F_\nu \rangle = bM_{\mu\mu}$.
- For quadratic forces $F_\nu = c\mu^2$ one gets $\langle F_\nu \rangle = cM_{\mu\mu}$ and $\langle \mu F_\nu \rangle = cM_{\mu\mu\mu}$.

Hence, the order of the moments has to be at least one order higher than that of the significant terms of the multipole expansion.

Space Charge Forces

In order to determine the space charge forces a model to reconstruct the charge distribution within the particle bunch from the moment description is needed. The simplest approach is to assume a homogeneously charged ellipsoidal bunch. Starting from the force acting between two particles moving together with constant speed one gets the space charge force $F(\vec{r})$ in an arbitrary point of observation \vec{r} within the bunch by integrating over the bunch volume. In order to avoid the singularities of the integrand in the proximity of the point of observation one can omit a symmetric area surrounding this point when integrating. Within this area the space charge forces compensate themselves. Following this approach one gets

$$\vec{F}(\vec{r}) \approx G \left(\frac{2\gamma\sigma_z}{\sigma_x + \sigma_y} \right) \cdot \frac{eQ_0}{\gamma^2} \cdot \frac{\vec{r} - \langle \vec{r} \rangle}{V_G} \quad (29)$$

Computer Codes (Design, Simulation, Field Calculation)

with

$$G(u) = (1 - \exp(-u)) \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \gamma/u \end{pmatrix}. \quad (30)$$

The moment approach is not limited to this linear space charge model. The particle density reconstruction can also be done by other techniques like moment matching or the maximum entropy method.

Time Integration

The time evolution of all moments $\langle \mu \rangle$ is specified by the stated set of fundamental differential equations (9). By defining a single comprehensive vector $\vec{\psi}$ which includes all considered moments one can state the new time dependent variable $\vec{\psi}(t)$. The whole physical model can then be written in the standard mathematical form

$$\vec{\psi}'(t) = \vec{\phi}(t, \vec{\psi}(t)), \quad \vec{\psi}(t_0) = \vec{\psi}_0, \quad (31)$$

where $\vec{\psi}'$ denotes the time derivative and $\vec{\phi}$ summarizes all kinematic and kinetic effects of the forces mentioned above. If an initial set of moments $\vec{\psi}_0$ is provided the problem can be solved by standard time integration methods [1].

Multi Ensembles

In order to reproduce more complex particle interactions a multi ensemble model is possible. In this approach a particle distribution is represented by several sets of moments each of them defining an ensemble as shown in Fig. 1.

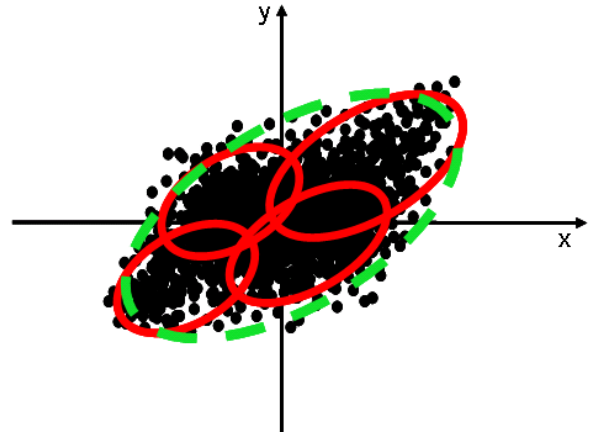


Figure 1: A 2D example of a multi ensemble setup. The red ellipsoids show four ensembles representing the particle distribution. The green dashed shape describes the set of moments obtained through weighted averages over the moments of the four single ensembles.

The time evolution of the several ensembles are calculated according to the procedure for single ensemble bunches described above. At any time a set of moments representing the whole bunch can be obtained from the moments of the

ensembles. The density distribution function for the whole bunch composed e.g. of two ensembles is defined as

$$f(\vec{r}) = f_1(\vec{r}) + f_2(\vec{r}). \quad (32)$$

This leads to the following equations for the charge Q :

$$Q = \underbrace{Q \int_{\Omega} f_1(\vec{r}) d\Omega}_{Q_1} + \underbrace{Q \int_{\Omega} f_2(\vec{r}) d\Omega}_{Q_2} \quad (33)$$

and defines the local density distribution functions via the relations

$$1 = \int_{\Omega} \frac{Q}{Q_1} f_1(\vec{r}) d\Omega \quad 1 = \int_{\Omega} \frac{Q}{Q_2} f_2(\vec{r}) d\Omega. \quad (34)$$

The first order moments of the whole bunch can then be obtained from the first order moments of the individual ensembles $\langle \mu \rangle_{f_1}$ and $\langle \mu \rangle_{f_2}$ as follows:

$$\langle \mu \rangle = \frac{Q_1}{Q} \langle \mu \rangle_{f_1} + \frac{Q_2}{Q} \langle \mu \rangle_{f_2}. \quad (35)$$

For the second order centralized moments

$$M_{uv} = \langle (u - \langle u \rangle) \cdot (v - \langle v \rangle) \rangle$$

with $u, v \in \{x, y, z, p_x, p_y, p_z\}$ one obtains

$$\begin{aligned} M_{uv} = & \frac{Q_1}{Q} \cdot M_{uv,f_1} + \frac{Q_1}{Q} (\langle u \rangle_{f_1} - \langle u \rangle) \cdot (\langle v \rangle_{f_1} - \langle v \rangle) \\ & + \frac{Q_2}{Q} \cdot M_{uv,f_2} + \frac{Q_2}{Q} (\langle u \rangle_{f_2} - \langle u \rangle) \cdot (\langle v \rangle_{f_2} - \langle v \rangle). \end{aligned} \quad (36)$$

Additional higher order moments can be calculated similarly.

APPLICATIONS

Based on the moment approach the fast online beam dynamics simulation tool V-Code was implemented at TEMF. The aim of several further developments was to increase the application range to various accelerator designs. For example dipole bending magnets were introduced in [6] in order to simulate recirculating machines. The V-Code can be utilized for the design phase as well as during the accelerator operation.

Beam Line Design and Optimization

At the Superconducting Darmstadt Linear ACcelerator S-DALINAC the V-Code has been used during the design process of the injector for the new 100 keV polarized electron source. Starting from an initial ensemble issued from a simulation of the electron source with the CST code MAFIA [4] the entire beam line was modeled within the V-Code and simulated as a whole. The results of detailed simulations can be found in [5].

Computer Codes (Design, Simulation, Field Calculation)

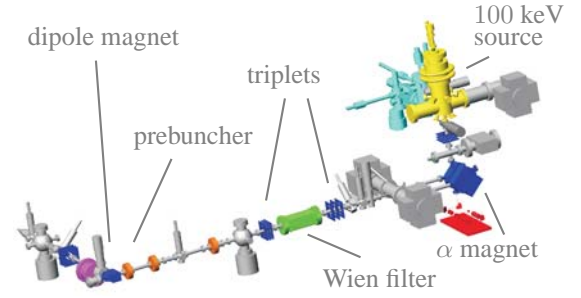


Figure 2: Schematic computational model of the compact injector design for the new polarized electron source at the S-DALINAC.

Operator Support

Fast beam dynamics simulations can advantageously assist the machine operators at various particle accelerator machines because of a flexible parameter variation combined with nearly simultaneous solution responses giving a detailed insight into the actual machine status. A user friendly front end was implemented for this purpose, giving the operator the possibility to comfortably change the parameter setup and receiving a well-arranged overview over the actual beam status.

Automatic Beam Adjustment

Finding an optimal parameter setup for an accelerator beam line is a tedious and laborious task as the number of variable parameters is typically very large. A fast simulation code opens the possibility to analyze a multitude of parameter sets in a reasonable time. By implementing objectives and valuation rules this procedure can be automated. Further, a connection between the simulation code and the accelerator diagnostic software allows to take into account measured parameters in the automated optimization process.

REFERENCES

- [1] W. Ackermann, T. Weiland: Efficient Time Integration for Beam Dynamics Simulations Based on the Moment Method, Proceedings of ICAP 2006, Chamonix Mont-Blanc, France, October 2006.
- [2] P.M. Bellan: Fundamentals of plasma physics, Chapter II, Cambridge University Press, 2006.
- [3] P.J. Channell: The Moment Approach to Charged Particle Beam Dynamics, *IEEE Trans. Nucl. Sci.*, NS-30, No. 4, Page 2607, 1983.
- [4] CST AG - Computer Simulation Technology, Bad Nauheimer Str. 19, D-64289 Darmstadt, Germany, <http://www.cst.com>.
- [5] B. Steiner: Strahldynamik-Simulation einer polarisierten Quelle für den S-DALINAC (SPIN), Dissertation, 2009.
- [6] S. Franke, W. Ackermann, B. Steiner, T. Weiland, J. Enders, C. Heßler, Y. Poltoratska: Implementation of Fringe Field Dipole Magnets into the V-Code Beam Dynamics Simulation Tool, Proceedings of the 11th European Particle Accelerator Conference (EPAC 2008), Genoa, Italy, June 2008.

DISCRETIZING TRANSIENT CURRENT DENSITIES IN THE MAXWELL EQUATIONS*

Mark L. Stowell, Daniel A. White,[†]

Lawrence Livermore National Laboratory, P.O. Box 808, Livermore CA, 94551, USA

Abstract

We will briefly discuss a technique for applying transient volumetric current sources in full-wave, time-domain electromagnetic simulations which avoids the need for divergence cleaning. The method involves both “edge-elements” and “face-elements” in conjunction with a particle-in-cell scheme to track the charge density. Results from a realistic, 6.7 million element, 3D simulation are shown. While the authors may have a finite element bias the technique should be applicable to finite difference methods as well.

INTRODUCTION

The Maxwell Equations

The Maxwell Equations with a current density source term can be written

$$\begin{aligned}\frac{\partial \vec{D}}{\partial t} &= \nabla \times \vec{H} - \vec{J} \\ \frac{\partial \vec{B}}{\partial t} &= -\nabla \times \vec{E}\end{aligned}$$

where

$$\vec{D} = \epsilon \vec{E} \text{ and } \vec{B} = \mu \vec{H}$$

These equations are commonly discretized using “edge-elements”, or discrete 1-forms, for the electric field and “face-elements”, or discrete 2-forms, for the magnetic flux density. This scheme requires that \vec{J} also be approximated with edge-elements, which works quite well in many situations. However, this scheme does have certain drawbacks.

One difficulty with 1-form current densities is that they can spread through material interfaces into non-physical regions. For example, consider a vacuum region abutting a weak conductor which contains a constant current density. What value for \vec{J} should be applied to the edges which are shared between these two regions? If the constant \vec{J} value is used, then the conducting region will contain the correct value but the vacuum region will also contain a non-zero current density. If a value of zero is applied on these edges, then the vacuum region will correctly have zero current but the conductor will contain less current density than desired.

Another difficulty, and the one we will focus on, arises if the current density is transient and the primary interest is to

determine how a cavity will resonate after a current pulse passes through it. The problem here is that the continuity equation for the electric charge,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{J} = 0,$$

is only weakly satisfied. Therefore, current densities can, and often do, leave behind non-physical charge densities after they pass through the computational mesh. These charge densities can, in turn, produce a non-physical, static, electric field which not only adds unexpected, mesh dependent, features to field plots but can also reduce the accuracy of the meaningful portion of the solution.

Integrating the charge continuity equation over time we obtain

$$\rho(t_b) - \rho(t_a) = - \int_{t_a}^{t_b} \nabla \cdot \vec{J} dt.$$

Assuming t_a and t_b are chosen such that \vec{J} is everywhere equal to zero before t_a and after t_b with no charges being left behind anywhere within the problem domain we would have $\rho(t_a) = \rho(t_b) = 0$ and so

$$\int_{t_a}^{t_b} \nabla \cdot \vec{J} dt = 0.$$

This is the constraint that we hope to satisfy. The accuracy with which this can be accomplished will hinge on our ability to accurately represent the divergence of the vector flux density \vec{J} .

Discrete Differential Forms

Differential Forms provide a general mathematical formalism to describe not only Div, Grad, and Curl but also integral relationships like the fundamental theorem of calculus, Kelvin-Stokes theorem, and the Divergence theorem. For example these three theorems can each be described by the generalized Stokes' theorem:

$$\int_{\Omega} d\omega = \oint_{\partial\Omega} \omega \quad (1)$$

Where “ ω ” is a differential form and “ d ” is the exterior derivative appropriate for that form type. Specifically, if ω is a 1-form (a standard vector field), this expression states that

$$\int_{\Sigma} \nabla \times \omega \cdot \vec{n} dA = \oint_{\partial\Sigma} \omega \cdot d\vec{r} \quad (2)$$

which is the classical Kelvin-Stokes theorem. Another important characteristic of differential forms is that for any k -form, ω , its exterior derivative, $d\omega$, is a $(k+1)$ -form.

* This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, UCRL LLNL-CONF-420323.

[†] stowell11@llnl.gov, white37@llnl.gov

Form Type	Field Type	Space	Meaning of DoF
0	Scalar	H(Grad)	Value at Point
1	Vector	H(Curl)	Path Integral Along Edge
2	Pseudo Vector	H(Div)	Surface Integral Over Face
3	Scalar Density	L_2	Volume Integral Over Cell

Table 1: This table lists the type of field that each form type is best suited for, the Hilbert space that the basis functions belong to, and how the degrees of freedom are computed.

Discrete Differential Forms [1][2], see Table 1, are finite element basis functions designed to satisfy the same differential and integral relations as their continuous counterparts. To achieve this the degrees of freedom of a k -form are associated with topological entities of dimension k : 0-forms are associated with nodes, 1-forms with edges, etc.. The value assigned to a degree of freedom (DoF) is the integral of the field over the corresponding entity: 0-form DoFs are point-wise evaluations of the field at the nodes, 1-form DoFs are line integrals along edges, etc.. These degree of freedom assignments make certain integrals trivial to compute. For example the surface integral of a 2-form field can simply be computed by summing up the DoF values from a series of faces approximating the surface of interest.

Form Type	Field Type	Exterior Derivative	Weak Derivative
0	Scalar	$\nabla \mapsto T_{01}$	None
1	Vector	$\nabla \times \mapsto T_{12}$	$\nabla \cdot \mapsto M_0^{-1} T_{01}^T M_1$
2	Pseudo Vector	$\nabla \cdot \mapsto T_{23}$	$\nabla \times \mapsto M_1^{-1} T_{12}^T M_2$
3	Scalar Density	None	$\nabla \mapsto M_2^{-1} T_{23}^T M_3$

Table 2: This table lists the discrete derivative operators appropriate for each form type. Where M_p is a p -form mass matrix and T_{pq} is a topological derivative matrix acting on p -forms and producing q -forms.

Certain derivative operators, see Table 2, are also quite simple to apply. This stems from the generalized Stokes' theorem, equation (1). Consider the Curl of a 1-form, equation (2) shows that the 2-form DoF of $\nabla \times \omega$ on a particular face is simply a linear combination of the 1-form DoFs of ω which form the boundary of that face. For first order basis functions the coefficients are just ± 1 . These coefficients are chosen to produce the correct orientation of the

Computer Codes (Design, Simulation, Field Calculation)

path for the line integral. In fact the Gradient of a 0-form, the Curl of a 1-form, and the Divergence of a 2-form can all be computed in this manner. These derivatives only require knowledge of the topology of the mesh and not its geometry. This follows from the fact that the geometry is incorporated into the value of the degree of freedom itself.

An important consequence of this separation of the geometry and topology of the mesh is that certain vector calculus identities can be satisfied to machine precision. Specifically, the identities $\nabla \times \nabla \phi = 0$ for all scalar functions ϕ and $\nabla \cdot \nabla \times \vec{F} = 0$ for all vector functions \vec{F} are precisely reproduced by the discrete operators T_{pq} , i.e. the matrix products $T_{12}T_{01}$ and $T_{23}T_{12}$ are equal to zero to machine precision.

For a more complete treatment of discrete differential forms see [3].

Example Problem

Consider a laser target chamber, which is roughly cylindrical with a height of nearly one meter and a radius of one meter. The chamber also has several port holes for diagnostic equipment as well as the input port for the laser beam. When a high power laser beam enters the chamber and strikes its target, it will partially vaporize the target and generate a flux of electrons which are propelled towards the outer walls of the chamber.

The charge packet in the simulations that will be discussed consists of 10^{12} electrons moving at essentially the speed of light from a target post in the center of the chamber towards the wall on the right in the upcoming images. The maximum current was 1.5 kA. The packet fans out in a conical shape as it progresses and has a Gaussian shape along the direction of propagation with a full width at half maximum of 3 cm. Therefore we expect the signal to have significant frequency content out to roughly 4.5 GHz.

We are primarily interested in the pulse of electromagnetic waves radiated by this charge packet so we do not model the incoming laser beam or the vaporization of the target. Also, we do not currently attempt to model the charge packet as a plasma, it is simply a known charge density moving through the mesh in a prescribed fashion. This approximation is valid for this particular problem because the liberated electrons have very high energies.

SOLUTION TECHNIQUES

Typical E/B Formulation

As mentioned previously a standard E/B formulation [4][5][6] of the problem requires that \vec{J} be approximated by discrete 1-forms with degrees of freedom on the edges of the mesh.

$$\begin{aligned} \frac{\partial}{\partial t} (\epsilon \vec{E}) &= \nabla \times \frac{1}{\mu} \vec{B} - \vec{J} \\ \frac{\partial}{\partial t} \vec{B} &= -\nabla \times \vec{E} \end{aligned}$$

In discrete form these equations become:

$$\begin{aligned} M_1(\epsilon) \frac{e_{n+1} - e_n}{\Delta t} &= T_{12}^T M_2(\mu^{-1}) b_{n+\frac{1}{2}} - M_1 j_{n+\frac{1}{2}} \\ \frac{b_{n+\frac{3}{2}} - b_{n+\frac{1}{2}}}{\Delta t} &= -T_{12} e_{n+1} \end{aligned}$$

Where $M_p(\alpha)$ represents a mass matrix computed using p -form basis functions and a material parameter α , T_{12} is the discrete Curl operator shown in Table 1, and lower case letters with subscript n , for example, obviously represent vectors of degrees of freedom at time $n\Delta t$.

If we take the divergence of Ampère's law and make use of the fact that the charge density is related to the electric displacement via $\nabla \cdot \vec{D} = \rho$, we derive the charge conservation equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{J} = 0.$$

In discrete form this becomes:

$$\frac{1}{\Delta t} M_0 (\rho_{n+1} - \rho_n) + T_{01}^T M_1 j_{n+\frac{1}{2}} = 0$$

Where ρ_n is a 0-form, i.e. nodal, representation of the charge density at time $n\Delta t$. The divergence of a 1-form can only be defined in a weak sense, i.e. as a type of least squares best fit. Hence this continuity equation for the electric charge may not be locally satisfied everywhere although it should be nearly satisfied globally.

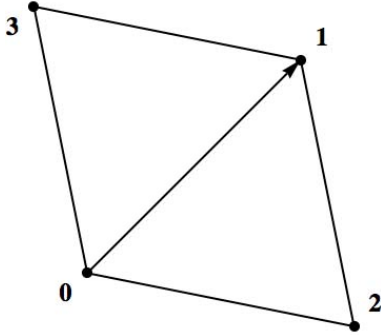


Figure 1: A toy problem illustrating the difficult of moving a charge density from one node to another using only a current density on the edge shared by the two nodes.

Consider the seemingly simple problem of moving a charge from one node to another along a particular edge of a mesh as shown in Figure 1. How much current density should be applied to the edge connecting nodes 0 and 1 to move all of the charge density from node 0 to node 1? Even this simple, two element, problem is over determined and cannot be solved.

$$\begin{aligned} \rho_n &= (\tilde{\rho}_0, 0, 0, 0) \\ \rho_{n+1} &= (0, \tilde{\rho}_1, 0, 0) \\ j_{n+\frac{1}{2}} &= (\tilde{j}_{01}, 0, 0, 0, 0) \end{aligned}$$

Computer Codes (Design, Simulation, Field Calculation)

We have two unknowns; the charge density $\tilde{\rho}_1$ deposited on node 1, and the current density $\tilde{j}_{n+\frac{1}{2}}$, but we have four equations leading to insolubility. This toy problem, even if it could be solved, would be difficult to efficiently extend to realistic situations involving charge densities traversing meshes in three space dimensions.

The simplest method to implement this type of current source is to simply compute the current density on each edge as the projection of a prescribed function onto the edges of the mesh and hope that the continuity equation will be satisfied closely enough that the errors will not be noticeable. In many cases this does indeed work well enough.

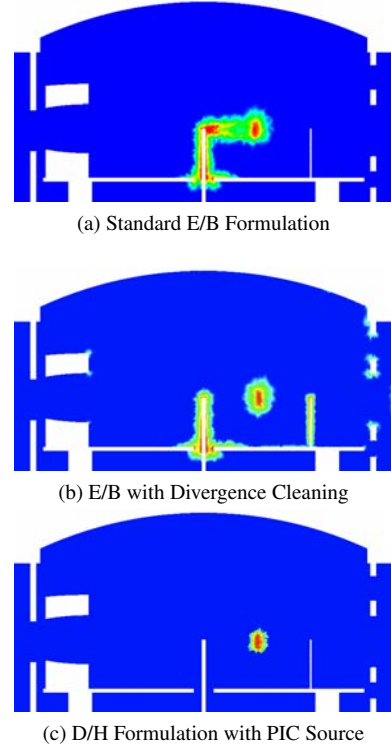


Figure 2: The Divergence of the vector field \vec{D} plotted on a logarithmic scale.

Figure 2a shows an example of a charge density plot for our model problem. The image clearly shows the charge packet itself just to the right of center. Unfortunately, it also shows a large non-physical charge buildup left behind in the wake of the packet. The boundary of the computational domain is assumed to be a perfect electrical conductor so the charge near the boundary can be interpreted as being related to the surface charge density. This is actually another oddity of the E/B formulation, surface charges appear smeared into the volume elements which touch the surface. This may not be an attractive feature of the image but at least it has a reasonable physical interpretation.

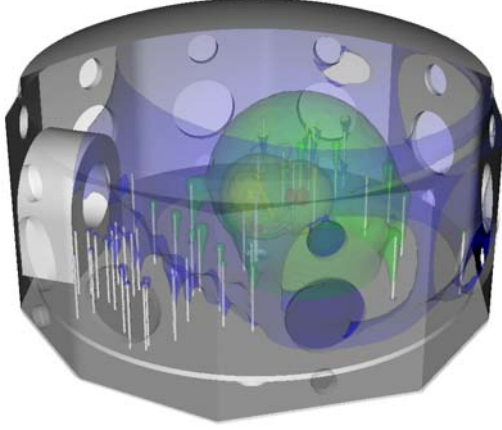


Figure 3: Logarithmically scaled contour plot of the magnitude of the electric field computed using the E/B formulation with divergence cleaning.

Divergence Cleaning

The non-physical charge buildup can be removed by performing divergence cleaning when deemed necessary or perhaps even at every time step. This is the process of adding something to the field so that its divergence has a desired value but its curl remains unchanged [7][8][9]. For the model problem we can add a $\vec{\tilde{J}}$ to the source so that the divergence will match the desired change in charge density given by $\dot{\rho}$. We assume that the correction to \vec{J} is the gradient of a scalar field ψ so that it will have zero curl.

$$\begin{aligned}\dot{\rho} &= -\nabla \cdot \vec{J} \text{ (the computed change in } \rho) \\ \dot{\tilde{\rho}} &= \dot{\rho} + \dot{\tilde{\rho}} = -\nabla \cdot \vec{J} - \nabla \cdot \vec{\tilde{J}} \text{ (the desired change)} \\ \nabla \cdot \vec{\tilde{J}} &= -\dot{\tilde{\rho}} - \nabla \cdot \vec{J} \text{ (the necessary correction to } \vec{J}) \\ \nabla^2 \psi &= -\dot{\tilde{\rho}} - \nabla \cdot \vec{J}\end{aligned}$$

In discrete form this becomes:

$$\begin{aligned}S_0 \psi &= -M_0 \dot{\tilde{\rho}} - T_{01}^T M_1 j_{n+\frac{1}{2}} \\ \tilde{j}_{n+\frac{1}{2}} &= T_{01} \psi\end{aligned}$$

Where S_0 is the 0-form stiffness matrix. Each divergence cleaning operation then requires an additional linear solve to compute the scalar field ψ . It should be noted that it is generally more difficult to solve a stiffness matrix than a mass matrix so obtaining this correction is not a trivial computation in comparison to updating the electric field at each time step using Ampère's law.

With this correction we see that the divergence of \vec{D} , shown in figure 2b, now matches the desired charge density. Again, note that the charge density near the surfaces is due to the presence of a surface charge density.

Computer Codes (Design, Simulation, Field Calculation)

Unfortunately, this method has a drawback when the charge density has a velocity near the speed of light. The correction introduces a small quasi-static field centered on the charge density, which appears to propagate faster than the speed of light. Figure 3 shows a logarithmically scaled contour plot of the electric field magnitude which clearly shows contours well beyond the charge packet, which is located near the innermost contour. In figure 2b this component of the field can also be seen because it introduces surface charge densities on the metal object ahead of the charge packet and on several sharp corners farther away. These non-physical charge densities are obviously due to the global solve necessary to compute ψ .

D/H Formulation

Obviously, the difficulties discussed in this paper stem from the treatment of \vec{J} as a 1-form vector field. Current density is, however, a flux vector, i.e. the amount of charge crossing a given area per unit time. Flux vector fields are more naturally described using 2-forms, so we should have more luck if we approximate Ampère's law using discrete 2-forms.

$$\begin{aligned}\frac{\partial}{\partial t} (\mu \vec{H}) &= -\nabla \times \frac{1}{\epsilon} \vec{D} \\ \frac{\partial}{\partial t} \vec{D} &= \nabla \times \vec{H} - \vec{J}\end{aligned}$$

In discrete form these become:

$$\begin{aligned}M_1(\mu) \frac{h_{n+1} - h_n}{\Delta t} &= -T_{12}^T M_2(\epsilon^{-1}) d_{n+\frac{1}{2}} \\ \frac{d_{n+\frac{3}{2}} - d_{n+\frac{1}{2}}}{\Delta t} &= T_{12} h_{n+1} - j_{n+1}\end{aligned}$$

In this formulation the curl of \vec{D} must be computed in the weak sense. This weak form requires the solution of a linear system to update \vec{H} using Faraday's law. In the standard E/B formulation it is the curl of \vec{B} that must be computed in the weak sense, requiring a linear solve in Ampère's law to update \vec{E} . Normally this linear solve allows us to apply voltage boundary conditions on \vec{E} where we can specify that the tangential component of \vec{E} is zero on perfect electrical conductors. In the D/H formulation this constraint becomes unnecessary because the natural boundary condition is that the tangential component of $\nabla \times \vec{H} = 0$ but, of course, this equation is consistent with the tangential component of $\vec{E} = 0$ on the boundary (assuming that, on the boundary, $\hat{n} \times \vec{E} = 0$ at time $t = 0$ and $\hat{n} \times \vec{J} = 0$ for all time.)

Simply treating \vec{J} as a 2-form does not magically solve all of our problems. What it does is convert our charge buildup problem from a global least-squares fit into much more simple local charge conservation problem. Computing the divergence of Ampère's law in discrete form now leads to:

$$\frac{1}{\Delta t} \left(\rho_{n+\frac{3}{2}} - \rho_{n+\frac{1}{2}} \right) + T_{23} j_{n+1} = 0.$$

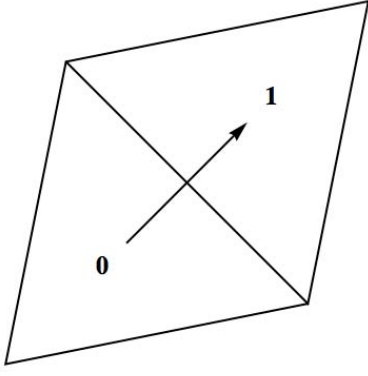


Figure 4: A toy problem illustrating the problem of moving a charge density from one element to another using only a current density on the face shared by the two elements.

Now our toy problem can be expressed in terms of a charge moving from one element to another, see Figure 4, by crossing the face shared by the elements. The problem is now well defined.

$$\begin{aligned}\rho_{n+\frac{1}{2}} &= (q_0, 0) \\ \rho_{n+\frac{3}{2}} &= (0, q_1) \\ \dot{j}_{n+1} &= (I_{01}, 0, 0, 0, 0) \\ T_{23} &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 \end{pmatrix}\end{aligned}$$

We still have two unknowns; the amount of charge deposited in element 1, and the current crossing the intervening face. However, we now have only two equations

$$\begin{pmatrix} 0 \\ q_1 \end{pmatrix} + \Delta t \begin{pmatrix} I_{01} \\ -I_{01} \end{pmatrix} = \begin{pmatrix} q_0 \\ 0 \end{pmatrix}$$

which in this case stipulate that the current is given by $I_{01} = q_0/\Delta t$ and the charge in element 1 is $q_1 = \Delta t I_{01}$ or simply $q_1 = q_0$. This trivial solution is exactly what we would expect and it is as simple as our intuition would suggest.

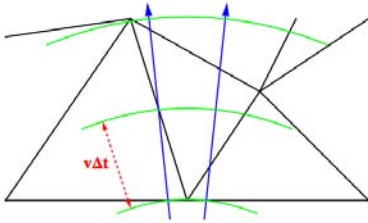


Figure 5: An example mesh illustrating charge carrying rays piercing the faces of an irregular mesh. The curved contours represent the locations at which the current fluxes must be computed as the charges progress through the mesh.

To extend this solution method to a realistic three dimensional mesh is fairly straightforward. One way to accomplish this is to use a particle-in-cell (PIC) technique. For our purposes a very rudimentary PIC method will suffice. Simply split up the trajectory of the charge packet into a group of rays and imagine the charges themselves as beads moving along these rays. At each time step we compute the fraction of each charge which crosses each face to compute the total current density. The use of fractional charges serves two purposes; to smooth out the charge packet even in coarse meshes, and more importantly to account for irregular mesh spacings. Figure 5 demonstrates a typical situation.

To achieve charge conservation we must ensure that the total amount of charge entering an element equals the that leaving. The use of rays makes this much easier as we can then require that the rays will enter and exit each element through unique faces rather than allowing charges to split and leave through two different faces. Of course this is a simplification which would have to be removed in a proper PIC simulation. The problem of charge conservation then becomes one of choosing current fluxes $I_i(t)$ which satisfy:

$$\sum_{n=0}^N I_i(n\Delta t) = \sum_{n=0}^N I_j(n\Delta t)$$

Where indices i and j indicate any two faces pierced by a particular ray and $T = N\Delta t$ is chosen large enough that the charge packet will have passed both faces by time T (we must also assume that $I_i(0) = 0$ for all faces but this is certainly reasonable.) If we use the first face pierced by the ray as a reference and label it with index 0 we can then chose all other current fluxes along that ray to be:

$$I_i(t) = \alpha_i I_0(t - n_i \Delta t) + (1 - \alpha_i) I_0(t - (n_i + 1) \Delta t)$$

Where n_i is the integer part of $d_i/(v\Delta t)$, α_i is the fractional part, d_i is the distance from face 0 to face i , and v is the speed of the charge packet. With these choices it is easy to show that the total current crossing each face along a particular ray, given by the above sums, must be equal aside from numerical errors due to finite precision arithmetic.

If enough rays are used and there are enough beads strung along each ray, then the source will appear reasonably smooth. The simulation discussed in this paper required fewer than 250 rays with 160 charges along each (i.e. less than 40,000 particles) to achieve acceptable results. We should emphasize that this does not constitute a self-consistent PIC simulation. The fields do not effect the motion of the charge packet in any way. We are simply using the PIC concept as a bookkeeping scheme to maintain charge conservation. Although this rigid beam approximation is valid for our test problem it is not a requirement of this method. The accuracy, and ease of implementation, of this method relates to the identification of the degrees of freedom of \vec{J} with the precise values of the charge fluxes across the mesh faces. In a more elaborate PIC simulation these fluxes could still be used to precisely balance the

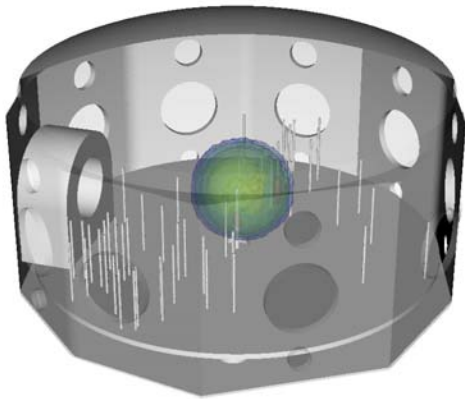


Figure 6: Logarithmically scaled contour plot of the magnitude of the electric field computed using the D/H formulation with a PIC source.

charges within each element although the bookkeeping and charge balance equations would become more complicated.

Figure 2c shows a charge density plot produced using this scheme. Clearly the image shows no sign of non-physical charge buildup. Additionally, the surface charge density does not appear. A further advantage of this method is that the charge density and current density on the surface of perfect electrical conductors can be more accurately computed, if desired. These surface fields can be directly computed from the surface degrees of freedom for \vec{D} and \vec{H} respectively.

Figure 6 again shows a logarithmically scaled contour plot of the electric field magnitude, analogous to that shown in figure 3. However, in the new plot the non-physical, quasi-static field contours are no longer present. The fields now properly propagate within a spherical shell which expands at the speed of light.

CONCLUSION

We have presented an outline for a charge conserving method of applying transient volumetric current sources to the Maxwell Equations in the time-domain. Some of the advantages of using a D/H formulation of the coupled first order wave equation have been discussed. The ability to run charge conserving simulations of transient current densities, while optionally computing accurate representations of surface currents and charge densities, is very appealing. The added benefit of more easily coupling to a PIC simulation, capable of more accurately modeling the motion of the charge packet itself, provides numerous avenues for enhancing the modeling of similar problems.

It should also be noted that the standard E/B formulation and the PIC method placed essentially equivalent demands on computing resources. Each simulation was performed Computer Codes (Design, Simulation, Field Calculation)

using the same number processors and ran for virtually the same length of time. Conversely, the divergence cleaning procedure, using an algebraic multi-grid solver, increased the run time by a factor of roughly 2.8.

REFERENCES

- [1] P. Castillo, R. Rieben, and D. White. FEMSTER: An object oriented class library of discrete differential forms. In *Proceedings of the 2003 IEEE International Antennas and Propagation Symposium*, volume 2, pages 181–184, Columbus, Ohio, June 2003.
- [2] P. Castillo, J. Koning, R. Rieben, and D. White. A discrete differential forms framework for computational electromagnetics. *Computer Modeling in Engineering & Sciences*, 5(4):331–346, 2004.
- [3] Douglas N. Arnold, Richard S. Falk, and Ragnar Winther. Finite element exterior calculus, homological techniques, and applications. *Acta Numerica*, 15:157–256, 2006.
- [4] D. White and M. Stowell. Full-wave simulation of electromagnetic coupling effects in RF and mixed signal IC's. *IEEE Trans. Microwave Theory Tech.*, 52(5):1404–1413, 2004.
- [5] R. Rieben, D. White, and G. Rodrigue. A high order mixed vector finite element method for solving the time dependent Maxwell equations on unstructured grids. *J. Comput. Phys.*, 204:490–519, 2005.
- [6] D. White, J. Koning, and R. Rieben. Development and application of compatible discretizations of Maxwell's equations. In D. Arnold, P. Bochev, R. Lehoucq, R. Nicolaides, and M. Shaskov, editors, *Compatible Spatial Discretizations*, volume 142 of *The IMA Volumes in Mathematics and its Applications*, pages 209–238. Springer, Berlin, 2006.
- [7] G.B. Jacobs and J.S. Hesthaven. Implicit-explicit time integration of a high-order particle-in-cell method with hyperbolic divergence cleaning. *Computer Physics Communications*, 180(10):1760 – 1767, 2009.
- [8] C.-D. Munz, P. Omnes, R. Schneider, E. Sonnendrücker, and U. Voß. Divergence correction techniques for maxwell solvers based on a hyperbolic model. *J. Comput. Phys.*, 161(2):484–511, 2000.
- [9] C. D. Munz, P. Ommes, and R. Schneider. A three-dimensional finite-volume solver for the maxwell equations with divergence cleaning on unstructured meshes. *Computer Physics Communications*, 130(1-2):83 – 117, 2000.

SIMULATION AND COMMISSIONING OF J-PARC LINAC USING THE IMPACT CODE

M. Ikegami, KEK, Tsukuba, Ibaraki 305-0801, Japan
H. Sako, T. Morishita, JAEA, Tokai, Naka, Ibaraki 319-1195, Japan
G. Shen, BNL, Upton, NY 11973-5000, USA

Abstract

The IMPACT code has been utilized for the beam commissioning of J-PARC linac. The activity is presented by reviewing two illustrative topics, where the experimental data is analyzed to realize a finer tuning. One is the RF set-point tuning for a DTL tank, where we have a significant discrepancy between the experimental result and prediction from a simple numerical model. The other is the beam profile measurement, where significant beam quality deterioration is found to develop in a characteristic way. In both cases, the IMPACT code has helped us to deepen our insight into the beam behavior.

INTRODUCTION

The beam commissioning of J-PARC linac was started in November 2006, and its initial stage was completed in October 2007 by achieving the linac beam power of 1.2 kW [1]. This beam power corresponds to 20 kW from the succeeding 3-GeV RCS (Rapid Cycling Synchrotron), and it is sufficient for the initial beam commissioning of the downstream facilities. Since then, J-PARC linac has been operated to provide a stable beam for the commissioning of downstream RCS, MR (Main Ring), and their beam lines to the experimental targets. After succeeding in delivering the first beams to all the experimental targets in May 2009, we are now in the next stage where we seek the operation with higher beam power.

J-PARC is a high-power frontier machine aiming at 1-MW beam power from RCS (133 kW from linac) in the final phase. Accordingly, it is of essential importance to reduce the uncontrolled beam loss, and hence, to avoid excess radio-activation of the accelerator components so as to maintain its hands-on maintenance capability. This is the case even in the early stages of the beam commissioning, and we need to reduce the integrated beam loss during the beam tuning. Therefore, it is required to realize a more sophisticated and efficient tuning rather than a traditional trial-and-error tuning. To this end, a simple and fast on-line numerical model plays an essential role in the beam commissioning of J-PARC linac.

On the other hand, the beams in a high-intensity linac are subject to strong space-charge forces. It often invokes collective and nonlinear phenomena, such as emittance growth and halo formation, being accompanied with the various operational errors. As these phenomena often lead to undesirable beam losses, we need to realize a precise tuning in Controls and Computing

beam-power ramp-up. Even a very small fraction of beam loss can cause serious radio-activation in a high-intensity operation. Therefore, more thorough and fine-grained understanding of the beam behavior and the space-charge-driven phenomena is required in ramping up the beam intensity. To this end, we need a precise and detailed simulation of the beam behavior with a time-consuming PIC (Particle-In-Cell) tracking.

In the beam commissioning, we fully utilize two numerical models which complement each other. One is an on-line envelope model, and the other is an off-line PIC model.

As an on-line model, we have adopted XAL originally developed for SNS [2]. XAL is a JAVA-based high-level software development framework dedicated to accelerator beam commissioning, and it includes an envelope model to be utilized as an on-line model. This model is capable of calculating the evolution of rms beam widths and a beam center orbit swiftly. However, it can not simulate the space-charge-driven emittance growth and halo development. This model has been used for various beam tuning in J-PARC linac directly connected with high-level software [3].

As an off-line model, we have mainly adopted the IMPACT code developed at LBNL [4]. IMPACT is a fully three-dimensional PIC code optimized for parallel computing, which is suitable for the detailed simulation for the space-charge-driven phenomena including emittance growth, halo formation, and resulting beam loss. We use IMPACT for the beam simulation from the RFQ (Radio Frequency Quadrupole linac) exit to the injection point to RCS. The initial distribution for the IMPACT simulation is generated with the PARMTEQM code [5].

In this paper, we show some examples of the studies

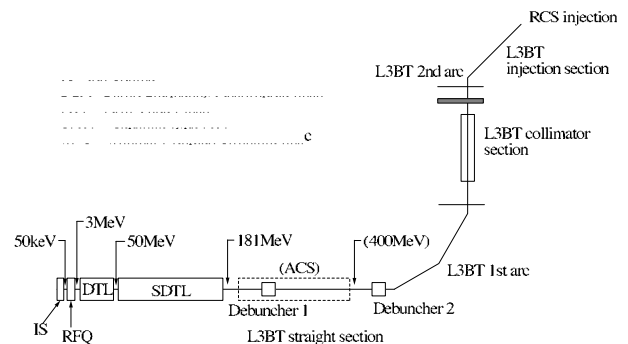


Figure 1: Schematic layout of J-PARC linac.

where the IMPACT code is used to analyze the experimental data obtained in the beam commissioning of J-PARC linac. In particular, we here take up the following two topics where the detailed simulation with the IMPACT code have helped us to deepen our understandings on the beam behavior in J-PARC linac. One is the phase-scan tuning of RF set-points for the first DTL tank (or DTL1), where the experimental observation shows significant discrepancy from a simple numerical model. The other is the analysis of the beam profile measurement at the exit of DTL section and SDTL (Separate-type DTL) section. In these measurements, we have observed a substantial beam quality deterioration which develops in a characteristic way. As we lack sufficient beam diagnostics in the DTL section, we have performed an IMPACT simulation to reproduce its characteristic features and then to infer its underlying mechanism.

These two topics have already been discussed in other literatures [6, 7] in detail. Then, we here try to present a brief review of these studies instead of delving into details.

LAYOUT OF J-PARC LINAC

Before proceeding to the simulation study, we here run through the basic layout of J-PARC linac. As shown in Fig. 1, J-PARC linac consists of a 50-keV negative hydrogen ion source, 3-MeV RFQ, 50-MeV DTL, and 181-MeV SDTL. We also have a 3-m long matching section between RFQ and DTL, to which we refer as MEBT (Medium Beam Energy Transport).

MEBT consists of eight quadrupole magnets, two buncher cavities, an RF chopper system, and various beam diagnostics including four wire-scanner beam profile monitors. It also has a 45-deg bend magnet followed by a transverse emittance monitor of double-slit type.

In contrast to MEBT, the DTL section has very limited number of beam diagnostics. The DTL section consists of three DTL tanks, and its total length is 27 m. As it has no available space for beam diagnostics inside the tank, it only has a beam current monitor and a beam phase monitor at each inter-tank spacing.

In SDTL section, the quadrupole magnets are placed between tanks instead of inside the drift tube. Therefore, it has a longer inter-tank spacing which can accommodate various beam diagnostics. In particular, we have an array of four wire scanners periodically placed at the inter-tank spacings at the DTL exit (or the most upstream portion of the SDTL section). From the measurement with these wire scanners, we can calculate the transverse Twiss parameters and the rms emittance assuming a design Twiss parameters and the emittance in the longitudinal direction. We also have a similar setup at the exit of the SDTL section and some downstream locations. The SDTL section consists of 30 tanks, and its total length is about 84 m.

To be noted here is that we lack the instrumentation for longitudinal profile measurement throughout the linac, while we are planning to introduce a few bunch shape monitors of INR type [8].

Controls and Computing

DTL PHASE SCAN

The set-points of RF phase and amplitude for a cavity are important parameters to be determined with a beam-based tuning. To find an adequate set-point, we have performed a so-called phase scan tuning. In this tuning, the phase set-point of a cavity is scanned with a fixed amplitude while measuring the output beam energy with the TOF (Time Of Flight) method. The phase scan provides us with a dependence of the output energy on the tank phase, to which we refer as a “phase scan curve”. We iterate the same procedure with different amplitude settings, and then compare the obtained phase scan curves with a numerical model. As

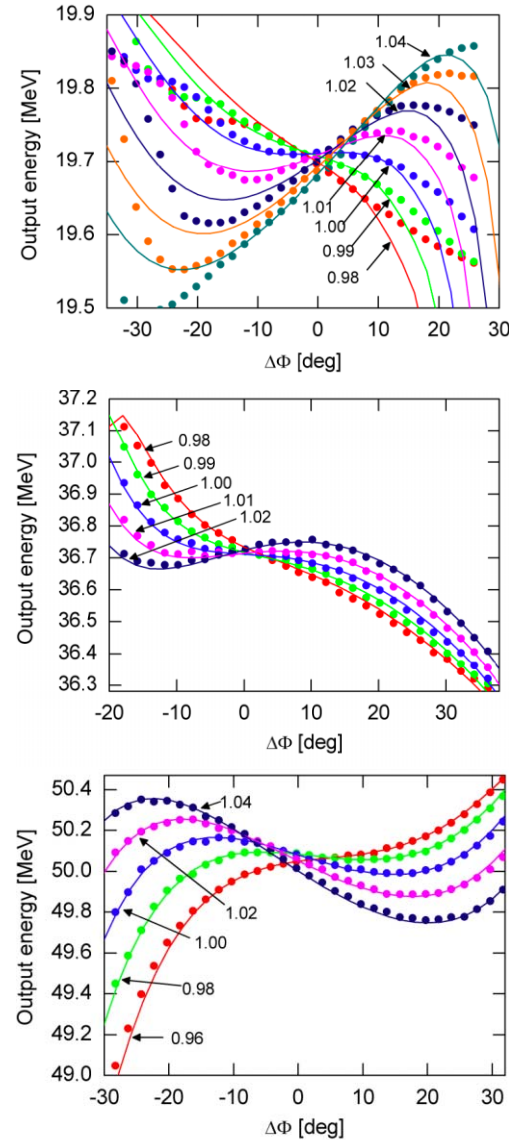


Figure 2: Measured and simulated phase scan curves for DTL1 (top), DTL2 (middle), and DTL3 (bottom). The scaled RF amplitude A is annotated for each curve. The measured results are shown with circle markers, and those from PARMILA modeling are shown with solid lines.

the phase scan curve of a DTL tank has a peculiar shape, we can find an adequate RF set-point with their signature matching. This tuning is performed one klystron at a time from the upstream end.

Figure 2 shows the phase scan result for DTL tanks, where we adopt the simple beam centroid motion simulated with PARMILA [9] as the reference for the tuning. In this figure, $\Delta\Phi$ denotes the phase shift from the design phase setting and A is the RF amplitude scaled by its design value. It is readily seen in Fig. 2 that the measured phase scan curves are thoroughly reproduced by the numerical model for DTL2 and DTL3. However, it shows notable deviations in DTL1 especially for the case with an RF set-point away from its design value. Furthermore, the trend of the phase scan curve is totally different when the amplitude setting is lower than the design value. Even with higher amplitude, the experimental curve shows a large deviation from the modeling for a large phase shift from the design value. These disagreements prevent us from performing an accurate phase signature matching, and has motivated us to establish a more rigorous numerical model employing a fully three-dimensional multi-particle tracking.

To investigate the disagreement between the simulated beam centroid motion and the measurement for DTL1, a parallel PIC simulation has been performed with the IM-

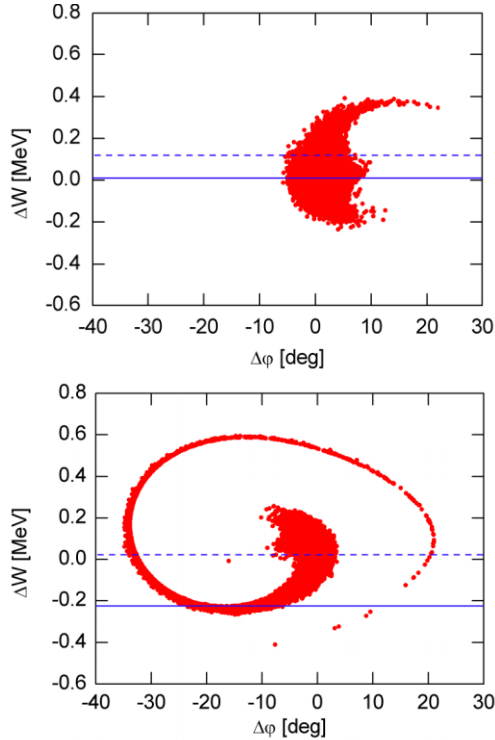


Figure 3: The longitudinal phase space distribution at the DTL1 exit simulated with the IMPACT code. Top: $\Delta\Phi = -29.5$ deg and $A = 1.02$, and bottom: $\Delta\Phi = 25$ deg and $A = 1.00$. The measured beam energy is shown with a broken line, and the centroid of the simulated distribution with a solid line.

Controls and Computing

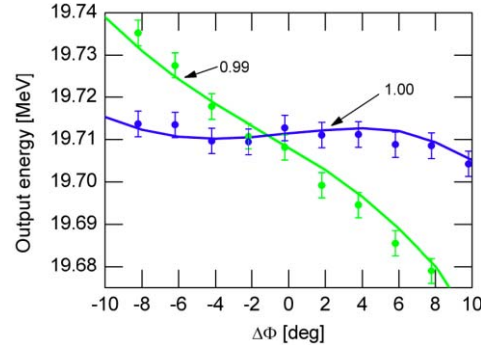


Figure 4: Measured and simulated phase scan curves for DTL1 with a restricted parameter range. Two phase scan curves are shown for $A = 0.99$ and 1.00 as annotated in the figure. The measured results are shown as circle markers, and the curves from PARMILA modeling are shown with solid lines.

PACT code. The tracking is performed from the exit of the RFQ with the initial distribution obtained with the PARMTEQM [10] (which is the same with the PARMILA simulation). The nonlinear Lorentz map integrator is utilized to deal with the highly nonlinear RF force which arises from unusually large RF set-point deviations involved in the phase scan tuning. To attain a reasonable accuracy, the integration step width is set to about $\beta\lambda/100$ with β and λ being the particle velocity scaled by the speed of light and the RF wave length, respectively. In the simulation, 95,322 simulation particles and $32 \times 32 \times 64$ meshes are employed.

To illustrate the findings in the simulation study, we show in Fig. 3 the simulated longitudinal distribution at the DTL1 exit for two cases with $(\Delta\Phi, A) = (-29.5$ deg, $1.02)$ and $(25.0$ deg, $1.00)$. These settings corresponds to the case where the measured phase scan curves show a significant deviation from the model. It is clearly seen in Fig. 3 that the longitudinal distribution is subject to significant filamentation with these settings. Furthermore, it has been confirmed that the extent of the filamentation is substantially dependent on the assumed initial distribution. The observed discrepancy of several tens of keV in the output energy is easily caused with a modest difference in the initial distribution. Meanwhile, we are unable to confirm the credibility of the assumed initial longitudinal distribution at the RFQ exit due to lack of longitudinal diagnostics in MEBT.

In conclusion, the discrepancy between the measured phase scan curve and that from a numerical model observed in the DTL1 tuning seems to be mainly attributable to the generation of a significant filamentation. Because the filamentation depends on the initial distribution, the discrepancy of several tens of keV is unavoidable with the RF set-point far from the design value. Then, the phase signature matching is valid only in the narrow region around the design set-point where the filamentation is sufficiently modest and insensitive to the initial distribution.

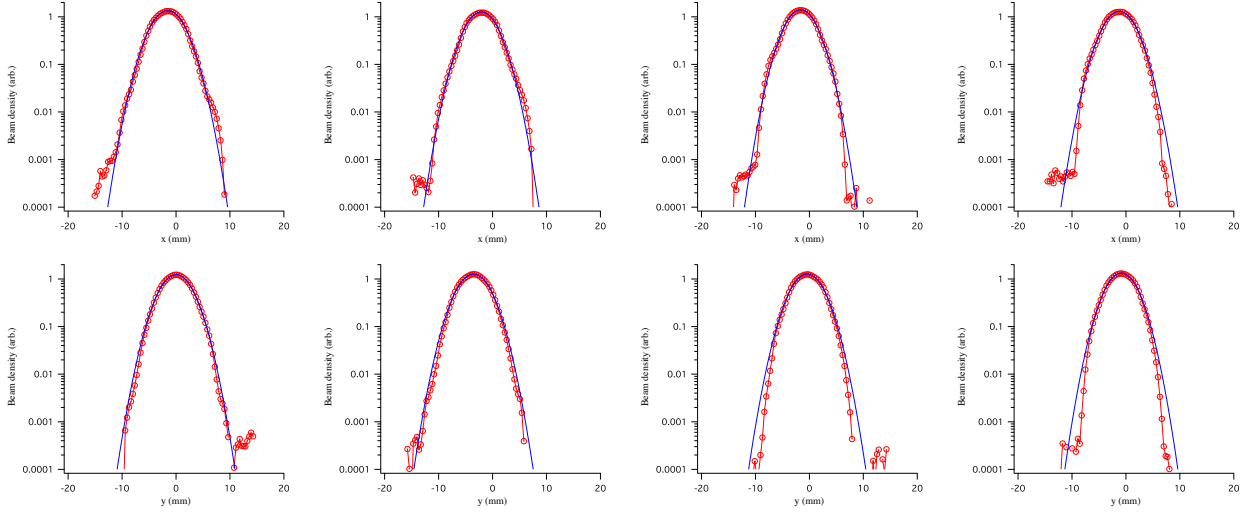


Figure 5: Typical beam profile measured at DTL exit with 30 mA peak current. Red circle: measurement, blue line: Gaussian fit. The beam profile measured with an array of four wire scanners are shown. The left two figures are the results with the first (the most upstream) wire scanner with the horizontal profile on the top and the vertical profile on the bottom. The mid-left, mid-right, and right figures are respectively results with the second, third, and fourth wire scanners. The same notation is adopted in Figs. 6, 7, and 8.

The phase scan curves for DTL1 are shown again in Fig. 7, where the parameter range is limited to have modest filamentation. It is readily seen in this figure that the experiment and the numerical model show a reasonable agreement. As seen in this figure, the goal tuning accuracy of 1 deg in phase and 1 % in amplitude is confirmed to be attainable for DTL1 with the narrow range analysis with two reference curves. It should be stressed here that the adequate phase scan range has been found with a help of the IMPACT simulation.

BEAM PROFILE MEASUREMENT

To mitigate the beam loss in the linac and the succeeding RCS, it is practically important to suppress excess emittance growth and beam halo formation. To measure the transverse emittance and the beam tail shape, we use beam profile monitors of wire-scanner type installed along the beam line. As mentioned above, four wire scanners are periodically installed at the exit of DTL. Then, the rms emittance can be calculated from the rms beam widths measured with this wire scanner array. With the design peak current of 30 mA, the obtained rms emittance at the exit of DTL is $0.42 \pi \text{ mm-mrad}$ in horizontal and $0.36 \pi \text{ mm-mrad}$ in vertical. On the other hand, the measured emittance at MEBT is around $0.22 \pi \text{ mm-mrad}$. All the emittance values are normalized. These observation indicates that we have a significant emittance growth in DTL. Besides, the emittance growth is found to be modest with lower peak current of 5 mA.

We also have similar setups of wire scanners at the exit of SDTL. In the observation with these wire scanners, there is no significant emittance growth after the DTL exit in both Controls and Computing

5-mA and 30-mA cases. This tendency has also been confirmed in more downstream sections.

Another interesting feature of the measurement is the shape of beam profile. Figure 5 shows a typical beam profile measured at the DTL exit. The beam profile is measured with four wire scanners in this section, and each wire scanner is $7 \beta \lambda$ apart. As readily seen in this figure, the beam profile is virtually Gaussian in spite of the significant emittance growth in DTL. Contrary to our expectations, the measured beam profile at the DTL exit lacks beam halo or “shoulder-like structure”. As the phase advance between wire scanners is about 60 deg in this region, the halo is supposed to be detected by some of these wire scanners if it has been generated. Meanwhile, the halo-like structure is clearly seen at the SDTL exit as shown in Fig. 6. It should be stressed here that the halo is generated despite the absence of significant emittance growth in the SDTL section.

This interesting feature has motivated us to perform particle simulations. To reproduce the experimental result, IMPACT simulations have been performed with various mismatch conditions in MEBT. Needless to say, it is of practical importance to understand the mechanism of the emittance growth and find the way to avoid it. Especially, reduction in the transverse emittance enables more flexible painting injection into RCS, and it is expected to contribute to the beam loss mitigation in RCS.

We have performed IMPACT simulation with the same simulation conditions with those in the previous section except for the choice of integrator and its step width. We here adopt the linear map integrator with the step width of $\beta \lambda / 10$ so that we can survey a wider parameter space.

We have tried several kinds of mismatch at MEBT artificially introduced in both of the transverse and longitudinal

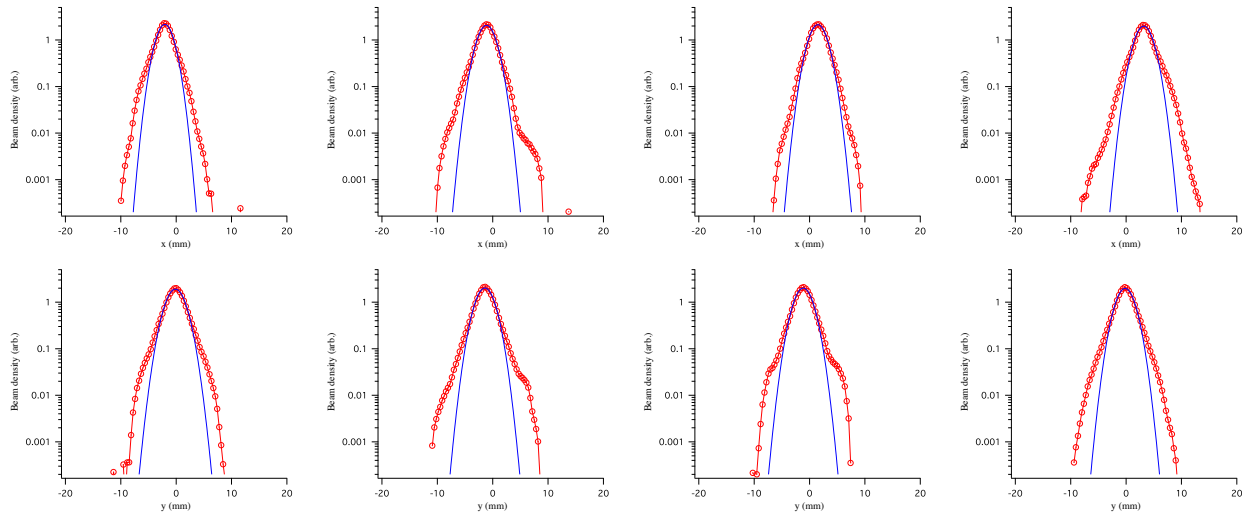


Figure 6: Typical beam profile measured at SDTL exit with 30 mA peak current.

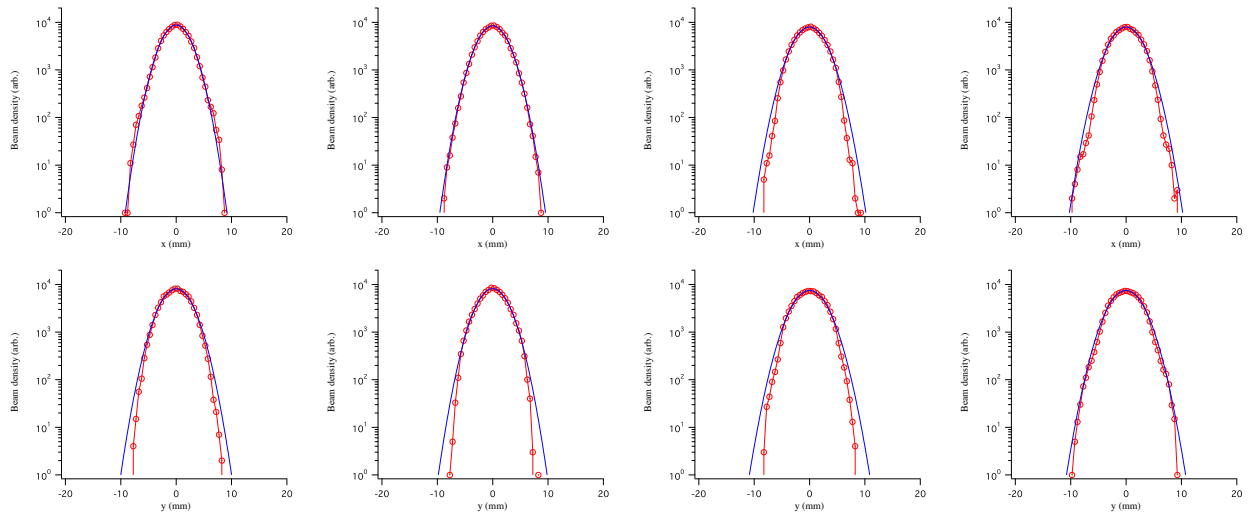


Figure 7: Simulated beam profile at DTL exit with 30 mA peak current assuming a larger longitudinal emittance than the PARMTEQ prediction.

nal directions. IMPACT simulations reveal that 30 to 40 % transverse mismatch oscillation at the upstream portion of DTL is anticipated to account for the observed emittance growth, where we define the degree of mismatch as the mismatch oscillation amplitude in the rms beam width. Either of the transverse and longitudinal mismatch in MEFT can drive the transverse mismatch oscillation in DTL through the space-charge coupling. We have also found in the simulation study that the halo develops more rapidly in most cases than the experimental observation with the assumed level of initial mismatch. In these cases, a clear halo has already been generated at the DTL exit, which disagrees with the experimental observation.

An extensive simulation study reveals that the onset of halo generation has a certain sensitivity to the kind of mismatch assumed in the simulation. Actually, the onset is Controls and Computing

delayed in some cases with certain types of longitudinal mismatch. Figures 7 and 8 show an example of these cases, where a larger longitudinal emittance is assumed than the PARMTEQM prediction. As readily seen in these figures, the beam profile at the DTL exit is virtually Gaussian, while that at the SDTL exit has a clear halo. The emittance growth in SDTL is also confirmed to be modest in this case.

It is demonstrated in this case that the experimentally observed beam behavior can be qualitatively reproduced with a particle simulation assuming a certain type of longitudinal mismatch at MEFT. The similarity in Figs. 5, 6, 7, and 8 is significant, while the simulated halo at the SDTL exit is a little less pronounced than the measurement. This finding does not exclude the possibility that the actual cause of mismatch is different from that assumed in this case. However,

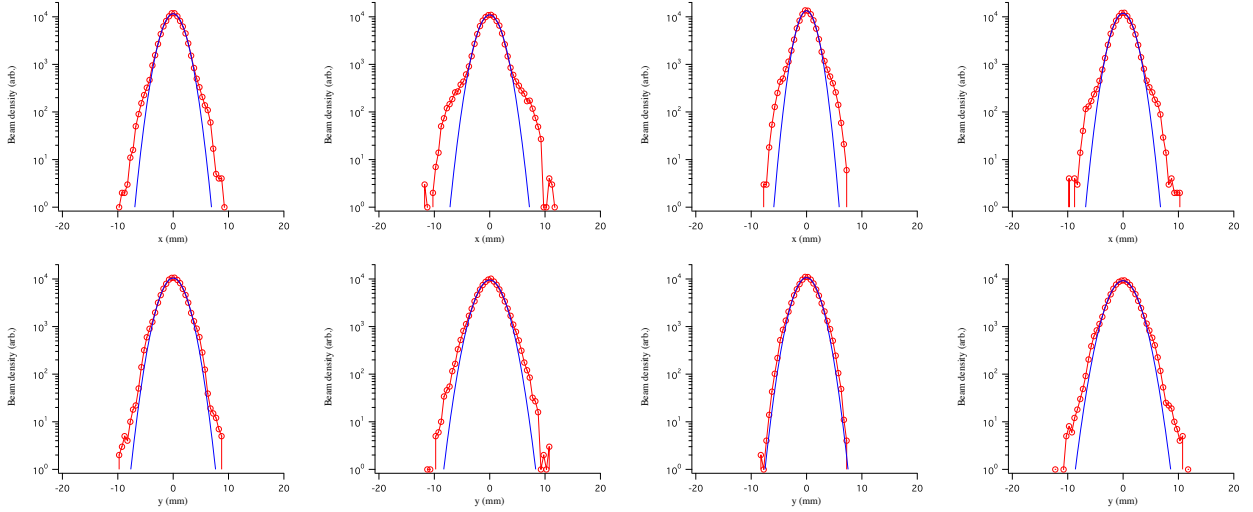


Figure 8: Simulated beam profile at SDTL exit with the same conditions as Fig. 7.

it suggests that we can narrow down the possible source of mismatch by surveying the parameter space with an extensive and comprehensive simulation. Then, it is supposed to contribute to identifying the actual cause of the mismatch utilizing the IMPACT results.

According to the observation with MEBT wire scanners, it is not likely to have a transverse mismatch of 30 to 40 % at the DTL entrance. Meanwhile, the lack of the longitudinal diagnostics at MEBT is a potential cause of excess longitudinal mismatch. These also support the hypothesis that the emittance growth is caused by a large longitudinal mismatch at MEBT. We are planning to improve the longitudinal matching in the coming beam commissioning runs by adjusting the amplitude of MEBT buncher cavities.

SUMMARY

We have adopted the IMPACT code as an off-line model to analyze the experimental results obtained in the beam commissioning of J-PARC linac. The beams in J-PARC linac are subject to the strong space-charge forces, and various undesirable phenomena arise due to its profoundly nonlinear nature. In the beam commissioning of J-PARC linac, it is of critical importance to mitigate the uncontrolled beam loss below an extremely low level so as to secure the hands-on maintenance capability. Then, deep understanding of the space-charge-driven phenomena is essential for its beam commissioning. To this end, detailed PIC simulation is an essential tool to analyze the experimental data to help understand its underlying physics.

The experimental data obtained in the beam commissioning often fails to be comprehensive due to lack of beam diagnostics. To make up the insufficient data and to get physical understanding of it, it is often required to perform a systematic simulation study covering a wide parameter space. In this paper, we have shown two illustrative topics where the IMPACT code is used in the J-PARC linac beam

commissioning. It should be stressed here that we adopt rather modest number of simulation particles and mesh grids in these studies compared to the typical numerical study for the space-charge beam dynamics. Adopting the modest simulation condition, the IMPACT code can provide us with the simulation result in a few tens of minutes to a few years even with a multi-core PC thanks to its fully optimized feature for parallel computing. This prompt response extends the parameter space we can cover in the simulation study, and widen the possibility of realizing extremely fine tuning required for a high-power frontier machine.

REFERENCES

- [1] M. Ikegami, "Progress in the beam commissioning of J-PARC linac and its upgrade path", to be published in Procs. of LINAC'08.
- [2] J. Galambos et. al., "XAL application programming structure", PAC'05, Knoxville, May 2005, p. 79.
- [3] H. Sako et. al., "Development of commissioning software system for J-PARC linac", PAC'07, Albuquerque, June 2007, p. 257.
- [4] J. Qiang, R. D. Ryne, S. Habib, V. Decyk, J. Comput. Phys. 163 (2000) 434.
- [5] K. R. Crandall et. al., "RFQ design codes", LA-UR-96-1836.
- [6] G. Shen, M. Ikegami, Chinese Physics C, Vol. 33 (2009), p. 577-582,
- [7] M. Ikegami, H. Sako, T. Morishita, Y. Kondo, "Measurement and simulation in J-PARC linac", to be published in Procs. of ICFA ABDW HB2008.
- [8] A. V. Feschenko, "Methods and instrumentation for bunch shape measurement", PAC'01, Chicago, June 2001, p. 517.
- [9] T. Takeda, "PARMILA", Los Alamos National Laboratory Report LA-UR-98-4487.
- [10] Y. Kondo et. al., "Particle distribution at the exit of the J-PARC RFQ", LINAC'04, Luebeck, August 2004, p. 78.

PHYSICS PROBLEM STUDY FOR A 100 MEV, 500 MICROAMP H⁺ BEAM COMPACT CYCLOTRON *

Technology Division of BRIF, CIAE (written by Tianjue Zhang^{1,‡}, Jianjun Yang^{1,2,†}, and Hongjuan Yao¹)

¹. China Institute of Atomic Energy, Beijing 102413, China

². Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Abstract

A high intensity compact cyclotron, CYCIAE-100, is selected as the driving accelerator for Beijing Radioactive Ion-beam Facility (BRIF). At present the physics design of this machine has been accomplished. This paper gives a brief review of the general design of this machine. For further intensity upgrade of this compact machine in the future, it is crucial to carry out in-depth study on the self fields effects including the contributions of single bunch space charge and the interaction of many radially neighboring bunches. In order to include the neighboring bunch effects fully self-consistently in compact cyclotrons, a new physical model is established for the first time and implemented in the parallel PIC code OPAL-CYCL. After that, the impact of the single bunch space charge and neighboring bunches on the beam dynamics in CYCIAE-100 for different intensity levels are studied by the simulations using the new model.

INTRODUCTION

Since 2004 a new exotic beam project, Beijing Radioactive Ion-beam Facility (BRIF), has been started at CIAE. As a driving accelerator for BRIF[1], CYCIAE-100 adopts a compact structure with 4 straight sectors. The H⁺ ions produced by the multi-cusp ion source are accelerated, and the high intensity proton beams are extracted through dual stripping. The extracted beam is 200–500 μA featured with energy of 75–100 MeV, which is continuously adjustable. Figure 1 shows the overall structure of CYCIAE-100 and Table 1 lists its key parameters. From the view of beam dynamics, the physics problem of this machine is composed of several aspects, including axial injection, central region, acceleration, stripping extraction and beam lines, which have been described in several papers published formerly[2]–[5]. The basic physics design and current construction status of machine will be briefly reviewed in the following section.

Table 1: Key Parameters of CYCIAE-100

Item	value
ion source type	multi-cusp
injection current	> 5 mA

*Work supported by the NSFC, under contract 10775185

‡tjzhang@ciae.ac.cn

†yangjianjun00@mails.tsinghua.edu.cn

number of poles	4
angle of poles	~47°
radius of poles	2000 mm
outer diameter of yoke	6160 mm
height of magnet	2820 mm
total iron weight	~415 t
field range	0.15–1.35 T
gap between hills	50–60 mm
injection energy	40 keV
rf frequency	44.32 MHz
Dee Voltage	60–120 kV
number of cavities	2
harmonic number	4
extraction type	multi-turn stripping

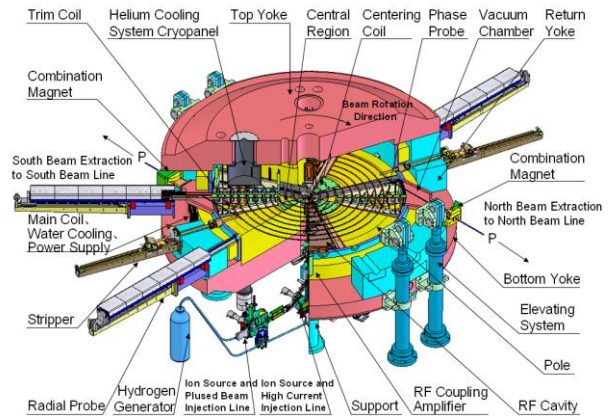


Figure 1: Sketch of the major parts of CYCIAE-100.

Beam loss is the key factor which limits the beam current of a high intensity cyclotron. Space charge effects, being one of the most significant collective effects, play an important role in high intensity cyclotron. Space charge may cause massive beam loss at the low- and middle-energy accelerator. In CYCIAE-100, the injection energy is only 40 keV and the maximal energy is 100 MeV ($\gamma=1.106$), and therefore, space charge can be remarkable under high current conditions. In addition, a common

characteristic of high intensity compact cyclotron is that dozens of radially neighboring bunches overlap at large radii region (therefore multi-turn extraction is adopted). Simulation shows that the particles in the extracted proton bunch by the stripper come from about 30 injected bunches in CYCIAE-100. This is mainly caused by two factors. Firstly, limited by space, the Dee voltage of RF cavity is relative small, accordingly, the radial gain per turn, or turn separation, is small. Simulation shows that the turn separation is reduced to only 3mm at the extraction point. Secondly, in order to achieve high current, the RF phase acceptance in the central region reaches 40° , and limited by space, no flat-top cavity is installed. Consequently, the particles at head and tail of the bunch gain quite different energy during each gap-crossing and energy spread is inevitably large, accordingly, the radial distribution range of a bunch is big. As a result, it is necessary to include the mutual interactions of radially neighboring bunches, or neighboring bunch effects for short, in space charge studies.

REVIEW OF PHYSICAL DESIGN OF CYCIAE-100

The adoption of a compact structure for CYCIAE-100 can provide strong vertical focusing to meet the requirement of intensive beam acceleration. The carbon foils are used to strip the H^- with very small beam loss during the extraction, the efficiency of which can reach 99% from our calculation and test. In order to reduce the beam loss induced by the Lorentz stripping, the hill field at the outer region is less than 1.4T. The vacuum dissociation is the other reason inducing beam loss in this H^- machine, which requires the vacuum in the tank better than 5×10^{-6} Pa on average. For the RF system, it adopts double Dee structure, and the fourth harmonics w.r.t revolution period of accelerated particles. The resonant cavity of half wave length is completely installed and fixed in the valley of the magnet.

During the acceleration, the beam corresponding to different energy has fixed equilibrium orbit. The betatron oscillations around equilibrium orbits at different energies up to 100MeV are investigated in detail with many magnet structures and their fields. The vertical oscillation frequency is higher than 0.5 at most of the acceleration region and towards 0.7. This is of advantage to upgrade the beam intensity later on.

After the static study, the accelerating beam dynamics is also done in detail. The transverse ellipses along the AEO are matched step by step from the central region to final energy. The vertical beam profiles with different RF phases are simulated and multi-particle tracking is carried out to control the beam loss in this small hill gap machine. In practice, the imperfection magnetic field exists and the deviation of the orbit centre takes place. Under the influence of the 1st and 2nd harmonic fields, the beam will oscillate about the deviated orbit centre and thus result in its radial dimension growth. In order to obtain a good

beam quality, the simulation results show that the magnetic field of the cyclotron should satisfy the following condition: the 1st harmonic is less than 2 gauss, the 2nd harmonic is less than 40 G, and the gradient of the 2nd harmonic is less than 8 G/cm. To comply with the requirement for being isochronous, it is demanded that the deviation between the measured field at medial plane and the idea field be approximately below 1.05×10^{-4} .

NEW PHYSICAL MODEL AND RELATIVE CODE DEVELOPMENT

In 2008, a 3D physical model was built to include neighboring bunch effects in the high intensity separated-sector cyclotrons using single turn extraction [6]. In this model, at the beginning only a single bunch is tracked until the turn separation is small enough, and then a new bunch is injected per revolution period. Two parameters M and N_B are introduced to set the time of starting injecting new bunches and the maximal bunch number respectively. The proper settings of these two parameters are crucial to precisely evaluate neighboring bunches effects. In this model N_B must be an odd number. Our study object is the central bunch and the other $(N_B - 1)$ bunches are auxiliary. Therefore we call it “central bunch” model hereafter. This model was implemented in the code OPAL-CYCL[6], which is one of the flavours of OPAL framework[7] and tracks particles with 3D space charge including neighboring bunches effects in cyclotrons, with time as the independent variable. Simulation results using OPAL-CYCL shows that the setting with $N_B = 9$ and $M = 4.5$ gives convergent results for the PSI 590MeV Ring cyclotron with 3mA beam current[6].

However, it is quite difficult to study the neighboring bunch effects in CYCIAE-100 and other similar compact cyclotrons by using “central bunch” model, because in high intensity compact cyclotrons, all the radially neighboring bunches overlap heavily and there is no clear borderline between the central bunch and the others bunches. On top of this, the study on central bunch is worthless and insufficient, because in this type of machines, multi-turn stripping extraction is used and the extracted beam include the contributions of dozens of the neighboring bunches. From the view of extraction, study on the behaviour of the extracted bunch is more significant and valuable.

“Start-to-Stop” model

Recently a new physical model was established for compact cyclotrons such as CYCIAE-100. In this model the multi-bunch tracking is divided into three stages in sequence: startup-running-stop. This model imitates the stages of an accelerator’s operation period and called “Start-to-Stop” model:

- Startup stage: with bunch injected and without bunch extracted

A new bunch is generated in the injection point per revolution period. The existing bunches has not fill

all the turns and the first bunch has not reached the stripper, so no particle is extracted.

- **Running stage:** with bunch injected and bunch extracted

A new bunch is generated in the injection point per revolution period; meanwhile, the particles which have reached the stripper are extracted. When a macro-particle is extracted, its tracking is finished. Its phase space variables and the extraction time are written into disk and the memory blocks which stores its relative data are free. When the extracted particle number is equal to the injected particle number, the simulation is running under the steady state.

- **Stop stage:** without bunch injected, with bunch extracted

No bunch is injected, which imitates the ion source stops providing particle and the accelerator is still working. During every revolution period, there are still particles reaching the stripper and being extracted. The total particle number is constantly reduced until all the existing particles are extracted, then the entire simulation is finished.

Comparing with the “central bunch” model, this model still holds the original meaning of the parameter M ; meanwhile, the parameter N_B here means the total injected bunch number from start to stop. Its value is a much larger one, i.e. larger than the total turn number of a cyclotron at least.

When space charge field is solved using quasi-static approximation, one needs to assure the relative motion of particles is non-relativistic, i.e. in the beam rest frame, the formula

$$\eta \equiv \frac{\Delta p}{m_0 c} \ll 1 \quad (1)$$

should be fulfilled, where Δp is the momentum spread in the beam rest frame. In the PSI Ring and similar separated-sector cyclotrons with single turn extraction, the energy gain per turn is large and energy spread is small. To meet the requirement of formula (1), “central beam” model follows the rule that each energy bin corresponds one bunch. Whereas, in the CYCIAE-100 and similar compact cyclotrons, the energy gain per turn is relatively small and energy spread is relatively large. Therefore in the “Start-to-Stop” model, the adaptive energy binning technology, which is usually adopted to deal with a single beam with large momentum spread[8], is introduced in the multiple beams issue of compact cyclotrons. The momentum range of each energy bin η_{bin} to some value far smaller than 1, then the momentum of the k th energy bin is

$$p_{k,\text{bin}} = \sinh(k\eta_{\text{bin}} + \sinh^{-1} p_{\text{min}}) \quad (2)$$

Where p_{min} is the lowest momentum of the all the existing particles and it defines the first energy bin. The bin index k of a given particle with momentum p_i is given by

$$k = \left\lfloor \frac{\sinh^{-1} p_i - \sinh^{-1} p_{\text{min}}}{\eta_{\text{bin}}} \right\rfloor \quad (3)$$

After binning we perform the Lorentz transformation, calculate the space charge field using FFT based solver and perform back-transformation for each bin respectively. Finally the field data is summed up to give the total space charge force imposed on each particle.

Recent development of OPAL-CYCL

Recently we implemented above “Start-to-Stop” model in the code OPAL-CYCL, to make it applicable to high intensity compact cyclotrons such as CYCIAE-100.

In order to achieve the best balance between accuracy, stability and efficiency, recently a second order Leap-Frog integrator is implemented into OPAL-CYCL.

In addition, the computation load and memory load on computer nodes now are better balanced by taking advantages of the dynamic mesh grid repartition functionality of IP²L[9]. Figure 2 shows the speedup at different repartition frequency for a production run setup. The timings were obtained on the Cray XT5 of the CSCS, Switzerland. Each of the computer nodes consists of 2 quad-core AMD Opteron 2.4 GHz Shanghai processors giving 8 cores in total per node with 16 GBytes of memory. We can see a significant gain of speedup when the repartition is performed once per 20 integration steps.

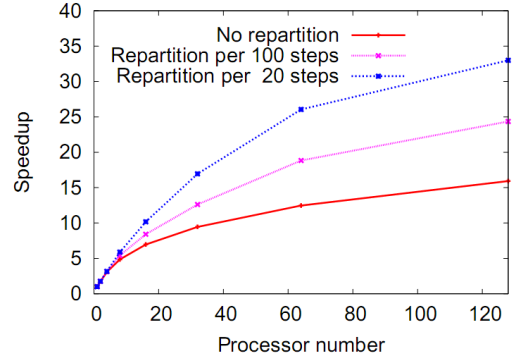


Figure 2: Speedup of OPAL-CYCL as a function of processors at different repartition frequency for a production run setup.

START-TO-STOP SIMULATION

In this section, we utilize the “Start-to-Stop” model to study the high intensity issues in CYCIAE-100. The motivation for this study is to predict the beam’s behavior at different current level so as to help the builders achieve the aim of the project. Since the aimed beam current of this cyclotron is 0.2–0.5 mA and it is planed to improve the current to 1mA during future upgrade, the simulation is done for 0mA, 0.2mA, 0.5mA and 1mA beam current respectively and the results are compared and analysis together.

In CYCIAE-100, the 40keV DC beam is transported from the ion source to the central region by the axial injection line and the spiral inflector. After the phase

selection and acceleration during the first several turns in the central region, at the exit of central region we get a bunched beam with energy equal to 1.49MeV and initial position $R = 23.14\text{cm}$, $\theta = 0^\circ$. The simulation starts from this position. Since it is hard to obtain the precise distribution of the beam at present, a 6D Gaussian ellipsoid distribution is assumed as the initial distribution. The initial phase width (6σ) is set to 6° and initial energy spread is zero. Both on the radial and vertical phase space, the particle distributions match the eigen-ellipses and the rms emittances are 1.2π mm-mrad and 0.4π mm-mrad respectively. The beam sizes (6σ) on both directions are set to 12mm and the initial distribution is assumed uncorrelated in phase space.

In the simulation ten thousand macroparticles per bunch are employed and ultimately more than two million macroparticles are injected under the running stage. Considering all bunches should lie along the radial direction approximately, the mesh size along radial direction is set to 256 and on the two other directions is 32. Figure 3–Figure 6 shows the results given by the simulation.

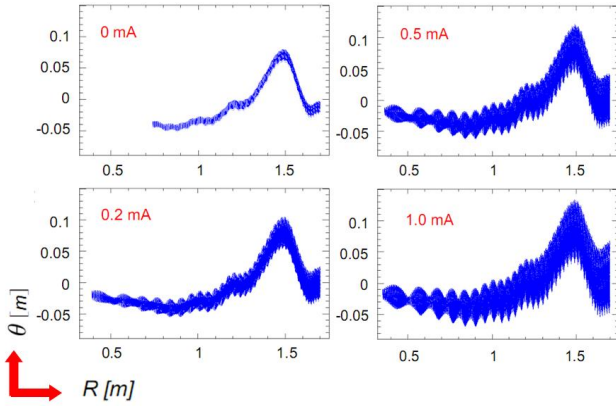
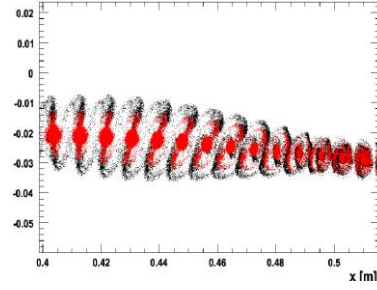


Figure 3: Top view snapshot at the time point bunches crossing 0° azimuth under the steady state

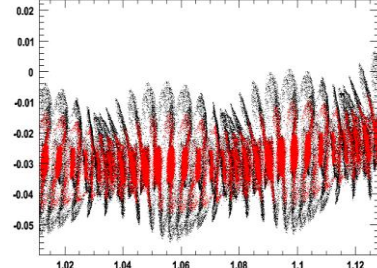
Figure 3 shows the top view snapshot at the time point bunches crossing 0° azimuth under the steady state of running stage. It is clear from the simulation that the phase width becomes larger along with the current increasing. In the final turn the phase width are 2.5° , 6.5° , 10.4° and 15.0° for 0mA, 0.2mA, 0.5mA and 1mA respectively. It is noted that despite the same value of M , the transition from single bunch simulation to multi-bunch simulation happens at earlier turn for higher current, because neighboring bunches effects rise up at earlier turn for high current.

In Figure 4 three typical local regions of Figure 3 are zoomed in and the distribution for 0.2mA and 0.5mA are shown together for comparison. The “vortex” motion can be observed clearly at the low energy region. Looking into Figure 3 and Figure 4, we can find an interesting phenomenon that the beam phase width oscillates along the radius when the current is larger than 0.2mA, which is believed by the mismatch caused by space charge and

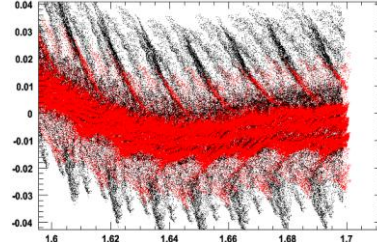
neighboring bunch effects. This will be studied in detail in the future.



(a) 4~8MeV



(b) 31~40MeV



(c) 86~100MeV

Figure 4: Zoom-in plot (using the same scales) of Figure 3 for 0.2mA (red) and 0.5mA (black) beam currents

In CYCIAE-100, the stripping probe is placed at the position $R = 187.55\text{cm}$, $\theta = 59.6^\circ$ for 100MeV beam extraction. Two electrons of a H^- particle are stripped by the carbon foil and the resultant proton beam is transported out of cyclotron. The distribution of extracted beam is crucial for the beam line design. Therefore the influence of space charge on the extracted beam is an interesting problem. Figure 5 shows the simulation result of the R - Z distribution of the extracted beam for different beam currents and Figure 6 shows its histograms along the radial and axial directions respectively. The beam sizes on both directions are expanded along with the current increase. On the axial directions more and more halo particles are generated along with the current increase. So the extraction beam line designer should take the beam current into account during beam upgrade so as to improve the transmission efficiency and reduce beam loss. Fortunately, up to 1mA, the axial beam size is less than 1.5cm, still smaller than the gap between hills (5cm) and the vertical distance of RF linear (4cm) at this region. Therefore, during the acceleration and extraction no massive beam loss is caused by the space charge and neighboring bunch effects under 1mA current.

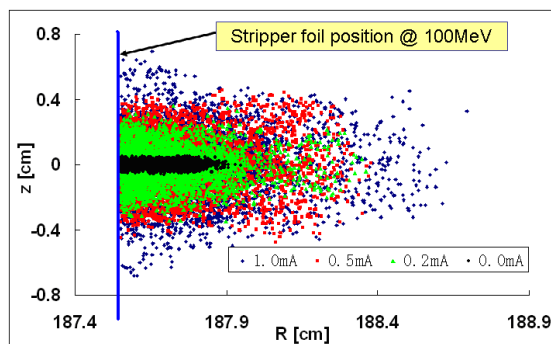


Figure 5: R - Z distribution of the extracted beam for different beam currents under the steady state.

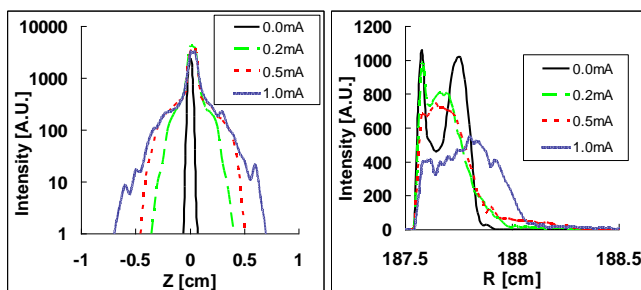


Figure 6: Comparisons of the histograms along the Z and R directions for the extracted beam for different beam currents under the steady state.

CONCLUSION AND FUTURE WORKS

The physics design for CYCIAE-100 has been accomplished. A brief review of the physics design on this cyclotron is presented. Then a new established “Start-to-Stop” model, which was recently implemented in OPAL-CYCL, is depicted. The “Start-to-Stop” simulation results of CYCIAE-100 shows that space charge and neighboring bunch effects will lengthen the phase width during acceleration and expand the beam sizes of the extracted beam both in the axial and radial directions. However, no massive beam loss is caused under 1mA current.

To improve the accuracy of the simulation, further investigation will be performed on the beam evolution at

the central region to obtain the more practical initial conditions. In addition, limited by the computer resource, only 10 thousand macroparticles per bunch were employed to do the above simulation, and the mesh size is $256 \times 32 \times 32$. A larger scale simulation is planned to improve the precision of the simulation. After that, the current limit of CYCIAE-100 will be also studied by the simulations.

ACKNOWLEDGMENTS

The authors would like to thank A. Adelman (PSI) for the collaboration on the OPAL-CYCL development and T. Schietinger (PSI) for providing the visualization tool HSPartROOT. We are also indebted to Y.Z. Lin and C.X. Tang (Tsinghua University) for useful discussions on the new “Start-to-Stop” model.

Some of the computations were performed on the Cray XT5 at CSCS, Switzerland and some were performed on the cluster PANDA which was just finished at CIAE, China.

REFERENCES

- [1] T.J. Zhang, Z.G. Li, C.J. Chu, et al., 17th ICC'04, Tokyo, 2004, p. 497
- [2] T.J. Zhang, Z.G. Li, J.Q. Zhong, et al., Chinese Physics C, 33(S2) (2009), p.33
- [3] H.J. Yao, T.J. Zhang, Y.L. Lu, et al., 18th ICC'07, Catania, 2007, p. 57
- [4] S.Z. An, F.P. Guan, H.D. Xie et al., Chinese Physics C, 33(S2) (2009), p.42
- [5] S.M. Wei et al., NIM-B, 266(2008), p. 4697
- [6] J.J. Yang, A. Adelman, M. Humbel, et al., Proc. HB'08, Nashville, 2008, in press.
- [7] A. Adelman, Ch. Kraus, Y. Ineichen and J. J. Yang, PSI-PR-08-02, Paul Scherrer Institut, 2008
- [8] G. Fubiani, J. Qiang, E. Esarey, et al., Phys. Rev. ST Accel. Beams, 2006, 9(6):064402.
- [9] A. Adelman, PSI-PR-09-05, Paul Scherrer Institut, 2009

SPACE CHARGE SIMULATIONS FOR ISIS

BG Pine, DJ Adams, B Jones, CM Warsop, RE Williamson
Rutherford Appleton Laboratory (STFC), Oxfordshire, UK.

Abstract

The ISIS Facility at the Rutherford Appleton Laboratory in the UK produces intense neutron and muon beams for condensed matter research. It is based on a 50 Hz proton synchrotron which accelerates $\sim 3 \times 10^{13}$ protons per pulse (ppp) from 70 to 800 MeV, corresponding to beam powers of ~ 0.2 MW. Studies are under way for major upgrades in the Megawatt regime. Underpinning this programme of operations and upgrades is a study of the high intensity effects that impose limitations on beam power.

The behaviour of the beam in the 50 Hz rapid cycling synchrotron (RCS) is largely characterised by high space charge levels and the effects of fast ramping acceleration. High intensity effects are of particular importance as they drive beam loss, but are not fully understood with only limited analytical models available. This paper reviews several methods by which these effects are explored numerically on ISIS, and compares them where possible with experimental or analytical results. In particular we outline development of a new space charge code Set, which is designed to address key issues on ISIS and similar RCS machines.

INTRODUCTION

ISIS high intensity operation is restricted by beam loss, as irradiation of equipment limits access for essential maintenance. Understanding beam loss is therefore of vital importance, however due to the complex interactions between the beam particles and their environment such understanding is challenging both analytically and numerically.

The ISIS Synchrotron Group is actively studying high intensity effects of the beam in a number of different ways, both to improve performance of the accelerator and also to enable the design of upgrades which can achieve significantly higher beam intensities. Aspects of this work are reported here, including closely related profile monitor simulations, injection painting, beam dynamics, half integer studies and developments of codes.

PROFILE MONITOR

ISIS profile monitors are important for studies of injection painting, space charge, beam halo, betatron motion and instabilities, as well as suffering space charge effects of their own. The profile monitors on ISIS use ions, liberated from the residual gas by passage of the beam, to reconstruct transverse beam distributions. A (near) uniform electric field, perpendicular to the direction of the beam, accelerates ions to a suitable detector. The number of ions detected is assumed to be proportional to the local beam intensity. This process

suffers from two major sources of error: 1) irregularity in the electric field used to drive the ions to the detector; 2) broadening effects produced by the space charge field of the beam, which at high intensities can dominate the measured profile width. Fortunately correction schemes for both of these phenomena have been found, and are discussed below.

Drift field effects

A model of an ISIS profile monitor (see Figure 1) was constructed in CST Studio Suite [1]. Figure 2 displays the potential produced by the electrodes in both transverse and longitudinal cross-sections. As can be seen the required linear field is not achieved perfectly – both transverse and longitudinal sections show deviation from the ideal behaviour.

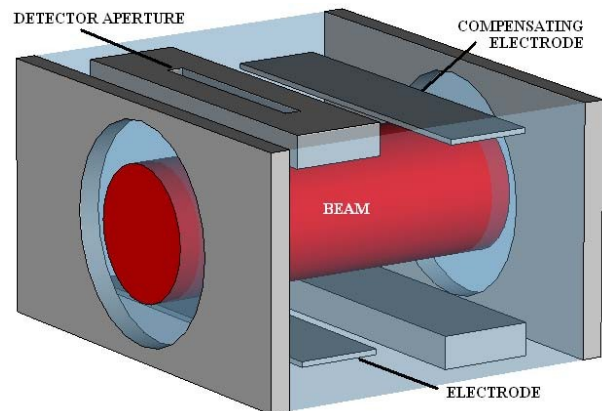


Figure 1: Residual Gas Profile Monitor.

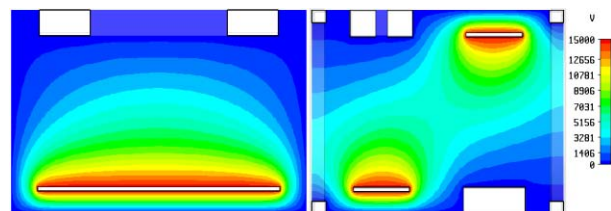


Figure 2: Electrostatic potentials: transverse - left, longitudinal - right.

In order to study these effects in more detail, potentials were calculated then extracted from the CST model, and used as the field source in a specially developed 2D particle tracker. Realistic beam distributions (parabolic, elliptic) were used as the source of particles. The results showed that a simple scaling correction was effective for reasonably well centred and behaved beam distributions [2]. 3D simulations [3] showed considerably more complex behaviour, as particles may oscillate along longitudinal field saddle points between the two electrodes in the monitor body. On investigation however this more complex behaviour could be accounted for by a

modification of the scaling rule calculated for the 2D case. Figure 3 shows the result of a correction using the scaling rule on a simulated profile including what the ideal measurement would be, and the corresponding corrected profile.

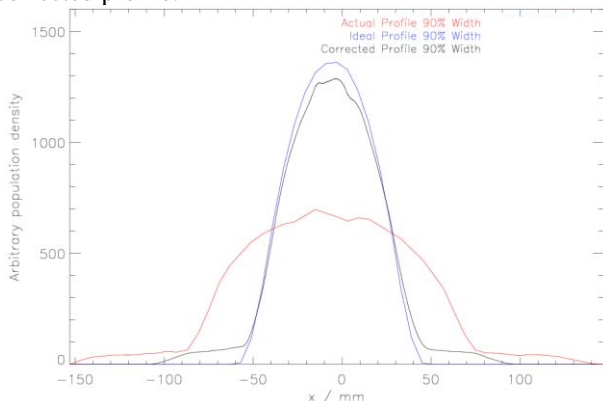


Figure 3: “Measured” - red, ideal - blue and corrected - black beam profiles, all from simulation.

Space charge broadening

Space charge broadening effects have been studied both by simulation [2, 3] and experiment [3, 4]. Simple models suggest that the measured profile width ought to be inversely proportional to the magnitude of the drift field. This was confirmed both experimentally and in simulations, and holds for all percentage widths of the beam, once modified for relativistic effects. An experimental measurement of the profile width relative to applied drift field voltage is displayed in Figure 4.

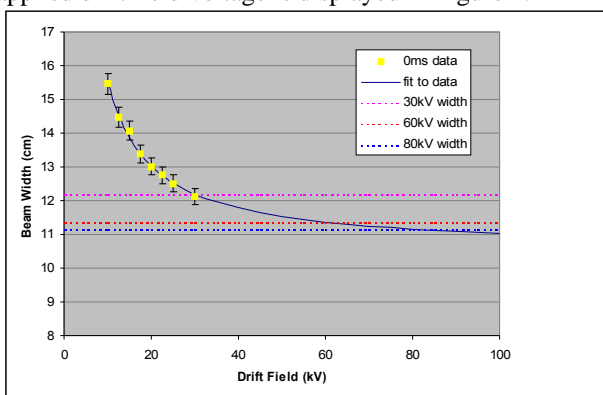


Figure 4: Measured profile width versus drift field voltage at 0 ms and a beam intensity of 2.12×10^{13} ppp.

INJECTION

The ISIS multi-turn, charge exchange injection process accumulates $\sim 3 \times 10^{13}$ ppp over 130 turns, or 200 μ s, at a constant energy of 71.5 ± 0.5 MeV. The injection region is located in the middle of a straight section, to minimise interaction with the rest of the machine. The beam is effectively unbunched, as RF runs at low levels throughout injection. Injection begins -0.4 ms before field minimum of the main dipole magnets in the ring, i.e. on a falling magnet field (ISIS runs with a sinusoidal main magnet field). This naturally provides horizontal painting

Computer Codes (Design, Simulation, Field Calculation)

of the beam through the horizontal dispersion at the foil. Vertical painting is achieved using a programmable sweeper magnet in the injection line (see Figure 5). For standard operations vertical sweeping is anti-correlated relative to the horizontal, and is intended to produce a hollow beam distribution which will then be filled in under the action of the beam's space charge.

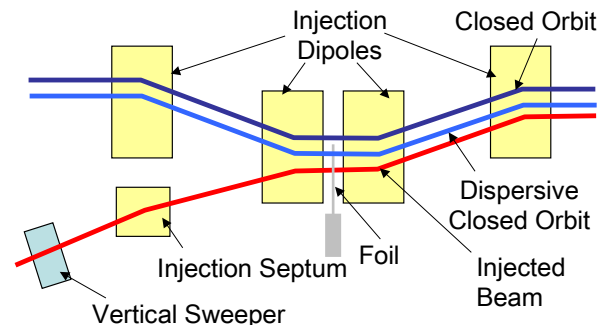


Figure 5: ISIS injection system.

The ISIS injection painting process has also been studied [5, 6] with the ORBIT simulation code [7]. Figure 6 shows betatron amplitudes during injection, measured using chopped beams, which show the anti-correlated painting dynamics. These are compared with simulated values.

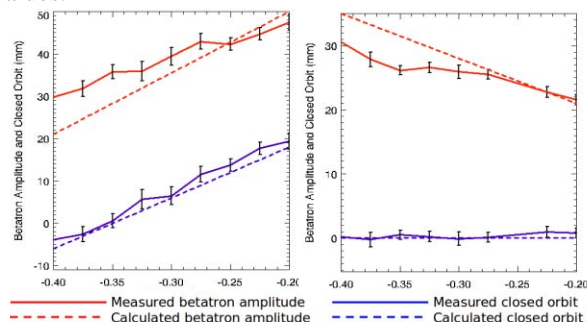


Figure 6: Anti-correlated painting on ISIS.

Further simulations include the action of space charge on the beam profile during injection. The results have been compared with experimental data taken with the profile monitors mentioned above, at both low (2.5×10^{12} ppp) and high (2.5×10^{13} ppp) proton intensities. The experimental data were corrected for drift field effects and the widening effect of space charge on the measured profile, and the results compared with profiles extracted from the simulation data (Figure 7). As can be seen there is a good agreement between the simulated and experimental results, which show the hollow painted beam being filled in under the influence of space charge.

These simulations were used to generate the beam phase space distributions shown in Figure 8. Transverse calculations include space charge; longitudinal dynamics reproduce expected horizontal movement, but without space charge.

Work is well advanced for longitudinal simulations of ISIS in ORBIT with space charge. Essential features of the beam loss versus time profile have been reproduced.

Models of the collimator system and the MICE [8] internal target have also reproduced observed loss patterns reasonably.

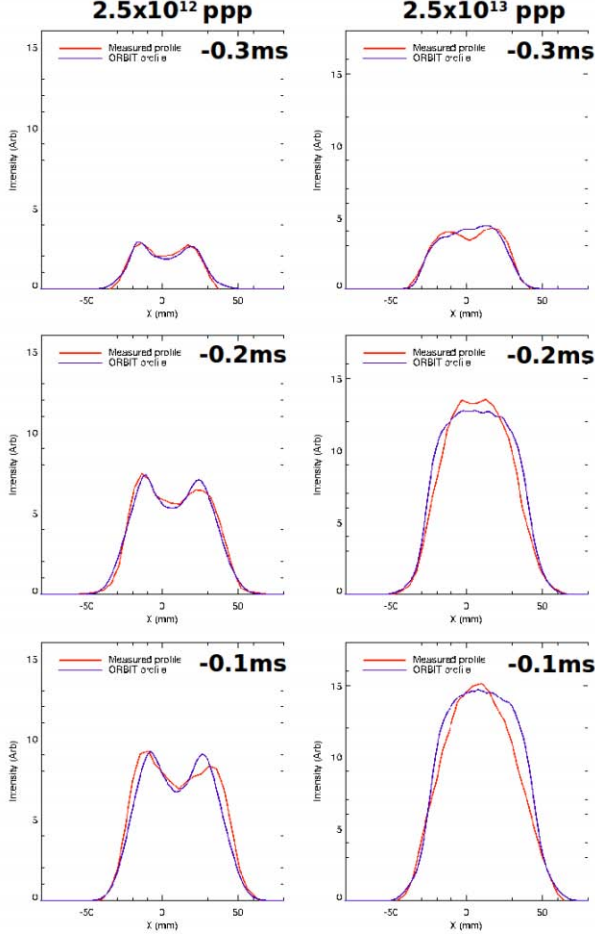


Figure 7: Anti-correlated painting – horizontal profiles.

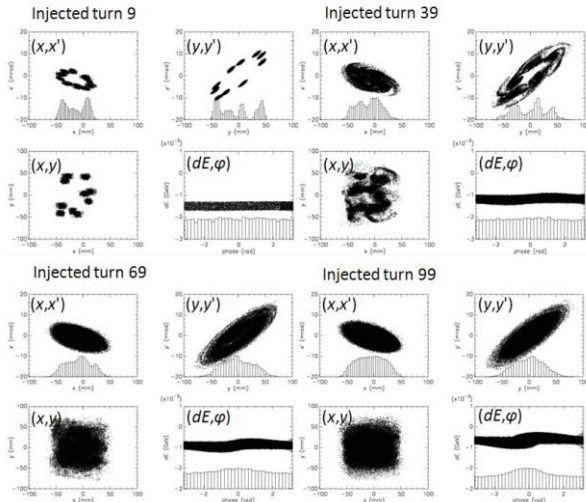


Figure 8: Simulations of injection: Normal anti-correlated case at 2.5×10^{13} ppp.

HALF INTEGER

Transverse space charge and half integer resonance have been identified as the main loss mechanisms on the Computer Codes (Design, Simulation, Field Calculation)

ISIS RCS. During injection and trapping, bunch intensity reaches a maximum and incoherent tune depression due to space charge can reach ~ 0.4 in both planes. The beam is pushed towards the $2Q_h = 8$ and $2Q_v = 7$ resonance lines. Initial work has focused on analytical studies and 2D simulations to confirm the predicted effects [9, 10].

2D ORBIT simulations were run with an RMS matched waterbag beam circulating in an alternating gradient model of the ISIS lattice. The driving terms (representing the main resonance lines) reproduced many expectations from studies of the envelope equation. In particular, stability was observed on the incoherent resonance. Emittance growth started to rise before the coherent resonance condition occurred however (Figure 9). This suggests that significant emittance growth, and hence the practical space charge limit, lie between the incoherent and coherent predictions.

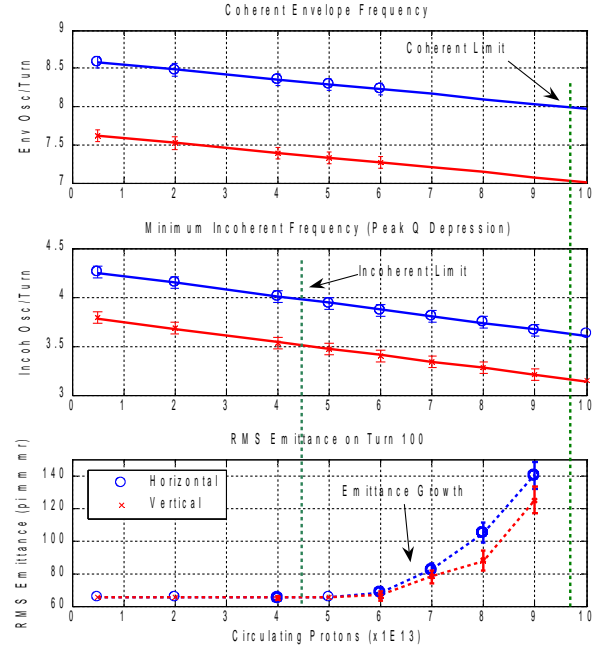


Figure 9: Intensity limits associated with half integer resonance and RMS emittance growth.

More detailed studies comparing halo structure as predicted by ORBIT and available theory have confirmed expected behaviour [11]. Simulations including the effects of momentum spread and longitudinal motion are in development.

Extensive experimental work is under way, running ISIS in storage ring mode (SRM) without RF or the AC component of the main magnets, enabling a beam to be stored for many turns as a coasting unbunched beam. This is ideal for studying 2D aspects of beam behaviour. SRM is becoming a rich experimental system for studying high intensity effects [12]. Figure 10 shows that beam intensity reduction and corresponding profile growth coincide with the start of an instability. Simulation and experimental studies are presently under way to understand in more detail the action of both instabilities and envelope resonance under high space charge. This

work also clearly depends on a detailed knowledge of the profile monitors used to take measurements.

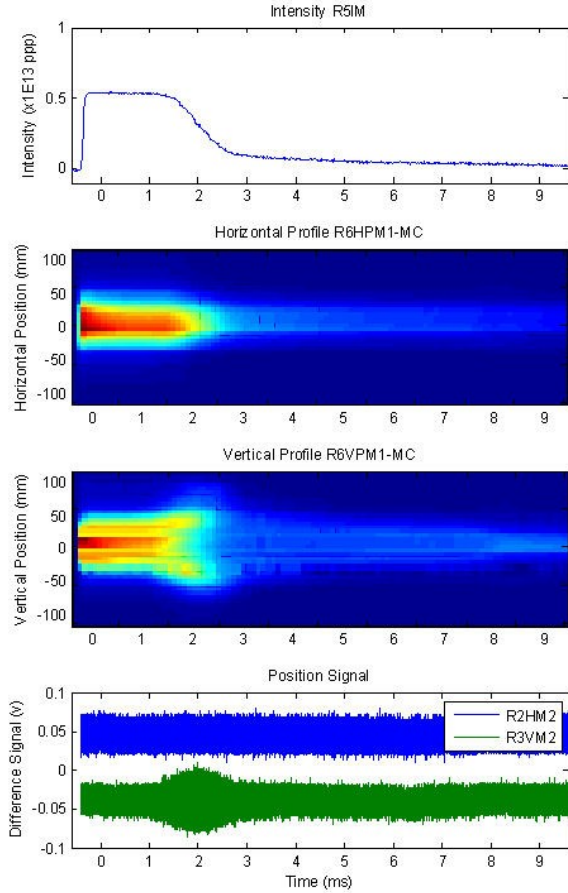


Figure 10: Beam intensity, horizontal and vertical profiles, horizontal and vertical position signals vs time.

SET

A new code Set is under development at ISIS. This code is intended to supplement the use of ORBIT for 2D and 3D beam tracking simulations, as a tool that can be readily modified and redeployed as required to meet a given purpose. In particular, the focus is on the challenges of the ISIS RCS, including image forces from the unique profiled vacuum vessel, halo predictions, 2D and 3D RCS space charge effects and overall to understand and predict beam loss. Set works using either MAD input data or its own matrix routines for generating lattices, and has an FFT based Poisson-solver for calculating the beam's space charge. Early simulation work [10, 11] focused on replicating ORBIT results for the half integer resonance. Example results for the ISIS lattice (2D, coasting beam) driven with a $2Q_v = 7$ resonance are shown in Figures 11, 12 and 13. Figure 11 shows Set and ORBIT envelope frequencies as the intensity is swept from 1 – 14E13. Figure 12 shows the incoherent tune footprints after 100 turns, as the intensity is varied from 6E12 ppp to 13E13 ppp. Figure 13 shows beam phase and real space on the 100th turn for an intensity of 6E13.

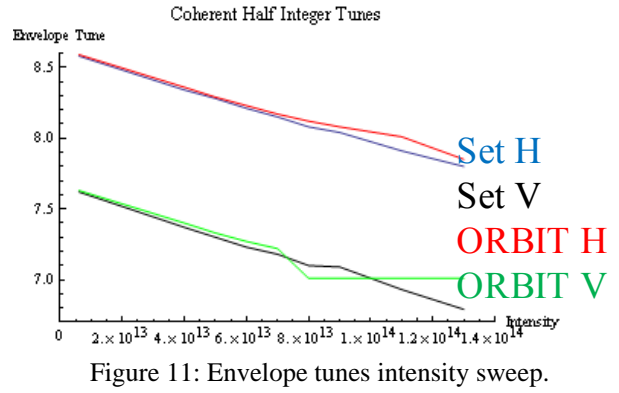


Figure 11: Envelope tunes intensity sweep.

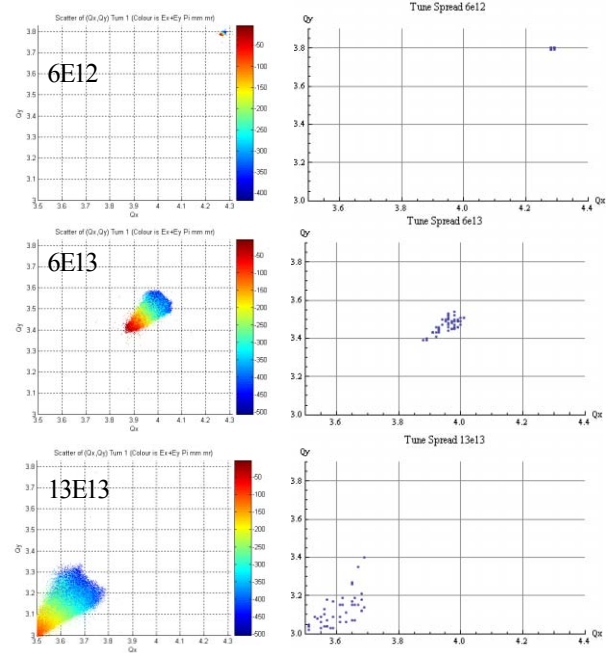


Figure 12: Incoherent tune comparison ORBIT - left, Set - right.

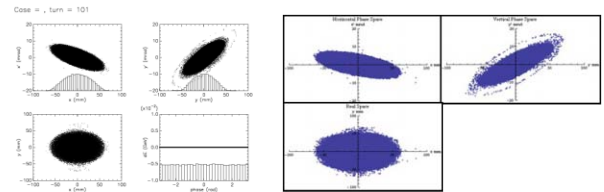


Figure 13: Phase space on 100th turn ORBIT - left and Set - right.

Set has been used to study tune shifts from image forces and closed orbits [13] and the results compared with Laslett theory (Figure 14). Direct space charge should have no influence on the coherent dipole tune, as the charge distribution of the beam moves with the centre of charge. However image forces will affect the coherent tune, as the centre of charge does move relative to the vacuum vessel. This is of particular interest on ISIS due to the vacuum vessel which follows the design beta function of the beam. At high intensities the machine is very sensitive to closed orbit changes, which may indicate beam loss driven by image forces. Present upgrade studies

are investigating the benefits of increasing injection energy from 70 to 180 MeV. Set has been used to study the space charge limitations at this higher energy. The half integer simulations carried out for the nominal ISIS ring were reproduced, but for an injection energy of 180 MeV rather than 70 MeV. Simple scaling of the space charge force indicates peak intensities should increase by a factor of 3 over 70 - 180 MeV. As can be seen from Figure 15, the RMS emittance begins to rise at 3 times the intensity seen in the previous case.

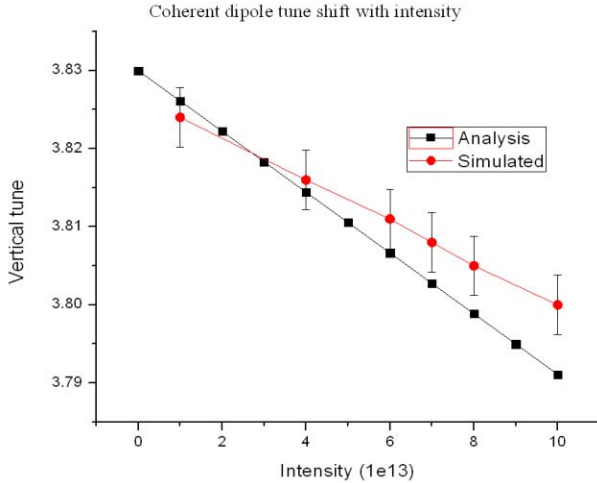


Figure 14: Simulated versus analytical tune shift.

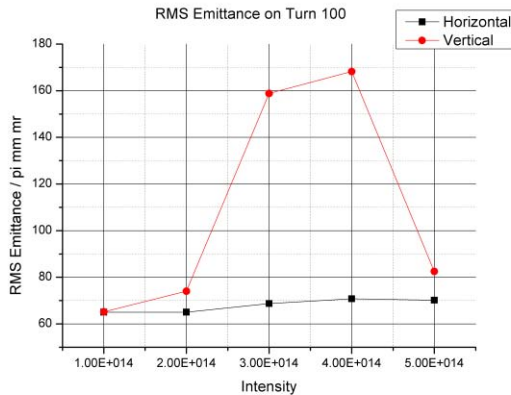


Figure 15: RMS emittance growth associated with half integer resonance for 180 MeV injection energy.

Image forces become more significant when the beam is executing a closed orbit, as image forces from the beam pipe only cancel when the beam is well centred. Figures 16 and 17 show the results of simulation runs including half integer driving terms from the trim quadrupoles (ISIS has special programmable quadrupoles distinct from the main lattice), and also an angular kick once per turn. Each set of simulations was run twice, to allow the resulting perturbed beam distribution to be matched into the lattice. Figure 16 shows the variation of closed orbit (RMS position) around the ring as a function of intensity, from 1 - 5E14 ppp. Figure 17 compares RMS emittance with and without a closed orbit at an intensity of 2E14 ppp. Image forces are clearly influencing the behaviour of the beam,

much as we expect on ISIS. A more complete analysis, and eventually experimental work on the ISIS synchrotron, are to follow.

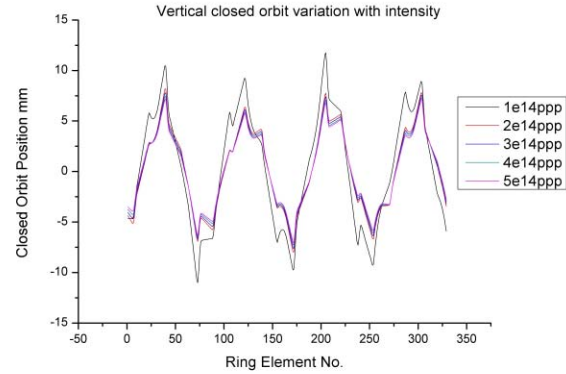


Figure 16: Matched closed orbit variation with intensity.

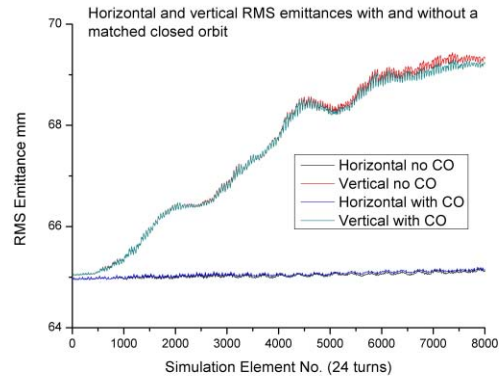


Figure 17: RMS emittance with and without matched closed orbit errors for 2E14 ppp intensity.

FUTURE WORK

ORBIT

Extensive work has been carried out developing a 3D model of ISIS in ORBIT that can run on a parallel system of computers, the SCARF cluster at Rutherford Appleton Laboratory [14]. This is largely complete, including injection and longitudinal motion with space charge, though fundamental RF only (simulations with the second harmonic RF are in development). A recent area of progress involved modifying ORBIT run scripts to allow for the ramping ISIS Q values, as shown in Figure 18. ISIS Q values are ramped to move the tunes away from dangerous resonances during injection and trapping. Previous simulations had used the fixed Q values (close to the initial values from Figure 18). Beforehand the lattice creation algorithms had to be run for each turn of the simulation for the new Q values. Beam was run around the first turn, beam distributions saved and then input into the next turn with a modified lattice. Results for the 99% emittance of the beam, comparing constant Q with ramping Q are shown in Figure 19. It appears that with the real (ramping) Q values, emittance in the horizontal

plane is increased relative to simulations with a constant Q , whereas it is decreased in the vertical plane. These are preliminary results and it has not yet been ascertained what driving terms or growth mechanisms might be responsible for these phenomena.

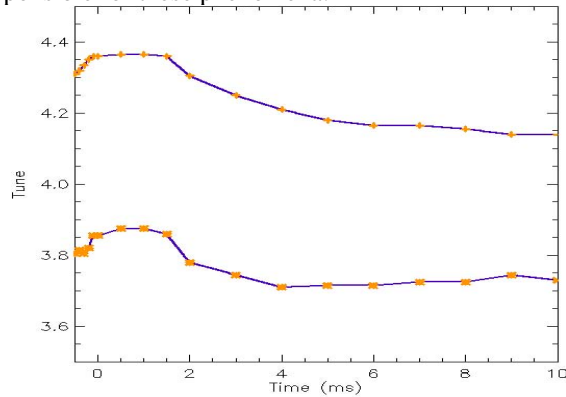


Figure 18: ISIS Q values during the cycle.

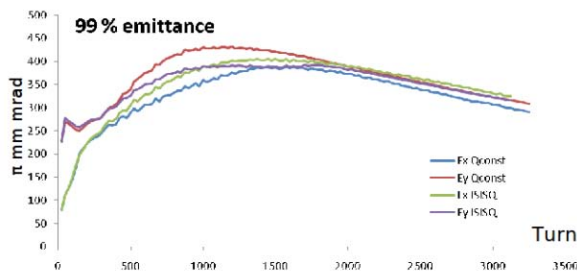


Figure 19: 99% emittance evolution comparing ramped Q with constant Q values.

Set

Work is also under way on a 1D longitudinal code which includes space charge and impedances. This will eventually be added to Set to make it fully 3D. Results are shown in Figure 20 for bunch length and phase convergence versus macroparticle number.

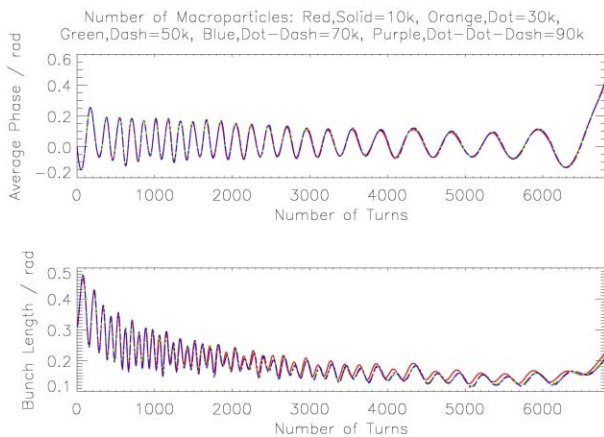


Figure 20: 1D convergence tests showing results for phase and bunch length.

A parallel version of 2D Set has been implemented, and successfully run on SCARF. Further work will add a

realistic injection scheme, including the effect of foil scattering. The long term goal is to carry out a full simulation of the ISIS cycle and recreate the beam loss patterns seen on the real machine.

SUMMARY

Understanding space charge, and hence beam loss, is essential for the operation of a high intensity RCS, and even more important for the design of an upgrade.

Space charge simulations are being developed in a number of crucial areas on ISIS: to correct profile monitor measurements, to study injection processes, for the examination of half integer resonance effects, for longitudinal dynamics, and finally for 3D studies of loss patterns on ISIS. A new code Set is being developed to enable further study of key beam dynamics issues that are important for ISIS, such as image effects and any dominant loss mechanisms.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge essential collaborations with ISIS Diagnostics, ISIS Operations and STFC e-Science. Thanks also to SNS/ORNL for continuing use of the ORBIT code.

REFERENCES

- [1] <http://www.cst.com>
- [2] BG Pine *et al*, "Modelling of Diagnostics for Space Charge Studies on the ISIS Synchrotron", Proc. of EPAC06, Edinburgh, UK.
- [3] RE Williamson *et al*, "Analysis of Measurement Errors in Residual Gas Ionisation Profile Monitors in a High Intensity Proton Beam", Proc. of EPAC08, Genoa, Italy.
- [4] SJ Payne *et al*, "Investigation into the Relationship Between the Drift Field Voltage and Measured Beam Width on the ISIS Ring Profile Monitor System", Rutherford Appleton Laboratory ISIS Machine Physics Report, IADM/MPS/2.1/03 (2003).
- [5] B Jones *et al*, "Injection studies on the ISIS Synchrotron", Proc. of PAC07, Albuquerque, USA.
- [6] B Jones *et al*, "Injection optimisation on the ISIS Synchrotron", Proc. of EPAC08, Genoa, Italy.
- [7] JA Holmes *et al*, ORBIT User Manual, ORNL Tech. Note SNS/ORNL/AP/011.
- [8] <http://mice.iit.edu/>
- [9] CM Warsop, "Transverse space charge studies for the ISIS Synchrotron", Proc. of EPAC06, Edinburgh, UK.
- [10] CM Warsop *et al*, "Studies of space charge loss mechanisms on the ISIS synchrotron", Proc. of PAC07, Albuquerque, USA.
- [11] CM Warsop *et al*, "Space charge loss mechanisms associated with half integer resonance on the ISIS synchrotron", Proc. of EPAC08, Genoa, Italy.
- [12] CM Warsop *et al*, "Studies of space charge loss mechanisms associated with half integer resonance on the ISIS RCS", Proc. of PAC09, Vancouver, Canada.
- [13] BG Pine *et al*, "Image simulations on the ISIS synchrotron", Proc. of EPAC08, Genoa, Italy.
- [14] <http://sct.esc.rl.ac.uk/SCARF/index.html>

AN INTEGRATED BEAM OPTICS-NUCLEAR PROCESSES FRAMEWORK IN COSY INFINITY AND ITS APPLICATIONS TO FRIB*

B. Erdelyi#, Northern Illinois University, DeKalb, IL 60115. and ANL, Argonne, IL 60439

L. Bandura, NSCL, Michigan State University, East Lansing, MI 48824

Abstract

When faced with the challenge of the design optimization of a charged particle beam system involving beam-material interactions, a framework is needed that seamlessly integrates the following tasks: 1) high order accurate and efficient beam optics, 2) a suite of codes that model the atomic and nuclear interactions between the beam and matter, and 3) the option to run many different optimization strategies at the code language level with a variety of user-defined objectives. To this end, we developed a framework in COSY Infinity with these characteristics and which can be run in two modes: map mode and a hybrid map-Monte Carlo mode. The code, its applications to the FRIB, and plans involving large-scale computing will be presented.

INTRODUCTION

The next generation of nuclear physics research will require advanced exotic beam facilities based on heavy ion driver accelerators. There are many next-generation facilities that are currently under commissioning, construction, or envisioned [1-5]. Included amongst these is the future Facility for Rare Isotope Beams (FRIB) at the National Superconducting Cyclotron Lab at Michigan State University. These facilities are capable of producing exotic beams composed of rare nuclei in large quantities. The exotic isotopes are produced via projectile fragmentation and fission in targets. High-performance fragment separators, a key component of all rare isotope facilities, consist of superconducting magnets that are used for the capture, selection, and transport of rare isotopes. Large aperture magnets are necessary in order to accept rare isotope beams with large emittances resulting from their production mechanism.

The beam optics code COSY INFINITY uses powerful differential algebraic (DA) techniques for computing the dynamics of the beam in the fragment separator through high order transfer maps [6]. However, until now it has lacked the ability to calculate the beam-material interactions occurring in the target and energy absorbers. Here, a hybrid map-Monte Carlo code has been developed and integrated into COSY in order to calculate these interactions. The code tracks the fragmentation and fission of the beam in target and absorber material while computing energy loss and energy and angular straggling as well as charge state evolution. This is accomplished by implementing auxiliary codes such as ATIMA [7] and GLOBAL [8]. EPAX [9] is utilized to return cross

sections of fragmentation products. The special case of fission has been treated by using the code MCNPX [10] to accurately predict the cross sections and dynamics of exotic beams produced by a ^{238}U beam incident on a Li or C target. The extensions to the code have made it possible to simultaneously compute high order optics and beam-material interactions in one cohesive framework.

The hybrid map-Monte Carlo code can calculate important quantities that describe the performance of the fragment separator. These include the transmission and the separation purity. In a map-only approach, calculations such as these are not possible. Experimental planning and optimization is possible with the hybrid map-Monte Carlo code, as various fragment separator settings can be readily adjusted. Here we present a description of the code, examples of calculations with it, and its application to the separation of rare isotopes.

IMPLEMENTATION

A solely map-based approach is not sufficient to model the evolution of an exotic beam in the fragment separator. It is impossible to take into account fragmentation and fission of the beam in matter in such an approach. There are also many other effects that are nondeterministic. Stochastic effects such as energy and angular straggling in matter and charge exchange demand a Monte Carlo method. To compute the extent of the stochastic effects, the most up-to-date programs such as ATIMA for calculating energy loss and energy and angular straggling have been integrated into COSY as simple procedures.

To get an accurate view of the evolution of the beam, any material that the beam passes through must be divided up into "slices." There are a couple of reasons to do this. One reason is that some of the rarer isotopes would not be produced at all if the whole target or wedge material thicknesses were used. By the same argument, each slice cannot be too thick as it won't account for multiple fragmentations or fissions. Having slices that are too thin increases the run time of the program. Also, the data acquired from MCNPX assumes a very thin thickness (0.1068 g/cm^2), so any deviation from this thickness per slice will give increasingly inaccurate results. The approximation for the cross sections and dynamics will be worse. A target thickness on this order will not be used for a FRIB, so for the most accurate approximations, more than one slice per target is used. The target thicknesses would typically be about 30%-40% of the range of the primary beam in the target material. Convergence tests have been performed to determine how many slices are necessary for a normal target thickness. This value is approximately one slice per 10% of the

*This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357
#erdelyi@anl.gov

projectile's range in the target material. The number of slices, however, is input by the user so more slices or fewer slices may be used. Particles are transported through the target by computing the map of each target or wedge slice and, in addition, the beam is allowed to fragment or fission only once per slice. The results of the creation of the particles in each slice, and the dynamics that occur must be composed slice by slice to get the full results of the beam's isotopic composition and dynamics for a whole target or absorber.

RESULTS

The performance of the fragment separator can be described quantitatively by two values, namely the separation purity and transmission. These quantities can only be determined with the map-Monte Carlo code. The separation purity is key to showing how much background contamination exists at the end of a separation. The transmission indicates the ratio of the number of particles of a particular rare isotope at the end of a separation stage to the number that is formed in the target. In some cases the separation purity may be good but the transmission is poor or vice versa. The two quantities must be evaluated together to effectively evaluate the performance of the separator.

In addition to having a desirable separation purity and transmission, it is also necessary that in some cases that all of the particles of the separated rare isotope beam have the same energy. In cases such as these, additional optics and a properly shaped energy absorber must be used to achieve a monochromatic beam.

Transmission

There are four general reaction mechanisms that take place when the primary beam interacts with a target. These include light and heavy nuclear fragmentation and light and heavy nuclear fission. The transmissions of isotopes produced by these mechanisms are shown in Table 1. The transmissions were calculated for a target thickness of 20% of the range of the primary beam in a Li target and wedge thicknesses of 30% of the range of the selected isotope beam energy in a two-stage separator.

Table 1: Transmissions According to Production Mechanism

Production Mechanism	Isotope	Transmission (%)
Light Fragmentation	^{14}Be	90.6
Heavy Fragmentation	^{100}Sn	91.0
Light Fission	^{78}Ni	21.5
Heavy Fission	^{132}Sn	42.9

There is a great dependence of the transmission on the reaction mechanism which produces each isotope. The best transmission results from fragmentation due to the low initial emittance of the rare isotope beam. The transmission is lower for fission products as they are initially emitted from the target with large angular and energy spread. This causes the loss of the isotope since it travels beyond the apertures of the fragment separator.

Due to the challenges of fission products in particular, a comprehensive transmission study was conducted for the fission product ^{132}Sn . This isotope was chosen for its importance to the nuclear physics community and the fact that a beam of ^{132}Sn has a large emittance and, hence, is one of the most difficult to capture. For this study, a 200 MeV/u ^{238}U beam is incident on a Li target of variable thickness. It fissions, producing ^{132}Sn among thousands of other isotopes. The thicknesses of the target and the first and second wedges of a two-stage separator are varied in increments of 10% of the range of ^{238}U in a Li target and ^{132}Sn in the two Al wedges for all thickness combinations between 10% and 60% of the range for the target thickness and 10% and 70% of the range for the wedges. The transmission of ^{132}Sn after two separation stages is shown in Figure 1 for constant target thickness and varied wedge thicknesses. While it is obvious that with increasing material in the system the transmission decreases, the transmission depends slightly more on the first wedge thickness than the second. This difference is more pronounced for thin targets.

Separation Purity

The separation purity, ϕ , describes the fragment separator's ability to select one isotope from all

others: $\phi = \frac{N_{A_0, Z_0}}{N_{A, Z}}$, where N_{A_0, Z_0} is the total number of

particles of the separated isotope at the end of a separation stage. N_{A_0, Z_0} is the total number of particles of all isotopes at the end of a separation stage.

This quantity depends on many things which are not explicit in this expression. First, the optics of a fragment separator system must be optimal to focus the separated isotope in as small as a region in x as possible. The ability to do this differs according to the isotope and the reaction mechanism by which it is produced. The primary beam also plays a role in producing the type and quantity of background impurities that exist in the system. If the primary beam is of low Z, the background impurities must have equal or lower Z and the contamination will be low to insignificant. If, however, a high Z beam such as ^{238}U is used, the range of contaminating isotopes produced is vast. Also, a radioactive beam such as this will produce background that would not otherwise be seen with a beam that is produced solely by fragmentation. The addition of fission as a production mechanism leads to an even broader range of isotopes produced and, in addition, these production rates are dependent on the energy of the primary beam.

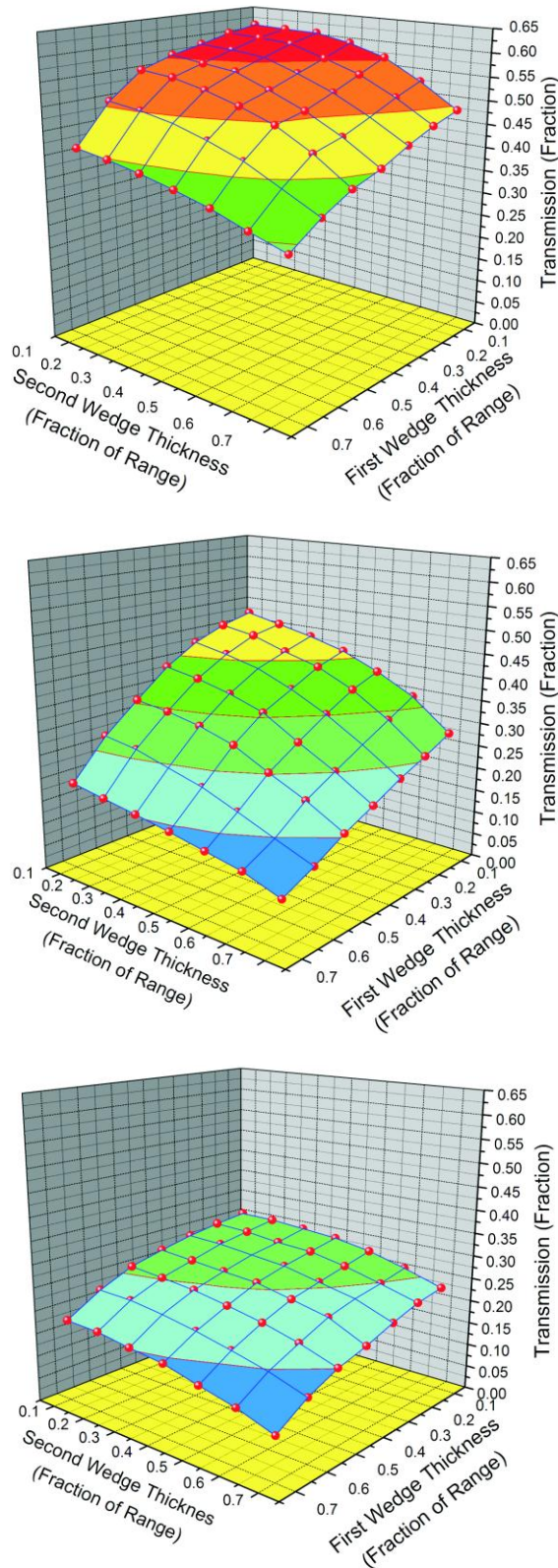


Figure 1: Transmission as a function of first and second wedge thicknesses in a two-stage separator. Target thicknesses are 10, 30, and 50% of the primary beam's range in target material (from top).

There are four general production mechanisms that have different beam dynamics and background that either complicate or make the separation easier. These mechanisms represent the extremes of the dynamics in the separator. All other isotope production mechanisms fall between these extremes in beam dynamics. For each of these four reaction mechanisms, one isotope was selected to be studied in detail. In each of these cases, the energy of the beam is limited by the parameters of the FRIB linear accelerator. The maximum energy that a primary beam attains is ^{238}U accelerated to 200 MeV/u. The optimal target and wedge thicknesses in each case are computed using the program LISE++ [11].

The optimization is completed for a one stage separation; the second stage has the same wedge thickness in terms of fraction of the range of the rare isotope beam in Al.

The separation purity of many rare isotopes has been calculated for four types of production mechanisms representing the extremes in beam dynamics (Table 2). The results yielded 100% purity for a one stage separation of ^{14}Be , a light fragmentation product. For ^{100}Sn (Figure 2), the separation purity is 7.5% after two separation stages, with only one contaminant. The heavy fission product, ^{132}Sn , has a separation purity of 4.04% after two stages. The light fission product, ^{78}Ni is challenging to separate with a separation purity of only 0.003% after two separation stages. Separation purity with the gas cell branch has also been computed for each of these isotopes.

Table 2: Separation Purity According to Production Mechanism

Production Mechanism	Isotope	One Stage	Two Stage	Gas Cell Branch
		ϕ (%)	ϕ (%)	ϕ (%)
Light Fragmentation	^{14}Be	100	100	96.7
Heavy Fragmentation	^{100}Sn	7.73×10^{-5}	5.89×10^{-2}	1.48×10^{-3}
Light Fission	^{78}Ni	2.79×10^{-2}	0.364	2.94×10^{-2}
Heavy Fission	^{132}Sn	1.15	4.04	1.52

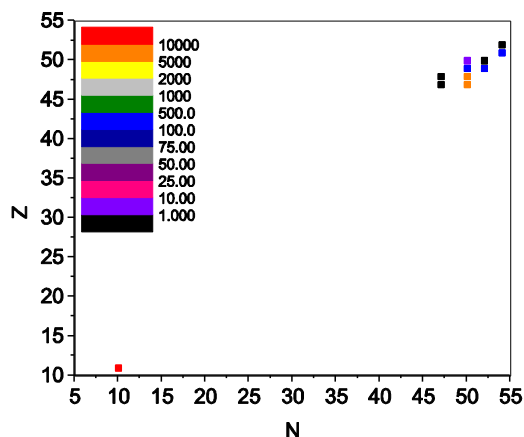


Figure 2: Distribution of isotopes that remain after the one stage separation of ^{100}Sn . Remaining isotopes are plotted in the N-Z plane. The quantities of the isotopes at the end of the system are indicated by the color of the box with all isotope quantities scaled such that one ^{100}Sn particle is at the end of the separation stage.

GAS CELL BRANCH

ISOL (Isotope Separation On Line) is not sufficient to study some isotopes. These isotopes need to be studied at a lower energy, and therefore are stopped in a neutral He gas cell. This low-energy regime is key for many nuclear physics and astrophysics experiments. For these cases, the second separation stage of a two-stage fragment separator is replaced with a monochromatic gas cell branch. This is necessary in order to stop all of the particles selected in the separation in as small of a region as possible in the He gas cell. In some cases, after stopping the isotopes in the gas cell, they may be reaccelerated to the desired energy. After the achromatic image of the first stage, the optics of the first half of the first stage are repeated, followed by a monochromatic wedge. A monochromatic absorber is obtained by shaping the wedge to cancel the wedge map element $(\delta|\delta)$ at first order and potentially further shaping it to cancel the map aberrations $(\delta|\delta^n)\delta_i^n$, where n is the order of the aberration.

CONCLUSION

A hybrid map-Monte Carlo code has been developed to accurately model beam-material interactions for the purpose of fragment separator beam dynamics simulations. This code has allowed for the calculation of important quantities that determine quality of the separation. These include the transmission and separation purity. Using the code, one may simulate a variety of exotic beam experiments. Future work will focus on more detailed optimization to find the best fragment separator settings for rare isotopes to be captured for experiment.

Computer Codes (Design, Simulation, Field Calculation)

REFERENCES

- [1] RISAC. Scientific Opportunities with a Rare-Isotope Facility in the United States. Technical report, NRC, 2006.
- [2] J. Nolen. Overview of the U.S. Rare Isotope Accelerator Proposal. Nucl. Phys., A734:661.668, 2004.
- [3] Y. Yano. RI Beam Factory project at RIKEN. In Proc. 17th Int. Conf. Cycl. Appl., pages 169.173, Tokyo, Japan, 2004.
- [4] T. Kubo, K. Kuska, and K. Yoshida et al. Status and overview of superconducting radioactive isotope beam separator BigRIPS at RIKEN. IEEE Transactions on Applied Superconductivity, 17:1069.1077, 2007.
- [5] W. Henning. The GSI project: An International Facility for Ions and Antiprotons. Nucl. Phys., A734:654.660, 2004.
- [6] M. Berz. Modern Map Methods in Particle Beam Physics. Academic Press, San Diego, 1999.
- [7] H. Weick. <http://www-linux.gsi.de/~weick/atima/>.
- [8] C. Scheidenberger, T. Stohlker, W. Meyerhof, H. Geissel, P. Mokler, and B. Blank. Charge states of relativistic heavy ions in matter. NIM B, 142:441.462, 1998.
- [9] K. Summerer and B. Blank. EPAX version 2: a modified empirical parameterization of fragmentation cross sections. Nuclear Physics A, 701:161C.164C, 2002.
- [10] M. James, G. McKinney, J. Hendricks, and M. Moyers. Recent enhancements in mcnpix: Heavy-ion transport and the LAQGSM physics model. NIM A, 562(2):819.822, 2006.
- [11] O. Tarasov and D. Bazin. LISE++: design your own spectrometer. Nuclear Physics A, 746:411.414, 2004.

THE STUDY ON THE SPACE CHARGE EFFECTS OF RCS/CSNS

S. Xu, S. Fang, S. Wang,* Institute of High Energy Physics (IHEP), Beijing, 100049, China

Abstract

The China Spallation Neutron Source (CSNS) is an accelerator-based facility. It operates at a 25 Hz repetition rate with an initial design beam power of 100 kW and is upgradeable to 500 kW. The accelerator of CSNS consists of a low energy linac and a Rapid Cycling Synchrotron (RCS). The RCS is a key component of CSNS. In this kind of high intensity RCS, the beam is space charge dominated, and the space charge effects are the main source of beam loss. Many simulation works were done for the study of space charge effects for CSNS/RCS by using the codes ORBIT and SIMPSONS. Various conditions are considered in the simulations, including the effects of different lattice structure, different tunes, the combine effect of sextupole field and space charge, different painting beam distribution, etc. The beam loss and emittance growth are compared for different conditions.

INTRODUCTION

The China Spallation Neutron Source (CSNS) is based on a high power accelerator, which consists of a 80 MeV linac, a 1.6 GeV rapid cycling synchrotron and beam transport lines [1]. The accelerator complex is designed to deliver a beam power of 100 kW at a 25 Hz repetition rate, with an upgrade capability of up to 500 kW by raising the linac output energy and increasing the intensity. The RCS is a key component of CSNS. It accumulates a beam injected at 80 MeV, accelerates the beam to a design energy of 1.6 GeV, and extracts the high energy beam to the target. Due to the high beam density and high repetition rate, the rate of beam loss must be controlled to a very low level. In this kind of high power RCS, especially at the low energy end, the beam is space charge dominated, and the space charge effects are the main source of beam loss. The space charge effects limit the maximum beam density, as well as beam power. Many simulations were done to study space charge effects in CSNS/RCS by using the codes ORBIT and SIMPSONS. Various conditions, which may influence the space charge effects and beam loss, are considered, including the effects of different lattice structure, different tune, the combine effect of sextupole field and space charge, different painting beam distribution, etc. The beam loss and emittance growth are compared for different conditions. The simulation results are the foundation of physics design and the choice of design parameters. To control the uncontrolled beam loss, the transverse and momentum beam collimation systems are designed. With the beam collimation, the uncontrolled beam loss can be compressed to less than 1W/m.

The present lattice of the CSNS/RCS is a triplet based fourfold structure, as shown in Fig. 1. Table 1 shows the main parameters of the lattice.

Table 1: Main Parameters of the Lattice

Circumference (m)	228
Superperiod	4
Number of dipoles	24
Number of long drift	12
Total Length of long drift (m)	75
Betatron tunes (h/v)	4.82/4.80
Chromaticity (h/v)	-4.3/-8.2
Momentum compaction	0.041
RF harmonics	2
RF Freq. (MHz)	1.0241~2.3723
Trans. acceptance ($\mu\text{m}\cdot\text{rad}$)	540
RF Voltage (kV)	165

SPACE CHARGE EFFECTS DURING INJECTION

Painting Schemes

Two painting schemes—correlated and anti-correlated painting—are both available in the injection of CSNS/RCS. For correlated painting, both the emittance ϵ_x and ϵ_y are painted from small to large during injection. It produces a rectangular transverse beam profile without space charge effects. For anti-correlated painting, the emittance is painted from small to large in one direction, and from large to small in the other direction (vertical direction here). It produces an elliptical transverse beam profile without space charge effects. In the case of disregarding space charge effects, the painting beam density is uniform. Figs. 2 and 3 show simulation results of painting with and without space charge effects. The upper left and right graphs show the particles distribution in (x, x') and (y, y') phase space respectively. The lower graphs show the distribution in (x, y) space and the emittance evolution during painting.

* wangs@ihep.ac.cn

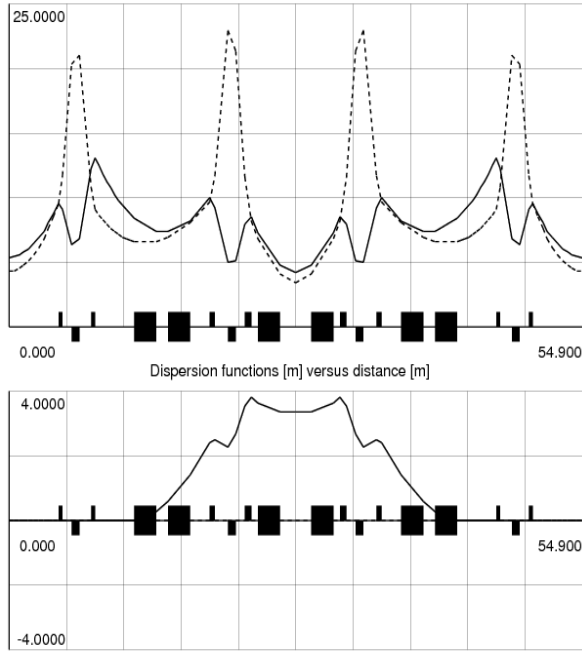


Fig. 1: Twiss parameters of one superperiod.

As shown in Fig. 2 (a) and Fig. 3 (a), the results of painting without space charge effects are just as predicted. Due to space charge effects, some halo particles are generated, and beam emittance is increased somehow.

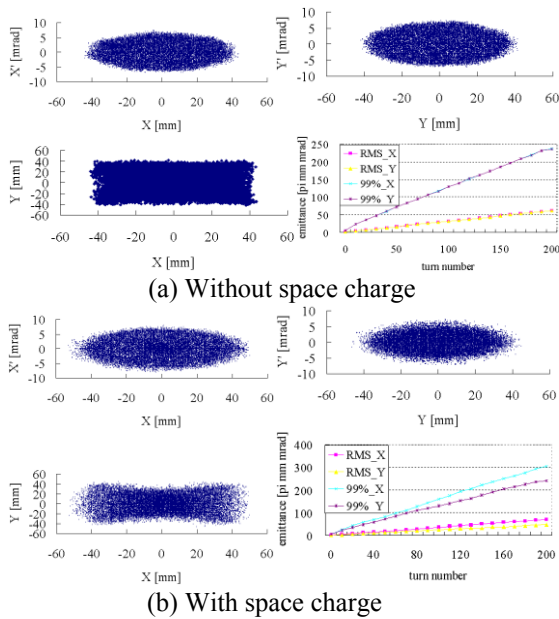


Fig. 2: Particles distributions and emittance evolution from correlated painting without (a) and with (b) space charge effects.

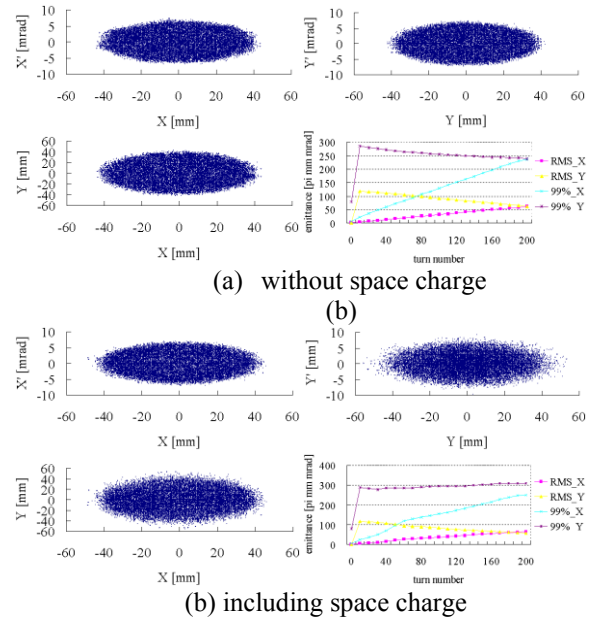


Fig. 3: Particles distributions and emittance evolution from anti-correlated painting without (a) and with (b) space charge effects.

Fig. 4 shows Poincaré maps of one test particle during different injection time-periods by correlated painting, in which the filamentation is observed. For anti-correlated painting, since both x and y emittance are painted from small to large during injection, the beam profile is susceptible to transverse coupling. However it has the capability of painting over the halo, while in the case of anti-correlated painting, it is immune to the transverse coupling, but it does not have the capability of painting over the halo in one direction (vertical shown here) [2].

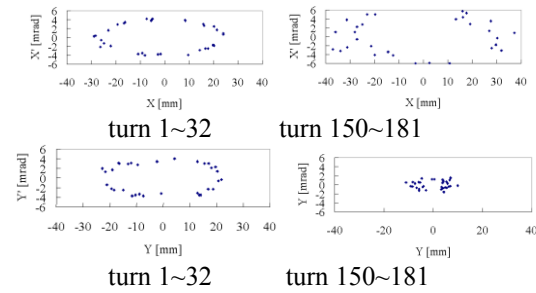


Fig. 4: Poincaré maps of one test particle during different injection time-periods from correlated painting.

Injection Painting Optimization

In order to reduce the transverse space charge effects, it is essential to produce a uniform longitudinal distribution. As shown in Figure 5, obtained by using ORBIT, much uniform longitudinal distribution can be obtained by using momentum offset in the injection painting procedure.

To investigate the diffusion motion of particles, all the particles are divided into three parts in (y, y') phase

space, outer part, middle part and inner region, as shown in Fig. 6, indicated by different colours. Due to space charge effects, Diffusion occurs among different parts, as shown in Figure 6(a). The diffusion results in the change of beam distribution, and the generation of halo particles. In order to produce much uniform transverse distribution and reduce halo production, two ways were tried to optimize the painting orbit bumps: painting a smaller emittance than the target value (case 1), injecting less particles in the inner and outer region of the emittance space (case 2). However, for case 1, this way results in small RMS emittance and large tune shift, which may be troublesome in the presence of magnet errors [3]. Fig. 7 and Fig. 8 show beam distributions and tune spreads of test particles with different injection painting respectively. The beam distribution in case 1 has a peak top, and large tune spread. The beam distribution of case 2 with a relatively flat beam density is more uniform.

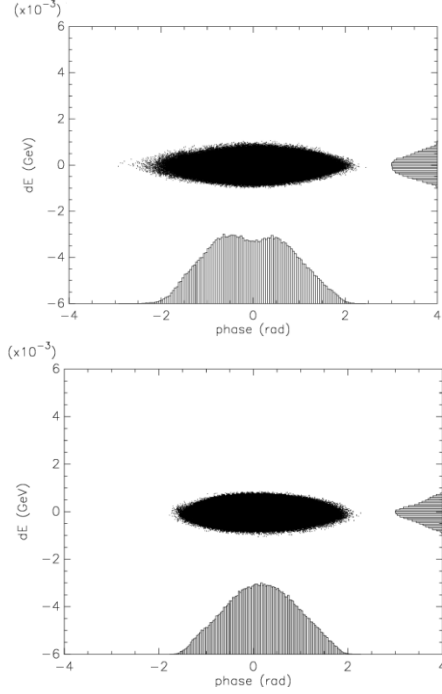


Fig. 5: Particle distributions in $(\phi, \Delta E)$ phase space after painting with (top) and without (bottom) momentum offset.

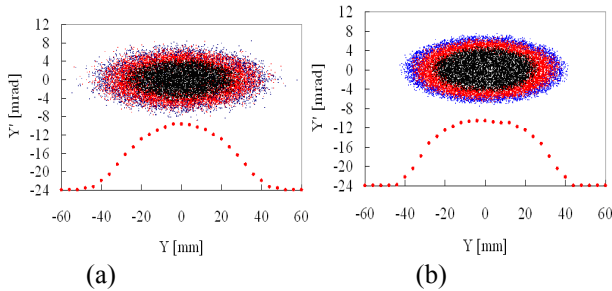


Fig. 6: Particles distribution in (y, y') phase space after painting with (a) and without (b) SC from the anti-correlated painting.

SPACE CHARGE EFFECTS DURING RAMPING

The initial beam distribution is produced by anti-correlated painting with a momentum offset. Fig. 9 shows the time evolution of the 99% unnormalized emittance in early-stage ramping. The unnormalized emittance should be depressed with $1/\gamma\delta\zeta$ without space charge effects. Due to space charge effects, there is apparent emittance growth in horizontal direction at the beginning of ramping. The study is focused on this time-period.

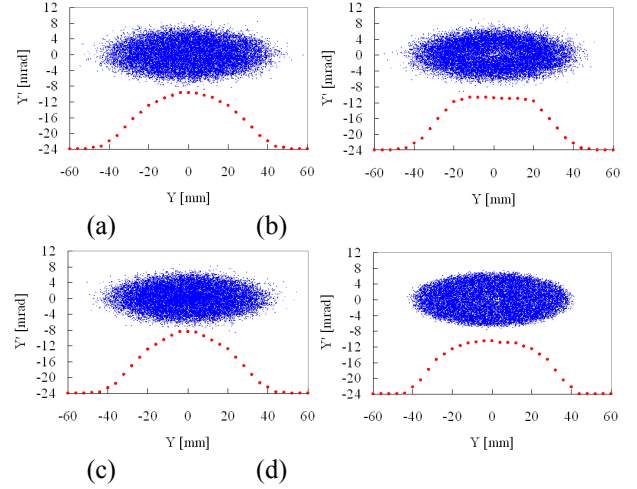


Fig. 7: Beam distributions in (y, y') phase space after painting by different injection orbit bumps: (a) uniform painting; (b) injecting less particles in the inner and outer region of the emittance space; (c) painting a smaller emittance than the target value; (d) uniform painting without SC.

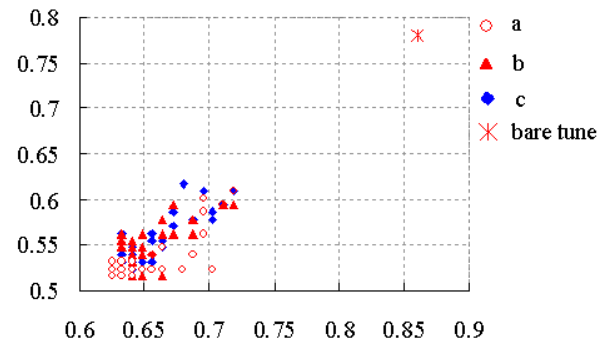


Fig. 8: Tune spreads of test particles with different painting distributions: (a) injecting particles a smaller emittance than the target value; (b) injecting less particles in the inner and outer region of the emittance space; (c) uniform painting.

Fig. 10 shows diffusion of particles among different parts during ramping. The red dots indicate the distribution just after the painting (200 turns). The blue dots indicate the distribution after 400 turns. Some particles move towards the inner region of the phase

space during ramping, while some towards the outer region. The diffusion among three parts is clearly observed.

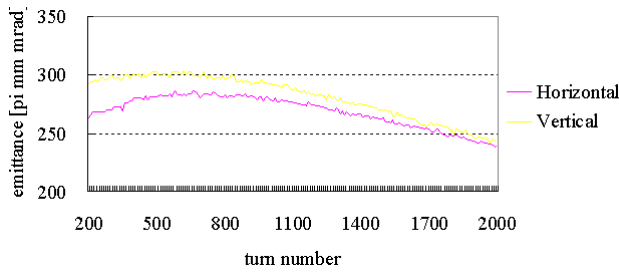


Fig. 9: Time evolution of the 99% emittance.

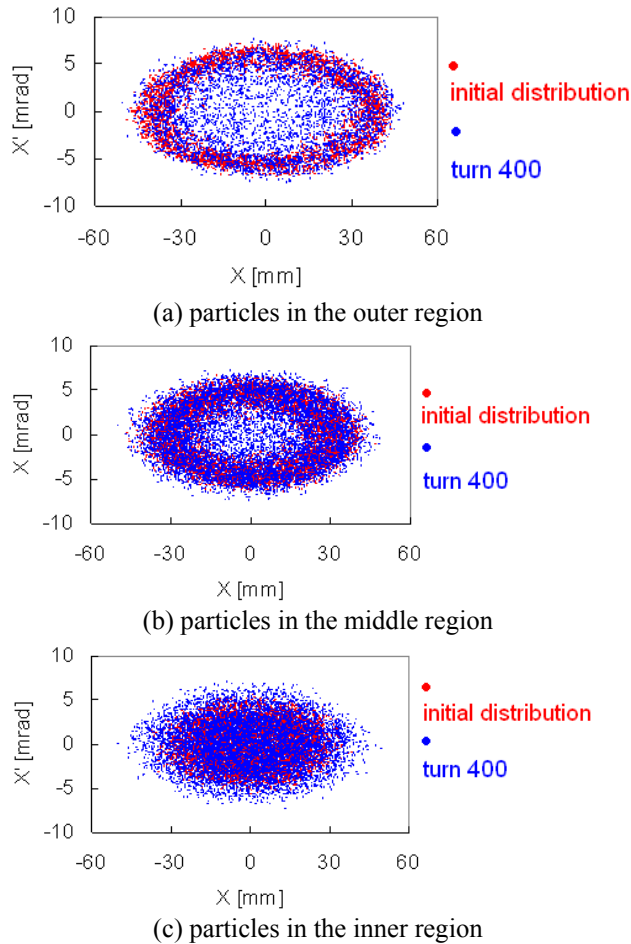


Fig. 10: Diffusion of particles among different parts during ramping.

Since the simulation does not include magnet errors, the emittance growth and particle distribution change must be driven by the beam itself (space charge force). There are several possible driving mechanisms such as (1) pure evolution of any coherent mode not related to resonance of an individual particle; (2) pure nonlinear resonance originating from the space charge; (3) coupling resonance

with the coherent modes [4]. More work will be done to study the mechanism here.

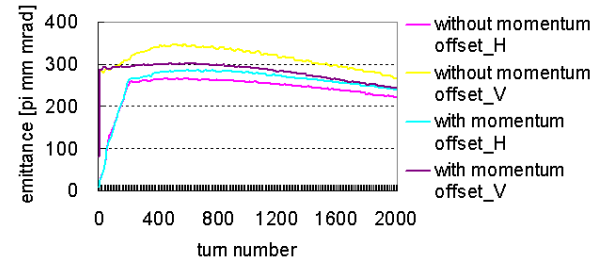


Fig. 11: Time evolution of the 99% emittance with and without momentum offset during injection.

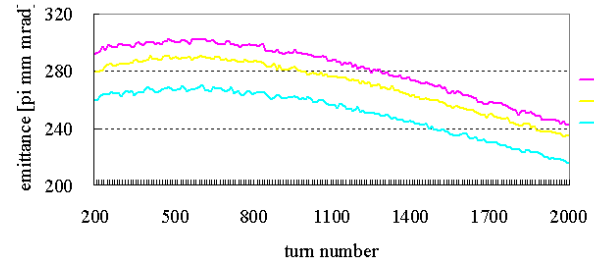


Fig. 12: Time evolution of the vertical 99% emittance with different initial distributions. Lines (a) (b) (c) indicate using the particles distribution of (a) (b) (c) in Fig. 7 as the initial distributions, respectively.

Simulations were also made to study emittance growth due to different painted distributions. The effects of different longitudinal distribution are first investigated. As described in Figs. 5 and 11, with a momentum offset during the injection painting, the longitudinal distribution after painting is much uniform and the emittance growth during ramping is less. Fig. 12 shows the time evolution of the vertical 99% emittance during ramping for different initial distributions. All the initial distributions are produced by anti-correlated painting (by 200 turns) with a momentum offset, but with different injection orbit bumps. All the initial transverse distributions show a similar trend of emittance growth without magnet errors.

EMITTANCE GROWTH VS. TUNES

The dependence of emittance growth on tunes was considered. Tunes around design values of 4.82/4.80 and with a split of 0.5 were compared. In simulations, anti-correlated painting scheme without momentum offset was adopted. Further work will be done to study the tune effects in the case of using momentum offset in the injection painting procedure. Fig. 13 plots the maximum 99% emittance during ramping vs. the bare horizontal tune, where the vertical tune was fixed to 4.78. Fig. 14 presents the time evolution of the 99% emittance with different bare tunes. There is less emittance growth at the tune of just below the integer tune. For the tune 4.82/4.36, there is large emittance growth. Fig. 15 shows the Poincaré maps of two test particles during ramping with

the bare tune 4.82/4.36, in which particles are trapped by the structure resonance $Q_y=4$. Because the magnet errors are not included in the simulations, the structure resonance is excited by space charge force [5]. It is apparent that the choice of tune is very important.

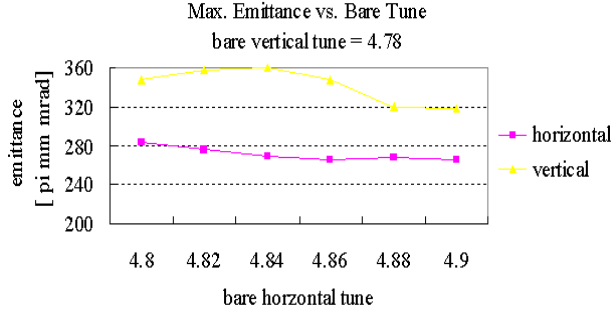


Fig. 13: Maximum 99% emittance during ramping vs. the bare horizontal tune.

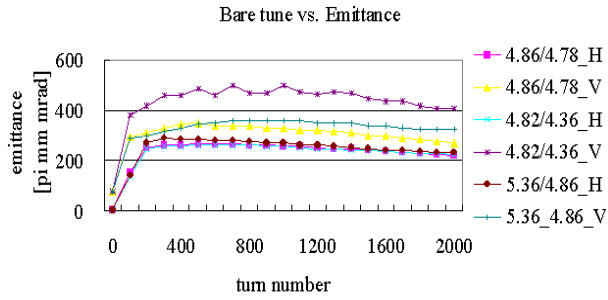


Fig. 14: Time evolution of 99% emittance with different bare tunes.

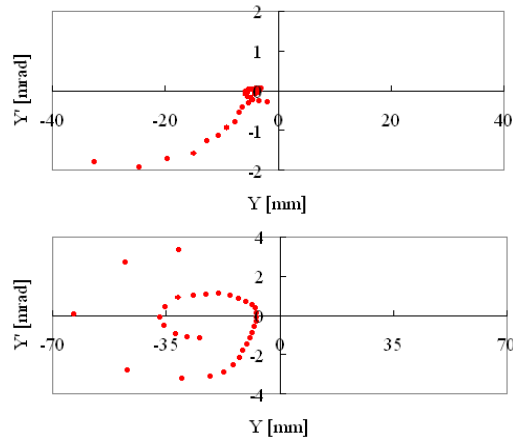


Fig. 15: Poincaré maps of two test particles during ramping with the bare tune 4.82/4.36.

DEPENDENCE ON LATTICE TYPE

To investigate the dependence of emittance growth on the lattice type, another two lattices, named as Lattice-B and Lattice-C, were introduced in simulation for compare. As shown in table 2, the lattice B consists of FODO cells in the arcs and doublet in the dispersion free straight sections, while Lattice-C consists of full FODO structure.

Transverse and Longitudinal Instabilities and Cures

In this section anti-correlated painting scheme without momentum offset were still adopted. Fig.16 shows the time evolution of the 99% emittance for different lattice types. Lattice-C looks better in the point of view of emittance growth.

SUMMARY

Space charge effects have been studied by simulations, including the space charge effects in painting, ramping, and the dependence of space charge effects on the bare tune and lattice type. Some injection painting optimizations were made for decreasing halo formation and tune spread. The choice of bare tune is very important. The choice of lattice structure also influences the emittance growth. All the simulations in this paper do not include the magnet field errors. Further work will involve the combine effects of space charge effects magnet field errors.

ACKNOWLEDGMENT

The authors would like to thank Shinji Machida for providing the SIMPSONS code and help on the test run.

Table 2: Parameters of lattices

	Lattice-B	Lattice-C
Structure	FODO+ Doublet	FODO
Circumference (m)	248	239.6
Super-periods	4	4
Qx/Qy	5.86/5.78	5.86/5.78
Max. straight (m)	9.3	4.5

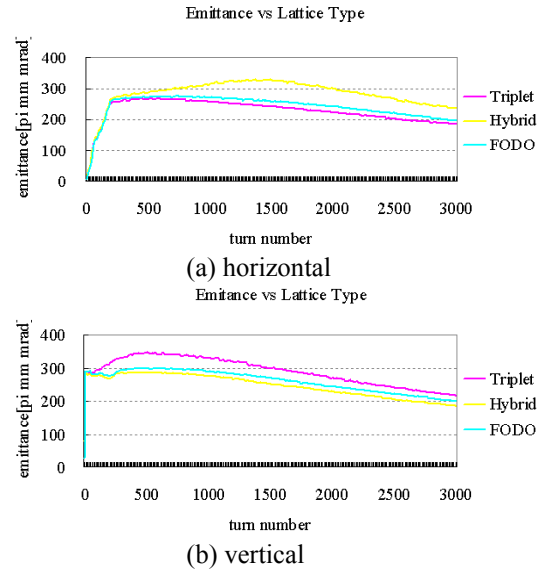


Fig. 16: Time evolution of the (a) horizontal and (b) vertical 99% emittance.

REFERENCES

- [1] CSNS Feasibility Study Report, June, 2009, IHEP.
- [2] J. Beebe-Wang *et al*, “Space Charge and Magnet Error Simulation for the SNS Accumulator Ring”, p. 1286, Proc. of EPAC2000.
- [3] J. Galambos *et al*, “SNS Injection Simulations with Space Charge”, p. 3140, Proc. of PAC99.
- [4] Accelerator Technical Design Report for High-Intensity Proton Accelerator Facility Project, J-PARC.
- [5] S. Machida, “Space Charge Effects in Low-Energy Proton Synchrotrons”, Nucl. Instr. & Meth. A309, (1991) 43-59.

OPTIMIZATION ALGORITHMS FOR ACCELERATOR PHYSICS PROBLEMS*

B. Mustapha[#] and P. N. Ostroumov

Physics Division, Argonne National Laboratory, IL 60439, U.S.A.

Abstract

Optimization tools are needed in every step of an accelerator project, from the design to commissioning to operations. However, different phases have different optimization needs that may require different optimization algorithms. For example, a global optimizer is more appropriate in the design phase to map the whole parameter space whereas a local optimizer with a shorter path to solution is more adequate during operations to find the next best operating point. Different optimization algorithms are being used in accelerator physics, we mention in particular standard algorithms such as least square minimization and evolutionary algorithms such as genetic optimization. Over the years, we have developed several optimization tools for beam tracking codes to include 3D fields and SC effects. Including particle tracking in the optimization process calls for parallel computing. We will review the different algorithms and their implementation and present few highlight applications.

OPTIMIZATIONS IN ACCELERATOR PHYSICS

Optimizations are heavily used in the design phase of an accelerator project, but they are much less used to support the commissioning and operations once the machine is built. During the design phase, optimizations are used in the design of the different beam line elements: magnets, rf cavities, etc. They are also used for the lattice optimization to find the appropriate sequence of elements and drift spaces. Once the lattice is defined more optimizations are used to determine the appropriate element settings for optimal beam dynamics and beam quality. This is often iterated with the lattice design. Once the accelerator is built, more effort is dedicated to hardware problems than to developing a realistic model of the machine. We believe that using the appropriate optimization tools during the commissioning should help better understand the machine's behaviour and expedite the delivery of the first beam. Fits to reproduce the experimental data using a model should significantly improve the predictability of the model to use for real-time machine operations. Often, simplified models (1D, single particle) are used to support daily machine operations [1]. Simple models have the advantage of being fast and able to describe the overall behaviour of the machine while detailed 3D models are slow and still cannot reproduce the details seen in the data [2]. We believe that a significant effort should be dedicated to

developing more realistic 3D models before being able to use them to support real-time machine operations. Once such models are fully developed large scale parallel computing could be used for fast turn-around simulations and optimizations.

ELEMENTS OF AN OPTIMIZATION PROBLEM

The first important step is the proper definition of the optimization problem. An optimization problem has one or more objectives which are the important quantities or qualities characteristics of the problem that you would like to optimize. These objectives depend on the parameters of the problem which are the variables affecting the outcome or the solution to the problem. It is usually a good practice to choose the parameters to which the solution is more sensitive. These parameters could be subject to constraints and or correlations which define the limits of the parameter space. The simplest case is where the parameters are independent with lower and upper bounds. If the parameters are correlated, it is usually recommended to reduce them into a set of independent parameters. The last and most important element of an optimization problem is the choice of the appropriate optimization algorithm. Depending on the nature of the problem, the most appropriate algorithm could be a local optimizer, a global one, a standard or an evolutionary.

LOCAL VERSUS GLOBAL OPTIMIZATION ALGORITHMS

It is usually not hard to find a local minimum of an objective function. What is hard is to prove that the minimum found is a good one and it is even harder to prove that a minimum is an absolute or a global one. A local optimizer usually starts with a first guess then finds a direction that minimizes the objective function and moves one step in that direction. The procedure is repeated iteratively until no more progress could be made. A local algorithm is usually fast because it explores the parameter space along a single path defined by the minimization direction adjusted at every step. In contrast, a global optimizer should explore the entire parameter space and eventually find all local minima before finding a global one. It should also prove that the minimum found is a global one which makes it much slower than a local optimizer. Luckily, not all problems or applications require global optimizations. A global optimizer is more appropriate to use in the design phase of an accelerator project to map the whole parameter space and make sure not to miss the best set of design parameters. Such a global optimizer should also find all feasible solutions to

* This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357.

[#] Corresponding author: brahim@anl.gov

Beam Dynamics, Other

help make compromises if needed. On the other hand, a local optimizer with a shorter path to solution is more adequate to support accelerator operations. In this case, we usually start from a good starting point to find the next best operating point by returning few elements on the accelerator. The time to solution is a very important parameter in the choice of the appropriate optimization algorithm. A good optimizer should also optimize the path to the best solution.

STANDARD VERSUS EVOLUTIONARY OPTIMIZATION ALGORITHMS

Standard optimization algorithms [3] are the most common and widely used. They usually have a single objective function to optimize which could be a weighted sum of multiple objectives. Derivatives of the objective function with respect to the optimization parameters are usually required. A single trial solution is evaluated at every iteration. Evolutionary algorithms are more recent [4] and are based on the “Theory of Evolution and Natural Selection”, where only the best survive. Multiple objective functions could be included in the optimization without the requirement of knowing their derivatives. Multiple trial solutions are evaluated at every iteration which gives the global nature of evolutionary algorithms.

Examples of Standard Optimization Algorithms

The first example of standard optimization algorithms is the simplex method [5], which does not require the function derivatives. It starts by building a simplex consisting of the first guess and a base of feasible solutions in the parameter space. Iteratively, the worst solution is replaced by a better one built from the base. The iterations stop when no more progress could be made or at a predefined cut-off error on the solution. The second example is the gradient descent method [6] which require the first derivatives of the objective function on the optimization parameters. The derivatives are used to determine the minimization direction p_i at every iteration i : $p_i = -B_i^{-1} \nabla f_i$ where f is the objective function and B is a symmetric non singular matrix. In the case where B is the matrix identity I the method is called the steepest descent method. The third example is the Newton method [7] which requires both the first and second derivatives of the objective function. In this case B is the matrix of second derivatives also known as the Hessian matrix H . The fourth example is the Quasi-Newton method [8] which uses the first derivatives but keeps updating the matrix of second derivatives at every iteration. In this case B is an approximation to the Hessian matrix and does not require extra computing effort for the second derivatives as for the Newton method. Each of these methods is actually a class of methods with a variety of implementations.

Example of Evolutionary Algorithm: A Genetic Optimizer

A genetic optimizer [9] starts with a set of solutions randomly generated inside the parameter space. The

solutions are evaluated and ranked based on the objectives and constraints of the problem to select a subset of best solutions. The selected solutions are then used to generate the next population of solutions by crossover, mutation or other using predefined rates. The evaluation and ranking procedure is repeated for the new set of solutions until no progress could be made or a stable set of best solutions is obtained. For a given solution, the array of parameter values plays the role of a gene. Figure 1 shows the different ways of generating the next generation of solutions from the selected set of best solutions, namely crossover, mutation and random. The random generation of new solution could be turned on at the beginning for better sampling of the parameter space then turned off.

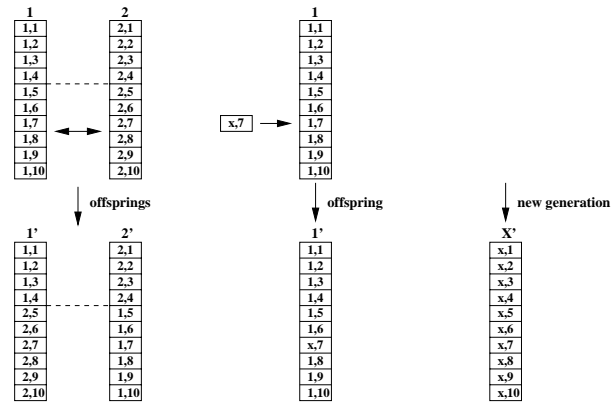


Figure 1: Three mechanisms to generate off-springs from selected best solutions. Crossover of two solutions on the left, mutation of a single solution in the middle and new random solution on the right.

PARALLEL OPTIMIZATION ALGORITHMS

Standard algorithms are serial in nature because the direction of the next iteration is decided based on the outcome of the current one. The fact that a single solution is evaluated per iteration makes these algorithms parallelizable only at the level of the objective function and derivatives evaluation. For example, a least square minimization where the objective function is of the form:

$$f = \sum_{i=1}^N \frac{\Delta X_i^2}{\epsilon_X^2}$$

could be parallelized by parallelizing the sum for large N . Optimizations using multi-particle tracking could be parallelized by parallelizing the tracking, the Poisson solver and the statistics calculations for large number of particles. A global optimizer may be parallelized by subdividing the parameter space and assigning the different sub-spaces to different processors.

In contrast Evolutionary algorithms are parallel in nature because at every iteration multiple solutions are evaluated independently which makes them well suited

for parallel processing with minimal communication. It is however not easy to parallelize the ranking, selection and offspring generation which are usually assigned to the master process. This makes evolutionary algorithms more suitable for optimization using multi-particle codes with realistic 3D external and space charge fields. Usually no parallel particle tracking is required unless a very large number of particles is needed for the optimization problem. In practice, any tracking code with space charge calculations could be used. A higher level parallel layer is often used to manage the generation, ranking and selection of trial solutions and every process calls the code when needed. In our beam dynamics code TRACK [10] this layer was built-in within the code.

APPLICATIONS IN BEAM DYNAMICS OPTIMIZATION

Automatic longitudinal tuning of a multiple charge state ion beam before a stripper

In this application, a longitudinal fine-tuning procedure was developed specifically for a multiple charge state beam to minimize its longitudinal emittance right before a stripper [11]. The beam should reach the stripper in the form of an up-right ellipse in the $(\Delta\phi, \Delta W)$ plane to minimize the emittance growth from the energy straggling effect in the stripper. This could be realized by matching the beam centroids and Twiss parameters of the different charge state beams. The objective function in this case is:

$$F = \frac{(W_{q0} - W_0)^2}{\epsilon_w^2} + \sum_{qi} \frac{\Delta W_{qi}^2}{\epsilon_{\Delta W}^2} + \sum_{qi} \frac{\Delta\phi_{qi}^2}{\epsilon_{\Delta\phi}^2} + \sum_{qi} \frac{\alpha_{qi}^2}{\epsilon_{\alpha}^2} + \sum_{qi} \beta_{qi}$$

where W_0 is the desired beam energy and ϵ_w is the associated error. $\epsilon_{\Delta W}$, $\epsilon_{\Delta\phi}$ and ϵ_{α} are the allowed errors on the relative energy, phase and α shifts of the individual charge state beams from the central beam. The fit parameters in this case are the RF cavities phases and amplitudes in the section up-stream of the stripper. Figure 2 shows the results of the fit for a five charge state uranium beam in the medium energy section of the RIA driver linac. This optimization reduced beam losses in the high-energy section of the linac by a significant factor as seen on figure 3.

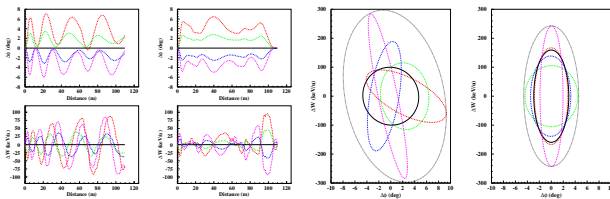


Figure 2: The left 4 plots show the phase and energy oscillations of the five charge states around the central charge state before and after applying the tuning procedure. The right 2 plots show the corresponding beam ellipses on the stripper before and after tuning.

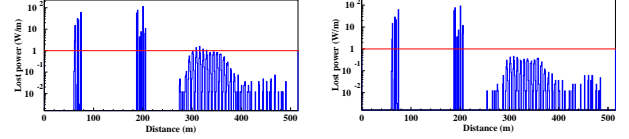


Figure 3: Beam loss in the RIA driver linac before and after applying the longitudinal tuning procedure. The two peaks correspond to the location of the strippers and the scatter loss is in the high energy section which has reduced significantly after fine tuning is applied.

A realistic corrective steering algorithm

We have recently developed a realistic corrective steering procedure [12]. A simplified algorithm is presented in figure 4.

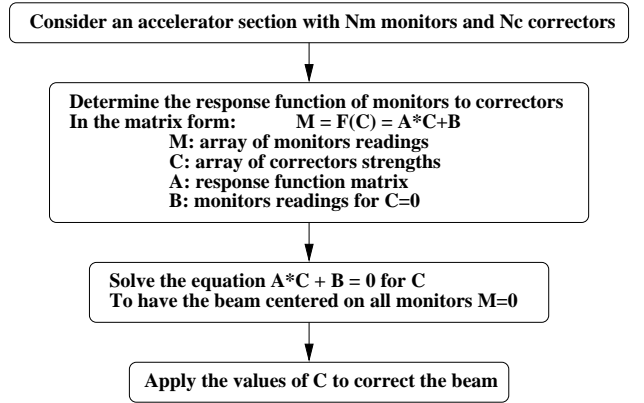


Figure 4: Algorithm for the corrective steering procedure.

In TRACK implementation, instead of solving the matrix equation $A*C+B=0$ for the array of corrector field strengths C , with A being the response function of monitors to correctors and B the monitors readings for $C=0$, we perform a least square minimization of the equivalent function given below:

$$f(C_{i_c}, i_c = 1, N_c) = \sum_{i_m=1}^{N_m} \frac{(\sum_{i_c=1}^{N_c} A_{i_c i_m} * C_{i_c} + B_{i_m})^2}{\sigma_{i_m}^2}; \text{ for } |C_{i_c}| \leq C_{\max}$$

In this way, we can include the monitor precision σ_{im} and the maximum corrector field strength C_{\max} in the solution. Monitors with different precisions will have different weights in the minimization procedure. The minimization should lead to an approximate solution in the case of an over-determined problem (more equations than variables) and to the best solution in the case of multiple solutions (under-determined problem). This realistic corrective steering procedure could very well be applied in the design phase of an accelerator project to determine the monitors and correctors requirements for an effective beam center correction as well as in the control room of an operating accelerator. The results of its application during the design of the HINS project front-end linac [13]

being built at Fermilab are shown on figure 5. After multiple iterations, the optimum numbers and locations of correctors and monitors for an effective correction were determined.

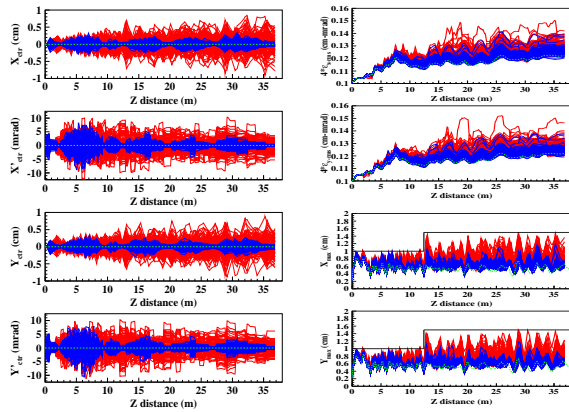


Figure 5: Beam position and angle centers (left) and beam emittances and envelopes (right) before (red curves) and after (blue curves) applying the correction procedure for 100 randomly misaligned linacs. In green is the ideal case without misalignment and in black is the aperture.

Design optimization of a chicane in an ultra-low emittance electron injector

We have recently implemented a parallel genetic optimizer into the beam dynamics code TRACK. As a first application we used it for the design optimization of the first chicane of an electron injector for an X-FEL-O linac [14]. The objective was to minimize the transverse emittance at the end of the chicane. The parameters were the quadrupoles and solenoids strengths as shown in figure 6. We were able to reduce the transverse emittance by about 10% from a manually optimized case which is very critical for an ultra-low emittance injector.

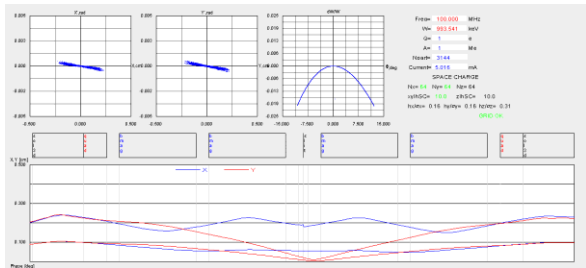


Figure 6: Optimized beam in the first chicane of an electron injector for an X-FEL-O linac.

To study the convergence of the genetic optimizer we compared the corresponding results to the results from a 2D map in the case of two parameter optimization. Figure 7 shows the results for both cases proving the

convergence of the genetic optimizer in 5 iterations corresponding to a total of 500 trial solutions. Large scale optimizations are underway for a larger section of the linac.

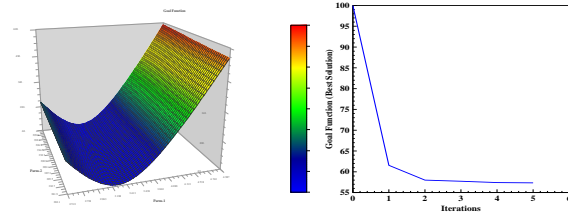


Figure 7: Comparing the result of a 2D map (left) with the result of the genetic optimizer (right) for a two parameter optimization problem.

POTENTIAL APPLICATION: THE MODEL DRIVEN ACCELERATOR

Concept and Motivations

The concept of the model driven accelerator is to develop and use a computer model to support real-time accelerator operations. Presently, no accelerator in the world could fully rely on a computer model for its operations. The main reason is a discontinuity between the design and operation phases of an accelerator project. Among the factors contributing to this discontinuity are: 1- Simulations in the design phase assume almost perfect conditions and cannot reproduce the real machine, 2- Actual elements specification and performance are usually different from their original design and in most cases 3- Not enough diagnostic devices to characterize the machine. The lack of a realistic model to support the commissioning and operations results in significant delay in the deployment of a new machine and a lot of time spent on machine tuning during operations. This usually leads to low availability and high operating cost of the machine. For example, a complex system such as the proposed FRIB facility [15], where primary beams from proton to uranium up to 600 MeV/u are used to produce beams of rare isotopes all over the map, cannot afford not to have a computer model to support its operations.

To bridge the gap between the design and operation phases we propose to develop a realistic model of the machine. Among the benefits of such a model is fast tuning for the desired beam conditions and fast retuning to restore the beam after a failure. This should significantly improve the availability of the machine and reduce its operating cost.

Requirements for the realization of the Model Driven Accelerator

The main requirements for the realization of the model driven accelerator could be summarized in the development of a 3D beam dynamics code with the appropriate set of optimization tools and large scale computing capabilities. A multi-particle beam dynamics code is more realistic than matrix-based and single-

particle codes because it supports 3D fields, includes fringe fields and appropriate space charge calculations. It also allows more detailed simulations necessary to study eventual beam loss and produce data similar to the measured data. Such a code should also include a large set of optimization tools. Optimization tools are needed not only for design optimization but also to tailor the computer model to the actual machine to be useful for real-time operations. Multi-particle optimizations usually involve tracking a large number of particles for large number of iterations which is very time consuming and requires large scale parallel computing. Therefore the beam dynamics code should have parallel computing capabilities.

A small scale realization of the Model Driven Accelerator

We have recently succeeded to extract, accelerate, analyze and recombine a two-charge state DC bismuth beam from an ECR ion source [16]. The beam is perfectly combined at the point of injection into a subsequent RF accelerator. We consider this as a small scale realization of the concept of the model driven accelerator. The beam dynamics code TRACK was used first to design the beam line and then to support the operation by predicting the elements settings required to recombine the two charge state beam at the end of the LEBT. Figure 8 is a general 3D view of the prototype 2Q-LEBT beam line.

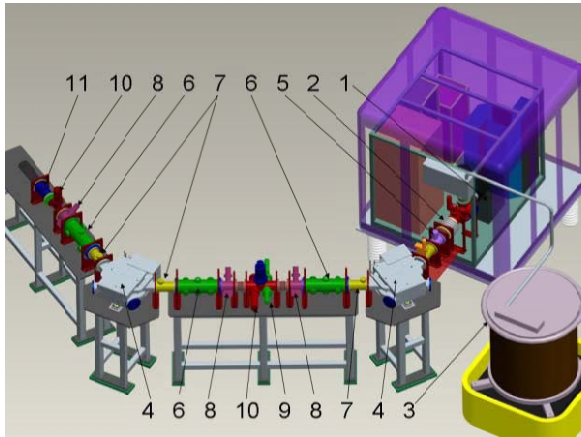


Figure 8: General 3D view of the 2Q-LEBT

For a realistic beam dynamics simulation, 17 beams (O and Bi) are tracked simultaneously from the ion source through the LEBT with a total current at the source of about 2 mA. We assumed a 4D water-bag initial distribution for all beams and a 50 % charge compensation factor in non-electric devices and 0% in electric devices. Realistic 3D models were developed and used for all beam line elements. In order to tailor the TRACK model to the actual beam line we had first to determine the initial beam parameters at the source. To do so we had to develop a new procedure to fit the beam profiles measured at the middle plane by varying the

beam parameters at the source. Figure 9 shows the result of the fit for a two-charge state 75-kV bismuth beam (20+, 21+). The transverse emittances and Twiss parameters obtained by fit show that despite the axial symmetry of the extraction region, the beam is not axial symmetric which could be explained by a non symmetric plasma boundary inside the ion source.

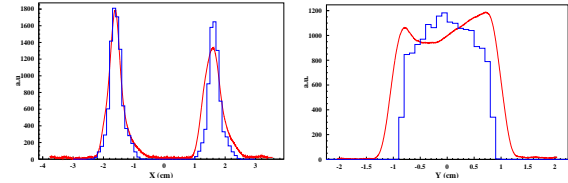


Figure 9: Horizontal (left) and vertical (right) beam profiles. The curves are the measured profiles and the histograms are the result of the TRACK fit.

Once the initial beam conditions are known, we used the computer model to find the element settings for the desired operation mode. A new fit procedure was developed to produce symmetric beam dynamics between the two bending magnets as this is a necessary condition to recombine the multiple charge state beams. Another fit was also used to find the setting of the last triplet for a perfect combination at the end of the LEBT where a beam profile monitor and a Pepper-Pot emittance meter are installed. Figure 10 shows the measured beam profiles and figure 11 shows the Pepper-Pot images at the end of the LEBT. A comparison of the element settings predicted by TRACK and the actual setting to combine the two beam shows maximum deviation of $\sim 10\%$ which could be improved by checking the assumptions made in the simulations. We notice that the two charge state beams are almost perfectly combined.

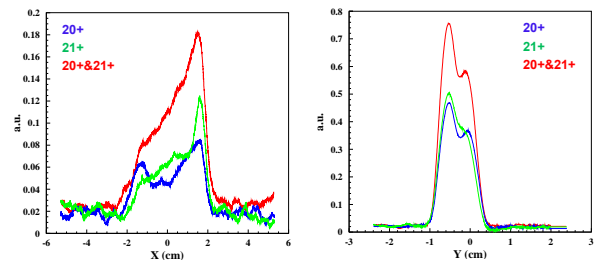


Figure 10: Measured beam profiles at the end of the LEBT for the individual Bi 20+ and 21+ beams and the combined beam.

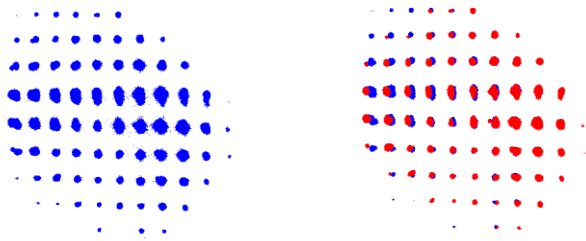


Figure 11: Pepper-Pot images of the combined beam (left) and the individual beams (right). Bi 20+ is in blue and 21+ is in red.

Further developments towards the Model Driven Accelerator

To be able to use a realistic 3D model online for real-time machine operations, we should be able to perform large scale optimizations on large number of processors (32768 processors or more). More optimization tools need to be developed for the commissioning phase to tailor the computer model to the actual machine by fitting the measured data. For this purpose, interfaces between the beam diagnostic devices and the computer model are needed to calibrate and analyze the data to input to the code. Numerical experiments could be used to test and fine tune the tools before implementation to the real machine by producing detector-like data. Only after all these developments, that a realization of the model driven accelerator will be possible. As a full scale application, we are proposing to apply this concept to the superconducting linac ATLAS at Argonne and to other existing machines.

SUMMARY

Optimization tools and methods are needed in every phase of an accelerator project, namely the design, commissioning and operations. No single algorithm could satisfy all these optimization needs. Different algorithms are being used in accelerator physics: local, global, standard and evolutionary. We have briefly reviewed and compared different classes of optimization algorithms and presented few applications in beam dynamics optimization. The ultimate goal of realizing the concept of the “Model Driven Accelerator” will require the development of a realistic 3D beam dynamics code with the appropriate set of optimization tools and large scale parallel computing capabilities. For new machines, we should take advantage of the commissioning phase to bridge the gap between the original design and the actual machine by tailoring the computer model to the machine. Obviously, a significant development effort is still needed for a full scale realization of the concept of the “Model Driven Accelerator”.

REFERENCES

- [1] “Using Online Single Particle Model for SNS Accelerator Tuning”, A. Shishlo, Proceedings of HB-2008, Nashville, Tennessee, USA.
- [2] “Measurement and Simulations of the J-PARC Linac”, M. Ikegami, Proceedings of HB-2008, Nashville, Tennessee, USA.
- [3] “Numerical Optimization”, a book by J. Nocedal and S. J. Wright, Springer Series in Operations Research.
- [4] “Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach”, E. Zitzler and L. Thiele, IEEE Transactions on Evolutionary Computation 3 (1999) 257–271.
- [5] “A simplex method for function minimization”, J. A. Nelder and R. Mead, Comp. J. 7 (1965) 308.
- [6] “Multiplier and gradient methods”, M. R. Hestenes, Journal of Optimization Theory and Applications, 4 (1969), p. 303-320.
- [7] “Newton’s method for constrained optimization”, J. Goodman, Mathematical Programming, 33 (1985), p. 162-171.
- [8] “Quasi-Newton methods, motivation and theory”, J. E. Dennis and J. J. More, SIAM Review, 19 (1977), p. 46-89.
- [9] “Multivariate optimization of a high brightness dc gun photoinjector”, I. V. Bazarov and C. K. Sinclair, Phys. Rev. ST Accel. Beams 8, 034202 (2005).
- [10] “TRACK: The New Beam Dynamics Code”, V. N. Aseev et al, Proceedings of PAC-2005 Conference, Knoxville, Tennessee.
- [11] “Automatic longitudinal tuning of a multiple-charge-state heavy-ion beam”, B. Mustapha and P. N. Ostroumov, Phys. Rev. ST Accel. Beams 8, 090101 (2005).
- [12] “A Realistic Corrective Steering Algorithm: Formalism and Applications”, B. Mustapha et al, Proceedings of PAC-2009, Vancouver, BC, Canada.
- [13] “Front-End Design of a Multi-GeV H-minus Linac”, P. N. Ostroumov et al, Proceedings of PAC-2005, Knoxville, Tennessee.
- [14] “Optimized Design of an Ultra-Low Emittance Injector for Future X-ray FEL Oscillator”, P. N. Ostroumov et al, Proceedings of PAC-2009, Vancouver, BC, Canada.
- [15] “FRIB: A New Facility for Rare Isotope Beams”, R. C. York, Proceedings of PAC-2009, Vancouver, BC, Canada.
- [16] “Analysis and recombination of multiple-charge-state beams from an electron cyclotron resonance ion source”, P. N. Ostroumov et al, Phys. Rev. ST Accel. Beams 12, 010101 (2009).

Application of Multiobjective Genetic Algorithm in Accelerator Physics*

L. Yang[†], NSLS-II, Upton, NY 11973, USA

D. Robin, F. Sannibale, C. Steier, W. Wan, ALS, Berkeley, CA 94720, USA

Abstract

The optimization of an accelerator system is important in both design and upgrade stage, and many of them are Multiobjective problems, i.e. searching for a balance between several quantities. A full understanding of this balance could provide the decision maker more information on the final choice. In this paper we present the application of an optimization algorithm called Multiobjective Genetic Algorithm (MOGA) in two problems. One is the lattice of a synchrotron light source (take ALS as an example) and the other is a VHF gun.

INTRODUCTION

The optimization of an accelerator system is obviously an important problem in both design and upgrade stage. Depending on different system, storage ring or LINAC, collider or light source, this could be minimizing the emittance, optimizing beta functions and bunch length. For a optimization algorithm, the challenges come from the convergence of solutions, constraints on variables and objective functions, conflicting objective functions. In this paper we will introduce an algorithm called multiobjective genetic algorithm (MOGA), show the applications on two problems, one is the lattice optimization for the Advanced Light Source (ALS), a problem with strong constraints in both variable space and objective space. The other is VHF Gun, in which a single simulation cost a couple of hours, therefore in order to get result in a reasonable amount of time, the convergence speed becomes very important.

GENETIC ALGORITHM AND MULTIOBJECTIVE OPTIMIZATION

Genetic algorithm (GA) is a search technique in optimization, it was developed in 1970s [8, 7, 4] and now as a class of evolutionary algorithms (EA). The outline of Genetic Algorithm (GA) usually has four steps, first a set of numbers in parameter space are chosen, i.e. the initial population, then they are paired to produce new candidate, we call them parents and children. This is called crossover. The third step is mutation, where children are given a random change according to certain strategy. The last step mimics the nature select process, where the objective functions are evaluated for each child, and the children are sorted according to their corresponding objective

functions. This is a complete generation, and some good children candidate are allowed to continue the evolution.

In the early development, the multiobjective optimization problems (MOP) was converted to a single objective optimization problem by weighted sum method. Later, the truly multiobjective optimization with nondominated sorting was developed based on GA [5]. The detailed mathematical definition of dominance can be found in ref. [4, 2, 10]. It extends the comparison between two scalars to two vectors.

MOGA has been introduced into photoinjector design [2] and accelerator lattice optimization [6, 10]. The comparison of MOGA and GLASS is also shown in [9].

Algorithm 1 Multi-Objective Genetic Algorithms

- 1: Initialize population (first generation, random)
 - 2: **repeat**
 - 3: select parents to generate children (crossover)
 - 4: mutation(children)
 - 5: evaluate(children)
 - 6: merge(parents, children).
 - 7: non-dominated sort(rank)
 - 8: select half of (parents, children)
 - 9: **until** reach a generation with the desired convergence to the PO set
-

The structure of our MOGA implementation is shown in algorithm. 1. The first population is initialized with uniformly distributed random numbers, as we will see in storage ring lattice optimizations, most of these random populations at first did not give physical solutions due to transverse stable condition.

Two parents are chosen from the population, and used to generate two children. The newly generated values follows certain probability density function (PDF) as shown in Fig. 1. Following Ref. [4] we are using polynomial PDF with one parameter η to control the shape. This form is convenient to include the boundaries without artificial cuts when the new values are outside of it.

The “new born” children are applied with an operation of mutation, this mimics the effect from nature environment. We also choose a polynomial PDF to describe it. Fig. 2 shows the probability of the old value $x = -1$ will be mutated to. It has equal probability to go less or greater than -1 .

After the new generation is produced, we then evaluate the objective functions, which are the lattice functions in our case. The results are ranked based on their objective functions and the violation to the constraints. Here we also follow Dr. Deb’s approach [5], where Nondomi-

* Work supported by the Director, Office of Science, U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

[†] lyang@bnl.gov

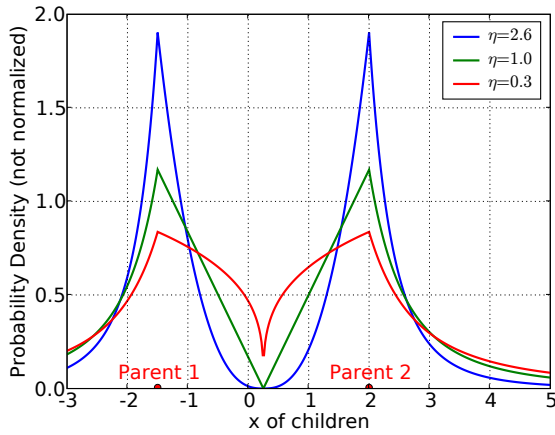


Figure 1: PDF for crossover. This shows the variable within range $[-3,5]$ and two parents are at -1.5 and 2 . η controls different shape of PDF.

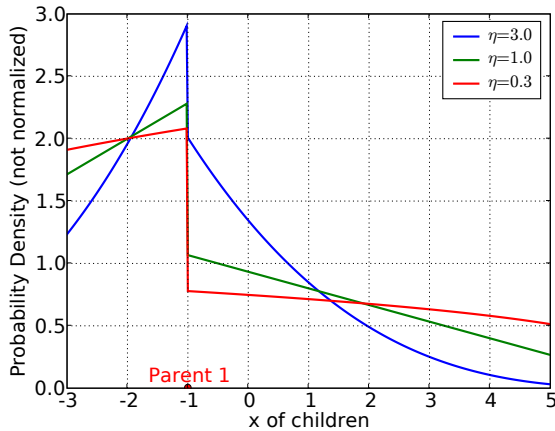


Figure 2: PDF for mutation. Here shows PDF of an example variable defined in range $[-3,5]$ with a value -1 . η controls the shape of PDF.

nated sorting is used to find the Pareto optimal. This process is repeated for the non-Pareto set. Those violating the constraints have the lowest rank. A simple way to do this would be building up a table (or matrix):

The table represents “dominance” relations between each pair. Those candidates with empty column are not dominated by any one and are called “nondominated set”. Obviously the nondominated set has a rank 1. The candidates who are violating the constraints are not shown in Table. 1. They can have a similar table, but anyone of them are dominated by anyone in Table. 1.

LATTICE OPTIMIZATION FOR ALS

We use Advanced Light Source (ALS) as an example to apply MOGA in lattice optimization problems. The ALS is a 3rd generation synchrotron light source located Computer Codes (Design, Simulation, Field Calculation)

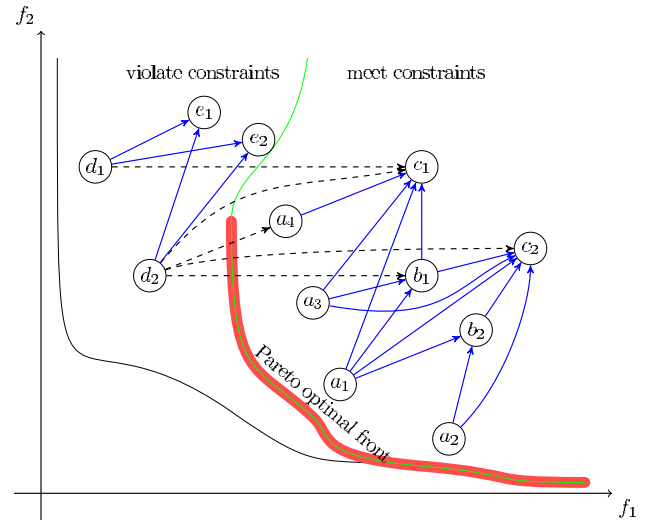


Figure 3: Nondominated Sorting. The candidates are in two group separated by the constraints. The arrows represent the relation of dominance, and the dashed arrow is valid if there are no constraints.

Table 1: Dominance table.

	a_1	a_2	a_3	a_4	b_1	b_2	c_1	c_2
a_1					1	1	1	1
a_2						1		1
a_3					1		1	1
a_4							1	
b_1							1	1
b_2								1
c_1								
c_2								

at Lawrence Berkeley National Laboratory optimized for the generation of soft x-ray. The ALS is 200 m in circumference and consists of 12 sectors. The lattice structure of each of the sectors is a triple bend achromat. All sectors are the same with the exception of three sections symmetrically distributed along the ring where the central dipole is superconductive. In this paper we only optimize the sectors with normal conducting dipoles, and the three superconductive bends then can be matched.

The first problem we applied to optimize is the emittance and beta function. In this problem, the emittance as one of the most important quantities of all light sources need to be minimized, while the beta function in this case want to be around 1 meter. The constraints are transverse stability, i.e. the one turn transfer matrix should have a trace in range $[-2,2]$, the maximum beta function less than 30 meters, and the maximum dispersion less than a few centimeters. The optimal results are shown in Fig. 4 and the corresponding brightness change are shown in Fig. 5.

The optimal results are in two isolated regions, and the history of evolution is presented in Fig. 6.

A second optimization on high-low beta is also carried

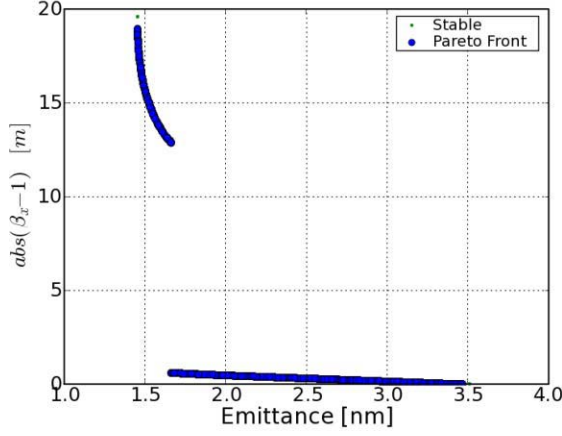


Figure 4: Emittance and beta function of the optimal solutions.

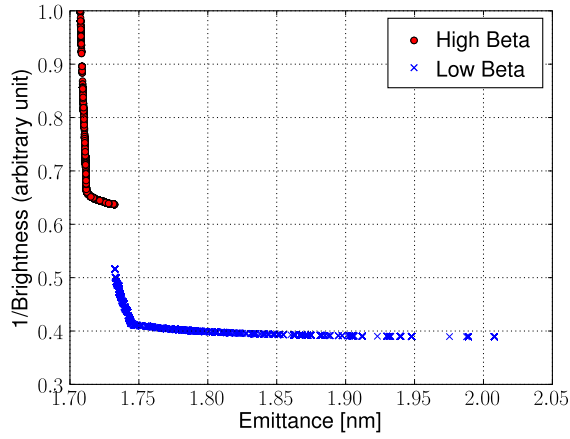


Figure 5: Brightness and emittance. The brightness is normalized by the maximum one in this plot.

out on ALS lattice. Two sectors are treated as one with high and low beta functions in each straight section. The low beta is about 1 meter and the high beta is about 10 meter. The third objective function is still emittance. Fig. 7 shows Pareto optimal set (now is a surface in 3D) projected into $\beta_{\text{high}}-\beta_{\text{low}}$ plane.

The lattice optimization problems usually have many constraints, eight in the dynamics or the practical way. As a dynamics system, many of the randomly generated value can not give a physical solution. This could be a serious problem for deterministic algorithms where extra efforts are needed when the predicted solution fails. This effort is first considering the constraints instead of objective functions, therefore is quite different direction from the original setup for objective functions. For MOGA, this kind of problem does not exist, since it is population based and uses the sorting to select/deselect candidates for the next iteration. In this way, it can easily survive the objective-functions-constraints conflict situations mentioned before.

Computer Codes (Design, Simulation, Field Calculation)

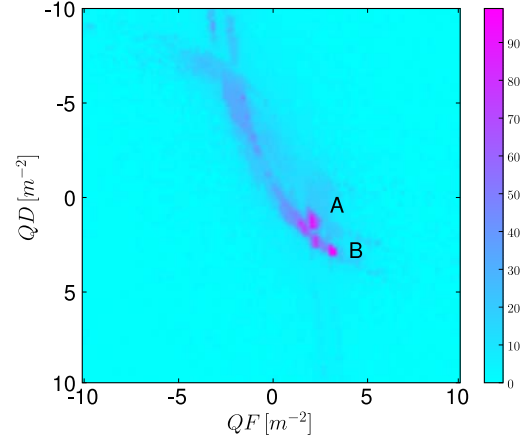


Figure 6: Evolution history in parameter space. The color represents how many generation the candidate in this region can survived without being replaced by better candidates

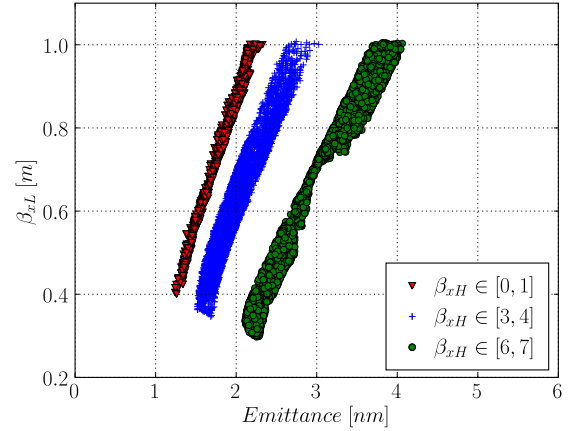


Figure 7: Low beta

In next section we also show a VHF gun optimization problem, which does not have this dynamics stability problem, but the computing time is long and fast convergence is quite important.

VHF PHOTOINJECTOR OPTIMIZATION

MOGA was also applied on VHF photoinjector optimization at LBNL [12]. The single calculation of beam quantity needs a few hours. This makes the algorithm requiring derivatives (approximation by finite difference method) not practical.

Since MOGA is population based, no interaction between evaluation of each candidate, it is very suitable for parallelization. We used the master-slave model to run MOGA on a cluster with 128 CPUs. All nodes carry out the beam simulation, while the master node do extra MOGA optimizations which is significantly small effort compared

with dynamics simulations.

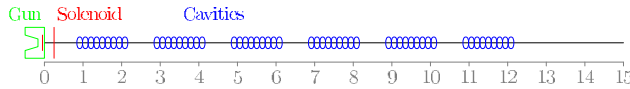


Figure 8: Caption

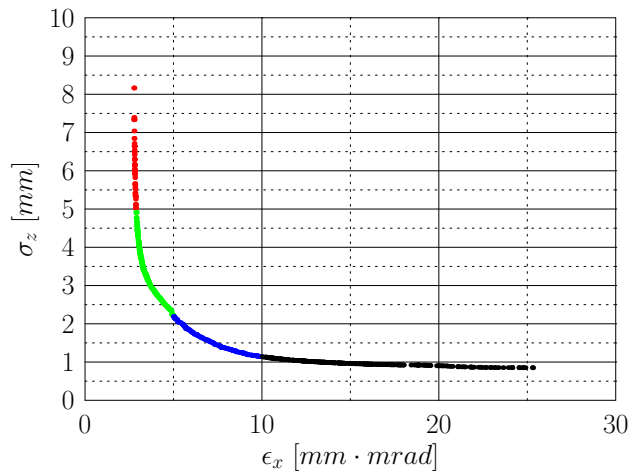


Figure 9: Caption

Fig. 8 shows the layout of a very simple case, to help our understanding of dynamics and various limitations []. Beam is launched from the left cavity, and pass through 6 cavities. The emittance and bunch length at $s = 15$ meter is obtained from Astra as objective functions. The final optimal solutions is shown in Fig. 9.

ACKNOWLEDGMENT

We thank the Susan James and Krishna Muriki from Scientific Cluster Support group.

REFERENCES

- [1] T. Back, U. Hammel, and H. Schwefel. Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1(1):3, 1997.
- [2] I. Bazarov and C. Sinclair. Multivariate optimization of a high brightness dc gun photoinjector. *Phys. Rev. ST Accel. Beams*, 8(3):034202, March 2005.
- [3] I. Das and J. E Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural and Multidisciplinary Optimization*, August 1997.
- [4] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Ltd, 2001.
- [5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation (IEEE-TEC)*, 6(2):182–197, 2002.

- [6] L. Emery. Global optimization of damping ring designs using a multi- objective evolutionary algorithm. Prepared for Particle Accelerator Conference (PAC 05), Knoxville, Tennessee, 16-20 May 2005.
- [7] D. E. Goldberg. *Genetic Algorithms for Search, Optimization, and Machine Learning*. MA: Addison-Wesley, 1989.
- [8] J. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.
- [9] D. Robin, F. Sannibale, C. Steier, W. Wan, and L. Yang. Global Optimization of the Magnetic Lattice Using Genetic Algorithms. EPAC'08, 11th European Particle Accelerator Conference, 23- 27 June 2008, Genoa, Italy.
- [10] L. Yang, D. Robin, F. Sannibale, C. Steier, and W. Wan. Global optimization of an accelerator lattice using multiobjective genetic algorithms. *Nuclear Instruments and Methods in Physics Research Section A*, 2009.
- [11] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: improving the strength pareto evolutionary algorithm for multiobjective optimization. In K.C. Giannakoglou, D.T. Tsahalis, J. Priaux, K.D. Papaliliou, and T. Fogarty, editors, *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems. Proceedings of the EUROGEN2001 Conference, Athens, Greece, September 19-21, 2001*, page 95100, Barcelona, Spain, 2002. International Center for Numerical Methods in Engineering (CIMNE).
- [12] F. Sannibale, K. Baptiste, J. Corlett, T. Huang, S. Kwiatkowski, D. Li, J. Qiang, J. Staples, R. Wells, L. Yang, A. Zholents, J. McKenzie. *Status of the LBNL Normal-Conducting CW VHF Photo-Injector*

APPLICATION OF DIRECT METHODS OF OPTIMIZING STORAGE RING DYNAMIC AND MOMENTUM APERTURES

M. Borland[†], L. Emery, V. Sajaev, A. Xiao, ANL, Argonne, IL 60439, USA
W. Guo, BNL, Upton, NY, 11973, USA

Abstract

Optimization of dynamic and momentum apertures is one of the most challenging problems in storage ring design. For storage-ring-based x-ray sources, large dynamic aperture is important in obtaining high injection efficiency, which leads to efficient operation and protects components from radiation damage. X-ray sources require large momentum aperture to obtain sufficiently long Touschek lifetimes with low-emittance beams. We have developed effective methods of optimizing dynamic and momentum apertures that rely directly on tracking using a moderately sized Linux cluster. After reviewing the method, we describe examples of its application to APS operations, upgrades, and next-generation storage rings.

INTRODUCTION

One of the most desirable characteristics of storage-ring-based x-ray light sources is low emittance. To achieve this, lattice designers use strong focusing to obtain large horizontal phase advance per cell, leading to large chromatic aberrations and thus strong chromaticity correcting sextupoles in order to obtain adequate momentum aperture (MA). In addition, low emittance means small dispersion, requiring yet stronger chromatic sextupoles. This leads to small dynamic aperture (DA), making it more difficult to accumulate beam. In extreme cases, the dynamic aperture may be so small that sufficient lifetime is not achieved.

Ring designers commonly add extra families of sextupoles to correct the effect of the chromatic sextupoles [1]. The challenge is to adjust the sextupoles to simultaneously maximize both DA and MA. Perhaps the most common approach is to minimize many resonance and tune variation driving terms [2]. However, one must carefully choose the weights for these terms, based on experience and, ultimately, tracking. Further, we commonly want non-zero linear chromaticity to suppress instabilities, which challenges the assumptions of the perturbative approach, since then one does not want the higher-order chromaticities to be minimized, but rather one needs to use them to reduce the chromatic tune spread.

In this paper, we discuss further a tracking-based optimization method [3] that has proven very successful and is a considerable improvement over previous attempts dis-

cussed in [4] and, in part, in [3]. Following an explanation of the method, we discuss application to the Advanced Photon Source (APS) storage ring and the NSLS-II ring.

Although our method could use any tracking code, the ability to create fully scripted simulations is essential, since matching and tracking must run without human intervention. Thus, we use the tracking program *elegant* [5, 6], as well as the SDDS Toolkit [7] and *geneticOptimizer* [8].

OPTIMIZATION METHOD

In this method we use many computers simultaneously to evaluate the DA and MA for various lattice tunings (e.g., tunes and sextupole settings). DA and MA computation includes radiation damping, synchrotron oscillations, and physical apertures. After completion of a sufficient number of evaluations, a genetic algorithm is used to “breed” more candidate configurations based on the best configurations seen so far. The process continues until a sufficiently good solution is obtained or until the results stop improving.

Dynamic Aperture

For the DA, we’ve found that the area of the stable region is a good parameter to use, with some limitations and conditions. We first determine the DA by performing line scans outward from the origin. (Scanning outward is used instead of scanning inward in order to avoid being fooled by stable islands.) Once the stable boundary is found, we analyze the boundary points to clip off any regions that “stick out” in a manner that indicates a poorly behaved boundary. An example is shown in Figure 1: The region that sticks out on the right side is probably related to a stable island and is not considered a useful contribution to the DA. Finally, having found the clipped DA boundary, we compute its area and its contribution to the penalty function. Because the area computation ignores contributions from useless regions, the optimized results are unlikely to display such regions.

The contribution to the penalty function is computed by comparing the area A to the desired area A_d using a weighting factor ΔA

$$P(A) = \begin{cases} (A - A_d)^2 / \Delta A^2 & A < A_d \\ 0 & A \geq A_d \end{cases} \quad (1)$$

For APS we typically want an aperture $-13\text{mm} \leq x \leq 7\text{mm}$ and $|y| \leq 1.5\text{mm}$, giving $A_d = 30\mu\text{m}^2$.

In some cases, the DA area may be misleading, for example, a solution with large vertical aperture but small hori-

* Work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[†] borland@aps.anl.gov

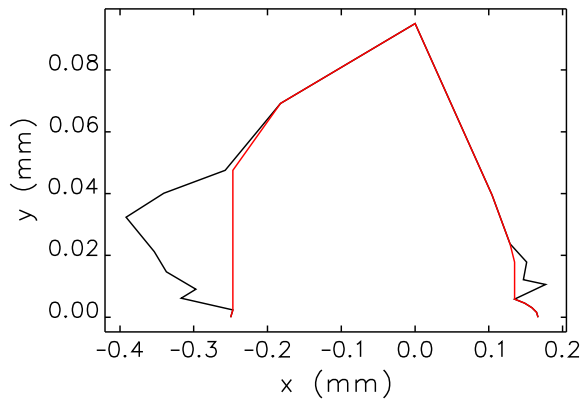


Figure 1: Illustration of algorithm for removing pathological features from DA results. The red boundary is used for DA area computations.

zontal aperture. To prevent this, we limit the vertical search region to the maximum desired vertical DA.

Momentum Aperture

To assess the Touschek lifetime, we need to know the MA for particles scattered at many positions around the ring, i.e., the s -dependent MA [9]. We don't need the MA for every lattice element, but simply for a representative set. A good choice for a double-bend cell is to compute the MA at several locations on either side of the dipoles. For the APS, we compute the MA at the exit of the S1, S3, and S4 sextupoles in the first six (out of 40) sectors.

We compute the MA's contribution to the penalty function in a fashion similar to what was done for the DA:

$$P(\delta_{min}) = \begin{cases} (\delta_{min} - \delta_{des})^2 / \Delta\delta^2 & \delta_{min} < \delta_{des} \\ 0 & \delta_{min} \geq \delta_{des} \end{cases} \quad (2)$$

where δ_{min} is the minimum absolute value of the MA at any location, δ_{des} is the desired value, and $\Delta\delta$ is a weighting factor. For APS, we typically choose $\delta_{des} = 2.35\%$, equal to the bucket half-height.

An alternative is to compute the Touschek lifetime from the s -dependent MA using the `touschekLifetime` program [10], which operates on output from `elegant`. As of now, this additional step has not seemed necessary.

Importance of Errors

It is well known that the DA and MA are affected by magnet strength and alignment errors. For example, without errors the effect even of the half-integer resonance may not be seen in a tracking simulation. Hence, we must include lattice errors in the simulations. It would seem that in order to be realistic, we must not only include errors, but also correct those errors using real-world techniques.

Effective methods exist for correcting linear optics [11] and coupling (e.g., [12, 13, 14]), which is important in light Single Particle Beam Dynamics and Optics

sources because of the small insertion device vertical apertures. In the APS, for example, we correct lattice function errors to the 1% rms level [15] and coupling to the 1% level.

However, correction is not essential in the simulations. Instead, we simply use random errors that give lattice function and coupling errors at post-correction levels. This neatly side-steps a considerable complication.

To prevent the optimizer from being misled by variations resulting from different ensembles, we use a fixed error ensemble for all simulations. We impart strength and roll errors to quadrupoles and sextupoles only, which gives all the essential features of a post-correction lattice. Strength errors are typically 0.02% rms, which typically gives lattice function errors of 1% rms or more. Roll errors are typically 0.5 mrad rms, which gives coupling of roughly a few percent. We prefer to make the errors somewhat on the high side, since this helps ensure a robust solution. Following optimization, we evaluate the lattice with typically 20 to 50 ensembles to verify that the solution is robust. This is a simple precaution that can be easily carried out with, e.g., `elegantRingAnalysis` [16]. To date, we have seen only one case where this step yielded a surprise. This was traced to a poor choice of integer tunes.

Software Details

The optimization uses the general-purpose script `geneticOptimizer`. The user provides an input file listing the independent variables along with their initial values, allowed ranges, and randomization levels. Typically the independent variables are the tunes and the sextupole strengths. (Although not required by our method, allowing the tune to vary has been very effective and is advisable.) The user also provides two scripts, one to run a configuration and another to postprocess it. The first script (the "run script") performs any required matching and assembles the full lattice, then performs tracking to determine the DA and MA. The second script postprocesses the results and returns the value of the penalty function.

As an example, for some future long-straight-section lattices for the APS, the run script invokes `elegant` four times to perform matching of different types of sectors. Rematching permits wide variation of the tune without changing other essential lattice features, such as maximum lattice functions or lattice functions at the insertion devices. It also ensures that there are no spurious lattice function errors. Following matching, the sector solutions are combined into a ring solution that is evaluated for DA/MA in a single run. Often this run also performs the final chromaticity correction using whichever sextupoles are left free.

Executing and combining the results of multiple program runs is greatly facilitated by `elegant`'s thorough use of SDDS files and the SDDS toolkit. It permits complete automation of the process, regardless of the complexity, using only commands in a script.

APS OPTIMIZATION

The APS storage ring has 280 sextupole magnets with individual power supplies. Because of the symmetry of the lattice, in the past these sextupoles have been powered in four families. Because we run in modes with fairly high single-bunch current while lacking a bunch-by-bunch feedback system, we operate with significant non-zero chromaticities $\xi_x = d\nu_x/d\delta$ and $\xi_y = d\nu_y/d\delta$. In 24-bunch, 100-mA mode, we have $\xi_x \approx 7$ and $\xi_y \approx 6$, while for hybrid mode we require $\xi_x \approx \xi_y \approx 11$ to achieve the required 16-mA single-bunch current.

Operational Lattices

We previously reported [3] on successful application of this technique to improvement of the APS operational lattice, where we realized a 25% improvement in lifetime for the 24-bunch mode. One of the most surprising aspects of this work was that in several cases the optimization clearly favored a sextupole configuration that did not have the same symmetry as the lattice itself. One advantage of the new configuration is that the lifetime is now long enough to be used in 324-bunch, non-top-up mode, which reduces the number of lattices we must maintain. Since then, we've also optimized the hybrid mode lattice and obtained a 10% improvement in lifetime, which is less than the predicted improvement of 20%, but still significant.

APS Renewal Lattices

The APS Renewal is a project to update the APS accelerators and beamlines. One of the most interesting aspects of the accelerator improvements is the provision of a number of long straight sections (LSS). These will allow 7.7 m for insertion devices compared to 4.8 m at present. We previously reported [3] on development of several such lattices and successful testing of a mockup lattice with eight symmetrically placed, emulated long straight sections.

Since then, we've explored several additional lattices. Among these are four groups of two long straights separated by a short straight (4x2LSS) and 10 symmetrically placed long straights (10LSS). Both of these yielded solutions that can be expected to work in practise. One advantage of our optimization method is that it works with little human involvement beyond initial setup, which is similar for most configurations. Hence, we can quickly look at many different possibilities. Further, if the optimization converges, the solution is almost certain to be valid as it is based to begin with on tracking.

The 4x2LSS configuration is a complex case that illustrates how the commandline nature of `elegant` and SDDS contribute to the implementation. APS has two types of sectors: those with Decker distortion [17] (sectors 1 through 35) and those without (sectors 36-40). Hence, we must match two kinds of Decker-distorted sectors, one with a short straight on both ends and a second with a short straight on one end and long straight on the other. A third

solution is required for the non-distorted sector. We constrained all sectors to have very similar tunes, finding a solution for the normal sector first and then matching the other sectors to that solution. Using SDDS tools allowed us to overlay the reflected short-to-long solution on the sectors requiring long-to-short optics, which is necessary because we wish to independently vary all the sextupoles without symmetry constraints and hence have different element names in these sectors. Following this, we match the non-distorted sector to the same initial and final lattice functions and the same phase advance as the Decker-distorted sectors. The linear optics solutions are then loaded over the full lattice and the sextupoles are set according to the values provided by the optimizer. The free sextupoles are adjusted to give the desired chromaticity, subject to sextupole strength limits. Finally, we track to obtain the DA and MA. The DA, MA, and chromaticity are then used in the penalty function computation, as described above.

NSLS-II OPTIMIZATION

The NSLS-II ring is currently in the construction stage. The lattice of this 3-GeV machine features 30 sectors with alternating long and short straight sections. Strong damping wigglers are employed to bring the emittance below the 1-nm level. Considerable work has been done on design and optimization of this lattice for near-zero chromaticity [18, 19] using minimization of resonant driving terms. Here, we report results of optimization using our direct technique. Unlike the previous work, we allowed all sextupoles to vary independently. (Hence, our results do not necessarily indicate an advantage of our method.) This choice was based on experience with APS optimization, which taught us that symmetry constraints on the sextupoles are not necessarily desirable [3].

We began with a somewhat mistuned lattice having $\xi_x \approx 1$ and $\xi_y \approx 3$, but exhibiting poor DA and MA. The first stage of optimization targeted $\xi_x = \xi_y = 2$, a DA area of $100 \mu\text{m}^2$, and an energy aperture of 3% using 40 processors, with tunes free to vary. The optimization converged after about 500 runs (see Figure 2), moving the tunes from $\nu_x = 33.1$ and $\nu_y = 16.20$ to $\nu_x = 33.25$ and $\nu_y = 16.09$. This optimization gave an unnecessarily large vertical aperture, so we modified the parameters to only scan to 1.5 mm in the vertical instead of 3.0 mm, and re-optimized starting from the previous best value, thus increasing the horizontal aperture. We evaluated the result with 20 random ensembles (using the typical error levels given above), giving excellent results as seen in Figures 3 and 4. Note that multipole errors have yet to be included.

Starting from the $\xi_x = \xi_y = 2$, we continued the optimization with a target of $\xi_x = \xi_y = 4$. The resulting dynamic aperture was nearly identical to that shown in Figure 3, while the momentum aperture was slightly reduced. Obtaining this new solution was relatively effortless, requiring changing only a few values in the penalty function.

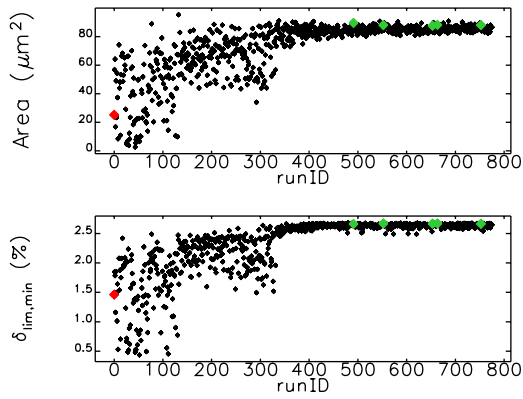


Figure 2: Progress of initial optimization of NSLS-II for $\xi_x = \xi_y = 2$. To aid convergence, the error level was adjusted by hand at around run 150 and again at around run 350.

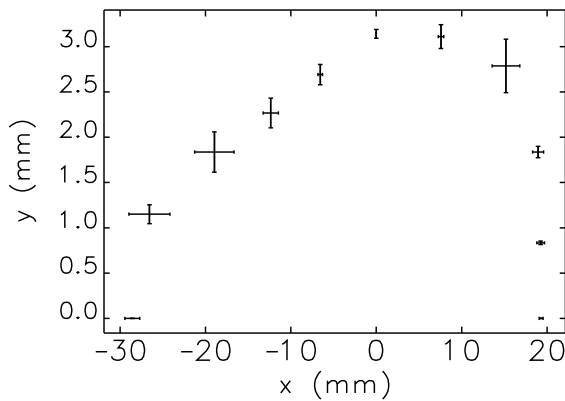


Figure 3: Dynamic aperture for 20 ensembles for final optimization of NSLS-II for $\xi_x = \xi_y = 2$.

FUTURE DEVELOPMENT

Although this technique is highly successful, there are still opportunities for improvement. Using parallel DA and MA computation[20], we could use many cores for each job submitted by the genetic optimizer, which promises convergence in hours instead of weeks. We could also perform optimization with several error ensembles rather than a single ensemble, in order to further ensure robustness of the final result. We are interested in trying other optimization algorithms besides the genetic technique, e.g. parallel simplex [4]. A related option is to use elegant's internal optimizer, which will be possible in the near future once DA and MA optimization is supported in the parallel version. As mentioned above, a refinement of the penalty function would be to compute the momentum aperture directly using the `touschekLifetime` program. One could also envision simulating injection efficiency instead of only the dynamic aperture, which might yield interesting new solutions with highly asymmetric dynamic apertures (i.e., large on the side where beam is injected, but smaller on the

Single Particle Beam Dynamics and Optics

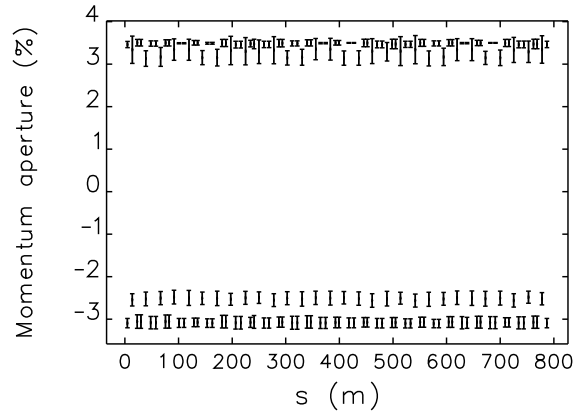


Figure 4: Momentum aperture for 20 ensembles for final optimization of NSLS-II for $\xi_x = \xi_y = 2$.

other side). Incorporating lattice and coupling correction is also desirable because it will ensure that choices of tune are not overly constrained by correctable effects of errors.

CONCLUSION

We have developed a practical and robust tracking-based method of optimizing storage ring nonlinear dynamics. Details of the algorithm and implementation were presented. An update on APS-related applications was given, followed by examples of successful application to the NSLS-II ring design.

REFERENCES

- [1] E. A. Crosbie, Proc. PAC87, 443-445 (1988).
- [2] J. Bengtsson, SLS Note 9/97 (1997).
- [3] M. Borland *et al.*, Proc. PAC09, TH6PFP062, to be published.
- [4] H. Shang and M. Borland, Proc. PAC05, 4230-4232 (2005).
- [5] M. Borland, APS LS-287, September 2000.
- [6] M. Borland, these proceedings.
- [7] M. Borland, Proc. PAC95, 2184-2186 (1996).
- [8] M. Borland and H. Shang, unpublished program.
- [9] M. Belgroune *et al.*, Proc. PAC03, 896-898 (2003).
- [10] A. Xiao and M. Borland, Proc. PAC07, 3453-3455 (2007).
- [11] J. Safranek, NIM A 388, 27 (1997).
- [12] P. Bagley *et al.*, Proc. PAC89, 874-876 (1990).
- [13] J. Safranek and S. Krinsky, Proc. PAC93, 1491-1493 (1993).
- [14] L. Emery *et al.*, Proc. PAC03, 2330-2332 (2003).
- [15] V. Sajaev and L. Emery, Proc. EPAC 2002, 742-744 (2002).
- [16] M. Borland, Proc. PAC05, 4200-4202 (2005).
- [17] G. Decker and O. Singh, Phys. Rev. ST Accel. Beams 2, 112801 (1999).
- [18] S. Kramer and J. Bengtsson, Proc. PAC05, 3378-3380 (2005).
- [19] W. Guo *et al.*, Proc. PAC09, TU5RFP008, to be published.
- [20] Y. Wang *et al.*, these proceedings.

RadTrack: A USER-FRIENDLY, MODULAR CODE TO CALCULATE THE EMISSION PROCESSES FROM HIGH-BRIGHTNESS ELECTRON BEAMS*

M. Ruelas, G. Andonian, RadiaBeam Technologies, LLC, Marina Del Rey, CA, USA
S. Reiche, Paul Scherrer Institute, Switzerland

Abstract

One of the most important goals of simulations is to accurately model beam parameters and compare results to those obtained from real laboratory diagnostics. Many codes are specialized to either model beam dynamics or emitted radiation. For meaningful physical results, the output of these codes are stitched together in start-to-end fashion. This procedure, which is often employed by simulation experts, is cumbersome, and has wide room for error in data entry or file parsing. This paper describes the development and deployment of RadTrack: a user-friendly code, with start-to-end support of typical accelerator and radiation codes to accurately model laboratory diagnostics.

INTRODUCTION

The code RadTrack was developed to accurately model observable beam parameters in a real laboratory environment. The code emphasizes modularity to address a comprehensive set of problems and an easily navigable user-interface to attract a wide user base. The graphical user-interface is built on a visualization canvas that easily generates and displays important information. The interface is intuitive for seamless management of start-to-end simulations, which incorporate several codes of varying I/O context. The interface allows for simple parallelization for complex, memory demanding calculations. RadTrack was developed as a code that can calculate beam dynamics and emitted radiation processes in a transparent, intuitive manner accessible to most accelerator scientists and students.

RADTRACK CORE

The code RadTrack was first developed as an extension to the radiation code QUINDI [1] to calculate the radiative effects of bending beam trajectories. The code QUINDI was developed for a specific problem and its results have been benchmarked to experiments at the Brookhaven National Laboratory Accelerator Test Facility [2]. RadTrack builds upon the code in a number of ways, while also incorporating other desirable features.

The RadTrack core code is broken down into a number of modular steps. The particle trajectories are calculated using Q-Tracker, an extension to the code QUINDI. Q-Tracker is a simple particle tracker, with trajectories determined by the Lorentz force law, which outputs the 6-dimensional phase space used by RadTrack. The radiation

field solver is a modified version of the existing code QUINDI. The radiation emission is calculated using the Lienerd-Wiechert potentials [3]:

$$\vec{E}(r, t) = \frac{e}{\sqrt{4\pi\epsilon_0}} \left[\frac{\vec{n} - \vec{\beta}}{\gamma^2 (1 - \vec{\beta} \cdot \vec{n})^3 R^2} \right]_{\text{ret}} + \left[\frac{\vec{n} \times [(\vec{n} - \vec{\beta}) \times \dot{\vec{\beta}}]}{c (1 - \vec{\beta} \cdot \vec{n})^3 R} \right]_{\text{ret}}$$

where \vec{n} is the unit vector pointing from the radiation point to an observation point and R is the distance to the observation point. The magnetic field is derived from

$$\vec{B}(r, t) = [\vec{n} \times \vec{E}(r, t)]_{\text{ret}}$$

The fields in the above relations are calculated at the retarded time $t' = t + R(t)/c$.

The RadTrack modular approach separates the functions of particle trajectory calculation and radiation field solving. Figure 1 displays the modular philosophy employed by the code where individual functions are separated to allow for in-depth, comprehensive problem analysis. This is advantageous for implementation of the start-to-end function, where multiple outputs of codes are parsed as inputs into subsequent codes. For example, the user may use particle trajectories from other codes, like TREDI [4], in conjunction with the radiation solver QUINDI or the trajectories from Q-Tracker with another radiation code. Efforts

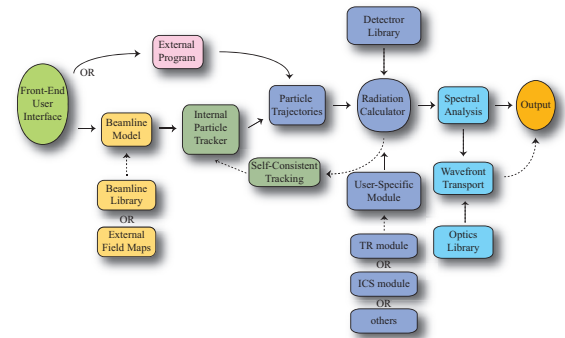


Figure 1: Flow diagram of the RadTrack code design. The ultimate goal is to simulate real laboratory diagnostic observables using the computation tools available to the user (acceleration, radiation, transport, etc.).

* Work supported by DOE Grant No. DE-FG02-08ER85018

have been directed at including parsers for trajectories from well-established, popular codes like PARMELA [5], ELEGANT [6], and TREDI.

User Interface

The RadTrack user interface is also modular, separated into panels with distinct features. The panels are intuitively designed to contribute to a smooth workflow.

Particle Distribution RadTrack incorporates a useful beam distribution panel. Initial beam parameters (used for the source particles) are defined and a 6-dimensional phase space distribution is generated. The user specified inputs include beam moments, correlations, various degrees of noise, and complex modulations or the Courant-Snyder parameters. This panel also supports the importing of 6-dimensional particle distributions from some external codes. The built-in parser also offers rudimentary distribution analysis and graphical representations of the beam. Alteration of the beam parameters is handled by textual input or by sliders and buttons on the graphical display of the beam, allowing a very intuitive method of creating unique and specific particle distributions (such as rescaling Twiss parameters).

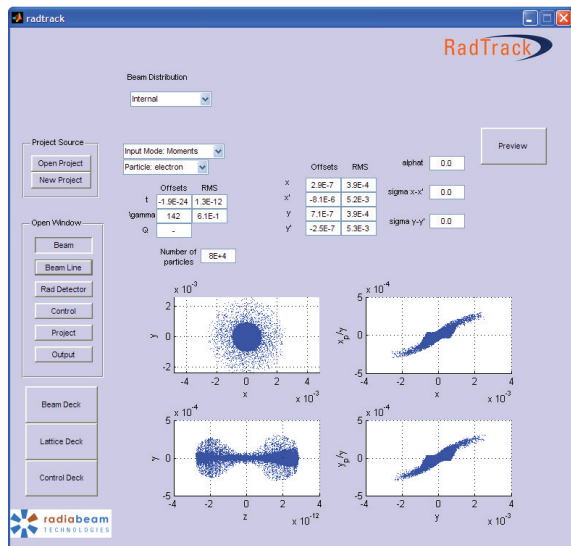


Figure 2: Screenshot of the beam distribution panel in RadTrack displaying transverse profiles, and transverse and longitudinal correlations of a user generated source beam.

RadTrack supports a number of particle distribution sets. Gaussian beam distributions are generated by a normalized Box-Muller transformation [7] of a uniform distribution. The final distribution is formed by dialation, rotation and shearing of the distribution. Other distributions (such as waterbag, K-V, boxcar, etc.) are also supported, however in a basic, limited fashion. The quiet start [8] has been investigated and has recently been shown to be of great importance because it reduces artificial beam microbunching

Computer Codes (Design, Simulation, Field Calculation)

that leads to overestimates of the beam emittance and quality.

Beamline Visualization Aide The definition of the beamline lattice is addressed graphically in RadTrack. Beamline elements, such as drifts, bending magnets, and focusing magnets, are selectable and editable via a palette. The reference particle trajectory is calculated and plotted through the displayed beamline for straight-line visualization. The graphical output of this aide is a beamline of block shapes representing different magnetic elements while the textual output is tab-delimited text file (input deck) of the beamline lattice for use in Q-Tracker. Parsers for other codes are also available.

RadTrack allows for the placement of the detector for radiative processes at arbitrary locations, defined by beam offsets and rotations in the from of Euler angles. This allows for the straightforward modeling of bending radiation, such as synchrotron or edge radiation, as the user may lock the detector plane tangential to the reference trajectory. Modeling specific diagnostics, such as those for other types of radiation sources (including Compton, THz, etc.) requires using this visually descriptive method for detector placement (Figure 3).

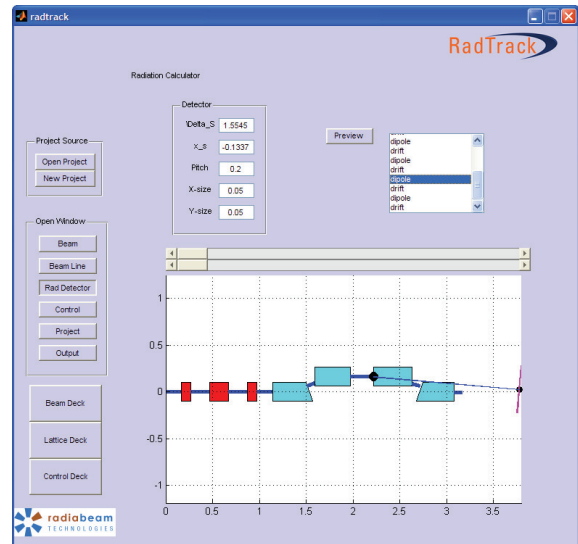


Figure 3: Screenshot of the user interface displaying the beamline constructor with a radiation detection screen. This example of the BNL ATF chicane compressor was created to benchmark the code to laboratory data.

Parallel version Simulation project management is built directly into the RadTrack framework to aid in extending the computational capabilities from a single processor to a multi-node computing cluster. The management of files allows for multiple runs based on previous input data sets and outputs in a variety of formats.

The ability to run single particle calculations and highly detailed design studies, requiring the use of a cluster, is ac-

complished using a single interface. Also, start-to-end simulations are executed, and I/O file parsing is accomplished, through the project manager. This feature is useful because the user will experience a consistent, undisturbed workflow while the code handles the necessary parsing and I/O interplay between the codes. Additionally, the file management of the various output files of each code are easily accessible, stored and organized by the software. The parallel version addresses the need to simulate large number of particle sets for complex problems (such as microbunching, and coherent transition radiation at optical wavelengths).

Post-processor The RadTrack post-processor was inspired by the demand to retrieve useful information for the simulation in a timely fashion. It displays spectral information using simple, built-in mathematical function tools. This gives the user valuable information on the validity of the performed run before further, extensive analysis is completed. It is envisioned that this unit will be incorporated into a real laboratory control room to provide near real-time beam information to be used by the experimenter for data taking or for use in a feedback loop to ensure high quality beam control.

Due to its modular nature, new module development to solve specific radiation problems is straightforward to incorporate. Efforts in radiation transport and Inverse Compton Scattering solvers are underway. Improvements in space charge and coherent synchrotron radiation effects are also being investigated to enhance the capabilities of RadTrack.

REFERENCES

- [1] D. Schiller, et al., Proc. Particle Accelerator Conference 2007, Albuquerque, N.M., p.3612 (2007).
- [2] G. Andonian, et al., Phys. Rev. ST Accel. and Beams, vol.12, p.030701 (2009).
- [3] J.D. Jackson, *Classical Electrodynamics*, Wiley, New York (1999).
- [4] Giannessi, et al., AIP Conference Proceedings 1997, vol. 413, p. 313 (1997).
- [5] Billen and Young, Los Alamos National Laboratory Tech. Report LA-UR-96-1835, LANL (rev. 2000).
- [6] M. Borland, APS Tech. Report LS-287 (2000).
- [7] Box and Muller, Ann. Math. Stat. 29, p.610-611 (1958).
- [8] J. Gonichon, et al. Proc. Particle Accelerator Conference 1993, Washington D.C., p. 2696 (1993)

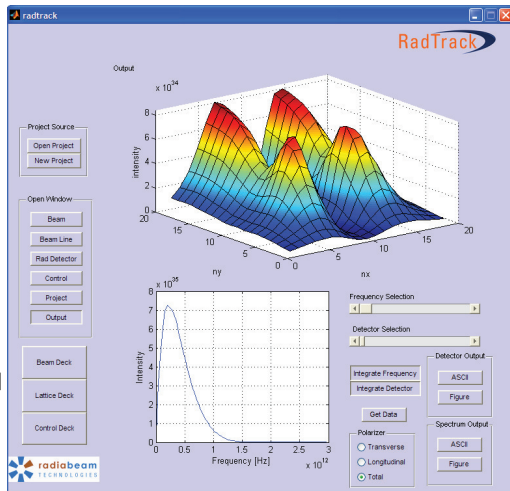


Figure 4: Screenshot of the post-processor unit displaying a calculation performed for the BNL ATF experiment..

CONCLUSIONS

RadTrack is a user-friendly tool used for the calculation of beam trajectories and emitted radiation of high brightness beams. The novel user-interface is accessible to a wide range of users and incorporates intuitive features such as the visualization of beam phase space densities and the graphical display of beamline lattice files in real time. It also incorporates a seamless method for start-to-end simulations and parallel extensions via the project management aide.

Computer Codes (Design, Simulation, Field Calculation)

TOMOGRAPHIC RECONSTRUCTION OF A BEAM PHASE SPACE FROM LIMITED PROJECTION DATA*

G. Asova[†], S. Khodyachykh[‡], M. Krasilnikov, F. Stephan, DESY, 15738 Zeuthen, Germany
I. Tsakov, INRNE BAS, Sofia, Bulgaria

Abstract

The production of electron beams suitable for the successful operation of the European XFEL is studied at the Photo-Injector Test Facility at DESY, Zeuthen site (PITZ). The PITZ beamline is equipped with three dedicated stations for transverse emittance measurements and in the forthcoming shutdown period a section for transverse phase-space tomography diagnostics will be installed. The module contains four observation screens and therefore only four projections can be used in order to reconstruct an underlying phase-space density distribution.

This work presents the performance of a number of reconstruction algorithms on limited projection sets using numerical data applied to the PITZ operating conditions. Different concepts for comparison between an original phantom and the reconstructed distribution are presented.

INTRODUCTION

The PITZ facility is dedicated to the development and optimization of electron sources subsequently to be used in FELs like the FLASH and the future European XFEL. Such goals require detailed knowledge of the electron beam properties according to which the PITZ beamline is equipped with extensive diagnostics components. A key element for the performance of a FEL is the small transverse emittance, wherefrom the transverse phase space is a central point in the electron source characterization at PITZ. Currently, the transverse phase space is being reconstructed using single slit scan technique [1] and a new module for transverse phase space tomography diagnostics will be installed in the forthcoming 2009-upgrade.

The module consists of four stations as each two surround a FODO cell. Correspondingly, four projections are to be used for tomographic reconstruction. The design has been discussed in [2] and expectations towards its performance with nominal beam parameters of 1 nC bunch charge, 32 MeV/c momentum and normalized transverse emittance of 1 mm mrad can be found in [3]. The setup will also be used in a combination with a transverse deflecting cavity structure to study the longitudinal phase space of individual pulses within the bunch train. In any case the choice of proper reconstruction algorithm is of great importance.

This work focuses on the performance of a few reconstruction algorithms with respect to their applicability to limited input projection data. The methods discussed are Filtered Backprojection (FBP), Constrained Additive Algebraic Reconstruction Technique (caART) and Maximum Entropy (MENT). Several approaches to quantify the quality of the reconstruction conclude the contribution.

TRANSVERSE PHASE-SPACE TOMOGRAPHY OF AN ELECTRON BEAM

Tomography deals with the reconstruction of an n -dimensional object knowing an infinite number of its $(n - 1)$ -dimensional projections calculated at different view angles in $[0, \pi]$. A great number of scientific and practical areas are using the tomography ideas - medical imaging is interested in innocuous cross sectioning of the human body, archaeology needs non-destructive material inspection.

The object of interest in the transverse beam dynamics is an underlying density distribution $\rho(x, x', y, y')$ at a given position along the beamline. The density distribution cannot be obtained instantly but its spatial components are directly measurable by means of screens, wire scanners, etc. Meeting an observation screen, for instance, the four-dimensional phase space is projected onto a spatial distribution (x, y) . A number of projections of the spatial distribution, taken at different angles, are needed for the reconstruction and, therefore, one needs to vary the orientation of the phase space on the screen. The last is equivalent to rotation of the beam in the phase space and is achievable by altering the focusing conditions using magnets. Let the system be linear such that M denotes a valid 2×2 transformation matrix from the position of reconstruction z_i to the position of observation z_f and $p(x_f)$ is a projection onto the horizontal axis at z_f . The condition on the linearity should be interpreted so that the matrices M describe well the transport between the two longitudinal positions. The projection can be written as a function of the initial phase-space coordinates as the Radon transform

$$p(x_f) = \iint \rho(x_i, x'_i) \delta(x_f - M_{11}x_i - M_{12}x'_i) dx_i dx'_i. \quad (1)$$

The problem to be solved is, having a number of $p(x_f)$ with different matrices M , to find a unique inversion of the Radon transform. Disregarding any intrinsic measurement errors, the singularity of the solution depends on the number of projections, the equidistant steps between each

* This work has partially been supported by the European Community, contract No. RII3-CT-2004-506008 and 011935.

[†] galina.asova@desy.de

[‡] Presently at Siemens AG, Rudolstadt, Germany.

two of them and the linearity of the system. While the latter might be hard to cope with and the first two strongly depend on the hardware setup, the overall result can be optimized applying a suitable reconstruction algorithm.

RECONSTRUCTION ALGORITHMS

Backprojection (BP) and its derivative Filtered Backprojection (FBP) [4] attain fast computations and are simple to implement making them two of the most commonly used reconstruction methods. BP employs direct inversion of the Radon transform (1) as it iteratively smears each projection onto the position of reconstruction according to the fact that the density at a point can be defined as integration over the line integrals from different projections passing through this point or

$$\rho(x, x') = \int_0^\pi p_\theta(x_\theta) d\theta \quad (2)$$

for θ determining the transformation. Defined in this way, the inversion of the Radon transform is influenced by blurring effects due to the fact that each point might be added more than once since it might contribute not to a single line integral in the space (x, y) . The FBP uses an additional low pass filter which introduces negative values in each projection. The latter ones are needed in order to correct for projections from other angles. The filter is applied in advance to the iterative smearing.

When applied to limited data sets the performance of BP and FBP is unsatisfactory. This is shown in Fig. 1, where

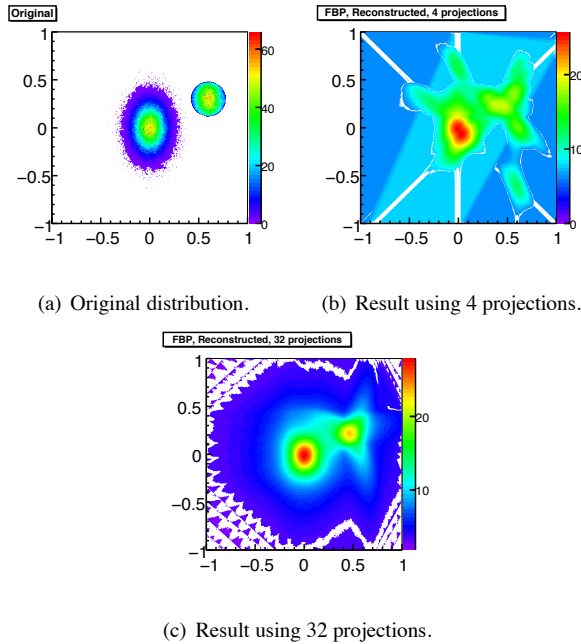


Figure 1: Comparison between original distribution and its reconstructed ones using FBP with different number of input projections. The ridges on the reconstructed distribution decrease with the number of angular steps.

an original phantom distribution is given beside the reconstruction result of FBP having four projections, obtained with equidistant rotations over π - Fig. 1(b). Fig. 1(c) shows that the quality of the reconstruction in terms of streaking artifacts improves with the increase of the number of projections. For that particular example each projection is convoluted with a Butterworth low-pass filter [4] prior the integration. Filtering out low intensity values would not be sufficient since the uniform spot on the right hand side of Fig. 1(a) is still not well defined even in the case of 32 projections.

FBP is not a recursive algorithm - the process of reconstruction depends on one projection at an iteration step and can consequently be completed disregarding any supplementary input. The availability in advance of all projections to be used is required for recursive algorithms like ART [5], as MENT [6] is regarded here as a derivative of the algebraic techniques solving a minimization task in a different manner.

The ART, as the name implies, uses a matrix-like indirect approach to the inversion problem - the different projections are considered as a set of linear equations with the values of the function to be reconstructed as unknown variables. If the wanted density distribution is described as constituted of K pixels and w_{nk} represents the contribution of the k -th pixel to the n -th projection, with n denoting some of the available N projections, a projection can be written in the form

$$p_n = \sum_{k=1}^K w_{nk} \rho_k. \quad (3)$$

Solution of a system of linear equations like Eq. (3) is to be found.

ART is an iterative algorithm - the wanted density of a bin is calculated over a number of steps as on each step projections of the current guess are snapped. The repetitive procedure continues until the computed projections resemble the given ones according to some set of criteria. Additive ART (aART) applies a correction to the k -th pixel on the $(i + 1)$ iteration step of the kind

$$\Delta \rho_k^{(i+1)} = \left(\frac{p^{(i+1)} - q^{(i+1)}}{\sum_{j=1}^K w_{(i+1)j}^2} \right) w_{(i+1)k}, \quad (4)$$

where $q^{(i+1)}$ is the projection calculated from an iterative guess $(i + 1)$ and $p^{(i+1)}$ is the one, given in advance from the measured data.

If the system described by the equations like (3) is underdetermined, i.e. the number of pixels is more than the number of projections, a unique solution does not exist. Such is the case for the PITZ setup of four screens and also for reconstruction applied on double-quadrupole scan data. On the other hand, there might be none or multiple solutions for an overdetermined system. The MENT algorithm gives

a possibility to select an outcome most consistent with the measured data. As it is used here MENT has already been described in [6, 3].

Fig. 2 shows the results of aART and MENT applied to the same four projection as in Fig. 1(b). An additional constraint for non-negative pixel content has been applied to the ART, usually known as Constrained Additive ART (caART), and MENT requires such by definition. The qual-

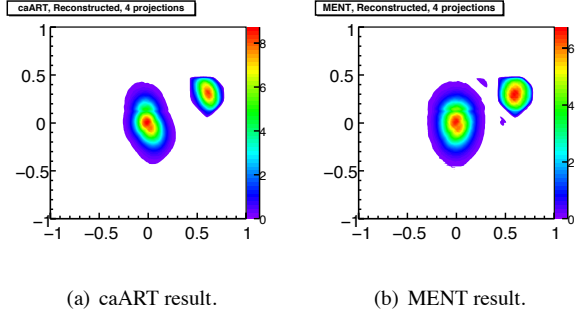


Figure 2: caART and MENT applied to four equidistant rotations. The object of reconstruction is the one in Fig. 1(a).

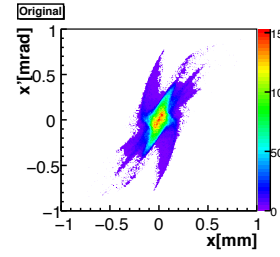
ity of the resulting distributions is visibly improved as expected according to the fact that a minimization task would be needed for an underdetermined system. Smearing artifacts like those that can be seen in the outcome of FBP are not present, the projected area of the central gaussian distribution is refined as well as is the uniform spot on the right hand side.

RECONSTRUCTION FROM NUMERICAL DATA

The decision for choosing a suitable reconstruction algorithm depends on its performance on numerical data. Here the caART and MENT are applied on a simulated electron beam distribution, matched to the optics of the lattice and tracked with ASTRA [8]. The influence of space-charge forces is included in the tracking as they tend to be significant for the PITZ operating conditions. The periodicity of the particle trajectories is expressed with the mismatch of the measured Twiss β -functions from the design values at the positions of observation. The numerical tracking reveals maximum β -mismatches of 3 and 6% for the horizontal and the vertical planes respectively. Fig. 3 shows the original phase-space and the resulting reconstructions for the horizontal plane. Both methods manage to reproduce the double-like structure in the core of the beam as visually MENT surpasses in better restoration of the density in the tails. Table 1 contains the relative deviations of the reconstructed distributions with respect to the original.

As an alternative measure of the quality of the reconstruction, constructive for distributions non-symmetric in neither of the x or x' planes, the skewness could be used.

Except with the second central moments and the covariance, needed in order to determine the transverse emit- Beam Dynamics, Other



(a) Original distribution.

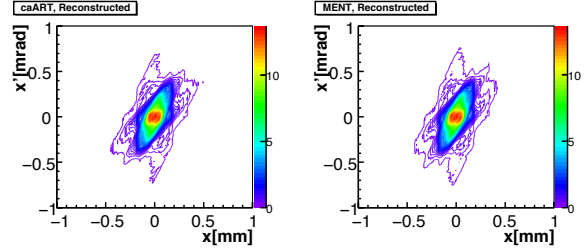


Figure 3: Realistic distribution and its reconstructed using caART and MENT.

Table 1: Relative errors between the original distribution and its reconstructed from Fig. 3.

Algorithm	σ_x [%]	$\sigma_{x'}$ [%]	$\sigma_{xx'}$ [%]
caART	1.59	3.57	5.94
MENT	0.75	0.64	2.33

tance, the quality of the reconstruction should be judged from point of view of what charge density it represents. A convenient interpretation is offered by the mean square norm

$$\|\Delta\| = \sqrt{\frac{\sum_x \sum_{x'} (\rho_{orig} - \rho_{recon})^2}{\sum_x \sum_{x'} \rho_{orig}^2}}, \quad (5)$$

where ρ_{recon} and ρ_{orig} denote the resulting reconstruction and the original distribution correspondingly. For such an estimation to be valid, effects from low-density bins should be discarded and the bin size of the two objects has to be equal. ASTRA calculates the moments of the distribution in a statistical manner, whereas here two discrete binned distributions are compared. If the binning does not represent the underlying data according to its specific features, lateral bins with low content would introduce gaussian tails and consequently differently calculated beam sigma matrix elements. This has been taken into account in advance as the bin width δ has been optimized using a minimization of a cost function $F(\delta)$ according to

$$F(\delta) = \frac{2\mu - var}{\delta^2} \quad (6)$$

for μ and var being the mean and the variance of the underlying spatial distribution [7]. Using the norm and the

reconstructions above, MENT and caART yield equal values until the second digit after the decimal point. In the first case the norm is $\|\Delta\| = 0.247$ and at the second $\|\Delta\| = 0.258$.

An alternative for the comparison is based on the equivalent ellipse emittance representation of the phase space, i.e. compare the resulting and original distributions from which only parts within ξ times the projected emittance are taken into account. Low intensity bins outside the contour defined by $\xi \cdot \varepsilon$ are discarded. The projected emittance is calculated separately for each of the two distributions - the original ASTRA one, describing the real phase space in some of the transverse planes, and the reconstruction. By this, the total area in the phase space is defined and a two-dimensional 'peel-off' cut on the tails is applied afterwards. The resulting fractional areas inside the contour are considered in order to calculate the norm. The outcome in such a case slightly favors the MENT - $\|\Delta\| = 0.249$, while for the ART this value is 0.260. This can also be seen in Fig. 4 where different fractions of the distributions are taken into account. The horizontal lines represent the case when the two distributions are compared within their total area. For fractional area of above six the norm converges to the value for which no cut has been done as MENT solutions are closer. i.e. the core part is always reconstructed smoother with better accuracy than using ART.

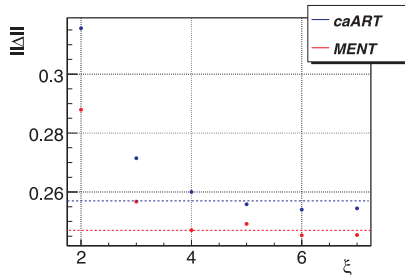


Figure 4: Mean square norm for different cuts of the phase-space distributions. A cut is done simultaneously on the original and the reconstructed phase spaces. The transverse emittances defining the area of the cut are calculated separately from the data describing the original and the reconstructed distributions. The horizontal lines indicate that the full phase spaces have been taken into account.

CONCLUSIONS

The work presents some investigations done in order to find a tomographic reconstruction algorithm suitable to be used with limited projection data. Several algorithms have been tested, namely Filtered Backprojection, Constrained Additive Algebraic Reconstruction Technique and Maximum Entropy. The last two inherit the ideas behind the FBP and as such they outperform it - a major reason for that is the fact that both are trying constructively to discard pixel content which are not consistent for any of the projection data. A number of concepts on how the reconstruction

Beam Dynamics, Other

results have been evaluated are presented as well. A conclusion that MENT represents the underlying phase space with better accuracy can be drawn.

REFERENCES

- [1] L. Staykov, "Characterization of the transverse phase space at the Photo-Injector Test Facility in DESY, Zeuthen site", PhD thesis, Universität Hamburg, 2008.
- [2] G. Asova et al., "Design considerations for phase space tomography diagnostics at the PITZ facility", proceedings of DIPAC 2007, Mestre, Italy.
- [3] G. Asova et al., "Phase space tomography diagnostics at the PITZ facility", proceedings of ICAP 2006, Chamonix, France.
- [4] R. M. Rangayyan, "Biomedical Image Analysis", CRC Press, 2005.
- [5] R. Gordon, "A tutorial on ART (Algebraic Reconstruction Technique)", *IEEE Trans. Nucl. Sci.*, NS-21(78), 1974.
- [6] G. Minerbo, "A Maximum Entropy Algorithm for Reconstructing a Source from Projected Data", *Comp. Graphics Image Proc.* 10 (1970) 48.
- [7] H. Shimazaki, "Recipes for Selecting the Bin Size of a Histogram", PhD thesis, Kyoto University, 2007.
- [8] K. Flöttmann. <http://www.desy.de/~mpyflo/>.

PARTICLE-IN-CELL SIMULATION OF ELECTRON-HELIUM PLASMA IN CYCLOTRON GAS STOPPER*

Y.K. Batygin[#], G. Bollen, C. Campbell, F. Marti, D.J. Morrissey, G. Pang, S. Schwarz, NSCL,
Michigan State University, East Lansing, MI 48824, USA

Abstract

The cyclotron gas stopper is a newly proposed device to stop energetic rare isotope ions from projectile fragmentation reactions in a helium-filled chamber [1, 2]. The radioactive ions are slowed down by collisions with a buffer gas inside a cyclotron-type magnet and are extracted via interactions with a Radio Frequency (RF) field applied to a sequence of concentric electrodes (RF carpet). The present study focuses on a detailed understanding of space charge effects in the ion extraction region. The space charge is generated by the ionized helium gas created by the stopping of the ions and eventually limits the beam rate. Particle-in-cell simulations of a two-component (electron-helium) plasma interacting via Coulomb forces were performed in the space charge field created by the stopping beam.

INTRODUCTION

The cyclotron gas stopper is a device for the deceleration of radioactive ions created by the projectile fragmentation (see Fig. 1). Fast ions (~ 100 MeV/u) are injected into a helium-filled chamber inside a vertical magnetic field where they immediately enter a solid degrader so that they can be captured by the magnetic field. The fast ions lose the remainder of their kinetic energy in collisions with the helium buffer gas. This process ionizes the helium atoms. An electric field parallel to the magnetic field is used to remove electrons and move positively charged ions to the RF-carpet. At high incoming particle rates, the amount of ionization becomes so large that the stopped ions cannot be completely removed. As a result, a neutralized plasma accumulates in the center of the stopping chamber and additional fast ions are not or are only slowly extracted because they come to rest in the plasma-shielded region. This present work analyzed the overall process of charge migration to provide estimates of ion stopping efficiency as a function of incoming particle rate.

NUMERICAL METHOD

Present simulations are based on a preceding detailed numerical study of rare isotope production, transport, and stopping in a gas-filled magnetic field [1]. The program LISE++ [3] was used to calculate the transmission, yields, and ion-optical properties of the projectile fragment beam. A C++ version of the ATIMA code [4]

was used to calculate the energy lost by the incoming beam in the solid degrader. The Stopping and Range Tables from the SRIM package [5] were used to calculate the energy loss in the helium gas. The CycStop code [6] combined this input to calculate the fast ion stopping

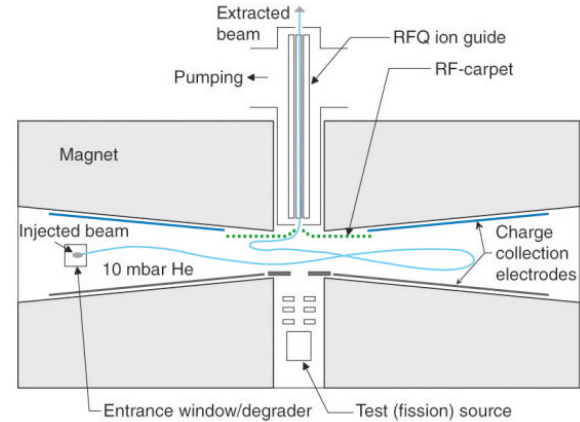


Figure 1: Schematic layout of cyclotron gas stopper [1]. The fast projectile fragments are incident horizontally at the left and after stopping are moved to the center for axial extraction.

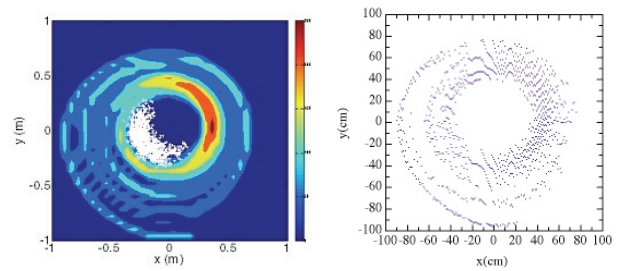


Figure 2: (Left) Top view of the energy loss density in color and the positions of the stopped ions in white, and (right) the distribution of e^-/He^+ ion-pairs created by the stopping of a ^{79}Br beam from the CycStop code [6].

distribution, the losses, and the spatial deposition of energy in the helium. The energy distributions were the input for the space charge phenomena in the present work (see Fig. 2).

The calculation of space charge effects in e^-/He^+ plasma were performed with a modified version of the BEAMPATH code [7]. The simulations were performed by simultaneous tracking of He^+ and electrons in the field created by their own space charge forces, \vec{E}_{sc} , and applied external electric field, \vec{E}_o , with a velocity \vec{v} given by

$$\vec{v} = k(\vec{E}_o + \vec{E}_{sc}), \quad (1)$$

*Work supported by the US Department of Energy under Contract No DE-FG02-06ER41413

[#]batygin@nscl.msu.edu

where k is the mobility of ion or electron. The mobility of He^+ ions at room temperature and pressure of $P = 200$ mbar was taken as $k_{He^+} = 10^5 \text{ cm}^2/(\text{kVs})$ and is inversely proportional to gas pressure. The mobility of electrons is typically three orders of magnitude higher than that of He^+ under the same conditions. However, direct simulation of e^-/He^+ plasma with such large differences in ion mobilities is problematic due to the dramatic difference in the time scales of the dynamics of electrons and He^+ . Therefore, in the present work the mobility of the electrons was artificially reduced by a factor of ten to $10^7 \text{ cm}^2/(\text{kVs})$ to simultaneously follow the motion of the ions and electrons in the same calculation.

Radioactive ions were injected into the gas at a rate of dN/dt_{in} . Each incoming radioactive ion created a cloud of He^+ and electron charge at the rate dQ/dN , each. The product of these two rates gives the rate of creation of He^+ and electrons in the gas-filled chamber $dQ/dt = (dN/dt_{in})(dQ/dN)$. The simulation time step, $\Delta t = \Delta T/N_{step}$, was defined as a ratio of the simulation period, ΔT , and the number of integration steps, N_{step} . At each integration step the entire spatial distribution of e^-/He^+ particles was injected into the system such that the incoming charge of each component was $\Delta q = (dQ/dt) \Delta t$. This distribution is added to that remaining from the previously injected particles (see Fig. 3). The electrons and positive ions move in opposite directions in the net field created by applied collection field and by their space charge.

Space charge field is calculated via solution of Poisson's equation in two-dimensional cylindrical coordinates:

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial U}{\partial r} \right) + \frac{\partial^2 U}{\partial z^2} = Q(r, z), \quad (2)$$

where $Q(r, z) = -\rho(r, z)/\epsilon_0$ with Dirichlet boundary conditions at the surface of a tube of radius a , and Neumann condition at the axis:

$$U(a, z) = 0, \quad U(r, 0) = U(r, L) = 0, \quad \frac{\partial U}{\partial r}(0, z) = 0. \quad (3)$$

In Eq. (2), Q is the total instantaneous value of charge of electrons and ions presented in the system. While at every time step, the new fraction of e^-/He^+ is injected into the system, some fraction of particles is removed through the extraction region and through the walls of the gas cell. The Poisson's equation (2) is substituted by finite-difference analog:

$$U_{k,j+1} \left(1 + \frac{1}{2(j-1)} \right) - 2U_{k,j} \left(1 + \frac{h_r^2}{h_z^2} \right) + U_{k,j-1} \left(1 - \frac{1}{2(j-1)} \right) + U_{k+1,j} \left(\frac{h_r}{h_z} \right)^2 + U_{k-1,j} \left(\frac{h_r}{h_z} \right)^2 = -Q_{k,j} h_r^2. \quad (4)$$

Poisson's equation is solved via Fourier transformation in z -direction and solution of three-diagonal matrix equation in radial direction. Accuracy of calculations is controlled by calculation of error of the Gauss theorem

$$\vartheta = 1 - \left| \int_S \vec{E} d\vec{S} / \int_v Q dv \right|, \quad (5)$$

which is usually of the order of 10^{-3} (see Fig.4).

SIMULATION RESULTS

Fig. 3 illustrates the dynamics of e^-/He^+ cloud formation for the injection of ^{79}Br ions into the system with an axial electric field of $E_0 = 10 \text{ V/cm}$ for an incoming particle rate of $dN/dt_{in} = 10^5 \text{ part/sec}$ and energy of $E = 1.69 \cdot 10^9 \text{ eV}$. The number of e^-/He^+ pairs per stopped ion is $N = E/E_i = 4.23 \cdot 10^7$ where $E_i = 40 \text{ eV}$ is the effective ionization potential in helium gas. Each incoming beam particle creates e^-/He^+ pairs equal to $dQ/dN = 6.76 \cdot 10^{-12} \text{ C/part}$ so that the increase in charge due to e^-/He^+ pairs in the system is $dQ/dt = 6.76 \cdot 10^{-7} \text{ C/s}$. As illustrated in Fig. 3, shortly after starting the injection of beam, the electrons quickly travel to the anode wall of gas cell at $z = 10 \text{ cm}$ (the right edge in this view). After $t = 5 \cdot 10^{-3} \text{ sec}$, the system has relaxed to steady-state condition for this beam rate. Under these conditions the outgoing particle rate is equal to incoming rate $dN/dt_{out} = dN/dt_{in}$. The charged particles are considered to be extracted from the gas if they reach the walls (anode or cathode used to apply the drift field).

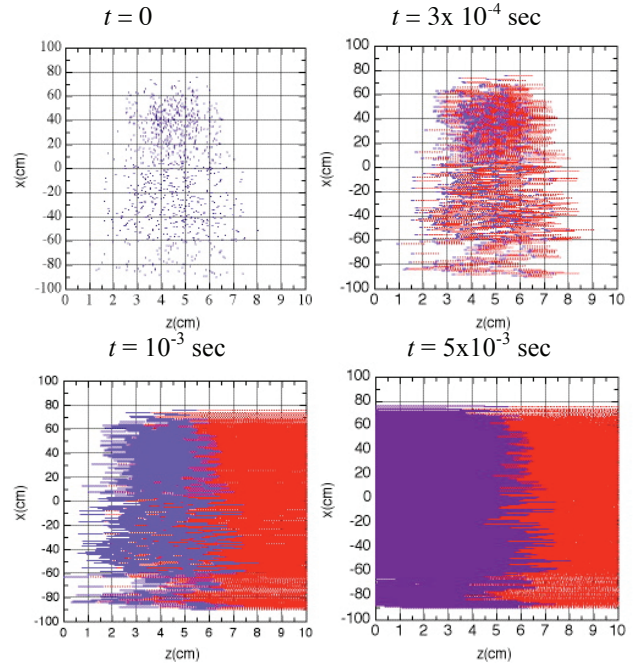


Figure 3: Time sequence of the motion of He^+ ions (blue) and e^- (red) created by a ^{79}Br beam at $dN/dt_{in} = 10^5 \text{ part/sec}$. The applied (horizontal) electric field was 10 V/cm .

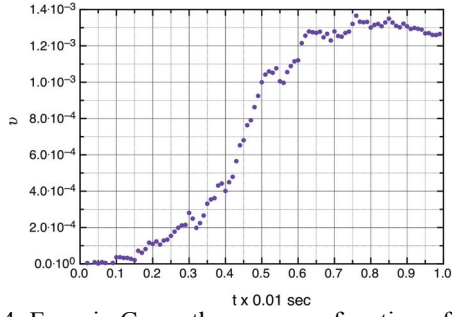
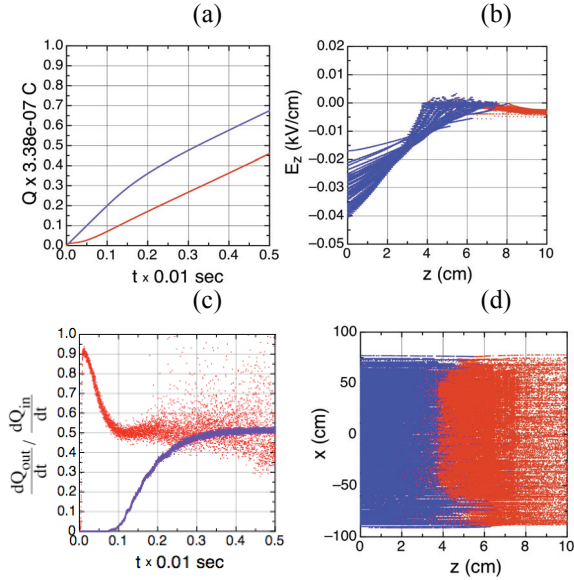


Figure 4: Error in Gauss theorem as a function of time.

Table 1: Simulation Parameters

Energy per stopped ion	$1.69 \cdot 10^9$ eV
He^+ mobility coefficient	$k_{\text{He}^+} = 10^5 \text{ cm}^2/\text{kVs}$
e^- "mobility coefficient"	$k_e = 10^7 \text{ cm}^2/\text{kVs}$
Number of macroparticles per filling	$2 \cdot 10^3$
Number of integration steps	$5 \cdot 10^3$
Total number of macroparticles	10^7
Box for Poisson's equation $R \times Z$	150 cm x 10 cm
Mesh $N_R \times N_Z$	1024 x 1024

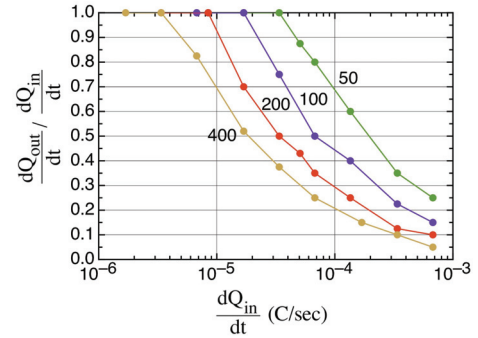
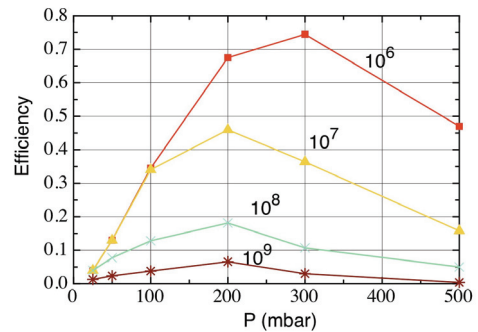
Figure 5: Conditions for space-charge neutralized plasma with e^- (red) and He^+ (blue) at incoming beam rate of $dN/dt_{in} = 10^7 \text{ part/s}$: (a) accumulated charge, (b) total electric field, (c) flux of electrons through $z=0$ and that of He^+ through $z=10 \text{ cm}$, (d) particle distribution at $t = 5 \cdot 10^{-3} \text{ s}$.

At significantly higher incoming beam rates, the fraction of outgoing ions becomes lower than the creation rate of ionized gas. This results from the formation of charged neutralized area in the regions of the chamber with the highest ionization densities (see Fig. 5). This limits the extraction of radioactive ions as well because some of the beam stops in the neutralized area and cannot leave the region. Fig. 5d illustrates the calculated motion of e^-/He^+ ions at an incoming particle rate of $dN/dt_{in} = 10^7 \text{ part/s}$. This problem becomes more pronounced with increasing incoming particle rate. Fig. 6 illustrate the dependence of extraction efficiency of e^-/He^+ ions as a

function of the incoming charge rate for different values of buffer gas pressure and applied electric field. The space charge forces were found to significantly limit the extraction efficiency when the incoming rate exceeded $dN/dt_{in} = 5 \cdot 10^6 \text{ part/s}$.

The present calculations were used to estimate the total collection and extraction efficiency of radioactive beam from the cyclotron gas stopper. As a test, the actual spiral distribution of ionized buffer gas was replaced by a uniform disk of 75 cm radius, with an axial ionization distribution centered at $z = 5 \text{ cm}$ and $\sigma_z = 1.5 \text{ cm}$. Both the actual and simple distributions give a similar dependence of the extraction efficiency on the incoming particle rate. The results for the extraction of e^-/He^+ ions were combined with the predicted stopping distribution of high energy ^{16}O ions in the cyclotron stopper to obtain a total efficiency for stopping and collection (see Fig. 7). The figure show the advantage of using a cyclotron gas stopper over a traditional linear gas stopper for high-intensity, light ion beams.

The present results should be considered with some caution due to the reduced electron mobility and the omission of any diffusion processes. Continued development of the cyclotron gas stopper design is underway including detailed studies of the effects of space charge on the ion motion at the RF-carpet and investigations into techniques to perform calculations with the true electron mobility.

Figure 6: The fraction of extracted charge, i.e., efficiency of e^-/He^+ extraction, as a function of incident beam rate for different values of buffer gas pressure (in mbar).Figure 7: The total efficiency for ^{16}O heavy ions as a function of gas pressure and incident beam rate (in part/s).

REFERENCES

- [1] G. Bollen, D.J. Morrissey, S. Schwarz, Nucl. Instrum. Meth. A550, (2005) 27.
- [2] G.K.Pang, et. al, Proc. of PAC07, Albuquerque, New Mexico, USA (2007) 3588-3590.
- [3] O.Tarasov and D.Bazin, Nucl. Phys. A746 (2004) 411.
- [4] H. Weick, <http://www-linux.gsi.de/~weick/atima/>
- [5] J.F. Ziegler, <http://www.srim.org/>
- [6] C. Campbell, et al., Nucl. Instr. Meth. A, to be published.
- [7] Y.Batygin, Nucl. Instrum. Meth. A539 (2005) 455.

INCORPORATING PARTIAL SIBERIAN SNAKES INTO THE AGS ONLINE MODEL*

V. Schoefer[†], L.Ahrens, K.Brown, A. Luccio, W. MacKay, T. Roser
BNL, Upton, NY 11973, USA

Abstract

In order to preserve polarization during polarized proton operation for RHIC, two partial Siberian Snakes are employed in the AGS, where a large number of strong spin depolarizing resonances must be crossed. These snakes cause a significant distortion to the injection lattice of the AGS and must be included in the online model. In this report, we discuss the problem of modeling snakes as optical elements, particularly as MAD-X elements, and present results comparing measurement to the AGS online model.

OVERVIEW

Polarized proton beam in the RHIC complex is created in the OPPIS source and accelerated through a 200 MeV Linac. The beam is then accelerated in the Booster and subsequently injected into the Alternating Gradient Synchrotron (AGS) at a $G\gamma = 4.5$, and accelerated to a $G\gamma$ of 45.5 Here G is the anomalous g-factor magnetic moment of the proton ($G = 1.7928$) and γ is the relativistic Lorentz factor.

In order to preserve polarization of the beam during acceleration through intrinsic and imperfection depolarizing resonances, the AGS lattice has been outfitted with two partial Siberian snakes. The snakes magnets are helical dipoles which, to provide sufficient spin rotation in the limited physical space available in the AGS lattice, have a “double pitch” structure [1]. That is, the far upstream and downstream regions of each snake are helices of one pitch and the central regions are of different, slower, pitches. One snake is superconducting and the other is normal conducting and they are called the ‘cold’ and ‘warm’ snakes respectively.

The central helical field of the cold snake can be run as high as 2.5 T, but it typically operated at 2.1 T. These field strengths correspond to rotations of the proton spin vector of 10 % (or 18 degree) and 15% (or 27 degree), respectively, around the longitudinal axis. The warm snake is operated with a central helical field of 1.53T, which corresponds to a spin rotation of 5% (9 degrees) about the longitudinal axis. Both snakes are run with constant current throughout the AGS acceleration cycle.

Each of the two snakes is strongly focusing in both planes and they represent a significant perturbation to the AGS optics. Both snakes require external magnetic ele-

ments to provide matching to the typical AGS lattice and each therefore has four quadrupoles near it used to compensate for perturbations to the linear optics [2].

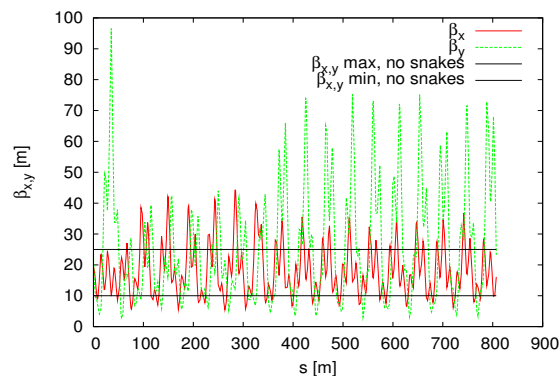


Figure 1: Modeled AGS β functions at injection energy with both snakes included and using operational currents in the compensation quadrupoles. Black lines show the maximum and minimum beta functions in a lattice without snakes. The compensation quadrupole currents that are ultimately determined to be optimal are often far from the modeled fit.

The beam orbit inside the helical dipoles is itself a helix. At injection energy, this helix has a radius of approximately 2 cm and ideally beam is delivered into the snake displaced horizontally by that amount, with no vertical displacement. As the beam rigidity increases, the radius of the helix decreases like γ^{-1} .

The cold snake also has a significant off-axis longitudinal magnetic field component. A 1 meter long superconducting solenoid has been included in the design of the snake to compensate for that effect. However, since the beam’s offset from the central axis is a function of energy and the solenoid can only be operated DC, the coupling contribution from the cold snake’s longitudinal field can only be completely cancelled at a single beam rigidity.

Accurate modeling of the snakes is critical to polarized proton operation because avoidance of intrinsic and imperfection depolarizing resonances simultaneous requires tight control of the vertical closed orbit and a vertical betatron tune near an integer value (9 in the case of the AGS). This is a region of configuration space that tends to be both physically and numerically sensitive.

* Work performed under Contract Number DE-AC02-98CH10886 with the auspices of the US Department of Energy

[†] schoefer@bnl.gov

HELICAL DIPOLE FIELDS

Following Ptitsyn [3], in right-handed polar coordinates (r, ϕ, z) with the z coordinate pointing along the direction of beam motion we may write the solution to Laplace's equation for the scalar potential in the current free region in the aperture of the helical dipole as

$$\psi = \sum_{m=1}^{\infty} I_m(mkr)(a_m \cos(m\theta) + b_m \sin(m\theta)) \quad (1)$$

Here one takes advantage of the helical symmetry to use $\theta = \phi - kz$ as the polar coordinate and k is the pitch of the helix which is here taken to be constant over the length of the magnet. I_m is the modified Bessel function.

The magnetic field derived from this potential is:

$$\begin{aligned} B_r &= -k * \sum_{m=1}^{\infty} I'_m(mkr)(a_m \cos(m\theta) + b_m \sin(m\theta)) \\ B_z &= k * \sum_{m=1}^{\infty} m I_m(mkr)(b_m \cos(m\theta) - a_m \sin(m\theta)) \\ B_\phi &= -\frac{1}{kr} B_z \end{aligned} \quad (2)$$

The simplified relationship between B_ϕ and B_z is a result of the helical symmetry of the field.

It is apparent from this form that there are intrinsic nonlinearities to the field since I_m is proportional to r^m as r approaches zero.

The non-zero off axis longitudinal field contribution is also visible.

CURRENT STATUS

The current AGS online model is implemented using a CDEV server that takes real-time snapshots of magnet current settings and uses MAD-X to calculate model parameters.

The snakes are currently implemented in the AGS online model in MAD-X using only the linear matrices produced by integration of the field maps [4]. The transfer maps so calculated are numerical Taylor expansions about an assumed off-axis ideal orbit $(x, y) = (20 \text{ mm}, 0 \text{ mm})$. Figure 2 shows the level of disagreement in the measured and modeled betatron tunes in a typical operational polarized proton lattice. Since the fields of both snakes are constant in time, their effect on the lattice decreases as the beam becomes more rigid and the snakes become more transparent to betatron motion. The disagreement between measurement and model is thus most significant early in the acceleration cycle, where it is near 0.05 units. Gaps in the model data, where there are measurements but no model prediction, are points in the cycle for which MAD-X could not find a closed orbit given the supplied magnet currents. The large horizontal disagreements near $\gamma = 8$ are the result of

the transition γ_{tr} jump, which is not yet included in the online model.

It was shown by Luccio et al in 2006 that the second-order expansion of the snake field about the ideal trajectory results in a significant dependence of the vertical focal length on the horizontal orbit position [5]. Correct implementation of the calculated second-order matrices in MAD-X is made difficult by the fact that one cannot offset arbitrary matrices, and so the model interprets the expansion as being around the origin rather than around a point 2 cm to the outside of the ring. Additionally, the measured closed orbit in the AGS is not included in the model, so strong sensitivity to closed orbit offsets caused by the snake are not properly accounted for. This is particularly relevant for calculating closed orbit bumps that traverse the snakes.

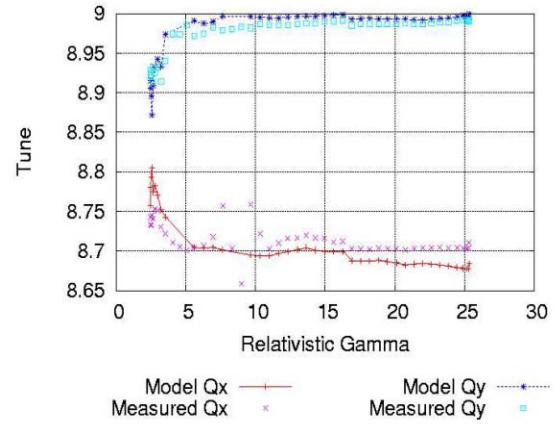


Figure 2: AGS online model tune predictions and measurements for operational currents, calculated using linear snake transfer matrices.

EQUIVALENT ELEMENT APPROACH

One approach used primarily to gain the ability to offset the snake from the central axis was to develop a sequence of normal and skew quadrupoles of a combined length equal to that of the snake and to fit their focusing strengths to reproduce the linear snake transfer matrix.

This approach is adequate for offline studies where the closed orbit at the snake is predictable and static and when the nonlinear effects of the snake are not important. For on-line modeling, this approach fails to take into account the possibility that the actual closed orbit through the snake may not be ideal and is not constant throughout the cycle. Even in the case where the orbit can be assumed to be near the ideal trajectory, the offset changes as the beam is accelerated and becomes more rigid and the ideal trajectory moves closer to the central axis. There is also no natural scaling for the parameters of these elements as a function of the beam rigidity, and so the individual element strengths must be re-fit for each beam energy one wishes to model.

SUMMARY AND PLANS

The above outlined challenges indicate that accurate modeling of the AGS snake elements requires a full non-linear description of the fields to be implemented in the model. The closed orbit can then be calculated using the entire field, at which point a Taylor expansion can be performed around online calculated orbit, rather than around a predetermined ideal orbit. The Polymorphic Tracking Code [6] seems ideally suited to such a task, given its separation of tracking coordinate system from the magnetic field and its ability to perform high order expansions.

This is a primary motivation for the upgrade of the AGS online model to include multiple modeling platforms [7].

REFERENCES

- [1] J. Takano et al, Field Measurements in the AGS Warm Snake, Proc. EPAC 2004, Lucerne, Switzerland.
- [2] E.D. Courant, Matching Quadrupoles for AGS Helical Snake, SPIN2002 International Spin Physics Symposium Conference Proceedings, Upton, NY, vol. 675, 799-8003
- [3] V. Ptitsyn, Notes on the Helical Field, RHIC/AP/41, Brookhaven National Laboratory, Upton, NY, October 1994.
- [4] A. U. Luccio et al, Cold Snake Optimization by Modeling. Tech. Report C-A/AP/128, Brookhaven National Laboratory. Upton, NY, December 2003.
- [5] A. U. Luccio, N. Malitsky, N. Tsoupas, Focusing of the AGS Cold Snake. Tech. Report C-A/AP/248, Brookhaven National Laboratory. Upton, NY, July 2006.
- [6] E. Forest, F. Schmidt, and E. McIntosh. Introduction to the Polymorphic Tracking Code. Technical Report CERN-SL-2002-044, KEK-Report 2002-3, 2002.
- [7] V. Schoefer et al. RHIC Injector Complex Online Model Status and Plans, Particle Accelerator Conference. Vancouver, Canada, 2009.

A FAST POINT TO POINT INTERACTION MODEL FOR CHARGED PARTICLE BUNCHES BY MEANS OF NONEQUISPACED FAST FOURIER TRANSFORM (NFFT)

T. Flisgen*, G. Pöplau, U. van Rienen, Rostock University, 18059 Rostock, Germany

Abstract

Demanding applications such as heavy ion fusion, high energy colliders and free electron lasers require the study of beam phenomena like space-charge induced instabilities, emittance growth and halo formation. Numerical simulations for instance with GPT (General Particle Tracer, Pulsar Physics) calculate the mutual Coulomb interactions of the tracked particles [5]. The direct summation of the forces is rather costly and scales with $\mathcal{O}(N^2)$. In this paper we investigate a new approach for the efficient calculation of particle-particle interactions: the fast summation by Nonequispaced Fast Fourier Transform (NFFT) [3, 4], whereas the NFFT is a generalization of the well known Fast Fourier Transformation (FFT). We describe the algorithm and discuss the performance and accuracy of this method for several particle distributions.

INTRODUCTION

The design of particle accelerators requires a sophisticated understanding of the dynamic behaviour of the particle bunch. Therefore several algorithms have been developed to determine the trajectories of the particles in the six-dimensional phase space.

Assuming the energy spread of the charged particles to be small, the space-charge forces may be computed in the bunch's rest frame by superposing the electrostatic field of each particle. The electric field at the position of the j -th particle $\mathbf{r}_j \in \mathbb{R}^3$ in the rest frame is given by

$$\mathbf{E}(\mathbf{r}_j) = \frac{1}{4\pi\epsilon_0} \sum_{\substack{\ell=1 \\ j \neq \ell}}^N q_\ell \frac{\mathbf{r}_j - \mathbf{r}_\ell}{\|\mathbf{r}_j - \mathbf{r}_\ell\|^3}, \quad j = 1, \dots, N, \quad (1)$$

where N denotes the number of particles, q_ℓ the charge of the ℓ -th particle, ϵ_0 the permittivity of vacuum and $\|\cdot\|$ the Euclidean norm. Since $(N-1)$ interactions have to be taken into account for each of the N particles, the direct evaluation of the sum in Eq. (1) reaches a disadvantageous numerical complexity of $\mathcal{O}(N^2)$. Note that the evaluation of the electric field strength has to be performed in each discrete time step of the tracking to determine the forces acting on the particles.

FAST SUMMATION USING THE NFFT

The presented method calculates the electric fields of the bunch approximately using the Nonequispaced Fast

Fourier Transform [3, 4]. The algorithm overcomes the quadratic runtime behaviour of the direct field evaluation and scales with $\mathcal{O}(N \log N)$.

Splitting of Potential Function

To describe the NFFT-based fast field calculation, the potential of a charged particle is separated into a short-range and a long-range effect:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q}{r} = \phi_{sr}(\mathbf{r}) + \phi_{lr}(\mathbf{r}). \quad (2)$$

Note that $\mathbf{r} \in \mathbb{R}^3$ denotes the point in the space, where the potential is evaluated and $r = \|\mathbf{r}\| \in \mathbb{R}_{\geq 0}$ the distance between the charged particle (here located at the origin) and the point of field estimation.

We demand the short-range effect $\phi_{sr}(\mathbf{r})$ to have compact support, such that $\phi_{sr}(\mathbf{r}) = 0 \quad \forall \quad r \geq \epsilon_I$ and the long-range effect $\phi_{lr}(\mathbf{r})$ to be bounded and $(p-1)$ times differentiable. The variable ϵ_I denotes the near field radius.

To cope with the singularity at $r = 0$ and to ensure the smoothness of the long-range effect, we regularize the potential at $r = \epsilon_I$ using an ansatz function (see the dashed, the crossed and the dotted curves in Fig. 1).

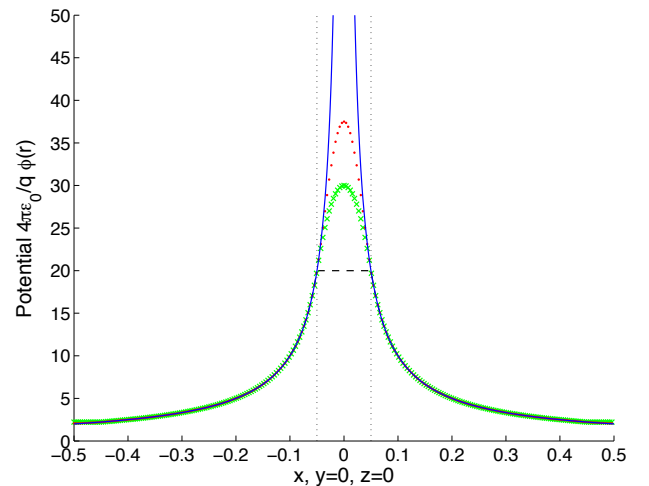


Figure 1: Potential $\phi(\mathbf{r})$ (solid) and long-range effect $\phi_{lr}(\mathbf{r})$ with $p = 1$ (dashed), $p = 2$ (crossed), $p = 3$ (dotted), where $\epsilon_I = 1/20$ and $\mathbf{r} = (x \ 0 \ 0)^T$.

Notice that the potential function is regularized at the boundary $r = l_B = 9/20$ as well to obtain a periodic smooth long-range contribution $\phi_{lr}(\mathbf{r})$. The deviation be-

* thomas.flisgen@uni-rostock.de

tween $\phi(\mathbf{r})$ and $\phi_{lr}(\mathbf{r})$ for $r < \varepsilon_I$ (see Fig. 1) is corrected by the near field potential $\phi_{sr}(\mathbf{r})$.

In summary the far field is given by the case differentiation

$$\phi_{lr}(\mathbf{r}) = \begin{cases} T_I(\mathbf{r}) & \text{if } r < \varepsilon_I, \\ \phi(\mathbf{r}) & \text{if } \varepsilon_I \leq r \leq l_B, \\ T_B(\mathbf{r}) & \text{else} \end{cases} \quad (3)$$

and the near field is defined by

$$\phi_{sr}(\mathbf{r}) = \begin{cases} \phi(\mathbf{r}) - T_I(\mathbf{r}) & \text{if } r < \varepsilon_I, \\ 0 & \text{else,} \end{cases} \quad (4)$$

where $T_I(\mathbf{r})$ and $T_B(\mathbf{r})$ denote the ansatz functions for the inner and outer regularization. It has to be mentioned that the sum of both effects (see Eq. (2)) yields the potential $\phi(\mathbf{r})$ only for $r \leq l_B$. Thus we have to guarantee by a scaling strategy, which is specified in [1], that the potentials will not be evaluated for particle distances $r > l_B$.

Fourier Construction

Since the long-range potential is $(p - 1)$ times continuous differentiable, it can be expressed approximately by the multivariate Fourier sum

$$\phi_{lr}(\mathbf{r}) \approx q \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} e^{2\pi i \mathbf{k} \mathbf{r}}, \quad (5)$$

where $\hat{b}_{\mathbf{k}} \in \mathbb{C}^{n \times n \times n}$ are the Fourier coefficients of the long-range effect. These coefficients are obtained by equispaced sampling of $\phi_{lr}(\mathbf{r})$ and Fast Fourier Transform of the resulting sample values [1, 3, 4]. It is worth to mention that the multi-index $\mathbf{k} = (k_x, k_y, k_z)$ runs over the finite set

$$I_n := \{-n/2, \dots, n/2 - 1\}^3, \quad (6)$$

where n^3 is the total number of Fourier coefficients.

Superposition of Field Contributions

Due to the fact that we have to compute potentials of charges located at different positions \mathbf{r}_ℓ in the bunch, we generalize Eq. (2) by replacing \mathbf{r} with $\mathbf{r} - \mathbf{r}_\ell$. This yields

$$\phi(\mathbf{r} - \mathbf{r}_\ell) = \frac{1}{4\pi\varepsilon_0} \frac{q_\ell}{\|\mathbf{r} - \mathbf{r}_\ell\|} \quad (7)$$

$$= \phi_{sr}(\mathbf{r} - \mathbf{r}_\ell) + \phi_{lr}(\mathbf{r} - \mathbf{r}_\ell). \quad (8)$$

The equation determines the potential of the ℓ -th particle at the position \mathbf{r} . To compute the potential at the position of the j -th particle in the bunch, we need to sum up the field contributions of the remaining $(N - 1)$ particles:

$$\phi_{bu}(\mathbf{r}_j) = \frac{1}{4\pi\varepsilon_0} \sum_{\substack{\ell=1 \\ j \neq \ell}}^N \frac{q_j}{\|\mathbf{r}_j - \mathbf{r}_\ell\|} \quad (9)$$

$$= \sum_{\substack{\ell=1 \\ j \neq \ell}}^N \phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) + \sum_{\substack{\ell=1 \\ j \neq \ell}}^N \phi_{lr}(\mathbf{r}_j - \mathbf{r}_\ell). \quad (10)$$

As a result of the compact support of $\phi_{sr}(\mathbf{r})$ it has a contribution to the left sum of Eq. (10) only for small distances $\|\mathbf{r}_j - \mathbf{r}_\ell\| < \varepsilon_I$. Thus we do not need to sum over all ℓ (except $\ell = j$), which would lead to an $\mathcal{O}(N^2)$ runtime behaviour of the algorithm. Instead we need to sum up over the index set

$$I_{\varepsilon_I}^{\text{NE}}(j) = \{\ell \in \{1, \dots, N\} : 0 < \|\mathbf{r}_j - \mathbf{r}_\ell\| < \varepsilon_I\}, \quad (11)$$

where only the cases $\phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) \neq 0$ are considered. It is worth to mention that a sorting algorithm has to be implemented, which determines the near field particles for all of the N particles.

According to the previous considerations the long-range effect in the right sum of Eq. (10) can be replaced by the Fourier sum Eq. (5):

$$\begin{aligned} \phi_{bu}(\mathbf{r}_j) &= \sum_{\ell \in I_{\varepsilon_I}^{\text{NE}}(j)} \phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) \\ &+ \sum_{\substack{\ell=1 \\ j \neq \ell}}^N q_\ell \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} e^{2\pi i \mathbf{k}(\mathbf{r}_j - \mathbf{r}_\ell)}. \end{aligned} \quad (12)$$

The difference in the exponent of the e-function will be separated and written as a product of two exponential functions. This yields

$$\begin{aligned} \phi_{bu}(\mathbf{r}_j) &= \sum_{\ell \in I_{\varepsilon_I}^{\text{NE}}(j)} \phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) - q_j \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} \\ &+ \sum_{\ell=1}^N q_\ell \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} e^{2\pi i \mathbf{k} \mathbf{r}_j} e^{-2\pi i \mathbf{k} \mathbf{r}_\ell}. \end{aligned} \quad (13)$$

It is highlighted that the summation $\sum_{\ell=1}^N \dots$ in (13) does not exclude the case $j = \ell$ anymore. For the sake of equality we have to preclude this case manually by subtracting the contribution $q_j \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}}$.

Now we rewrite our formula such that it can be expressed by two multivariate Nonequispaced Fast Fourier Transforms:

$$\begin{aligned} \phi_{bu}(\mathbf{r}_j) &= \sum_{\ell \in I_{\varepsilon_I}^{\text{NE}}(j)} \phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) - q_j \sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} \\ &+ \underbrace{\sum_{\mathbf{k} \in I_n} \hat{b}_{\mathbf{k}} \left(\underbrace{\sum_{\ell=1}^N q_\ell e^{-2\pi i \mathbf{k} \mathbf{r}_\ell}}_{\text{NFFT}^T} \right) e^{2\pi i \mathbf{k} \mathbf{r}_j}}_{\text{NFFT}}. \end{aligned} \quad (14)$$

The summation in the inner brackets denotes a transposed version of the NFFT. The outer summation denotes a NFFT. These transforms can be calculated very efficiently by using the software library of Kunis and Potts [2].

Determination of Electric Field Strength

Since the electric field strength of the bunch at the position \mathbf{r}_j is needed for the tracking procedure, we have to

evaluate the gradient of our potential:

$$\mathbf{E}(\mathbf{r}_j) = -\nabla \phi_{bu}(\mathbf{r}_j). \quad (15)$$

Exemplarily only the x -component $E_x(\mathbf{r}_j)$ of the electric field is discussed. Therefore we have to estimate the partial derivation $\frac{\partial}{\partial x_j}$ of the derived potential given in Eq. (14):

$$E_x(\mathbf{r}_j) = - \sum_{\ell \in I_{\varepsilon_1}^{\text{NE}}(j)} \frac{\partial}{\partial x_j} \phi_{sr}(\mathbf{r}_j - \mathbf{r}_\ell) - \underbrace{\sum_{\mathbf{k} \in I_n} 2\pi i k_x \hat{b}_{\mathbf{k}} \left(\underbrace{\sum_{\ell=1}^N q_\ell e^{-2\pi i \mathbf{k} \mathbf{r}_\ell}}_{\text{NFFT}^T} \right)}_{\text{NFFT}} e^{2\pi i \mathbf{k} \mathbf{r}_j}. \quad (16)$$

Again the summation in the inner brackets denotes a transposed NFFT. This is followed by n^3 multiplications with $2\pi i k_x \hat{b}_{\mathbf{k}}$. Finally the outer summation is computed by a NFFT. The transforms are performed by the library [2].

To ensure an advantageous runtime behaviour of the algorithm, we define the near field radius depending on the number of particles:

$$\varepsilon_I = \frac{p}{2\sqrt[3]{N}}. \quad (17)$$

The order of smoothness is set to $p = 3$ and the total number of Fourier coefficients to $n \approx \sqrt[3]{N}$.

BENCHMARKING THE ALGORITHM

The presented algorithm has been implemented in C and tested on an Intel(R)-Xeon(TM)-3 GHz machine with 4 GB RAM using Windows Server 2003.

Error Definitions

Beside the runtime behaviour of the algorithm which we compare with the runtime of the direct summation Eq. (1), we discuss the accuracy of the method. Therefore we define the relative error in the electric field strength by

$$f_j = \frac{\|\mathbf{E}_{nfft}(\mathbf{r}_j)\| - \|\mathbf{E}(\mathbf{r}_j)\|}{\|\mathbf{E}(\mathbf{r}_j)\|}, \quad j = 1, \dots, N, \quad (18)$$

where $\|\mathbf{E}_{nfft}(\mathbf{r}_j)\|$ is the absolute value of the strength computed by Eq. (16). The field strength $\|\mathbf{E}(\mathbf{r}_j)\|$ is determined by the direct summation Eq. (1). Additionally we consider the maximum of the relative error given by

$$f_{max} = \max_{j=1}^N |f_j|. \quad (19)$$

Spherical Distribution

Initially a spherical particle distribution is considered. The radius of the sphere amounts to $R = 2.2$ mm. Fig. 2 shows the runtime for the direct summation (see Eq. (1)) and the NFFT-based summation (see Eq. (16)).

Computer Codes (Design, Simulation, Field Calculation)

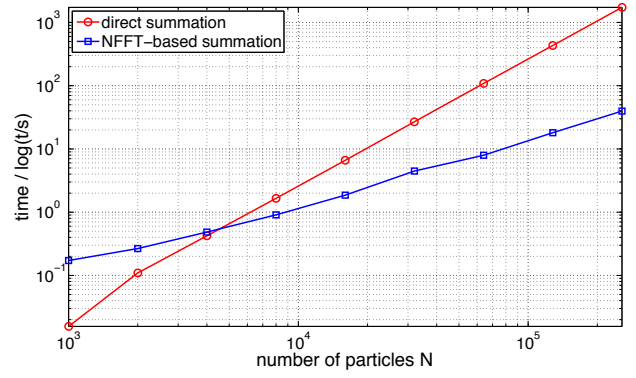


Figure 2: Runtime comparison for a spherical particle distribution using a logarithmic scaling of axes.

Note that logarithmic scales are used for both axes. The \circ -curve documents the $\mathcal{O}(N^2)$ runtime behaviour of the direct field evaluation, whereas the \square -curve shows the beneficial $\mathcal{O}(N \log N)$ performance of the NFFT-approach. Moreover Fig. 2 demonstrates that the approach becomes faster than the direct summation for $N > 4500$ particles. For less particles the proposed algorithm is slower due to its overhead e.g. the computation of the Fourier coefficients in Eq. (5).

Fig. 3 plots the error in the field strength in dependence on the particle location \mathbf{r}_j . The error is encoded by the colour and the size of the particle. It is obvious that the field values in the center of the distribution are afflicted with larger relative errors. The maximal relative error in the electric field strength is $f_{max} \approx 0.0188$.

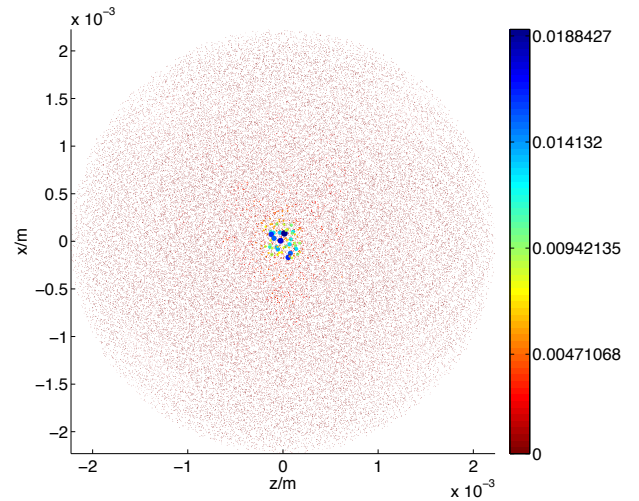


Figure 3: Locations of the relative errors $|f_j|$ for the spherical distribution with $N = 64000$ particles encoded by the colour and the size of the particles.

It is mentionable that larger relative errors in the center of the distribution result in small absolute errors, since the field strengths are small in the center of the charged bunch.

Cylindrical Distribution

Secondly a cylindrical distribution is taken into account. It has a length of $L = 3.5$ mm and a radius of $R = 2$ mm. For this case the runtime behaviour of the algorithm is very similar to Fig. 2 [1].

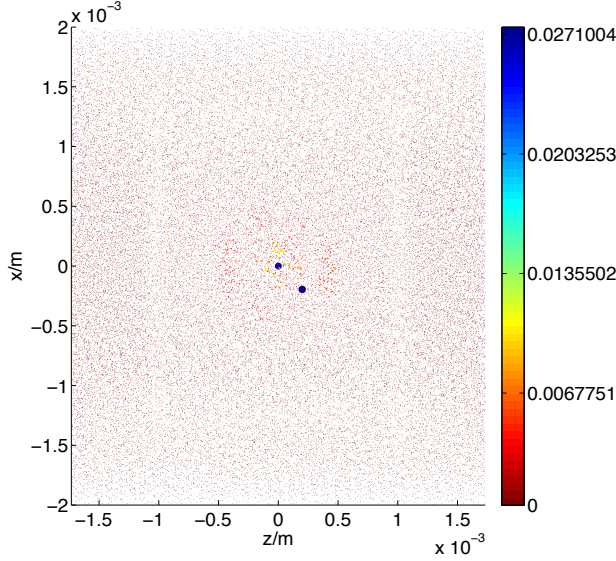


Figure 4: Locations of the relative errors $|f_j|$ for the cylindrical bunch with $N = 64000$ particles encoded by the colour and the size of the particles.

Fig. 4 reflects the error $|f_j|$ for the cylindrical particle distribution. The maximal error is $f_{max} \approx 0.027$. Note that the "hard edges" of the cylinder are difficult to treat numerically (especially with mesh-based strategies for field evaluation) and analytically. Despite this fact the largest relative error is again in the center of the distribution.

Sandwich Distribution

Finally a sandwich distribution is examined. It is constructed of ten flat ellipsoidal bunches arranged in series. The semi-axes of the bunches are $R_x = R_y = 1$ mm in the transversal direction and $R_z = 0.025$ mm in the longitudinal direction. The distance between the bunches is $\Delta d = 0.111$ mm. The runtime behaviour of the algorithm for the sandwich distribution is analog to Fig. 2, but in contrast to that plot the point of intersection is now at $N \approx 6000$ [1].

Fig. 5 shows the geometry and illustrates the error $|f_j|$ in the absolute electric field strength for each particle. The maximal relative error is $f_{max} \approx 0.032$. It is remarkable that larger errors are not located on the boundary of the flat ellipsoidal bunches, but again in the center of the whole distribution.

CONCLUSIONS

In this paper a method for the efficient calculation of electric fields inside a bunch of charged particles is presented. Computer Codes (Design, Simulation, Field Calculation)

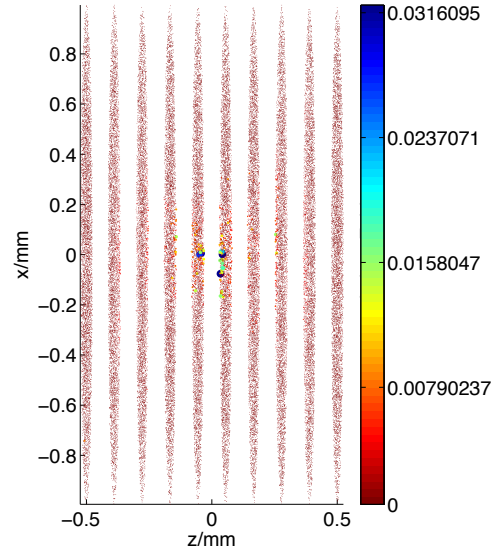


Figure 5: Locations of the relative errors $|f_j|$ for the sandwich bunch with $N = 64000$ particles encoded by the colour and the size of the particles.

constructed. The presented algorithm overcomes the quadratic runtime of the direct summation and scales for the proposed examples with $\mathcal{O}(N \log N)$. Although it is still slower compared to mesh-based methods, it has the advantage that it copes very well with "hard edges" and discontinuities in the charge density ρ .

It is spotlighted that the time evolution of the bunch's shape is mainly determined by the forces on the particles located at the boundary of the distribution. Therefore especially the field values at the edges have to be computed with small errors. Hence larger relative errors in the center of the distribution pose no crucial problem for the tracking.

REFERENCES

- [1] T. Flisgen. *Implementation of a Fast Point to Point Interaction Model for Charged Particle Bunches by means of NFFT*. Master thesis, Rostock University, 2009.
- [2] S. Kunis and D. Potts. NFFT, Softwarepackage, C subroutine library. <http://www-user.tu-chemnitz.de/~potts/nfft/>, 2002 – 2004.
- [3] G. Pöplau, D. Potts, and U. van Rienen. Calculation of 3D space-charge fields of bunches of charged particles by fast summation. In A. Anile, G. Ali, and G. Mascaly, editors, *Scientific Computing in Electrical Engineering*, number 9 in Mathematics in Industry, pages 241–246, Berlin, 2006. Springer-Verlag.
- [4] D. Potts and G. Steidl. Fast summation at nonequispaced knots by NFFTs. *SIAM J. Sci. Comput.*, 24:2013 – 2037, 2003.
- [5] S. van der Geer and M. de Loos. *The General Particle Tracer Code. Design, implementation and application*. PhD thesis, TU Eindhoven, 2001.

TRIUMF-VECC ELECTRON LINAC BEAM DYNAMICS OPTIMIZATION

Y. Chao, F. Ames, R. Baartman, I. Bylinskii, S. Dechoudhury, G. Goh, S.R. Koscielniak, R.E. Laxdal, M. Marchetto, L. Merminga, V. Naik, V.A. Verzilov, F. Yan, V. Zvyagintsev
TRIUMF, 4004 Wesbrook Mall, Vancouver, BC, Canada, V6T 2A3

Abstract

The TRIUMF-VECC Electron Linac is a device for gamma-ray induced fission of actinide targets, with applications in nuclear physics and material science. A phased construction and commissioning scheme will eventually lead to a 50 MeV, 10 mA CW linac based on superconducting RF technology. Using this linac to deliver high intensity electron beams for applications such as an energy-recovered light source is a possibility integrated in the design study. The multitude of design and tuning parameters, diverse objectives and constraints require a comprehensive and efficient optimization scheme. For this purpose we adopted the genetic optimization program developed at Cornell University as a prototype. Feature extensions were developed to accommodate specifics of the Electron Linac design, provide framework for more generic and integrated design process, and perform robustness/acceptance analyses. In this report we will discuss the method and its application to the design optimization of the Electron Linac. [4].

OVERVIEW

TRIUMF and VECC of Kolkata, India are signing an MOU to jointly develop Injector Cryo-Modules for an electron linac (E Linac) for radioactive ion beam (RIB) production via photo-fission of ^{238}U . This provides a source of neutron-rich isotopes complementary in character to those produced by proton beams.

The E Linac accelerates 10 mA CW e^- beam (16 pC/bunch) to 50 MeV with 1.3 GHz superconducting RF cavities housed in three cryo-modules. The beam is

Table 1: Beam parameters for the E Linac

RIB 16 pC per bunch	100 keV	10 MeV
RMS ϵ_N transverse (μm)	7.5	12.5
Bunch length (cm)	2.8 ($\pm 20^\circ$)	0.6
Energy spread	± 1 keV	± 40 keV
High brightness 100 pC per bunch	200 -300 keV	50 MeV
RMS ϵ_N transverse (μm)	1.0	10.0
Bunch length (mm)	4.0	1.0
Energy spread	± 0.5 keV	± 50 keV

generated at a 100 keV grid-modulated thermionic gun with a 650 MHz pulse structure. A normal conducting buncher and two 1.3 GHz SRF single cell cavities provide graduated bunching and longitudinal matching into the main accelerating structure. Transverse focusing is provided by solenoids or quadrupoles.

Coupled to a high brightness photo injector, the E Linac can potentially be used in applications beyond RIB production, such as an X-ray source through Compton scattering. It is therefore interesting and relevant to investigate if, and how, the same configuration can deliver both the 16-pC/bunch RIB beam and a 100-

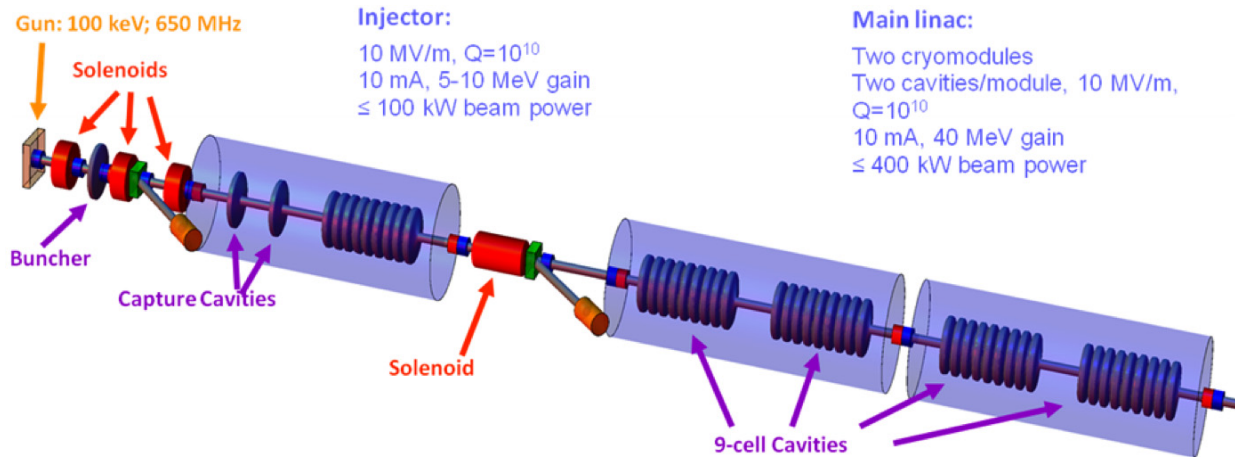


Figure 1: Schematic of the TRIUMF E Linac.

* TRIUMF receives funding via a contribution agreement through the National Research Council of Canada. This work is supported in part by funding from VECC, Kolkata, India.

pc/bunch high brightness (HB) beam, achieving final beam specs suitable for their respective applications. The commissioning plan may dictate different setup parameters at the front end for optimized performance at different stages of the project. Such questions can best be answered by a systematic optimization program accounting for a broad variety of objectives and constraints.

Beam dynamics modeling is done with Astra [1], Parmela [2] and Track [3]. To explore the multi-dimensional parameter space in a systematic and comprehensive manner, a genetic optimization program, originally developed at Cornell University [4], was adopted and modified to run on the Western Canada Research Grid [5]. A typical optimization involves the selected evolution of design parameters (tuning parameters and element locations), toward progressively improved design objectives (beam parameters and performance measures such as beam loss), subject to constraints mostly reflecting limitations in geometry and hardware.

OPTIMIZATION - GENETIC ALGORITHM WITH PARETO DOMINATION CRITERION

A genetic algorithm is superior in its robustness against near singularities in the modeling process. The particular algorithm adopted performs selection based on the Pareto domination criterion, with the additional advantage of allowing overview of multiple competing objectives, and avoidance of artificial cut-off in constrained parameters. This algorithm has proved quite competent in achieving design goals set for the current study. Figure 2 shows a well defined Pareto front plot representing trade-off between 3 competing objectives from a typical sufficiently evolved run.

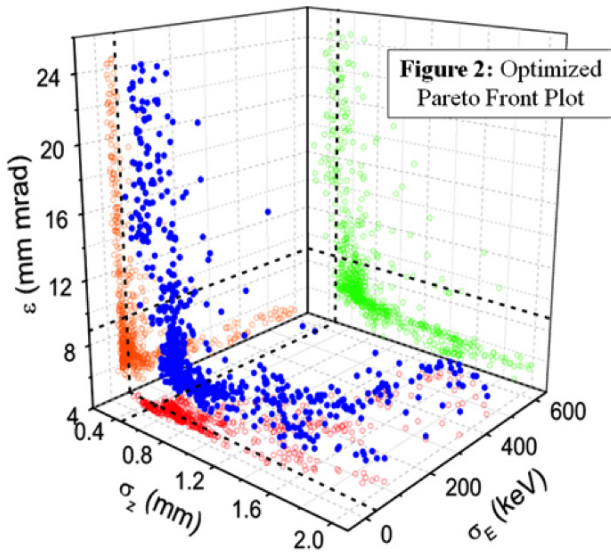


Figure 2: Optimized Pareto front plot.

OPTIMIZED SOLUTIONS

The genetic algorithm was used to find the following E Linac configurations. Figure 3 shows a 50 MeV RIB solution where the nonlinearity in the RF waveform (a T_{655} term) was used to cancel the nonlinear momentum compaction due to non-relativistic dynamics at low energy (a T_{566} term) while performing efficient bunching and acceleration at the same time. The normalized longitudinal emittance decreased by $\sim 50\%$ from the otherwise inevitable peak growth. Figure 4 shows a 10 MeV high bunch charge solution where the reduction in transverse emittance was accomplished through realignment of longitudinal slices in the transverse phases space by optimized RF focusing.

The genetic algorithm was used to find the following E Linac configurations. Figure 3 shows a 50 MeV RIB solution where the nonlinearity in the RF waveform (a T_{655} term) was used to cancel the nonlinear momentum compaction due to non-relativistic dynamics at low

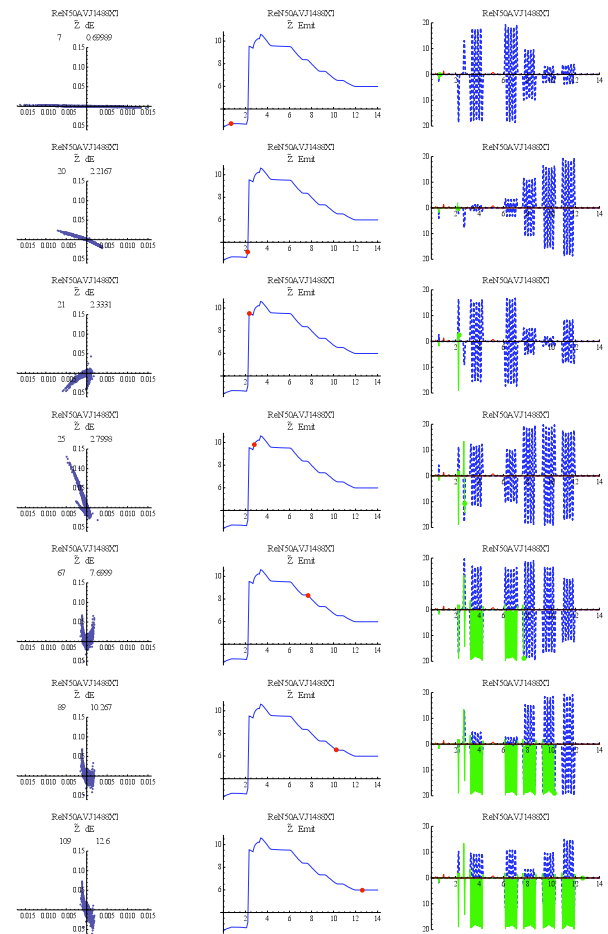


Figure 3: Longitudinal space 100 keV -50 MeV RIB; Left: Phase space distribution E (MeV) vs Z (m); Center: ϵ_z^N in keV-mm vs distance in m; Right: Progress of beam along Z and RF waveform.

energy (a T_{566} term) while performing efficient bunching and acceleration at the same time. The normalized longitudinal emittance decreased by $\sim 50\%$ from the otherwise inevitable peak growth. Figure 4 shows a 10 MeV high bunch charge solution where the reduction in transverse emittance was accomplished through realignment of longitudinal slices in the transverse phase space by optimized RF focusing.

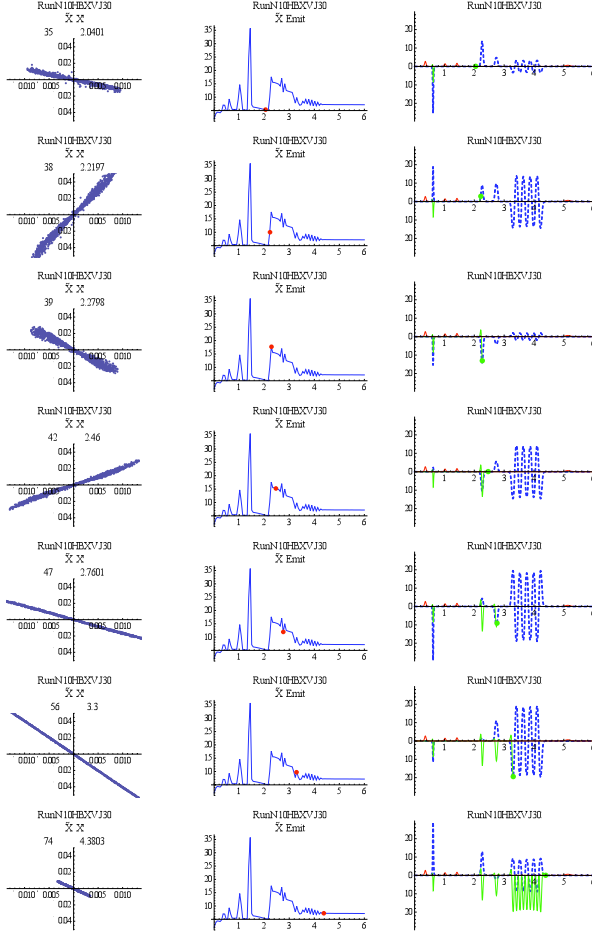


Figure 4: Transverse space 200 keV -10 MeV HB; Left: Phase space distribution X' (rad) vs X (m); Center: ϵ_N^X in mm-mrad vs distance in m; Right: Progress of beam along Z and RF waveform

APPLICATION OF THE METHOD

Besides providing globally optimized solutions for a given configuration, this method proved valuable in resolving other design issues and providing insights into the underlying physical mechanism of solutions. A few examples are shown in the following.

- Performance comparison between different designs: As the level and detail of optimization can be controlled better than many other methods, one can compare the relative merits of different designs optimized to the same level. Figure 5 shows such comparisons between different choices of the capture cavity configurations, and between different

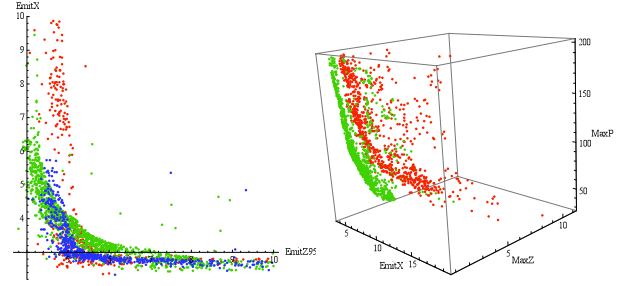


Figure 5: Left: Pareto fronts for $\beta=0.7+\beta=1.0$ capture cavities (red), $\beta=0.7+\beta=0.85$ (green) and $\beta=0.7+\beta=0.7$ (blue). Right: Pareto fronts for 1.55 m inter-bunch-capture distance (red) and 1.05 m inter-bunch-capture distance (green).

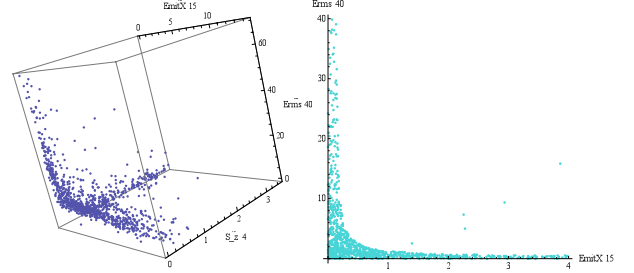


Figure 6: Left: Trade-off among solutions attempting to satisfy 3 parameters at the same time: $\epsilon_N^X = 15$ mm-mrad, $\sigma_Z = 4$ mm & $\sigma_E = 40$ keV. Right: Solutions attempting to satisfy 2 parameters: $\epsilon_N^X = 15$ mm-mrad & $\sigma_E = 40$ keV, as well as additional constraints.

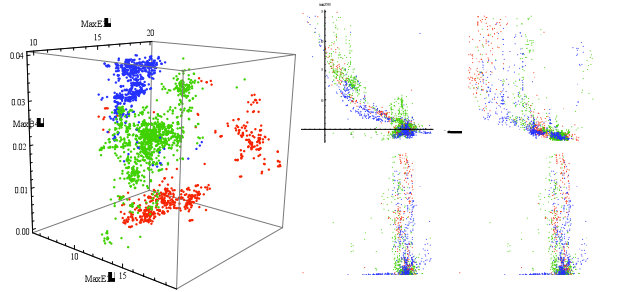


Figure 7: Left: Distribution of capture cavity fields (MaxE(2/3): MV/m) & last solenoid field (MaxB(4): T) optimized for 3 different capture configurations. Right: Various correlations between buncher field (MaxE(1): MV/m), first solenoid field (MaxB(1): T) and longitudinal & transverse emittances.

geometries of the design.

- Solving for externally imposed design goals: The method can be trivially extended to solve for design parameters satisfying externally imposed design goals, or provide insight on trade-off between parameters in meeting such goals through Pareto-front plots. Figure 6 shows a case where exact solutions can be obtained for given design goals in terms of beam parameters, and in the case of over-constrained goals, the Pareto front mapping out best-achievable options in the parameter space.

- Correlation between parameters: Due to its ability to efficiently carry along high dimension of parameters, and the large statistics readily available from the globally optimized gene pool, this method provides insight into correlation and interplay between variables, objectives and constraints. Figure 7 shows an example of the different optimization strategies, in terms of the capture cavity fields and solenoids, for the 3 capture configurations described in Figure 5, and another example of how final beam parameters are correlated to various tuning parameters.
- Exception handling of design. Figure 8 shows the compromise in performance of various designs when one of the capture cavities is not used. Again the comparison is more rigorous because all cases are optimized to the same level. It also shows, through inspection of the optimized gene pool, that in such exception cases the second capture cavity has to do more acceleration at the expense of bunching, and the opposite for the 9-cell cavity, in order to achieve improved beam parameters. It is also clear that the first capture cavity ($\beta=0.7$ in all cases) is more important than the second capture cavity. This is mostly due to its more favorable distance to the buncher in satisfying the longitudinal matching condition.

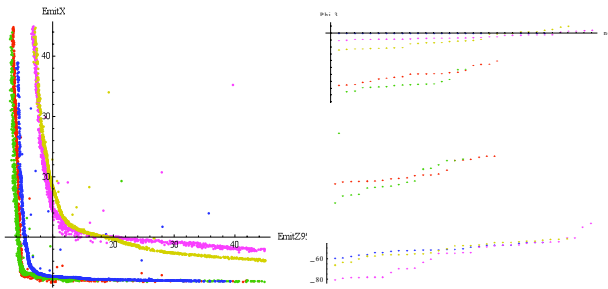


Figure 8: Left: Performance comparison between normal and exception cases: red: $\beta=0.7+\beta=1.0$ capture cavities, green: $\beta=0.7+\beta=0.7$, blue: 1st capture only ($\beta=0.7$), pink: 2nd capture only ($\beta=1.0$), olive: 2nd capture only ($\beta=0.7$). Right: Corresponding RF phases for the 2nd capture & the main (9-cell) cavities.

EXTENSION OF THE METHOD

Extension to the optimization program for the current design effort, either developed or under development, include the following:

- A generic framework based on python scripts allowing the definition of optimization objectives and constraints through user defined operation on arbitrary code-generated files.
- Method to incorporate different design prototypes into a single selection process subject to common selection criteria. This can be useful when other criteria such as cost become relevant.
- A flexible structure based on XML and python scripts allowing evaluation of a design by arbitrary processing modules. This extends the concept of model into an integrated design process where for example, Astra, is only one of many modules called on to return a complete set of performance metrics as input to the selection mechanism. This feature has been used for efficient benchmarking between Astra and Track.
- These modules can be invoked in parallel or in series in a single optimization run. In the former case complementary performance metrics can be obtained from different modules, while in the latter the user can perform high level analysis on raw simulated parameters, or end-to-end optimization over an integration of successive components.
- The optimization program is being integrated as a tuning component in the study of machine acceptance and robustness, where one defines the part of parameter space spanned by input beam and machine error that can be handled by tuning within operating range. In this context local-minimum algorithms such as the Levenberg-Marquardt method may be considered as an efficient alternative.

REFERENCES

- [1]. <http://tesla.desy.de/~lfroehli/astra/>
- [2]. <http://epaper.kek.jp/p03/PAPERS/FPAG029.PDF>
- [3]. <http://www.phy.anl.gov/atlas/TRACK/>
- [4]. I. Bazarov and C. Sinclair, PRST-AB **8**, 034202 (2005), and references therein.
- [5]. <http://www.westgrid.ca/>

DESIGN OF 10 GEV LASER WAKEFIELD ACCELERATOR STAGES WITH SHAPED LASER MODES *

E. Cormier-Michel, E. Esarey, C.G.R. Geddes, C.B. Schroeder, W.P. Leemans,
Lawrence Berkeley National Laboratory, Berkeley CA, 94720, USA
D.L. Bruhwiler, B. Cowan, K. Paul,
Tech-X Corporation, Boulder CO, 80303, USA

Abstract

We present particle-in-cell simulations, using the VORPAL framework, of 10 GeV laser plasma wakefield accelerator stages. Scaling of the physical parameters with the plasma density allows us to perform these simulations at reasonable cost and to design high performance stages. In particular we show that, by choosing to operate in the quasi-linear regime, we can use higher order laser modes to tailor the focusing forces. This makes it possible to increase the matched electron beam radius and hence the total charge in the bunch while preserving the low bunch emittance required for applications.

INTRODUCTION

Laser driven wakefield accelerators (LWFAs) are able to produce accelerating gradients thousands of times higher than conventional accelerators, making them suitable to build compact devices. In a LWFA the radiation pressure of the laser pulse induces charge separation, producing a plasma wave (wake) traveling at the group velocity of the laser pulse, close to the speed of light, and hence able to accelerate particles to relativistic velocities (see [1] for a complete review). Energies up to a GeV have been obtained in only a few centimeters [2]. Recently, experiments have shown that it is possible to control the injection and the acceleration of electrons [3, 4], providing a path towards high quality electron beams that can be used for applications, including free electron lasers [5], gamma ray sources [6] and colliders for high energy physics [7]. Light sources need stable electron bunches of the order of a GeV. A multi-TeV collider was designed using staging of several 10 GeV LWFA accelerator modules at a density of $n_0 \sim 10^{17} \text{ cm}^{-3}$, each about a meter long [7]. Efficient transfer of the laser energy to the accelerated beam, acceleration of positrons and conservation of a low emittance must be considered for applications. In this paper we present the design of these stages using Particle-In-Cell (PIC) simulations with the VORPAL framework [8]. We show that by using higher order laser modes, in the quasi-linear regime, the focusing forces in the wake can be controlled in order to improve the stage efficiency.

RESULTS

The PIC method is a fully self-consistent algorithm which allows non-linear evolution of the plasma wake and of the laser pulse simultaneously. In PIC simulations the smallest dimension, i.e., the laser wavelength ($\lambda \sim 1 \mu\text{m}$), needs to be resolved, whereas the box size increases with the plasma wave wavelength $\lambda_p = (\pi c^2 m / e^2 n_0)^{1/2}$, where n_0 is the plasma density. The acceleration length also increases with higher energy stages, i.e., lower plasma densities, making the simulations more computationally intensive. Simulations of a 1 GeV stage, to model the recent experiments or gamma ray sources, with a density of $n_0 \sim 10^{18} \text{ cm}^{-3}$, require of the order of 10^6 processor-hours and $\sim \text{TB}$ of storage. This allows only a few runs in three dimensions (3D), and parameter scans for stage optimization can be done in two dimensions (2D) only. Because the size of the box in 3D and the simulation length each scale as $n_0^{-3/2}$, a 10 GeV stage at $n_0 = 10^{17} \text{ cm}^{-3}$ would require 10^9 processor-hours which is not yet achievable with today's computational facilities. Approximations are then necessary to simulate such stages at the nominal density and reduced models, such as envelope and quasi-static models [9, 10] or calculation in a Lorentz boosted frame [11, 12, 13], can be used.

The approach used here to design high energy modules is to simulate shorter, higher density stages with scaling of the physical parameters with the plasma density [14]. In the scaled simulations the dimensionless parameters $k_p L$, $k_p r_0$ and a_0 , where $k_p = 2\pi/\lambda_p$ is the plasma wave number, L and r_0 are the laser length and spot size respectively, and $a_0 = 7.2 \times 10^{-19} \lambda^2 [\mu\text{m}] I [\text{W}/\text{cm}^2]$ is the normalized laser intensity, are kept constants. PIC simulations in the quasi-linear regime ($a_0 \simeq 1$) at different densities and comparison with a quasi-static code in 2D cylindrical geometry at $n_0 = 10^{17} \text{ cm}^{-3}$ show that the wake structure stays constant under these conditions [15]. Simulations also show that laser evolution, self-focusing and depletion, and electron beam dephasing scale as predicted by the linear theory, even though this theory is strictly valid in the low intensity limit ($a_0 \ll 1$), thus allowing scaled design of multi-GeV stages. Reduction of wake amplitude due to the presence of a charged beam (beam loading) also scales predictably for a wide range of parameters, allowing prediction of beam charge in unscaled stages [14]. Fig. 1(a) shows the accelerating wake structure at $n_0 = 10^{18} \text{ cm}^{-3}$

* The author acknowledge the assistance of the VORPAL development team. Work supported by the U.S. Department of Energy, HEP Contract No. DE-AC02-05CH11231, and the COMPASS SciDAC project, and by NA-22, and used computational facilities at NERSC.

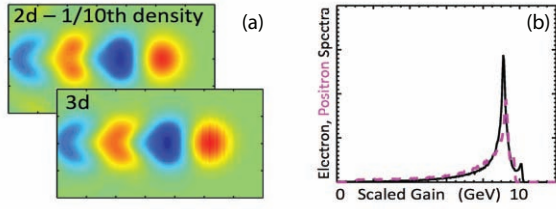


Figure 1: (a) Accelerating field of a scaled stage in 2D at $n_0 = 10^{18} \text{ cm}^{-3}$ and in 3D at $n_0 = 10^{19} \text{ cm}^{-3}$ with an electron beam located in the first wave bucket after the driver. (b) Momentum spectrum of an electron (solid line) and positron (dashed line) beam for $a_0 = 1$, $k_p L = 2$, $k_p r_0 = 5.3$, $n_0 = 10^{19} \text{ cm}^{-3}$ initially on axis followed by a longitudinal linear increase (plasma taper), and beam parameters $k_p \sigma_{LRMS} = 0.5$, $k_p \sigma_{TRMS} = 1$ and $Q = 22.8 \text{ pC}$ (corresponding to $Q = 228 \text{ pC}$ at $n_0 = 10^{17} \text{ cm}^{-3}$).

in 2D and $n_0 = 10^{19} \text{ cm}^{-3}$ in 3D with an electron beam of transverse size $k_p \sigma_{TRMS} = 1.78$ and of charge 130 pC and 50 pC respectively, corresponding to the same amount of beam loading, located in the first wave bucket after the driver. Fig 1(a) shows that the accelerating structure is preserved for different densities and also between 2D and 3D. 2D simulations at $n_0 = 10^{19} \text{ cm}^{-3}$ can then be used at low cost for optimization of the electron beam shape and longitudinal plasma profile to reduce the energy spread and increase the acceleration length [14], showing that it is possible to accelerate a beam to 10 GeV, with 4% energy spread in 0.7 m, as seen in Fig. 1(b). Because the wake is quasi-symmetric in this regime, positrons can be accelerated in a similar way with the same energy gain and energy spread, also shown in Fig. 1(b). Simulations with various pulse lengths, keeping the laser energy constant, show that efficient stages are obtained at $k_p L = 1$, where the laser depletes most of its energy at the dephasing length [15], allowing for 30% more beam charge at the same energy gain. Hence, we were able to increase the efficiency of the stage compared to the initial point design by using scaled PIC simulations, which allows propagation of the laser pulse until depletion.

In the quasi-linear regime, the transverse size of the bunch must be matched to the focusing fields to prevent oscillation of the beam spot size, which can degrade emittance and energy spread [16]. Past simulations used gaussian laser modes, which produce high transverse fields and hence small matched beam sizes. The matched beam spot size, i.e. with no transverse oscillation of the beam, depends on the focusing force strength and is typically of the order of a fraction of a micron for $n_0 = 10^{17} \text{ cm}^{-3}$, $\gamma = 2 \times 10^4$ and $\epsilon_n = 1 \text{ mm mrad}$, where γ is the relativistic factor and ϵ_n the normalized emittance of the electron beam. This small spot size, in turn, limits beam charge, and hence stage efficiency, because high charge density creates a blow out disrupting the wake strongly. In the linear or quasi-linear regime ($a^2 \lesssim 1$) the transverse field is directly

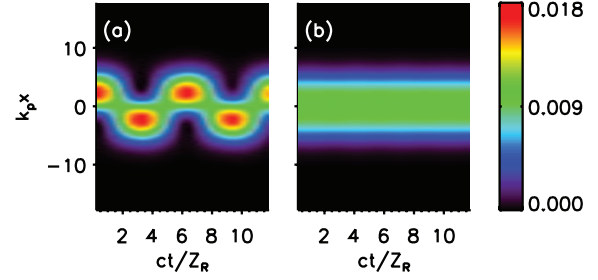


Figure 2: Transverse intensity profile, integrated longitudinally, as a function of propagation distance for the fundamental and first order Hermite-Gaussian mode with parallel (a) and cross (b) polarization, propagating in a matched plasma channel with a density on axis $n_0 = 10^{19} \text{ cm}^{-3}$, with $a_0 = 0.1$, $a_1 = a_0/\sqrt{2}$ and $k_p L = 2$ and $k_p r_0 = 5.3$ for both modes.

proportional to the transverse gradient of the laser intensity profile, $E_{\perp} \sim \nabla_{\perp} a^2$. It is then possible to reduce the focusing forces, and hence increase the matched beam spot size and the beam charge, by tailoring the transverse profile of the laser pulse.

The transverse laser intensity profile can be shaped by using combinations of higher order laser modes. Solutions of the paraxial wave equation, describing the evolution of the slow varying transverse envelope \hat{a}_{\perp} of a low intensity ($|\hat{a}_{\perp}|^2 \ll 1$) laser pulse propagating in a matched plasma channel, expressed in the cartesian coordinate system (x, y, z) are of the form:

$$\hat{a}_{x,(m,p)}(x, y, z) = \frac{a_{m,p}}{(2^{m+p} m! p!)^{1/2}} \times H_m \left(\sqrt{2} \frac{x}{r_0} \right) H_p \left(\sqrt{2} \frac{y}{r_0} \right) \times e^{-(x^2 + y^2)/r_0^2 + i\theta_{m,p}} \quad (1)$$

where H_m are the Hermite polynomials, r_0 is the laser spot size and $\theta_{m,p} = (-1/2k)[k_p^2 + 4(m+p+1)/r_0^2]z$, $k = 2\pi/\lambda$, is the phase shift. In solving the paraxial wave equation one can find that the condition for the laser pulse to be matched, i.e., to have no spot size variation, is to propagate in a plasma channel of the form $n(r) = n_0 + \Delta n_c r^2/r_0^2$, with $\Delta n_c = 1/\pi r_e r_0^2$ where r_e is the classical electron radius [17]. This condition is independent of the mode number: hence, several modes can be propagated in the same plasma channel. On the other hand the phase velocity $\beta_{ph} \simeq 1 - (1/k)\partial\theta_{m,p}/\partial z$ depends on the mode number, inducing a modulation of the intensity profile with propagation distance when adding two modes with the same polarization. This is shown in Fig. 2(a) where addition of the fundamental and first order Hermite-Gaussian modes with parallel polarization causes the transverse intensity profile to oscillate at the frequency $k_{\text{beat}} = 1/Z_R$, where $Z_R = kr_0^2/2$ is the Rayleigh length. This can be avoided by using orthogonal polarization between the two modes, as shown in Fig. 2(b). For sim-

plicity we used 2D simulations, for which $\hat{a}_{x(m)}(x, z) = (a_m/\sqrt{2^m m!})H_m(\sqrt{2}x/r_0)\exp(-x^2/r_0^2 + i\theta_m)$, and $\theta_m = (-1/2k)[k_p^2 + 2(2m+1)/r_0^2]z$. Generalization to 3D is straightforward and presented in detail elsewhere [18].

In 2D, when adding the fundamental ($m = 0$) and first order ($m = 1$) Hermite-Gaussian mode with crossed polarization, the intensity profile is of the form:

$$\hat{a}(x)^2 = a_0^2 e^{-2x^2/r_0^2} \left(1 + \frac{4a_1^2 x^2}{a_0^2 r_0^2} \right) \quad (2)$$

Considering a gaussian profile longitudinally, $\hat{a}(\zeta, x)^2 = \hat{a}(x)^2 \exp(-2\zeta^2/L^2)$, where $\zeta = z - ct$ and L is the laser pulse length, the linear plasma response behind the driver is given by [1]:

$$\phi = -\hat{a}(x)^2 \sqrt{\pi/2} (k_p L/4) e^{-k_p^2 L^2/8} \sin(k_p \zeta) \quad (3)$$

where $\phi = e\Phi/mc^2$ is the normalized electric potential and $|\hat{a}|^2 \ll 1$ is assumed. The transverse electric field is then given by $E_x/E_0 = -(1/k_p)\nabla_x \phi$, where $E_0 = mc^2 k_p/e$ is the cold non-relativistic wavebreaking limit, i.e.,

$$\begin{aligned} \frac{E_x}{E_0} &= a_0^2 \sqrt{\pi/2} (k_p L/4) e^{-k_p^2 L^2/8} e^{-2x^2/r_0^2} \frac{4x}{k_p r_0^2} \\ &\times \left(1 - \frac{2a_1^2}{a_0^2} + \frac{4a_1^2 x^2}{a_0^2 r_0^2} \right) \sin(k_p \zeta) \end{aligned} \quad (4)$$

We see that the transverse field is 0 near axis ($x^2/r_0^2 \ll 1$), for all phases $k_p \zeta$, for $a_1/a_0 = 1/\sqrt{2}$, corresponding to a flat top profile near axis of the intensity profile [Fig. 3(a)].

When the laser pulse is propagating in a plasma channel one has to take into account the curvature of the fields transversely. This can be evaluated to first approximation, assuming a broad channel $k_p^2 r_0^2 \gg 1$, by using $k_p(x) = k_{p0} \sqrt{1 + \Delta n x^2 / n_0 r_0^2}$ in eq. (3) [19]. Assuming $\Delta n_c x^2 / n_0^2 r_0^2 = 4x^2 / k_p^2 r_0^4 \ll 1$ the transverse electric field can be expressed, to the first order in x/r_0 ,

$$\begin{aligned} \frac{E_x}{E_0} &= \frac{-a_0^2 L}{4r_0} \sqrt{\frac{\pi}{2}} e^{-2x^2/r_0^2} e^{-k_p^2 L^2/8} \\ &\times \left[\left(4 - \frac{\Delta n}{n_0} + \frac{k_p^2 L^2}{4} \frac{\Delta n}{n_0} - 8 \frac{a_1^2}{a_0^2} \right) \sin(k_p \zeta) \right. \\ &\left. - \frac{\Delta n}{n_0} k_p \zeta \cos(k_p \zeta) \right] \frac{x}{r_0} + \mathcal{O}(x^3/r_0^3) \end{aligned} \quad (5)$$

Contrary to the solution with a flat plasma profile there is no value of a_1/a_0 that gives a null transverse field for all phases, as shown in Fig. 3(a) for $a_0 = 0.1$. However we can reduce the transverse field to 0 over a small length of phase if we consider the response of the fundamental and the first order Hermite-Gaussian mode separately and by noticing that the transverse fields driven by each of these modes do not cross the axis at the same phase. By introducing a delay ($k_p \zeta_s$) between the two modes and by adjusting their relative intensity (a_1/a_0), with values derived by setting to 0 the first order term in x/r_0 in eq. (5), the

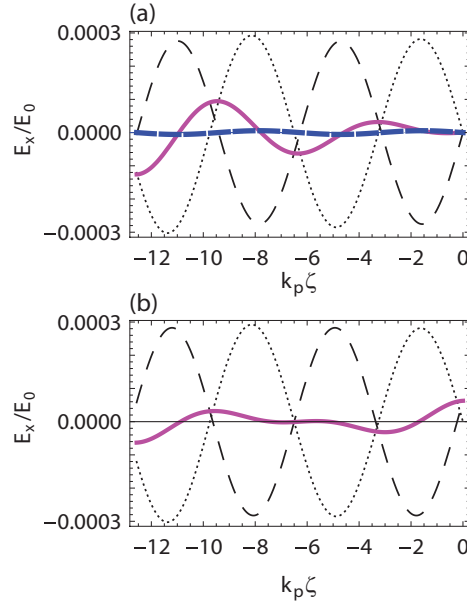


Figure 3: Transverse electric field using higher order modes, as a function of the longitudinal coordinate $k_p \zeta$ at $k_p x = 0.1$ and for $a_0 = 0.1$, $k_p L = 2$ and $k_p r_0 = 5.3$. (a) Without (thick dashed line) and with (solid line) including corrections of the plasma channel. Contribution of the fundamental (dotted line) and Hermite-Gaussian mode (thin dashed line), with $a_1 = a_0/\sqrt{2}$, in the plasma channel are also shown. (b) The first order Hermite-Gaussian mode is delayed by $k_p \zeta_s = 0.22$ and the intensity ($a_1 = 0.72a_0$) adjusted such that the sum (solid line) of the field from the fundamental (dotted line) and the first order mode (dashed line), in the plasma channel, is $\simeq 0$ around $k_p \zeta = -6.5$.

two responses cancel each other and the focusing field can be reduced to 0 over an interval of phase $k_p \zeta$, as seen in Fig. 3(b). This produces constant focusing forces over the length of the electron beam and minimizes the field variations as the beam dephases.

PIC simulations show that the two modes can be propagated simultaneously in the same plasma channel with cross-polarization in the quasi-linear regime ($a_0 \simeq 1$), and that the mode does not evolve, i.e., without oscillations. This allows conservation of a reduced transverse field until the laser depletes, as shown in Fig. 4. Injection of a test electron beam shows that the matched beam radius is increased by almost a factor of 3 when using adjusted higher order modes compared to using a gaussian pulse only. Fig. 5 shows the evolution of the beam radius, with normalized emittance $\epsilon_n = 0.014$ mm mrad ($k_p \epsilon_n = 6 \times 10^{-3}$), accelerated by a plasma wave driven by the fundamental mode only, with $a_0 = 1$, $k_p L = 1$, $k_p r_0 = 5.3$ and $n_0 = 5 \times 10^{18}$ cm $^{-3}$. The electron beam radius oscillates around $\sigma_x \simeq 0.1$ μ m ($k_p \sigma_x \simeq 0.042$) $\pm 30\%$, while the emittance stays constant, with only 0.01% variation. When adding the first order Hermite-Gaussian mode with $a_1 = 0.7a_0$ and a delay of $k_p \zeta_s = 0.2$, the matched beam

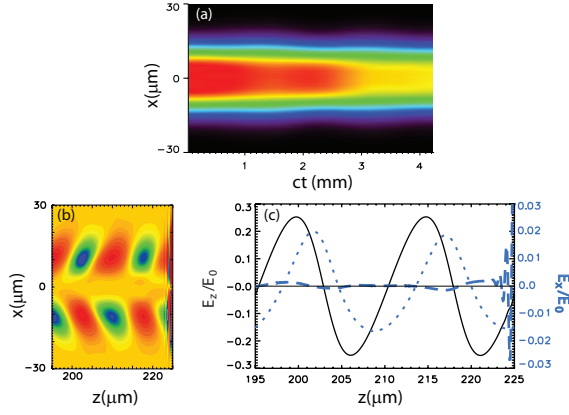


Figure 4: (a) Transverse laser intensity profile, integrated along the longitudinal direction, as a function of the propagation distance, of a fundamental, with $a_0 = 1$, plus a first order Hermite-Gaussian mode, with $a_1 = 0.7$, delayed by $k_p \zeta_s = 0.2$. Both modes verify $k_p L = 1$ and $k_p r_0 = 5.3$, and propagate in a plasma channel with a density on axis $n_0 = 5 \times 10^{18} \text{ cm}^{-3}$. (b) Transverse electric field driven by the fundamental plus first order Hermite-Gaussian mode whose transverse profile is shown in (a). (c) Lineout of the transverse field shown in (b) at $x = 1 \mu\text{m}$ (dashed line) and corresponding accelerating field (solid line). The transverse field driven by the fundamental mode only (dotted line) is also shown.

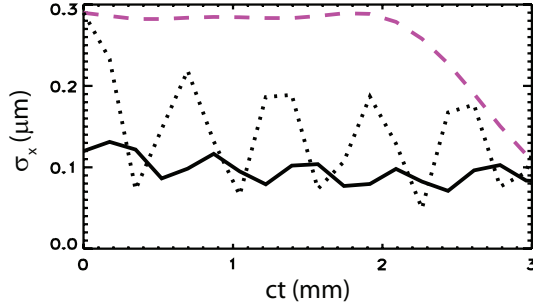


Figure 5: Evolution of a test electron beam spot size in a wakefield driven by the gaussian only, the beam being matched (solid line) and unmatched (dotted line), and in a wakefield driven by the fundamental plus delayed first order Hermite-Gaussian mode (dashed line).

radius is increased to $\sigma_x \simeq 0.285 \mu\text{m} \pm 1.5\%$ for $ct < 2$ mm, where the beam slips in a different focusing phase, and has gained 96% of its maximum energy. This represents almost a factor of 3 increase compared to the fundamental mode only, for which a beam with this radius would be highly mismatched (130% variation). This means that we can potentially increase the charge of the bunch (in 3D) by a factor of 9, without increasing the electron beam peak density. The laser power is twice the gaussian case, hence increasing the efficiency by a factor 4.5.

Advanced Concepts

CONCLUSION

We use PIC simulations for the design of GeV LWFA stages and future 10 GeV modules which will be driven by a 40J-1PW laser system. Stages have been designed in the quasi-linear regime, i.e., moderate laser intensities ($a_0 \simeq 1$), which enables symmetric acceleration of electron and positron beams, advantageous for the high energy particle colliders. We use shorter simulations of stages at higher density and scaling of the physical parameters to predict the behavior of the particle beam in the accelerating structure at reasonable computational cost. This allows us to follow the laser evolution up to depletion for optimization of the stage design.

In the quasi-linear regime, we show that we can tailor the focusing forces acting on the particle beam, by shaping the transverse profile of the laser driver with higher order modes. We show that higher order modes can be propagated simultaneously over many Rayleigh lengths in the same plasma channel, independently of the mode number. However, to avoid intensity modulation due to the difference in phase velocity, cross polarization of the different modes must be used. Linear theory calculations and simulations in the quasi-linear regime show that it is possible to reduce the transverse fields to ~ 0 in the plasma channel by adjusting delay and amplitude of the modes. This allows increase of the matched beam spot size, limited to the region where the fields are linear, of the accelerated electron beam, while keeping the same normalized emittance, enabling increase of the beam charge and stage efficiency important for high energy colliders and gamma ray sources. Future work will include beam loading effects with the use of higher order modes to flatten and reduce the focusing field inside the beam.

REFERENCES

- [1] E. Esarey *et al.*, Rev. Mod. Phys., 81, (2009), 1229.
- [2] W.P. Leemans *et al.* Nature Phys., 2, (2006), 696.
- [3] J. Faure *et al.*, Nature, 444, (2006), 737.
- [4] A.J. Gonsalves *et al.*, Proc. PAC'09, (2009).
- [5] C.B. Schroeder *et al.*, Proc. AAC Wkshp, 1086, (2009), 637.
- [6] C.G.R. Geddes *et al.*, Proc. CAARI'08 (2008).
- [7] C.B. Schroeder *et al.*, Proc. AAC Wkshp, 1086, (2009), 208.
- [8] C. Nieter and J. Cary, J. Comp. Phys., 196, (2004), 448.
- [9] B. Cowan *et al.*, Proc. AAC Wkshp, 1086, (2009), 309.
- [10] C. Huang *et al.*, J. of Compt. Phys., 217, (2006), 658.
- [11] J.-L. Vay, Phys. Rev. Lett., 98, (2007), 130405.
- [12] J.-L. Vay *et al.*, Proc. PAC'09, (2009).
- [13] S. Martins *et al.*, Proc. PAC'09, (2009).
- [14] E. Cormier-Michel *et al.*, Proc. AAC, 1086, (2009), 297.
- [15] C.G.R. Geddes *et al.*, Proc. PAC'09, (2009).
- [16] P. Michel *et al.*, Phys. Rev. E, 74, (2006), 026501.
- [17] E. Esarey and W.P. Leemans, Phys. Rev. E, 59, (1999), 1082.
- [18] E. Cormier-Michel *et al.*, in preparation.
- [19] N.E. Andreev *et al.*, Phys. Plasmas, 4, (1997), 1145.

MULTIPOLE EFFECTS IN THE RF GUN FOR THE PSI INJECTOR

M. Dehler, Paul Scherrer Institut, Switzerland

Abstract

For the 250 MeV test injector at PSI, it is planned to use a 2.6 cell RF gun originally developed for high charge operation in the CLIC test facility CTF-2. First start-to-end simulations assuming perfect field symmetries show, that this gun should be able to generate bunches at 200 pC with an emittance of below 400 nm rad, which would be compatible with the requirements for the SwissFEL. This gun uses double side coupled RF feeds in the last cell as well as asymmetrical tuners in the last two cells, which lead to transverse multipole effects in the field and phase space distribution and may lead to a deteriorated emittance. Since the beam in the last cells is already relativistic at energies between 4 and 6.4 MeV, this effect can be computed in a clean way by looking at the distributions of the integrated beam voltage at the cavity iris and deriving any transverse kicks via the Panovsky-Wenzel theorem. Doing this approach for the various operation modes planned for the SwissFEL shows an emittance dilution well below critical thresholds.

INTRODUCTION

Within the framework of the SwissFEL project at PSI, a 250 MeV test injector facility (see fig. 1 on the next page) is under construction, which will be used to develop test techniques to create and transport high brilliance electron beams suitable for short wave length free electron lasers.

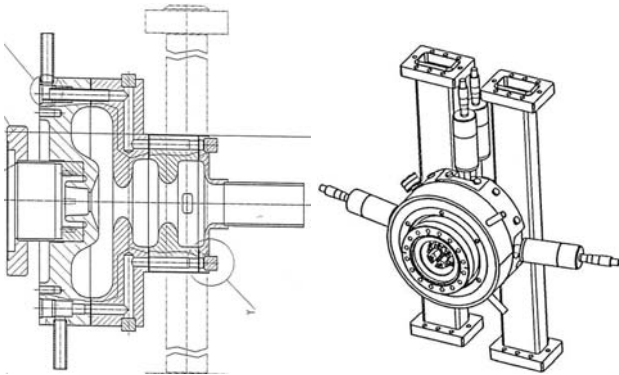


Figure 2: Geometry of the RF gun

Initially it will operate as a stand-alone machine. It must produce the ultra-high brightness electron beam and permit an objective assesment of the technological risks, which are associated with the construction of a low-energy XFEL user facility. Later it is intended to use it as the injector for RF Guns and Linac Injectors

Table 1: Baseline operation modes of the SwissFEL

	High	Small
Bunch charge Q (pC)	200	10
Laserspot σ_r (μm)	270	100
Pulse length FWHM (ps)	9.9	3.7
Acc. gradient (MV/m)	100	100
$\epsilon_{N,slice}$ @ 150 MeV (nm rad)	320	80
$\epsilon_{N,proj}$ @ 150 MeV (nm rad)	330	96

the main linac of the future SwissFEL free electron laser facility.

For the electron source itself, two options are foreseen. The first uses a pulsed DC gun in combination with a two-frequency cavity; the high initial gradient in the gun is supposed to give a high brilliance beam at a reduced current of 5.5 Amperes, which is compressed ballistically by the subsequent two-frequency RF cavity in combination with a drift to approximately 20 A[1]. The alternative consists in using a more convential S-band RF gun running at 100 MV/m and to generate this 20 A beam current directly at the exit of the gun[2]. Recent results from LCLS[3] make this option look rather promising.

For first tests, it is planned to use a 2.6 cell gun originally developed for high current operation in the CLIC test facility CTF-2. The general geometry of this gun[4] is shown in fig. 2. The specialty compared to other design is the large diameter first half cell, where the TM_{02} resonance is used for the main accelerating mode. The original reason for this choice is, that this resonance is particularly well suited for the operation with extremely high beam charges and currents. For the operation at the modest currents of the SwissFEL (Tab. 1), this feature has no influence.

The structure is rotationally symmetric, with perturbations introduced by tuners, field sensors and the holes of the power couplers. These introduce field asymmetries in the monopole type accelerating mode, the main effect of these being transverse kicks on the beam. Small dipole and quadrupole corrector magnets after the gun can easily compensate the integral average kick over the bunch length. What remains, is the transient, time varying part, which leads to emittance growth. More recent designs[5] avoid these problems by obviating the need for tuners all together through more precise manufacturing and by compensating the field perturbation coming from the power coupler with a more complicated race track geometry of the cells.

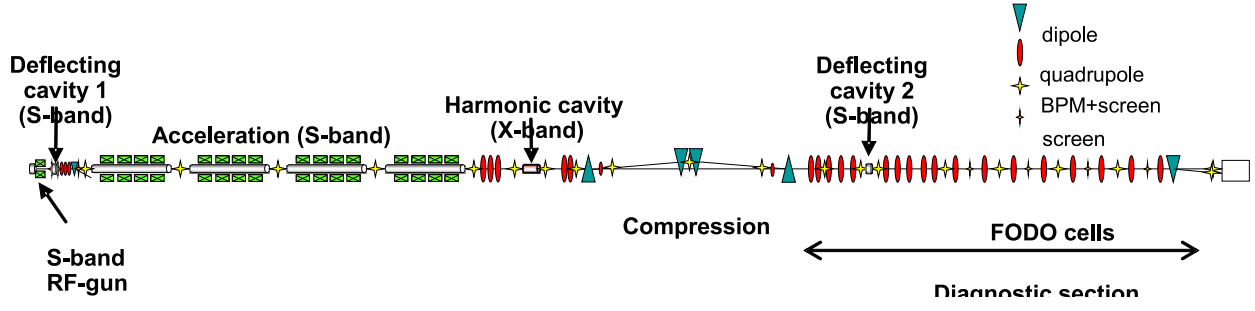


Figure 1: Layout of 250 MeV test injector

METHODOLOGY

Apart from minor perturbations from tuners and field sensors, the first cell is rotationally symmetric in a very good approximation and standard 2 1/2 D beam dynamics simulations will predict the beam dynamics in a relatively precise manner. In cells two and three, we have piston tuners and in cell three, the power coupler with its double holes, which are placed at a 90° angle to the tuners. Cell one also has two, symmetrically placed tuners, but their influence is negligible and was omitted in the analysis.

The input coupler feeds RF power into the gun via a U-turn shaped waveguide, which is side coupled via two symmetric coupling slots (Fig. 3) to the last cell. Due to symmetry, all odd number multipole fields as dipoles, sextupoles etc. cancel out. What is left, are even number multipoles as quadrupoles, which affect the beam transversally. An interesting fact is, that the main contribution to the field distortion is not due to the real power flux through the slots filling the gun and compensating for internal losses, but caused by the perturbation of the cell geometry which would correspond to a reactive power flux.

We have only one tuner of 6.5 mm diameter per cell, so all multipoles are present, the strongest one being the dipole. In this case, the beam will see a transverse kick due to the tuners.

The average kicks and focusing experienced by the beam can be easily compensated by small corrector magnets following the gun. The problem is the amplitude variation over the bunch length giving an uncorrectable increase in the transverse momentum spread and emittance.

Computing the effect directly using a 3D particle-in-cell code becomes challenging – the effect of the coupler needs to be cleanly separated from the emittance dilution due to RF fields, wakes and numerical noise. So an indirect approach was employed here, which, while being approximate, gives a clean picture of the multipolar contribution.

As can be seen in figure 4, the beam in the second and third cells has already a relativistic energy of roughly $\gamma = 9$ and 11 respectively. If, in addition, we approximate the particle trajectories to be parallel to the cavity axis, we can use two techniques to accurately compute the kick distribution coming from the coupler.

The first is an algorithm[7] used for the computation of RF Guns and Linac Injectors

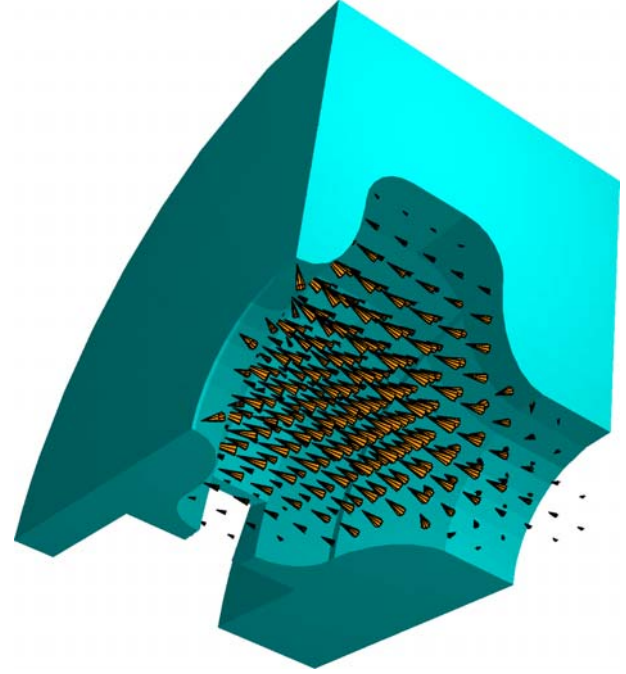


Figure 3: Symmetric quarter of the third cell with coupling slot showing the fundamental mode.

wake potentials. Essentially, it states, that the transverse distribution of the accelerating voltage seen by a relativistic beam at an offset (r, ϕ)

$$V_{||}(r, \phi, t) = \int E_z(r, \phi, z - ct, t) dz$$

follows a potential distribution in two dimensions:

$$\Delta_{r,\phi} V_{||}(r, \phi) = 0 \quad (1)$$

This means in practice, that we just need to integrate the voltages in a numerically advantageous way at the beam pipe radius itself. The values inside the beam pipe are obtained from solving eq. 1 with the given boundary values at the beam pipe. With a cylindrical beam pipe, this is done by computing the Fourier series of the boundary values over the angle ϕ . Given the Fourier coefficients A_n and B_n , the

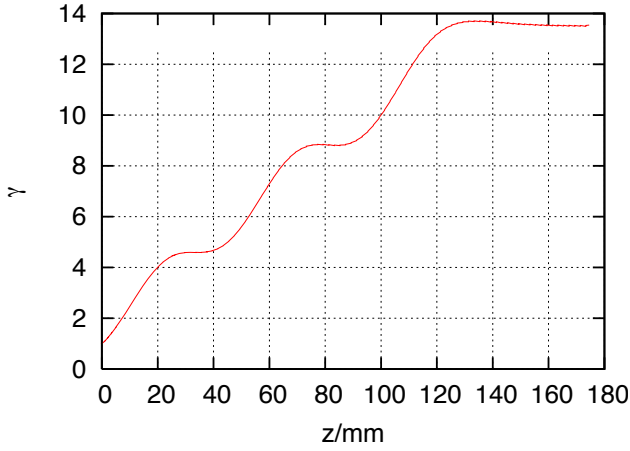


Figure 4: Evolution of beam energy γ during acceleration in the gun. The third cell extends from $z = 78 \text{ mm}$ to $z = 125 \text{ mm}$.

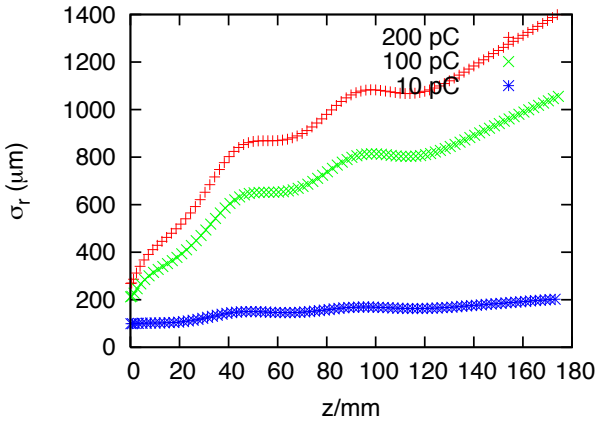


Figure 5: Rms beam radius σ_r during acceleration in the gun. The third cell extends from $z = 78 \text{ mm}$ to $z = 125 \text{ mm}$.

voltage inside the beam pipe is given as:

$$V_{\parallel}(r, \phi) = \sum_{n=0}^{\infty} \frac{r^n}{r_0^n} (A_n \sin n\phi + B_n \cos n\phi)$$

As the second tool, the Panovsky Wenzel theorem [6] relates the transverse kick voltage to the longitudinal one as:

$$\vec{V}_{\perp} = \int \nabla_{r,\phi} V_{\parallel}(r, \phi) dt \quad (2)$$

For the case of a quadrupole kick, the resultant emittance dilution is obtained by calculating the standard deviation of the transverse kick over the whole bunch (transversally and longitudinally) and weighting it with the rms transverse size. Dipoles kicks don't vary with the radial offset, so one only has to compute the transverse kick using the longitudinal distribution.

RF Guns and Linac Injectors

SIMULATION RESULTS

Power couplers

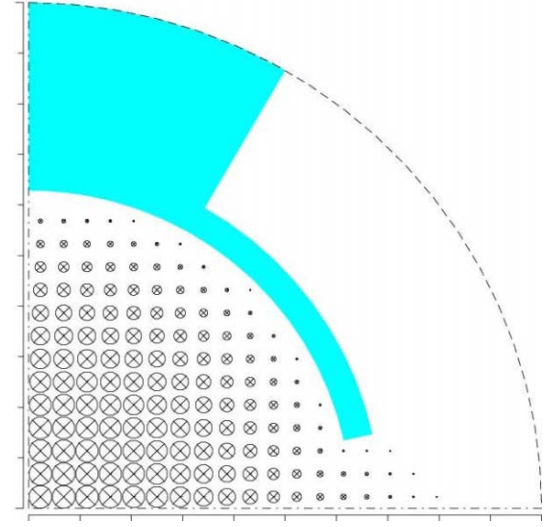


Figure 6: Cut through third cell with coupling hole showing accelerating mode

For the numerical simulations, a quarter of the last cell was modeled, which included the coupling slot and a segment of the U-type waveguide. The modeling was done using a three dimensional cylindrical grid (mesh lines in r , ϕ and z direction). The big advantage of this type of grid is, that the field solution for a completely rotational symmetric cell without coupling slots shows a perfect symmetry. We do not have to worry about numerical artifacts creeping into the field resulting in artificial multipolar kicks. Figure 6 shows the electric field in the equatorial plane of the cell.

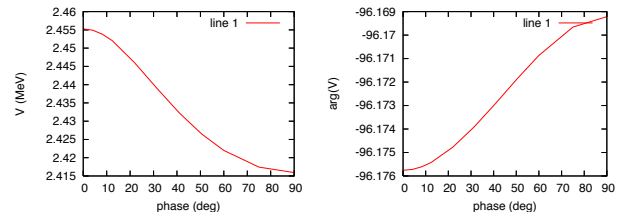


Figure 7: Amplitude and complex phase of integrated beam voltage at the beam pipe wall versus phase ($0 < \phi < 90$).

Figure 7 shows the variation of the integrated beam voltage at the beam pipe wall. Decomposing it into multipole components give the following dependency at $r = 20 \text{ mm}$:

$$V_{\parallel}(\phi) = 2.466 \text{ MeV} + 21.48 \text{ keV} \cdot \cos(2\phi) \dots$$

The solution is dominated by monopole and quadrupole fields, higher multipoles can be safely ignored. The phases of the monopole and quadrupole voltages are aligned better than 0.2 degrees, which via the Panovsky Wenzel theorem means, that the bunch traverses the cavity in the

Table 2: Emittance dilution ϵ_N due to coupling slots for the two operation modes. For comparison, the initial thermal emittance ϵ_{th} is also included.

	High	Small
Bunch charge Q (pC)	200	10
Beam radius in cell 3 (μm)	1530	240
Beam energy in cell 3 (γ)	11	11
Bunch length (ps)	10	4
$\sigma_{V\perp}$ (eV/c)	50.5	3.2
ϵ_N (nm rad)	13	0.3
ϵ_{th} (nm rad)	195	72

zero crossing of the transverse kick. The kick distribution within the beam pipe comes out to be

$$V_{\perp}(r, \phi, t) = 1.7\text{keV/c/mm} \cdot r \cdot \cos(2\phi) \cdot \sin(\omega t)$$

with $\omega = 2\pi \cdot 3 \text{ GHz}$ as the resonance frequency.

For simplicity, we assume a bunch with homogeneous charge distribution of radius r_0 and length Δt . For the variance of the transverse kick we solve

$$\sigma_{V\perp}^2 = \frac{1}{\pi r_0^2 \Delta t} \int_{-\Delta t/2}^{\Delta t/2} \int_0^{2\pi} \int_0^{r_0} V_{\perp}^2(r, \phi, t) r \, dr \, d\phi \, dt$$

to obtain:

$$\sigma_{V\perp} = 3.3 \frac{\text{eV}}{\text{c pC mm}} r_0 \Delta t$$

The resulting emittance dilution comes out as:

$$\epsilon_N = \frac{1}{mc} \sigma_r e \sigma_{V\perp}$$

Table 2 summarizes the results for the different operation modes.

Tuners

All cells contain stub like piston tuner to adjust the resonant frequency and the field balance within the structure. In the TM_{02} half cell, two pistons of 9.2 millimeter diameter are needed to give the required tuning range. These are placed symmetrically at opposite locations on the cell diameter, the resultant perturbation contains only even-order multipoles (quadrupoles etc.) and was omitted in the following calculation due to their small amplitude.

The following two cells each contain one piston of 6.5 millimeter diameter at the cell diameter. From the theoretical design, they should reach into the cavities by two millimeters. The real, actual values have not been measured. So, to be on the safe side, they were assumed to be positioned three millimeters in for the calculation. With only one tuner per cell, we have asymmetry and the effect will be a dipole kick to the beam. Also, both tuners are aligned, so that both contributions will add up linearly.

The calculation follows essentially the same approach as that for the coupling holes, the difference being, that one does it for a dipole kick instead of a quadrupolar distribution. So the intermediate results are omitted and only the final results are listed in table 3.

RF Guns and Linac Injectors

Table 3: Kicks and emittance dilution ϵ_N due to the cell tuners in cell 2 and 3 (tuner position at 3 mm) for the two operation modes. For comparison, the initial thermal emittance ϵ_{th} is also included.

	High		Small	
Bunch charge Q (pC)	200		10	
Tuner in cell	2	3	2	3
Beam radius (μm)	790	142	790	142
Beam energy (γ)	9	9	9	9
Bunch length (ps)	10	10	4	4
$\sigma_{V\perp}$ (eV/c)	137	100	55	40
ϵ_N (nm rad)	24	28	1.7	1.8
ϵ_{th} (nm rad)	195	72	195	72

CONCLUSION

A concern for using CERN gun 5 in the PSI 250 MeV injector was a possible emittance dilution due to transient multipole kicks coming from the RF input coupler. The effect has been computed approximatively via a combination of indirect techniques. The additional dilution due to the coupler geometry is well below the initial thermal emittance to be expected from the cathode and will fulfill the specifications of the machine. Also the transverse kicks coming from the tuners in the last two cells create dipole kicks within the required tolerances.

REFERENCES

- [1] A. Oppelt et al., "Towards a low emittance X-ray FEL at PSI", Proc. of the 2007 Free-Electron Laser Conference (FEL2007), Novosibirsk (Russia)
- [2] Y. Kim et al., "Linear Accelerator for the PSI-XFEL FEL3 Beamline", Proc. of the LINAC 2008, Vancouver, Canada (2008)
- [3] Y. Ding et al., "Measurement and Simulations of Ultra-Low Emittance and Ultra-Short Electron Beams in the Linac Coherent Light Source", Phys. Rev. Lett. 102, 254801 (2009), SLAC-PUB-13525
- [4] R. Bossart and M. Dehler, "Design of a RF Gun for Heavy Beam Loading", proc. of the European Particle Accelerator Conference EPAC, June 10-14 1996, Barcelona, Spain, pp. 1544-1546.
- [5] Z. Li et al., "Coupler design for the LCLS injector S-band structure", proc. of the Particle Accelerator Conference (PAC05), 16-20 May 2005, Knoxville, TN, USA (2005), SLAC-PUB-11728
- [6] W.K.H. Panofsky and W.A. Wenzel, Rev. Sci. Instr. 27 (1956), p. 967.
- [7] T. Weiland, NIM 216 (1983), 31-34

AN APPLICATION OF DIFFERENTIAL ALGEBRAIC METHODS AND LIOUVILLE'S THEOREM: UNIFORMIZATION OF GAUSSIAN BEAMS

Bela Erdelyi^{†‡} and Shashikant Manikonda[‡]

[†]Department of Physics, Northern Illinois University, DeKalb, IL 60115

[‡]Physics Division, Argonne National Laboratory, Argonne, IL 60439

Abstract

Most charged particle beams under realistic conditions have Gaussian density distributions in phase space, or can be easily made so. However, for several practical applications beams with uniform distributions in physical space are advantageous or even required. Liouville's theorem and the symplectic nature of beam's dynamic evolution pose constraints on the feasible transformational properties of these density distribution functions. Differential Algebraic methods offer an elegant way to investigate the underlying freedom involving these beam manipulations. Here, we explore the theory, necessary and sufficient conditions, and practicality of the uniformization of Gaussian beams from a rather generic point of view.

INTRODUCTION

Several practical applications such as irradiation of targets for isotope production, uniform irradiation of detectors for improved efficiency, irradiation of biological samples and materials for testing require manipulation of beam density distributions. Typically, these applications require uniform spatial distributions at the target location. However, most beams delivered by accelerators to these targets are Gaussian. There are several approaches for uniformization of Gaussian beams. One such method, the so-called nonlinear focusing method, uses higher order multipoles to provide a material-less, elegant, purely optical solution. Prior work done in this direction, using nonlinear focusing methods, can be found in [?] and references therein. In this paper we present a new approach based on differential algebraic (DA) techniques to investigate the underlying freedom involving these beam manipulations.

Background

Detailed understanding of the beam dynamics requires the study of the motion of the reference particle as well as the motion of the particle in the relative coordinates. The position and momenta are usually sufficient to describe the motion. Usually the arclength s along the reference orbit is used as the independent variable. At each point on the reference orbit it is possible to define a unique orthogonal coordinate system, denoted by $(\hat{e}_x, \hat{e}_y, \hat{e}_s)$, satisfying a certain set of conditions [?, ?]. In this coordinate system the motion of the particles in the beam can be described using

relative coordinates, which are given by

$$\vec{z}(s) = \begin{pmatrix} x, a = \frac{p_x}{p_0}, y, b = \frac{p_y}{p_0}, \\ l = k(t - t_0), \delta = \frac{(E - E_0)}{E_0} \end{pmatrix}$$

where the position (x, y) describe the position of the particle in the local coordinate system, p_0 is a fixed momentum and E_0 and t_0 are the energy and the time of flight of the reference particle, a and b are the momentum slopes, E is the total energy, and k has a dimension of velocity which makes l a length like coordinate. The point $\vec{z} = 0$ corresponds to the reference particle.

Let position s_i, s_f be the initial and final position on the reference orbit. The transfer map or transfer function \mathcal{M} relates initial conditions at s_i to the conditions at s_f via

$$\vec{z}(s_f) = \mathcal{M}(s_i, s_f) (\vec{z}(s_i)). \quad (1)$$

For weakly non-linear systems, like an accelerator system, the map can be expanded as a Taylor series. Implementation of such a map on a computer would require the map to be truncated at a certain order. A detailed discussion of the properties and use of the Taylor transfer maps can be found in [?].

Beam Phase Space Density Function

Beam production mechanism usually determines the phase space density function describing the distribution of particles in the beam. Let function $f(\vec{z}_i)$ be the initial phase space density function of the beam. According to Liouville's theorem, as long as the system can be considered a Hamiltonian system, the phase space distribution of the beam will stay constant along the trajectories. It also implies that the volume of phase space occupied by the beam is conserved. Hence, it can be written that

$$f(\vec{z}_i) = g(\vec{z}_f), \quad (2)$$

where g is the final phase space density function at any point s_f along the reference orbit. In terms of the transfer map of the system (1), (2) becomes

$$g(x_f, a_f, y_f, b_f, \delta_f) = f \circ \mathcal{M}(s_i, s_f)^{-1}(x_f, a_f, y_f, b_f, \delta_f) \quad (3)$$

where $(x_f, a_f, y_f, b_f, \delta_f)$ are the initial and final phase space coordinates. The function $g(x_f, a_f, y_f, b_f, \delta_f)$ is the new phase space density function, $\mathcal{M}(s_i, s_f)$ is the transfer map of the system. For most practical application the

quantity that is more useful is the density function in position variables. This can be obtained by integrating the phase space density function, g , with respect to the momentum and energy spread variables,

$$\rho(x_f, y_f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_f, a_f, y_f, b_f, \delta_f) \cdot da_f \cdot db_f \cdot d\delta_f$$

Gaussian Density Function Most charged particle beams under realistic conditions have Gaussian density distributions in phase space, or can be easily made so. To keep the discussion reasonably simple, we restrict ourselves to an initial uncorrelated Gaussian distribution. This is without loss of generality, since the formalism can be applied essentially without change to the general, correlated Gaussian situation. The Gaussian density function in phase space, $f(x, a, y, b, \delta)$, is given by

$$f(x, a, y, b, \delta) = K \exp \left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(a-\mu_a)^2}{2\sigma_a^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} - \frac{(b-\mu_b)^2}{2\sigma_b^2} - \frac{(\delta-\mu_\delta)^2}{2\sigma_\delta^2} \right)$$

where $\mu_x, \mu_a, \mu_y, \mu_b, \mu_\delta$ and $\sigma_x, \sigma_a, \sigma_y, \sigma_b, \sigma_\delta$ are the means and variances in x, a, y, b, δ respectively, and the factor $K = 1 / \left((2\pi)^{5/2} \sigma_x \sigma_a \sigma_y \sigma_b \sigma_\delta \right)$ is a normalization constant.

Computation of Beam Density Function Using Differential Algebra

For a beam with initial Gaussian density distribution in phase space a DA based technique has been developed to compute the density function at end of a beam-optics system, given by a transfer map \mathcal{M} . For simplicity we describe the technique for a one dimensional case. The initial phase space density function at point s_i is given by,

$$f(x_i, a_i) = K_2 \exp \left(-\frac{(x_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} - \frac{(a_i - \mu_{a_i})^2}{2\sigma_{a_i}^2} \right), \quad (4)$$

where $K_2 = 1 / (2\pi\sigma_{x_i}\sigma_{a_i})$. Without loss of generality we can assume $\mu_{x_i} = \mu_{a_i} = 0$. We denote the exponential in the above equation as $h(x, a)$,

$$h(x_i, a_i) = -\frac{1}{2} \left(\frac{x_i^2}{\sigma_{x_i}^2} + \frac{a_i^2}{\sigma_{a_i}^2} \right).$$

To compute the density function, $\rho(x_f)$, at point s_f on the reference orbit, we need to solve the integral

$$\begin{aligned} \rho(x_f) &= \int_{-\infty}^{\infty} K \exp(h) \circ \mathcal{M}(s_i, s_f)^{-1}(x_f, a_f) da_f \\ &= \int_{-\infty}^{\infty} K \exp(h \circ \mathcal{M}(s_i, s_f)^{-1})(x_f, a_f) da_f. \end{aligned} \quad (5)$$

Since the integration limits tend to infinity the above integral is difficult to solve directly using numerical techniques. By changing the form of the integral to a Gaussian integral and applying a perturbation method to evaluate the integral term-by-term, one can use the following closed form solution for each term [?]:

$$\int_{-\infty}^{\infty} z^n \exp(-\alpha z^2 + \beta z + \gamma) dz = \sqrt{\frac{\pi}{\alpha}} \exp\left(\frac{\beta^2}{4\alpha} + \gamma\right) \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{k!(n-2k)!} \frac{(2\beta)^{n-2k}}{(4\alpha)^{n-k}}. \quad (6)$$

Below we describe the steps to change the form of the integral in equation 5:

1. Compute the inverse map, $\mathcal{M}(s_i, s_f)^{-1}$, to a pre selected truncation order N [?].
2. Compute $h \circ \mathcal{M}(s_i, s_f)^{-1}$ and arrange the resulting polynomials in powers of a_f by collecting its coefficients A_n . We express this as a sum of second order polynomial and remainder term \mathcal{R}_3 ,

$$h(x_i, a_i) \circ \mathcal{M}(s_i, s_f)^{-1} = A_0(x_f) + A_1(x_f)a_f + A_2(x_f)a_f^2 + \mathcal{R}_3(x_f, a_f)$$

$$\mathcal{R}_3(x_f, a_f) = \sum_{n=3}^N A_n(x_f) a_f^n.$$

3. Expand the exponential of the remainder term $\mathcal{R}_3(x_f, a_f)$ in Taylor series and arrange the resulting polynomial in powers of a_f by collecting its coefficients C_j , $\exp(\mathcal{R}_3(x_f, a_f)) = \sum_{j=3}^N C_j(x_f) a_f^j$. The equation 5 can now be rewritten as,

$$\begin{aligned} \rho(x_f) &= K \sum_{j=3}^N C_j(x_f) \int_{-\infty}^{\infty} a_f^j \exp(A_0(x_f) + A_1(x_f)a_f + A_2(x_f)a_f^2) da_f. \end{aligned} \quad (7)$$

The equation is the desired form where we can utilize the Gaussian integral formula, equation 6. The same procedure, essentially without any change, will work for the 5D case too, where the steps shown above are performed one-by-one for each integration variable a , b and δ .

APPLICATIONS

We consider a system with an octupole magnet and a quadrupole magnet, separated by a 20cm drift. The setup is followed by a 110 cm drift to the final image. Each magnet is 25cm in length and 20cm full aperture. Both the transverse beam optics, in coordinates (x, a, y, b) , and kinetic energy spread are considered for the simulation. A

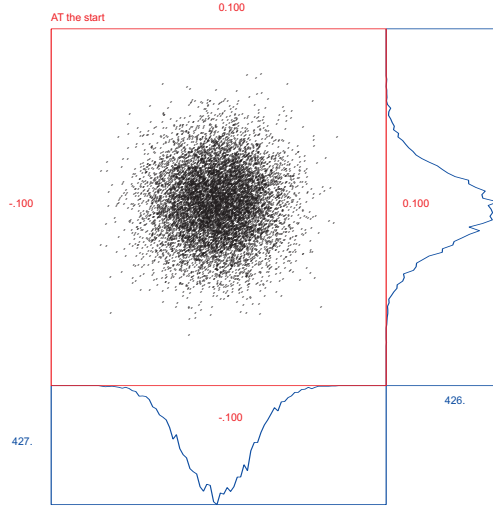


Figure 1: XY distribution plot and the histogram plot of 10000 particles at the entrance of the system. The beam has Gaussian density distribution in phase space.

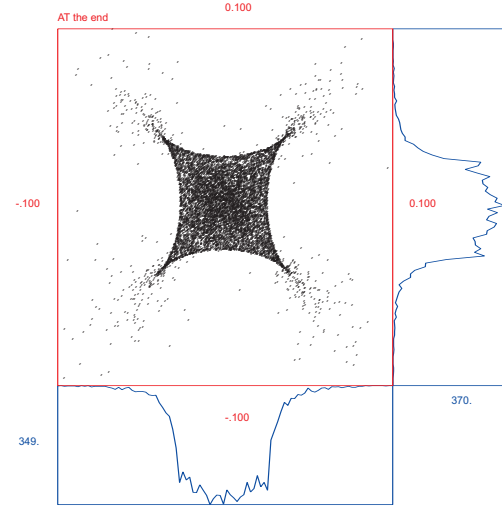


Figure 2: XY distribution for 10000 particles at the final image.

100MeV proton beam with RMS emittance, $\varepsilon_x = \varepsilon_y = 2 \times 10^{-5}$ and 5% kinetic energy spread is considered. Using the DA technique described in the section, the beam density function at the final image is computed as a Taylor expansion in position variables. By using the poletip field of the octupole magnet as a fit parameter, the lower order coefficients in the Taylor expansion of the beam density function are minimized leading to a near uniform beam density distribution with resonable octupole strengths. Figures 1 and 2 show the $x - y$ distribution and histogram in x and y for the beam at the start and end of the system. It can be noticed from the figures that the beam density function changes from a Gaussian distribution to a near uniform one. The uniformity can be improved in principle by the use of even higher order multipoles, if necessary.

SUMMARY

We showed that DA-based methods are powerfull and very general methods for applications to beam uniformization of Gaussian beams. The main ingredients of the method are DA-based methods that allow transfer map computation and inversion of arbitrary lattice maps, Liouville's theorem, and a novel application of closed form Gasussian integrals to this setting. Unlike previous methods, our method is free of assumed correlations in the initial beam distributions, it is as easy to work with in 5D as it is in 2D, takes into account arbitrary lattice nonlinearities to any desired order, and the optimization of the system can be automated. The method has been implemented in the code COSY Infinity [?, ?]. More detailed studies and applications, including the available freedom in phase space manipulations will be published in forthcoming papers.

Beam Dynamics, Other

Acknowledgement

This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357.

COSY EXTENSIONS FOR BEAM-MATERIAL INTERACTIONS*

L. Bandura[#], NSCL, Michigan State University, East Lansing, MI 48824
B. Erdelyi, Northern Illinois University, DeKalb, IL 60115. and ANL, Argonne, IL 60439

Abstract

While COSY INFINITY provides powerful DA methods for the simulation of fragment separator beam dynamics, the master version of COSY does not currently take into account beam-material interactions. These interactions are key for accurately simulating the dynamics from heavy ion fragmentation and fission. In order to model the interaction with materials such as the target or absorber, much code development was needed. There were four auxiliary codes implemented in COSY for the simulation of beam-material interactions. These include EPAX for returning the cross sections of isotopes produced by fragmentation and MCNPX for the cross sections of isotopes produced by the fission and fragmentation of a ^{238}U beam. ATIMA is implemented to calculate energy loss and energy and angular straggling. GLOBAL returns the charge state. The extended version can be run in map mode or hybrid map-Monte Carlo mode, providing an integrated beam dynamics-nuclear processes design optimization and simulation framework that is efficient and accurate. The code, its applications, and plans for large-scale computational runs for optimization of separation purity of rare isotopes at FRIB will be presented.

INTRODUCTION

The next generation of nuclear physics research will require advanced exotic beam facilities based on heavy ion driver accelerators. There are many next-generation facilities that are currently under commissioning, construction, or envisioned [1-5]. Included amongst these is the future Facility for Rare Isotope Beams (FRIB) at the National Superconducting Cyclotron Lab at Michigan State University. These facilities are capable of producing exotic beams composed of rare nuclei in large quantities. The exotic isotopes are produced via projectile fragmentation and fission in targets. High-performance fragment separators, a key component of all rare isotope facilities, consist of superconducting magnets that are used for the capture, selection, and transport of rare isotopes. Large aperture magnets are necessary in order to accept rare isotope beams with large emittances resulting from their production mechanism.

The beam optics code COSY INFINITY uses powerful differential algebraic (DA) techniques for computing the dynamics of the beam in the fragment separator through high order transfer maps [6]. However, until now it has lacked the ability to calculate the beam-material interactions occurring in the target and energy absorbers. Here, a hybrid map-Monte Carlo code has been developed

and integrated into COSY in order to calculate these interactions. The code tracks the fragmentation and fission of the beam in target and absorber material while computing energy loss and energy and angular straggling as well as charge state evolution. This is accomplished by implementing auxiliary codes such as ATIMA [7] and GLOBAL [8]. EPAX [9] is utilized to return cross sections of fragmentation products. The special case of fission has been treated by using the code MCNPX [10] to accurately predict the cross sections and dynamics of exotic beams produced by a ^{238}U beam incident on a Li or C target. The extensions to the code have made it possible to simultaneously compute high order optics and beam-material interactions in one cohesive framework.

The hybrid map-Monte Carlo code can be used to calculate important quantities that describe the performance of the fragment separator. These include the transmission and the separation purity. In a map-only approach, calculations such as these are not possible. Experimental planning and optimization is possible with the map-Monte Carlo code, as various fragment separator settings can be readily adjusted. Here we present a description of the code and how it is implemented in COSY.

DESCRIPTION OF HYBRID MAP-MONTE CARLO CODE

While COSY INFINITY possesses a powerful DA framework for accurate simulation of beam dynamics in electromagnetic fields, the master version does not allow for the simulation of beam-material interactions. This ability is necessary, however, in order to model the dynamics of fission and fragmentation products. In order to track heavy ions through target and absorber material, much code development to COSY was needed.

New additions made to the code include the implementation of auxiliary codes to determine how many of each type of isotope are produced from the fragmentation and fission of an energetic heavy ion beam of a given nuclear mass A and nuclear charge Z incident on a specified target of a given thickness. Also, the dynamics of these new particles need to be determined. It is necessary to model the cross sections and dynamics of fragmentation and fission separately due to the auxiliary codes available.

Fragmentation Cross Sections

In the case of any primary beam that has nuclear charge $Z < 92$, the secondary particles of interest are fragmentation products. The cross sections of these

*This work was supported by the U.S. Department of Energy, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357
#bandura@anl.gov

particles are found via the code EPAX, which has been integrated into the code. The parameters that must be input to the code are the nuclear mass A and nuclear charge Z of the primary beam and the nuclear mass and charge of the product. Also, it is necessary that the target's nuclear mass and charge and thickness are input. EPAX will then return the cross section of the product in millibarns.

Fragmentation Dynamics

The secondary particles that emerge from the target have different kinematics depending on the production mechanism by which they were formed. For nuclear fragmentation, the angular divergence and momentum deviation of the secondary particle will solely be based on its mass and on the initial mass of the nucleus that fragments. This is called the “fireball” method, where the momentum is given by a Gaussian distribution with the standard deviation given by.

$$\sigma = \frac{85}{c_{light}} \sqrt{\frac{A_c(A_p - A_c)}{A_p - 1}},$$

where A_c is the child particle and A_p is the parent particle from which the child particle fragments. The momentum of the new fragment is modified by adding a random number chosen by Gaussian distribution with standard deviation σ . The parallel component of the momentum is used to calculate energy loss and straggling, and the perpendicular component is used to calculate the angular divergence of the particle and angular straggling.

Fission Cross Sections

In contrast to the fragmentation process, the cross sections of fission products are energy-dependent. The map-Monte Carlo code utilizes MCNPX in order to find the cross sections of all the isotopes produced by a ^{238}U beam incident on a Li or C target. These are two targets under development for the FRIB (Figure 1). MCNPX was run for four different beam energy and target combinations, namely 200, 400, 800, and 1500 MeV/u incident on both Li and C targets of thickness 0.1068 g/cm². The ^{238}U beam was assumed to be point-like and have no angular divergence or energy spread. The output from MCNPX is the number of particles of each isotope produced N_{prod} . This number includes all the isotopes produced from all nuclear processes. With this number, the cross section of any isotope may be computed by the formula:

$$\sigma_{cs} = \frac{N_{prod}A}{N_0 x \rho N_A},$$

where A is the nuclear mass of the target, N_0 is the number of source ^{238}U particles, x is the target thickness, ρ is the target density, and N_A is Avogadro's number. The number of source particles in each run was between 6×10^8 and 1×10^9 .

The cross section data provided by MCNPX were used to interpolate the cross section of the fission products as a second order polynomial in energy given by

$$\sigma_{cs} = c_0 + c_1 E + c_2 E^2,$$

where E is the energy of the ^{238}U beam and the c_n 's are the coefficients of the interpolation for a given isotope with nuclear mass A and nuclear charge Z . The coefficients of this polynomial are listed in a file that is read once by COSY and stored in an array each time the code is run. This method is fast and uses very little memory. Another benefit to this method is the in which a more up-to-date fission model may be included in the future as developments in the field occur.

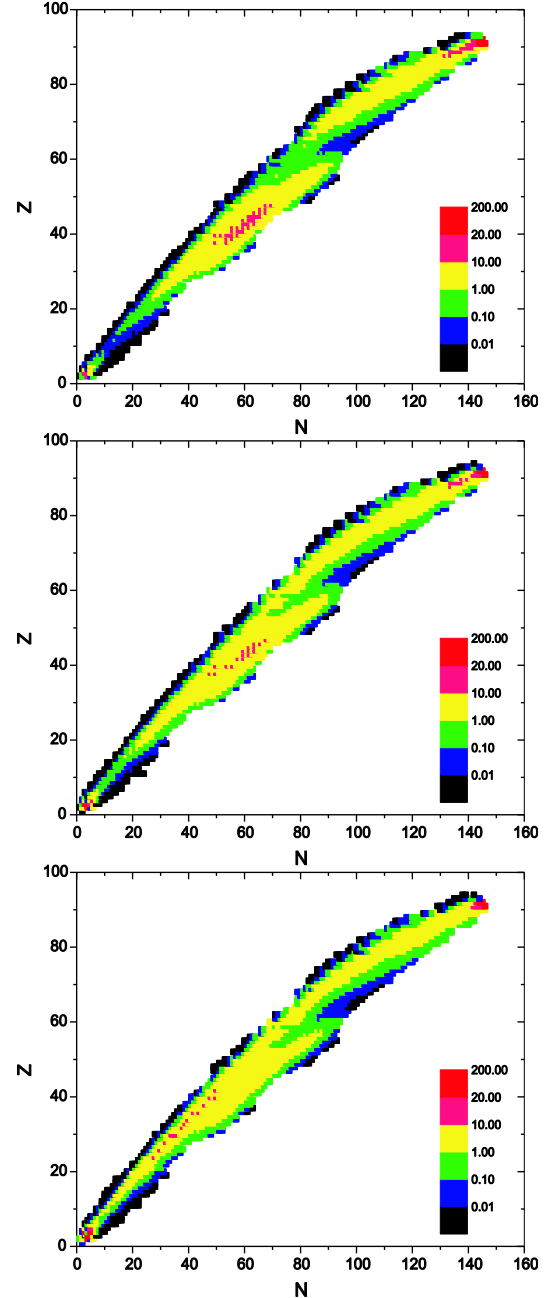


Figure 1: Cross sections of all isotopes produced by 200, 400, and 1500 MeV/u ^{238}U beam incident on a 0.1068 g/cm² Li target.

Fission Dynamics

For fission, the coordinates are not only based on the masses of the nuclei as is the case for fragmentation. There is an extra energy release that results from the fissioning nucleus. This extra energy release means that the products will have large δ , or energy spread, and also large angular divergences. If δ and the angular coordinates are plotted, then the result is a “fuzzy” spherical shell that represents the phase space that is occupied by fission products (Figure 2).

The fact that all fission products are emitted from the target in a sphere can be used to model the dynamics of the emitted beam. More precisely, the coordinates of these fission products are best represented by a spherical shell with some thickness. In order to determine the initial coordinates of each isotope that will pass through the fragment separator, we must have a method to obtain the “sphere” of each isotope in a random manner. At low energies, the thickness of the spherical shell is large, with the most particle density at larger radii. As the energy increases, the thickness of the spherical shell becomes thinner and more dense and, hence, represents a lower beam emittance.

CONCLUSION

A hybrid map Monte Carlo code has been developed to accurately model beam-material interactions for the purpose of fragment separator beam dynamics simulation. A comprehensive fission model was developed to accurately model cross sections and kinematics of isotopes produced from a ^{238}U primary beam. Using the code, one may simulate a variety of exotic beam experiments and compute important quantities such as the separation purity and transmission of various rare isotopes.

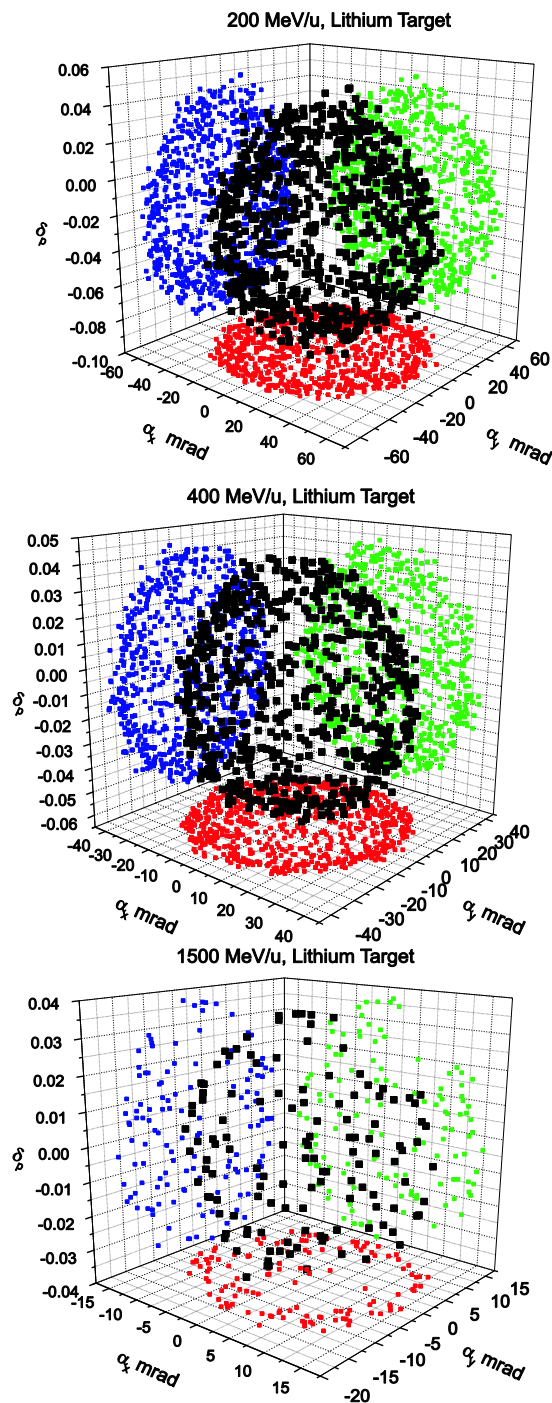


Figure 2: Kinematics of a ^{132}Sn beam resulting from the fission of 200, 400, and 1500 MeV/u ^{238}U primary beam upon interaction with a 0.1068 g/cm^2 Li target. The ^{132}Sn coordinates are shown in black with the colored circles representing the projection of these coordinates on the various coordinate planes. The coordinates shown are the momentum deviation (δ_p) and scaled horizontal and vertical momenta (α_x, α_y).

REFERENCES

- [1] RISAC. Scientific Opportunities with a Rare-Isotope Facility in the United States. Technical report, NRC, 2006.
- [2] J. Nolen. Overview of the U.S. Rare Isotope Accelerator Proposal. Nucl. Phys., A734:661.668, 2004.
- [3] Y. Yano. RI Beam Factory project at RIKEN. In Proc. 17th Int. Conf. Cycl. Appl., pages 169.173, Tokyo, Japan, 2004.
- [4] T. Kubo, K. Kuska, and K. Yoshida et al. Status and overview of superconducting radioactive isotope beam separator BigRIPS at RIKEN. IEEE Transactions on Applied Superconductivity, 17:1069.1077, 2007.
- [5] W. Henning. The GSI project: An International Facility for Ions and Antiprotons. Nucl. Phys., A734:654.660, 2004.
- [6] M. Berz. Modern Map Methods in Particle Beam Physics. Academic Press, San Diego, 1999.
- [7] H. Weick. <http://www-linux.gsi.de/weick/atima/>.
- [8] C. Scheidenberger, T. Stohlker, W. Meyerhof, H. Geissel, P. Mokler, and B. Blank. Charge states of relativistic heavy ions in matter. NIM B, 142:441.462, 1998.
- [9] K. Summerer and B. Blank. EPAX version 2: a modified empirical parameterization of fragmentation cross sections. Nuclear Physics A, 701:161C.164C, 2002.
- [10] M. James, G. McKinney, J. Hendricks, and M. Moyers. Recent enhancements in mcnpX: Heavy-ion transport and the LAQGSM physics model. NIM A, 562(2):819.822, 2006.

OPTIMIZING SRF GUN CAVITY PROFILES IN A GENETIC ALGORITHM FRAMEWORK*

A.S. Hofler[#], P. Evtushenko, F. Marhauser, Thomas Jefferson National Accelerator Facility,
Newport News, VA 23606, U.S.A.

Abstract

Automation of DC photoinjector designs using a genetic algorithm (GA) based optimization is an accepted practice in accelerator physics. Allowing the gun cavity field profile shape to be varied can extend the utility of this optimization methodology to superconducting and normal conducting radio frequency (SRF/RF) gun based injectors. Finding optimal field and cavity geometry configurations can provide guidance for cavity design choices and verify existing designs. We have considered two approaches for varying the electric field profile. The first is to determine the optimal field profile shape that should be used independent of the cavity geometry, and the other is to vary the geometry of the gun cavity structure to produce an optimal field profile. The first method can provide a theoretical optimal and can illuminate where possible gains can be made in field shaping. The second method can produce more realistically achievable designs that can be compared to existing designs. In this paper, we discuss the design and implementation for these two methods for generating field profiles for SRF/RF guns in a GA based injector optimization scheme and provide preliminary results.

OPTIMIZATION SYSTEM OVERVIEW

Alternative Platform and Programming Language Independent Interface for Search Algorithms (APISA) [1] builds on the Platform and Programming Language Independent Interface for Search Algorithms (PISA) [2] system. PISA provides a modular way to combine GAs and problems. It uses two communicating state machines to separate the GA implementation from the problem model evaluation. It is easy to apply different GAs to a given problem because the state machine structures are well defined and the files used to communicate between the two state machines are standardized. Changing the GA only requires running the optimization scheme with a different GA state machine; the problem model is unchanged. APISA takes advantage of this compartmentalization and provides problem model evaluations customized for accelerator physics. APISA uses A Space Charge Tracking Algorithm (ASTRA) [3] or General Particle Tracer (GPT) [4] to simulate particle dynamics making it a suitable tool for injector design optimization. The version of APISA described in this paper relies on ASTRA for the beam dynamics simulations.

*Authored by JSA, LLC under U.S. DOE Contract DE-AC05-06OR23177. The U.S. Govt. retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce this for U.S. Govt. purposes.
[#]hofler@jlab.org

FIELD MORPHING

The original version of APISA assumes that the field descriptions provided for the magnets and rf accelerating components are fixed and that the optimization can vary the amplitude and/or phase of these elements. This version of APISA, which is geared toward designing SRF/RF gun based injectors, allows the functional form of the on-axis field description of the gun to be varied.

Under the assumption that the desired field pattern resembles a π mode, the software uses a sine wave as the fundamental form for the field description. A truncated Fourier series,

$$f(z) = 1 + \sum_{n=1}^{15} a_n \cos\left(2\pi n \frac{z}{L_{\text{cavity}}}\right) + \sum_{n=1}^{15} b_n \sin\left(2\pi n \frac{z}{L_{\text{cavity}}}\right)$$

where L_{cavity} is the length of the cavity, is then applied to the fundamental form to produce the field description used in the beam dynamics simulation. Each coefficient of the series can be designated as a variable controlled by the optimization scheme or fixed to a specified value. The default value for all coefficients is zero. Other variables that can be fixed or varied are the frequency of the underlying sine function and the number of cells the underlying sine function should represent. The fractional part of the number of cells is interpreted as a gun cell, that incorporates the beam emitting cathode and generally precedes the full cells. The number of cells and the sine frequency are used to calculate L_{cavity} and the free space wavelength of the cavity.

The system computes characteristics of the generated field profile and the morphing function, $f(z)$, and these characteristics can be used as constraints or objectives in the optimization. For example, to preserve the nodes that occur between cells in a π mode, the minimum of $f(z)$ must be positive; otherwise, additional unwanted zero crossings are introduced in the generated field profile. Because $f(z)$ can change the frequency of the generated field, the system determines the resonance frequency from a Fourier transform of the field profile. The frequency can be used as a constraint and an objective to guide and restrict the frequencies of the fields produced.

Preliminary results for a PITZ-like 1.5 cell RF gun operating at 40 MV/m followed by an emittance compensation solenoid [5] indicate that the field amplitude in the half cell should be much larger than in the full cell. These results are obtained using 128 nodes of a Jefferson Lab cluster computer. Each case represents

14 generations with 128 individuals in each generation. The optimization attempts to minimize the beam size and emittance for a 1.65 nC 24 ps FWHM plateau electron bunch 1.618 m downstream of the cathode [6] for a fixed RF phase and solenoid setting. The variables in the optimization are the coefficients of $f(z)$ subject to the constraints that the beam size, emittance, and the minimum of $f(z)$ are all positive. Two cases are shown in Figures 1 and 2 where the half cell amplitude is more than twice the full cell amplitude.

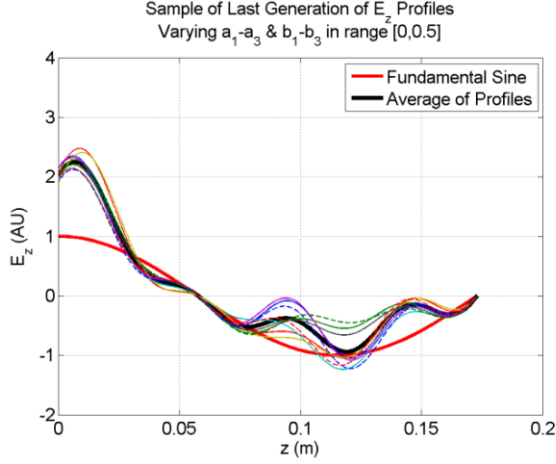


Figure 1: Varying the first three Fourier coefficients.

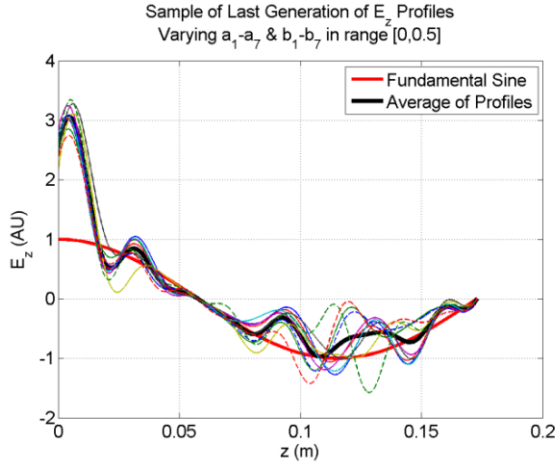


Figure 2: Varying the first seven Fourier coefficients.

In Figure 1, the optimization is changing six coefficients, a_1 through a_3 and b_1 through b_3 of $f(z)$, while in Figure 2 the optimization is varying fourteen coefficients, the first seven for each a_n and b_n . In Figure 2, with more and higher frequencies available to include in the field profile, the optimization pushes the peak field closer to the cathode. Both cases strongly indicate that high electric fields on the cathode yield better gun

performance. These cases, also, have significant particle loss which requires further study.

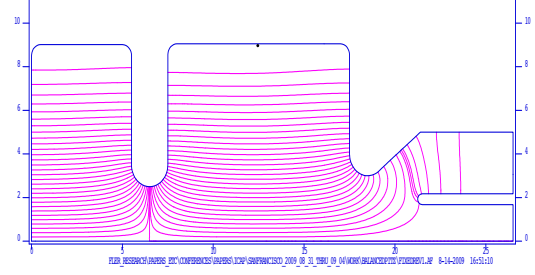


Figure 3: Standard balanced field PITZ geometry.

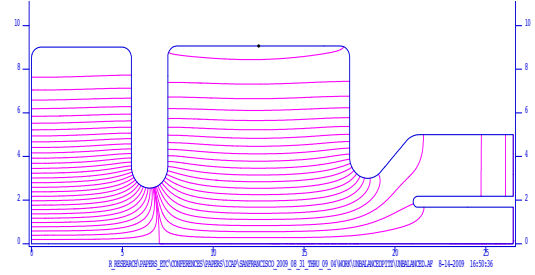


Figure 4: Unbalanced field geometry.

Table 1: Changes relative to the balanced geometry

Element	Dimension	Change
Half Cell	radius	-37.4 μm
Iris	radius	+0.5 mm
Full Cell	radius	+162.6 μm

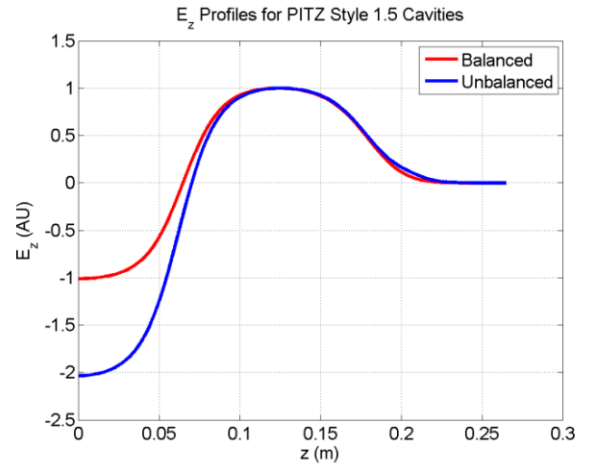


Figure 5: Normalized E_z profiles for the two geometries.

To test the validity of this conclusion in a realistic cavity design, two variations of the PITZ 1.5 cell cavity have been modelled. The first yields the standard well balanced field profile design in Figure 3. Figure 4 shows

the second case where the geometry is modified slightly to create a field with the amplitude in the half cell being roughly twice the full cell amplitude. Table 1 summarizes the relative changes in the physical dimensions for the unbalanced geometry relative to the balanced case. The on-axis field profiles for both geometries are shown in Figure 5.

As in the optimization, the main solenoid strength and RF injection phase are fixed in the ASTRA simulations. In addition, simulations using ASTRA's autophase feature to find the phase for maximum energy gain are provided for comparison. For the balanced field cases, the peak electric field is 40 MV/m whereas it is 80 MV/m in the unbalanced case. Figures 6 and 7 show the transverse normalized emittance and beam size, respectively, for all cases.

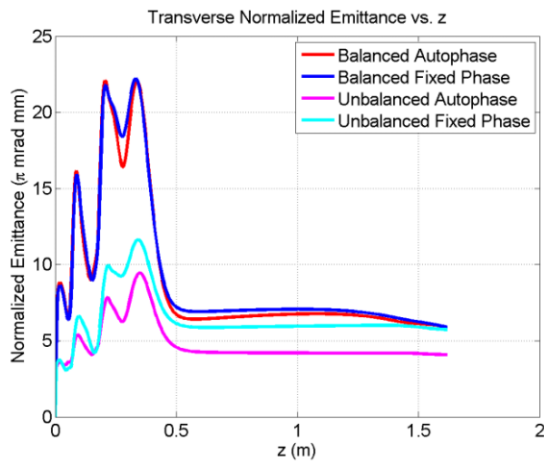


Figure 6: Transverse emittance along beam line.

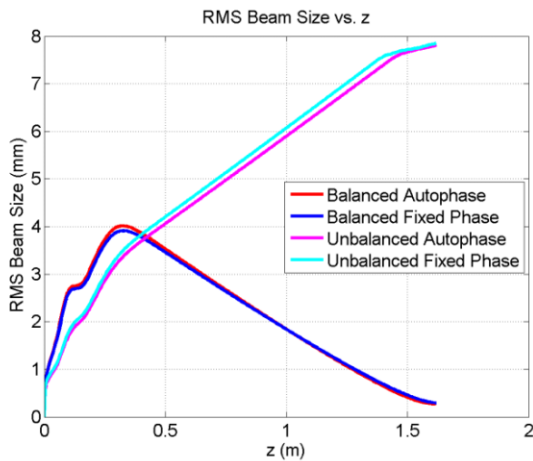


Figure 7: RMS beam size along beam line.

While the final emittance for the fixed phase unbalanced field case is comparable to both balanced field results, the unbalanced results show a general improvement in emittance due to the increased RF

focusing from the higher gradient in the half cell. The charge transmission is significantly better in the unbalanced case where particles are lost on apertures. In contrast, the balanced case particle loss is due to backward travelling particles. The difference in beam size is due to changes in RF focusing and can be managed by increasing the solenoid strength.

CAVITY GEOMETRY MORPHING

Cavity geometry determines the field characteristics, so it is necessary to consider a system that varies the cavity geometry. Including the geometry configuration in the optimization allows for studying the impact of changes in the cavity geometry on the beam dynamics of an injector design. It is also a step forward in automating the injector design and optimization process since the cavity shape can be developed and tested concurrently with the other injector elements.

Including cavity geometry in the optimization framework requires incorporating into APISA a field solver to compute the field from the specified geometry. A goal of this effort is to use free software packages as much as possible. Poisson Superfish [7] is a generally accepted tool for computing the field information for cylindrically symmetric cavities and will be the field solver used in this system. Using Wine (Wine Is Not an Emulator) [8] with Xvfb, the X Window's virtual frame buffer [9], to capture the graphics output, Poisson Superfish can run on a monitorless Linux machine in a cluster computer.

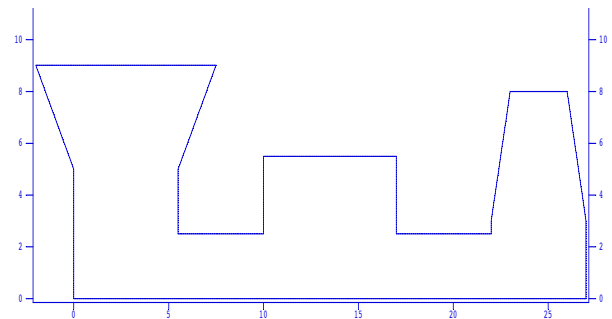


Figure 8: Straight line approximations for re-entrant (left), pillbox (center) and elliptical (right) cavities.

Poisson Superfish relies on a drive point that is treated as a source of a fictitious magnetic current to compute the cavity fields [7]. The drive point must be placed in an area of the cavity geometry that has a sufficiently high magnetic field. Changing the drive point location can significantly affect the cavity mode that is excited in the structure. Also, the search frequency can impact the field mode Poisson Superfish finds. To isolate these issues from the optimization, the Poisson Superfish processing, which will consider the results of several drive point locations and search frequency choices, will be encapsulated in a separate program that feeds back to the optimization scheme the generated field profile for the

most π -like mode and characteristics of the field and cavity properties that can be used as constraints or objectives. The Poisson Superfish processing will not include tuning the cavity geometry as that can change parameters that are under the control of the optimization and therefore mislead the optimization.

The geometry file generation will be broken into two phases. The first will use straight line cavity geometries shown in Figure 8 that can be easily morphed from a pillbox cavity to approximations for elliptical and re-entrant cavities by changing the tilt of the cavity end caps. The second phase will generate more realistic cavities using smooth elliptical curves to describe the geometry.

CONCLUSION

Two methods for bringing cavity field characterization into an automated injector optimization framework have been presented. The first method assumes a general underlying form of the optimal cavity field profile that the optimization can transform by varying the coefficients of a truncated Fourier series used to morph the fundamental form. Applying this approach to a PITZ style 1.5 cell RF gun based injector leads to the conclusion that higher accelerating field in the cathode region of the gun improves gun performance. An unbalanced field profile, with the field stronger in the half cell, is preferred. Because the field morphing method does not consider the boundary conditions of the cavity, a second approach that works with cavity geometries is needed to develop more realistic optimized injector designs. The design for a scheme that morphs the cavity geometry has been described. The two methods balance each other because the first concentrates on what the injector designer wants

from an optimal field profile to achieve the desired injector beam characteristics and the second considers the optimal performance that can be realized subject to the physics of the cavity.

REFERENCES

- [1] I. V. Bazarov and C. K. Sinclair, "Multivariate optimization of a high brightness dc gun photoinjector," *Phys. Rev. Special Topics - Accelerators and Beams* 8 (2005), 034202.
- [2] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, "PISA - a platform and programming language independent interface for search algorithms," *Evolutionary Multi-Criterion Optimization (EMO 2003)*, Faro, Portugal, 2003, pp. 494-508 (2003).
- [3] K. Floettmann, "ASTRA: A Space Charge Tracking Algorithm," <http://www.desy.de/~mpyflo>.
- [4] General Particle Tracer, <http://www.pulsar.nl/gpt>.
- [5] K. Abrahmian, *et al.*, "Characterization of the electron source at the photo injector test facility at DESY Zeuthen," *Nuclear Instruments and Methods in Physics Research A* 528 (2004), pp. 360-365.
- [6] A. Hofler, P. Evtushenko, and M. Krasilnikov, "RF gun optimization study," *PAC'07*, Albuquerque, N.M., June 2007, TUPMS064 pp. 1326-1328 (2007); <http://www.JACoW.org>.
- [7] J. Billen and L. Young, "Poisson Superfish," LA-UR-96-1834, Revised January 13, 2006.
- [8] Wine HQ, <http://www.winehq.org/>.
- [9] Xvfb, <http://ftp.x.org/pub/X11R7.0/doc/html/Xvfb.1.html>.

COMPUTATIONAL MODELS FOR MICRO CHANNEL PLATE SIMULATIONS*

V. Ivanov, Muons, Inc. 522 N. Batavia Ave., Batavia, IL 60510, U.S.A.[#]

Abstract

Many measurements in particle and accelerator physics are limited by the time resolution with which individual particles can be detected. This includes particle identification via time-of-flight in major experiments like CDF at Fermilab and Atlas and CMS at the LHC, as well as the measurement of longitudinal variables in accelerator physics experiments. Large-scale systems, such as neutrino detectors, could be significantly improved by inexpensive, large-area photo-detectors with resolutions of a few millimetres in space and a few picoseconds in time. The invention of a new method of making micro-channel plates (MCP) promises to yield better resolution and be considerably less expensive than current techniques.

INTRODUCTION

One of the first full numerical models for MCP simulations was suggested by A.J. Guest [1]. Further improvements of this model were done by Y. Kulikov [2] in simulation of spatial resolution for light amplifiers of static images. Here, two different numerical models for short-pulse MCP simulations are suggested [3]. The semi-analytical approach is a powerful tool for the design of static image amplifiers (night vision devices, electron optical converters, streak cameras etc.). Monte Carlo simulations can be successfully used for large area photo detectors with micron and pico-second resolution range. Both approaches have been implemented in the computer codes MCPS [4] and MCS (Monte Carlo Simulator). The results of computer modelling for electric fields and MCP parameters are presented.

ELECTRIC FIELD DISTRIBUTION FOR TILTED CYLINDRICAL CHANNELS IN A DIELECTRIC MEDIUM

The chevron pair is a typical MCP configuration which can prevent ion feedback and increase the efficiency of the first strike problem. It consists of two glass plates with tilted cylindrical pores with a different orientation for each plate. The side surfaces of the plates are metalized, and the voltage V applied to them. The internal surfaces of the pores are coated with a resistive layer and secondary emitter material. Typical dimensions are: plate thickness – 0.5 mm, pore diameter 5-10 μm , coatings – 10 nm, tilt angle – 8° . The field distribution in the pore has a complex structure compared with straight channels, where the electric field vector is parallel to the z -axis, and the field is a 1D one. This tilted field can change the gain factor of secondary emission in the pore. One can show

that the field for the most internal part of the pore can be described analytically, but the fringe fields should be evaluated numerically.

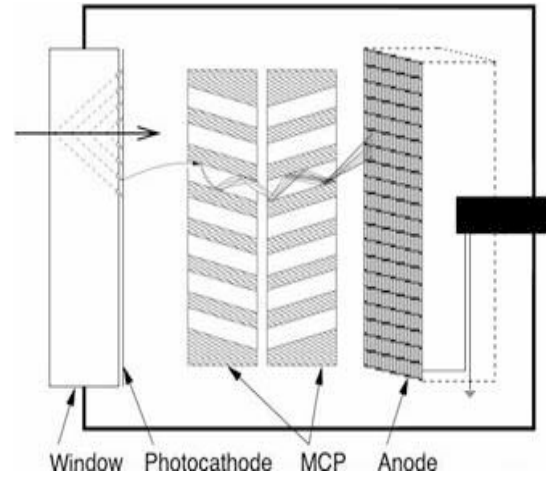


Figure 1: Chevron type MCP.

Problem 1

The potential distribution and electric field for a straight cylindrical pore in a uniform external field (Figure 1) are given by formulae

$$\varphi(\rho, \theta) = \begin{cases} -\frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2} E \rho \cos \theta, & \rho < R, \\ -Ex - E \frac{\varepsilon_1 + \varepsilon_2}{\varepsilon_1 + \varepsilon_2} \frac{R^2}{\rho} \cos \theta, & \rho > R. \end{cases} \quad (1)$$

$$E(\rho, \theta) = \begin{cases} \frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2} E \cos \theta, & \rho < R, \\ E - E \frac{\varepsilon_1 + \varepsilon_2}{\varepsilon_1 + \varepsilon_2} \frac{R^2}{\rho^2} \cos \theta, & \rho > R. \end{cases} \quad (2)$$

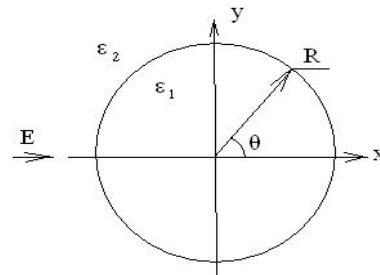


Figure 2: Cylindrical channel of radius R with dielectric constant ε_1 is in the medium ε_2 in an external electric field of strength E .

*Work supported by DoE grant #2005170.

[#]ivanov@muonsinc.com

Problem 2

The external field can be expanded into parallel and perpendicular components

$$E_0 = E_{\parallel} + E_{\perp}, E_{\parallel} = E_0 \cos \alpha, E_{\perp} = E_0 \sin \alpha. \quad (3)$$

The parallel component does not make a perturbation, but the perpendicular one excites the field of Problem 1. The actual field in a cylindrical channel is

$$E_{\perp}(\rho, \theta) = \frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2} E_0 \cos \theta \sin \alpha, \quad (4)$$

$$E_{\parallel} = E_0 \cos \alpha.$$

Finally, the angle between the cylindrical axis and the vector of the internal electric field is

$$\tan \gamma = \frac{E_{\perp}}{E_{\parallel}} = \frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2} E_0 \cos \theta \tan \alpha. \quad (5)$$

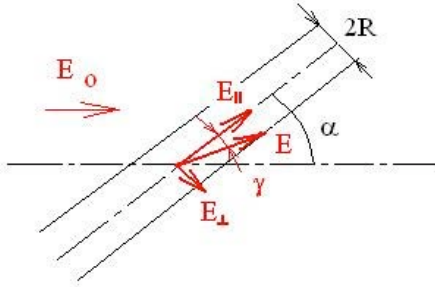


Figure 3: Cylindrical channel in tilted electric field E_0 with angle α .

Problem 3

The presence of conductive material excites surface currents, which increase the electric field from the external side of the layer and decrease it from the internal one – the analogue of a dipole charge layer. It creates the screening effect of lowering for the external field E_{ext} in the pore. The field relaxation ratio for an infinitely thin layer with conductivity σ is

$$k(\sigma) = \frac{E_{int}}{E_{ext}} = \frac{\sigma_0}{\sigma}, \quad (6)$$

$$\sigma_0 = \frac{1}{Z_0} = \sqrt{\frac{\varepsilon_0}{\mu_0}} = \frac{1}{376.73\Omega}.$$

An example of fringe-field numerical simulation is presented in Figure 5. Some simulations of the electric field in a tilted pore were done by E.Gatti [12], but no fringe fields were studied there.

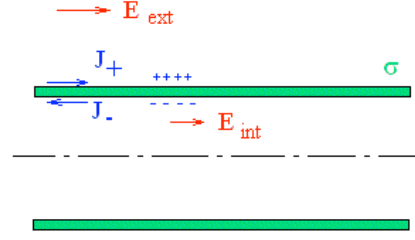


Figure 4: Screening effect of a resistive layer.

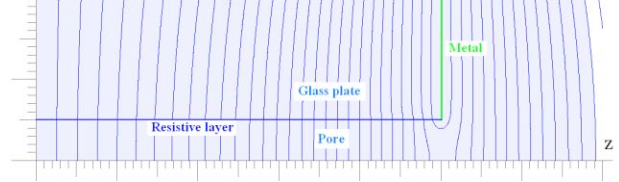


Figure 5: Field map for the fringe field.

DIFFERENT MODELS OF SECONDARY EMISSION

Publication [1] describes a complete model of secondary emission (SE). Here, the energy and angular dependence for true SE is presented by Guest's formula

$$X \equiv \frac{V}{V_{max}} \sqrt{\cos \theta}, \quad (7)$$

$$\sigma = \sigma_m X^{\beta} \exp[\alpha(1 - \cos \theta) + \beta(1 - X)]$$

where θ – incident angle, V – impact energy, V_{max} – impact energy corresponding to a maximum SE yield of σ_m , α – surface absorption factor, β – smoothness factor ($\beta = 0.55$ for $V < V_{max}$, and $\beta = 0.25$ for $V > V_{max}$). Numerical results of the gain evaluation for MCP thickness $L = 0.5\text{mm}$, voltage $V = 2\text{kV}$, $\sigma_m = 3$ using Guest's model are shown in Figure 6.

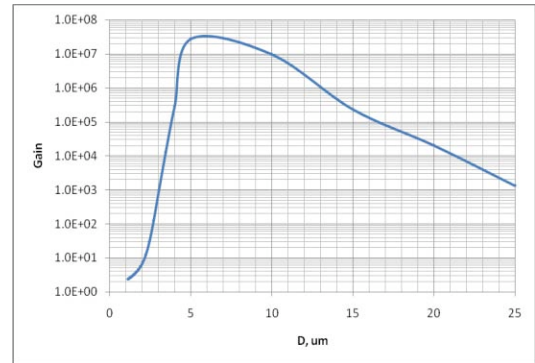


Figure 6: Gain factor vs. MCP diameter D .

The method of MCP simulation [2] was used in modelling of “Planacon” light amplifier (Figure 7) by Burle/Photonics [13]. Our results for simulation (Figures 8-10) show good agreement with experimental data [3], [6]-[7].

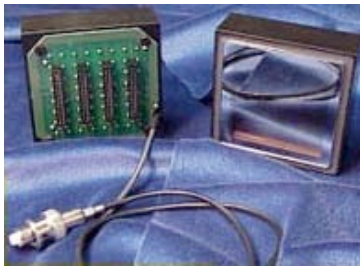


Figure 7. Chevron-type photo multiplier 85022. The MCP is 2"x2" large, there are 1024 anodes, all equally spaced by 1.6 mm. Pore diameter D=25 μ m.

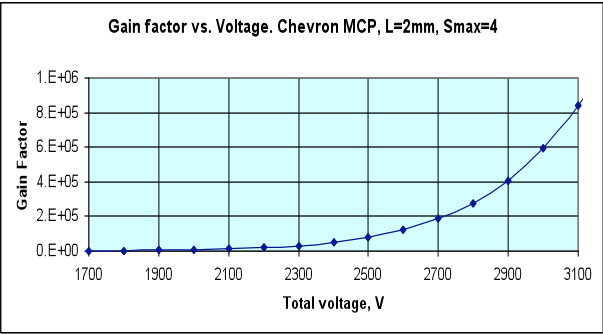


Figure 8. Gain factor vs. MCP voltage.

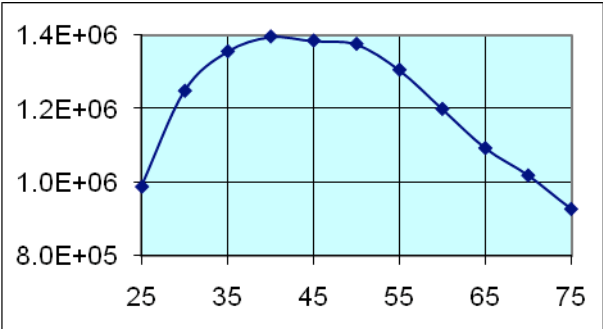


Figure 9. Gain factor vs. ratio L/D. L – length of pore, D – its diameter.

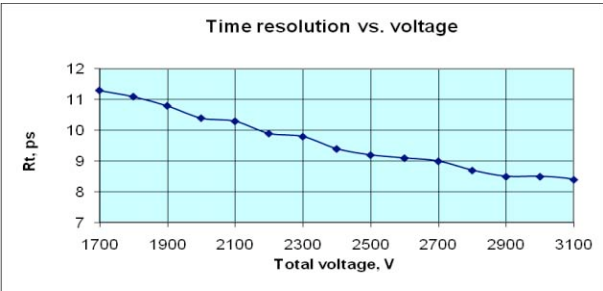


Figure 10. Time resolution including photo cathode – MCP gap,

There are number of other SE models [8-11]. All of them more or less agree each with other for different materials (Figure 11).

Computer Codes (Design, Simulation, Field Calculation)

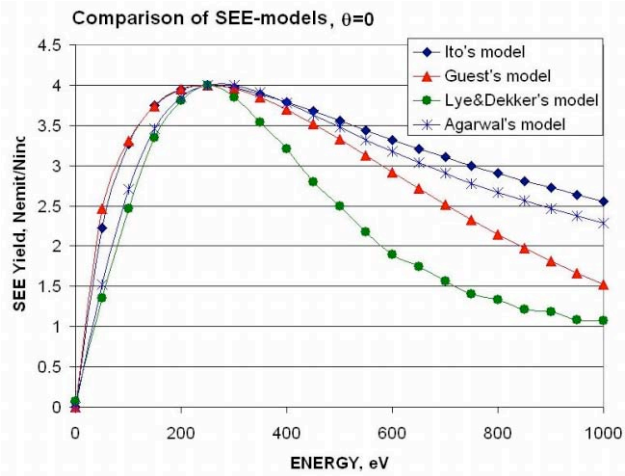


Figure 11: Different models of SE (Courtesy of Z.Insepov).

As for composite materials, these analytical models (Figure 12) do not give good agreement with experimental data.

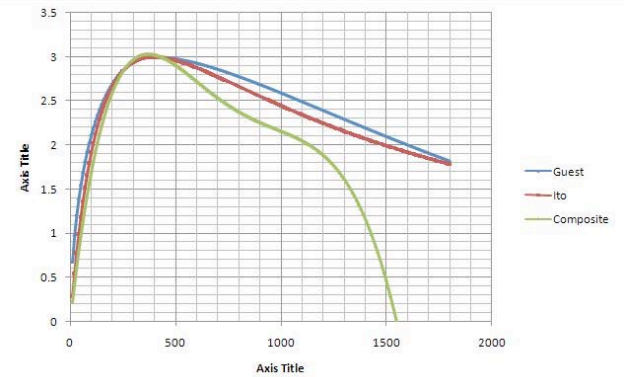


Figure 12: SE curves for a composition of 30% Al₂O₃ + 70% ZnO. Blue – Guest's model; magenta – Ito's models, green – experimental data.

These approximations were used in simulation of the INCOM MCP with parameters: D = 40 μ m, L/D = 40, L = 1.6 mm, voltage U=1kV. The results of numerical simulations of gain and time resolution for different SEE models are presented in Table 1.

Table 1: Comparative analysis of SEE models

Model	Guest	Ito	Composite
Gain	1179	1132	1016
T res., ps	28.3	32.9	26.7

FUNNEL-TYPE MCP SIMULATION

There are ideas for improving the open-area ratio for better determination of the location the first strike, including making the entrance to the channel into a funnel, so that the electron strikes the surface of the

funnel and secondary electrons are sucked into the channel. Alternatively, the funnel could be directly coated with the photo-cathode material, with the photo-electron or electrons then initiating the shower in the channel, as shown in the simulation in Figure 13. The funnel solution is attractive in that it hides the photo-cathode from ion feedback- ions that are created on the channel walls and accelerated back up the channel. These ions are a cause of aging of MCP's, and can be a problem at high gain. Figure 14 demonstrates the result of optimization for photo electron capturing in varying the funnel diameter and photo cathode resistance. The general-purpose electromagnetic and electron optic code "POISSON-2" [14] was used to simulate the two dimensional electrostatic fields and particle tracking.

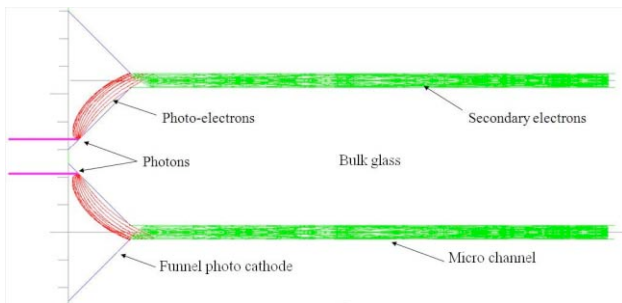


Figure 13: A two of funnel-type cells. Magenta – incoming photon bunches; red – photo electrons; green – cascades of secondary electrons.

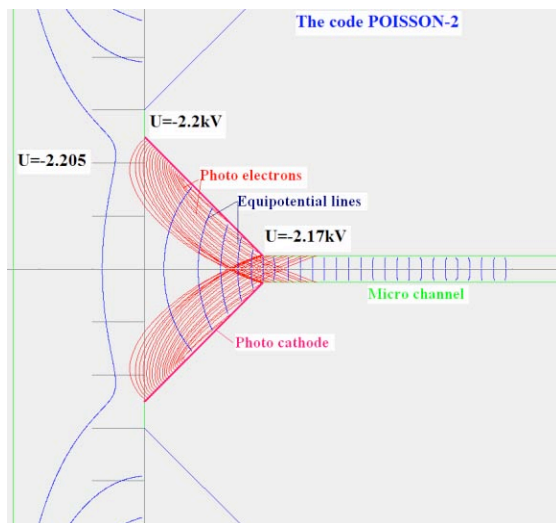


Figure 14: Geometric parameters of the funnel and photo cathode resistance were varied to optimize the capture ratio for photoelectrons.

REFERENCES

- [1] A.J. Guest, ACTA Electronica, V.14, N1, 1971, pp. 73-97.
- [2] Y.V.Kulikov, Radiotekhnika i Svetotekhnika, **4**, 1967, pp.376-380.
- [3] V.Ivanov, T.Roberts, R.Abrams, H.Frish, PAC'09, 4-8 May, Vancouver, Canada.
- [4] V. Ivanov, Micro Channel Plate Simulator, User's Guide, Muons, Inc., 2009.
- [5] V. Ivanov. 20-21 July, 2009, Chicago, 1st Workshop on Photo-cathodes: 300nm-500nm.
- [6] V. Ivanov, Z.Insepov, Pico-Second Workshop VII, February 26-28, 2009; Argonne National Lab.
- [7] V.Ivanov, H.Frish, T.Roberts et al., 27-29 May, 2009, Moscow, IX Seminar on Theoretical & Applied Electron & Ion Optics.
- [8] M.Ito, H.Kume, K.Oba, IEEE Trans. on Nucl. Sci., NS-31, No 1, (1984), pp.408-412.
- [9] B.K.Agarwal, Proc. Phys. Soc. (1958), pp.851-852.
- [10] E.H.Eberhardt, Appl. Optics, **18**, 9 (1979), pp. 1418-1423.
- [11] R.G. Lie, A.J.Dekker, Phys. Rev., Vol.107, N4, 1957.-pp.977-981.
- [12] E.Gatti, K.Oha, P.Rehak, IEEE Trans. On Nucl. Sci., NS-30, No 1, (1983), pp.461-468.
- [13] http://www.burle.com/prod_showcase.htm.
- [14] V. Ivanov. Proc. 2nd Int. Conf. on Computations in Electromagnetism, Nottingham, May 13-15, 1994.

TGE GP V'KO RTQXGO GP V'QHVTCEMPI 'EQF G'BBSIMC

H. J. Kim and T. Sen, Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA

Abstract

The beam-beam simulation code (BBSIMC) is a incoherent multiparticle tracking code for modeling the nonlinear effects arising from beam-beam interactions and the compensation of them using an electromagnetic lens. It implements short range transverse and longitudinal wakefield, dipole noise to mimic emittance growth from gas scattering, beam transfer function, and wire compensation models. In this paper, we report on recent improvements of the BBSIMC including a beam-beam compensation model using a low energy electron beam and a current carrying wire.

INTRODUCITON

A beam-beam simulation code BBSIMC has been developed at FNAL over the past few years to study the effects of the machine nonlinearities and the beam-beam interactions. The code is under continuous development with the emphasis being on including the important details of an accelerator and the ability to reproduce observations in diagnostic devices. At present, the code can be used to calculate tune footprints, dynamic apertures, beam transfer functions, frequency diffusion maps, action diffusion coefficients, emittance growth and beam lifetime. Calculation of the last two quantities over the long time scales of interest is time consuming even with modern computer technology. In order to run efficiently on a multiprocessor system, the resulting model was implemented by using parallel libraries which are MPI (interprocessor Message Passing Interface standard) [?], state-of-the-art parallel solver libraries (Portable, Extensible Toolkit for Scientific Calculation, PETSc) [?], and HDF5 (Hierarchical Data Format) [?]. The following section describes the physical model used in the simulation code. Some applications are presented for the Large Hadron Collider (LHC) wire compensator and the Relativistic Heavy Ion Collider (RHIC) electron lens.

PHYISCAL MODEL

In the collider simulation, the two beams moving in opposite direction are represented by macroparticles of which the charge to mass ratio is that of each beam. Fewer number of macroparticles are chosen than bunch intensity of the beam because it becomes prohibitive for few revolutions around accelerator even with modern supercomputers. They are generated and loaded with an initial distribution for a specific simulation purpose according to the beam parameters at the interaction point, for example, six-dimensional Gaussian distribution for long-term beam evolution. The transverse and longitudinal motion of particles is calculated by transfer maps which consist of linear and Computer Codes (Design, Simulation, Field Calculation)

nonlinear maps. In the simulations, the following nonlinearity is included: head-on and long-range beam-beam interactions, external electromagnetic force by current carrying wire, multipole errors due to quadrupole triplets, and sextupole strength of chromaticity correction. In the following, linear and nonlinear tracking models are described in detail.

Transportation through arc

The six-dimensional accelerator coordinates $\mathbf{x} = (x, x', y, y', z, \delta)^T$ are applied, where x and y are horizontal and vertical coordinates, x' and y' the trajectory slopes of each coordinates, $z = -c\Delta t$ the longitudinal distance from synchrotron particle, and $\delta = \Delta p_z/p_0$ the momentum deviation from the synchrotron. The linear rotation between two elements denoted by i and j can be written as

$$\mathbf{x}_j = \begin{pmatrix} \mathcal{M} & \hat{\mathcal{D}} \\ 0 & \mathcal{L} \end{pmatrix} \mathbf{x}_i. \quad (1)$$

Here, \mathcal{M} is coupled transverse map of *off-momentum* motion defined by $\mathcal{M} = \mathcal{R}_j \hat{\mathcal{M}}_{i \rightarrow j} \mathcal{R}_i^{-1}$, where $\hat{\mathcal{M}}_{i \rightarrow j}$ is the uncoupled linear map described by twiss functions at i and j elements, and the transverse coupling matrix \mathcal{R} is defined as

$$\mathcal{R} = \frac{1}{\sqrt{1+|C|}} \begin{pmatrix} I & C^\dagger \\ -C & I \end{pmatrix}, \quad (2)$$

where C^\dagger is the 2×2 matrix and the symplectic conjugate of the coupling matrix C . The dispersion matrix is described by $\hat{\mathcal{D}} = (0, \mathbf{D})$, and the dispersion vector $\mathbf{D} = (D_x, D_{x'}, D_y, D_{y'})$ is characterized by the transverse dispersion functions and the map \mathcal{M} :

$$\mathbf{D} = \mathbf{D}_j - \mathcal{M} \mathbf{D}_i. \quad (3)$$

\mathcal{L} is a longitudinal map and a nonlinearity of synchrotron oscillations is applied by adding the longitudinal momentum change at rf cavity.

Beam-beam interactions

For head-on and long-range beam-beam interactions, we assume that one beam is strong and is not affected by the other beam while the other beam is weak and experiences a beam-beam force due to the strong beam during the collision, so called weak-strong beam-beam model. Besides, the charge distribution of the strong beam is assumed to be

Gaussian:

$$\rho(x, y, z) = \frac{nq}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right), \quad (4)$$

where n is the number of particles and q is the electric charge of the beam. Note that the coordinates (x, y, z) , denote the rest frame of the strong beam. The beam-beam force between two beams with transverse Gaussian distribution $\rho(x, y) = \int dz \rho(x, y, z)$ is well-known, and the expression for the slope change is given by, for elliptical beam with $\sigma_x > \sigma_y$:

$$\Delta x' = \frac{2\tilde{n}r_0}{\gamma} \frac{\sqrt{\pi}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \Im F(x, y), \quad (5a)$$

$$\Delta y' = \frac{2\tilde{n}r_0}{\gamma} \frac{\sqrt{\pi}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \Re F(x, y), \quad (5b)$$

where $F(x, y)$ is a complex function defined in [?]. New constants are defined as $r_0 \equiv qq_*/4\pi\epsilon_0 m_0 c^2$ and $\tilde{n} \equiv n \left(|\vec{\beta}|^{-1} + |\vec{\beta}_*| \right) / \left(|\vec{\beta}| + |\vec{\beta}_*| \right)$. Here, a subscript asterisk designates a variable of weak beam.

Electromagnetic lens

It is well known that for a large separation distance ($\gg \sigma$) at parasitic crossings, the strength of long-range interactions is inversely proportional to the distance. Its effect on a test beam can be compensated by current carrying wires which create just the $\frac{1}{r}$ field. The advantage of such an approach consists of the simplicity of the method and the possibility to deal with all multipole orders at once. For a finite length l_w embedded in the middle of a drift length L , the transfer map of a wire can be obtained by

$$\mathcal{M}_w^{(L)} = D_{L/2} \circ \mathcal{M}_k^{(L)} \circ D_{L/2}, \quad (6)$$

where $D_{L/2}$ is the drift map with a length $\frac{L}{2}$, and $\mathcal{M}_k^{(L)}$ is the wire kick integrated over a drift length. The change in slopes of a test beam is

$$\begin{pmatrix} \Delta x' \\ \Delta y' \end{pmatrix} = \frac{\mu_0 I_w l_w}{4\pi (B\rho)} \frac{u-v}{x^2 + y^2} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (7)$$

where I_w is the current of wire, u and v are $\sqrt{(\frac{L}{2} + l_w)^2 + x^2 + y^2}$ and $\sqrt{(\frac{L}{2} - l_w)^2 + x^2 + y^2}$ respectively. Besides, we take into account the wire placement including pitch and yaw angles. The transfer map of wire can be written by

$$\mathcal{M}_w = D_{-L/2} \circ S_{\Delta x, \Delta y} \circ T_{\theta_x, \theta_y}^{-1} \circ D_{L_2} \circ \mathcal{M}_k \circ D_{L_1} \circ T_{\theta_x, \theta_y} \circ D_{-L/2}, \quad (8)$$

where T_{θ_x, θ_y} represents the tilt of the coordinate system by horizontal and vertical angles θ_x, θ_y to orient the coordinate system parallel to the wire, and $S_{\Delta x, \Delta y}$ represents a Computer Codes (Design, Simulation, Field Calculation)

shift of the coordinate axes to make the coordinate systems after and before the wire agree. When the wire is parallel to the beam, Eq. (??) becomes $\mathcal{M}_w = \mathcal{M}_k$. For cancelling the long-range beam-beam interactions of the round beam with the wire, one can get the desired wire current and length; the integrated strength of the wire compensator should be commensurate with the integrated current of the beam bunch, i.e., $I_w l_w = cq_* n_*$.

Electron lens

An electron lens is expected to improve beam lifetime and luminosity of the colliding beams by reducing the betatron tune shift and spread from the head-on collisions. A space charge force of low-energy electron beam is acting as a focusing or defocusing lens depending on the high energy bunches. In BBSIMC, two electron beam distribution functions are implemented: (a) Gaussian distribution and (b) Smooth-edge-flat-top (SEFT) distribution. The transverse kick on the high energy beam from the electron beam is given by

$$\Delta \vec{r}' = \frac{2\tilde{n}r_0}{\gamma} \frac{\vec{r}_\perp}{r_\perp^2} \zeta(r_\perp : \bar{\sigma}),$$

where \tilde{n} is the number of electrons of the electron beam adjusted by the electron speed, r_0 is the classic particle radius, $\bar{\sigma}$ is the electron beam size, and γ is the Lorentz factor. The function ζ is given by

- for Gaussian distribution

$$\zeta(r_\perp : \bar{\sigma}) = \left[1 - \exp\left(-\frac{r_\perp^2}{2\bar{\sigma}^2}\right) \right],$$

- for SEFT distribution

$$\zeta = \frac{\sqrt{2}\tilde{\rho}_0}{8} \left[\frac{1}{2} \log\left(\frac{\theta_+^2 + 1}{\theta_-^2 + 1}\right) + \tan^{-1} \theta_+ + \tan^{-1} \theta_- \right],$$

where $\tilde{\rho}$ is a constant, and $\theta_\pm = \sqrt{2} \left(\frac{r}{\bar{\sigma}}\right)^2 \pm 1$.

Finite bunch length

The bunch length effect needs to be considered in case of (1) the longitudinal bunch length σ_z is comparable to the transverse lengths σ_x and σ_y , (2) the orbit function β_x and β_y are not constant through beam-beam interactions, and (3) the transverse beta functions are small and comparable to σ_z . We make slices of both beams moving in opposite directions. Each slice is integrated over its longitudinal boundary, and has only transverse charge distribution at the center of its longitudinal boundary. We take into account the collision between a pair of slices: each slice in a beam interacts with particles in the other beam in turn at the collision points. In addition, electric field energy varies along the bunch due to the inhomogeneity of beam parameters in the longitudinal direction, and couples transverse and longitudinal motions. The coupling can be modelled by including beam-beam interactions due to the longitudinal component of the electric field as well as the transverse components [?].

Particle Loading

At the start of the simulation, the particles in the weak beam are distributed over the phase space. The number of simulation particles is limited by the computational power. In order to make the best use of a relatively small number of simulation particles compared to the bunch intensity, the initial distribution should be optimized. Indeed the initial distribution is very important because a proper choice can reduce the statistical noise in the physical quantities. The simulation particles are generated in two steps: (i) The particle coordinates (x, y, z) of particles can be directly generated from the spatial Gaussian distribution, $\bar{\rho}(x, y, z) = \bar{\rho}_x(x) \bar{\rho}_y(y) \bar{\rho}_z(z)$, where $\bar{\rho}_\zeta(\zeta) = \bar{\rho}_{\zeta 0} \exp\left(-\frac{\zeta^2}{2\sigma_\zeta^2}\right)$. Since the particle coordinates are not correlated, one can generate them by inverse mapping of each cumulative distribution function of horizontal, vertical, and longitudinal Gaussian distributions, using bit-reversed sequence to minimize nonphysical correlations [?]. (ii) An equilibrium distribution in transverse phase space e.g. in the (x, x') plane is $\hat{\rho}(x, x') = \hat{\rho}_0 \exp\left(-\frac{x^2 + (\beta_x x' + \alpha_x x)^2}{2\sigma^2}\right)$. Since the spatial coordinate x is determined at the first step, the slope x' can be obtained from the random variate r of a univariable Gaussian, i.e., $x' = (r - \alpha_x x) / \beta_x$.

APPLICATIONS

LHC

Long-range beam-beam interactions are known to cause emittance growth or beam loss in the Tevatron and are expected to deteriorate beam quality in the LHC. Increasing the crossing angle to reduce their effects has several undesirable effects, the most important of which is a lower luminosity due to the smaller geometric overlap. For the LHC, a wire compensation scheme has been proposed to compensate the long-range interactions by applying external electromagnetic forces. At large beam-beam separation, the electromagnetic force which a beam exerts on individual particles of the other beam is proportional to $\frac{1}{r}$, which can be generated and canceled out by the magnetic field of a current-carrying wire. However, several issues need to be resolved for efficient compensation. With the present bunch spacing, there are about 30 long-range interactions on both sides of an interaction point (IP). The beam-beam separation distance varies from 6.3σ to 12.6σ . The resulting beam-beam force is not identical to that generated by a single or multiple wire(s).

The wire-beam separation distance is one of major parameters which determine the performance of a wire compensator. Figure ?? shows the angle-averaged dynamic aperture for off-momentum particles with $3 \sigma_{\Delta p/p}$ for different wire-beam separations. The reference separation (9σ) is chosen as the average of beam-beam separations. The dynamic aperture calculated at different phase angles is the largest radial amplitude of particles that survive up to a certain time interval; in this simulation, 10^6 turns. When the beam-beam compensation is not present, the dynamic aperture is around 8σ . However, for a wide separation range, the dynamic aperture is smaller than 8σ by about $2-4 \sigma$. The dynamic aperture decreases linearly as the separation decreases. Figure ?? shows the results of particle loss in 1×10^6 turns for different wire-beam separations. The particle loss saturates at large separation while there is a sharp increase of particle loss at small separation. We directly see the minimum particle loss between 0.9 and 1.0 of the reference separation. It reveals that the average of beam-beam separations is close to an optimal separation between the wire and the high energy bunch.

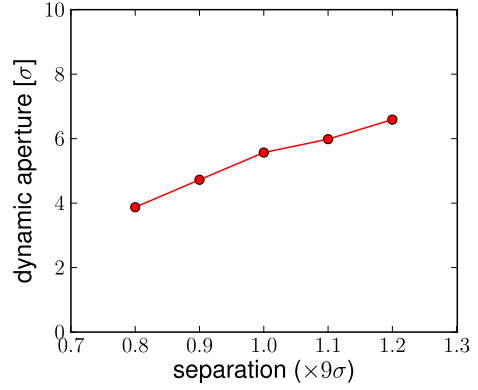


Figure 1: Plot of angle-averaged dynamic apertures according to wire separation distance with wire strength 82.8 Am.

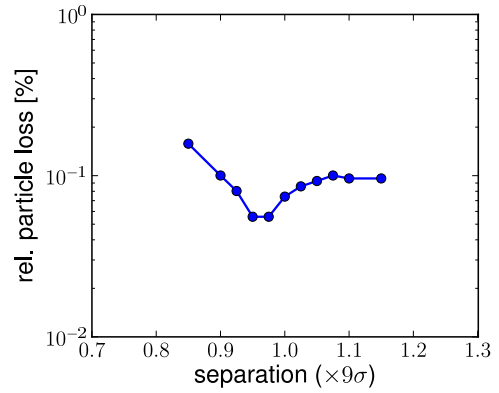


Figure 2: Plot of particle loss according to wire-beam separation distance with wire strength 82.8 Am.

tain time interval; in this simulation, 10^6 turns. When the beam-beam compensation is not present, the dynamic aperture is around 8σ . However, for a wide separation range, the dynamic aperture is smaller than 8σ by about $2-4 \sigma$. The dynamic aperture decreases linearly as the separation decreases. Figure ?? shows the results of particle loss in 1×10^6 turns for different wire-beam separations. The particle loss saturates at large separation while there is a sharp increase of particle loss at small separation. We directly see the minimum particle loss between 0.9 and 1.0 of the reference separation. It reveals that the average of beam-beam separations is close to an optimal separation between the wire and the high energy bunch.

RHIC

Increasing the luminosity requires higher beam intensity and often focusing the beam to smaller sizes at the interaction points. The effects of head-on interactions become even more significant. The head-on interaction introduces a tune spread due to a difference of tune shifts between small and large amplitude particles. In the proton-proton run of RHIC, the maximum beam-beam parameter reached so far is about $\xi = 0.008$. The combination of beam-beam and

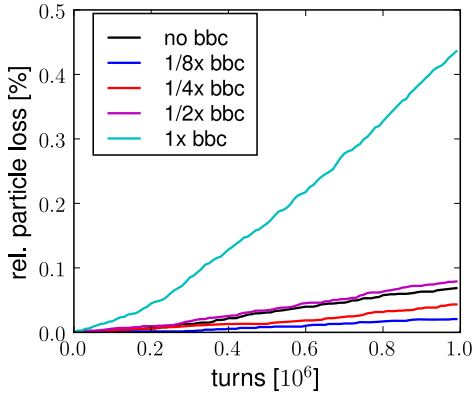


Figure 3: Plot of particle loss according to electron beam intensity for a 1σ Gaussian electron beam profile .

machine nonlinearities excite betatron resonances which diffuse particles into the tail of beam distribution and even beyond the stability boundary. It is therefore important to mitigate the head-on beam-beam effect.

In order to seek the electron lens parameters at which the beam life time is improved, we choose three different electron beam distribution functions: (a) 1σ Gaussian distribution with the same rms beam size as that of the proton beam σ , (b) 2σ Gaussian distribution with rms size twice that of the proton beam, and (c) Smooth-edge-flat-top (SEFT) distribution with an edge around at 4σ . When the electron beam profile matches the proton beam, the full compression of the tune spread requires the electron beam intensity $N_e = 4 \times 10^{11}$ which is defined as the electron beam intensity required for full compensation or $1x$ bbc. Figure ?? shows the results of particle loss for different intensities with the 1σ Gaussian electron beam profile. At an intensity of $1x$ bbc, the particle loss is nearly six times the loss without beam-beam compensation. The beam lifetime of $\frac{1}{2}x$ bbc however is comparable with that of no bbc. As the electron beam intensity is decreased, the particle loss decreases significantly below $\frac{1}{4}x$ bbc, and is reduced to 30% of no bbc at $\frac{1}{8}x$ bbc.

For the 2σ Gaussian and SEFT electron beam profiles, we calculated dynamic apertures and particle loss for different electron beam intensities. The results are summarized in Table ?. The upper limits of the electron beam intensity for the two distributions are chosen so that peak of the electron profile is matched to that of $1x$ bbc at 1σ Gaussian. For the $\frac{1}{2}x$ bbc and $1x$ bbc of 2σ Gaussian profile, there is a small increase in the dynamic aperture of off-momentum particles. There is however a significant reduction in beam loss, for example, below 10% of the particle loss without beam-beam compensation when the electron beam intensity is $\frac{1}{2}x$ bbc. The dynamic aperture obtained with the SEFT profile remains almost the same up to $2x$ bbc. Nevertheless a significant improvement of beam lifetime is also observed below $2x$ bbc. There is a threshold electron beam intensity below which beam life time is increased: $\frac{1}{2}x$ bbc for the 1σ Gaussian, $2x$ bbc for the 2σ Computer Codes (Design, Simulation, Field Calculation)

Profile	Intensity (4×10^{11})	DA (σ)	Particle loss [†] (%)
1σ Gaussian	1	4.48	635
	1/2	5.10	115
	1/4	5.44	63
	1/8	5.63	30
2σ Gaussian	4	3.53	93
	2	5.05	10
	1	5.40	8
	1/2	5.63	6
SEFT	8	3.60	330
	4	4.77	21
	2	5.46	22
	1	5.47	6
	1/2	5.57	6

[†]relative to that without beam-beam compensation

Table 1: Comparison of dynamic apertures and particle loss for different electron beam profiles and intensities.

Gaussian, and $4x$ bbc for the SEFT profile. Particle loss is relatively insensitive to electron lens current variations below threshold current with the 2σ Gaussian and SEFT profiles. This looser tolerance on the allowed variations in electron intensity is likely to be beneficial during experiments.

SUMMARY

In order to study the effects of the machine nonlinearities and the beam-beam interactions, and the effectiveness of compensation schemes of beam-beam interactions, we have developed a six-dimensional weak-strong code BBSIMC. The simulations are carried out using both LHC and RHIC. The results of LHC simulation show that the particle loss is minimized at the wire separation which corresponds to the average of beam-beam separations. We observed, from the results of RHIC, that there is a threshold electron beam intensity below which proton beam life time is increased. A wider electron beam profile than the proton beam at the electron lens location is found to increase beam life time.

REFERENCES

- [1] M. Bassetti and G.A. Erskine, CERN-IRS-TH/80-06, CERN (1980).
- [2] MPI-2: Extensions to the Message-Passing Interface, U. of Tennessee, Knoxville, TN (2003).
- [3] S. Balay, et al, Tech. Rep. ANL-95/11 - Revision 2.3.3, ANL (2008).
- [4] *HDF5 User's Guide*, NCSA, UIUC, Urbana, IL (2005).
- [5] H.J. Kim and T. Sen, in *Proceedings of PAC09*, WE6PFP032.
- [6] K. Hirata, H. Moshhammer, and F. Ruggiero, *Particle Accel.* **40**, 205 (1993).
- [7] C.K. Birdsall and A.B. Langdon, *Plasma Physics via Computer Simulation* (McGraw-Hill, 1985).

A NEW MODEL-INDEPENDENT METHOD FOR OPTIMIZATION OF MACHINE SETTINGS AND ELECTRON BEAM PARAMETERS*

Martin J. Lee, GO AI Services, Los Altos, CA 94024, USA

Jeff Corbett and Juhao Wu, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Abstract

Nonlinear programs are widely employed in particle accelerators and storage rings to compute machine settings for optimal model-predictive control of beam parameters. Conventional iterative methods today suffer from problems with finding the global optimal solution when the start solutions are outside the basin-of-attraction for a given objective function to be minimized. A new iterative matrix inversion global optimization (IMIGO) method [1] has been developed to overcome this limitation. IMIGO unlike the existing iterative nonlinear solvers, it calculates only the Jacobian vector of the objection function and not the Hessian matrix at each iteration-this unique feature has led to a new application of IMIGO for optimization of electron beam parameters for cases when a model is unavailable or only an inaccurate model is available. Some possible applications of this IMIGO-based model-independent optimization method will be presented in the paper.

INTRODUCTION

A nonlinear program is a solver typically employed to find the global minimum of a given objective function subjected to certain conditions known as constraints. For optimization of beam parameters, (b_1, b_2, \dots, b_n) , the control variables are the strengths or settings of a group of accelerator elements, (a_1, a_2, \dots, a_m) , that are used to control these beam parameter values. In general, an objective function is defined as a function of the beam parameters. Since each of the beam parameters is a function of the control variables, the value of a given objective function is determined by the values of the control variables, $f_{\text{obj}}(a_1, a_2, \dots, a_m)$. In beam parameter optimization, the start values of the control variables $(a_1^{\text{start}}, a_2^{\text{start}}, \dots, a_m^{\text{start}})$ are known. Nonlinear programs are used to find the lowest value of a given objective or 'cost' function for the values of the control variables within given bounds: $\Delta_k > (a_k - a_k^{\text{start}}) > -\Delta_k$ for $k = 1, 2, \dots, m$. When the absolute minimum value of the objective function is found, $f_{\text{obj}} \Rightarrow f_{\text{obj}}^{\text{min}}$, the set of variable values is commonly referred to as the global minimum solution: $(a_1^{\text{sol}}, a_2^{\text{sol}}, \dots, a_m^{\text{sol}})$.

The inherent difficulty of using an iterative method to find the global-minimum solution is well known. In general, an iterative method requires an initial guess solution. If this start solution is too far from the global-minimum solution, the program will find only a local-minimum solu-

tion. This problem is known as the basin-of-attraction limit (BOA). A BOA is defined to be the biggest region around a given minimum solution. The problem with the existing iterative nonlinear programs is that they will only find the actual solutions for a special case in which the start solution is inside the BOA corresponding to the global-minimum solution. The new nonlinear programming method IMIGO provides a mitigation to this limitation.

Existing nonlinear programs can be classified into two basic types: One uses an analytical iterative method and the other relies on a stochastic search method such as a genetic algorithm. The inherent difficulty of using an iterative method to find the global-minimum solution is well known. In general, an iterative method requires an initial guess solution. If this start solution is too far from the global-minimum solution, the program will find only a local-minimum solution. As an illustration, a surface plot of the objective function for a minimization problem with two variables is shown in Fig. 1. This figure shows the locations of local-minimum points and the global-minimum point.

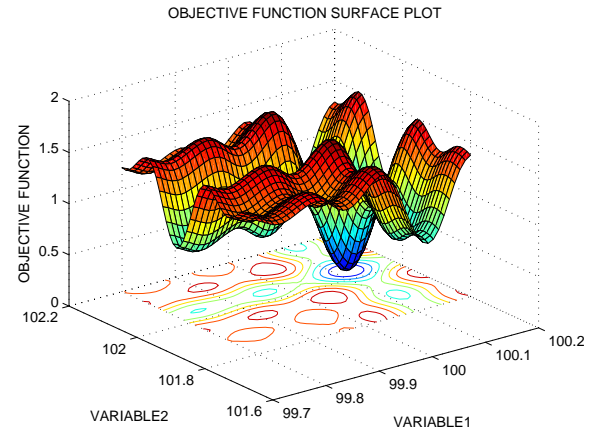


Figure 1: Object function of two variables.

AN OVER VIEW OF IMIGO

IMIGO, like conventional solvers, finds the solution by solving the following set of equations iteratively starting from a given start solution:

$$f_k(a_1, \dots, a_m) = \partial f_{\text{obj}} / \partial a_k = 0 \quad (1)$$

for $k = 1, 2, \dots, m$.

One unique feature of IMIGO is that it solves these equations without the using the values of the derivatives:

*JC and JW's work is supported by the US Department of Energy under contract DE-AC02-76SF00515.

$\partial f_j / \partial a_k = 0$ for $j, k = 1, 2, \dots, m$, *i.e.*, Hessian matrix [1]. In other words, IMIGO is a non-derivative based solver. The iterative process in OASIS is shown in Fig. 2. When the values of the variables converge, the set of values of the variables at the end point is a solution that corresponds to a minimum, maximum, or saddle point of the objective function. This unique feature has led to a simple two parameter (s, p) search method to find the global minimum of the objective function.

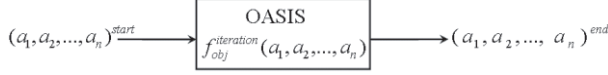


Figure 2: Block diagram of the iterative process to find a solution that minimizes the value of a given objective function.

A main advantage of IMIGO is that it can find the global-minimum solution even when the start solution is not within the BOA corresponding to the global-minimum solution. Another salient feature is that it can search for a path that ends at the global-minimum solution independent of the size and complexity of a given problem, *i.e.*, the problem and the objective function can include many variables and the problem may be very non-linear. To use IMIGO, the user makes a guess on the variable values at the start point for a given objective function to be minimized. The user also imposes specific upper and lower bounds on each of the variables. IMIGO first uses an exhaustive search method to find the start values of the two convergence control parameters.

A Two-Variable Problem

As an illustration of how OASIS works, the results obtained for a typical small-scale minimization problem with two variables a_1 and a_2 are presented [2]. In this example, the same bounds, $\Delta = 0.2$, are imposed on the values of both bounded variables: $\Delta > (a_k - a_k^{\text{start}}) > -\Delta$ for $k = 1, 2$ with $a_1^{\text{start}} = 100$ and $a_2^{\text{start}} = 102$. Figure 3 shows a plot of the objection function for a solution path starting at a given point and ending at the global minimum point. It can be seen from this plot that, because IMIGO is a non-Hessian-based algorithm, the objective function values it finds for points on the solution paths first rise above the objective function value at the start point before falling toward zero at a minimum point: $a_1^{\text{end}} = 100.13$ and $a_2^{\text{end}} = 102.04$. Figure 4 also shows another special feature of OASIS Pathfinder-Its unique ability to find the global-minimum solution when the start point is outside of the BOA of the global-minimum point.

An Eight-Variable Problem

As an application of IMIGO to a real accelerator project, we have used IMIGO to optimize the two bunch compressor setting for the LCLS at SLAC [3]. The objective function is formed to set the final electron rms bunch length, Accelerator/Storage Ring Control Systems

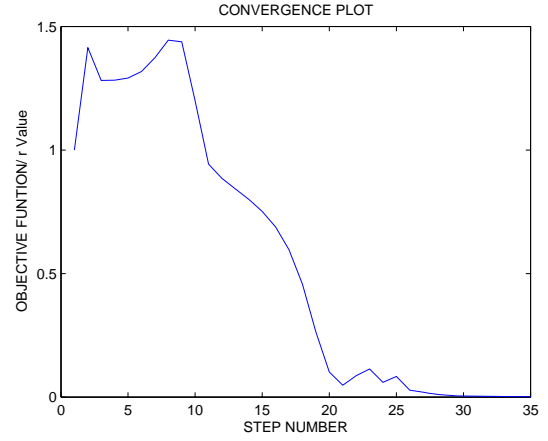


Figure 3: Variation of objective function on solution paths to the global minimum.

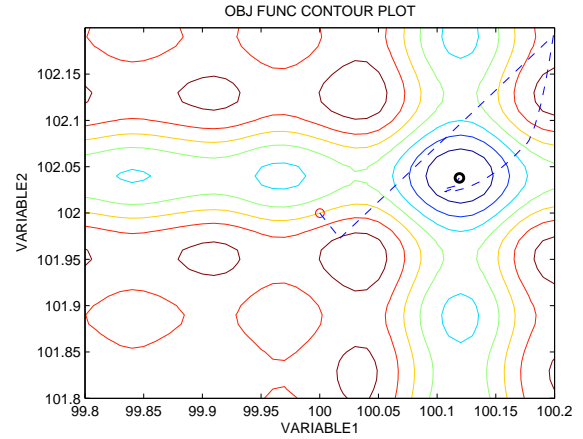


Figure 4: Plot of a solution path that ends at the global minimum point.

the final centroid energy, and the final energy chirp along the electron bunch; and simultaneously minimize the rms bunch length fluctuation and the final energy chirp fluctuation. The objective function is a function of eight variables: the LINAC acceleration phase and total acceleration voltage in the three linac section, and the R_{56} of the two bunch compressors. As an example, the objective function as a function of the two LINAC section (L1 and L2) phases is shown in Fig. 5. However, as described about, even though this is an eight-dimensional optimization, IMIGO in fact does the search in two-dimension, namely in the (s, p) 2-dimensional space. The IMIGO was able to find minimum solution. Yet, the model in Ref. [3] does not include the space charge effect or the coherent synchrotron radiation (CSR) effect. Going into more detailed study with space charge and CSR effects are time consuming with numerical simulation, and is not easy to get a closed analytical expression for the objective function. Hence in the following we will discuss the possibility of using IMIGO to do a model independent beam optimization. The machine is in

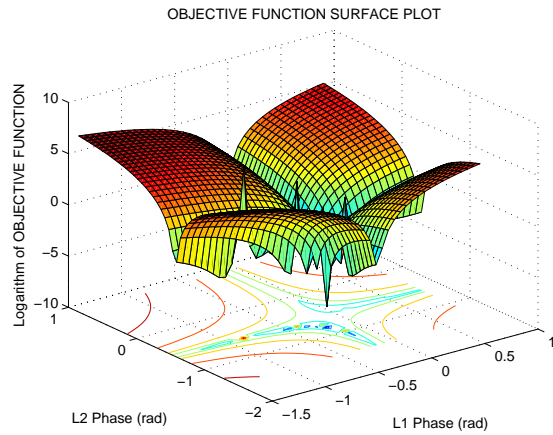


Figure 5: Objective function varies with L1 and L2 phases.

fact the ‘model’, and the measurement data is the output of the ‘model’. Directly working with the machine measurement data, one can also optimize the system by running IMIGO to tell how to set the machine parameters. Hence this is the model-independent approach.

MODEL-INDEPENDENT BEAM OPTIMIZATION

In reality, as described above, normally the model is either too simplified compared to real situation, if an analytical expression for the objective function is needed; or the model can be time-dependent, which will need the control system to be self-adaptive or at least to have a dynamics response function. Therefore, staying with an over-simplified model, or using a static model for a dynamic system will lead to the malfunction of the accelerator system. A model-independent analysis is therefore needed.

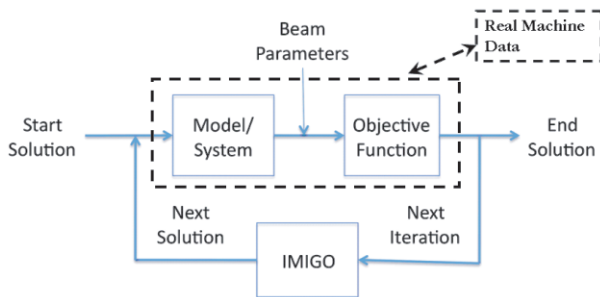


Figure 6: Illustration of the flowchart for model-independent beam optimization.

Shown in Fig. 6, we show the difference between a model-independent optimization and a conventional model-based optimization. In the model-based case, the model is first validated and then the objective function is either constructed analytically or numerically. While in the model-independent case, the real machine measurement data are used to form the objective function.

Accelerator/Storage Ring Control Systems

As emphasized above, since IMIGO is a non-Hessian algorithm, the real measurement data even though will error, can provide objective function accurate enough for IMIGO to find the global minimum. As the follows, we describe a Gedanken experiment to do model-independent beam optimization for a space charge dominated beam. We assume that the machine will deviate from the single particle model significantly, *i.e.*, the space charge is not negligible.

Since the space charge effects are small at low beam current, the objective function predicted from the single-particle model is approximately equal to the value measured on the real accelerator. As the beam current increases, the measured value of the objective function changes. Thus, the size and shape of the BOA of the objective function are different for different beam currents. Since the global-minimum solution is at the ‘bottom’ of the BOA, its value is also different for different beam currents. By replacing the model-predicted objective function with the measured objective function, IMIGO can be used to find the global minimum solution for any desired value of beam current without using the model as shown in Fig. 6. In practice, the optimization process can be carried out incrementally by repeating it over many min-step current changes. As long as the global minimum solution for the previous step is within the BOA of the objective function of the next step, the global minimum solution for the desired beam current will be found.

DISCUSSION

In this paper, we introduce a new nonlinear optimization package, called IMIGO. We demonstrate the power of IMIGO by both working out a detailed challenge example normally associated with the famous Levenberg-Marquardt method. It is shown that IMIGO can find the global minimum even if the start value is out of the BOA of the global minimum. We also demonstrate that IMIGO can solve real accelerator optimization problem as in Ref. [3]. Furthermore, with IMIGO a model-independent beam optimization is possible. Additional research is being conducted to find the global optimal solution for other objective functions such as luminosity and beam lifetime in colliders, and Free Electron Laser peak power and brightness for light sources.

REFERENCES

- [1] M.J. Lee, SLAC-R-758, “A Proposed Global Optimum Algebraic Iterative Solver for Modeling of Lattice Element Errors,” (2009).
- [2] K. Levenberg, “A Method for the Solution of Certain Problems in Least Squares,” *Quart. Appl. Math.* Vol. 2, pp 164-168, 1944; D. Marquardt, “An Algorithm for Least-Squares Estimation of Nonlinear Parameters,” *SIAM J. Appl. Math.* Vol. 11, pp 431-441, 1963.
- [3] M.J. Lee and J. Wu, “Optimal Control of Electron Beam Parameters and Machine Setting With a New Nonlinear Program,” PAC’09, paper ID: FR5REP026.

RF-KICK CAUSED BY THE COUPLERS IN THE ILC ACCELERATION STRUCTURE

A. Lunin, I. Gonin, A. Latina, N. Solyak, and V. Yakovlev, Fermilab, IL 60510, U.S.A.

Abstract

In this paper new results of calculation of the RF kick from the power and HOM couplers of the ILC acceleration structure are presented. The RF kick is calculated by HFSS and CST codes. Special measures allowing the calculation of the effect are described.

INTRODUCTION

The standard 1.3 GHz SC RF cavity of the ILC linac contains 9 cells, an input coupler, and two HOM couplers, upstream and downstream, see Fig. 1.

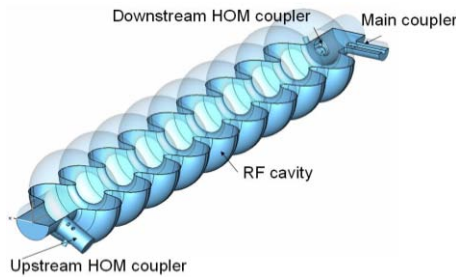


Figure 1: The ILC RF cavity with the main and HOM couplers.

The couplers break the cavity axial symmetry that causes a) main RF field distortion and b) transverse wake field. These effects may cause beam emittance dilution. RF-kick and coupler wake increase with the bunch length [1]. Calculations of the RF kick for the ILC cavity have been performed by different groups, with mismatching results, see Tab. 1.

Table 1: Results of RF-kick calculations.

	FNAL [1] $Q=3.5 \times 10^6$ HFSS	DESY [2] $Q=2.5 \times 10^6$ MAFIA	SLAC [3] $Q=3.5 \times 10^6$ OMEGA3P
$\frac{10^6 V_x}{V_z}$	-105.3+69.8i	-82.1+58.1i	-86.0+60.7i
$\frac{10^6 V_y}{V_z}$	-7.3+11.1i	-9.2+1.8i	-4.6+5.6i

The main reasons of the disagreement are the following: transverse fields caused by the couplers are extremely small (about 5-6 orders of magnitude smaller than the longitudinal fields); cancellation takes place between upstream and downstream coupler. Such characteristics demand for very high precision simulations of the field, better than 10^{-6} . This is a severe challenge for all numerical methods and codes.

GENERAL

In order to achieve reliable estimation for the RF kick, we used the following approaches: (i) different mesh geometry, (ii) different mesh size, (iii) different order of Linear Colliders

finite elements, (iv) different methods of the kick calculations (direct and Panofsky – Wenzel theorem), (v) different number of cells (from $\frac{1}{2}$ cell to entire 9-cell geometry), and (vi) different codes (HFSS and CST).

HFSS code allows the use of a non-uniform mesh. A special three-zone mesh (see Fig. 1) was used in order to improve the field approximation near the axis. Intermediate mesh is necessary to match the fine mesh near the axis and regular mesh in the rest of the cavity.

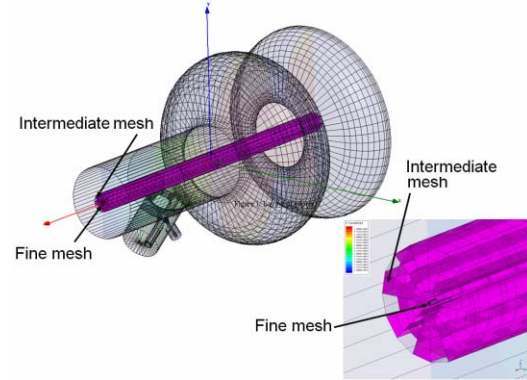


Figure 2: The three-zone mesh for HFSS used in order to improve the field approximation near the axis. Fine mesh repeats the pattern of the intermediate one.

A special symmetric mesh pattern was used in order to reduce the mesh noise. Different techniques of mesh symmetrization were used near the axis. The number of mesh nodes was up to 0.8×10^6 . Cross-check of the direct RF kick were performed applying the Panofsky – Wenzel theorem. Fig. 3 shows the field pattern near the coupler.

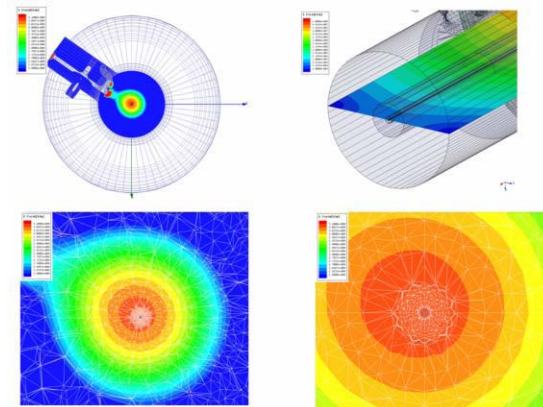


Figure 3: The field pattern near the coupler. The field asymmetry causes RF-kick.

Fig. 4 shows the results of the RF-kick HFSS simulations for upstream and downstream couplers for three cases: different finite element orders, different mesh

numbers and direct calculation and Panofsky – Wenzel (PW) theorem. One can see that convergence takes place for large number of mesh nodes, and that both 1st and 2nd

order elements and direct and PW method give the same results.

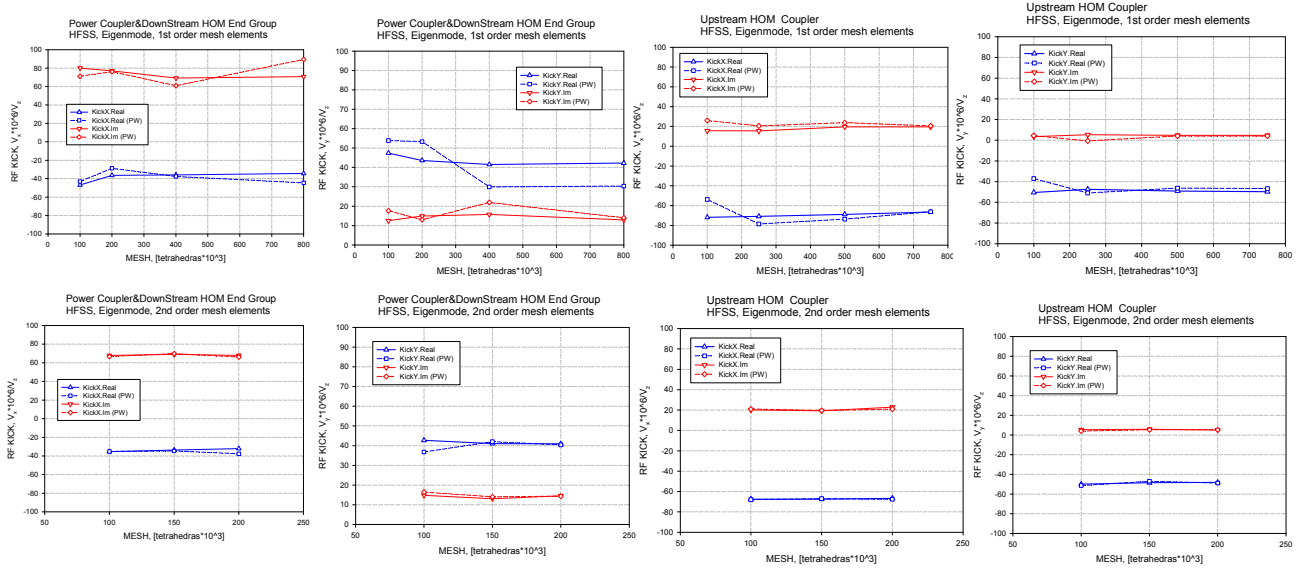


Figure 4: The results of the RF-kick simulations for upstream and downstream couplers for different finite element order, different mesh number and for direct calculation and Panofsky – Wenzel (PW) theorem. One and half cells were calculated.

In the Fig. 5 the results for the RF-kick for upstream coupler are presented for different number of the cells. Calculations were made by both HFSS and CST codes. In the CST case the mesh position was adjusted to the cavity axis in order to achieve symmetry in this region. One can see that the results are very close for all the cases and that the only difference is in the first cell, where the transverse field's components are not completely dumped.

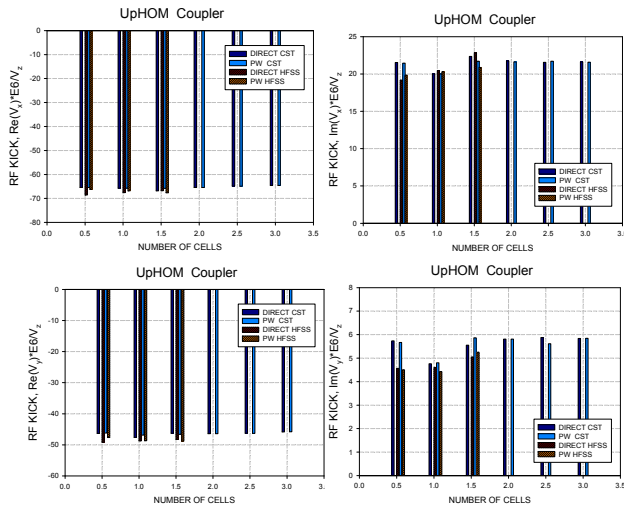


Figure 5: RF-kick for upstream coupler for different number of the cells calculated by both HFSS and CST codes.

In order to cross check the results, the entire structure with the couplers were simulated, see Figure 1. The total RF-kick is close to what was calculated separately, see Tab. 2. Thus, all the results show the same values of the

transverse kick that however differ from the previous results for the vertical kick, that is the most critical for ILC.

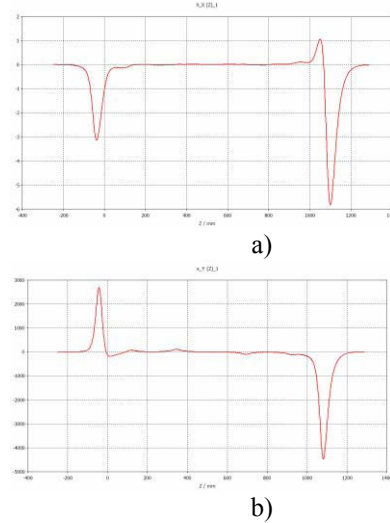


Figure 6: Transverse magnetic field (a) and electric field (b) of the 9-cell ILC structure axis.

The vertical kick ratio to the energy gain per cavity is $(-6.8+18.4i) \times 10^{-6}$ versus $(-7.4+11.1i) \times 10^{-6}$ in our old calculations. However the beam vertical dilution is still small for the main linac. For the bunch compressor the girder optimization technique [4] still allows to compensate the emittance dilution

Table 2. RF-kick calculated separately for Upstream Coupler (a), Downstream End (b) and Total RF-kick (c) of ILC structure .

a)

	NEW FNAL* (HFSS & CST MWS)		OLD FNAL (HFSS)		SLAC (Omega3P)	DESY (Mafia)
	Direct	PW	Direct	PW	Direct	Direct
KickX $\frac{10^6 \cdot V_x}{V_z}$	-62.8+21.1i (CST) -62.7+21.4i (HFSS)	-61.7+20.4i (CST) -63.5+19.6i (HFSS)	-68.8+3.7i	-65.6+7.6i	-57.8+7.0i	-57.1+6.6i
KickY $\frac{10^6 \cdot V_y}{V_z}$	-43.5+5.2i (CST) -45.3+4.7i (HFSS)	-43.8+5.5i (CST) -45.9+4.9i (HFSS)	-48.3-3.4i	-53.1-2.1i	-40.9-3.5i	-41.4-3.5i

b)

	NEW FNAL* (HFSS)		OLD FNAL (HFSS)		SLAC (Omega3P)	DESY (Mafia)
	Direct	PW	Direct	PW	Direct	Direct
KickX $\frac{10^6 \cdot V_x}{V_z}$	-34.8+70.1i	-35.6+69.6i	-36.5+66.1i	-27.3+67.2i	-25.1+51.4i	-25.0+51.5i
KickY $\frac{10^6 \cdot V_y}{V_z}$	41.1+14.1i	42.1+13.1i	41.0+14.5i	40.9+12.8i	36.5+8.9i	32.2+5.2i

c)

	NEW FNAL* (HFSS & CST MWS)		OLD FNAL (HFSS)	SLAC (Omega3P)	DESY (Mafia)
	Direct	PW	Direct	Direct	Direct
KickX $\frac{10^6 \cdot V_x}{V_z}$	-97.5+91.5i (HFSS)	-99.1+89.2i (HFSS)	-105.3+69.8i	-86.0+60.7i	-82.1+58.1
KickY $\frac{10^6 \cdot V_y}{V_z}$	-4.2+18.8i (HFSS) Full ILC Cavity 0.4+20.6i (CST)	-3.8+18.0i (HFSS) Full ILC Cavity 0.1+21.2i (CST)	-7.3+11.1i	-4.6+5.6i	-9.2+1.8

* The End Group effect is taken into account during V_z calculation

REFERENCES

- [1] N. Solyak, et al, EPAC2008, MOPP042
- [2] I. Zagorodnov, and M. Dohlus, LCWS/ILC 2007
- [3] K.L.F. Bane, et al, EPAC2008, TUPP019
- [4] N. Solyak, et al, PAC2009, TU5PFP061

COMPUTATION OF A TWO VARIABLE WAKE FIELD INDUCED BY AN ELECTRON CLOUD *

A. Marković, G. Pöplau, U. van Rienen, Rostock University, Germany

Abstract

The instability of a single positron or proton bunch caused by an electron cloud has been studied using analytical and semi-analytical methods which model the influence of the cloud with the wake field to the bunch. Usually these simulations are fast because the transverse wake due to the electron cloud is being pre-computed and then it is being applied to the bunch turn after turn to simulate the head tail instability. The wake field [1] in these cases is computed in the classical sense as excited electromagnetic field that transversally distorts those parts of the bunch trailing a certain transversal offset in the leading part of the same bunch. The transversal wake force depends only on the longitudinal distance between the leading part of the bunch producing the wake force and the trailing parts of the bunch feeling the wake force. However during the passage of the bunch through the electron cloud the density of the cloud near the beam axis changes rapidly. That means that the environment changes in the time as the bunch proceeds through the cloud and therefore it is not sufficient to apply the single variable (the distance) approximation for the wake field.

In this paper pursuing the idea of K. Ohmi [2] we compute the wake forces numerically as two variable function of the position of the leading part of the bunch and the position of the bunch parts trailing the leading offset in the bunch.

INTRODUCTION

In order to simulate a single bunch instability due to the electron cloud (e-cloud) the bunch movement should be followed turn by turn until the synchrotron tune of the bunch has been resolved, which may take some thousands of turns of the bunch in the ring. At each turn along the ring, the bunch interacts with the e-cloud. A fully self-consistent beam – electron cloud interaction simulation at every turn, even with only one interaction point per turn, would inevitably lead to high computational costs. An idea to speed up the single bunch instability simulation would be to pre-compute the transverse kick of the e-cloud on the bunch. Such a pre-computed kick will be later applied on the bunch at each turn during the tracking of the bunch with the appropriate transport matrices.

Because of the nature of the beam – e-cloud interaction there is a dipole kick on the bunch only if a part of the bunch perturbs the cloud, typically if a slice of the bunch has a slight transversal offset at the entrance in the cloud

of homogeneously distributed electrons. However, if the bunch enters the e-cloud with no parts transversely displaced, it does not perturb the e-cloud asymmetrically. During its passage, it only destroys the homogeneous distribution of the electrons because it attracts them towards the beam axis. As a result, the concentration of electrons near the beam axis grows very fast during the bunch passage. As a matter of fact the electrons near the beam axis start oscillating in the beam potential while the electrons from higher radii are approaching the beam axis and constantly increase the number of electrons near the beam axis. Thus, if the transversal offset in the bunch occurs in the rare part of the bunch the number of electrons on the beam axis which will be perturbed by the beam offset is very high. Consequently the kick from the cloud on the following bunch slices would be expected to be stronger. On the contrary, if the transversal offset occurs in the front part of the bunch then the number of electrons that will be perturbed is not going to be that large and so the expected transverse kick on the following bunch slices would not be as strong as if the electron perturbation happens later during the bunch passage.

From this very simple consideration it is obvious that the pre-computed kick due to the interaction with the e-cloud, would depend on the longitudinal position of the bunch slice with the transverse offset and the longitudinal position of the slice that receives the transversal kick. Hence it is necessary to pre-compute the matrix of kicks from every transversally sliced slab of the bunch to the trailing bunch slabs. The resulting triangular matrix can be used for the single bunch instability tracking.

3D SELF-CONSISTENT PIC SIMULATION

The interaction of two different particle species is being simulated by the particle in cell program MOEVE PIC Tracking [3]. The bunch and the cloud are represented by a 3D distribution of macro-particles in a beam pipe with elliptical cross-section. The macro-particles are defined in the six-dimensional phase space $\Psi(x, p_x, y, p_y, z, p_z)$ and typical values of their number are of order 10^6 for both species. Usually the bunch particles have a Gaussian spatial distribution. The cloud particles are assumed to be homogeneously spreaded in the 3D space bounded by the beam pipe in the transverse plane and with a certain size in longitudinal direction. The interaction is simulated during the bunch passage through the e-cloud. Figure 1 shows the longitudinal profile of the ILC bunch (blue) before it has entered in a 3D homogeneously distributed e-cloud (red).

In this paper the interaction with the electron cloud is

* Work supported by DFG under contract number RI 814/20-1

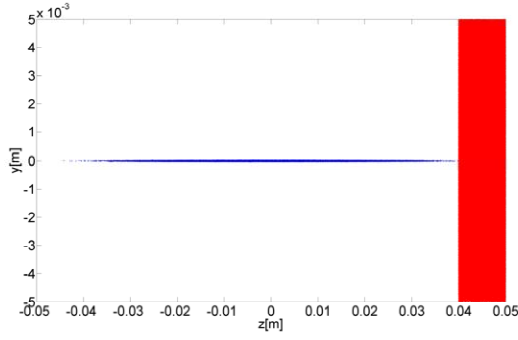


Figure 1: Longitudinal profile of the ILC bunch and the e-cloud, before the bunch enters in the cloud.

simulated for a region without external magnetic field, hence only the fields of the beam and the e-cloud are acting on the both particle species. Beside the strong transversal electrical field \mathbf{E}_b , the beam, being ultra relativistic $\mathbf{v}_b \approx c$, evokes also a strong transversal magnetic field $\mathbf{B}_b = (\mathbf{v}_b \times \mathbf{E}_b)/c^2$. The e-cloud produces only its own space-charge field \mathbf{E}_e . Because of the cancelling of the own magnetic and the electric forces the only force that affects the bunch particles \mathbf{F}_b is the space-charge force of the e-cloud:

$$\mathbf{F}_b = q(\mathbf{E}_b + \mathbf{E}_e + \frac{\mathbf{v}_b \times \mathbf{E}_b}{c^2}).$$

On the other hand, the electrons feel their own and the space-charge forces of the beam:

$$\mathbf{F}_e = q(\mathbf{E}_b + \mathbf{E}_e + \frac{\mathbf{v}_e \times \mathbf{E}_b}{c^2}).$$

Since the electrons are relatively slow (approximately $\mathbf{v}_e \approx 0.01c$) the magnetic force from the beam can be neglected. Consequently in order to integrate the trajectory of the both species it is necessary to compute the electrical fields of the beam \mathbf{E}_b and of the e-cloud \mathbf{E}_e . The time integration of the particle trajectory is computed for a discrete time step with typical values of $dt = 1$ ps. Before the particles are pushed, following algorithmic steps are performed at every time step in the simulation:

- Definition of the 3D laboratory frame grids for the distributions of the bunch and the cloud which is followed by weighting the particle charge on the grid nodes as an input for the discretized Poisson equation.
- The computation of the Poisson equation for the ultra-relativistic bunch takes place on its centre-of-mass system which does not correspond with the laboratory frame grid in which the grid Poisson equation for the e-cloud will be solved. Hence two separate computations of the grid Poisson equation for both species are performed in parallel.
- Interpolation of the grid fields \mathbf{E}_b and \mathbf{E}_e on each particle position in the space and optionally superposition with external fields if present.

Two Stream Instabilities and Collective Processes

Once the forces are computed for every macro-particle of the bunch and the cloud, their trajectories will be all pushed by the leap-frog method for one time step dt further. The total time simulated equals the time the bunch needs to cross the thickness of the defined e-cloud.

In order to evaluate the interaction we first simulate a symmetrical passage of the ILC bunch (parameters given in Table 1) through a homogeneous e-cloud with a density $\rho_e = 10^{12} \text{ m}^{-3}$. The bunch is represented by 10^6 and the 10 mm thick e-cloud slab by $0.5 \cdot 10^6$ macro-particles.

Figure 2 shows a vertical stripe of electrons in the transverse plane gathered during the bunch passage through the initially homogeneous e-cloud. Figure 3 represents the vertical phase space of the initially static electrons after the bunch passage. It displays qualitatively two types of electron motion during the bunch passage. The electrons from the periphery are attracted to the bunch centroid position. The more they approach the beam axis the more energy they win. By the time they reach the beam axis the bunch moved further and these particles continue non-braked until they hit the chamber wall on the opposite side.

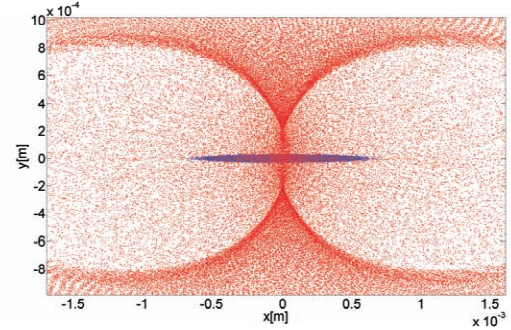


Figure 2: Transversal profile of the ILC bunch (blue dots) and the e-cloud (red dots) after the bunch passes through the cloud.

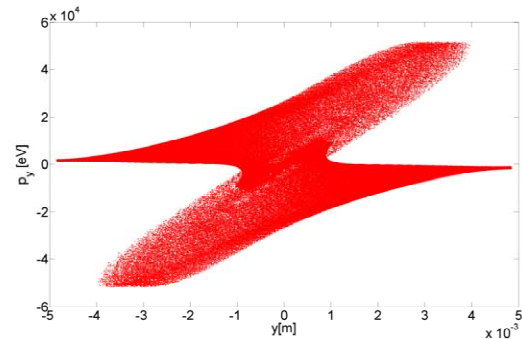


Figure 3: Electron distribution in the transversal (y) phase space after the passage of the ILC bunch.

On the other hand the fine time discretization of the simulation allows us to monitor the oscillations of the electrons of the cloud near the beam axis. First those electrons are attracted in the beam potential. Later as the electron concentration in a very small space grows, the repelling own space-charge force prevails over the beam attractive force

and the electrons would disintegrate. This attraction and repulsion of the electrons to and from the beam centroid position in the transverse plane continues until the whole bunch passes by that certain longitudinal position. The frequency and the amplitude of this oscillations depend mainly on the longitudinal charge profile of the bunch, of course the longer the bunch the higher the number of oscillations will be. An indication of the e-cloud oscillations around the beam axis can be also seen at Figure 4, where the transversal momentum of the bunch particles is displayed in the function of the bunch length. The transversal momentum of the bunch particles at the beginning of the interaction is set to zero. As the bunch passes through the cloud the head particles receive a kick towards higher radii. As the time passes and the electron concentration on the beam axis grows the bunch particles will be kicked towards the beam axis. The envelope of the distribution of the transverse impulse of the bunch particles along its length reflects the e-cloud oscillations. Nevertheless in the case of a symmetrical bunch passage through an initially homogeneous cloud the net transverse force on the bunch over its length remains around zero (red line in Figure 4).

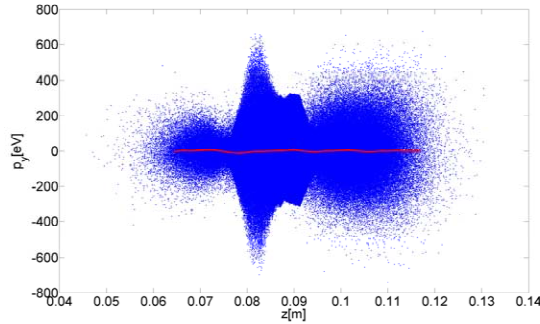


Figure 4: Transversal impulse (p_y) of the ILC bunch particles after the passage through the e-cloud.

Simulation with parts of the bunch offset

The following interaction simulations are done with a bunch which has a 3D slab shifted in the transversal plane. The offset in the y -plane for the examples presented in Figure 5 and 7 is equal to σ_y of the ILC bunch.

Figure 6 displays the transversal impulse of the bunch particles after the passage through an initially homogeneous e-cloud as shown in Figure 1. The red line is the average value of the transversal impulse of the particles at a certain longitudinal position in the bunch. It can be observed that the particles of the offset part got a vertical kick towards a higher radius of the pipe. This is due to the fact that the electrons from the higher radii started moving towards the beam axis attracted from the previous head parts of the bunch. Since this electrons did not had the time to reach the beam axes until the offset part arrived in the e-cloud, the offset part has been attracted towards the approaching electrons from the higher radii.

Two Stream Instabilities and Collective Processes

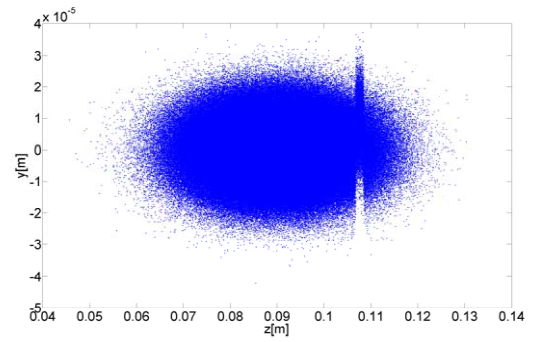


Figure 5: Profile of the ILC bunch with an offset slab in the leading part of the bunch.

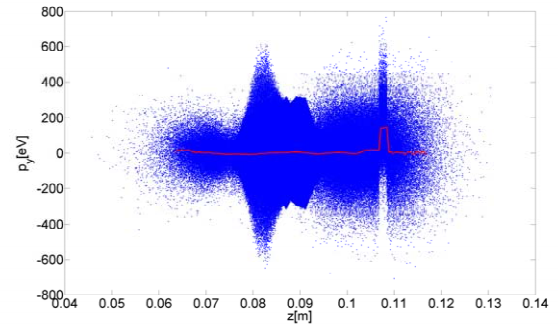


Figure 6: Transversal impulse (p_y) of the ILC bunch particles from Figure 5 after the passage through the e-cloud.

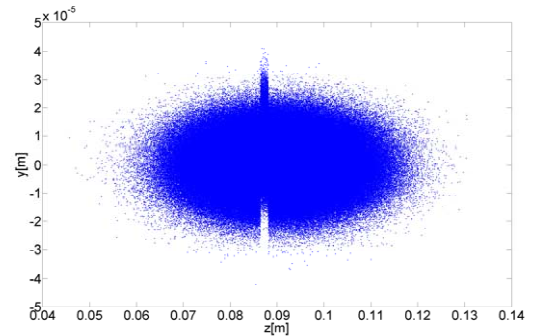


Figure 7: Profile of the ILC bunch with an offset slab in the middle part of the bunch.

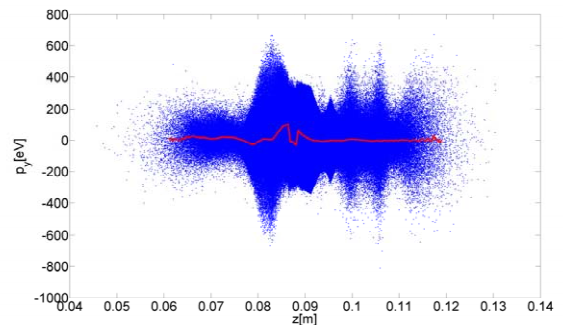


Figure 8: Transversal impulse (p_y) of the ILC bunch particles from Figure 7 after the passage through the e-cloud.

On the other hand, Figure 8 displays the transversal impulse of the bunch for the case that it's offset part is in its middle as shown on the Figure 7. Here by the time the offset part arrives in the e-cloud enough electrons gathered on the beam axis to pull the offset part down towards the beam axis. The offset particles receive an impulse in the negative direction as it can be seen from the trend of the red line at the longitudinal position of the offset part. In the same time the electrons on the beam axis will be perturbed and shifted towards the centroid position of the offset part of the bunch. Thus the following bunch parts will receive an transversal kick in the positive direction (towards the centroid position of the previous offset part). This sort of a head-tail coupling between the bunch part perturbing the cloud and the following bunch parts happens on a very short time scale, due to a very rapid movement of the electrons in the transversal plane around the beam axis. The transversal movement of the beam particles is comparatively slow.

TWO VARIABLE WAKE FIELD INDUCED BY AN ELECTRON CLOUD IN KEKB-LER AND ILC E^+ DR

Parameter	symbol	KEKB-LER	ILC e+ DR
Circumference	L	3016 m	6695 m
Beam energy	E_b	3.5 GeV	5 GeV
Population	Nb	$3.3 \cdot 10^{10}$	$2.0 \cdot 10^{10}$
Charge	Q	5.28 nC	3.22 nC
Length (rms)	σ_z	6 mm	9 mm
Beam size(rms)	σ_x	420 μ m	156 μ m
	σ_y	60 μ m	7.8 μ m
Damping time	$\tau_{x(y)}$	4000 turns	1150 turns

Table 1: Bunch parameters of the low energy ring of the KEK B-factory and the ILC positron damping ring.

The interaction simulations presented in Figure 6 and 8 showed that the transverse kick of the e-cloud on the bunch depends strongly on the longitudinal position of the shifted part of the bunch z_i . In order to compute the kick from a transversally slid slab i on the trailing bunch slabs j ($z_i > z_j$), both bunches (ILC DR and KEKB LER) are longitudinally divided into N_s slabs. Thus we performed a series of N_s interaction simulations. For each simulation a single slab i ($i = 1 \dots N_s$) was shifted by $\Delta y_i = \sigma_y$ and sent through an initially homogeneous e-cloud.

As a result we obtained a matrix with the average dipole kicks $\Delta p_y(j, i)$, from every transversally shifted slab i , on the trailing bunch slabs j ($z_i > z_j$). This matrix is triangular with non-zero entries for $j = i \dots N_s$ and $i = 1 \dots N_s$. The equation of motion for the beam in the y -plane is given by

$$\frac{\Delta p_y(j, i)}{p_b} = \frac{F_y T_t}{p_b} = \frac{\gamma E_e q T_t}{m_o c \gamma} = r_e W_1(z_j, z_i) \Delta y_i N_i. \quad (1)$$

Two Stream Instabilities and Collective Processes

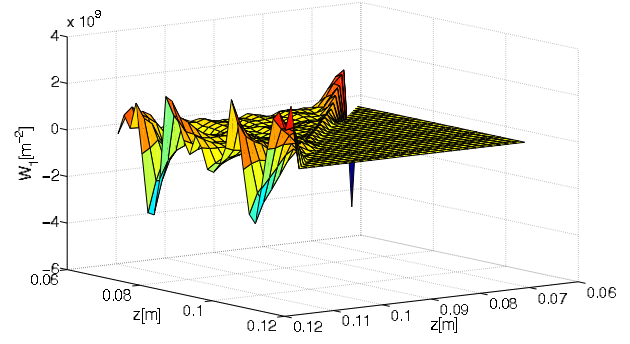


Figure 9: Two variable wake field $W_1(z_j, z_i)$ for the ILC damping ring, $N_s = 50$ slabs.

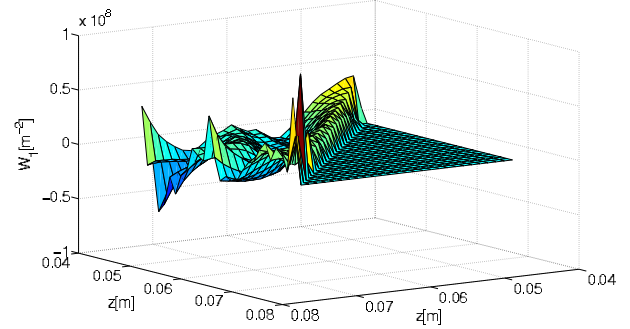


Figure 10: Two variable wake field $W_1(z_j, z_i)$ for the KEKB, $N_s = 50$ slabs.

The computed kick $\Delta p_y(j, i)$ is extrapolated for the total length L of the corresponding ring, thus T_t is the revolution period of the bunch in the ring. On the left hand side the change of the transverse impulse $\Delta p_y(j, i)$ is given relative to longitudinal impulse of the bunch p_b . On the right hand side the classical electron radius r_e is multiplied by the wake function $W_1(z_j, z_i)$ and $\Delta y_i N_i$. The parameter $\Delta y_i N_i$ represents a measure of how many particles N_i are shifted for a Δy_i in the slab i . Note that because of the Gaussian profile of the particle distribution in longitudinal direction N_i varies considerably. Finally the wake function

$$W_1(z_j, z_i) = \frac{\Delta p_y(j)}{p_b r_e \Delta y_i N_i} \quad (2)$$

was computed for both rings as shown in Figure 9 and 10.

Since the computed two variable wake matrix scales linearly with the transversal shift Δy in the next step it could be applied during the tracking simulation with the program of K. Ohmi to investigate the single bunch instability.

REFERENCES

- [1] K. Ohmi, F. Zimmermann and E. Perevedentsev, Phys. Rev. E **65**, 016502 (2001).
- [2] K. Ohmi, private communication, 2007.
- [3] G. Pöplau. *MOEVE: Multigrid Poisson Solver for Non-Equidistant Tensor Product Meshes*. Universität Rostock, 2003.

A HIGH-LEVEL INTERFACE FOR THE ANKA CONTROL SYSTEM*

S. Marsching¹, M. Fitterer¹, S. Hillenbrand¹, N. Hiller¹, A. Hofmann¹, E. Huttel², V. Judin¹,
M. Klein¹, A.-S. Müller^{1,2}, N. J. Smale², K. G. Sonnad¹

¹University of Karlsruhe, Karlsruhe, Germany

²Research Center Karlsruhe, Karlsruhe, Germany

Abstract

ANKA is a synchrotron radiation source located near Karlsruhe, Germany. While the control system has always provided access to technical parameters, like power supply currents or RF frequency, direct access to physical parameters like tune or chromaticity has been missing. Thus the operator has to change and monitor the technical parameters manually and to calculate the physical parameters using separate tools. Therefore effort has been made to integrate the monitoring of physical parameters and simulation tools into the control system. At ANKA the MATLAB-based Accelerator Toolbox is used for simulation purposes, however the control system framework ALMA Common Software (ACS) does not support MATLAB natively. For this reason, a software bridge has been created, which provides direct access to control system components from MATLAB. Thus operators can write their own MATLAB code simultaneously using simulation code and components from the control system. This system has already been used to automate measurements, thus allowing unattended long-term measurements, which have not been possible before. Future plans include creating a graphical user interface and various monitoring and stabilization loops.

ABOUT ANKA

The ANKA facility is a synchrotron radiation facility located near Karlsruhe, Germany. A 2.5 GeV electron storage ring is used to generate synchrotron radiation for various X-ray and IR beamlines. The storage ring can be operated with different optics, in order to provide radiation for different purposes. A low emittance optic is used for normal user operation at a beam energy of 2.5 GeV [1]. A special “low alpha” optic is used to generate coherent THz synchrotron radiation for the IR beamlines in the so called “special user operation” mode [2].

CONTROL SYSTEM

The ANKA control system [3] is based on the software framework ALMA Common Software (ACS) [4, 5]. ACS uses CORBA as a communication link between components. At the moment, components written in C++, Java or Python are supported by the framework. The central ACS manager provides a naming service which is used by clients to find a component by name. The ACS manager

also stores configuration information for components, thus centralizing all configuration information.

The ACS components represent technical components of the storage ring (e.g. power supplies, beam position monitors, RF generators) and are written in C++ and Java. The ACS clients (at ANKA they are all written in Java) provide a GUI for displaying data provided by the components as well as controls for changing the settings of the devices represented by the components. Apart from the lookup and configuration process, the ACS manager is not involved in the communication between clients and components. The architecture of the control system is shown in Fig. 1 (components within the box “Legacy Components”).

Limitations of the legacy control system

The legacy control system is build around the hardware of the accelerator / storage ring. This means that it basically provides remote control and monitoring of all relevant devices, but it is not aware of the physics of the accelerator. The only exception is a software client specially built for performing orbit corrections. However, this client basically is just a special GUI and the code concerning physics is not part of the control system backend components.

As the ANKA storage ring has a flexible lattice and is often operated with different optics (a low-emittance, a low-alpha or a low-beta optics), having a good model for calculating the accelerator settings for different modes of operation is very important. However, without a direct connection between the model and the real accelerator, all changes calculated in the model have to be transferred into technical parameters manually.

Features of the new high-level interface

The new high-level interface allows for an easy integration of new diagnostic components. As the diagnostic components can easily access all control system parameters, task like logging tunes against beam energy can be performed very easily. Even measurements that require the change of machine parameters (e.g. chromaticity measurements) can easily be automated using the new high-level interface as it also allows for write access to most parameters.

ACCELERATOR MODEL

The Accelerator Toolbox for MATLAB (AT) provides tools to create a generic accelerator model [6]. This model

* This work has been supported by the Initiative and Networking Fund of the Helmholtz Association under contract number VH-NG-320.

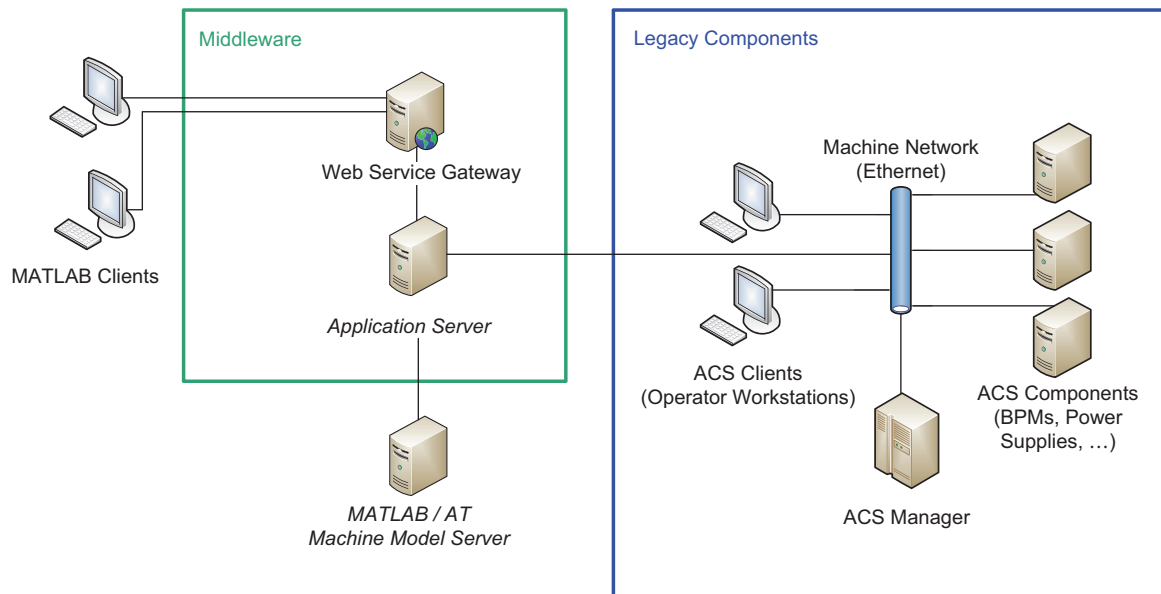


Figure 1: Architecture of the ANKA control system

can be used for many simulation tasks required in accelerator physics. However, real world accelerators always differ from the model because of effects which cannot easily be measured directly or change over time and can hardly be monitored automatically (e.g. small asymmetries, imperfect alignment of magnets, higher order effects in the magnet fields, collective effects).

Therefore, LOCO (Linear Optics from Closed Orbits [7]) is used to fit parameters (i.e. the quadrupole gradients) in the model to match the real accelerator [8]. This is done by measuring an orbit response matrix (change of orbit at beam position monitors against kick at corrector magnets) and then fitting the model to make the model's orbit response matrix match the measured one.

Connecting the model with the control system

While the typical accelerator model parameters are magnet fields or kick angles, the parameters in the control system are magnet currents. Thus a conversion between model parameters and control system parameters is necessary. While an absolute conversion is error-prone, a relative conversion works quite well. Therefore a model that was fitted using a response matrix measured at a similar accelerator state (especially at a similar beam energy) is used as a base. For a magnet the current that was in effect when the model's response matrix was measured is compared to the present current and the difference is calculated. This current difference is then converted to a difference in the magnetic field, which is applied to the model.

This means the following process is used in order to get Accelerator/Storage Ring Control Systems

a model that matches the current state of the accelerator:

1. Load the accelerator model best fitting the current state.
2. Compare magnets' currents in model with currents from control system and calculate difference.
3. Convert current difference to field difference.
4. Apply field difference to model.

The model resulting of this process is a good approximation of the real state of the accelerator. It can be used to predict effects of changes in the magnets' currents or, what is even more important, to calculate changes in the magnets' current needed to attain certain changes in the accelerator optics. The results of these calculations can then be used to apply the changes using the reverse process:

1. Compare magnets' fields with fields from initial model.
2. Convert field differences to current differences.
3. Apply current differences to control system.

By using this process iteratively many small changes can be applied in order to finally attain a big change.

TECHNICAL REALIZATION

As the control system cannot be directly accessed from MATLAB, a "Web Service Gateway" is used. This

software connects to control system components using CORBA and (from the view of the legacy control system) works like a usual ACS client. The gateway exposes the components of the control system using a simplified facade that can be accessed through a web service protocol (at the moment Hessian is used). This allows for a simple, lightweight client which can easily be integrated into MATLAB, using MATLAB's support for integrating Java components.

This approach allows for automation of measurements and integration of simulation methods without having to change or even rewrite the legacy control system. On one hand the complexity of the communication protocol of the control system (CORBA) is hidden from the high-level applications. On the other hand the control system can focus on its main task, the control of many hardware-devices in a reliable and efficient way. Thus the most important goals are attained with minimal effort and virtually all existing components of the control system can be reused by the high-level interface.

EXAMPLE USAGE

The working point of the storage ring has a significant effect on the beam life-time. Therefore it is important to choose a good working point and take care of stabilizing at or near to this working point.

However, changing the gap of a wiggler in the storage ring has major impact on the vertical tune, thus moving the working point (see Fig. 2).

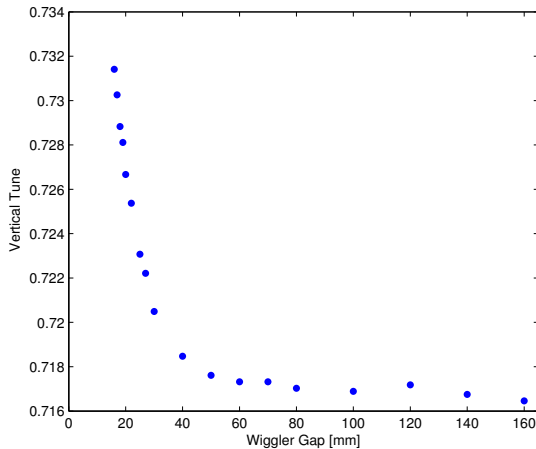


Figure 2: Change of vertical tune against gap of the SUL wiggler in the ANKA storage ring.

The new high-level interface can now be used to calculate changes for the quadrupole magnets in order to compensate the tune drift caused by the wiggler and stabilize at a chosen working point.

A similar problem occurs in the energy ramp: Different working points are chosen for different energies. Besides, the quadrupole power supplies have to be ramped in order

to account for the changed beam energy. This ramp uses a pre-defined table, so the working point is well-known for certain energies. However, between the defined data points the working point not controlled explicitly and therefore might move on a suboptimal path. Furthermore, the values for the ramping table, which is defined in the term of magnet currents, have to be changed manually.

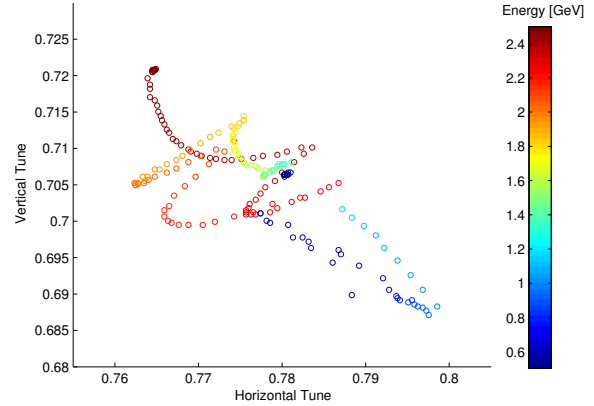


Figure 3: Working point of the storage ring against beam energy.

Fig. 3 shows the horizontal and vertical tune against the beam energy in the ramp from the injection energy of 505 MeV to the final energy of 2.5 GeV. In the future the high-level control system might be used to perform the energy ramp instead of using predefined tables for the magnet currents. This energy ramp algorithm could take care of choosing a working point that is not too close to one of the resonances or other instabilities.

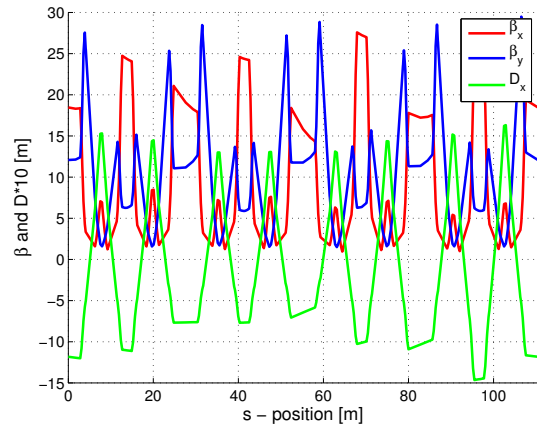


Figure 4: Betatron functions and dispersion along the ring for the low-alpha mode.

BENEFITS FOR THE LOW-ALPHA MODE

The new high-level interface can be used to further develop and investigate the low-alpha optic. For example using the fitted model the optics functions (shown in Fig. 4) can be calculated, taking into account individual gradient errors of each quadrupole.

Another use of the model integration is the calculation of the momentum compaction factor α . The bunch length correlates with α and getting a short bunch length is the main goal of using the low-alpha optic. Therefore a good estimation of α from the model can help in estimating the expected bunch length for a certain optic.

A more detailed description of the low-alpha optic can be found in [2].

FUTURE PLANS

While the present solution is a good start for a more physics-aware control system, the architecture is not yet suitable for a long-term solution. As the accelerator model is integrated within the client applications, the synchronization between model and control system has to be performed for each client individually. Besides there is no abstraction layer between model and client code, thus the client software is tied to the model implementation.

Therefore a Java-based application server will be added to the system. This application server connects the clients to the accelerator model and the legacy control system. Due to the abstraction layer provided by the application server, the actual implementation of the model or the control system could be changed, without having to modify the client software. For example there are plans to try using MAD [9] instead of AT for the accelerator model.

Having only a single component for the accelerator model will also mean, that the synchronization between model and accelerator is performed at a single place and thus different clients are less likely to interfere with each other.

REFERENCES

- [1] E. Huttel *et al.*, "Operation with a Low Emittance Optics at ANKA", PAC'05, Knoxville, May 2005, RPAE037, p. 2467.
 - [2] A.-S. Müller *et al.*, "Far Infrared Coherent Synchrotron Edge Radiation at ANKA", PAC'05, Knoxville, May 2005, RPAE038, p. 2518.
 - [3] I. Verstovsek *et al.*, "ANKA Control System Takes Control", PCaPAC00, Hamburg, October 2000
 - [4] M. Plesko *et al.*, "ACS the Advanced Control System", PCaPAC02, Frascati, October 2002
 - [5] ALMA Common Software website, <http://www.eso.org/~almamgr/AlmaAcs/>
 - [6] A. Terebilo, "Accelerator toolbox for MATLAB", SLAC-PUB-8732, May 2001.
 - [7] J. Safranek *et al.*, "MATLAB-based LOCO", EPAC'02, Paris, June 2002, p. 1184.
- Accelerator/Storage Ring Control Systems

- [8] A.-S. Müller *et al.*, "Linear and Non-Linear Optics Studies in the ANKA Storage Ring", PAC'03, Portland, May 2003, RPP043, p. 3273.
- [9] MAD-X Home Page, <http://mad.web.cern.ch/mad/>

PteqHI DEVELOPMENT AND CODE COMPARING

J. Maus*, R. A. Jameson, A. Schempp, IAP, Frankfurt, Germany

Abstract

For the development of high energy and high duty cycle RFQs accurate particle dynamic simulation tools are important for optimizing designs, especially in high current applications. To describe the external fields in RFQs, the Poisson equation has to be solved taking the boundary conditions into account. In PteqHI this is now done by using a finite difference method on a grid. This method will be described and simulation results will be compared to different RFQ particle dynamic codes.

PTEQHI WITH POISSON SOLVER

PteqHI is a program to simulate particle dynamics in RFQs. It has its roots in PARMTEQ and has continuously been developed and adapted to meet several problems by R. A. Jameson [1]. It describes the external field with the same multipole expansion method than PARMTEQM and it also uses the SCHEFF routine for its space charge calculation, but it uses time as the independent variable. Simulations of a set of 10 RFQs, which are similar to the IFMIF designs in terms of final energy, frequency, emittance, beam current, but with changing aperture have revealed same limitations of these original methods.

This was one of the reason to change the way the electric field is calculated along the RFQ. The new routines solve the Poisson equation directly. This is done by a Solver that uses the finite difference method on a grid.

Generating the grid

The first step is to set up the grid with the boundary conditions figuring out which grid points lie in or on the electrodes. The tip of the electrodes are found using the cell table for the geometry data and interpolating them at each z position using cubic splines. Once the tip position is known the electrode is represented by an arc with a selectable brake out angle. Since the boundary conditions at the electrodes are Dirichlet boundary condition the voltage $\pm \frac{U_0}{2}$ is assigned to the grid points which lie inside the electrodes. In order to describe the surface as smooth as possible the grid points are shifted in such manner that there is always a grid point on the surface. Longitudinal boundary conditions are more difficult to realize, because it cannot be assumed that the structure is symmetric in longitudinal direction. The small changes of the aperture and modulation which disturb the symmetry can be seen in results of

the solver. To overcome this problem many cells are combined to a segment and calculated at the same time. Since the segments overlap, the regions which are influenced by the asymmetry are never used for the dynamic calculation.

Transition Region A real RFQ does not start directly with the electrodes, but with a tank wall. So we decided to let the particles start outside the tank wall, where the potential is close to zero. Then they drift through the small gap between the electrodes and the tank wall seeing the rising RF-field. The first segment of the RFQ includes the tank wall, the gap, the radial matching section and the first two regular cells of the RFQ to be able to simulate the effects of the rising RF-field.

Space charge grid The space charge effect is also calculated by solving the Poisson equation which a charge density $\rho \neq 0$. Therefore a second grid is needed which is generated in the same way as the grid for the external field, but with zero potential on the electrodes. By forcing the potential to be equal to zero on the electrodes the image effect is also taken into account directly, since the purpose of the image effect is to make sure that the potential of a conducting surface vanishes. There is also the option to "turn off" the image effect by setting the boundary condition of the grid to a cylinder (e.g., radius $m \cdot a$) with zero potential on its surface. So the effect of the image charges can be studied.

Poisson solver

For solving the Poisson equation the finite difference method is used. This method is an iterative method where at each iteration step the new value for a grid point is a function of the old value of that point and the values of the neighboring points

$$\varphi_{0,n+1} = F(\varphi_{0,n}, \sum_{i=1}^6 a_i \cdot \varphi_{i,n}, \rho_0), \quad (1)$$

where φ is the potential at the point 0 and ρ_0 is the charge density at that point. Each pair of grid points has a certain distance h_i between them. The a_i are a function of these distances. In general the h_i can vary from one pair of nodes to the next, so that the shifted grid points can be taken into account in order to represent the electrodes correctly without introducing some kind of steps. From one iteration to the next the value at each node converges to the exact answer. The accuracy is limited by the h_i . This basic method is known as the Gauß-Seidel relaxation. For speeding up

* maus@iap.uni-frankfurt.de

the time the solver needs to converge, a successive overrelaxation (SOR) method can be used. The new value for the node is then calculated as a combination of the old value and the value from the neighboring nodes.

$$\varphi_{0,n+1} = \varphi_{0,n} + \omega (\psi - \varphi_{0,n}), \quad (2)$$

where ω , ψ are a fixed relaxation parameter and the combination of neighboring nodes. A further improvement in calculation time can be achieved by increasing the relaxation parameter from 1 to its final value. This is called Chebyshev acceleration. For further details see [2] [3].

SIMULATIONS

For studying the influences of the different simulation methods and different simulation programs a set of 12 RFQs was designed with the same design strategy but with different values for the minimum aperture. That leads to a set of RFQs with different performances. Some have an aperture which is too small (cases with a high a-factor) and therefore a bad transmission. Other RFQs have quite big aperture, but a bad ability to catch the beam longitudinally (small a-factor). Overall, RFQs with a high or a low aperture have a bad performance and in the middle there is optimum value for the aperture. The question is, does that optimum depend on the simulation method which has been used?

Figure 1 and 2 show the external field for the synchronous

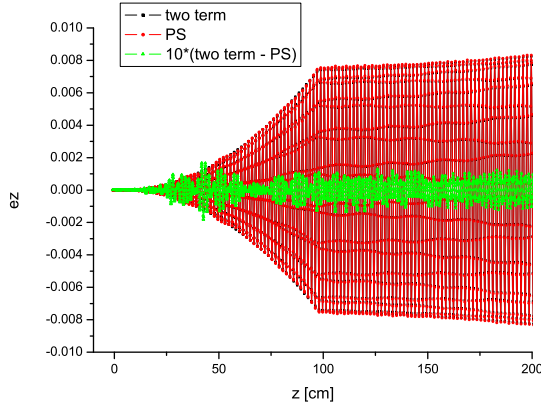


Figure 1: Two Term field and field from the Poisson Solver in the beginning of the RFQ for the synchronous particle. Green curve is 10 times the difference of the two methods.

particle at the beginning and for the entire structure for one of the RFQs with a high transmission. The black curve relates to the field from the two term potential and the red curve to the field found from the Poisson solver. The oscillation comes from the sinusoidal RF. The green curve shows ten times the difference of the other two curves. The differences appear mostly in the beginning of the RFQ and at the end where the amplitude of the field has increased. The major discrepancy is in the region closer Linear Accelerators

to the electrodes where the Two-term-potential assumes a pure quadrupole shape which can differ a lot from the real geometry.

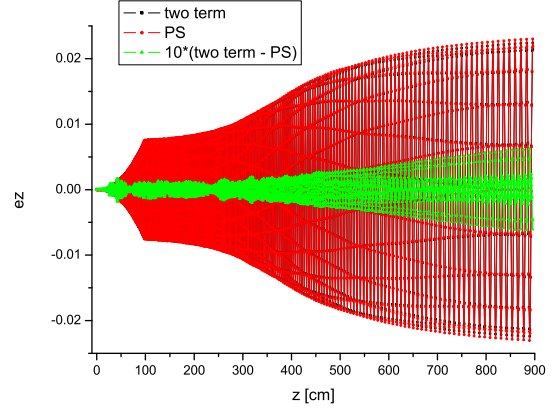


Figure 2: Two Term field and field from the Poisson Solver for the entire RFQ for the synchronous particle. Green curve is 10 times the difference of the two methods.

Space charge and image effect

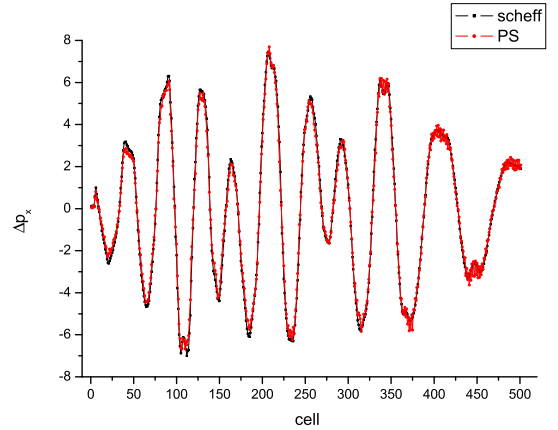


Figure 3: Momentum Δp_x for a specific particle, comparing SCHEFF and Poisson space charge solver with image effect.

The basic routine for space charge calculation in pteqHI as well as PARMTEQM [4] is SCHEFF routine. It is a two dimensional routine assuming cylindric symmetry. So it is normally called once per cell when the shape of the beam is round. There are transformations for a non-cylindrical beam. SCHEFF represents the beam by charged rings and calculates analytically the effect of charged rings on rings. Figure 3 and 4 show the resulting change of the momentum (Δp_x and Δp_z) for a specific particle over the entire RFQ

with image effect turned on and figures 5 and 6 with the image effect turned off. The Poisson solver was used to drive the space charge calculation and SCHEFF ran passively to get the corresponding momentums. The basic shape of the curves are very similar, but when the particle has left the center of the beam and the amplitude of the space charge effect increases a difference between the two curves can be seen in any case.

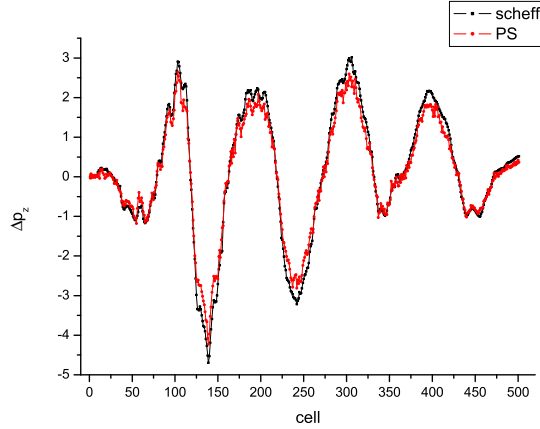


Figure 4: Momentum Δp_z for a specific particle, comparing SCHEFF and Poisson space charge solver with image effect.

In figure 6 it can be seen that the amplitude of the longitudinal momentum change from the Poisson solver with image effect turned off is lower than the with image effect on (Fig. 4). This is due to the fact that the potential was forced to be zero on the cylinder with the radius of $m \cdot a$, but the transverse results are very similar (Figs. 3 and 5).

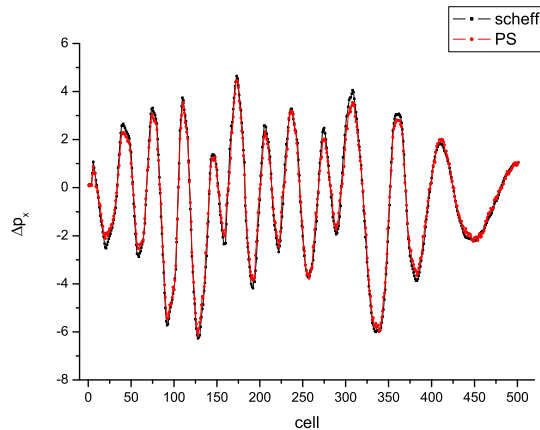


Figure 5: Momentum Δp_x for a specific particle, comparing SCHEFF and Poisson space charge solver with image effect turned off.

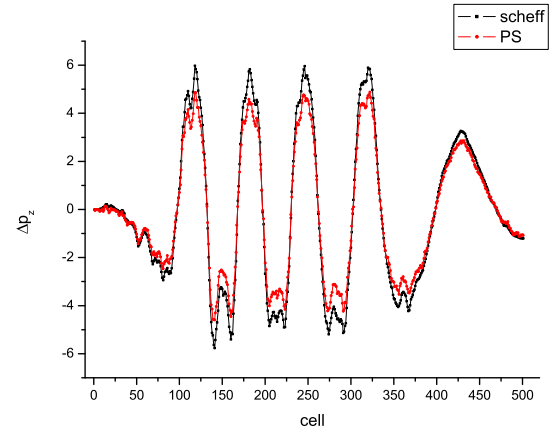


Figure 6: Momentum Δp_z for a specific particle, comparing SCHEFF and Poisson space charge solver with image effect turned off.

Results

Figure 7 and 8 show the transmission and the percentage of accelerated beam for the set of RFQs. All curves have in common, that they have a peak for a certain RFQ. The transmission and the percentage of accelerated beam differ for the RFQs with a small a-factor and therefore with a rather big aperture. Once the aperture is small enough, all particles get accelerated or radically lost. The black curve is the standard pteqHI using the multipole expansion method for the external field and SCHEFF for the space charge routine without the image effect. Its peak is at a a-factor of 41 to 48 and then it falls off in both directions. For the red curve the external field was described using the Poisson solver and SCHEFF was used for the space charge calculation. The curve is similar to the black one with the same peak, but it does not fall off as fast on the right side as the multipole expansion method curve. When the Poisson solver (with image effect turned off) is used for the space charge calculation as well (green curve), the shape of the curve and the position of its optimum change. The peak of the curve is shifted to the left to wider apertures. But the value for the maximum transmission is nearly the same.

This suggests that the treatment of the space charge effect has big influence on the simulation and should therefore be the focus for further studies.

Image charge

As shown above the inclusion of the image charge has an effect on the corresponding change of momentum. In Fig. 9 the effect of the image charge is shown. For a big a-factor the image charge will cause the transmission to drop down (small aperture). For the optimum of the curve the transmission increases with the image charge and on the left hand side it just has a minor effect. The image charge becomes important when particles come close to the

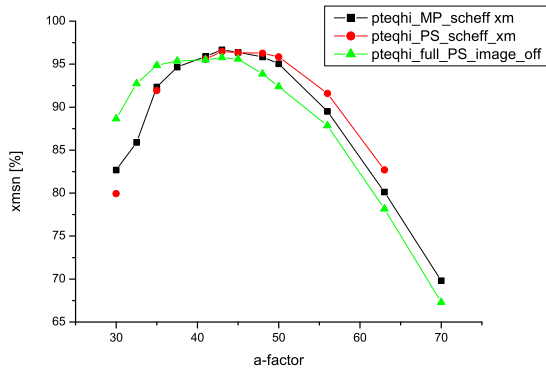


Figure 7: Transmission for the set of RFQs with different simulation methods

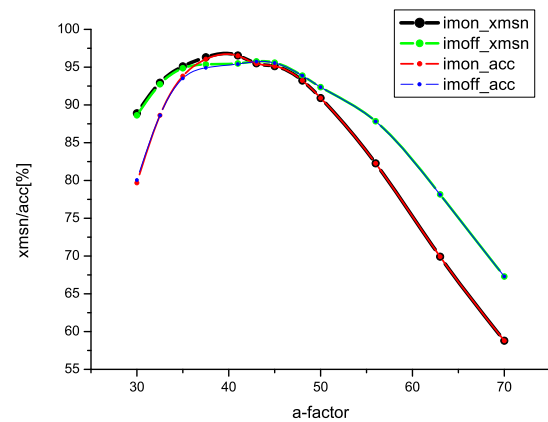


Figure 9: Effect of the image charge on transmission and accelerated particles.

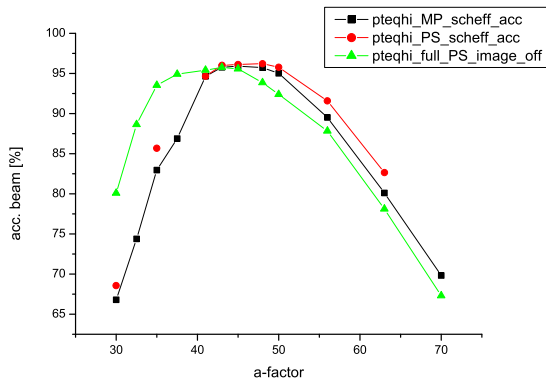


Figure 8: Percentage of accelerated beam for the set of RFQs with different simulation methods

electrodes. This happens more easily when the aperture is small. That can explain, why the transmission drops on the small aperture side. When the aperture is wide enough and the core of the particles stays away from the electrodes the image effect should not make a big difference.

COMPARISON WITH OTHER CODES

Comparing different codes has been a tedious task. One has to make sure, that the different programs simulated the same RFQ with the best match. Some codes start directly with the electrodes while others start the beam inside the tank wall. There were some hidden tricks one needs to know to make the program do what it is supposed to do. Often one must deal with a “black box” with sometimes no support from the coder. That was one major reason for writing an open-source Poisson solver where it is known, what is inside and what decisions have been made.

Linear Accelerators

CONCLUSION

For detailed beam dynamic simulation of high current, high power and high energy applications, accurate simulation tools are needed that treat the involved physics correctly and use as few assumptions as possible. One step in that direction has been made by replacing the old multipole expansion method, the SCHEFF routine and the image charge routine with a 3D Poisson solver. It has been shown, what the effects of the routines are and how they change the results of the simulation.

REFERENCES

- [1] R.A. Jameson, “LA-UR-07-0876”, Los Alamos National Laboratory
- [2] R.W. Hockney and J.W. Eastwood, “Computer Simulation Using Particles”, IOP Publishing Ltd, Bristol and Philadelphia, 1988
- [3] U. Trottenberg and C.W. Oosterlee and A. Schüller, “Multigrid”, Academic Press, London, 2001
- [4] K. R. Crandall and T. P. Wangler and L. M. Young and J. H. Billen and G. H. Neuschaefer and D. L. Schrage, “RFQ Design Codes, Los Alamos National Laboratory, 2005

TRACY#

H. Nishimura

Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.

Abstract

Tracy# is a C# class library for single particle beam dynamics in full 6-dim canonical phase space. The code is based on Goemon that is a C++ version of the Tracy2 library. This paper describes the new features in Tracy# from a software engineering aspect.

INTRODUCTION

During the ALS[1] design phase, Tracy[2] was developed in Pascal for modeling and tracking studies by using embedded Pascal compiler/interpreter to parse the user logic. It evolved to a 6-dimensional version called Tracy2[3]. Subsequently its accelerator physics library was separated from Tracy and ported to C/C++ independently at multiple light sources, including SLS[4], Diamond[5], Soleil[6] and NSLS-2[7], mostly for model-based accelerator controls. At the ALS, the library was re-written in C++ as Goemon[8], which has now been re-written in C# [9] and called Tracy#.

FEATURES

Library Layers

Tracy# is a library that implements single particle beam dynamics for modeling, simulation and analysis studies.

An application built with Tracy# has 3 layers:

- Physics layer, which is Tracy# itself.
- Accelerator layer to model particular accelerators.
- Application layer.

The Physics layer is for beam dynamics and uses the following integrators:

- The 4x5 linear matrix formalism.
- The 2nd and the 4th-order symplectic integrators[10] in 6-dim.
- K-pot Hamiltonian[11] that models small rings properly.

In Tracy# these integrators were implemented in C# taking advantage of some of the important language features. For example, operator overloading is used for:

- Vector and matrix calculations.
- Differential algebra (DA) [12].
- Lattice definitions.

The second layer is to model particular accelerator structures. This is the place where virtual accelerators are

built in forms of C# classes. Starting with an ideal lattice, a virtual machine is enhanced to be a practical one that provides realistic error emulations, various control knobs, and customized physics routines to calculate a range of quantities including dynamic apertures.

The application layer is for client application programs that access the virtual accelerators. These programs can use any features of the .NET libraries that cover database access, XML, networks and graphics. Tracy# is designed to be compatible with these standard .NET libraries. Actually, Tracy# needs only one external routine that is for the singular value decomposition to invert the large sensitivity matrices. We chose an open-source math library in C#[13].

Environment

Tracy# is built on the .NET Framework 3.5, and works on Windows XP SP2 as well as Vista. The development environment is Visual Studio 2008. The programming language for the physics layer is C# 3.0. The client programs are also developed in the same development environment and usually in C# 3.0. However, they can be in other .NET languages, such as Visual Basic.NET or IronPython[14] that we mention later.

Tracy# can also work non-Windows platform by using MONO[15] that is an independent, open-source implementation of the .NET Framework that covers various platforms, including Linux, Solaris and Macintosh. As Tracy# itself is a plain C# code, it is compatible with MONO. It is also possible to make its application program portable by carefully choosing the graphics components for GUI programming.

Implementation

As mentioned, Tracy# uses advanced features of the .NET Framework. In particular, .NET generics proved invaluable especially List< >. In case of Goemon in C++, we did not use C++ generics called template to keep the compile-and-link time reasonably short.

Porting the routines for vector/matrix and DA was not trivial. Goemon used local variables allocated on the memory stack to carry out calculations rapidly. This trick in C/C++ does not work with C#. Therefore the C# routines needed fine tuning to restore execution speed[9]. Currently, we are rewriting routines further to enable parallel computing as mentioned later.

Graphics is not a core part of Tracy#. Instead, the accelerator layer uses it extensively. The graphics programming was in WinForm of .NET 2.0, and is migrating to Windows Presentation Foundation (WPF) of .NET 3.0.

Tracy# also uses relational databases and XML as described in the following sections.

*Work supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DATABASE

There has been an increasing need for data management to track lattice configurations, simulation conditions, and the results of calculations with proper context. Tracy# makes use of standard SQL and XML for these management tasks. Both technologies are well supported on .NET.

Currently, Tracy# uses ADO.NET 2.0 for database access. Its DataSet component is a kind of on-memory database embedded in a client program. Its external data source can be almost any major SQL database including MySQL.

At the ALS MySQL is the most common database system. It is used by controls applications including various Matlab programs dealing with automated machine control[16]. Tracy# reads tables produced by these programs to retrieve the current operational reference values. For example, there is a table for the measured sensitivity matrix produced by a Matlab program. This is the table most referred by Tracy#.

Tracy# usually restores the data of a database table to an ADO.NET DataTable object in memory, and then they can be stored in other database, or to XML. This part is not limited to any particular database system.

XML

Tracy# uses XML extensively to manage its own data through a new API called LINQ to XML[17]. The merit of using XML rather than SQL is the flexibility to support complex data structures. One XML file can replace multiple SQL data tables and relations among them. Although XML may look complicated, it simplifies the data management significantly. This is especially true with LINQ to XML as it provides an API that is much simpler than other XML APIs like DOM or SAX. For example, an XML node can be created independently from the XML document, and later it can be inserted to another XML node that may already be in the XML document. Therefore, we can even use an XElement object like an ordinary variable. Here XElement is a new class of LINQ to XML that supports XML nodes.

Currently, Tracy# uses XML for the following functions:

- Lattice description
- Algorithm description
- Data storage

Lattice Definition in XML

In constructing a virtual accelerator using Tracy#, a lattice structure is typically defined in the creator of its custom ring class by using operator overloading like: SEC=SYM+L1+QF+L2+BL2 where SEC is a new beam line object and the right-hand side is a series of either single elements, such as magnets, or lines of them. So the lattice is defined in C# which allows the lattice to be defined in a very detailed fashion.

There is also a complementary way of defining a simple lattice in a descriptive manner. For example:

```
<Lattice Name="BR0">
  <Element Class="Quad" Name="QF" L="0.15"
    K="2.62488449385119" />
  <Element Class="Quad" Name="QD" L="0.1"
    K="-2.44458454557616" />
  <Element Class="Bend" Name="B" L="1.05"
    K="0" T="15" T1="7.5" T2="7.5" />
  <Element Class="Drift" Name="L1" L="0.5469" />
  <Element Class="Drift" Name="L2" L="0.4969" />
  <Element Class="Drift" Name="L3" L="2.0938" />
  <Eline Name="S1" Line="QD, L1,B,L2,QF"/>
  <Eline Name="S2" Line="QF,L3, QD"/>
  <Eline Name="S3" Line="QD,L1,B,L2,QF"/>
  <Eline Name="SGA" Line="S1,S2,S3,-S1"/>
  <Element Class="Marker" Name="SYM"/>
  <Eline Name="SEC" Line="SYM,SGA,-SGA" />
</Lattice>
```

Tracy# accepts such a lattice defined in XML that follows the same convention of defining a lattice in Tracy2 and Goemon. Without writing a complex lattice parser, or defining a special grammar, we can use XML to define a lattice as a data input to Tracy# at runtime.

XML can also be used to define a very complex lattice. We had such scenario with the radiation safety studies for top-off injection at the ALS[18]. This was a tracking study far off the reference orbit. The horizontal offset can be over 50 cm which is enough to bypass the bending magnet (the vacuum chamber has an ante-chamber that allows this path). Therefore magnetic field profiles and the chamber apertures must be included in the lattice. We created a special version of Tracy# for this study. Transferring the lattice in XML, extra nodes were added to assign the magnet field profiles and chamber geometries. XML was flexible enough to accommodate complexities of the lattice definitions.

Simulation Logic in XML

Similar to the case of a simple lattice definition in XML, simple simulation logic can be described in XML. Here is an example of initializing the ideal lattice of the ALS Booster Ring that uses the lattice just defined above.

```
<Ring Name="BR0" Lattice="BR0" ELine="SEC"
  Symmetry="4">
  <Energy Value="1.9E9" />
  <GetTwiss Value="1" />
  <SetQforTune QF="QF" QD="QD" />
  <FitTune Nux="5.25" Nuy="2.75" />
</Ring>
```

This example shows that XML can describe simple logic in a reasonable manner.

Note that XML is used for both lattice definitions and algorithms, which means that they can be in the same XML document. Currently, we have very simple parsers

that use LINQ to XML, and add to their functionality whenever required.

XML also served an important role for providing very complex logic as in the case of the top-off radiation safety studies. There was no established methodology from other accelerators that could be applied to the ALS. Therefore the specification of the safety calculation evolved as our understanding of the problem grew. The flexibility of XML was crucial to accommodate the evolving nature of the simulation.

XML for Data Storage

Tracy# uses XML as its primary data format of files. The most recent example is for our sextupole upgrade project[19] to reduce the storage ring emittance down to 1/3 of the current value. We are in a process of optimizing the lattice with two new families of sextupole magnets. There is a need to store various lattice configurations, the result of simulations for each one of them, and retrieve the stored content. The resulting XML file contains over 1 million lines. We wrote several client programs of Tracy# to read and update the XML file.

An interesting observation is that there is no real distinction between input and output data in the case of XML. This is similar to the case of using a relational database. A program reads what it needs from the XML files, avoiding redundant calculations and keeping the previous context, then carries out certain tasks and updates XML by changing the attribute values, and also adding new nodes. This means that it grows as the simulation evolves. If we use a relational database, the process of adding nodes corresponds to modifying a table design (schema) or creating a new table with new relations which will be far more complex than a case of XML.

Potential problems with XML can be its file size and processing speed. Fortunately, we have not yet experienced any significant limitation.

The XML effort happened first with the control system upgrade in C#[20]. Tracy# is benefitting from our previous and continuing work on the control system. Routines for database access using ADO.NET, and graphics programming in WPF that were first used for the machine controls have also been migrated to Tracy#. Concerning resource sharing, our C# development is catching up with Matlab that has been used for both physics[21] and controls[16] at the ALS.

INTERACTIVE SCRIPTING

Interactive scripting offers a convenient option to the traditional software development cycle which includes editing, compiling before execution. Our early attempts were TracyM and TracyML[22] that have been superseded by AT[16] in Matlab. An inconvenience is that an interactive programming language isolates itself and requires specially compiled modules to access external routines that are written in compilers. However, on the .NET Framework, it can access the .NET libraries,

including Tracy#, normally and directly without any extra layer.

IronPython Example

Below is an example of using Tracy# interactively from IronPython[14]. Invoking IronPython, establishing the link to the Tracy# library, and importing its name spaces, we can use all the routines directly without any special layer. This example calculates the emittance of the nominal and the new low-emittance lattices of the ALS storage ring. (>>> is a prompt.)

```
>>> import sys
>>> sys.path.append('T:/Tracy/IronPython')
>>> from ipTracy import *
>>> SR=ALSSRW()
>>> SR.getTwiss(1)
True
>>> SR.fitNuxNuyEta(14.25, 9.20, 0.06)
True
>>> SR.getTwiss(1)
True
>>> SR.calcIntegral()
>>> SR.calcEmittance(1.9E9)
>>> print 'Emittance=', SR.NtlEmittance
Emittance= 6.8094734722e-009
>>> SR.fitNuxNuyEta(16.25, 9.20, 0.15)
True
>>> SR.getTwiss(1)
True
>>> SR.calcIntegral()
>>> SR.calcEmittance(1.9E9)
>>> print 'Emittance=', SR.NtlEmittance
Emittance= 2.16587971373e-009
>>>
```

Here T:/Tracy/IronPython is the location of the directory of the Tracy# library and the module ipTracy.py contains lines to import the several .NET assemblies of Tracy#. This kind of capability of interactive scripting is one of the major merits of the .NET Framework.

ONGOING EFFORT

There are multiple efforts in progress to upgrade Tracy# by taking newer functions of the .NET Framework. The most important one is the parallelism to make use of modern multi-core CPUs. We have just added new routines to track particles simultaneously and testing it[23]. A parallelized for statement has tripled the execution speed in case of tracking for dynamic aperture calculations on a PC with a quad-core CPU.

ACKNOWLEDGEMENTS

The author thanks C. Steier for his encouragement and useful advices. He also appreciates C. Timossi for technical discussions on computational issues and useful comments.

REFERENCES

- [1] LBL PUB-5172 Rev. LBL, 1986
A. Jackson, PAC'93, Washington, D.C, USA, p.1432, (1993)
- [2] H. Nishimura, EPAC'88, Rome, Italy, p.803(1989)
- [3] J. Bengtsson, E. Forest and H. Nishimura, "Tracy User Manual", unpublished, ALS, LBNL
- [4] M. Böge, J. Chrin, ICALEPCS'01, San Jose, USA, p.430 (2001)
- [5] M. Heron, et al, EPAC'06, Edinburgh, UK, p.3068 (2006)
- [6] P. Brunelle, A. Loulergue, A. Nadji, L.S. Nadolski, EPAC'04, Lucerne, Switzerland, p.2032 (2004).
- [7] L. B. Dalesio, ICALEPCS'07, Knoxville, USA, p.253(2007)
- [8] H. Nishimura, PAC'01, Chicago, USA, p.3006, (2001)
- [9] H. Nishimura and T. Scarvie, EPAC'06, Edinburgh, UK, p.2263 (2006)
- [10] E. Forest and R. Ruth, Physica D, vol. 43(1) p.105(1990)
- [11] E. Forest, "Beam Dynamics, A New Attitude and Framework", Chap.12, 1998, Harwood Academic Publications
- [12] M. Berz, SSC-152, 1988
Leo Michelotti, IEEE PAC89, CH2669-0(1989)839.
N. Malitskey, A. Reshetov and Y. Yan, SSCL-659, 1994.
- [13] P. Selormey, "DotNetMatrix: Simple Matrix Library for .NET", <http://www.codeproject.com/KB/recipes/psdotnetmatrix.aspx>
- [14] <http://ironpython.codeplex.com>
- [15] <http://www.mono-project.com>
- [16] J. Corbett, G. Portmann, A. Terebilo, PAC'03, Portland, USA, p.2369(2003)
- [17] <http://msdn.microsoft.com/en-us/library/bb387098.aspx>
- [18] H. Nishimura, R. J. Donahue, R. M. Duarte, D. Robin, F. Sannibale, C. Steier, W. Wan, PAC'07, Albuquerque, USA, p.1173(2007)
- [19] H. Nishimura, S. Marks, D. S. Robin, R. D. Schlueter, C. A. Steier, W. Wan, PAC'07, Albuquerque, USA, p.1170(2007)
- [20] H. Nishimura, C. Timossi, G. Portmann, M. Urashka, C. Ikami and M. Beaudrow, PCaPAC'08, Ljubljana, Slovenia, p.122(2008)
- [21] G. Portmann, J. Corbett and A. Terebilo, PAC'05, Knoxville, USA, p.4009(2005)
<http://www-ssrl.slac.stanford.edu/at/>
- [22] H. Nishimura, ICAP'98, Monterey, USA(1998).
<http://www.slac.stanford.edu/xorg/icap98/papers/F-Tu06.pdf>
H. Nishimura and W. Decking, EPAC'98, Stockholm, Sweden, p.1224(1998)
- [23] <http://msdn.microsoft.com/en-us/concurrency>

MODELING SINGLE PARTICLE DYNAMICS IN LOW ENERGY AND SMALL RADIUS ACCELERATORS

E. Nissen, B. Erdelyi, Department of Physics, Northern Illinois University, Dekalb, IL 60115, USA

Abstract

This research involves the development of a model of the small circumference (11.5 m) accelerator in which the earth's field has a strong effect, and in which image charge forces are also included. The code used for this simulation was COSY Infinity 9.0 which uses differential algebras to determine high order map elements, as well as quantities such as chromaticity. COSY also uses Normal Form algorithms to determine the betatron tune and any amplitude dependent tune shifts which may result. The power of COSY is that it can derive the required quantities directly from the map without costly integration and tracking. Thus determining the map for both the default elements of the ring, plus the effects of image charge forces, and the earth's magnetic field is both non-trivial, and important. This research uses the Baker Campbell Hausdorff method to determine the map of the ring with the external fields included. Furthermore COSY has the ability to directly implement misalignments within the beamline itself allowing for a study of their effects on beam dynamics. The presentation will include both coding development and applications to the University of Maryland Electron Ring.

INTRODUCTION

With the increase in demand for high current accelerators, methods for determining the effects of space charge become more important. One method for gaining experimental data on the effects of space charge is to use a low energy electron beam to model a high energy heavy ion beam, this is the approach used by the University of Maryland Electron Ring (UMER) [?]. UMER sends 10 Kev electrons through a storage ring that is only 11.5 meters in circumference. Currently all acceleration occurs in the electron gun, which sends the beam through a matching section and into the Y-shaped injector. This injector involves offsets on both the injection and recirculation sides, the ring then uses an additional 17 sections comprised of a bending dipole between a pair of quadrupoles, followed by a diagnostic chamber, followed by another set of quadrupoles enclosing a dipole. The arrangement of the elements in the ring are shown in Fig. 1. Due to the small radius and low energy of the beam the effects of the earth's field on the trajectory of the beam is nontrivial. Furthermore, the offsets in the injection and recirculation parts of the Y-shaped section mean that the effects of image charge on the beam should also be taken into account.

First there will be a brief introduction to COSY Infinity and its unique properties, then there will be an overview of how the earth's field, the image charge force, and unique el-
Computer Codes (Design, Simulation, Field Calculation)

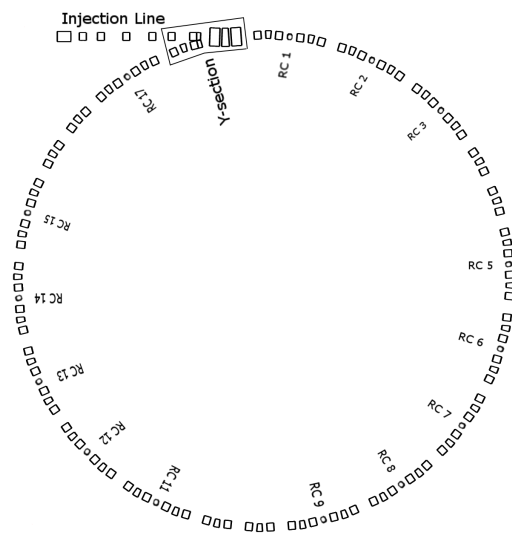


Figure 1: COSY Infinity produced diagram of the University of Maryland Electron Ring. Sections marked RC contain ring chambers which house both non-intercepting Beam Position Monitors, and intercepting Phosphor screens. Sections not marked with an RC number contain glass gaps for current monitors.

ements contained in this particular beam are implemented. Finally there will be a brief look at some experimental observations.

CODE DEVELOPMENT

The code used in this study is COSY Infinity 9.0. This code uses differential algebraic vectors which allow not only for an accurate calculation of numerical derivatives, but also carries them through the various mathematical operations. This behavior means that COSY can integrate a test particle through an electromagnetic field and all of the variable dependencies will be preserved, allowing for fast accurate computation of maps [?]. COSY also has a large library of default beam elements, so in this study they were used as often as possible.

Short Solenoid

One issue that was dealt with early on was the field profile of a short solenoid in the injection line. The field profile that was measured for the physical element was very different from the kind used in the available COSY solenoids, so the open architecture of COSY allowed us to model the solenoid using a field profile provided by fitted data. The

field profile of the short solenoid is shown in Fig 2.

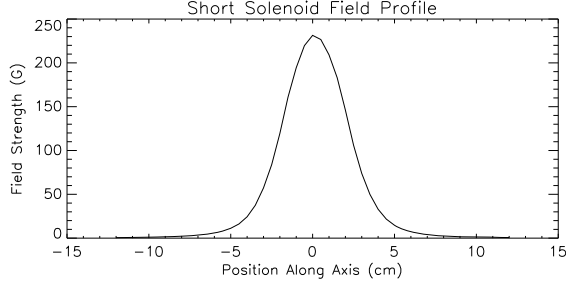


Figure 2: This is a representation of the field of the UMER short solenoid.

Earth's Magnetic Field

The earth's magnetic field accounts for approximately 20% of the bending in the ring, and its direct effects, as well as the methods used to counteract them must be modeled. The earth's field is implemented as a kick using the Baker Campbell Hausdorf (BCH) method. This method shows that if we have a set of two differential equations using the independent variable s , and that they each have a solution that is valid over a given interval in s that is represented by ℓ .

$$\frac{d\vec{z}}{ds} = \vec{g}_1(\vec{z}, s) \Rightarrow \vec{f}_1(\ell), \quad (1)$$

$$\frac{d\vec{z}}{ds} = \vec{g}_2(\vec{z}, s) \Rightarrow \vec{f}_2(\ell), \quad (2)$$

Then the best way to find a solution for the combination of the two equations is to compose the solutions in this manner, with increasing accuracy with decreasing ℓ [3].

$$\begin{aligned} \frac{d\vec{z}}{ds} &= \vec{g}_1(\vec{z}, s) + \vec{g}_2(\vec{z}, s) \\ \Rightarrow \vec{f}_1\left(\frac{\ell}{2}\right) \circ \vec{f}_2(\ell) \circ \vec{f}_1\left(\frac{\ell}{2}\right). \end{aligned} \quad (3)$$

In the case of an accelerator element that is being acted on by the earth's magnetic field, the two solutions are expressed as a kick placed between two maps. The number of kicks was determined by increasing the number for the various elements one by one until the answers had converged past the number of significant figures available for the field. The drifts and dipoles each used 6 kicks per element, and the quadrupoles used 15 kicks per element.

The strength of the kick is determined from the equations of motion for the beam, and the strength of the earth's magnetic field is interpolated from a series of measurements previously taken at UMER. The magnetic field kicks are given as:

$$\Delta a(s) = \left(b_i - \frac{B_y}{B_z}\right) \sin\left(\frac{B_z}{\chi_{m0}}(1 + hx_i)s\right) +$$

$$+ \left(a_i - \frac{B_x}{B_z}\right) \cos\left(\frac{B_z}{\chi_{m0}}(1 + hx_i)s\right) + \frac{B_x}{B_z}, \quad (4)$$

$$\begin{aligned} \Delta b(s) &= \left(\frac{B_x}{B_z} - a_i\right) \sin\left(\frac{B_z}{\chi_{m0}}(1 + hx_i)s\right) + \\ &+ \left(b_i - \frac{B_y}{B_z}\right) \cos\left(\frac{B_z}{\chi_{m0}}(1 + hx_i)s\right) + \frac{B_y}{B_z}. \end{aligned} \quad (5)$$

In this case COSY is using a coordinate system in which x and y are the horizontal and vertical positions respectively, and a and b are analogous to the momentum for each variable. In the case above B_x , B_y , and B_z are the measured field strengths. χ_{m0} is the magnetic rigidity of the beam, and h is the inverse of the radius of curvature, h is zero for a straight element.

In the interests of completeness, and in the unlikely case that there is a zero B_z field, the following equations were also calculated.

$$\Delta a(s) = \frac{B_y}{\chi_{m0}}(1 + hx_i)s, \quad (6)$$

$$\Delta b(s) = \frac{B_x}{\chi_{m0}}(1 + hx_i)s. \quad (7)$$

Image Charge Kick

This BCH method was also used to implement the image charge force in the injection section. The image charge implementation assumed cylindrical charge symmetry of the beam, which places the image charge at a distance of R^2/ξ where R is the radius and ξ is the offset of the center of the beam. The number of kicks required was analyzed in the same manner as the earth's field, it was found that the number of kicks for the quadrupole and drifts/dipoles used for the earth's field was more than sufficient for the image charge. This then leads to a kick on the beam center of:

$$\Delta a = \frac{\hat{x}}{v_0} \sqrt{-\frac{q\lambda}{2\pi\epsilon_0 m} \ln\left(\frac{R^2 - (x^2 + y^2)}{R^2}\right)}, \quad (8)$$

$$\Delta b = \frac{\hat{y}}{v_0} \sqrt{-\frac{q\lambda}{2\pi\epsilon_0 m} \ln\left(\frac{R^2 - (x^2 + y^2)}{R^2}\right)}. \quad (9)$$

This allows for the image charge force to be calculated for the center of the beam.

Dipole Modification

As previously stated the earth's magnetic field accounts for approximately 20% of the bending of the beam, in UMER this is counteracted by changing the current to the bending dipoles in the ring. These dipoles have a design bending radius of 10 degrees, but they have their magnetic fields reduced to counteract the earth's magnetic field. This will lead to a small angular and positional offset at the exit of the magnet. COSY Infinity always assumes that if the beam enters at the center of the element then it will exit at the center of the element, so it was necessary to give the bending dipoles in this simulation an offset and a tilt at

their exit as a geometric correction. This effect can be seen in Fig 3.

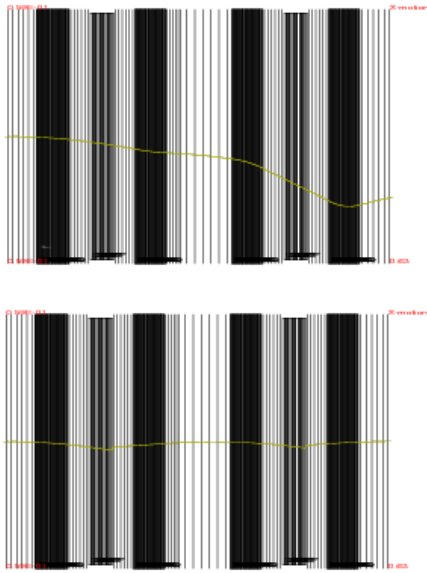


Figure 3: The top shows the COSY generated path of the beam centroid as it moves uncorrected through a ring section. The second shows the same beam with the dipoles adjusted to counteract the earth's magnetic field.

OBSERVATIONS

A series of experiments was recently performed to determine how accurately the COSY model portrayed real life. Fig 4 shows that the agreement between theory and tracking is accurate through RC 9, after this point unresolved issues make accurate simulations difficult. Also a compar-

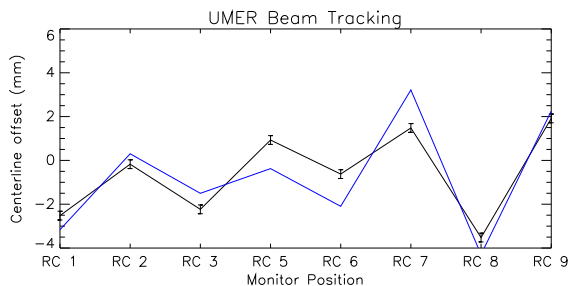


Figure 4: Comparison of horizontal tracking data using the beam position monitors (black) and the predicted location of the beam centroid using COSY Infinity (blue).

ison was made between one set of earth's field compensation values and those calculated using COSY. The previous values were designed to have the bending dipole reduced by an angle equal to the bending caused by the earth's field in that region. The values determined by COSY were found by fitting the bending dipoles such that a beam that entered Computer Codes (Design, Simulation, Field Calculation)

a ring section with zero angle and zero offset would exit that section with zero angle and zero offset. A comparison of the tracking data is shown in Fig 5.

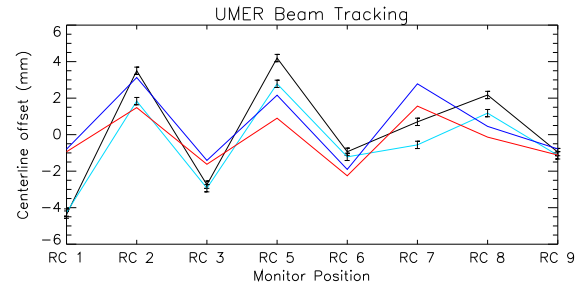


Figure 5: This is a comparison of the measured and predicted trajectories for both methods of earth-field compensation. The black line uses the integrated offset method, while the light blue line uses the COSY calculated values. The dark blue and red lines are the cosy predicted values for the compensation and the COSY calculated settings respectively.

CONCLUSIONS

The adaptability and solid theoretical underpinnings of the code COSY Infinity allow for accurate simulations to be made not only of well behaved large accelerators, but also small radius, low energy accelerators, which have non-trivial complications such as image charge forces and the earth's magnetic field. Using this code on the University of Maryland Electron Ring will allow for a better understanding of high space charge systems without the cost of a heavy ion accelerator.

ACKNOWLEDGEMENTS

This work was supported by the department of Energy under contract DE-FG02-08ER41532 with Northern Illinois University. The funding of UMER is provided by the Department of Energy offices of high energy and fusion sciences. The data shown in this work would not have been possible without the help of the UMER staff, especially R. Kishek, D. Sutter, S. Bernal, W. Chao, B. Beaudoin, and K. Fiuza.

REFERENCES

- [1] R. A. Kishek et. al., "Scaled Models: Space-Charge Dominated Electron Storage Rings," International Journal of Modern Physics A, 2007
- [2] M. Berz "Advances in Imaging and Electron Physics vol 108," 1999
- [3] A. J. Dragt, "Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics," 2009

POSSIBILITY OF ROUND BEAM FORMATION IN RIBF CYCLOTRONS

H. Okuno[#], RIKEN Nishina Center, Wako, Saitama, 351-0198, Japan

A. Adelman, PSI, 5232 Villigen, Switzerland

J. J. Yang, CIAE, Beijing, 102413, China & Tsinghua University, Beijing, 100084, China.

Abstract

Since 1997 RIKEN Nishina center has been constructing a next-generation exotic beam facility, RI beam factory (RIBF), based on a powerful heavy ion driver accelerator. Its accelerator complex was successfully commissioned at the end of 2006 and started supplying heavy ion beams in 2007. The four ring cyclotrons (RRC, IRC, SRC and SFC) connected in series accelerate the energy of the heavy ion beams up to 400 MeV/u for the lighter ions such as argon and 345 MeV/u for heavier ions such as uranium. Intensity upgrade plans are under way, including the construction of a new 28 GHz superconducting ECR ion source. The new ECR will take all the succeeding accelerators and beam transport lines to a space charge dominant regime, which should be carefully reconsidered to avoid emittance growth due to space charge forces. Beam dynamics in the low energy cyclotron, RRC was studied with OPAL-cycl a flavor of the OPAL. The simulation results clearly show vortex motions in the isochronous field, resulting in round beam formation within the first 10 turns after the injection.

INTRODUCTION

RIKEN Nishina center has undertaken construction of an RI Beam Factory (RIBF) [1] since April 1997 aiming to realize a next generation facility that is capable of providing the world's most intense RI beams at energies of several hundred MeV/nucleon over the whole range of atomic masses. The RIBF requires an accelerator complex which would accelerate the full mass range of

ions and deliver ~80 kW of uranium beam at energy of 345 MeV/nucleon. Figure 1 shows a bird's eye view of

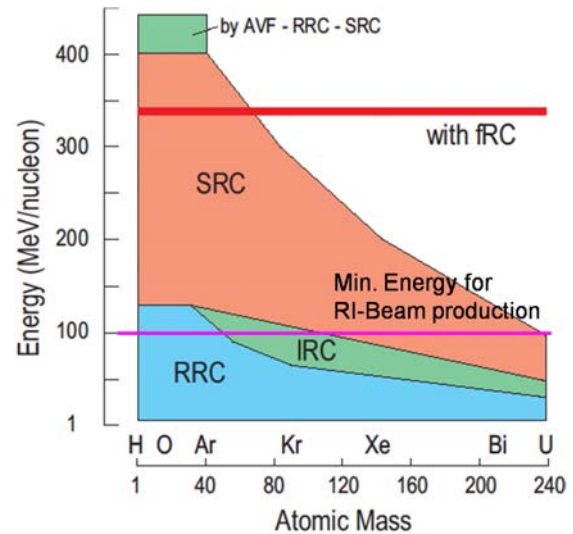


Figure 2: Performance of the RIBF accelerator complex.

RIBF. The left part is the old facility completed in 1990. Using the four-sector K540-MeV ring cyclotron (RRC), many experiments were carried with RI beams of light ions because RRC can accelerate relatively light ions up to 100 MeV/u, which is the lower limit for the RI-beam production as shown in Fig. 2. At first, the two ring cyclotrons, Intermediate Ring Cyclotron (IRC). Superconducting Ring Cyclotron (SRC) were designed as energy boosters for the RRC in order to expand the

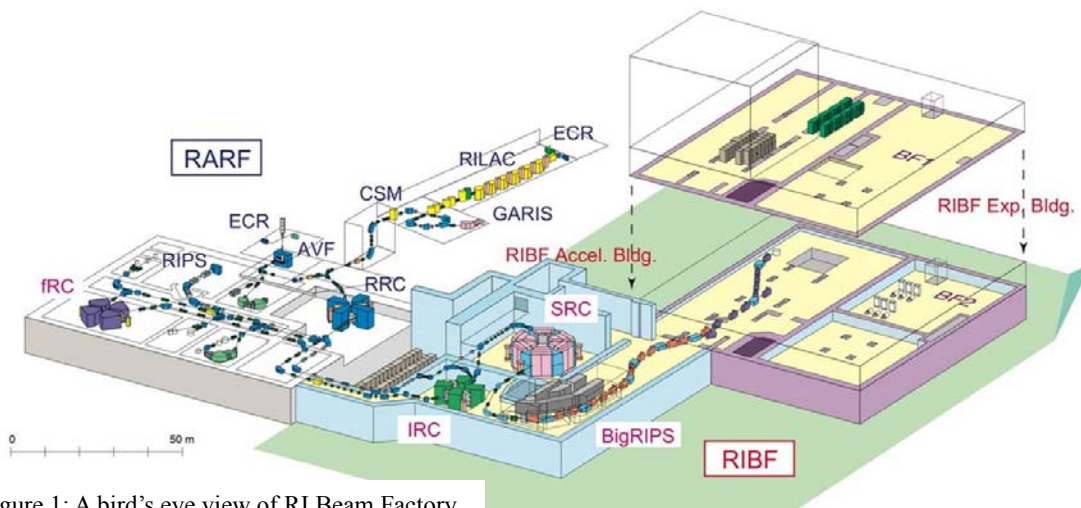


Figure 1: A bird's eye view of RI Beam Factory.

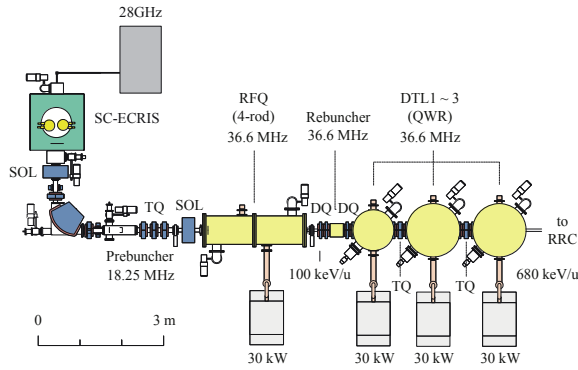


Figure 3: A plan of the new injector.

mass range where RI beams can be produced as shown in Fig. 2. In 2000, after the appearance of the RIA project in the US, we realized that all the ion beams including uranium can be accelerated up to the maximum energy of 345 MeV/u by adding another cyclotron, a fixed-frequency ring cyclotron (fRC), before the IRC. In this way, we are now able to generate a wide variety of RI beams from hydrogen to uranium. The intensity of the goal is more than 1 μA (6.0×10^{12} pps).

The RIBF accelerator complex was successfully commissioned at the end of 2006 and started supplying heavy ion beams in 2007. In 2008 intensity of Ca beam extracted from SRC reached 170 μA (1.1×10^{12} pps) which is close to the goal intensity. Uranium beam of 0.4 μA (2.5×10^9 pps) could be accelerated up to 345 MeV/u in November 2008. [2] This intensity is a value of the world's top level, but far from our goal.

Key issues to increase the intensity of uranium ion beams can be clearly pointed out as follows. First, we need more beams from the ion source. A new 28GHz superconducting ECR ion source has been constructing. [3] This ion source is designed to have as large plasma volume as 1100 cm^3 and expected to produce U^{35+} ions at an intensity of more than 15 μA , which is necessary to obtain 1 μA beams from the SRC. Beam tests started from the last May with 18 GHz rf power source and operation with a 28 GHz source will start from the next year. Next, we need a new injector which efficiently accelerates ion beams from the new powerful ion source in order to avoid the emittance growths due to their

PSI Inj. 2	RRC	fRC	IRC	SRC
1	0.644	0.096	0.031	0.016

Table 1: List the value of the parameter of (1) in the text.

space charge forces. Figure 3 shows a plan for the new injector which is designed to accelerate ions with a mass-to-charge ratio of 7, aiming at heavy ions such as $^{84}\text{Kr}^{13+}$, $^{136}\text{Xe}^{20+}$ and $^{238}\text{U}^{35+}$, up to energy of 680 keV/nucleon. [4] It consists mainly of the SC-ECR, an RFQ linac based on the four rod structure and three drift-tube linac (DTL) based on the quarter-wavelength resonator (QWR). A four-rod structure RFQ denoted by Kyoto University will be recycled for this purpose after some modifications to change the resonance frequency by about 10 %. The new injector linac will be commissioned from the autumn of 2010. The following section of this paper will describe the space charge effects in the succeeding cyclotrons which may be taken to space charge dominant by the intensity upgrade program for uranium beam.

MOTIVATION

The intensity upgrade of uranium beam described in the previous section motivated us to study the vortex motions in the RIBF cyclotrons. Longitudinal space charge force causes additional acceleration for head particles and deceleration for tail particles. The accelerated or decelerated particles move to higher or lower radii due to isochronous condition in cyclotron, causing rotation of the bunch. The nonlinearity of space charge force produces spiral shaped halo, finally rotating sphere. These vortex motion phenomena were theoretically studied as shown in [5-9] and experimentally verified at the PSI Injector II.

Pozdeyev described in his thesis [8] that the effects of the space charge force for bunches of similar lengths approximately scale as:

$$\frac{qI}{\gamma^5 m h \omega^3} \quad (1)$$

where I is the total beam current, h is the harmonic number and the momentum is $p = m\gamma R\omega$. Table 1 lists the values of the parameter in the case of Injector II, RRC, fRC, IRC and SRC, indicating that RRC is approximately equivalent to the Injector II from the point of view of space charge force effects, while the effects in the other three cyclotrons are small. Therefore our beam dynamics study focused on the RRC, low energy cyclotron.

Item	Value
K value	540
Number of magnet sector	4
Sector Angle	50 deg
Inj. and ext. radius	0.39 and 3.56 m
Number of Cavity	2 (double gap)

Table 2: Main specification for RRC.

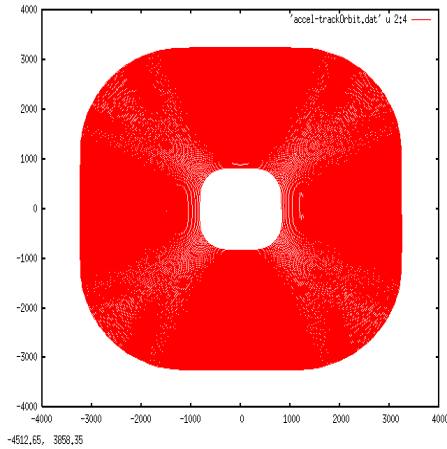


Figure 4: Top view of single-particle tracking of the reference particle up to the final turn in RRC.

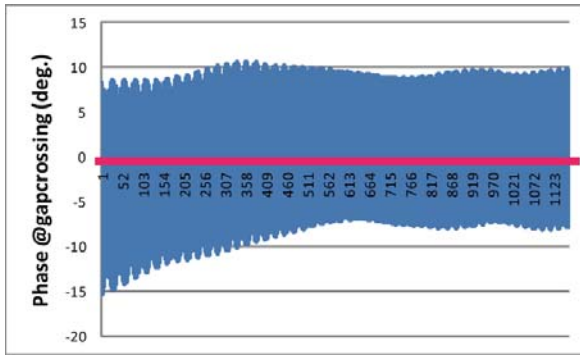


Figure 5: Phase slip at the gap crossing in RRC.

SIMULATIONS USING OPAL

OPAL-cycl [10] is one of the flavours of the Object Oriented Parallel Accelerator Library (OPAL) framework. It is a new 3D PIC based self-consistent numerical simulation code covering neighbouring bunch effects. Self consistency is to be understood in the electrostatic approximation. A more detailed description of the OPAL framework and OPAL-cycl code can be found in the User's Reference Guide [11].

Table 2 shows the main parameters of RRC which consists of the four-sector magnets and two double-gap

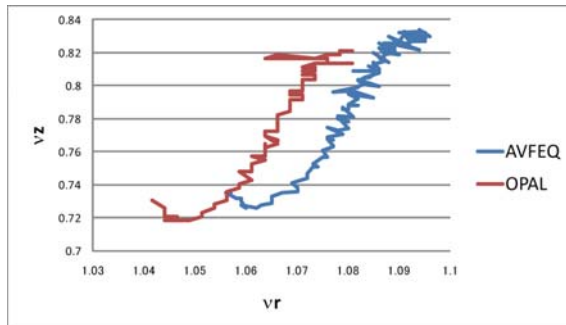


Figure 6: Tune diagram for RRC.

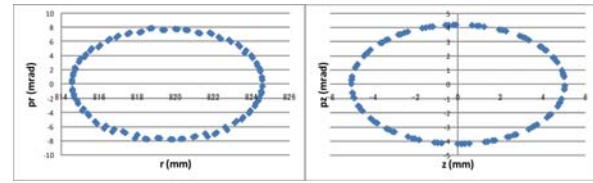


Figure 7: Eigen-ellipses at the first turn of RRC in the radial and vertical directions.

rf resonators. Simulations using OPAL-cycl need realistic isochronous field map and radial distribution of rf voltage at the gaps. The field map for isochronous field was calculated by ANSYS with high accuracy. [12] Radial distribution of rf voltage was assumed to be constant since the resonator of RRC shows almost constant radial distribution in the case of 18.25 MHz.

Single-particle tracking and tune calculation

OPAL-cycl prepares single-particle tracking and tune calculation which are important to check isochronism of the field map, validity of the initial condition and matching phase ellipse before multi-particle tracking including space charge forces. Figure 4 shows the top view of the acceleration up to the final turn. It took about 296 turns for the particle to reach the final energy of the cyclotron. Phase slip at the rf gap crossing is shown in Fig 5, showing that the used field map is sufficiently isochronous. Betatron frequencies in the radial and vertical direction were calculated using the tune calculation mode of OPAL-cycl, as shown in Fig. 6. The result from AVFEQ [13] and OPAL-cycl shows good agreements even though different numerical algorithms are used, however we are in the process understanding the visible differences. Single-particle tracking was carried out with initial offsets of $r(z) = 5.0$ mm, $pr(pz) = 0.0$ mrad from the static equilibrium orbit at the injection energy, in order to get eigen-ellipses in the radial and vertical directions. The results in Fig. 7 shows that the ratios between the semi major axis and semi minor axis are 5.0/4.2 and 5.0/2.5 mm/mrad in the radial and vertical direction, respectively. These ellipses were used as the initial conditions for multi-particle tracking described in the next subsection.

Multi-particle tracking including space charge force effects

Multi-particle tracking including space charge forces was carried out at the beam current of 0.5 mA after obtaining the initial conditions from the single-particle tracking studies. The first 10 turns after injection were simulated with acceleration because we are mainly interested in the behaviour of bunches just after the beam injection. Initial transversal rms emittance was assumed to be 2.5 mm mrad from the operational experiences. Initial rms bunch length was assumed to be 2 or 4 degree in order to study how the bunch length impacts on the vortex motion. Figure 8 shows the results for the simulations. The results for both cases of 2.0 and

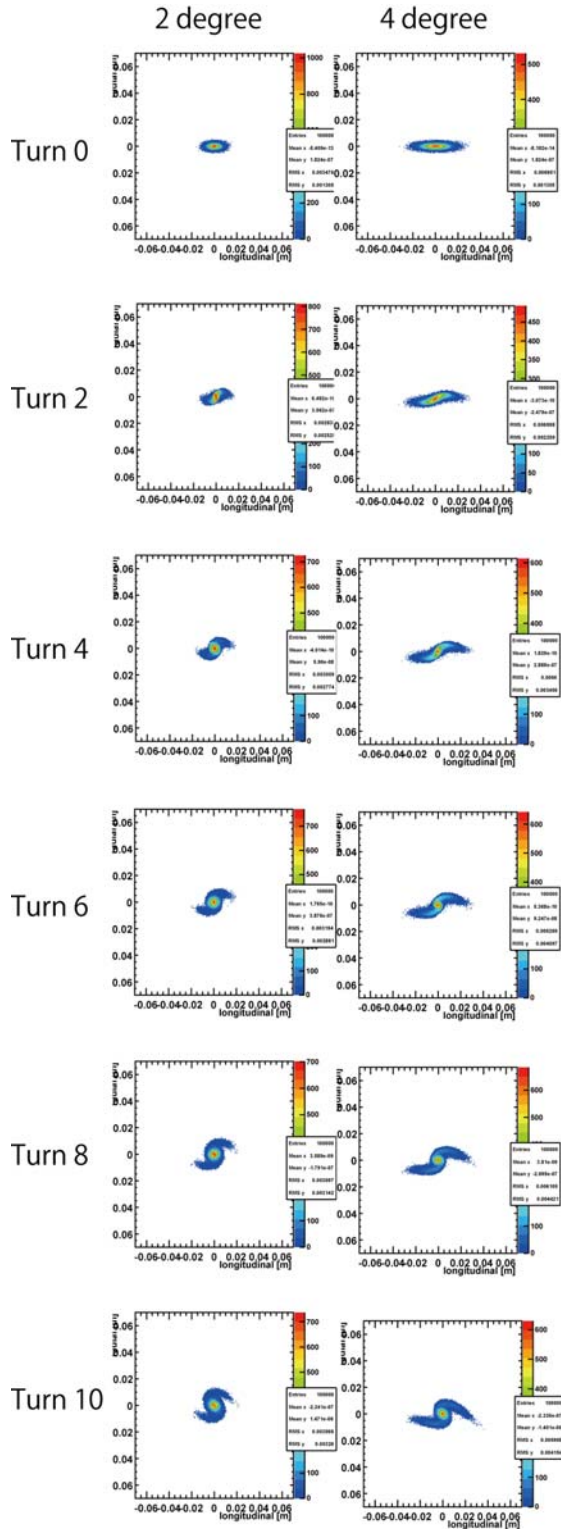


Figure 8: Result for multi-particle tracking. Turn numbers are shown in the left side of the contour plots. Rms of the initial bunch length is shown in the top of the contours.

4.0 degree clearly show the vortex motions of bunches toward stationary round distributions. The bunch in the case of 2.0 degree has smaller tails and rotates faster than that in the case of 4.0 degree. More detailed simulations up to the final turns including space charge effects from neighbouring turns are in progress.

SUMMARY AND FUTURE WORKS

The intensity upgrade of uranium beam is under way in the RIBF accelerator complex. This may be based on the vortex motions in the low energy cyclotron RRC from Pozdeyev's scaling law, which motivated us to simulate the beam dynamics in RRC using OPAL-cycl. After obtaining the initial conditions using single-particle tracking, as a first step we simulated the first 10 turn of the RRC. The results show the clear signs of vortex motions, motivating us to carry out a more detailed simulation campaign up to the final energy, including neighbouring turns.

ACKNOWLEDGMENTS

The authors thank L. Stingelin for providing an isochronous field for RRC.

REFERENCES

- [1] Y. Yano, Nucl. Instrum. Methods Phys. Res., Sect. B **261** (2007) 1009.
- [2] N. Fukunishi, et al., Proc. Particle Accelerator Conference 2009, in press, May 2009, Vancouver.
- [3] T. Nakagawa, et al., High Energy Physics and Nuclear Physics **31** (2007) 133.
- [4] O. Kamigaito, et al., 3rd Ann. Meeting of PASJ and 31th Linac meeting in Japan, Sendai, (2006) 502.
- [5] S. Adam, IEEE Trans. On Nuclear Science **32**, 2507 (1985).
- [6] S. Koscielniak and S. Adam, in Proc. Particle Accelerator Conf. (Washington, 1993), p. 3639.
- [7] A. Adelmann, Ph.D. thesis, ETHZ, Switzerland (2002), no14545.
- [8] E. Pozdeyev, Ph. D. Thesis, MSU, USA (2003)
- [9] S. Adam, Ph.D. thesis, ETHZ, Switzerland (1985), no 7694.
- [10] J.J. Yang, A. Adelmann, M. Humbel, et al., Proc. of HB2008, in press, August 2008, Nashville.
- [11] A. Adelmann, et al., Tech. Rep. PSI-PR-09-05, Paul Sherrer Institut (2009).
- [12] L. Stingelin, private communication.
- [13] A. Goto, private communication.

SET CODE DEVELOPMENT AND SPACE CHARGE STUDIES ON ISIS

BG Pine, DJ Adams, CM Warsop, RE Williamson
Rutherford Appleton Laboratory (STFC), Oxfordshire, UK.

Abstract

The ISIS Facility at the Rutherford Appleton Laboratory in the UK produces intense neutron and muon beams for condensed matter research. It is based on a 50 Hz proton synchrotron which accelerates $\sim 3 \times 10^{13}$ protons per pulse (ppp) from 70 to 800 MeV, corresponding to beam powers of ~ 0.2 MW. Studies are under way for major upgrades in the Megawatt regime. Underpinning this programme of operations and upgrades is a study of the high intensity effects that impose limitations on beam power.

The behaviour of the beam in the 50 Hz rapid cycling synchrotron (RCS) is largely characterised by high space charge levels and the effects of fast ramping acceleration. High intensity effects are of particular importance as they drive beam loss, but are not fully understood with only limited analytical models available. This paper reviews development of a new space charge code Set, which is designed to address key issues on ISIS and similar RCS machines.

INTRODUCTION

ISIS high intensity operation is restricted by beam loss, as irradiation of equipment limits access for essential maintenance. Understanding beam loss is therefore of vital importance, however due to the complex interactions between the beam particles and their environment such understanding is challenging both analytically and numerically.

The ISIS Synchrotron Group is actively studying high intensity effects of the beam in a number of different ways, both to improve performance of the accelerator and also to enable the design of upgrades which can achieve significantly higher beam intensities. This paper focuses on developments of the beam tracking code Set.

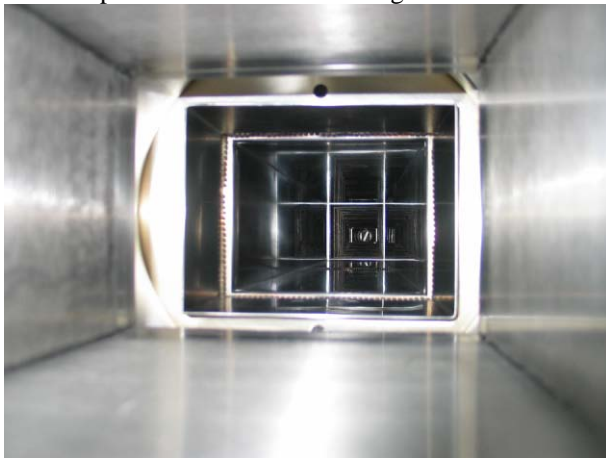


Figure 1: Rectangular ISIS vacuum vessel.

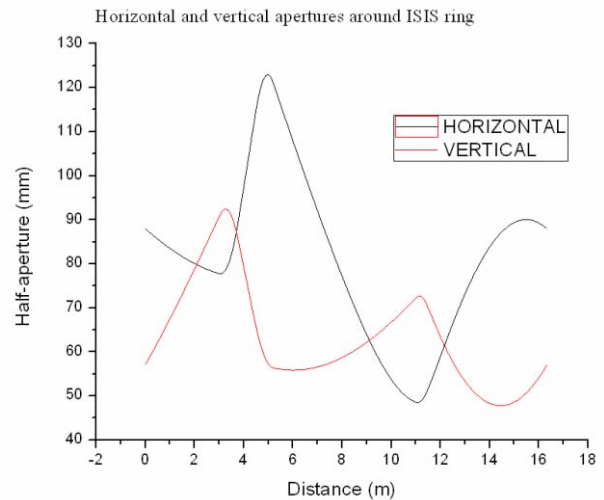


Figure 2: Profiled vacuum vessels in ISIS super-period.

SET

A new code Set is under development at ISIS. This code is intended to supplement the use of ORBIT [1] for 2D and 3D beam tracking simulations, as a tool that can be readily modified and redeployed as required to meet a given purpose. In particular, the focus is on the challenges of the ISIS RCS, including image forces from the unique profiled vacuum vessel (Figures 1 and 2), halo predictions, 2D and 3D RCS space charge effects and overall to understand and predict beam loss. Set works using either MAD input data or its own matrix routines for generating lattices, and has an FFT based Poisson-solver for calculating the beam's space charge. Early simulation work [2, 3] focused on replicating ORBIT results for the half integer resonance. Example results for the ISIS lattice (2D, coasting beam) driven with a $2Q_v = 7$ resonance are shown in Figures 3, 4 and 5.

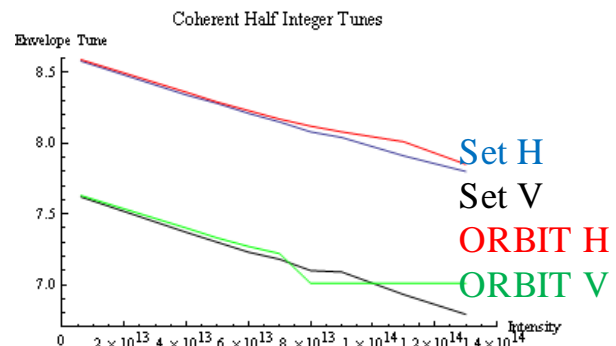


Figure 3: Envelope frequencies intensity sweep.

Figure 3 shows Set and ORBIT envelope frequencies as the intensity is swept from $1 - 1.4 \times 10^{13}$. Figure 4 shows the incoherent tune footprints after 100 turns, as the intensity

is varied from $6\text{E}12$ ppp to $13\text{E}13$ ppp. Figure 5 shows beam phase and real space on the 100th turn for an intensity of $6\text{E}13$.

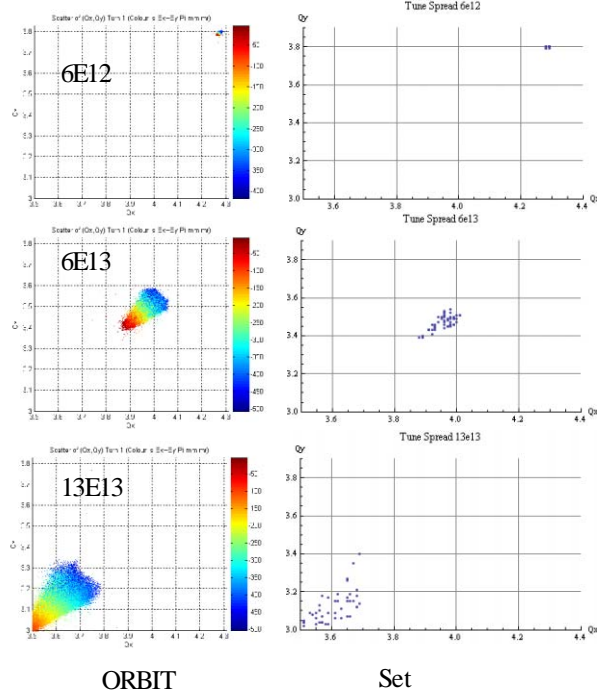


Figure 4: Incoherent tune comparison ORBIT - left, Set - right.

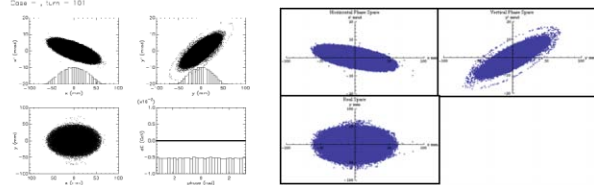


Figure 5: Phase space on 100th turn ORBIT - left and Set - right.

Set has been used to study tune shifts from image forces and closed orbits [4] and the results compared with Laslett theory (Figure 6). Direct space charge should have no influence on the coherent dipole tune, as the charge distribution of the beam moves with the centre of charge. However image forces will affect the coherent tune, as the centre of charge does move relative to the vacuum vessel. This is of particular interest on ISIS due to the vacuum vessel which follows the design beta function of the beam. At high intensities the machine is very sensitive to closed orbit changes, which may indicate beam loss driven by image forces.

Present upgrade studies are investigating the benefits of increasing injection energy from 70 to 180 MeV. Set has been used to study the space charge limitations at this higher energy. The half integer simulations carried out for the nominal ISIS ring were reproduced, but for an injection energy of 180 MeV rather than 70 MeV.

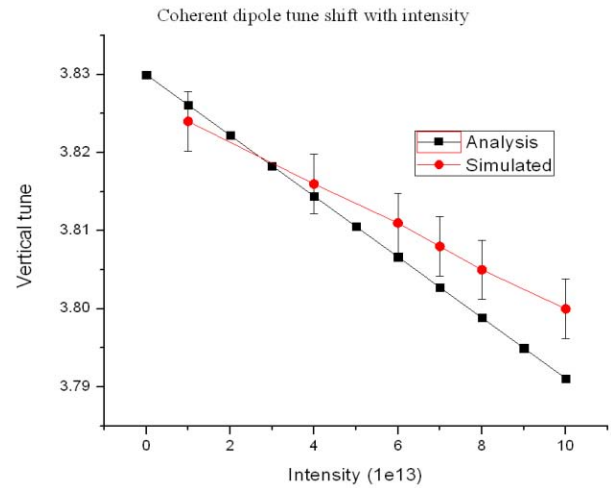


Figure 6: Simulated versus analytical tune shift.

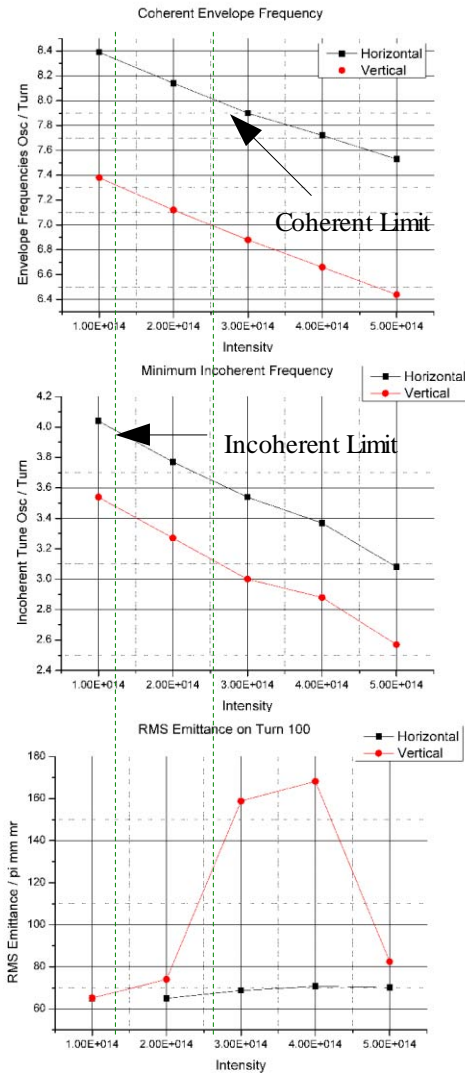


Figure 7: Envelope frequencies, incoherent tune and RMS emittance growth associated with half integer resonance for 180 MeV injection energy.

Simple scaling of the space charge force indicates peak intensities should increase by a factor of 3 over 70 - 180 MeV. As can be seen from Figure 7, the RMS emittance begins to rise at 3 times the intensity seen in the previous case. Image forces become more significant when the beam is executing a closed orbit, as image forces from the beam pipe only cancel when the beam is well centred. Figures 8 and 9 show the results of simulation runs including half integer driving terms from the trim quadrupoles (ISIS has special programmable quadrupoles distinct from the main lattice), and also an angular kick once per turn. Each set of simulations was run twice, to allow the resulting perturbed beam distribution to be matched into the lattice. Figure 8 shows the variation of closed orbit (RMS position) around the ring as a function of intensity, from 1 - 5E14 ppp. Figure 9 compares RMS emittance with and without a closed orbit at an intensity of 2E14 ppp. Image forces are clearly influencing the behaviour of the beam, much as we expect on ISIS. A more complete analysis, and eventually experimental work on the ISIS synchrotron, are to follow.

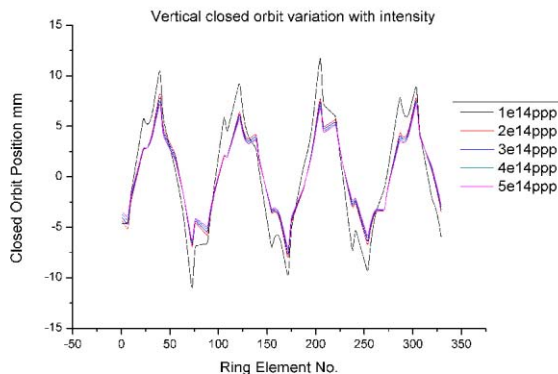


Figure 8: Matched closed orbit variation with intensity.

FUTURE WORK

Set

Work is also under way on a 1D longitudinal code which includes space charge and impedances. This will eventually be added to Set to make it fully 3D. Results are shown in Figure 10 for bunch length and phase convergence versus macro-particle number.

A parallel version of 2D Set has been implemented, and successfully run on the SCARF cluster [5]. Further work will add a realistic injection scheme, including the effect of foil scattering. The long term goal is to carry out a full simulation of the ISIS cycle and recreate the beam loss patterns seen on the real machine.

SUMMARY

Understanding space charge, and hence beam loss, is essential for the operation of a high intensity RCS, and even more important for the design of an upgrade.

A new code Set is being developed to enable further study of key beam dynamics issues that are important for ISIS, such as image effects and any dominant loss mechanisms.

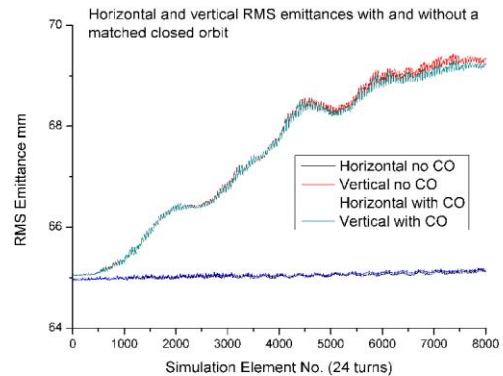


Figure 9: RMS emittance with and without matched closed orbit errors for 2E14 ppp intensity.

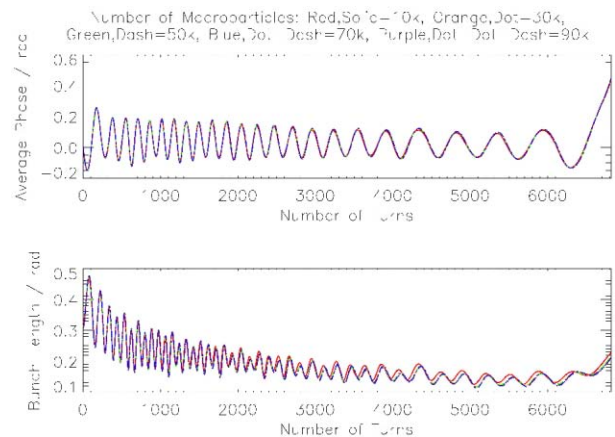


Figure 10: 1D convergence tests showing results for phase and bunch length.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge essential collaborations with ISIS Diagnostics, ISIS Operations and STFC e-Science.

REFERENCES

- [1] JA Holmes *et al*, ORBIT User Manual, ORNL Tech. Note SNS/ORNL/AP/011.
- [2] CM Warsop *et al*, "Studies of space charge loss mechanisms on the ISIS synchrotron", Proc. of PAC07, Albuquerque, USA.
- [3] CM Warsop *et al*, "Space charge loss mechanisms associated with half integer resonance on the ISIS synchrotron", Proc. of EPAC08, Genoa, Italy.
- [4] BG Pine *et al*, "Image simulations on the ISIS synchrotron", Proc. of EPAC08, Genoa, Italy.
- [5] <http://sct.esc.rl.ac.uk/SCARF/index.html>.

COMPLETE RF DESIGN OF THE HINS RFQ WITH CST MWS AND HFSS*

G. Romanov[#], A. Lunin, Fermilab, Batavia, IL 60510, USA.

Abstract

Similar to many other linear accelerators, the High Intensity Neutron Source requires an RFQ for initial acceleration and formation of the bunched beam structure. The RFQ design includes two main tasks: a) the beam dynamics design resulting in a vane tip modulation table for machining and b) the resonator electromagnetic design resulting in the final dimensions of the resonator. The focus of this paper is on the second task. We report complete and detailed RF modeling on the HINS RFQ resonator using simulating codes CST Microwave Studio (MWS) and Ansoft High Frequency Structure Simulator (HFSS). All details of the resonator such as input and output radial matchers, the end cut-backs etc have been precisely determined. Finally in the first time a full size RFQ model with modulated vane tips and all tuners installed has been built, and a complete simulation of RFQ tuning has been performed. Comparison of the simulation results with experimental measurements demonstrated excellent agreement.

INTRODUCTION

Within the framework of the High Intensity Neutrino Source (HINS) program at FNAL, we plan to build and operate a portion of the Front End (up to energy of 62 MeV) as a technical feasibility proof of the proposal. A detailed description of the project and the current status is given in [1]. In the Front End test stand a four vane 325 MHz Radio Frequency Quadrupole (RFQ) will be used for bunching the beam and accelerating it from 50 keV to 2.5 MeV.

The complete beam dynamics design, resulted in a vane tip modulation table for machining, is described in [2]. The mechanical design concepts for this RFQ, tuning results, manufacturing of the RFQ in industry and the preliminary results of initial testing of RFQ at the Front End test stand are discussed in [3].

The electromagnetic design of RFQ resonators is rather complicated and requires essentially three-dimensional modeling. That, and also an additional complication with RF tuning because of some blunder made in the mechanical design of RFQ, urged us to develop a full length 3D RFQ model for simulation. Modern three-dimensional electromagnetic codes are now available and successfully used for RFQ design [4, 5, and 6]. This paper focuses exclusively on the computational technique of electromagnetic design. We report complete and detailed RF modeling on the HINS RFQ resonator using

simulating codes CST Microwave Studio (MWS) and Ansoft High Frequency Structure Simulator (HFSS).

RFQ MODEL FOR ELECTROMAGNETIC SIMULATION

The basic parameters of the RFQ are given in table 1.

Table 1

Input energy	50 keV
Output energy	2.5 MeV
Frequency	325 MHz
Total length of vanes	302.428 cm
Average bore radius	3.4 mm

The RFQ design has several features that have been taken into account during electromagnetic simulations.

Instead of π -mode stabilizing loops (PISLs) usual for RFQs longer than $\sim 3\lambda$, where λ is the rf wavelength [7], FNAL's RFQ design uses the end-wall tuners - field stabilizers simpler than PISLs [8]. This method requires a precise knowledge of dipole mode spectrum, so simulating full length RFQ with end-wall tuners installed was needed.

Modulation of the vanes in the regular accelerating section of the RFQ is shown in Fig1. A variable modulation changes capacitive loading and therefore local frequency along RFQ as also reported elsewhere [4, 9, and 10]. In our RFQ the local frequency variation due to the modulation is significant, so the vane tip modulation has been included in the model.

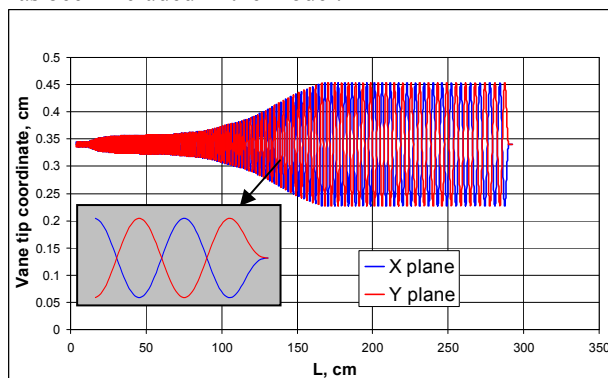


Fig.1 Vane tip modulation along RFQ. Radial matchers are excluded.

The output radial matcher is designed to form axially symmetric beam exiting the RFQ, and because of this special function it is different than the input radial matcher. Fig 2 shows profile of the output radial matcher and imposed profile of the input matcher to compare with. The RFQ ends (cutbacks) can be tuned in simulations individually, but their combined effect on field flatness must be evaluated. Besides the end-wall tuners have

*This work was supported by the U.S. Department of Energy under contract number DE-AC02-76CH03000.

[#]gromanov@fnal.gov.

different tuning range and sensitivity for input and output ends of the RFQ. So again, for proper RFQ end design a full length model had to be considered.

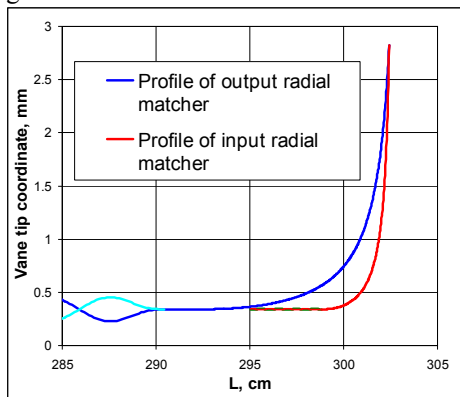


Fig.2 The radial matcher profiles.

The basis of the 3D RF model prepared for simulation was a solid engineering model built in SolidWorks and stored in SAT format. The SAT file was then imported into CST Microwave Studio. CST MWS has its own well developed tools to work with solid models and to heal imported objects, so the RF model was completely prepared directly in MWS environment.

PERIODS OF RFQ

Many basic RFQ parameters can be obtained and defined with 2D approach. CST MWS and HFSS are entirely 3D codes, but RFQ “slices” with thickness of one mesh step were effectively used to define the basic RFQ parameters. Some RF features like PISLs and slug tuners can be studied with periods of RFQ [6]. The period of RFQ defined by the slug tuner spacing was simulated to evaluate slug tuning sensitivity. Also magnetic field distortion around slug tuners was investigated since the bumps created by the tuners were visible at the bead pull axis which was close to the tuners.

The specific vane tip modulation (see Fig.1) attracted attention as a possible reason of the local frequency variation. To check whether the vane tip modulation affects local resonant frequency, one accelerating period with modulated vane tips (see Fig.3) has been simulated. The frequency of the model was found to be 324.7 MHz, while the frequency of the same model with identical average bore radius and no modulation was 323.5 MHz. The difference of 1.2 MHz is significant for such a long RFQ and must be taken into account.

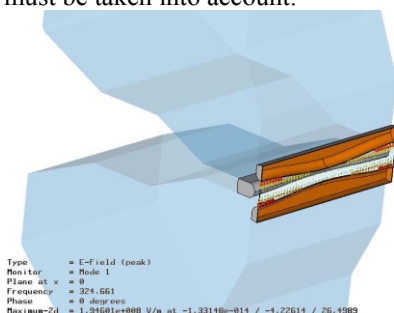


Fig.3. A model of accelerating period #267.

Computer Codes (Design, Simulation, Field Calculation)

RFQ END TUNING

Field flatness is always one of the most important tasks for RFQ tuning. This parameter is very sensitive (especially in long RFQs) to proper vane terminations. The RFQ ends are supposed to be tuned by appropriate geometry and dimensions of undercuts and the errors are usually hard to fix.

In the RF 3D model the end-wall tuners were set to a default value of 25.4 mm of penetration as it was during the initial RF measurements. The tuning slugs were flush with the inner wall of the RFQ cavity, so the slugs were not included in RF model for the vane end tuning. Each RFQ end can be tuned separately, so applying appropriate boundary conditions we used 1/8 of the full model, which is equivalent to full length RFQ with two input or two output ends. Actually the model can be of any length for this tuning, but we prefer to use field flatness as a tuning criterion which is more sensitive for longer cavity. If a local frequency is used as a tuning criterion, a reasonably short model should be used for higher sensitivity.

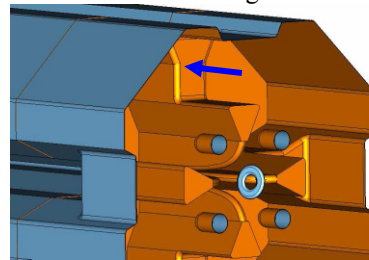


Fig.4. A model of RFQ output end. The back wall (marked by arrow) is moved for tuning.

After main dimensions of cut-backs have been found, the back wall as the most influential parameter was being moved inward (i.e. “removing” of material) for fine tuning (Fig.4). We monitored the electric field distribution of quadrupole mode along RFQ at 4 mm and 45° off the axis (in the gap between tips). The field distribution changes with cut-back variation as it is shown in Fig.5, and it gets flat at optimal cut-back of 65.6 mm.

COMPLETE SIMULATION

Initially the full length model was used to obtain realistic spectrum of quadrupole and dipole modes with end-wall tuners installed, since the exact mode spacing is a key for field stabilization in the RFQ. Then it was realized that the vane tip modulation cannot be ignored, and this feature has been added to the full length model.

The attempts to perform the simulation of the full model with MWS were not successful. It was decided to transfer the problem into HFSS. Using powerful MWS modeler the vane tip modulation was prepared in faceted representation of shape that is more appropriate for triangulated surface meshing in HFSS (see Fig.6).

The model parameters were set in accordance with actual initial RFQ settings during RF measurement just after final assembly [11]. Fig. 7 shows the field distributions as simulated with and without vane modulation in comparison to the previous actual

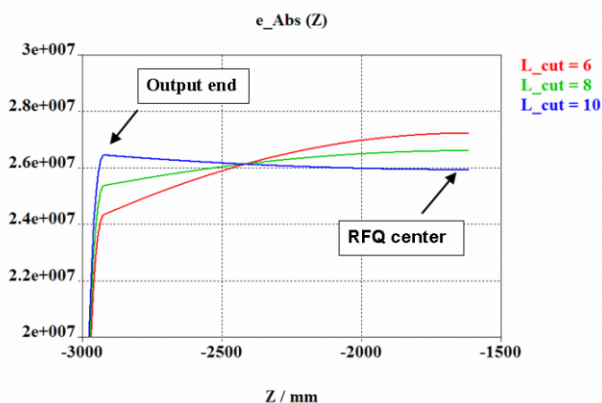


Fig.5 Field distribution with varying cut-back depths.

measurements. With the vane tip modulation included in the model, the simulation reproduces the measurements with high accuracy, including both total field distribution tilt and sinus-like shape. Therefore, the theory that the vane tip modulation is responsible for the additional field distribution distortion is supported. Without modulation, the field tilt in the simulation is due only to the detuned output matcher.

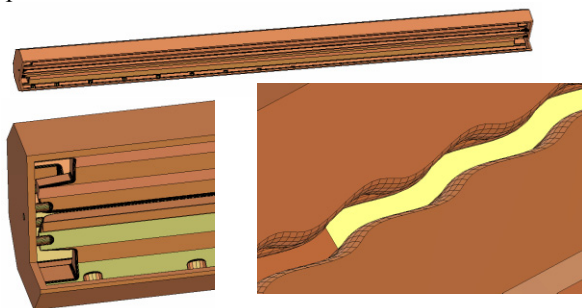


Fig.6. Full length RFQ model with vane tip modulation, radial end matchers, end-wall tuners and slug tuners.

To check final RFQ tuning in simulation, the output matcher cut-back dimensions in the solid model were set to the specified values and slug tuners were introduced. Skipping intermediate simulations of the tuning process with slug tuners, the result is summarized in Fig. 9.

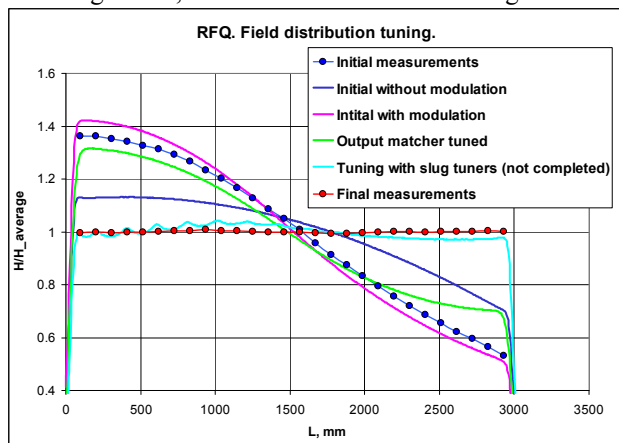


Fig.7 Evolution of field distribution. Lines – simulations, points – measurements. Due to the axial symmetry H-field distributions for only one quadrant are shown.

Computer Codes (Design, Simulation, Field Calculation)

The slug tuners can only increase the local and overall frequency. After field flattening in the simulation, the RFQ frequency was well above required value of 325 MHz. To decrease the initial frequency of the resonator a reduction of the average bore radius by 50-70 μm was recommended after test simulations.

The results of simulations and the recommendations were taken into account during the final vane machining, RFQ assembly and tuning [4].

CONCLUSIONS

In the first time a full size RFQ model with modulated vane tips and all tuners installed has been built, and a complete simulation of RFQ tuning has been performed. Results of this complete simulation are in excellent agreement with the measurements made on the RFQ during initial assembly and tuning. All recommendation and predictions were proven correct by the final tuning results.

This work is an important step toward entirely computational RF modeling on RFQ structures.

ACKNOWLEDGEMENTS

Special thanks to G.Apollinari, T.Page and R.Webber.

REFERENCES

- [1] R.Webber , Overview of the High Intensity Neutrino Source Linac R&D Program at Fermilab, this conference.
- [2] P.Ostroumov et al., Application of a new procedure for design of 325 MHz RFQ, JINST 1 P04002, April 2006.
- [3] G.Apollinari et al., The fabrication and initial testing of the HINS RFQ, this conference.
- [4] D.Li et al., Detailed Modeling of the SNS RFQ Structure with CST Microwave Studio, LINAC'06, THP008, Knoxville, Tennessee, USA, 2006.
- [5] A.Kurup et al., Electromagnetic Design of a Radio Frequency Quadrupole for the Front end Test Stand at RAL, EPAC'06, MOPCH116, Edinburgh, Scotland, 2006.
- [6] A.France et al., Design of Slug Tuners for the SPIRAL2 RFQ, PAC'07, Albuquerque, New Mexico, USA, TUPAN006, 2007.
- [7] A. Ratti et al., Conceptual design of the SNS RFQ, Proc. of the LINAC 98, p. 276.
- [8] L.Young, 25 Years of Technical Advances in RFQ Accelerators, PAC'03, Portland, Oregon, USA, MOPB001, 2003.
- [9] A. Ueno and Y. Kondo, RF-Test of a 324-MHz, 3-MeV, H- RFQ Stabilized with PISL's, LINAC'00, Monterey, California, TUD02, 2000.
- [10] O.Delferriere et al., A New RF Tuning Method for the End Regions of the IPHI 4-vane RFQ, EPAC'06 Edinburgh, Scotland, MOPCH105, 2006.
- [11] G.Romanov, A.Lunin and R.Webber. Recommendations on RFQ Final Tuning, Fermilab, Technical division technical note TD-08-008, February 2008.

H5PartROOT—A VISUALIZATION AND POST-PROCESSING TOOL FOR ACCELERATOR SIMULATIONS

Thomas Schietinger

Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland

Abstract

Modern particle tracking codes with their parallel processing capabilities generate data files of the order of 100 Gigabytes. Thus they make very high demands on file formats and post-processing software. H5PartROOT is a versatile and powerful tool addressing this issue. Based on ROOT, CERN's object-oriented data analysis framework developed for the requirements of the LHC era, and the HDF5 hierarchical data format, supplemented by an accelerator-specific interface called H5Part, H5PartROOT combines the statistical and graphical capabilities of ROOT with the versatility and performance of the HDF5 technology suite to meet the needs of the accelerator community. Providing the user with both a graphical user interface (data browser) and a shared library to be used in an interactive or batch ROOT session, H5PartROOT passes on the full power of ROOT without presupposing any knowledge about the intricacies of either ROOT or C++.

INTRODUCTION AND MOTIVATION

Three-dimensional particle simulations (e.g., OPAL [1]) follow the trajectories of a large number (up to 10^9 and more) of macro-particles through space as they are influenced by external (electro-magnetic and or gravitational) and internal (space charge) fields. The result of such a simulation is typically stored as a sequence of time steps. Each time step contains some quantities (often scalars or 3-vectors for the three spatial dimensions) describing properties of the macro-particle ensemble (bunch) as a whole (e.g., centroid position, mean particle energy etc.) and, if detailed analysis of the bunch evolution is desired, a full dump of the macro-particle phase space (i.e., positions and momenta of all macro-particles in the simulation).

With the increasing size of datasets produced by such simulations, swift post-processing becomes an issue of paramount importance. To address the problem we created a tool based on ROOT, the data analysis software used by CERN and its user community to analyze the vast amounts of data produced by the Large Hadron Collider, and HDF5, an extremely versatile and powerful data file format enjoying growing popularity throughout the scientific community. The considerable power and flexibility of both HDF5 and ROOT come at the prize of rather complex user interfaces. To spare the user the learning curves of these packages as much as possible, we built a ROOT applica-

tion which allows fast extraction of statistical data and generation of publication-quality plots with just a few mouse clicks. Since it makes use of the H5Part interface to HDF5, the application is called *H5PartROOT*.

BUILDING BLOCKS

HDF5

HDF5 (“Hierarchical Data Format 5”) is a highly sophisticated, “self-describing” data storage format. Originally created by NCSA, it is now supported by the HDF group [2]. The HDF5 technology suite allows the management of extremely large and complex data collections. Its versatile data model can represent very complex data objects and a wide variety of meta-data. The file format is completely portable and puts no limits on the number or size of data objects, making it an ideal format for large accelerator simulations. The HDF5 software library provides various high-level interfaces (C, C++, Fortran 90, Java) and runs on almost every computing platform, from laptops to massively parallel systems. Furthermore HDF5 comes with built-in performance features that optimize access time and storage space as well as a whole set of tools and applications for managing, manipulating, viewing, and analyzing the data.

H5Part

H5Part is a thin software layer on top of HDF5 to facilitate I/O for the simulation of particle accelerators (or any other multi-particle system that evolves in time) [3]. Designed as a portable, high-performance parallel data interface to HDF5 [4], it constrains HDF5's very general data format to a subset useful for three-dimensional particle accelerator simulations, i.e., it knows about time steps, phase space variables etc. H5Part is co-developed by LBNL and PSI.

ROOT

ROOT is an object-oriented data-processing framework developed at CERN for the requirements of the LHC era, i.e., to handle complex datasets of sizes measured in Terabytes [5]. Within the last decade, it has become the data analysis and visualization tool of choice in high-energy physics around the world.

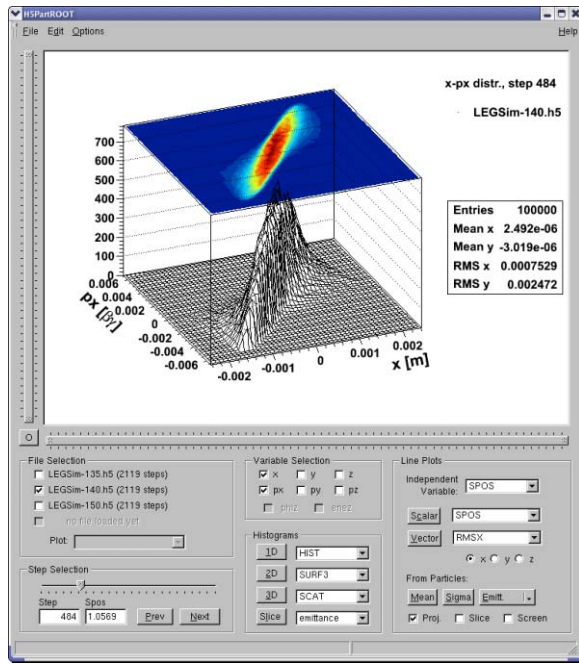


Figure 1: H5PartROOT main GUI showing an example phase space plot.

Written in C++, ROOT can be used either as an extensive class library to be linked to a C++ main program, or via its C++ interpreter (CINT) as an interactive command-line tool. The ROOT libraries include classes for plotting, fitting and other statistical analysis, as well as graphical user interfaces. ROOT is maintained by a core development team at CERN supported by numerous users around the world.

DESIGN OVERVIEW

The driving design philosophy behind H5PartROOT is the desire to pass on to the user the full power of ROOT without bothering him or her with the (sometimes rather cumbersome) intricacies of the ROOT (or generally C++) syntax. Since this is best achieved by a user-friendly graphical user interface (GUI) a main data browsing window is at the center of the H5PartROOT design. The GUI is implemented as a class, TH5MainFrame,¹ which processes all user requests via buttons or pull-down menus and displays information in a graphics window (“canvas”) embedded in the GUI frame. A second class, TH5Dataset, provides the link to a given HDF5 data file, and returns data processed as plots or statistical quantities as requested by the client. Figure 1 shows a screen-shot of H5PartROOT’s main GUI with an example phase-space distribution. By dragging the mouse, the user can rotate the plot or zoom in along the axes (functionalities inherited from ROOT).

While an intuitive GUI is convenient for rapidly browsing one or several datasets, it quickly reaches its limitations

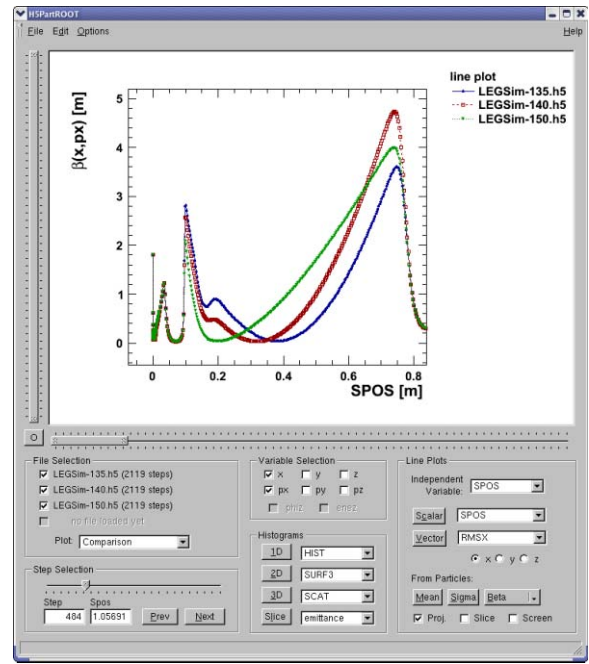


Figure 2: Example of a line plot produced with the H5PartROOT GUI (comparison of beta functions).

when it comes to producing elaborate plots combining information from different sources (e.g., measurement data). For this reason, H5PartROOT can also be used as a shared library within ROOT, which gives the user full access to all TH5Dataset methods from within the ROOT environment.

FEATURES

Besides simple data representations such as histograms and scatter plots of phase-space distributions, or line plots of time-step attributes, H5PartROOT provides a number of features relevant to particle accelerator physics:

- *Emittance:* computation of the rms emittance $\varepsilon(\vec{x}, \vec{x}') = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2}$ for any two phase space variables x, x' from the particle distribution.
- *Twiss parameters and phase space ellipses:* Computation of the rms Twiss parameters $\alpha(\vec{x}, \vec{x}')$, $\beta(\vec{x}, \vec{x}')$ from the particle distribution, graphical representation as a phase space ellipse.
- *Bunch slicing:* Evaluation of the above quantities or graphs for longitudinal slices (subsets of the particle distribution as a function of z) of the bunch (see Fig. 3). Both length and relative position of the slice within the bunch can be set by the user. Gauss-weighted slices are also supported.
- *Screen projection:* Evaluation of emittance and Twiss parameters for fixed longitudinal position s (as opposed to fixed time).

¹Like ROOT itself, H5PartROOT uses the Taligent coding convention.

- *Bunch clipping*: Particle distributions or bunch properties after particles in sparsely populated regions of the bunch have been removed. This feature is typically used to suppress the effect of bunch halos.

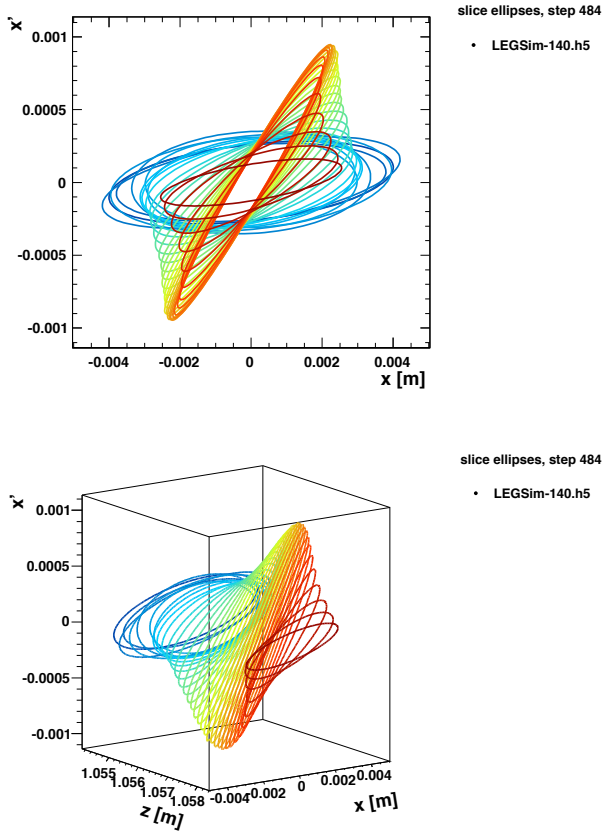


Figure 3: Twiss phase space ellipses for bunch slices in 2D (top) and 3D (bottom) representation.

USAGE

The Graphical User Interface

The main advantage of the GUI consists in the ease with which a user can analyze a dataset and produce publication-quality plots by clicking just a few buttons—without any knowledge of neither ROOT nor C++. The GUI has fields *File Selection*, *Step Selection* and *Variable Selection*, to tell H5PartROOT what to plot and *Histograms* and *Line Plots* to specify how to visualize the data. More sophisticated settings (choice of binning, log scale etc.) are available via an options menu.

Once a plot is done, its embellishments can easily be changed via context menus. Thanks to ROOT's many graphics interfaces, it can then be printed to almost any commonly known graphics format. A particularly useful print format is *C++ Macro*, which will produce the exact C++ code necessary to reproduce the plot inside a ROOT Computer Codes (Design, Simulation, Field Calculation)

session. This gives the user the possibility to save a result without fixing the exact plot style, which may have to be adapted later to the style of a larger document.

The Shared Library

To take full advantage of all ROOT features, however, it is advisable to use H5PartROOT as a shared library. Depending of the complexity of the task at hand this library may be

- loaded into an interactive ROOT session,
- used inside a ROOT macro, or
- linked to a compiled ROOT executable.

Figure 4 shows an example plot generated by the use of the H5PartROOT shared library.

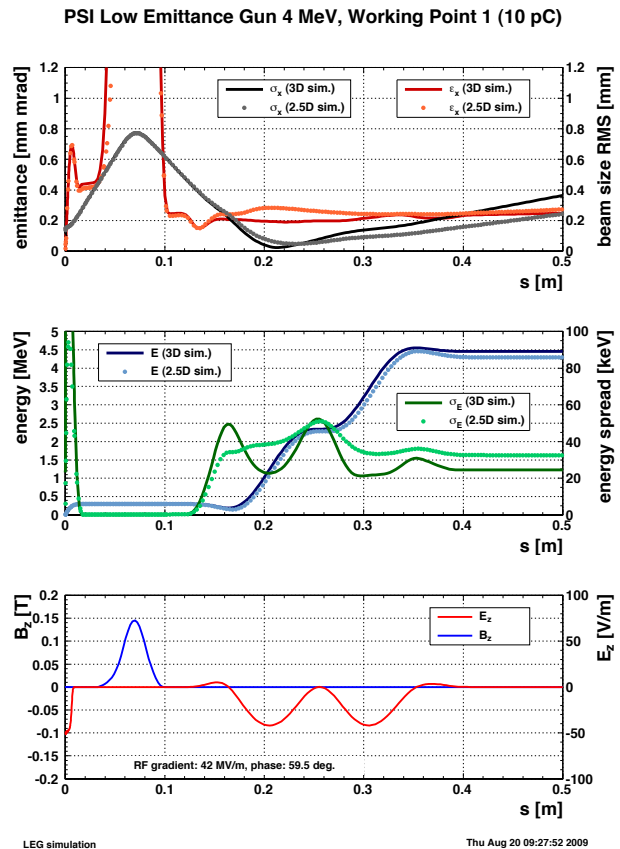


Figure 4: Code comparison plot produced by a ROOT macro after loading the H5PartROOT shared library.

In interactive sessions or macros the library is loaded with the command
`gSystem->Load("<libpath>/libh5root.so");`
 Once this is done, the user has access to the H5PartROOT classes with all their methods:

```
TH5Style::SetStyle()
    — set the H5PartROOT plotting style
```

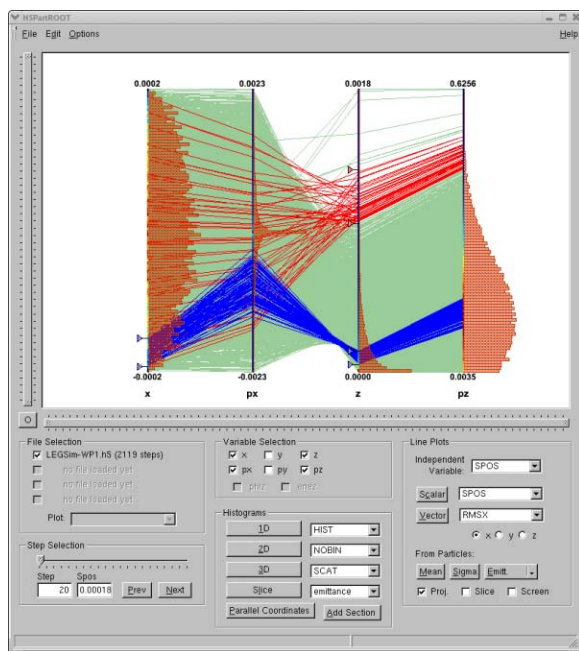


Figure 5: Prototype version of H5PartROOT featuring parallel coordinates. The plot shows a phase space distribution (x, p_x, z, p_z) with one selection in z (red) and another one in x and z (blue).

```
TH5Dataset data("example.h5");
    — load data from file example.h5
data.Histo2d("x", "y");
    — plot x and y (for default step 0, with default range)
```

INSTALLATION

The H5PartROOT source code and instructions on how to install and run the application can be downloaded from our web site [6]. Usage of the GNU build system (auto-tools) ensures straight-forward installation across different platforms.

FUTURE PLANS

Plans to improve and extend H5PartROOT include some more particle selection features in the GUI, better 3D plotting functionality (mainly by integration of the OpenGL graphics engine [7]) as well as new visualization concepts. An example are parallel coordinates, which have recently been added to ROOT [8]. Figure 5 shows a prototype version of the H5PartROOT data browser with parallel coordinates.

Further down the line we plan to adapt H5PartROOT to visualize slice-based simulations such as Homdyn [9]. Another possible development would be the parallelization of the analysis code given sufficient user interest.

Computer Codes (Design, Simulation, Field Calculation)

CONCLUSION

ROOT's excellent data analysis and visualization capabilities, developed for the high-energy physics community, can also be harnessed in the context of computational accelerator physics. In conjunction with HDF5 (and H5Part) it provides a powerful yet elegant solution to the post-processing needs of large-scale particle simulations. With H5PartROOT, we have presented our implementation of this promising approach.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge substantial contributions to H5PartROOT by Andreas Adelmann, Bo Liu and Yves Ineichen, as well as valuable feedback by many other users.

REFERENCES

- [1] A. Adelmann et al., "The Object Oriented Parallel Accelerator Library (OPAL), Design, Implementation & Application", these proceedings.
- [2] The HDF5 Group, <http://www.hdfgroup.org/HDF5/>.
- [3] H5Part: a Portable High Performance Parallel Data Interface to HDF5, <http://vis.lbl.gov/Research/AcceleratorSAPP/>; code available from <https://codeforge.lbl.gov/projects/h5part/>.
- [4] A. Adelmann et al., in Proceedings of Particle Accelerator Conference (PAC), Albuquerque, NM, USA, 2007, p. 3396.
- [5] ROOT: an object-oriented data analysis framework, <http://root.cern.ch>.
- [6] H5PartROOT: a ROOT Based Graphical User Interface to H5Part, <http://amas.web.psi.ch/tools/H5PartROOT>.
- [7] M. Tadel, J. Phys.: Conf. Ser. 119 (2008) 042028.
- [8] O. Couet, J. Phys.: Conf. Ser. 119 (2008) 042007.
- [9] M. Ferrario et al., SLAC-PUB-8400 (2000).

PARALLEL SDDS: A SCIENTIFIC HIGH-PERFORMANCE I/O INTERFACE*

Hairong Shang[†], Yusong Wang, Robert Soliday, Michael Borland, Louis Emery,
Argonne National Laboratory, Argonne, IL 60439, USA

Abstract

Use of SDDS, the Self-Describing Data Sets file protocol and toolkit, has been a great benefit to development of several accelerator simulation codes. However, the serial nature of SDDS was found to be a bottleneck for SDDS-compliant simulation programs such as parallel elegant. A parallel version of SDDS would be expected to yield significant dividends for runs involving large numbers of simulation particles. In this paper, we present a parallel interface for reading and writing SDDS files. This interface is derived from serial SDDS with minimal changes, but defines semantics for parallel access and is tailored for high performance. The underlying parallel I/O is built on MPI-I/O. The performance of parallel SDDS and parallel HDF5 are studied and compared. Our tests indicate better scalability of parallel SDDS compared to HDF5. We see significant I/O performance improvement with this parallel SDDS interface.

INTRODUCTION

SDDS [1] is a self-describing data file protocol developed at Argonne National Laboratory's Advanced Photon Source (APS). It is a standardized way to store and access data, and is the basis of a toolkit [2] of interoperable accelerator physics programs. Over the years, several SDDS-compliant accelerator programs (e.g. *clinchor* [3], *elegant* [4], and *shower* [5]) have been developed at the APS. Also, many existing accelerator design tools for which the source code is available have been converted to read and write SDDS files. This allows physicists to readily use several codes in combination, with greater speed, flexibility, and accuracy than otherwise possible. In addition to requiring accelerator codes to read and write SDDS files, we created a suite of generic data processing and display tools that work with SDDS files. In effect, we created a common pre- and postprocessing toolkit that is used by our codes and codes we have modified. This set of approximately 80 generic programs is referred to as the SDDS Toolkit [2].

A major advantage of using SDDS files is that data from one code can more readily be used by another. The self-describing nature of the files makes this robust, meaning that one code can be upgraded without requiring a change in the other code. The SDDS Toolkit also provides the ability to make transformations of data, which is useful when codes have different conventions (e.g., for phase-

space quantities). Finally, using SDDS means that adding capabilities to a simulation code is faster and easier. The new data is simply placed in SDDS files where it can be accessed with the existing suite of tools [2].

In addition to the SDDS Toolkit, users can import SDDS data directly into programming environments like C/C++, FORTRAN, IDL, Java, MATLAB, and Tcl/Tk, using libraries created and supported by APS. These libraries, like the rest of the SDDS software and our simulation codes, are covered by an Open Source license and are available for download from our web site. The codes discussed are all available for UNIX environments, including LINUX, Solaris, and MAC OS-X, and (usually) for Microsoft Windows. The program *elegant* [4] was the first of the SDDS-compliant accelerator codes, and it is widely used for accelerator design and simulation, and is at the center of the SDDS-compliant accelerator simulation codes. The computing power of *elegant* has been enhanced significantly through recent parallelizations and optimizations [6]. However, the SDDS tools with sequential execution are a bottleneck for both memory and I/O operations. Therefore, parallel SDDS is required for large simulations, as well as for analysis and visualizations of the resulting large data sets. This paper introduces the design, implementation, and performance study on parallel SDDS. Since HDF5 [7] is another popular scientific data format, the performance of parallel HDF5 is also studied on Jazz [8] for comparison. Although HDF5 already supports parallel I/O, it is not necessarily beneficial to switch from SDDS to HDF5, given the large number of programs and applications that already use SDDS. Only if HDF5 offers a significant performance advantage over parallel SDDS would such a conversion be considered.

SDDS File Format and Data Storage

An SDDS file is referred to as a "data set". Each data set consists of an ASCII header describing the data that is stored in the file, followed by zero or more "data pages". The data may be in ASCII or unformatted (i.e., "binary"). Each data page is an instance of the structure defined by the header. That is, while the specific data may vary from page to page, the structure of the data may not. Three types of entities may be present in each page: parameters, arrays, and columns. Each of these may contain data of a single data type, with the choices being long and short integer, single/double-precision floating point, single character, and character string. The names, units, data types, and other descriptions of these entities are defined in the header. Parameters are scalar entities. That is, each parameter defined in the header has a single value for each page. Ar-

* Work supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[†] shang@aps.anl.gov

rays are multidimensional entities with potentially varying numbers of elements. While there is no restriction on the number of dimensions an array may contain, this quantity is fixed throughout the file for each array. However, the size of the array may vary from page to page. All columns in a data set are organized into a single table, called the “tabular data section.” Thus, all columns must contain the same number of entries, that number being the number of rows in the table. There is no restriction on how many rows the tabular data may contain, nor on the mixing of data types in the tabular data. The tabular data is stored in the file by row-major order, which is partly a legacy of SDDS’s origins in the APS control system, where it is used to collect time-series data.

Obviously a column-major ordered data file would be read and written faster since the data is stored as column-major order in the memory. We will soon release in [9] a column-ordered serial SDDS library, which has much better performance over row-ordered serial SDDS library. For parallel I/O, the relative advantage of column-major ordering in data files is not a given since the MPI-I/O can be executed in two modes: independent and collective [10]. In collective mode, all processors pause until they are ready to execute the I/O together. We expect that independent MPI I/O benefits the row-major ordered SDDS files, while collective MPI I/O benefits the column-major ordered SDDS files.

We built parallel SDDS libraries in the four mode combinations of independent I/O or collective I/O and row-major ordered files or column-major ordered files for study and comparison with other implementations of parallel I/O, say, that of HDF5 [7]. In this reference collective I/O has been found to be much more effective than independent I/O for non-contiguous storage, though the authors didn’t specify whether HDF files were column- or row-major ordered.

PARALLEL SDDS IMPLEMENTATION

Parallel SDDS is built on top of MPI-I/O in either independent I/O or collective I/O modes, and derived from serial SDDS with minimal changes. In parallel SDDS, a file is opened, operated on, and closed by the participating processors in a communication group defined by the user interface. Other memory access functions are retained from serial SDDS.

In parallel SDDS, each processor holds the SDDS header data and the column data for only part of the rows, the total number of rows being the sum of the row numbers of all processors.

Similar to serial SDDS, parallel SDDS reads or writes a file page by page. For parallel SDDS page reading, we first read the header using the serial SDDS functions and then close the file. Depending on the input request, either all processors read the header or the master processor reads the header and then broadcasts it to the other processors, which reduces the file I/O load. Next, we use MPI-I/O to open the file and to read the page title information, which are the pa-

rameters (if any), arrays (if any), and the total number of rows (n_t) in the current page. Again, either all processors (n_p in number) read the title information or the master processor reads it and then broadcasts it to other processors if requested. Finally, each processor reads $n_t/n_p + r$, where r is 1 if the processor ID is less than or equal to $n_t \bmod n_p$, otherwise, 0.

For parallel SDDS page writing, all processors hold the layout information that is defined by the existing serial SDDS functions, and part of the tabular data partitioned by row. The file is opened for write with MPI-I/O. Only the master processor writes the ASCII layout, parameters (if any), arrays (if any), and the number of total rows, and then its own part of tabular data into the file. Other processors write their own part of the tabular data into the file at the same time.

PERFORMANCE COMPARISON

In this section we look primarily at row-major order SDDS library performance compared with row-major order HDF5, thus for clarity the term row-major order for HDF and SDDS is dropped.

In the SDDS test code all processors read the header, number of rows, parameters, and arrays in each page. Parallel SDDS was compiled with MPICH1 on ANL Jazz and the performance was studied with PVFS version 1 file system. There are 8 PVFS parallel file systems on Jazz running over 10/100 Ethernet. The theoretical peak I/O rate is 10 MB/sec per node.

In order to fairly compare parallel SDDS with HDF5, the parallel HDF5 write/read code (ph5example.c), which comes along with the parallel HDF5 package, was compiled with the same compiler used for parallel SDDS.

Reading Performance

Two HDF5 row-major-ordered data files were generated using ph5example, with sizes of 1.2GB (1245710336B) and 600MB (622856192B). Each file has one two-dimensional dataset, with dimension 811008x384 for the 1.2GB file and 811008x192 for the 600MB file. The dimensions are chosen by the requirement of ph5example that all dimensions must be a multiple of the number of processors. Here, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64 processors are used for performance study. However, SDDS does not have any limitations on the dimension sizes. The two HDF5 files were converted into two SDDS files using our hdf2sdds toolkit program. The SDDS file sizes were 1245722085B and 622861029B respectively, which are slightly bigger than the HDF5 files because the SDDS header is written in ASCII, and there are many columns in both files (making the header large). But in actual applications such as Pelegant [6], the SDDS files have only 8 columns, which produces less overhead than a HDF5 file would. The read performance of both parallel SDDS and parallel HDF5 was studied with the PVFS version 1 file

system on Jazz. The results of reading the two files are shown in Figure 1.

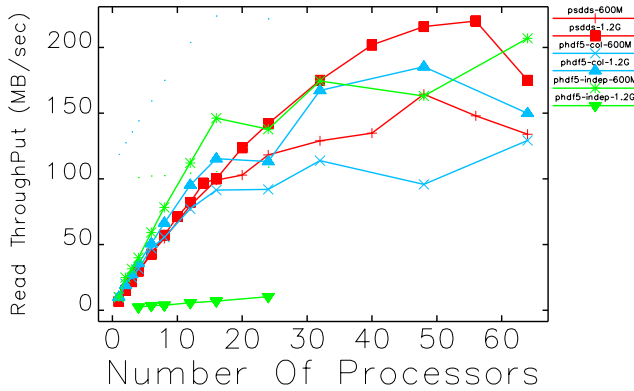


Figure 1: Parallel SDDS (psdds) and parallel HDF5 (phdf) I/O performance of reading 600MB and 1.2GB files on Jazz.

The figure includes a comparison of collective and independent I/O for HDF5. We didn't have the collective-I/O SDDS library run available for comparison. Figure 1 shows that independent HDF5 I/O has better performance than collective HDF5 for reading the 600MB file. In addition, independent HDF5 has better performance than parallel SDDS when the number of processors is less than 32. However, the speed of HDF5 starts drop after 32, and its performance is similar to parallel SDDS after that. The speed of independent-I/O SDDS continues increasing until the number of processors reaches 48 and then starts to drop. This may indicate that independent-I/O SDDS has better scalability than collective-I/O HDF5. Apparently, collective-I/O HDF5 is not a good choice for reading such a 600MB row-major ordered file.

However, the performance of independent-I/O HDF5 in reading a 1.2G file is so poor that our performance study could not be completed with available sources. It is much worse than collective-I/O HDF5, which is consistent with the results of parallel HDF5 [7]. The performance of collective-I/O HDF5 is slightly better than independent-I/O SDDS when the number of processors is less than 20. However, the performance of this SDDS library is consistently better than collective-I/O HDF5 when the number of processors is greater than 20.

Data access performance is affected by many factors, including caching, network bandwidth, and latency. Jazz has two kinds of nodes, large memory nodes, which have 2.4GB memory, and smaller memory node which have 1.2GB memory. The network bandwidth is 10 MB/s. The bandwidth per processor achieved by collective-I/O HDF5 is close to 10 MB/s with a small number of processors. However, it drops quickly to 3 MB/s as the number of processors increases. The bandwidth of parallel SDDS is about 6 MB/s from 1 processors to 56 processors, and drops

at 64 processors. The relatively low efficiency of SDDS at a low number of processors compared to parallel HDF5 may have two causes: First, reading SDDS data requires at least two times as much memory as the data size because of the way SDDS encapsulates the data. Therefore the nodes may not have enough memory to hold the data and swap space may be needed when the number of processors is small. Second, all processors read the SDDS layout at the same time using serial code. Therefore, the time spent in layout reading increases as the number of processors increases, which reduces the speed when the file header is big (as in our test files) and the number of processors is large. For example, the time to read the 1.2GB file header with one processor is 0.01 seconds, but increases to 2 seconds with 64 processors, while the data access time is only 3 seconds. The layout reading could be improved in the future.

Still, the results indicate that independent-I/O SDDS has better scalability than HDF of either I/O mode, and has better performance with large files. This may be due to the relatively simpler structure of SDDS data compared to HDF5.

Writing Performance

The writing performance of parallel HDF5 and SDDS was studied when writing 811008x192 and 811008x384 two-dimensional datasets into HDF5 files or SDDS files. Both collective-I/O and independent-I/O HDF5 writing were tested. Again results for collective-I/O SDDS was not available. The performance of parallel SDDS writing was studied by reading a previously generated SDDS file of 811008x192 data or 11008x384 data into an SDDS dataset, copying it into a new dataset in memory, and then writing the new dataset into an SDDS file. This doubles the memory size for storing two SDDS datasets in memory, so that memory requirements are more than four times the size of the data file. The purpose of copying in testing parallel SDDS writing is to verify that the write operation produces a file that is identical to the original (which was the case in all tests). The performance of parallel SDDS writing may be improved when writing data that is generated internally.

The results of writing files are shown in Figure 2. Unlike what we found for reading, independent-I/O HDF has better performance than collective-I/O HDF for writing the row-major ordered HDF5 file. Independent-I/O SDDS is also better than collective-I/O HDF. Similar to reading, independent-I/O HDF5 performs better than independent-I/O SDDS with a small number of processors, but as the number of processors increases, independent-I/O SDDS starts to perform better than independent-I/O HDF5 for writing both 600MB and 1.2GB files. The results again indicate that parallel SDDS has better scalability than HDF5 and better performance with large files in our tests.

Since only one processor writes the layout, the time spent in writing layout does not increase as the number of processors increases. However, the layout writing can be improved by buffered I/O, since right now each definition uses a separate write operation, and the I/O times are the

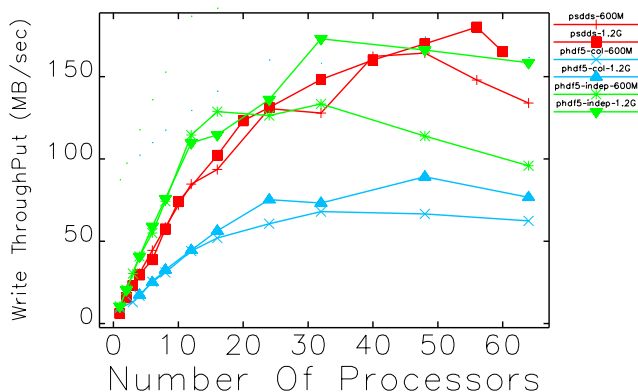


Figure 2: Parallel SDDS and parallel HDF5 I/O performance of writing 600MB and 1.2GB files on Jazz.

sum of number of parameters, arrays and columns, plus the time required to write other (generally small) parts of the SDDS header. Buffering could reduce this by a significant factor.

IMPROVEMENTS

We made further improvements in parallel SDDS that include 1) changing the header reading strategy so that only one processor reads the layout information, parameters, and total number of rows, and then broadcasts this information; 2) using buffered I/O for writing the layout, parameters, arrays, and the number of rows, and for reading parameters, arrays, and the total number of rows; and 3) parallel reading and writing of SDDS in column-major order.

Since the collective I/O seems to have better performance on the GPFS file system, we also implemented collective-I/O row-major SDDS. The performance was studied on the Intrepid (IBM Blue Gene/P) GPFS file system [11] with reading/writing a 2.4GB file. We have no performance of HDF5 on Intrepid due to lack of time. The results are as follows.

As expected, collective I/O does not benefit row-major SDDS data. But it does benefit the column-major SDDS data, especially in writing. The writing performance of column-major SDDS data is 1GB/s with 350 processors, which is close to the theoretical throughput (1GB/s for 320 processors).

Similar to the Jazz PVFS system, independent I/O row-major SDDS shows good performance on GPFS in both reading and writing. The maximum reading throughput is 600MB/s and the writing throughput is 370MB/s.

CONCLUSION

In this work, we implemented a parallel SDDS interface with independent I/O and completed a performance study of parallel SDDS and parallel HDF5 on Jazz with Computer Codes (Design, Simulation, Field Calculation)

PVFS version 1 file system based MPICH1 MPI-I/O. Parallel SDDS (for row-major ordered files) was found to have better scalability than HDF5 on a PVFS file system and better performance with large files. We also implemented parallel SDDS with independent I/O and collective I/O for row-major and column-major SDDS data, and studied the performance on the Intrepid (Blue Gene P) GPFS file system. The results show that collective writing of column-major ordered SDDS data reaches the theoretical throughput of the I/O nodes. Independent-I/O SDDS, which is currently being used in parallel applications such as Pelegant [12], shows good performance for both reading and writing row-major-ordered SDDS data.

ACKNOWLEDGMENTS

The ComPASS project is supported under the SciDAC program by the U.S. Department of Energy Office of High Energy Physics, Office of Nuclear Physics, Office of Basic Energy Sciences, and Office of Advanced Scientific Computing Research.

REFERENCES

- [1] M. Borland, "A Self-Describing File Protocol for Simulation Integration and Shared Postprocessors," Proc. of PAC95, Dallas, Texas, 2184 (1996); www.jacow.org.
- [2] M. Borland et al., "SDDS-Based Software Tools for Accelerator Design", Proc. of PAC03, Portland, Oregon, 3461 (2003); www.jacow.org.
- [3] L. Emery, "Required Cavity HOM deQing Calculated from Probability Estimates of Coupled Bunch Instabilities in the APS Ring," Proc. of PAC93, Dallas, Texas, 3360 (1993), www.jacow.org.
- [4] M. Borland, "elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation," Advanced Photon Source Note LS-287, September 2000.
- [5] L. Emery, "Beam Simulation and Radiation Dose Calculation at the Advanced Photon Source with shower, an Interface Program to the EGS4 Code System," Proc. of PAC96, Vancouver BC, 2309 (1996); www.jacow.org.
- [6] Y. Wang, M. Borland, Proc. of PAC07, Albuquerque, New Mexico, 3444 (2007); www.jacow.org.
- [7] C. Chilan et al., "Parallel I/O Performance Study with HDF5, A Scientific Data Package," www.spsscicomp.org/ScicomP12/Presentations/User/Yang.pdf
- [8] <http://www.lcrn.anl.gov/jazz/Documentation/index.php>
- [9] http://www.aps.anl.gov/Accelerator_Systems_Division/Operations_Analysis/oagSoftware.shtml
- [10] W. Gropp, E. Lusk, and R. Thakur, "Using MPI-2: Advanced Features of the Message-Passing Interface," MIT press, Cambridge, MA, 199.
- [11] https://wiki.alcf.anl.gov/index.php/File_Systems/
- [12] Y. Wang et al., these proceedings.

THE PYTHON SHELL FOR THE ORBIT CODE *

A. Shishlo, J. Holmes, T. Gorlov, ORNL, Oak Ridge, TN 37831, U.S.A.

Abstract

A development of a Python driver shell for the ORBIT simulation code is presented. The original ORBIT code uses the SuperCode shell to organize accelerator-related simulations. It is outdated, unsupported, and it is an obstacle to future code development. The necessity and consequences of replacing the old shell language are discussed. A set of core modules and extensions that are currently in PyORBIT are presented. They include particle containers, parsers for MAD and SAD lattice files, a Python wrapper for MPI libraries, space charge calculators, TEAPOT trackers, and a laser stripping extension module.

INTRODUCTION

The original ORBIT code has been very useful in the SNS ring design and in simulations of collective effects [1]. Thanks to a flexible structure, ORBIT can be extended very easily. After years of development by many scientists, ORBIT includes collimation, different types of space charge, impedances, electron-cloud effects, and numerous other features. These features are combined together by using a driving shell – the SuperCode (SC). SC is an interpreter programming language resembling C. At the time when ORBIT development started (1997), there were not many choices of driving shell language. SC was attractive because it is C-like, it is simple to learn, to understand, and to extend, and it has a set of effective auxiliary classes for arrays, vectors, strings, etc. As a result of deep integration, the ORBIT code has become inseparable from SuperCode, and SC has now become an obstacle to further ORBIT development.

There are several problems related to SC. First, SC is not an object-oriented language. This significantly slows down ORBIT development and limits the functionality of the code. All contemporary interpreters are object-oriented. Second, SC is not supported by anyone. Usually languages are surrounded by a community of users and developers, which facilitates an immediate response to problems and bugs. So, for SC the user is on his own. Finally, all auxiliary classes provided by SC have been implemented in the C++ Standard Template Library, and this implementation is probably more efficient. In SC none of these classes is protected by namespaces, and they could crush the ORBIT code compilation if there is a name conflict.

In an attempt to preserve the legacy of the ORBIT code, the PyORBIT project has been started. The motivation of PyORBIT is to replace the SuperCode driver shell by a modern interpreter language, Python [2]. Unfortunately, it is not possible to directly import the code of core ORBIT modules into the new project

because of ubiquitous SC dependencies. On the other hand, this gives us an opportunity to start from scratch in the architecture and the source code development and to keep all original ORBIT physical algorithms.

DRIVER SHELL PARADIGM

PyORBIT, like the original ORBIT code, uses a driver shell language approach instead of an input file analysis, as in traditional accelerator codes like MAD, MAD-X, PTC, PARMILA, Trace3D etc. These traditional codes construct an accelerator lattice and perform calculations according to information inside specialized input files. They each use their own language created for the particular code, and the list of possible tasks is predefined and limited. PyORBIT uses another approach. We use an existing programming language and extend it with specific accelerator-related functionalities. The user can create a unique simulation code in the form of a main program or script by using a predefined set of classes and methods.

There are several requisites for a programming language that can be used for this scheme:

- The program language should be popular among physicists. There are many languages that fall under this category: FORTRAN, C, C++, Ruby, Python, and Java.
- It should be an object-oriented language with an automated garbage collection. This condition eliminates FORTRAN, C, and C++.
- It should be fast. That will eliminate Ruby and Python, which are interpreted languages.
- It should be capable of an effective usage of the Message Passing Interface (MPI) library for parallel calculations. That will remove our last candidate – Java. There are several available Java wrappers for MPI, but the overhead for array exchange makes these packages unacceptable for us.

These constraints necessitate a two-language scheme. To provide the necessary speed we must use FORTRAN, C, or C++ at the low level, and Ruby or Python to organize the calculation at the upper level. For the PyORBIT project we chose C++ for its object-oriented nature, better standardization, and better free compiler availability than FORTRAN. For the upper level we preferred Python, because its pseudo-code compilation feature makes it significantly faster than Ruby. This combination of a scripting language for orchestrating simulations and a fast compilation language to perform calculations is very popular in scientific computing [3]. Generally, code development in a scripting language is considered 3-5-10 times faster than it is in languages like C++ or Java. The downside of the two-level approach is the necessity of a “glue” code to connect the codes in the two languages.

* ORNL/SNS is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725

PYORBIT CODE STRUCTURE

The directory structure of the PyORBIT code is shown in Fig. 1. PyORBIT consists of three main parts: a core, extensions, and pure Python classes.

The core includes C++ classes and wrappers for them. The wrappers define the Python user interfaces for underlying C++ classes. After compilation of the core source code and linking with the Python language static library and MPI libraries, the PyORBIT executable is placed in the “bin” directory (see Fig. 1). This executable is an extended Python language interpreter that has all the functionality of Python and that can dynamically operate with classes and methods from the core and extensions.

The extensions are independent packages that have no common classes. If two or more extensions use the same class, it should be moved to the core of PyORBIT. Each extension package is dedicated to some particular physical phenomena. At this moment PyORBIT has only one extension, a package that simulates different aspects of the laser-assisted stripping of H^+ ions [4]. The extensions are compiled into shared libraries and are placed into the “lib” directory. The libraries are dynamically loaded when invoked in the user’s Python script.

The pure Python classes’ directory in Fig. 1 has two subdirectories: one for core and one for extensions.

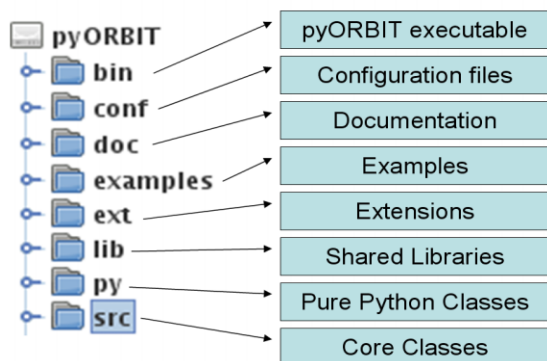


Figure 1: The directory tree of the PyORBIT code.

Today the C++ core of PyORBIT includes four components important for the future development: the MPI wrapper, the PyORBIT Bunch class, the TEAPOT elements library, and the 2D space charge package. The pure Python core components contain the accelerator lattice model, the TEAPOT-like implementation of the accelerator lattice, and the parsers for input files of MAD and SAD accelerator codes. Below we discuss these components.

PYTHON AND C++ CLASSES

To connect the Python and C++ levels, we want flexibility and full control and of our logic flow. Therefore, PyORBIT does not use an automated approach that extends the Python language with C++ classes. To create a wrapper class for a C++ class we follow the standard method described in the “Defining New Types” part of the Python documentation. Each wrapper class Computer Codes (Design, Simulation, Field Calculation)

inherits from a PyORBIT_Object class that extends the standard PyObject with one void pointer to the wrapped C++ class instance. In turn each C++ class inherits from the CppPyWrapper class that keeps a reference to the Python wrapping object. This cross-reference scheme allows access to Python and C++ objects from any level, and it is used everywhere in PyORBIT except for the MPI library wrapper, because MPI is a collection of functions, not classes.

PYORBIT MPI WRAPPER

From the beginning, PyORBIT was developed as a parallel code based on MPI. At the same time all parallel features can be switched off if the user wants to use only one CPU. To provide this functionality, PyORBIT has the MPI wrapper, which isolates the standard MPI functions from the rest of PyORBIT. It accomplishes this by wrapping the MPI functions into functions with different names, but the same signature, and exposes these wrappers to the Python level. At this moment 45 MPI functions are available from the Python script level. In addition to the MPI functions, the MPI wrapper also transforms the MPI communicators, groups, operations, and the MPI status to PyObjects that can be accessed from the Python and C++ levels. The ability to move MPI objects between Python and C++ was the main reason to create our own MPI wrapper instead of using one of the available open sources. It is expected that MPI on the Python level will be used only to perform small data exchange and to create necessary MPI communicators which later will be used on the C++ level for fast and massive data exchanges.

The PyORBIT MPI wrapper package is completely independent from the rest of the PyORBIT code and can be extracted and used anywhere.

BUNCH CLASS

PyORBIT is a particle tracking code, so a class representing a container for particles is the most important class of the code. The Bunch class of the PyORBIT core is this container class. By default it keeps 6D coordinates and one flag specifying an “alive/dead” status for each particle, and it has the following features:

- It is dynamic. The user can add or remove particles from this container. Its size will adjust to the number of particles.
- It is efficient. It provides fast access to the coordinates and it maintains spare room to accommodate additional particles without frequent memory resizing.
- It is extendable. The user can dynamically assign additional properties to each particle in the Bunch. This allows the Bunch class to be used for different kinds of physical problems. For instance, this additional information could be a macro-size of the particle, its spin, or amplitudes of different quantum states, as for the hydrogen atom in the Laser Stripping PyORBIT extension [4]. The possibilities

are numerous. The absence of this kind of extensibility is a big drawback in the original ORBIT code.

- It can be dumped and restored from a file.
- It has parallel capabilities. It automatically distributes particles among CPUs in its local communicator when it restores a bunch from a file.
- All methods of the C++ implementation are exposed to the Python level.

The additional information that can be attached to each particle in the bunch should be stored as a double array of the predefined length. This condition limits the user freedom, but it provides a fast way to exchange particle information between CPUs in parallel calculations. Still, this approach is general enough to be acceptable for all physical phenomena that we have in mind right now.

As said before, the Bunch class has 6 phase-space coordinates for each macro-particle. The Bunch class does not define the meaning of these coordinates, and it is up to the user to define them. In the TEAPOT-like tracking we follow the original ORBIT. We consider them as transverse displacements and angles for the transverse plane, and as position and energy deviation from the design energy for the longitudinal direction. However, in PyORBIT we change transverse units to meters and radians from millimetres and milliradians in ORBIT. The longitudinal position is also given in meters instead of radians.

ACCELERATOR LATTICE PACKAGE

The accelerator lattice package is a lightweight pure Python implementation of a structure shown in Fig. 2. The package includes three classes: the Accelerator Lattice class, the Accelerator Node class, and the Action Container class.

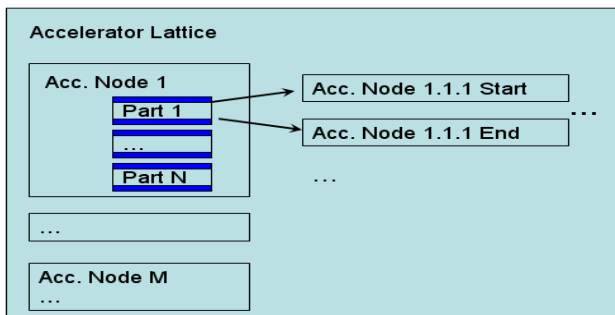


Figure 2: The PyORBIT accelerator lattice structure.

The Accelerator Lattice class is a container of the instances of the Accelerator Node class (nodes). The lattice class has methods to get the length of the lattice, to add a new accelerator node at any place in the lattice, to create a sub-lattice from the existing one, and to call the “trackAction” method for each accelerator node. This method accepts two objects: the instance of Action Containers and a dictionary with user parameters. The lattice puts into the parameters dictionary two references, one to itself and one to the accelerator node.

The Accelerator Node represents a single node in the lattice and is built according to E. Forest’s concept of “fibre bundle” [5]. Each part of the node is a container for references to child Accelerator Nodes. When the lattice calls for the “trackAction” method of the node, this method is performed recursively for each child node and executes actions that are in the Action Container. The user should consider the Accelerator Node class as an abstract class for subclasses that will perform meaningful actions.

The Action Container class is a keeper of user-defined functions (actions) that will be called upon entering the node, at each part of the node, and at the exit of the node. By default this container is empty, and it is up to the user to supply the calculations and their order inside the container. The Python mechanism of lexical closures allows one to define such actions in the source code of the class methods.

The accelerator lattice package is a very flexible construction that can accommodate almost any kind of functionality, but there is no restriction in PyORBIT to prevent other approaches to defining a model for the accelerator lattice.

TEAPOT-LIKE ACCELERATOR LATTICE

There are two PyORBIT components that enable TEAPOT-like tracking of the macro-particles. The first component is the collection of C++ functions that were developed for the original ORBIT to track 6D coordinates of charged macro-particles through simple accelerator elements including dipoles, drifts, quads, multipoles, solenoids, kickers, etc. These functions were thoroughly benchmarked against analytical models, and their source code was directly imported into PyORBIT. The second part is a collection of pure Python classes that are subclasses of the Accelerator Node class. These classes keep parameters of the nodes and call the C++ TEAPOT tracking functions.

The TEAPOT lattice can be built right in the script by adding accelerator nodes one by one or by analyzing a MAD input file. PyORBIT includes a MAD parser that can read a MAD input file specifying the structure of the accelerator. The parser is discussed below.

At present, the TEAPOT-like lattice does not have such nodes as a foil injection node, collimators, space charge nodes, diagnostics nodes, etc. that are in the original ORBIT code. We plan to import them into PyORBIT in the near future.

MAD-FILE PARSER

The original ORBIT does not have any internal tools to use MAD input files directly. The user has to run MAD to get MAD output files with transfer matrices and Twiss parameters that will be used by ORBIT. This dependency is one of the weak points of the ORBIT code. For PyORBIT the pure Python MAD parser is developed. It is completely independent from the rest of the code, and it is used as a generator for the TEAPOT-like input file for PyORBIT. The parser can perform mathematical calculations defined in the MAD file and can handle

insertions of external files. PyORBIT also has a modification of this parser for the SAD code input files. The SAD code is the accelerator design and simulation code developed by KEK accelerator theoretical group [8].

RUNGE-KUTTA TRACKER

In addition to the TEAPOT-like tracking, PyORBIT has a package to solve the equation of motion with arbitrary electric and magnetic fields

$$d\vec{p}/dt = q \cdot (\vec{E} + \vec{v} \times \vec{B}) \quad (1)$$

The package uses the Runge Kutta 4-th order (RK4) solver. The user must specify both electric and magnetic fields as functions of position and time. For prototyping, this can be done on the Python level, but the speed of calculations will be very slow. The user also can attach a custom implementation of the External Effects class. The user has to define the “applyEffects” method of this class which will be called at each time step of the RK4 solver. This allows the user to specify other things that can happen to the macro-particles during their motion through the electromagnetic field region. For instance, there may be decay, ionization, excitation, or interaction with a collimator material. In the laser-stripping package [4] this tracker is used to simulate the dynamics of the internal states of hydrogen atoms in the laser field.

2D SPACE CHARGE SOLVER

As a base for future space charge modules PyORBIT has a FFT-based Poisson solver package. The package includes three classes: Grid2D, PoissonSolver, and Boundary classes. The Grid2D class represents the two dimensional rectangular grid with a space charge density or the electrostatic potential. The PoissonSolver class calculates the potential on the grid

$$\phi_0(\vec{r}) = \int \frac{\rho(\vec{r} - \vec{r}') \cdot d\vec{r}'}{|\vec{r} - \vec{r}'|^2} \quad (2)$$

by using the Fourier convolution theorem and discrete transformation (FFT) [9].

The Boundary class is a container of arbitrary boundary points inside the defined grid. It modifies the potential on the grid by adding the potential in empty space [10]

$$\phi_{empty}(r, \theta) = \sum_{n=0}^N r^n [a_n \cos(n\theta) + b_n \sin(n\theta)] \quad (3)$$

where r and θ are the usual polar coordinates, N is a user defined maximum number of harmonics. The

coefficients a_n and b_n are found by minimizing the sum of potentials (2) and (3) at the boundary points in a least squares sense. The sum of two potentials is the solution of Poisson's equation with zero potential on boundary points. The number of boundary points and the number of harmonics determine the accuracy of the solution.

Again, this package is relatively independent from the rest of the PyORBIT code and can be used separately.

CONCLUSIONS

At present, PyORBIT does not have the full collection of physics modules of the original ORBIT code, but it has all the basic components to accommodate these modules. The new capabilities of PyORBIT include the customizable Bunch container, which provides the means to simulate a broader spectrum of physical problems. A fine example of these extended capabilities is the Laser Assisted Stripping module developed inside PyORBIT [4].

REFERENCES

- [1] A. Shishlo, S. Cousineau, V. Danilov, J. Galambos, S. Henderson, J. Holmes, M. Plum, “The ORBIT Simulation Code: Benchmarking and Applications”, ICAP 2006, Chamonix Mont-Blanc, France, 2-6 Oct 2006, pp. 53-58
- [2] <http://python.org>
- [3] <http://www.scipy.org>
- [4] T. Gorlov, A. Shishlo, “Laser stripping computing with the Python ORBIT code.” These proceedings.
- [5] <http://docs.python.org/extending>
- [6] E. Forest et al., “Polymorphic Tracking Code PTC,” KEK Report 2002-3
- [7] [http://en.wikipedia.org/wiki/Closure_\(computer_science\)](http://en.wikipedia.org/wiki/Closure_(computer_science))
- [8] H. Hirata, in Proceedings of the Second Advanced ICFA Beam Dynamics Workshop, Lugano, Switzerland, 1988 (CERN, Geneva, 1988), p. 62.
- [9] R. W. Hockney and J. W. Eastwood, Computer Simulation Using Particles, Institute of Physics Publishing (Bristol: 1988).
- [10] F. W. Jones, A Method for incorporating image forces in multiparticle tracking with space chargein, Proceedings of EPAC2000, (Vienna, 2000) 1381.

RECENT PROGRESS ON PARALLEL ELEGANT *

Y. Wang [†], M. Borland, H. Shang, R. Soliday, A. Xiao, ANL, Argonne, IL 60439, USA

Abstract

The electron accelerator simulation software `elegant` [1] is being parallelized in a multi-year effort. Recent developments include parallelization of input/output (I/O), frequency map analysis, dynamic aperture search, and position-dependent momentum aperture determination. Parallel frequency map, momentum aperture analysis, and dynamic aperture search provide rapid turnaround for important determinants of storage ring performance. The development of parallel Self-Describing Data Sets file (SDDS) I/O based on MPI-IO made it possible for parallel `elegant` (`Pelegant`) to take advantage of parallel I/O. Compared with previous versions of `Pelegant` with serial I/O, the new version not only enhances the I/O throughput with good scalability but also provides a feasible way to run simulations with a very large number of particles (e.g., 1 billion particles) by eliminating the memory bottleneck on the master with serial I/O. Another benefit of using parallel I/O is reducing the communication overhead significantly for the tracking of diagnostic optical elements, where the particle information has to be gathered to the master for serial I/O.

INTRODUCTION

The parallel version of `elegant`, `Pelegant`, has proved to be very beneficial to several computationally intensive accelerator research projects. Simulation with a very large number of particles is essential to study detailed performance of advanced accelerators. This was demonstrated in simulations of microbunching for FERMI [2]. In those simulations the maximum number of particles was reached at about 60M when the serial version of SDDS was used, which limited our ability to probe microbunching effects at shorter wavelengths. In the version of `Pelegant` used in those studies, the bottleneck came from the memory usage of the master CPU, which was required to hold all the particle information when simulating a diagnostic element, such as a watch point, where all the particles have to be gathered to master to be written on the disk.

The recent development of parallel SDDS [3] makes it possible for `Pelegant` to take advantage of parallel I/O through MPICH2 [4]. With parallel I/O, a common file is opened by all the processors, but each processor is only

responsible for reading/writing the particles allocated to it. This technique improved I/O throughput significantly, especially on some parallel file systems, such as Parallel Virtual File System (PVFS) [5] and General Parallel File System (GPFS) [6]. `Pelegant` can also run on Network File System (NFS) file system, although the I/O performance is not as good as on the parallel file systems. The overall performance of `Pelegant` on all the file systems mentioned above has also been improved due to reduced communication overhead compared with gathering particles to master before writing to the disk with serial I/O. A nice feature of this parallel SDDS I/O is that the output/input files are the same as the files for serial I/O, which is very convenient for data analysis and exchanging data in SDDS format with other related simulation programs.

In this paper, we first describe the effort we made to integrate parallel SDDS with `Pelegant`, then we report the progress made on the parallelization of frequency map, momentum aperture analysis, and dynamic aperture optimization in `Pelegant`.

IMPLEMENTATION OF PELEGANT WITH PARALLEL SDDS

The simulation code `elegant` is being gradually parallelized with particle-based domain decomposition to reduce the simulation time for multi-particle beams. Beamline elements are classified in the element dictionary as parallel-capable or serial-only. Particles will be gathered to the master CPU or scattered to slave CPUs when the beam encounters a serial element or a parallelized element [7, 8], respectively. As the majority of the frequently used elements has been parallelized, `Pelegant` has been efficiently used for several important accelerator research projects [2, 9, 10, 11, 12].

Even in cases where one must use beamline elements that are not yet parallel-capable, a very significant performance improvement can be realized. However, for simulations with large numbers of particles, I/O for input, intermediate output, and final output, can consume a significant portion of simulation time. In addition to the communication overhead of gathering particles to the master, memory also becomes a problem when we simulate a very large number of particles with a central process (i.e., Master) holding all the particle information for I/O operations.

To eliminate these bottlenecks, we developed parallel SDDS [3] with MPI-IO recently. The parallel SDDS is derived from the serial version of SDDS [13], which has been successfully applied to several accelerator simulation

* Work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[†] ywang25@aps.anl.gov

codes. We made significant efforts to integrate Pelegant with parallel SDDS:

1. For large-scale simulation, if one element in the lattice is not parallelized, the gather-scatter procedure will downgrade the performance of Pelegant, and the memory problem will appear. So we parallelized almost all of frequently used elements. A warning message will be given whenever a serial element is tracked.

2. Besides the final output of the simulation, some intermediate outputs from elements have to be adapted with parallel I/O, which was previously done by the master. Such I/O becomes impossible for simulations with extremely large numbers of particles.

3. The statistics information for the particles allocated to each slave will be calculated locally first, then the master will calculate the global statistics results based on the information provided by the slaves. Both the communication overhead and computation load on master are significantly reduced compared with previous versions, where the master gathers all the particles to perform the statistics calculation. The parallelization of the statistics calculation is not straightforward, as elegant provides more than 100 statistical parameters for the properties of the output beam. Parallelization of these statistical calculations requires a number of strategies to meet the different requirements of the specific statistic.

4. Validating results through numerical comparison with the serial version result is a challenge when simulating with a randomly generated beam, as the number of random numbers generated on each of the CPUs is unpredictable, especially for a beam with some cut-off criteria in more than one dimension. elegant provides several different types of beam generation. For certain types of beam distributions, we were able to make elegant generate the same random sequence as Pelegant by using the same set of random seeds as Pelegant and generating beam sequentially with several iterations. For some types of beams, it is essentially impossible in any straightforward way for Pelegant and elegant to generate the same sequence, and we have to do the reconciliation by visualization.

Pelegant with parallel SDDS has been successfully applied to accelerator research and operations at the Advanced Photon Source (APS) at Argonne National Laboratory. To test the performance of the new version of Pelegant, we chose a simulation requiring a significant amount of I/O operations. The system being modeled is a very large energy recovery linac (ERL) upgrade of the APS [11]. It includes a two-pass 7-GeV linac, nine 48-m-long undulators, and numerous transport-line magnets. Beam is accelerated from 10 MeV to 7 GeV and then decelerated through the same linac. Modeling includes rf acceleration with exact time dependence, coherent synchrotron radiation, longitudinal space charge, wakefields, quantum excitation, and beam apertures.

There are nearly 20 watch points in the system, providing valuable information about the phase space at important locations. In addition, statistics are computed at the exit of Computer Codes (Design, Simulation, Field Calculation)

every beamline element. The input beam is read from an SDDS file. Because of these factors, this simulation would be inefficient using the previous version of Pelegant for a large number of simulation particles, due to both I/O and memory bottlenecks.

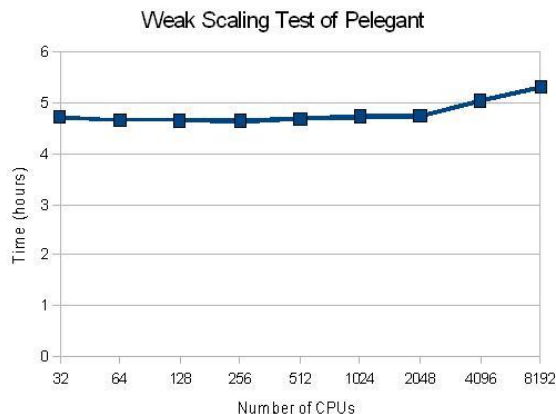


Figure 1: Weak scaling test of Pelegant on NERSC's Cray XT4 Franklin supercomputer.

We did a weak scaling test by increasing the number of particles in proportion to the number of CPU cores (i.e., keeping workload per CPU same) on the Cray XT4 Franklin supercomputer at National Energy Research Scientific Computing Center (NERSC). The performance test started from 448K particles on 32 cores and ended with 115 million particles on 8192 cores. This is significant because 115 million particles is approximately the actual number of electrons in the 19 pC beam.

From Figure 1, we can see that Pelegant achieved optimal performance when the number of cores is less than or equal to 2048. When the number of cores is above 4096, it takes a little bit more time than the test with fewer CPU cores due to the I/O scalability limit and communication overhead. A test with a quarter billion particles was also conducted on the Franklin supercomputer, but we ran out of allocated CPU hours. The simulation with a quarter billion particles is expected to be done in about 6.5 hours, which is still reasonably good from an efficiency point of view. Technically speaking, Pelegant should be able to simulate a billion particles efficiently. We can also increase the number of particles for each core (longer run time is expected) to run Pelegant more efficiently, as the memory requirement for the weak scaling test designed above is just under 1 percent of 2GB memory for each core on the Franklin supercomputer.

PARALLELIZATION OF FREQUENCY-MAP ANALYSIS

The frequency-map analysis command in elegant is very useful to quickly identify resonances in a circular accelerator. The task is to track particles with a grid of starting coordinates and determine frequencies of x and y mo-

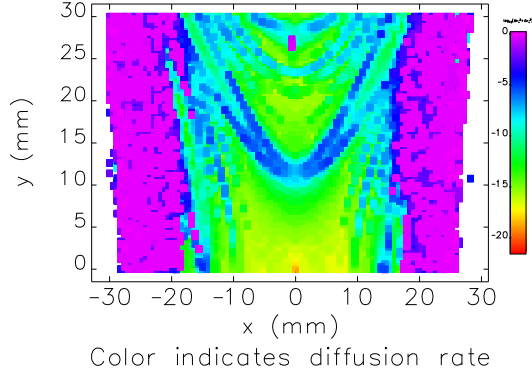


Figure 2: Result of parallel frequency map analysis with Pelegant.

tion. In the parallel version, the simulation is done by distributing the starting coordinates to different CPUs and tracking the particles simultaneously. We chose 2D domain decomposition to parallelize this operation. As there is no inter-process communication requirement for the tracking of each grid point, and the final frequency map results from all the processors are written with parallel I/O, a very good speedup of this type of simulation can be achieved. For example, we set the number of grid points in the x direction to 100 and the number in the y direction to 80. The simulation was done with 3 hours and 35 minutes with serial elegant, while it took 3.5 minutes on 100 CPUs on a cluster at APS. To use this operation efficiently, a user should choose the total grid points to be a multiple of the number of CPUs.

Figure 2 shows the frequency map for the APS Particle Accumulator Ring (PAR) lattice as computed with Pelegant. With this parallelized command, the user can choose very fine grids and get the frequency-map analysis result very quickly. This result took under 10 minutes on an 8-core desktop.

PARALLELIZATION OF MOMENTUM APERTURE SIMULATION

The position-dependent momentum aperture[14] is an important aspect for storage ring optimization, as it strongly affects the Touschek lifetime. The purpose of this simulation is to determine the range of momentum deviation for which the particle will survive as a function of starting longitudinal position. The end of each element will be used as the longitudinal position to scan, and hundreds of passes must be tracked for each starting longitudinal position. For a damping ring with a large number of elements, such as the International Linear Collider (ILC) damping ring, it would take several days or weeks to finish one simulation on a single processor. For this type simulation, we chose element decomposition for parallelization. The lattice is partitioned to several segments that are distributed to different CPUs. An equal number of elements are allocated

to each processor. The workload is largely independent of where the simulation starts, so the static load balancing will be good enough for efficiency considerations.

We did a test with 780 elements, which is a small portion of elements from the ILC damping ring, on a cluster at APS. The simulation was done in 55 hours and 50 minutes with the serial elegant. It took 7 hours and 13 minutes on 8 CPUs, and 17 minutes on 250 CPUs. This is another application where Pelegant allows finishing large simulations quickly enough to provide a useful design tool for day-to-day work.

DYNAMIC APERTURE OPTIMIZATION WITH PELEGANT

Dynamic aperture is another important aspect of storage ring optimization. We have successfully parallelized the line search mode to find the dynamic aperture. The line mode searches for the aperture boundary starting from the origin and moving outward. The exploration starts from $(0, 0)$ to $(x_{max} * \sin(\theta), y_{max} * \cos(\theta))$, where θ takes values from $-\pi/2$ to $\pi/2$. The area of the dynamic aperture is given in the output file with a parameter called “Area”[15]. The easiest way to parallelize this application is to distribute the workload to different processors with different θ s. In this case, the maximum number of CPUs would be limited to the number of lines to search. As the number of grid points in a line is usually greater than the number of lines to search, a better approach would be distributing different grid points of a line for each θ to different CPUs. First, the origin point is tracked by all the processors. If the particle in the origin survives from tracking, then the next n_c grid points close to the origin will be distributed to each CPU, where n_c is the total number of CPUs. This procedure will continue until a particle is lost for a grid point, or the boundary of the searching area is reached on a CPU. The last surviving grid point on each CPU will be returned to the master CPU. Finally, Pelegant finds the closest point to the origin from all the returned points, then searches the dynamic aperture boundary point for the next line.

The implementation mentioned above will not reach optimal performance as one processor has no knowledge of whether the other processors have reached the boundary of the aperture. But it still achieved very good efficiency when the number of grid points in each line is a multiple of the number of CPUs. Figure 3 shows an example the APS storage ring 400-turn dynamic apertures for 20 error ensembles using Pelegant. This took under 70 minutes on an 8-core desktop.

CONCLUSION

The capability of Pelegant running large-scale simulation has been significantly enhanced after successful integration with parallel SDDS, which eliminates the bottleneck caused by serial I/O. The program shows good

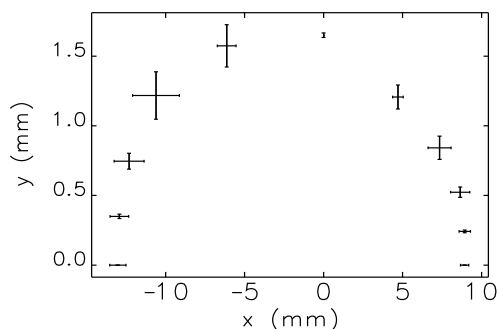


Figure 3: Dynamic aperture for 20 error ensembles for the APS storage ring.

scalability with a very large number of CPU cores on the Franklin supercomputer at NERSC. Simulation with hundreds of millions of particles can be done efficiently within reasonable time. Several important operations in *elegant*, such as frequency map analysis, dynamic aperture search and momentum aperture determination, can now be run on a multi-processor desktop or cluster with *Pelegant*.

ACKNOWLEDGMENT

These studies are partially supported by the SciDAC projects “Community Petascale Project for Accelerator Science and Simulation (ComPASS)” and “Frontiers in Accelerator Design: Advanced Modeling for Next-Generation BES Accelerator”. The ComPASS project is supported under the SciDAC program by the U.S. Department of Energy Office of High Energy Physics, Office of Nuclear Physics, Office of Basic Energy Sciences, and Office of Advanced Scientific Computing Research. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

- [1] M. Borland, “*elegant*: A Flexible SDDS-compliant Code for Accelerator Simulation,” APS LS-287, (2000).
- [2] M. Borland, “Modeling of the microbunching instability,” *Phys. Rev. ST Accel. Beams* 11, 030701 (2008).
- [3] H. Shang, Y. Wang, R. Soliday, M. Borland, “Parallel SDDS: A Scientific High-Performance I/O Interface,” these proceedings.
- [4] <http://www.mcs.anl.gov/research/projects/mpich2/>
- [5] <http://www.pvfs.org>
- [6] <http://www.ibm.com/systems/clusters/software/gpfs.html>
- [7] Y. Wang and M. Borland, “*Pelegant*: A Parallel Accelerator Simulation Code for Electron Generation and Tracking,” Proceedings of 12th Advanced Accelerator Concepts Workshop, (2006).
- [8] Y. Wang and M. Borland, “Implementation and Performance of Parallelized *Elegant*,” *Proc. PAC07*, (2007)
- [9] M. Borland and V. Sajaev, *Proc. PAC05*, 3886-3888 (2005).
- [10] R. Bartolini, *et al.*, *Proc. EPAC06*, 160-162.
- [11] M. Borland *et al.*, “Direct Methods of Optimization of Storage Ring Dynamic and Momentum Aperture,” *Proc. PAC09*, to be published.
- [12] M. Borland, *et al.*, “Possible Upgrade of the Advanced Photon Source with an Energy Recovery Linac,” *Proc. PAC09*, to be published.
- [13] M. Borland, “A Self-Describing File Protocol for Simulation Integration and Shared Postprocessors,” *Proc. PAC 95*, 2184-2186 (2006).
- [14] M. Belgroune *et al.*, “Refined Tracking Procedure for the SOLEIL Energy Acceptance Calculation,” *Proc PAC 2003*, 896-898 (2003).
- [15] http://www.aps.anl.gov/asd/oag/manuals/elegant_latest/elegant.html

BEAM FIELDS IN AN INTEGRATED CAVITY, COUPLER AND WINDOW CONFIGURATION *

Stephen Weathersby, Alexander Novokhatski
(SLAC National Accelerator Laboratory, Menlo Park, California)

Abstract

In a multi-bunch high current storage ring, beam generated fields couple strongly into the RF cavity coupler structure when beam arrival times are in resonance with cavity fields. In this study the integrated effect of beam fields over several thousand RF periods is simulated for the complete cavity, coupler, window and waveguide system of the PEP-II B-factory storage ring collider. We show that the beam generated fields at frequencies corresponding to several bunch spacings for this case gives rise to high field strength near the ceramic window which could limit the performance of future high current storage rings such as PEP-X or Super B-factories.

INTRODUCTION

The SLAC PEP-II asymmetric B-factory storage ring collider nominally collides 1700 bunches of 3.0 A of 3 GeV positrons on 2.0 A of 9 GeV electrons consisting of a low energy positron storage ring (LER) situated above a high energy electron storage ring (HER). The rings intersect at an interaction point (IP) within the BaBar detector sustaining a luminosity of $1.2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ at the $\Upsilon(4S)$ resonance.

Energy lost from synchrotron radiation and wake fields is replenished to the beam with high power RF supplied to cavities. Klystrons generate the 1 MW high power 476 MHz CW RF which is transported through WR2100 waveguides into the cavities through a 1.8 cm thick 24.8 cm diameter ceramic window. The window holds the ultra-high vacuum pressure required in the cavity from the near atmosphere pressure of the waveguide while transmitting 500 kW of RF power[1, 2]. The coupler geometry places the window at a half wavelength away from a detuned short position of the cavity field to minimize reflected power at the window position[1, 2]. This works well for reflected energy at harmonics of the main generator RF frequency. For high current storage ring B-factories and light sources, higher order modes (HOMs) excited by the beam constitute a significant portion of the cavity fields. The effect of such fields on the complete cavity/coupler/window/waveguide system is examined in this study. Fields produced by the beam in the cavity enter the waveguide through the cavity coupler and excite modes with fields near the window.

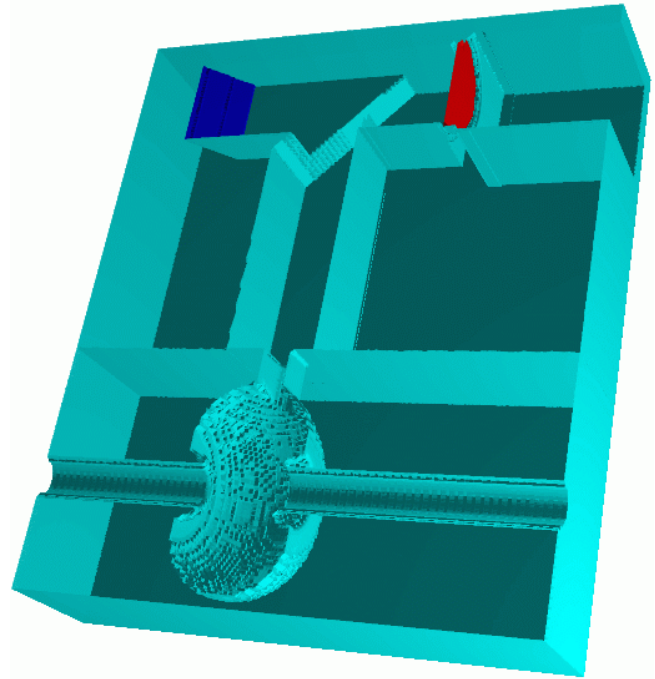


Figure 1: Model of PEP-II cavity/coupler/window/waveguide system.

RF SYSTEM MODEL

Figure 1 shows a cut plane through the full 3d model of the PEP-II cavity/coupler/window/waveguide RF system. A section of 1.5 meter long beam pipe is surrounded by a cavity which is connected through a small coupling iris into a rectangular volume which acts as a quarter-wave transformer. This volume intercepts a second rectangular volume, which functions as a filter, in which there exists a dielectric ceramic window (red) of relative permittivity $\epsilon_r=9$ and HOM absorbing ceramic tiles (blue) [6] with a relative permittivity of $\epsilon_r=30$ and conductivity $\sigma = 0.918 \text{ohm}^{-1}\text{m}^{-1}$. The model does not include the higher order mode dampers and detuning structures attached to the cavity. The WR2100 waveguide cutoff frequency for the TE₁₀ mode is 280 MHz. The beam pipe diameter is 9.5 cm with a cutoff frequency of 1.8 GHz.

RF SYSTEM EIGENMODES

The resonant modes in the cavity/coupler/window/waveguide RF system are identified in two ways. One is by performing an eigenmode deter-

* Work supported by Department of Energy Contract DE-AC02-76SF00515

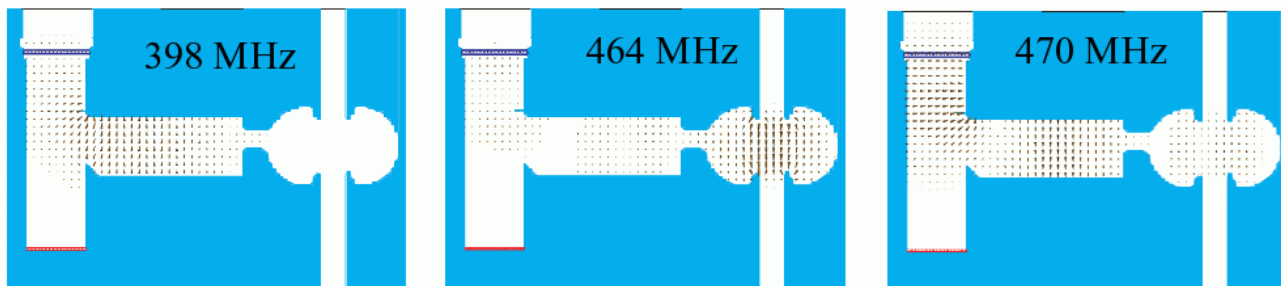


Figure 2: Eigenmode electric field distributions for the complete cavity-coupler-window-waveguide system from a MAFIA[4] calculation. Note the window is blue while the HOM absorbing tile is red in this figure.

mination with waveguide and beam ports shorted. The second is a scattering parameter simulation in which the waveguide port is excited with a matched broadband TE₁₀ waveguide mode to determine a reflected and transmitted power spectrum. The cavity resonant frequency found from these calculations is used to set the bunch spacing in resonance with the cavity for a time domain simulation.

Frequency domain solvers GdfidL[3] and MAFIA[4] both find the three eigenmodes shown in figure 2. The mode at 470 MHz exhibits substantial field near the window and in the cavity. The 464 MHz mode is the cavity accelerating mode.

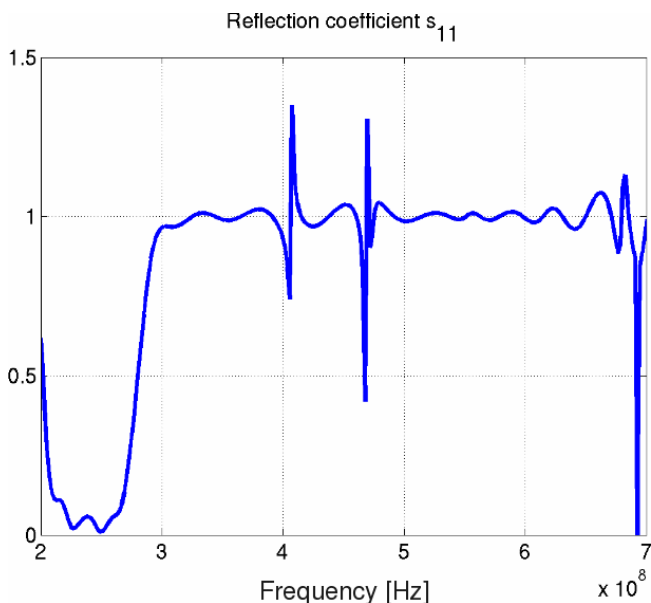


Figure 3: Reflection coefficient of the cavity/coupler/window/waveguide system at the waveguide port for a TE₁₀ waveguide mode showing resonances indicated near 400 MHz and 470 MHz.

Figures 3 and 4 show the reflection coefficient s_{11} and phase of the system from a Gaussian frequency domain WR2100 TE₁₀ waveguide excitation applied at the waveguide port as computed with GdfidL. This response is consistent with MAFIA results for the same geometry. The reflection coefficient indicates resonances near 400 and 470 High-Intensity Accelerators

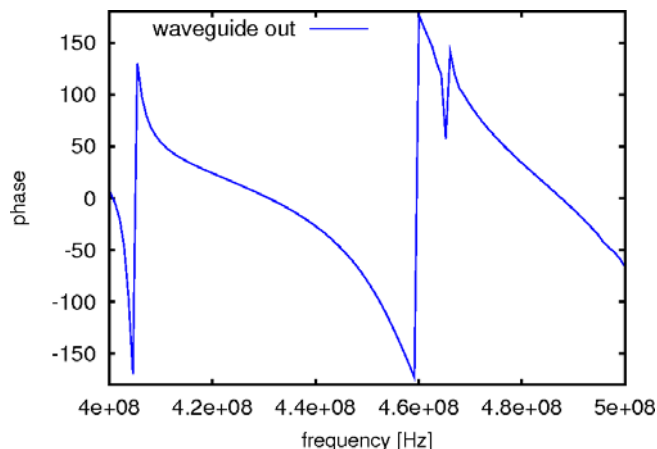


Figure 4: Reflection phase of the cavity/coupler/window/waveguide system at the waveguide port for a TE₁₀ waveguide mode showing resonances indicated near 400 MHz and 470 MHz.

MHz. They correspond to eigenmodes of the system found from GdfidL and MAFIA and indicate that these modes are not an artifact of boundary conditions imposed in the eigenmode determination.

TIME DOMAIN SIMULATIONS

To simulate nominal operating conditions, one thousand bunches are introduced into the beam pipe at a time interval resonant with the 464 MHz cavity resonance using the GdfidL time domain solver, with bunch spacing at every two 464 MHz RF buckets. Unlike normal operating conditions, no generator RF is applied at the waveguide port, which is simulated with perfectly matched layers (PML)[5] to absorb with negligible reflection any outgoing fields. The beam pipe boundaries are similarly treated with PMLs. The beam is the only source of fields in this simulation. Fields are solved for the entire cavity/coupler/window/waveguide system.

As shown in figure 5, the integrated electric field distribution after 1000 bunches shows a superposition of the main cavity mode and another mode with a similar field distribution to the 470 MHz mode with fields near the win-

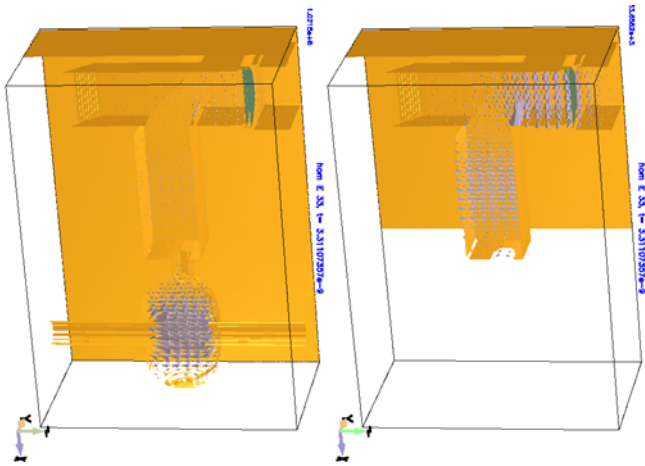


Figure 5: Electric fields after 1000 bunches determined from GdfidL. Left figure exhibits buildup of cavity fields in relation to coupler and waveguide. Right figure is same distribution auto-scaled to coupler and waveguide fields. The coupler waveguide field distribution has geometry similar to the 470 MHz mode of figure 2.

dow. Fields naturally build up in the cavity, but also in the coupler, waveguide and window areas. The waveguide and window fields exhibit the geometry of the 470 MHz mode of figure 2 which has substantial field near the window. Figure 6 plots the maximum electric field strength at

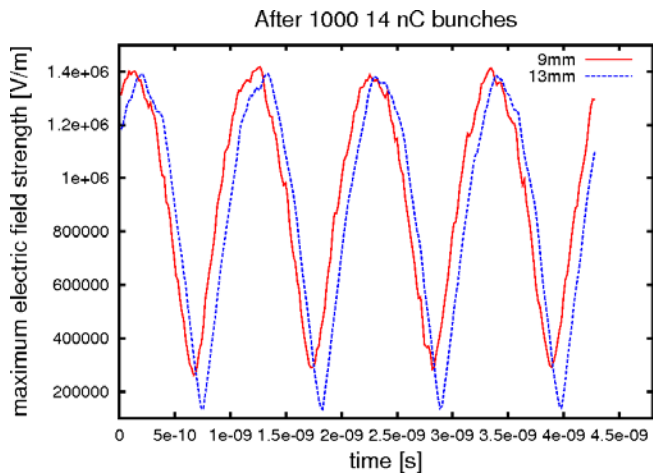


Figure 6: Maximum electric field as a function of time near the window after 1000 14 nC 9 mm and 13 mm bunches as calculated by GdfidL.

the location of the window as a function of time after passage of 1000 bunches of 14 nC bunch charge with bunch lengths 9 and 13 mm. The average field is higher for the shorter 9 mm bunch. A maximum electric field strength of 1.4 MV/m is generated by the beam near the window for both cases, which corresponds to nearly half of the copper breakdown voltage taken to be 30 kV/cm. The field distribution near the window is shown in figure 7 indicating large fields at both the center of the ceramic window and High-Intensity Accelerators

locations near the window/waveguide interface.

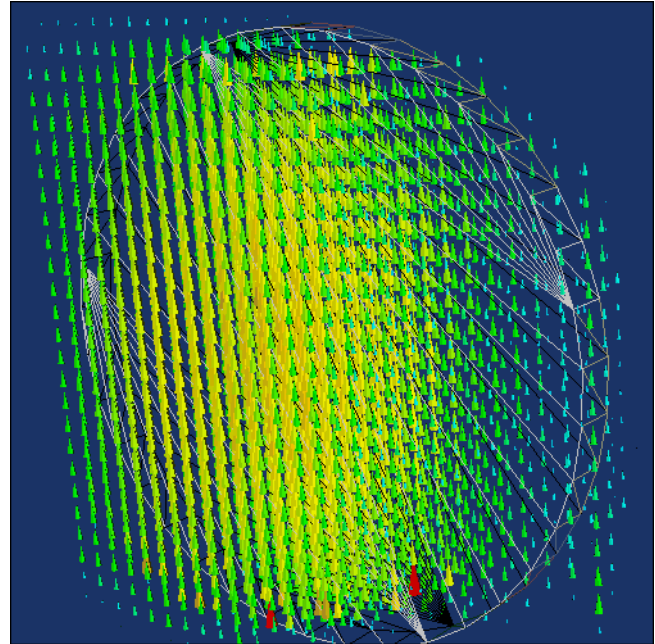


Figure 7: Snapshot of electric field distribution near the window after 1000 bunches looking from the waveguide vacuum side.

CONCLUSIONS

For high current storage ring B-factories and light sources, beam generated fields can compromise the performance of an RF system. Characterizing this effect requires modeling a complete cavity/coupler/window/waveguide RF system. In addition to the accelerating 464 MHz cavity mode, eigenmode and scattering parameter analysis performed on a model of the PEP-II cavity/coupler/window/waveguide RF system has found a 470 MHz mode with significant electric fields near the window. This mode is excited by the beam and produces a maximum electric field strength of 1.4 MV/m near the window after 1000 14 nC bunches at a two RF bucket spacing of the cavity resonance frequency. The average maximum electric field strength near the window increases with a shorter bunch.

REFERENCES

- [1] High power RF window and coupler development for the PEP-II B factory. M. Neubauer, K. Fant, J. Hodgson, J. Judkins, H. Schwarz, (SLAC), R.A. Rimmer, (LBL, Berkeley). SLAC-PUB-6897, SLAC-PUB-95-6897, LBL-37250, May 1995. 3pp.
- [2] An Asymmetric B factory based on PEP: Conceptual design report. SLAC-0372, SLAC-R-372, Feb 1991. 496pp.
- [3] <http://www.gdfidl.de>
- [4] The MAFIA collaboration "User Guide", CST GmbH, Darmstadt, Germany.

- [5] J. Berenger (1994). "A perfectly matched layer for the absorption of electromagnetic waves". *Journal of Computational Physics* 114: 185–200.
- [6] Coupling Network Simulation for the PEP-II RF Cavity. C.-K. Ng, K. Ko, N. Kroll, and R. Rimmer, SLAC-PUB-6565, LBL-35808, June 1994.

BPM BREAKDOWN POTENTIAL IN THE PEP-II B-FACTORY STORAGE RING COLLIDER

Stephen Weathersby, Alexander Novokhatski
(SLAC National Accelerator Laboratory, Menlo Park, California)

Abstract

High current B-Factory BPM designs incorporate a button type electrode which introduces a small gap between the button and the beam chamber. For achievable currents and bunch lengths, simulations indicate that electric potentials can be induced in this gap which are comparable to the breakdown voltage. This study characterizes beam induced voltages in the existing PEP-II storage ring collider BPM as a function of bunch length and beam current.

INTRODUCTION

The SLAC PEP-II asymmetric B-factory storage ring collider nominally collides 1700 bunches of 3.0 A of 3.1 GeV positrons on 2.0 A of 8.0-10.1 GeV electrons. It consists of a low energy positron storage ring (LER) situated above a high energy electron storage ring (HER). The rings intersect at an interaction point (IP) within the BaBar detector sustaining a luminosity of $1.2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ at the $\Upsilon(4S)$ resonance. To monitor the beam position, hundreds of beam position monitors (BPMs) line the beam vacuum chamber. Each BPM consists of a round button electrode 15 mm in diameter which is mechanically press fitted to the 50 Ohm feed-through connector as shown in figure 1.

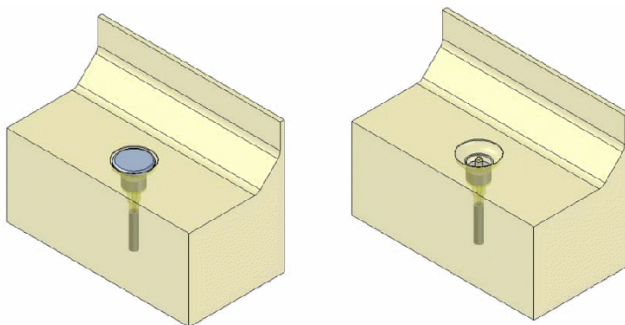


Figure 1: One quarter of the BPM geometry with and without a button. Chamber length is 9 cm. Button diameter is 15 mm.

While running at shortened bunch length (9 mm) some of the upper button electrodes heated up enough to fall off their mounts. The upper electrode fell onto the lower electrode as shown in Fig. 2 which not only shorted the underlying electrode but also became a large obstacle for the beam fields, increasing the current through the lower elec-

trode. This then melted the feed-through (Fig.3) causing a vacuum breach.

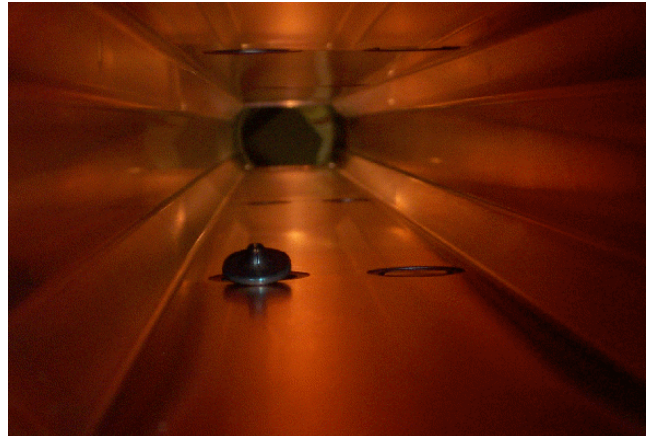


Figure 2: A button of an upper BPM fell off onto a lower button.



Figure 3: Melted feed-through of a lower button and the fallen upper button.

The origin of the heating is the wake field generated by an intense short bunch passing by the vacuum chamber discontinuity due to a BPM button. The effect of beam fields on a PEP-II BPM are examined[1]. Scattering parameter analysis reveals resonant behavior near the frequency of 7 GHz. Time domain simulations show that maximum electric fields in the BPM are located at the upbeam and downbeam extremes of the BPM button corresponding to an excited dipole resonant mode in the BPM environment. PEP-II has had a history of arcing and vacuum bursts caused by small geometric gaps in RF seals[3, 2]. It is natural to suspect that such small gap structures in the BPM design may cause the same problems. At a resonant condition when a

* Work supported by Department of Energy Contract DE-AC02-76SF00515

BPM mode is a harmonic of a bunch spacing frequency the fields of many bunches will be summed. Shorter bunches and higher currents will also raise the maximum electric field at the BPM, which can approach the breakdown voltage taken to be roughly 30 kV/cm in copper.

SCATTERING PARAMETER STUDIES

At LER currents of 2.4 Amperes with a bunch length of 0.8 cm, BPM buttons became hot enough to fall off their mounts near the IP region due to higher order mode (HOM) heating. In order to continue uninterrupted running, investigations were undertaken into the impact of missing buttons on machine performance. As shown in figure 1, the missing button leaves behind a stem onto which the button was mounted and the cavity which housed it. In order to quantify impedance presented to the beam by such a structure, a scattering parameter analysis is performed. Using electromagnetic solver MAFIA[4], a simulated wave of known power in the form of a lowest propagating vacuum chamber mode with a Gaussian frequency distribution is introduced into the geometries of figure 1. The vacuum chamber boundaries are matched to allow transmission of this mode in a smooth vacuum chamber without reflection. This mode with its radial electric field has similar physical characteristics to beam fields. Results indicate an impedance at 7 GHz for the intact button case as shown in figure 4.

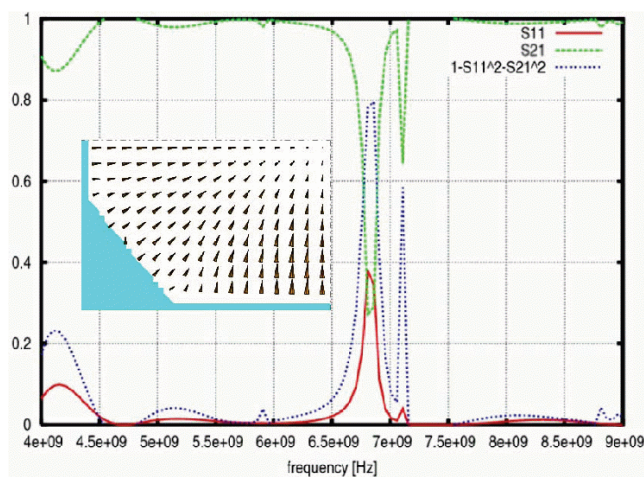


Figure 4: Fractional mode power transmitted s_{21} , reflected s_{11} and absorbed $1 - s_{11}^2 - s_{21}^2$ for the IP BPM geometry of figure 1 with button. Inset shows mode electric fields.

The missing button scenario presents essentially no impedance in this frequency range, and machine running in this configuration presented no heating issues. The loss factor as computed with GdfidL[5] is lower with the missing button as shown in figure 5. With the missing button the BPM was still functional despite an attenuated signal.

Lepton Accelerators and Colliders

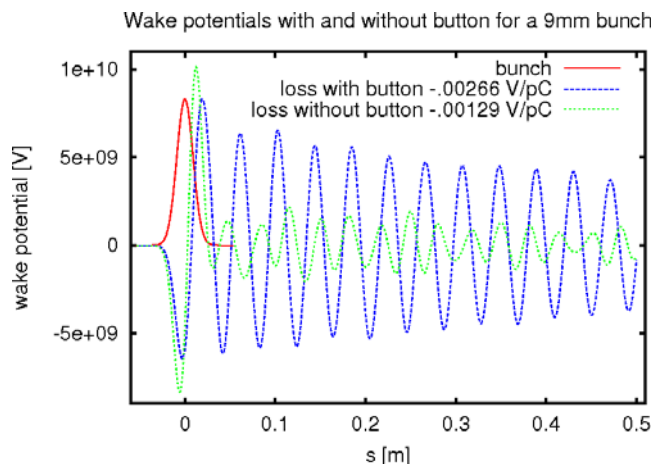


Figure 5: Wake potentials and loss factors for BPM with and without button for a 1 C 9mm bunch. Loss factor is lower without the button.

SINGLE BUNCH SIMULATIONS

A single 1 C 6 mm long Gaussian bunch propagating on axis through the 0.09 m long vacuum chamber of figure 1 with the intact button excites fields in the vicinity of the BPM. A simulation performed with the GdfidL[5] time domain solver yields the electric field strength shown in figure 6 at a particular time after the passage of the bunch (0.4 ns). The maximum fields occur in the gap between the button and the vacuum chamber along the direction of the beam. For this case an electric field strength of 4.59×10^{13} V/m is generated. Normalized to a typical 14 nC PEP-II bunch charge this yields a maximum 6 kV/cm. As a function of time the maximum electric field peaks shortly after the bunch passes as shown in figure 7, reaching 9×10^{13} V/m for a 1 C charge. This equates to 12.6 kV/cm for a 6 mm 14 nC PEP-II bunch.

The location of the maximum field changes with time. Figure 8 shows the position coordinates of the maximum field near the BPM as a function of time for a 6 mm Gaussian 1 C bunch. The largest coordinate variation is in z which is along the beam direction. The origin $z=0$ corresponds to the BPM z position. The z variation reflects the movement of charge back and forth between the front and rear of the BPM with a frequency of 7 GHz. Smaller variations are present for the x and y coordinates. The maximum field location correlates with the smallest part of the gap between the button and the vacuum chamber. Figure 9 is a cross section view of the geometry which shows the smallest gap is interior to the BPM cavity.

MULTI-BUNCH SIMULATIONS

To properly simulate B-factory operating conditions multiple bunches are required to traverse the vacuum chamber at the nominal bunch spacing of 4.2 ns, corresponding to two 476 MHz RF buckets. Initially a simulated Gaussian line charge of 1 C traverses the BPM vacuum chamber in

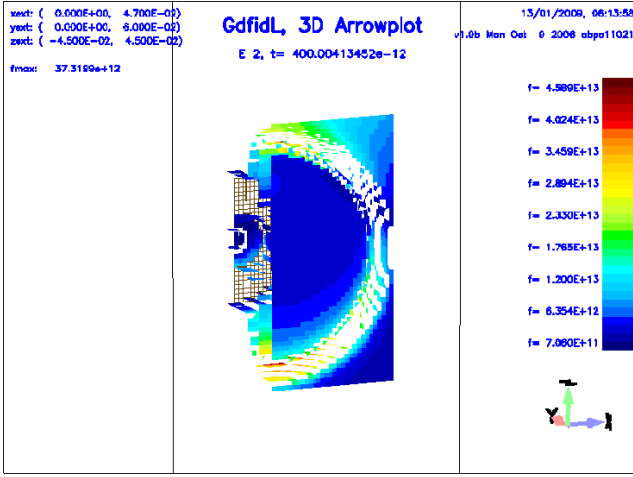


Figure 6: Electric field strength at the button as a function of time just after the arrival of one 9 mm 1 C bunch as computed with Gdfidl. Maximum field strength is 4.59×10^{13} V/m near the gap between the button and vacuum chamber interior to the button housing. Beam direction is along the z -axis.

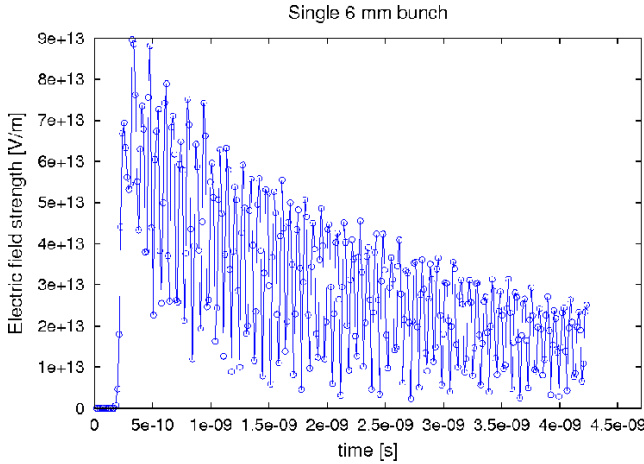


Figure 7: Maximum electric field strength at the BPM as a function of time for one 6 mm Gaussian 1 C bunch. The 14 GHz oscillation correlates with a 7 GHz half cycle of electric field maximum at two ends of the button. At $t = 0$ the head of the bunch arrives at the upbeam end of the BPM chamber.

a time domain simulation. The resulting electric and magnetic fields are computed and saved at a time when a subsequent bunch is about to enter the computational domain. This saved field distribution forms the initial conditions for the next bunch simulation.

Figure 10 shows the maximum field strength at two particular times as a function of the number of 9 mm bunches. After four bunches have past, the field strengths appear to approach an equilibrium.

The maximum field at the BPM depends on the bunch frequency and bunch length. The PEP-II RF frequency is Lepton Accelerators and Colliders

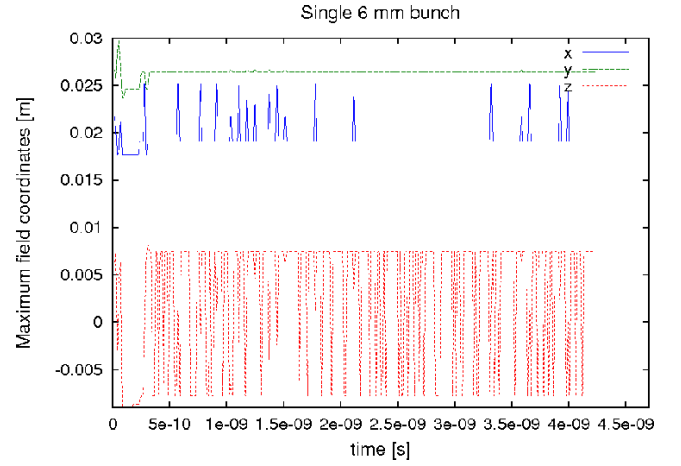


Figure 8: Coordinates of the maximum electric field strength in the vicinity of the BPM. $z = 0$ is the position of the BPM along the z -axis beam direction. The maximum field position oscillates between the upbeam and downbeam parts of the button.

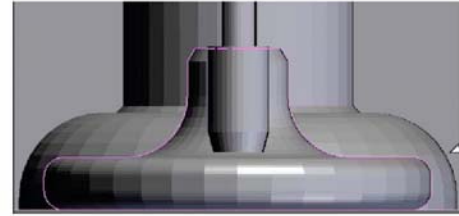


Figure 9: BPM cross sectional view. Maximum electric field strength at the BPM correlates with minimum gap spacing (~ 1 mm) between BPM button and vacuum chamber.

nominally 476 MHz and the nominal bunch spacing was 2 RF buckets or 4.2 ns. Changes in the RF frequency can move the bunch spacing closer to a BPM resonance. Figure 11 shows the maximum BPM electric field strength vs time for various RF frequencies and bunch lengths. The average field generated by four 9 mm bunches at various RF frequencies near the nominal 476 MHz is shown in figure 12.

A peak field of 1.2×10^{14} V/m for the shortest 4 mm 1C bunch at 1904 MHz translates to about 16.8 kV/cm at the nominal 14 nC bunch. Shorter bunches produce higher maximum electric fields as shown in figure 13 for the case of fixed 1904 MHz RF frequency.

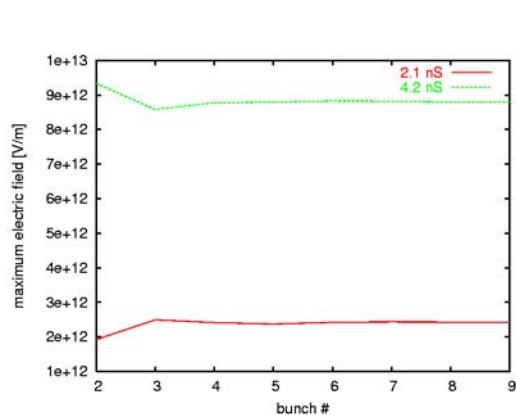


Figure 10: Maximum electric field at the button as a function of bunch number at two particular times for 1 C, 9 mm long Gaussian bunches. Solid trace is 2.1 ns, dashed trace is 4.2 ns. Fields have equilibrated after about 4 bunches.

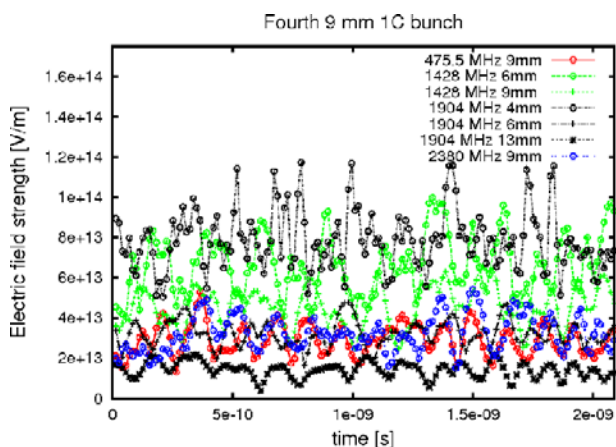


Figure 11: Maximum electric field at the button as a function of time during the passing of the fourth consecutive 1 C, at various RF frequencies and bunch lengths. Peak fields of 1.2×10^{14} V/m are achieved at 1904 MHz for the shortest bunch length of 4 mm.

CONCLUSION

Fields from short bunches can approach the breakdown voltage for the PEP-II B-factory button style BPM at a certain bunch spacing and bunch current. The position of the maximum electric field is internal to the BPM at the smallest gap along the beam direction and oscillates at 7 GHz between the upbeam and downbeam surfaces. The nominal PEP-II 14 nC, 9 mm bunch length at 4.2 ns bunch spacing will generate up to 6 kV/cm after about 4 bunches which is a factor of 5 below the 30 kV/cm copper breakdown voltage.

REFERENCES

- [1] Numerical simulation of the PEP-II beam position monitor. N. Kurita, D. Martin, C.K. Ng, S. Smith, (SLAC), T.

Lepton Accelerators and Colliders

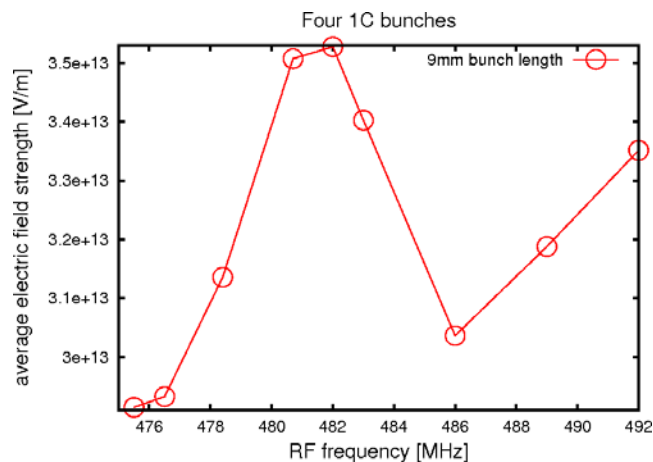


Figure 12: Average maximum electric field at the button as a function of frequency during the passing of the fourth consecutive 1 C 9mm bunch.

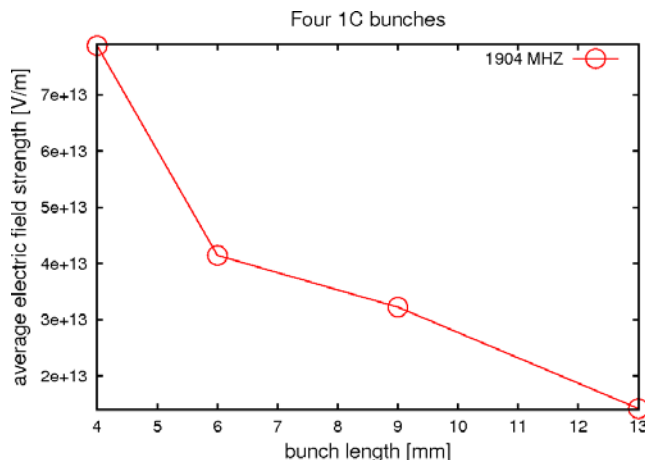


Figure 13: Average maximum electric field at the button as a function of bunch length during the passing of the fourth consecutive 1 C bunch for 1904 MHz RF frequency.

Weiland, (Darmstadt, Tech. Hochsch.) . SLAC-PUB-7006, SLAC-PUB-95-7006, Sep 1995. 10pp. Presented at International Workshop on Collective Effects and Impedance for B Factories, Tsukuba, Japan, 12-17 Jun 1995.

- [2] A Novokhatski, J. Seeman, M. Sullivan (SLAC). "A Model of an Electrical Discharge in the Flange Contacts With Omega Seals at High Currents in PEP-II". In the Proceedings of 11th European Particle Accelerator Conference (EPAC 08), Magazzini del Cotone, Genoa, Italy, 23-27 Jun 2008, pp TUPP054.
- [3] A. Novokhatski (SLAC). "HOM effects in vacuum system with short bunches". In the Proceedings of Particle Accelerator Conference (PAC 05), Knoxville, Tennessee, 16-20 May 2005, pp 289.
- [4] The MAFIA collaboration "User Guide", CST GmbH, Darmstadt, Germany.
- [5] <http://www.gdfidl.de>

RECYCLER LATTICE FOR PROJECT X AT FERMILAB *

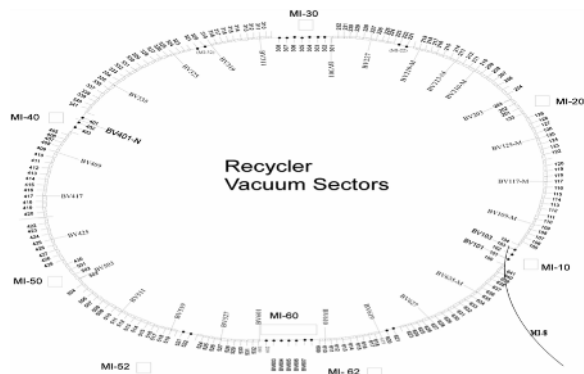
Meiqin Xiao, David E. Johnson, FNAL, Batavia, IL 60510, USA

Abstract

Project X is an intense proton source that provides beam for various physics programs. The source consists of an 8 GeV H^- superconducting linac that injects into the Fermilab Recycler where H^- are converted to protons. Protons are provided to the Main Injector and accelerated to desired energy (in the range 60 - 120 GeV) or extracted from the Recycler for the 8 GeV program. A long drift space is needed to accommodate the injection chicane with stripping foils. The Recycler is a fixed 8 GeV kinetic energy storage ring using permanent gradient magnets. A phase trombone straight section is used to control the tunes. In this paper, the existing FODO lattice in the RR10 straight section being converted into doublet will be described. Due to this change, the phase trombone straight section has to be modified to bring the tunes to the nominal working point. A toy lattice of recycler ring is designed to simulate the end-shim effects of each permanent gradient magnet to add the flexibility to handle the tune shift to the lattice during the operation of $1.6E14$ with KV distribution of the proton beam to give ~ 0.05 of space charge tune shift. The comparison or the combinations of the two modification ways for the Recycler ring lattice will be presented also in this paper.

INTRODUCTION

Project X [1] is an intense proton source that provides beam for various physics programs. The source consists of an 8 GeV H^- superconducting linac that injects into the Recycler where H^- are converted to protons. Protons are provided to the Main Injector and accelerated to desired energy or extracted from the Recycler for the 8 GeV program. The Recycler ring (shown in Fig. 1) is a fixed 8 GeV kinetic energy storage ring using permanent gradient magnets. RR10 is the straight section for placing the injection system and RR60 is a phase trombone straight section used to control the tunes.



RR10 STRAIGHT SECTION OF THE RECYCLER LATTICE FOR PROJECT X

The inject system for convert H^- to proton in Recycler is a multi turn stripping system which will be placed in RR10 straight section. Fig. 4 illustrates the injection Chicane with stripping foils

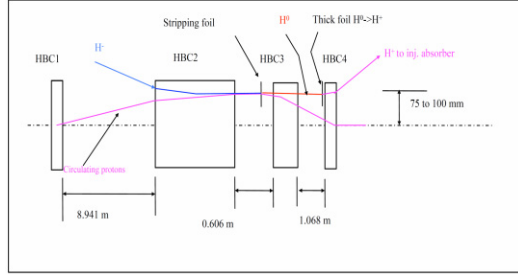


Fig.5: Injection system in RR10 straight section

To accommodate this, a 21.5 m long drift space is designed by converting the existing FODO lattice in RR10 straight section into a doublet, shown in Fig. 6.1 Meanwhile, constrain the $\beta_x, \beta_y < 55$ m in RR30, the Recycler lattice become as shown in Fig.6.2. The tunes now become (25.445, 24.134).

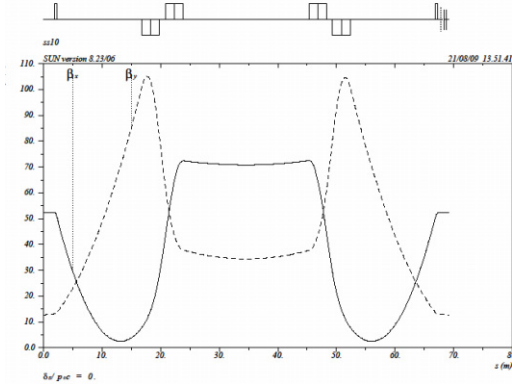


Fig. 6.1: Doublet in RR10 section

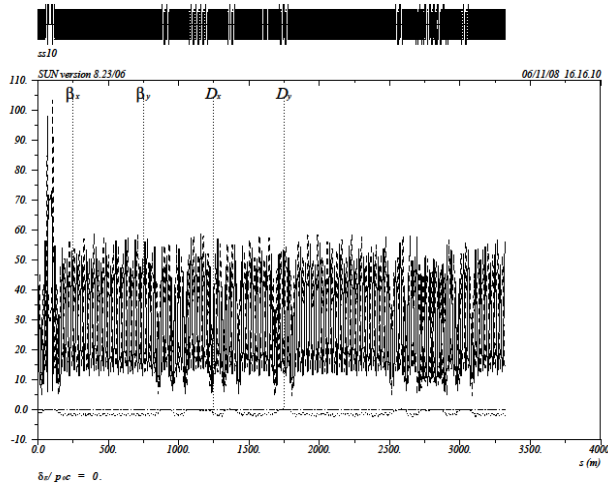


Fig. 6.2: Recycler lattice with a doublet in RR10 and RR30 constrained the $\beta_x, \beta_y < 55$ m

Accelerators and Storage Rings, Other

SOLUTIONS TO THE NOMINAL TUNES

RR 60 straight section is a phase trombone for Recycler tune control. It contains 32 permanent quads in 4 D-D FODO cells, and 36 trim quads in 9 families which are installed in the drift space of FODO cells, the basic settings of their currents are 0. Fig. 7 illustrates the two types of the quads.

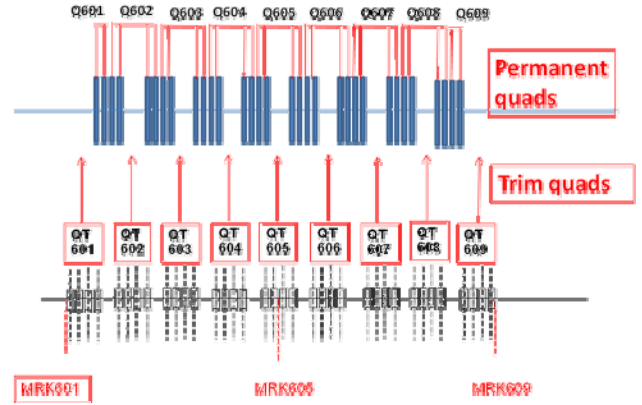


Fig. 7: Permanent quads and trim quads in RR60 phase trombone section

The First Solution

We kept the rest of the Recycler ring unchanged, but only varied the strengths of permanent quads in RR60 phase trombone straight section with the constrains:

- Keep $\alpha_{x,y}, \beta_{x,y}$ at two ends of the phase trombone RR60 unchanged
- $\beta_x, \beta_y < 55$ m
- Phase advance set to compensate the tune difference from the nominal tunes

We obtained (from MAD fitting):

$$\begin{cases} Q_x = 25.329 \\ Q_y = 24.350 \end{cases}$$

when

$$\begin{cases} k_1 Q_{60F} = 0.05404106 \\ k_1 Q_{60D} = -0.06359425 \end{cases}$$

Then we used trim quads to re-adjust base tunes. We obtained the tunes $Q_x=25.425, Q_y=24.415$ when

$$\begin{cases} QT601_I = -0.6149 \\ QT602_I = 0.7229 \\ QT603_I = -1.0923 \\ QT604_I = 2.9553 \\ QT605_I = -1.8319 \\ QT606_I = 2.8896 \\ QT607_I = -1.2121 \\ QT608_I = 0.7283 \\ QT609_I = -0.4635 \end{cases}$$

by a Mathematica program[3], which will be used in an application program for on-line tuning.

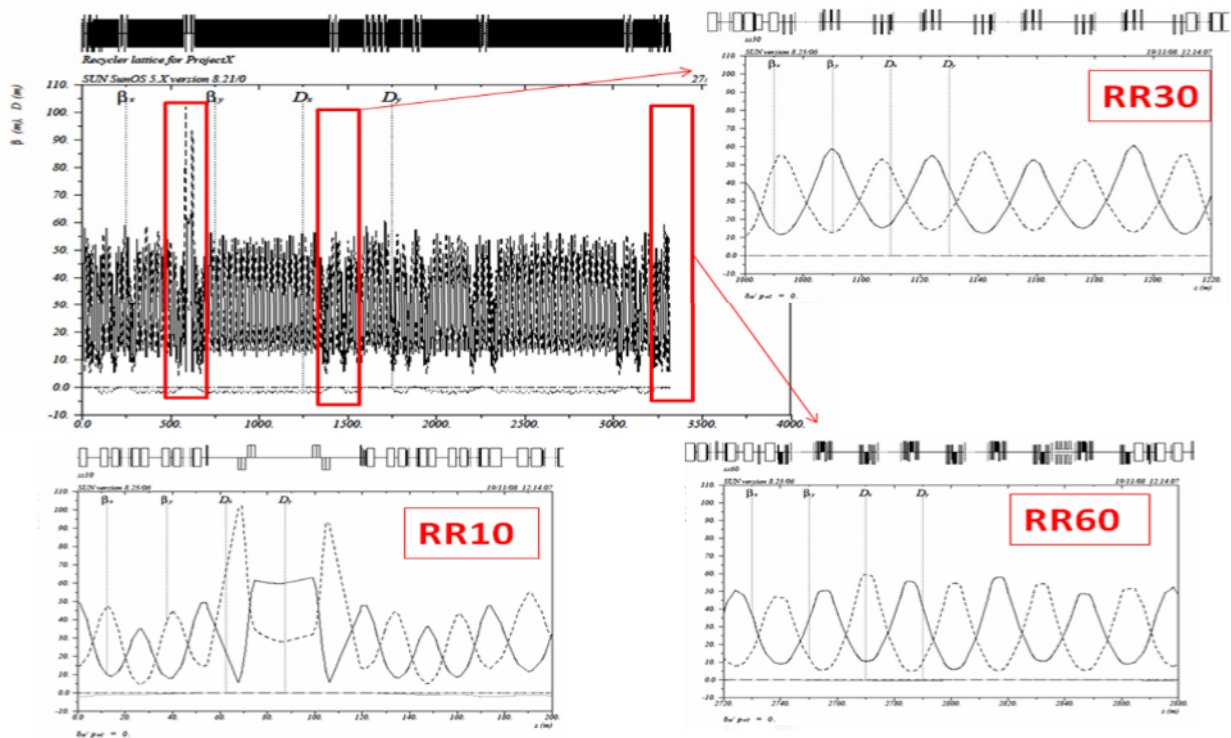


Fig. 8: Recycler lattice for Project X. Only the trombone section was modified.

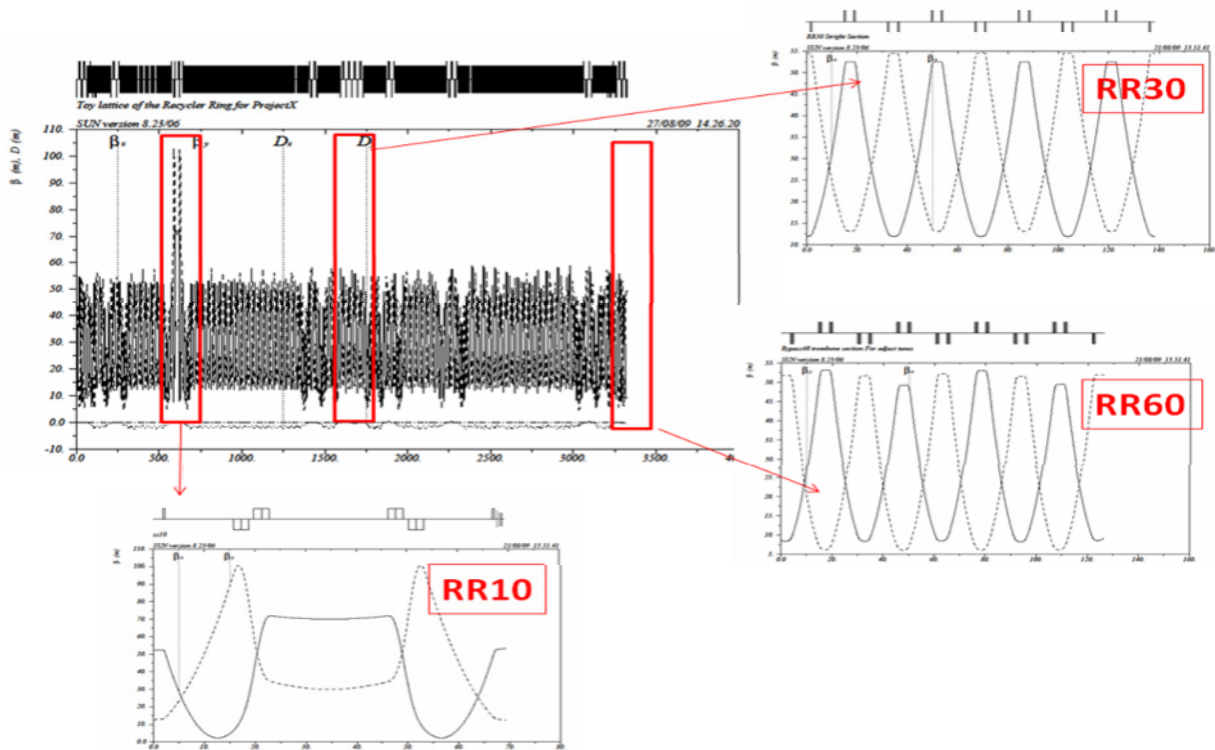


Fig. 9: Toy Recycler lattice. End-shim of each permanent gradient magnet in arc cell was tweaked so that the arc cell phase advance was slightly changed.

Notice that the currents of QT604 and QT606 are around 3 Amps. Actually the trim quads current could be changed up to 5 Amps. But the larger the currents, the larger the non-linear contribution to the operation program. With this solution, the recycler lattice for Project X shown in Fig. 8.

The Second Solution

To add the flexibility to handle the space charge tune shift (~ 0.05) due to the beam of 1.6×10^{14} with KV distribution, a toy lattice of the Recycler ring was built to simulate the end-shim effects of each permanent gradient magnet. It was found that the amount of the end-shim field tweaked is limited in order to eliminate the large beta-wave in the lattice. We then varied the permanent quads in RR60 again to obtain the nominal tunes (25.425, 24.415) with zero the trim quads current.

In the toy lattice:

- The phase advances of the arc cell were slightly changed from $\mu_x = 85.39^\circ$, $\mu_y = 79.22^\circ$ to $\mu_x = 84.45^\circ$, $\mu_y = 79.95^\circ$
- The phase advances of the normal straight section cell keep the same $\mu_x = 86.51^\circ$, $\mu_y = 80.34^\circ$
- The phase advances of the dispersion cell were changed slightly to $\mu_x = 87.85^\circ$, $\mu_y = 89.83^\circ$ to fit from arc cell to straight section cell.
- The phase advances of the RR60 trombone section straight section now are : $\mu_x = 97.53^\circ$, $\mu_y = 115.29^\circ$

This means:

- The amount of the end-shim tweaked is about 1% of the body quads field in horizontal and 0.5% in vertical planes.

- The amount of the strength of the permanent quads in RR60 is increased about 4% in horizontal and 13% in vertical planes.
- The currents of the trim quads in RR60 are 0 .

The toy lattice of the Recycler ring for Project X is shown in Fig. 9.

Apparently the second solution is more flexible to handle the space charge tune shift, but it needs a lot of work to adjust end-shim contributions. Technically it is doable.

CONCLUSION

Several considerations have been taken in modifying Recycler lattice for Project X. The Recycler lattice is manageable for Project X with some uncertainties on a. Beta functions on injection point (stripping foils) b. Nominal tunes considering the tune shift due to space charge effect. Toy lattice will be used do beam dynamics studies with space charge effects for looking for best nominal tunes.

REFERENCES

- [1] Project X ICD-1 . <http://projectx.fnal.gov/index.html>
- [2] Meiqin Xiao, "RR-30 straight section/Extraction Line Lattice Design". <http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=3099>
- [3] Meiqin Xiao, "Measurement and corrections of the Recycler Lattice at Fermilab". Proceeding of HB2008, August 25-29, 2008, Nashville

ARRAY BASED TRUNCATED POWER SERIES PACKAGE*

Lingyun Yang[†], NSLS-II, BNL, Upton, NY 11973, USA

Abstract

Truncated Power Series Algebra (TPSA) or Differential Algebra (DA) package has been a fundamental component for many accelerator physics simulation code, including FPP/PTC, MAD-X, BMAD and Tracy. We have developed a new algorithm to extend the ability of TPAS to handle problems with both large number of variables and high order. This package is implemented in C++ language with operator overloading, and has been integrated into PTC and MAD-X.

INTRODUCTION

Truncated Power Series Algebra (TPSA) or Differential Algebra (DA) package is a tool for Taylor series manipulation [1, 2]. It follows certain mathematical rules to do arithmetics among the Taylor series and can be used for map generating, sensitivity study, automatic differentiation and many other applications. Since it was implemented by Dr. Berz, it has been a fundamental component for many accelerator physics simulation codes, including FPP/PTC [4], MAD-X [3], BMAD and Tracy.

The package in Ref. [1] has two limitations:

- Limited index space at high v , e.g. 39 variables can only handle up to first order¹.
- Without operator overloading, Applying DA to an existing code will need many careful translation, and introduce temporary variable. (think about translate $x = (a + b) * c/d$).

We have developed a new TPSA package in C++ recently, with a new algorithm which can expand the ability to both high number of variables and high orders. It is also more user friendly and the speed are better than the FORTRAN 77 implementation.

ALGORITHM

The Taylor series of an given function $f(x_1, \dots, x_d)$ is given as:

$$f(x_1, \dots, x_d) = \sum_{n_1=0}^{\infty} \dots \sum_{n_d=0}^{\infty} \frac{(x_1 - a_1)^{n_1} \dots (x_d - a_d)^{n_d}}{n_1! \dots n_d!} \left(\frac{\partial^{n_1 + \dots + n_d} f}{\partial x_1^{n_1} \dots \partial x_d^{n_d}} \right) \quad (1)$$

Given the analytic form of $f(x_1, \dots, x_d)$, we can calculate the coefficient of Taylor series expansion, f_{x_1} , f_{x_2} , $f_{x_1 x_2}$, up to any order precisely. This can be done with finite difference method, but when going to high order, the truncation error and linear approximation may be a main obstacle. TPSA package uses a set of rule, which is complete for a field of real number and basic arithmetics. The rules can be found in Ref. [2], and we summerize here:

- Addition/Substraction. It is done between only corresponding terms (coefficients with same order or pattern of differentiation).
- Multiplication with constant. It scales all the coefficients.
- Multiplication with another series. The order of differentiation are added up for each variable. The new coefficient are added up to the new location with new pattern of order.
- Reciprocal. The coefficient is gained from the pattern of Taylor expansion of $1/x$.
- Basic functions. Such as $\sin(x)$, $\cos(x)$, \sqrt{x} are just deduced from their coefficients of Taylor series expansion together with the multiplication of two series.

As we can see that +/- are trivial, and every other calculation is depending on multiplication of two TPAS series. While for multiplication the key part is how to arrange the storage of each coefficients and locate them quickly.

IMPLEMENTATION

A TPSA vector with v independent variables, each up to d degrees, will have $C(v + d, d)$ or $C(v + d, v)$ elements, where $C(n + v, v)$ are binomial coefficients defined as

$$C(v + d, v) \equiv \binom{v + d}{v} = \binom{v + d}{d} = \frac{(v + d)!}{v!d!}$$

From this we can see that, for the length of a TPSA vector, the order and number of variables are symmetric. A code can do high order with low number of variables should be

* Work supported by the Director, Office of Science, U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

[†] lyang@bnl.gov

¹We can use more index function again, 3 instead of 2, to relax the limit from $(n + 1)^{v/2}$ to $(n + 1)^{v/3}$, but the performance will be lower.

Table 1: (v, d) Limited by Certain Number of Monomials N , e.g. $1k, 1M, 1G, 1T$.

v	d				
	$< 2^{10}$ (k)	$< 2^{20}$ (M)	$< 2^{30}$ (G)	$< 2^{40}$ (T)	$< 2^{50}$ (P)
4	10	68	398	$\approx 2^{40/4} * 4!$	$\approx 2^{50/4} * 4!$
6	6	26	92	300	962
8	4	16	46	115	282
10	4	12	30	67	139
12	3	10	23	46	88
14	3	9	19	36	64
16	3	8	16	30	51
18	2	7	15	26	42
20	2	7	13	23	36
30	2	5	10	16	23
40	2	4	8	13	18
50	1	4	7	11	16
60	1	4	7	10	14
70	1	3	6	9	13
80	1	3	6	9	12
90	1	3	6	8	11
100	1	3	5	8	11

able to do low order with high number of variables. Some examples are shown in Table 1. Where given the maximal number of terms in Taylor series, the table tells how many variables and upto which order the code can store.

The Taylor series expansion of a multi-variable function up to certain order has a *polynomial* structure, we write as (v, d) , where v is the number of variables, and d is the highest order considered. It is stored as a vector, where each element represents one term in Taylor Series Expansion.

In order to speed up the multiplication, we build up a look up table P . The element $P[i, j]$ is the location to save $v[i] \times v[j]$. A look up table for the above multiplication, 2 variables, order up to 3rd order looks like Table. 2

Table 2: Hash table for TPSA multiplication (2, 3)

		0	1	2	3	4	5	6	7	8	9
1	0	0	1	2	3	4	5	6	7	8	9
$\frac{\partial}{\partial x}$	1	1	3	4	6	7	8				
$\frac{\partial}{\partial y}$	2	2	4	5	7	8	9				
$\frac{\partial^2}{\partial x \partial x}$	3	3	6	7							
$\frac{\partial^2}{\partial x \partial y}$	4	4	7	8							
$\frac{\partial^2}{\partial y \partial y}$	5	5	8	9							
$\frac{\partial^3}{\partial x \partial x \partial x}$	6	6									
$\frac{\partial^3}{\partial x \partial x \partial y}$	7	7									
$\frac{\partial^3}{\partial x \partial y \partial y}$	8	8									
$\frac{\partial^3}{\partial y \partial y \partial y}$	9	9									

We should first notice that Table. 2 is symmetric, i.e. $P[i, j] = P[j, i]$, therefore only half of the storage is needed. Secondly, since we are truncating at certain order, the table is not squared but a step-like. The real size of

Computer Codes (Design, Simulation, Field Calculation)

this matrix $T(v, d)$ is

$$\begin{aligned}
 T(v, d) &= \sum_{k=0}^d \binom{v+k-1}{k} N(v, d-k) \\
 &\equiv \sum_{k=0}^d \binom{v+k-1}{k} \binom{v+d-k}{d-k} \\
 &= \sum_{k=0}^d \binom{v-1+k}{v-1} \binom{v+d-k}{v} \\
 &= \binom{2v+d}{d}
 \end{aligned} \tag{2}$$

where $N(v, d)$ is the length of TPSA vector (v, d) . From Table. 1 we can easily find the limit on (v, d) when using this hash table.

The above hash table can solve problems both low order but high number of variables and high order but small number of variables, but for both high order and large number of variables, the hash table becomes too large to store.

We can arrange the Taylor series in a way such that, the new location of multiplication of two TPSA coefficients can be identified recursively. The number of recursion is only linear with number of variables in this TPSA. Therefore we used a small hash table H which has a size of $v \times v$. This is not a constraint at all when comparing with the size of TPSA vectors. The location of derivative pattern b_k is

$$C(b_k) = \sum_{i=1}^v H(i, \sum_{j=v}^i b_{k,j})$$

where $H(v, d)$ is the hash function. One example is shown in Table 3. The derivative pattern b_k is a vector represents the order of derivative to each variable. For example, the index of (1 2 1 0) is 46 from Table 3:

$$\begin{aligned}
 C[(0010) * (1200)] &= C(1210) \\
 &= H(1, 0) + H(2, (0+1)) + \\
 &\quad H(3, (0+1+2)) + H(4, (0+1+2+1)) \\
 &= H(1, 0) + H(2, 1) + H(3, 3) + H(4, 4) \\
 &= 0 + 1 + 10 + 35 \\
 &= 46
 \end{aligned} \tag{3}$$

The overall complexity is $O(v)$ when $v \approx d$ and $\binom{d}{v+d}$ is too large to fit in memory.

Once the multiplication is well implemented, the rest is just a trivial application of it to Taylor series expansion. More detailed formula can be found in Ref. [2, 1].

BENCHMARK

This new package is compared with F77 implementation of TPSA (DA), and shown in Fig. 1. This speed test is very coarse, and only used system tools of Linux platform. Our conclusion here is that the new TPSA in C++ is faster than the F77 implementation, and with more features and abilities.

Table 3: Hash Table $H(i, j)$

i	j					
	0	1	2	3	4	5
1	0	1	2	3	4	5
2	0	1	3	6	10	15
3	0	1	4	10	20	35
4	0	1	5	15	35	70

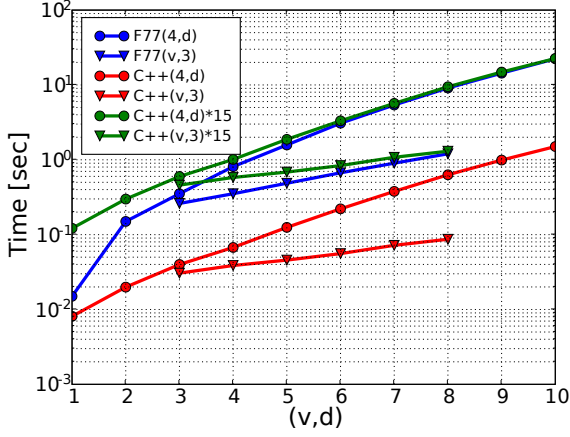


Figure 1: Performance comparison of F77 and C++ version.

CODE EXAMPLE

The following code shows a simple example calculating Taylor series of $f(x_1, x_2) = 1/(x_1 + x_2)$. With operator overloading the code is very straight forward. The package introduces a new type called TPSA, and tries to mimic the native C++ types, such as double, string. I am trying to make it transparent to users.

```
#include <iostream>
2 #include <ctime>
#include "tpsa.hh"
4
using namespace std;
6
int main()
8 {
    // number of variable, degree
10     const unsigned char cnv = 3, cnd = 10;

    // first independent variable, with index 0
    // initialized as 1.0
14     TPSA<double, cnv, cnd> a(1.0L, 0);
    // 2nd, as 2.0
16     TPSA<double, cnv, cnd> b(2.0L, 1);
    // 3rd, as 3.0, not used
18     // TPSA<double, cnv, cnd> c(3.0L, 2);

20     TPSA<double, cnv, cnd> aa; // normal variable

22     aa = 1.0/(a+b);

24     cout << aa << endl; // print out the result.

    // more example:
    // sin(aa), cos(aa), sqrt(aa), tan(aa).

28     return 0;
30 }
```

Computer Codes (Design, Simulation, Field Calculation)

The output of FORTRAN version TPSALib.f and C++ version is in Table. 4.

Table 4: Output of FORTRAN and C++ version

FORTRAN					C++		
I	COEFFICIENT	ORDER	EXPONENTS		V	Base	
1	0.3333333333333333E+00	0	0 0 0		3.333333333333e-01	0	0 0
2	-0.1111111111111111E+00	1	1 0 0		-1.111111111111e-01	1	0 0
3	-0.1111111111111111E+00	1	0 1 0		-1.111111111111e-01	0	1 0
4	0.3703703703703703E-01	2	2 0 0		3.7037037037e-02	2	0 0
5	0.7407407407407407E-01	2	1 1 0		7.4074074074e-02	1	1 0
6	0.3703703703703703E-01	2	0 2 0		3.7037037037e-02	0	2 0
7	-0.12345679012346E-01	3	3 0 0		-1.2345679012e-02	3	0 0
8	-0.3703703703703703E-01	3	2 1 0		-3.7037037037e-02	2	1 0
9	-0.3703703703703703E-01	3	1 2 0		-3.7037037037e-02	1	2 0
10	-0.12345679012346E-01	3	0 3 0		-1.2345679012e-02	0	3 0
11	0.41152263374486E-02	4	4 0 0		4.1152263374e-03	4	0 0
12	0.16460905349794E-01	4	3 1 0		1.6460905350e-02	3	1 0
13	0.24691358024691E-01	4	2 2 0		2.4691358025e-02	2	2 0
14	0.16460905349794E-01	4	1 3 0		1.6460905350e-02	1	3 0
15	0.41152263374486E-02	4	0 4 0		4.1152263374e-03	0	4 0
⋮					⋮		
56	0.56450292694768E-05	10	10 0 0		5.6450292695e-06	10	0 0
57	0.56450292694768E-04	10	9 1 0		5.6450292695e-05	9	1 0
58	0.25402631712645E-03	10	8 2 0		2.5402631713e-04	8	2 0
59	0.67740351233721E-03	10	7 3 0		6.7740351234e-04	7	3 0
60	0.11854561465901E-02	10	6 4 0		1.1854561466e-03	6	4 0
61	0.14225473759081E-02	10	5 5 0		1.4225473759e-03	5	5 0
62	0.11854561465901E-02	10	4 6 0		1.1854561466e-03	4	6 0
63	0.67740351233721E-03	10	3 7 0		6.7740351234e-04	3	7 0
64	0.25402631712645E-03	10	2 8 0		2.5402631713e-04	2	8 0
65	0.56450292694768E-04	10	1 9 0		5.6450292695e-05	1	9 0
66	0.56450292694768E-05	10	0 10 0		5.6450292695e-06	0	10 0

ACKNOWLEDGEMENT

The author would like to thank ALS accelerator physics group at LBNL for their support, Etienne Forest for integration with FPP/PTC, Frank Schmidt for integration with MAD-X.

REFERENCES

- [1] Martin Berz. Differential algebraic description of beam dynamics to very high orders. Particle Accelerators, pages 109–124, 1989.
- [2] Alex Chao. Notes on Topics in Accelerator Physics. 2002.
- [3] MAD-X website, <http://mad.web.cern.ch/mad/>.
- [4] E. Forest, Y. Nogiwa, F. schmidt. The FPP and PTC Libraries.

MOLECULAR DYNAMICS SIMULATION OF CRYSTALLINE BEAMS EXTRACTED FROM A STORAGE RING

Yosuke Yuri

Takasaki Advanced Radiation Research Institute, Japan Atomic Energy Agency,
1233 Watanuki-machi, Takasaki, Gunma 370-1292, Japan

Abstract

It is well-known that a charged-particle beam is Coulomb crystallized in the low-temperature limit. The feasibility of beam crystallization in a storage ring has been raised by the recent progress in beam cooling techniques and in understanding of the behavior of crystalline beams. To go a step further, we here investigate the extraction and transport process of crystalline ion beams, employing the molecular dynamics simulation technique. The dependence of the stability on the lattice of the extraction beam line is explored to show whether various crystalline beams can be transported stably without collapse of the ordered structure.

INTRODUCTION

Applying a dissipative force to a charged-particle beam circulating in a storage ring, we can reduce the emittance of the beam and even expect the occurrence of a type of phase transition. In fact, the beam finally exhibits an ordered structure at the low-temperature limit if some physical conditions are fulfilled. Such an ultimate state of the beam is known as a crystalline beam [1-3] whose spatial structure is determined by the line density [4]. When the line density is sufficiently low, a one-dimensional (1D) string is formed. By increasing the line density, we can transform the string into a two-dimensional (2D) zigzag crystal and, then, eventually into a three-dimensional (3D) shell crystal. In theory, the emittance of a crystalline beam is zero (except for quantum noise) at the low-temperature limit [5, 6].

In order to form and maintain various kinds of crystalline beams, the following two conditions must be satisfied [7, 8]; the storage ring must be operated below the transition energy for crystal formation, and the average betatron phase advance must be less than 127 degrees per lattice period for crystal maintenance. In an ideal crystalline state, the Coulomb repulsive force perfectly balances with the periodic external focusing force of the ring and random interparticle collisions disappear. The crystalline state is stable and thus lasts long even after the cooling force is removed [9]. As reviewed in Ref. [10], according to many advanced theoretical and numerical works, it is now considered that crystalline beams can exist in a storage ring, at least, in theory.

Here, consider the extraction of a crystalline beam from a storage ring. The crystalline beam is additionally kicked by an extraction device such as an electrostatic deflector or a septum magnet and then transported along the (usually, nonperiodic) beam line. The periodicity of the

focusing force is lost, and thus the above-mentioned maintenance condition of crystals is not fulfilled any more. We expect that the emittance of the crystalline beam is increased, or even the crystalline structure of the beam can be destroyed by the extraction process. In order to verify this expectation, molecular dynamics (MD) simulations were carried out. The present MD simulation results show the stability of extracted crystals is consistent with that defined by the maintenance condition of crystals in a ring.

SIMULATION PARAMETERS

For the present simulation study, the MD code "CRYSTAL" is employed [11]. As an extraction device, an electrostatic field can be assumed as well as a normal dipole magnetic field. The equation of motion is integrated by the code in a symplectic manner. In the present MD study, the beam is assumed to be bunched by a longitudinal radio-frequency field in the ring. The Coulomb forces among particles in a bunch are directly calculated. For more detail information on the MD code, see Ref. [10].

As a storage ring lattice, the parameters of S-LSR [12] have been adopted. The operating point of the ring assumed here satisfies the above two conditions. We have also considered several different test extraction lines that have different transverse phase advances from the extraction device to an ideal target position (length: 7.8 m) in order to see how the stability of the extracted beam depends on the lattice design. The main simulation parameters are summarized in Table 1. Here, we show only two cases of high and low phase advances.

MOLECULAR DYNAMICS RESULTS

High-phase-advance lattice

For extraction, a dipole magnetic magnet or

Table 1: Main simulation parameters.

Ring lattice	S-LSR
Ion species	40 keV $^{24}\text{Mg}^+$
Superperiodicity	6
Circumference	22.56 m
Ring tunes (ν_x, ν_y, ν_z)	(1.44, 1.44, 0.15)
Extraction device	Electrostatic or dipole magnetic field
Beam line length	7.8 m
Phase advances of the beam line [deg]	(110, 81), (161, 85), (174, 101), (310, 167), and, (363, 181)

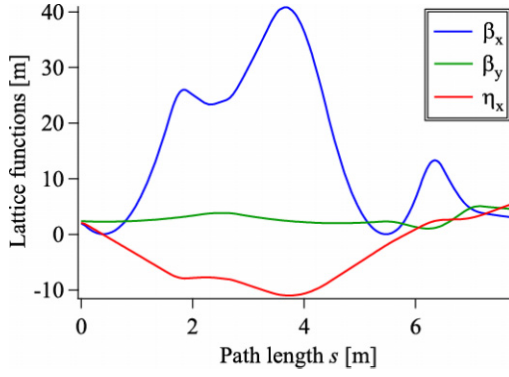


Figure 1: Beta and dispersion functions of the extraction beam line with a high phase advance. The ring betatron tunes have been set at 1.44 in both transverse directions. The horizontal and vertical phase advances from the extraction magnet to the target are, respectively, $\sigma_x = 363$ degrees and $\sigma_y = 181$ degrees.

electrostatic deflector with an extraction angle of 30 degrees and a bending radius of 0.57 m has been assumed in the middle of the straight section in S-LSR. The beta and dispersion functions are shown in Fig. 1. The ring tune has been set at 1.44 in both transverse directions in the present case. The horizontal and vertical phase advances, σ_x and σ_y , from the extraction position to the target are, respectively, 363 degrees and 181 degrees in the case of magnetic extraction. The transverse average phase advance $\sigma_{\perp} \equiv \sqrt{(\sigma_x^2 + \sigma_y^2)}/2$ is 287 degrees.

Figure 2 shows the spatial distributions of a bunched beam composed of 1000 $^{24}\text{Mg}^+$ ions before extraction (i.e., in the ring) and after extraction. The whole of the bunch has been extracted in a single turn. As is clearly seen in Fig. 2, the ordered configuration has been completely lost at the end of the beam line. The final transverse normalized rms emittance is 4×10^{-11} m.rad, which is two orders of magnitude higher than that in the ring.

The single-particle orbits of four ions arbitrarily picked up from the beam in Fig. 2 are shown in Fig.3. Because of the strong focusing of the extraction dipole magnet sitting at $s = 0$ m, the horizontal orbit of each particle crosses the design orbit. This behaviour is different from the fact that single-particle orbits of the crystalline beam in the ring are proportional to each other and oscillate not around the design orbit but around a certain position [5]. We have also confirmed that the emittance abruptly blows up at this position.

At sufficiently low line densities where the corresponding crystalline structure is a 1D string or a 2D zigzag in the ring, the situation has changed; the crystalline structure of a string or a zigzag can be maintained at the end of the beam line after extraction, while an emittance increase is observed. Figure 4 depicts the final spatial distribution of the very-low-intensity 1D string crystalline beam. The final normalized rms emittance of the beam is still very low (3×10^{-19} m.rad).

Beam Cooling

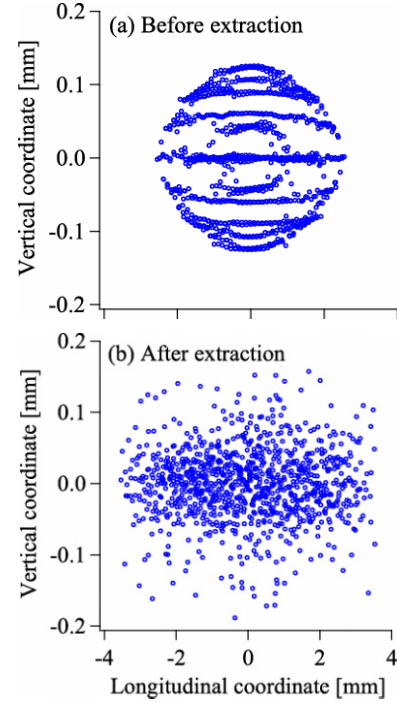


Figure 2: Side views of the bunched beam (a) in the ring before extraction, and (b) at the end of the extraction beam line ($s = 7.8$ m). The lattice of the beam line in Fig. 1 has been assumed. Each circle corresponds to a single $^{24}\text{Mg}^+$ ion with the kinetic energy of 40 keV. The number of ions in a bunch is 1000. The crystalline structure in the upper panel is a two-shell in the middle of the bunch.

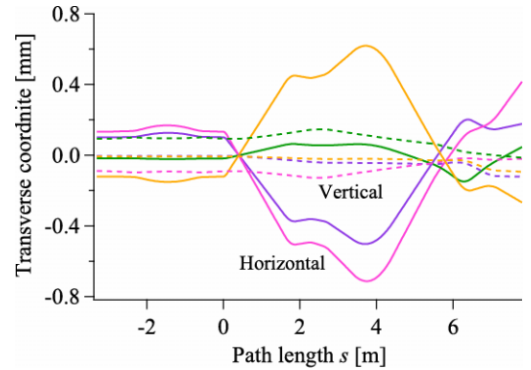


Figure 3: Transverse single-particle orbits of four ions in the beam in Fig. 2. The solid and dashed lines represent the horizontal and vertical orbits, respectively. The origin of the path length s has been set at the entrance of the extraction device. The negative value of s means the path in the ring.

The bunch length of the string becomes larger by about 20% because of no rf focusing in the beam line. The beam exhibits the horizontal head-tail oscillation induced by the momentum dispersion in the beam line [13].

We have replaced the extraction dipole magnet by an electrostatic deflector, and found that the stability of crystals is unchanged.

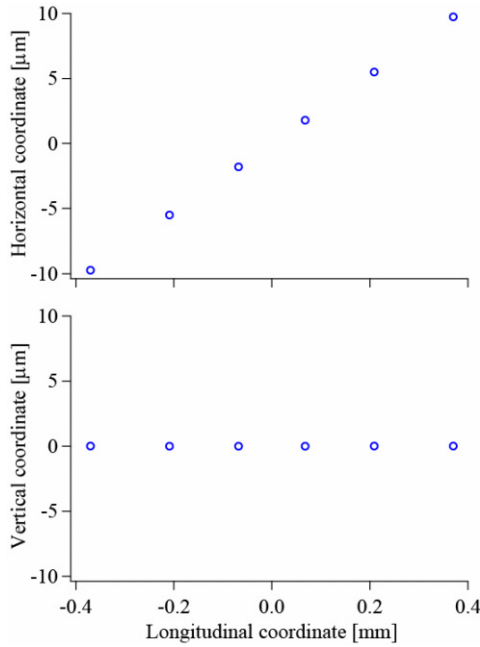


Figure 4: Real-space distribution of a bunched string crystal at the end of the beam line after the extraction. The number of ions in a bunch is six.

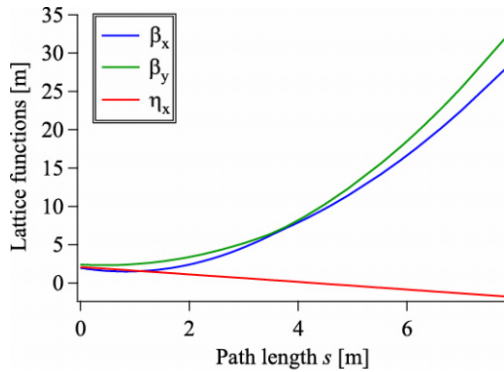


Figure 5: Beta and dispersion functions of the extraction beam line with a low-phase advance. The ring tunes have been set again at 1.44 in both directions. The horizontal and vertical phase advances from the extraction magnet to the target are, respectively, $\sigma_x = 110$ degrees and $\sigma_y = 81$ degrees.

Low-phase-advance lattice

The lattice functions of the low-phase-advance case are shown in Fig. 5. In this example, an extraction angle from the ring has been reduced from 30 deg to 7.5 deg to weaken the focusing effect seen in the previous case. In addition, the beam is not tightly focused to keep the phase advance reasonably low. The transverse average phase advance σ_\perp is 96 degrees in this case.

The final spatial distribution of the beam after extraction is shown in Fig. 6. Unlike the previous case of the high-phase-advance lattice (Fig. 2), the ordered structure of the 3D crystal can be still seen after the extraction. The final emittance of the beam is 3×10^{-12} m.rad. As shown in Fig. 7, the single-particle orbits of the

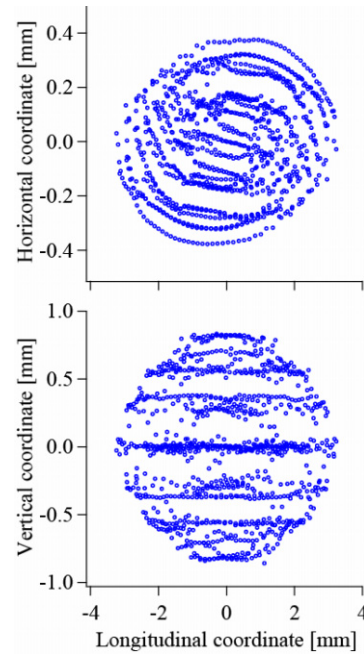


Figure 6: Real-space distribution of the bunched beam at the end of the beam line in the case of low phase advance. The initial distribution in the ring is the same as that in Fig. 2.

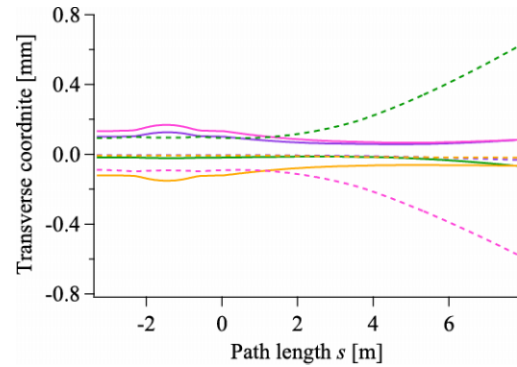


Figure 7: Transverse single-particle orbits of four ions in the beam in Fig. 6. The solid and dashed lines represent the horizontal and vertical orbits, respectively. The orbits of the same four ions as Fig. 3 are plotted. (The orbits of each particle for $s < 0$ are the same as those in Fig. 3.)

beam do not cross the design orbit (although they are not exactly proportional to each other).

At low line densities, it is also possible to extract 1D strings and 2D zigzag crystals stably.

DISCUSSION AND SUMMARY

We have investigated the extraction and transport of various bunched crystalline beams using the MD simulation technique. Once a crystalline beam is extracted from a storage ring, an increase of the emittance is inevitable. This is due to the non-periodic focusing provided by the extraction device and the beam line. However, the crystalline structures of a string or a zigzag

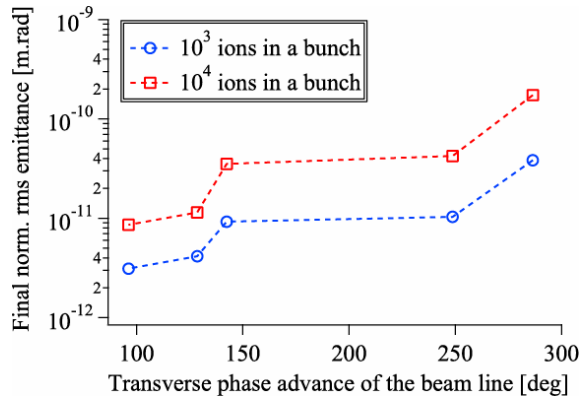


Figure 8: Final transverse rms emittance of the extracted 3D crystalline beam vs. the transverse phase advance of the beam line σ_{\perp} . The number of ions in a bunch are 1000 and 10000, respectively. In the lowest-phase-advance case ($\sigma_{\perp} = 96$ deg), the ordered structure of the beams has been maintained at the end of the beam line.

can be maintained after the extraction in spite of the emittance increase. The ordered configuration of 3D crystals easily melts away when the phase advance of the beam line is large. A reduction of the betatron phase advance of the extraction beam line can suppress the undesirable emittance increase, as shown in Fig. 8. According to the present MD simulation results, it is possible to maintain the 3D ordered configuration when the average phase advance of the beam line is less than 127 degrees, similarly to the maintenance condition in the

ring. For long-distance transport of crystalline beams, it may be necessary to make the lattice of the transport line periodic including the matching of dispersion.

REFERENCES

- [1] J. P. Schiffer and P. Kienle, Z. Phys. A **321**, 181 (1985).
- [2] A. Rahman and J. P. Schiffer, Phys. Rev. Lett. **57**, 1133 (1986).
- [3] J. Wei, X.-P. Li, and A. M. Sessler, Phys. Rev. Lett. **73**, 3089 (1994).
- [4] R. W. Hasse and J. P. Shiffer, Ann. Phys. (N.Y.) **203**, 419 (1990).
- [5] H. Okamoto, Phys. Plasmas **9**, 322 (2002).
- [6] H. Okamoto, H. Sugimoto, and Y. Yuri, to be published in Journal of Plasma and Fusion Research
- [7] J. Wei, H. Okamoto, and A. M. Sessler, Phys. Rev. Lett. **80**, 2606 (1998).
- [8] X.-P. Li, H. Enokizono, H. Okamoto, Y. Yuri, A. M. Sessler, and J. Wei, Phys. Rev. ST Accel. Beams **9**, 034201 (2006).
- [9] J. Wei and A. M. Sessler, in Proceedings of 5th European Particle Accelerator Conference EPAC'96, 1996, p.1179.
- [10] H. Okamoto, H. Sugimoto, Y. Yuri, M. Ikegami, and J. Wei, in these proceedings.
- [11] Y. Yuri and H. Okamoto, Phys. Rev. Lett. **93**, 204801 (2004).
- [12] A. Noda, M. Ikegami, and T. Shirai, New J. Phys. **8**, 288 (2006).
- [13] H. Okamoto, Y. Yuri, and K. Okabe, Phys. Rev. E **67**, 046501 (2003).

List of Authors

Italic papercodes indicate primary authors

— A —			
Ackermann, W.	TH1IODN01	De Gersem, H.	TU3IODN05 , TH4IOPK03
Adams, D.J.	TH3IODN02 , THPSC041	Dechoudhury, S.	THPSC012
Adelmann, A.	WE3IOPK01 , THPSC037	Decyk, V.K.	WE2IOPK03
Ahrens, L. A.	WE3IODN03 , THPSC010	Dehler, M.M.	WE4IOPK04 , THPSC017
Allen, C.K.	WE3IODN01	Dong, X.W.	TH2IOPK04
Ames, F.	THPSC012	Dorf, M.	TH1IOPK04
Andonian, G.	THPSC003 , THPSC003	Dragt, A.	MO3IOPK03
Asova, G.	THPSC004	Durant, M.A.	WE2IOPK05
— B —		— E —	
Baartman, R.A.	TH3IOPK04 , THPSC012	Ellison, J.A.	MO3IOPK04 , TH2IOPK01
Bandura, L.L.	THPSC019	Emery, L.	FR1IOPK09 , THPSC050
Barnard, J.J.	TH1IOPK04	Erdelyi, B.	TH4IODN02 , THPSC018 , THPSC019 , THPSC036
Barth, W.A.	TU1IOPK02 , TH4IOPK02	Esarey, E.	THPSC013
Bartolini, R.	TU3IOPK03 , TU3IOPK03	Evtushenko, P.	THPSC020
Bassi, G.	TH2IOPK01	— F —	
Batygin, Y.K.	THPSC006	Faltens, A.	TH1IOPK04
Bayer, W.B.	TU1IOPK02	Fang, S.X.	TH4IODN04
Beebe-Wang, J.	WE3IODN03	Fawley, W.M.	TU4IOPK02
Bengtsson, J.	TU3IOPK04	Ferracin, P.	TU3IODN03
Berz, M.	MO4IODN05	Fitterer, M.	THPSC030
Bi, Y.J.	WE3IOPK01	Flisgen, T.	THPSC011
Boine-Frankenheim, O.	TU2IOPK02	Forck, P.	TU1IOPK02
Bollen, G.	THPSC006	Franke, S.	TH1IODN01
Borland, M.	WE3IOPK02 , TH2IOPK04 , FR1IOPK09 , THPSC050 , THPSC054	Friedman, A.	TH1IOPK04
Brown, K.A.	WE3IODN03 , THPSC010	Friedman, S.A.	WE2IOPK03
Bruhwyler, D.L.	THPSC013	Fujita, Y.	MO4IODN03
Bylinskii, I.V.	THPSC012	Furman, M.A.	TU4IOPK02
— C —		— G —	
Cai, Y.	MO3IOPK04	Galambos, J.	WE3IODN01
Campbell, C.	THPSC006	Geddes, C.G.R.	TU4IOPK02 , THPSC013
Candel, A.E.	WE4IOPK02	Gerhard, P.	TU1IOPK02
Cary, J.R.	WE2IOPK05	Gjonaj, E.	MO4IODN02
Chao, Y.-C.	THPSC012	Goh, G.	THPSC012
Cheng, G.	TU1IOPK04	Gonin, I.G.	THPSC026
Chu, P.	WE3IODN01	Gorlov, T.V.	TH3IOPK03 , THPSC052
Clemente, G.	TU1IOPK02 , TH4IOPK02	Groening, L.	TU1IOPK02 , TH4IOPK02
Cohen, R.H.	TH1IOPK04	Grote, D.P.	TU4IOPK02 , TH1IOPK04
Corbett, W.J.	THPSC023	Guo, W.	WE3IOPK02 , FR1IOPK09
Cormier-Michel, E.	TU4IOPK02 , THPSC013	— H —	
Cousineau, S.M.	TU1IOPK01	Heinemann, K.A.	TH2IOPK01
Cowan, B.M.	THPSC013	Henestroza, E.	TH1IOPK04
— D —		Hillenbrand, S.	THPSC030
Dahl, L.A.	TU1IOPK02	Hiller, N.	THPSC030
Danilov, V.V.	TU1IOPK01	Hofler, A.S.	THPSC020
Davidson, R.C.	TH1IOPK04	Hofmann, A.	THPSC030
		Hofmann, I.	TU1IOPK02

Holmes, J.A. *TU1IOPK01*, *THPSC052*
Huttel, E. *THPSC030*

— I —

Ikegami, M. *TH1IOPK02*, *TH2IODN01*
Ineichen, Y. *WE3IOPK01*
Ivanov, V. *THPSC021*

— J —

Jameson, R.A. *THPSC031*
Jeon, D. *TU1IOPK02*
Johnson, D.E. *THPSC058*
Jones, B. *TH3IODN02*
Jones, F.W. *TH3IOPK04*
Judin, V. *THPSC030*
Jung, J.-Y. *TH1IOPK04*

— K —

Kabel, A.C. *WE4IOPK02*
Kaganovich, I. *TH1IOPK04*
Kaiser, M. *TU1IOPK02*
Kawaguchi, H. *MO4IODN03*
Khodyachykh, S. *THPSC004*
Kim, H.J. *THPSC022*
Klein, M. *THPSC030*
Ko, K. *WE4IOPK02*
Koch, S. *TU3IODN05*, *TH4IOPK03*
Kolomiets, A. *TH4IOPK02*
Kornilov, V. *TU2IOPK02*
Koscielniak, S.R. *THPSC012*
Krasilnikov, M. *THPSC004*
Kraus, C. *WE3IOPK01*
Kruger, S.E. *WE2IOPK05*
Kwan, J.W. *TH1IOPK04*

— L —

Latina, A. *THPSC026*
Laxdal, R.E. *THPSC012*
Lee, E. P. *TH1IOPK04*
Lee, L. *WE4IOPK02*
Lee, M.J. *THPSC023*
Leemans, W. *THPSC013*
Leitner, M. *TH1IOPK04*
Li, X.S. *TU4IODN01*
Li, Z. *WE4IOPK02*
Lin, F. *MO4IOPK04*
Liu, Y.D. *TH3IOPK01*
Liu, Z. *TU1IOPK01*
Luccio, A.U. *MO4IOPK04*, *THPSC010*
Lund, S.M. *TH1IOPK04*
Lunin, A. *THPSC026*, *THPSC047*

— M —

MacKay, W.W. *THPSC010*

Maier, M.T. *TU1IOPK02*
Makino, K. *MO4IODN05*
Malitsky, N. *MO4IOPK04*
Manikonda, S.L. *MO4IODN05*, *THPSC018*
Marchetto, M. *THPSC012*
Marhauser, F. *TU1IOPK04*, *THPSC020*
Markovik, A. *THPSC028*
Marsching, S. *THPSC030*
Marti, F. *THPSC006*
Maus, J.M. *THPSC031*
Merminga, L. *THPSC012*
Mickat, S. *TU1IOPK02*
Milosic, T. *TU1IOPK02*
Minaev, S. *TH4IOPK02*
Mitchell, C.E. *MO3IOPK03*
Morishita, T. *TH2IODN01*
Morris, J. *WE3IODN03*
Morrissey, D.J. *THPSC006*
Morse, W. *MO4IOPK04*
Müller, A.-S. *THPSC030*
Mustapha, B. *MO4IOPK02*, *FR1IOPK01*

— N —

Naik, V. *THPSC012*
Nemesure, S. *WE3IODN03*
Ng, C.-K. *WE4IOPK02*
Ng, E.G. *TU4IODN01*
Nishimura, H. *THPSC035*
Nissen, E.W. *THPSC036*
Nolen, J.A. *MO4IOPK02*
Novokhatski, A. *THPSC056*, *THPSC057*

— O —

Ohmi, K. *WE4IODN01*, *TH2IOPK02*
Oide, K. *TH2IOPK02*
Okamoto, H. *TH1IOPK02*
Okuda, N. *MO3IODN01*
Okuno, H. *THPSC037*
Onderwater, C.J. *MO4IOPK04*
Orlov, Y.F. *MO4IOPK04*
Orzhekhovskaya, A. *TH4IOPK02*
Ostroumov, P.N. *MO4IOPK02*, *FR1IOPK01*

— P —

Pang, G.K. *THPSC006*
Paul, K. *THPSC013*
Pelaia, T.A. *WE3IODN01*
Pine, B.G. *TH3IODN02*, *THPSC041*
Pöplau, G. *MO4IOPK05*, *THPSC011*, *THPSC028*

— R —

Rao, Y.-N. *TH3IOPK04*
Ratzinger, U. *TH4IOPK02*
Reiche, S. *THPSC003*

Riehl, G.A. **TU1IOPK02**
 Robert-Demolaize, G. **WE3IODN03**
 Robin, D. **FR1IOPK02**
 Romanov, G.V. **THPSC047**
 Roser, T. **THPSC010**
 Ruelas, M. **THPSC003**
 Russell, S.J. **WE3IOPK01**

— S —

Sajaev, V. **WE3IOPK02**, **FR1IOPK09**
 Sako, H. **TH2IODN01**
 Sannibale, F. **FR1IOPK02**
 Satogata, T. **WE3IODN03**
 Schauer, M. **WE2IOPK01**
 Schempp, A. **THPSC031**
 Schietinger, T. **THPSC049**
 Schmidt, F. **WE3IOPK04**
 Schnepf, S. **MO4IODN02**
 Schoefer, V. **WE3IODN03**, **THPSC010**
 Schroeder, C.B. **THPSC013**
 Schussman, G.L. **WE4IOPK02**
 Schwarz, S. **THPSC006**
 Semertzidis, Y. **MO4IOPK04**
 Sen, T. **THPSC022**
 Shang, H. **WE3IOPK02**, **THPSC050**,
THPSC054
 Sharp, W. M. **TH1IOPK04**
 Shasharina, S.G. **WE2IOPK05**
 Shen, G.B. **TH2IODN01**
 Shibata, K. **TH2IOPK02**
 Shishlo, A.P. **WE3IODN01**, **TH3IOPK03**,
THPSC052
 Singh, T.V. **WE2IOPK03**
 Smale, N.J. **THPSC030**
 Soliday, R. **WE3IOPK02**, **THPSC050**,
THPSC054
 Solyak, N. **THPSC026**
 Sonnad, K.G. **THPSC030**
 Steier, C. **FR1IOPK02**
 Stephan, F. **THPSC004**
 Stowell, M.L. **TH1IODN04**
 Suetsugu, Y. **TH2IOPK02**
 Sugimoto, H. **TH1IOPK02**

— T —

Talman, R.M. **MO4IOPK04**
 Tepikian, S. **WE3IODN03**
 Thoma, P. **WE2IOPK01**
 Tian, K. **TU1IOPK04**
 Tiede, R. **TU1IOPK02**, **TH4IOPK02**
 Tsakov, I.I. **THPSC004**

— U —

Uriot, D. **TU1IOPK02**

— V —

van Rienen, U. **MO4IOPK05**, **THPSC011**,
THPSC028
 Vay, J.-L. **TU4IOPK02**, **TH1IOPK04**
 Veitzer, S.A. **WE2IOPK05**
 Verzilov, V.A. **THPSC012**
 Vormann, H. **TU1IOPK02**

— W —

Waldron, W.L. **TH1IOPK04**
 Wan, W. **FR1IOPK02**
 Wang, H. **TU1IOPK04**
 Wang, N. **TH3IOPK01**
 Wang, S. **TH4IODN04**
 Wang, Y. **WE3IOPK02**, **THPSC050**,
THPSC054
 Warnock, R.L. **MO3IOPK04**
 Warsop, C.M. **TH3IODN02**, **THPSC041**
 Weathersby, S.P. **THPSC056**, **THPSC057**
 Wei, J. **TH1IOPK02**
 Weiland, T. **MO4IODN02**, **TU3IODN05**,
TH4IOPK03, **TH1IODN01**
 White, D.A. **TH1IODN04**
 Williamson, R.E. **TH3IODN02**, **THPSC041**
 Wipf, S.G. **WE4IOPK04**
 Wu, J. **THPSC023**

— X —

Xiao, A. **WE3IOPK02**, **TH2IOPK04**,
FR1IOPK09, **THPSC054**
 Xiao, M. **THPSC058**
 Xu, J. **MO4IOPK02**
 Xu, S.Y. **TH4IODN04**

— Y —

Yakovlev, V.P. **THPSC026**
 Yamazaki, I. **TU4IODN01**
 Yan, F. **THPSC012**
 Yang, J.J. **WE3IOPK01**, **TH2IODN04**,
THPSC037, **TH2IODN04**
 Yang, L. **FR1IOPK02**, **THPSC059**
 Yao, H.J. **TH2IODN04**
 Yaramyshev, S.G. **TU1IOPK02**, **TH4IOPK02**
 Yokoya, K. **MO3IODN01**
 Yuri, Y. **TH1IOPK02**, **THPSC061**

— Z —

Zhang, T.J. **TH2IODN04**
 Zhang, Y. **WE4IODN03**
 Zhou, D.M. **TH2IOPK02**
 Zvyagintsev, V. **THPSC012**

Institutes List

ANL

Argonne

- Bandura, L.L.
- Borland, M.
- Dong, X.W.
- Emery, L.
- Manikonda, S.L.
- Mustapha, B.
- Nolen, J.A.
- Ostroumov, P.N.
- Sajaev, V.
- Shang, H.
- Soliday, R.
- Wang, Y.
- Xiao, A.
- Xu, J.

BNL

Upton, Long Island, New York

- Ahrens, L. A.
- Beebe-Wang, J.
- Bengtsson, J.
- Brown, K.A.
- Guo, W.
- Lin, F.
- Luccio, A.U.
- MacKay, W.W.
- Malitsky, N.
- Morris, J.
- Morse, W.
- Nemesure, S.
- Robert-Demolaize, G.
- Roser, T.
- Satogata, T.
- Schoefer, V.
- Semertzidis, Y.
- Shen, G.B.
- Tepikian, S.
- Yang, L.

CEA

Gif-sur-Yvette

- Uriot, D.

CERN

Geneva

- Schmidt, F.

CIAE

Beijing

- Bi, Y.J.
- Yang, J.J.
- Yao, H.J.
- Zhang, T.J.

CLASSE

Ithaca, New York

- Orlov, Y.F.
- Talman, R.M.

CST

Darmstadt

- Schauer, M.
- Thoma, P.

DAE/VECC

Calcutta

- Dechoudhury, S.
- Naik, V.

DESY Zeuthen

Zeuthen

- Asova, G.
- Khodyachykh, S.
- Krasilnikov, M.
- Stephan, F.

Diamond

Oxfordshire

- Bartolini, R.

FZK

Eggenstein

- Smale, N.J.
- Huttel, E.

Fermilab

Batavia

- Gonin, I.G.
- Johnson, D.E.
- Kim, H.J.
- Latina, A.
- Lunin, A.
- Romanov, G.V.
- Sen, T.
- Solyak, N.
- Xiao, M.
- Yakovlev, V.P.

GSI

Darmstadt

- Barth, W.A.
- Bayer, W.B.
- Boine-Frankenheim, O.
- Clemente, G.
- Dahl, L.A.
- Forck, P.
- Gerhard, P.
- Groening, L.
- Hofmann, I.

- Kaiser, M.
- Kornilov, V.
- Maier, M.T.
- Mickat, S.
- Milosic, T.
- Orzhekhovskaya, A.
- Riehl, G.A.
- Vormann, H.
- Yaramyshev, S.G.

HU/AdSM

Higashi-Hiroshima

- Okamoto, H.
- Sugimoto, H.

IAP

Frankfurt am Main

- Jameson, R.A.
- Maus, J.M.
- Ratzinger, U.
- Schempp, A.
- Tiede, R.

IHEP Beijing

Beijing

- Fang, S.X.
- Liu, Y.D.
- Wang, N.
- Wang, S.
- Xu, S.Y.
- Zhang, Y.

INRNE

Sofia

- Tsakov, I.I.

ITEP

Moscow

- Kolomiets, A.
- Minaev, S.

IUCF

Bloomington, Indiana

- Liu, Z.

JAEA/J-PARC

Tokai-Mura, Naka-Gun, Ibaraki-Ken

- Morishita, T.
- Sako, H.

JAEA/Kansai

Kyoto

- Ikegami, M.

JAEA/TARRI

Gunma-ken

- Yuri, Y.

JAI

Oxford

- Bartolini, R.

JLAB

Newport News, Virginia

- Cheng, G.
- Evtushenko, P.
- Hofler, A.S.
- Marhauser, F.
- Tian, K.
- Wang, H.

KEK

Ibaraki

- Ikegami, M.
- Ohmi, K.
- Oide, K.
- Shibata, K.
- Suetsugu, Y.
- Yokoya, K.
- Zhou, D.M.

KIT

Karlsruhe

- Fitterer, M.
- Hillenbrand, S.
- Hiller, N.
- Hofmann, A.
- Judin, V.
- Klein, M.
- Marsching, S.
- Müller, A.-S.
- Sonnad, K.G.

KU Leuven

Kortrijk

- De Gersem, H.

KVI

Groningen

- Onderwater, C.J.

LANL

Los Alamos, New Mexico

- Russell, S.J.

LBNL

Berkeley, California

- Cormier-Michel, E.
- Esarey, E.
- Faltens, A.
- Fawley, W.M.

- Ferracin, P.
- Furman, M.A.
- Geddes, C.G.R.
- Henestroza, E.
- Jung, J.-Y.
- Kwan, J.W.
- Lee, E. P.
- Leemans, W.
- Leitner, M.
- Li, X.S.
- Ng, E.G.
- Nishimura, H.
- Robin, D.
- Sannibale, F.
- Schroeder, C.B.
- Steier, C.
- Vay, J.-L.
- Waldron, W.L.
- Wan, W.
- Yamazaki, I.
- Yang, L.

LLNL

Livermore, California

- Barnard, J.J.
- Cohen, R.H.
- Friedman, A.
- Grote, D.P.
- Lund, S.M.
- Sharp, W. M.
- Stowell, M.L.
- White, D.A.

MSU

East Lansing, Michigan

- Berz, M.
- Makino, K.

Muons, Inc

Batavia

- Ivanov, V.

Muroran Institute of Technology, Department of Electrical and Electronic Engineering

Muroran

- Fujita, Y.
- Kawaguchi, H.

NSCL

East Lansing, Michigan

- Batygin, Y.K.
- Bollen, G.
- Campbell, C.
- Marti, F.
- Morrissey, D.J.
- Pang, G.K.
- Schwarz, S.

Northern Illinois University

DeKalb, Illinois

- Erdelyi, B.
- Nissen, E.W.

ORNL

Oak Ridge, Tennessee

- Allen, C.K.
- Cousineau, S.M.
- Danilov, V.V.
- Galambos, J.
- Gorlov, T.V.
- Holmes, J.A.
- Jeon, D.
- Pelaia, T.A.
- Shishlo, A.P.

PPPL

Princeton, New Jersey

- Davidson, R.C.
- Dorf, M.
- Kaganovich, I.

PSI

Villigen

- Adelman, A.
- Dehler, M.M.
- Ineichen, Y.
- Kraus, C.
- Reiche, S.
- Schietinger, T.
- Wipf, S.G.

RIKEN Nishina Center

Wako

- Okuno, H.

RadiaBeam

Marina del Rey

- Andonian, G.
- Ruelas, M.

Rostock University, Faculty of Computer Science and Electrical Engineering

Rostock

- Flisgen, T.
- Markovik, A.
- Pöplau, G.
- van Rienen, U.

SFU

Burnaby, BC

- Goh, G.

SLAC

Menlo Park, California

- Cai, Y.
- Candel, A.E.
- Chu, P.
- Corbett, W.J.
- Kabel, A.C.
- Ko, K.
- Lee, L.
- Lee, M.J.
- Li, Z.
- Ng, C.-K.
- Novokhatski, A.
- Schussman, G.L.
- Warnock, R.L.
- Weathersby, S.P.
- Wu, J.

STFC/RAL/ISIS

Chilton, Didcot, Oxon

- Adams, D.J.
- Jones, B.
- Pine, B.G.
- Warsop, C.M.
- Williamson, R.E.

TEMF, TU Darmstadt

Darmstadt

- Ackermann, W.
- Franke, S.
- Gjonaj, E.
- Koch, S.
- Weiland, T.

TRIUMF

Vancouver

- Ames, F.
- Baartman, R.A.
- Bylinskii, I.V.
- Chao, Y.-C.
- Jones, F.W.
- Koscielniak, S.R.
- Laxdal, R.E.
- Marchetto, M.
- Merminga, L.
- Rao, Y.-N.
- Verzilov, V.A.
- Yan, F.
- Zvyagintsev, V.

TU Darmstadt

Darmstadt

- Schnepf, S.

TUB

Beijing

- Wei, J.
- Yang, J.J.

Tech-X

Boulder, Colorado

- Bruhwiler, D.L.
- Cary, J.R.
- Cowan, B.M.
- Durant, M.A.
- Kruger, S.E.
- Paul, K.
- Shasharina, S.G.
- Veitzer, S.A.

The University of Liverpool

Liverpool

- Bassi, G.

UCLA

Los Angeles, California

- Andonian, G.
- Decyk, V.K.
- Friedman, S.A.
- Singh, T.V.

UMD

College Park, Maryland

- Dragt, A.
- Mitchell, C.E.

UNM

Albuquerque, New Mexico

- Ellison, J.A.
- Heinemann, K.A.

University of Tokyo

Tokyo

- Okuda, N.

REGISTRATIONS AS OF AUGUST 25, 2009

Abell	Dan	Tech-X Corp.
Ackermann	Wolfgang	Technische Universitaet Darmstadt
Adelmann	Andreas	PSI
Alexahin	Yuri	Fermilab
Amundson	James	Fermilab
Asova	Galina	DESY
Austin	Travis	Tech-X Corporation
Bartolini	Riccardo	Diamond Light Source
Bassi	Gabriele	University of Liverpool
Batygin	Yuri	NSCL
Bauer	Carl	University of Colorado, Boulder
Bell	George	Tech-X Corp.
Bengtsson	Johan	Brookhaven National Laboratory
Berz	Martin	Michigan State University
Bethel	Wes	LBNL
Boehnlein	Amber	DOE/OHEP
Boman	Erik	Sandia National Labs
Borland	Michael	Argonne National Laboratory
Brown	Kevin	Brookhaven National Laboratory
Cai	Yunhai	SLAC
Campos	Pinto Martin	CNRS and IRMA (University of Strasbourg)
Candel	Arno	SLAC National Accelerator Laboratory
Canon	Shane	NERSC/LBL
Chao	Yu-Chiu	TRIUMF
Clemens	Markus	Helmut-Schmidt-University
Clemente	Gianluigi	GSI
Cowan	Benjamin	Tech-X Corporation
D'Angelo	John	USAF/AFRL
De Gersem	Herbert	Katholieke Universiteit Leuven
Decyk	Viktor	UCLA
DeFord	John	STAAR/ AWR Corporation
Dehler	Micha	Paul Scherrer Institut
Di Mitri	Simone	Sincrotrone Trieste

Dimitrov	Dimitre	Tech-X Corp.
Ding	Yuantao	SLAC
Dohlus	Martin	DESY
Dragt	Alex	Physics Department
Ellison	James	UNM
Erdelyi	Bela	NIU/ANL
Ferracin	Paolo	Lawrence Berkeley National Lab
Fiuza	Karen	UMER - UMD
Franke	Sylvain	Technische Universitaet Darmstadt
Friedman	Alex	LLNL & LBNL
Fulginiti	Aurelia	CST of America
Ge	Lixin	SLAC
Geddes	Cameron	LBNL
Geng	Huiping	SLAC
Gorlov	Timofey	ORNL
Groening	Lars	GSI Darmstadt
Grote	Dave	LLNL
Haenichen	Lukas	Technische Universitaet Darmstadt
Heinemann	Klaus	Dept. Mathematics and Statistics
Henshaw	William	Lawrence Livermore National Lab
Hess	Mar	Indiana University Cyclotron Facility
Hofler	Alicia	Jefferson Lab
Holmes	Jeffrey	Oak Ridge National Laboratory
Huang	Chengkun	UCLA/LANL
Ikegami	Masanori	KEK
Ineichen	Yves	ETH/PSI
Ivanov	Valentin	Muons, Inc.
Johnstone	Carol	Fermilab
Jones	Frederick	TRIUMF
Kabel	Andreas	SLAC National Accelerator Laboratory
Kase	Edwin	Tech-X Corporation
Kawaguchi	Hideki	Muroran Institute of Technology
Kim	Hyung	Fermilab
Koch	Stephan	Technische Universitaet Darmstadt

Kornilov	Vladimir	GSI Darmstadt
Krawczyk	Frank	Los Alamos National Laboratory
Lau	Thomas	Technische Universitaet Darmstadt
Limberg	Torsten	DESY
Lin	Fanglei	Brookhaven National Lab
Lin	Ming-Chieh	Tech-X Corporation
Liu	Yu Dong	Institute of High Energy Physics, P. R. China
Lunin	Andrei	Fermilab
Ma	Kwan-Liu	University of California at Davis
Makino	Kyoko	Michigan State University
Manikonda	Shashikant	Argonne National Laboratory
Markovik	Aleksandar	Rostock University
Marsching	Sebastian	Forschungszentrum Karlsruhe
Maus	Johannes	IAP Uni Frankfurt
Mayes	Christopher	Cornell University
Messmer	Peter	Tech-X Corp
Michel	Estelle	LBNL
Min	Misun	Argonne National Laboratory
Mustapha	Brahim	Argonne National Laboratory
Ng	Cho-Kuen	SLAC National Accelerator Laboratory
Nieter	Chet	Tech-X Corporation
Nishimura	Hiroshi	LBNL
Ohmi	Kazuhito	KEK
Okamoto	Hiromi	Hiroshima University
Okuda	Natsuki	University of Tokyo
Park	Chong Shik	Indiana University
Paul	Kevin	Tech-X Corporation
Pieloni	Tatiana	Ecole Polytechnique Federale de Lausanne
Pine	Ben	STFC
Poeplau	Gisela	Rostock University
Prior	Christopher	STFC/RAL
Qiang	Ji	LBNL
Rawat	Vineet	SLAC National Accelerator Laboratory
Reiche	Sven	Paul Scherrer Institut
Roberts	Tom	Muons, Inc.

Rowland	James	Diamond Light Source
Ruelas	Marcos	RadiaBeam Technologies
Rumolo	Giovanni	CERN
Russenschuck	Stephan	CERN
Ryne	Robert	LBNL
Samuel	F. Martins	GoLP/IPFN - Inst. Sup. Tecnico
Schauer	Martin	CST of America
Schietinger	Thomas	Paul Scherrer Institut
Schmidt	Frank	CERN
Schnepp	Sascha	Graduate School CE, TU Darmstadt
Schoefer	Vincent	Brookhaven National Laboratory
Schussman	Greg	SLAC
Shalf	John	LBNL/NERSC
Shang	Hairong	Argonne National Lab
Shasharina	Svetlana	Tech-X Corporation
Shishlo	Andrei	ORNL
Smithe	David	Tech-X Corporation
Solyak	Nikolay	Fermi National Accelerator Laboratory
Stern	Eric	Fermi National Accelerator Laboratory
Takeda	Hiroyuki	RIKEN Nishina Center
Tian	Kai	Jefferson lab
Tse	Eric	SLAC NAL
Valishev	Alexander	Fermilab
Vay	Jean-Luc	Lawrence Berkeley National Laboratory
Vorobiev	Leonid	Fermi National Accelerator Laboratory
Walstrom	Peter	Defense Threat Reduction Agency
Wang	Haipeng	Jefferson Lab
Wang	Nanbor	Tech-X Corporation
Wang	Yusong	Argonne National Laboratory
Warnock	Robert	SLAC National Accelerator Laboratory
Weathersby	Stephen	SLAC
Weiland	Thomas	Technische Universitaet Darmstadt
Werner	Gregory	University of Colorado
White	Daniel	LLNL
Wittberger	Marcus	Paul Scherrer Institute / ETH Zurich
Xiao	Aimin	APS

Xiao	Liling	SLAC
Xiao	Meiqin	Fermilab
Xu	Jin	ANL
Xu	Shouyan	IHEP
Yamazaki	Ichitaro	LBNL
Yang	Jianjun	THU / CIAE
Yang	Lingyun	LBNL
Yelick	Kathy	NERSC / LBNL / UC Berkeley
Yuri	Yosuke	Japan Atomic Energy Agency
Zhang	Yuan	Institute of High Energy Physics
Zhou	Demin	KEK