

**ICAP'18**  
**OCT 20 - 24, 2018**  
**KEY WEST, FL, USA**  
**CASA MARINA RESORT**

**13<sup>th</sup> International Computational Accelerator Physics Conference - [www.icap18.org](http://www.icap18.org)**



## Scientific Program Committee

Martin Berz	Michigan State University	USA
Oliver Boine-Frankenheim	GSI	Germany
Sarah M. Cousineau	ORNL	USA
Herbert De Gersem	Technische Universität Darmstadt	Germany
Kwok Ko	Stanford	USA
Kyoko Makino	Michigan State University	USA
Cho-Kuen Ng	SLAC	USA
Kazuhito Ohmi	KEK	Japan
Dmitri Ovsyannikov	Saint Petersburg University	Russia
Chris Prior	Rutherford Appleton Laboratory	UK
Stephan Russenschuck	CERN	Switzerland
Rob Ryne	LBNL	USA
Frank Schmidt	CERN	Switzerland
Pavel Snopok	Illinois Institute of Technology	USA
Chuanxiang Tang	Tsinghua University	China
Ursula van Rienen	Universität Rostock	Germany
Zhentang Zhao	SINAP	China

## Local Organizing Committee (LOC)

Kyoko Makino (Chair)	Michigan State University	USA
Martin Berz	Michigan State University	USA
Pavel Snopok	Illinois Institute of Technology	USA

## JACoW Editorial Team

Volker RW Schaa (Chief)	GSI	Germany
Pavel Snopok	Illinois Institute of Technology	USA
Joele Mira	iThemba LABS	South Africa

## Local Organizing Team

David Tarazona	Michigan State University	USA
Eremey Valetov	Michigan State University	USA
Adrian Weisskopf	Michigan State University	USA
Kazuko Fuchi	U Dayton	USA
Akiko Yamamoto	Cocokarafine	USA
Yuko Yamamoto	Cocokarafine	USA



# Contents

<b>Preface</b>	<b>i</b>
Committees . . . . .	iii
Contents . . . . .	v
<b>Papers</b>	<b>1</b>
SAPLG01 – Advances in Simulation of High Brightness/High Intensity Beams . . . . .	1
SAPAF01 – Genetic Algorithm Enhanced by Machine Learning in Dynamic Aperture Optimization . . . . .	8
SAPAF02 – Optimization of Heavy-Ion Synchrotrons Using Nature-Inspired Algorithms and Machine Learning . . . . .	15
SAPAF03 – Comparison of Model-Based and Heuristic Optimization Algorithms Applied to Photoinjectors Using Libensemble . . . . .	22
SAPAF04 – Single Objective Genetic Optimization of an 85% Efficient Klystron . . . . .	25
SAPAG03 – Mode-Analysis Methods for the Study of Collective Instabilities in Electron-Storage Rings . . . . .	30
SAPAG04 – HOM-Mitigation for Future SPS 33-Cell 200 MHz Accelerating Structures . . . . .	35
SUPAF04 – Symplectic and Self-Consistent Algorithms for Particle Accelerator Simulation . . . . .	42
SUPAF05 – Polarized Proton Beams From Laser-Induced Plasmas . . . . .	46
SUPAF06 – Simulations of Coherent Electron Cooling With Free Electron Laser Amplifier and Plasma-Cascade Micro-Bunching Amplifier . . . . .	52
SUPAF07 – High-Fidelity Three-Dimensional Simulations of Thermionic Energy Converters . . . . .	59
SUPAF08 – Particle-in-Cell Simulation of a Bunched Electrons Beam Acceleration in a TE113 Cylindrical Cavity Affected by a Static Inhomogeneous Magnetic Field . . . . .	64
SUPAF09 – Sparse Grid Particle-in-Cell Scheme for Noise Reduction in Beam Simulations . . . . .	71
SUPAF10 – Reconstruction of Particle Distributions at RFQ Exit at SNS Beam Test Facility . . . . .	76
SUPAG01 – Space Charge and Transverse Instabilities at the CERN SPS and LHC . . . . .	80
SUPAG03 – Challenges in Extracting Pseudo-Multipoles From Magnetic Measurements . . . . .	87
SUPAG05 – Muon Background Studies for Beam Dump Operation of the K12 Beam Line at CERN . . . . .	93
SUPAG06 – Simulation Challenges for eRHIC Beam-Beam Study . . . . .	99
SUPAG09 – Beam Dynamics Simulations of Medical Cyclotrons and Beam Transfer Lines at IBA . . . . .	104
SUPAG10 – Design Study of a Fast Kicker Magnet Applied to the Beamline of a Proton Therapy Facility . . . . .	110
SUPLG01 – Computational Accelerator Physics: On the Road to Exascale . . . . .	113
MOPLG01 – Challenges in Simulating Beam Dynamics of Dielectric Laser Acceleration . . . . .	120
MOPLG03 – Spin Dynamics in Modern Electron Storage Rings: Computational and Theoretical Aspects . . . . .	127
MOPAF02 – Realistic Modeling of the Muon g-2 Experiment Beamlines at Fermilab . . . . .	134
MOPAF03 – Polarization Lifetime in an Electron Storage Ring, an Ergodic Approach in eRHIC EIC . . . . .	140
MOPAF04 – Spin Dynamics in Modern Electron Storage Rings: Computational Aspects . . . . .	146
MOPAF05 – Approaches to Optimizing Spin Transmission in Lattice Design . . . . .	151
MOPAG01 – Plasma Wakefield Start to End Acceleration Simulations From Photocathode to FEL With Simulated Density Profiles . . . . .	154
MOPAG02 – Efficient Modeling of Laser Wakefield Acceleration Through the PIC Code Smilei in CILEX Project . . . . .	160
TUPAF01 – Upgrade of MAD-X for HL-LHC Project and FCC Studies . . . . .	165
TUPAF02 – SixTrack Project: Status, Runtime Environment, and New Developments . . . . .	172
TUPAF06 – Simulations of Longitudinal Beam Stabilisation in the CERN SPS With BLonD . . . . .	179
TUPAF08 – A Full Field-Map Modeling of Cornell-BNL CBETA 4-Pass Energy Recovery Linac . . . . .	186
TUPAF09 – Multi Pass Energy Recovery Linac Design With a Single Fixed Field Magnet Return Line . . . . .	191
TUPAF10 – Experience With CBETA Online Modeling Tools . . . . .	196
TUPAF12 – Longitudinal Beam Dynamics With a Higher-Harmonic Cavity for Bunch Lengthening . . . . .	202
TUPAF13 – Calculation of the AGS Optics Based on 3D Fields Derived From Experimentally Measured Fields on Median Plane . . . . .	209
TUPAF14 – Analytical Calculations for Thomson Backscattering Based Light Sources . . . . .	215
TUPAF15 – A Holistic Approach to Simulating Beam Losses in the Large Hadron Collider Using BDSIM . . . . .	221
TUPAF16 – Analysis of the Beam Loss Mechanism During the Energy Ramp-Up at the SAGA-LS . . . . .	227
TUPAF19 – pyaopt Optimization Suite and its Applications to an SRF Cavity Design for UEMs . . . . .	229
TUPAF20 – Mean-Field Density Evolution of Bunched Particles With Non-Zero Initial Velocity . . . . .	233
TUPAF22 – FEL Simulation Using the Lie Method . . . . .	240
TUPAF23 – Start-to-End Simulations of THz SASE FEL Proof-of-Principle Experiment at PITZ . . . . .	246
TUPAG01 – Computation of Eigenmodes in the BESSY VSR Cavity Chain by Means of Concatenation Strategies . . . . .	253

TUPAG02 – First Steps Towards a New Finite Element Solver for MOEVE PIC Tracking . . . . .	260
TUPAG04 – Statistical Analysis of the Eigenmode Spectrum in the SRF Cavities with Mechanical Imperfections	265
TUPAG07 – Efficient Computation of Lossy Higher Order Modes in Complex SRF Cavities Using Reduced Order Models and Nonlinear Eigenvalue Problem Algorithms . . . . .	270
TUPAG10 – Nonlinear Optics at UMER: Lessons Learned in Simulation . . . . .	278
TUPAG11 – ESS Accelerator Lattice Design Studies and Automatic Synoptic Deployment . . . . .	285
TUPAG13 – S-Based Macro-Particle Spectral Algorithm for an Electron Gun . . . . .	290
TUPAG14 – Constrained Multi-Objective Shape Optimization of Superconducting RF Cavities to Counteract Dangerous Higher Order Modes . . . . .	293
TUPAG17 – Beamline Map Computation for Paraxial Optics . . . . .	297
TUPAG20 – Computational Beam Dynamics Requirements for FRIB . . . . .	303
TUPAG21 – Novel, Fast, Open-Source Code for Synchrotron Radiation Computation on Arbitrary 3D Geometries	309
TUPAG22 – Main and Fringe Field Computations for the Electrostatic Quadrupoles of the Muon g-2 Experiment Storage Ring . . . . .	313
WEPAF01 – A Compact Permanent Magnet Spectrometer for CILEX . . . . .	320
WEPAF02 – Simulations of Beam Chopping for Potential Upgrades of the SNS LEBT Chopper . . . . .	325
WEPAF04 – Longitudinal Beam Dynamics in FRIB and ReA Linacs . . . . .	330
WEPLG01 – Analysis of Emittance Growth in a Gridless Spectral Poisson Solver for Fully Symplectic Multiparticle Tracking . . . . .	335
WEPLG03 – Theoretical and Computational Modeling of a Plasma Wakefield BBU Instability . . . . .	341
WEPLG05 – Review of Spectral Maxwell Solvers for Electromagnetic Particle-in-Cell: Algorithms and Advantages	345
<b>Appendices</b>	<b>351</b>
List of Authors . . . . .	351
Institutes List . . . . .	357

# ADVANCES IN SIMULATION OF HIGH BRIGHTNESS/HIGH INTENSITY BEAMS

Ji Qiang\*, LBNL, Berkeley, CA, USA

## Abstract

High brightness/high intensity beams play an important role in accelerator based applications by driving x-ray free electron laser (FEL) radiation, producing spallation neutrons and neutrinos, and generating new particles in high energy colliders. In this paper, we report on recent advances in modeling the high brightness electron beam with application to the next generation FEL light sources and in modeling space-charge effects in high intensity proton accelerators.

## START-TO-END SIMULATION OF MICROBUNCHING INSTABILITY EXPERIMENT IN AN FEL LINAC

The x-ray FEL provides a great tool for scientific discoveries in chemistry, physics, biology and material science. The microbunching instability seeded by shot noise and driven by collective effects (primarily space charge), can significantly degrade the quality of the electron beam before it enters the FEL undulators. Without proper control of the instability, the large final electron beam energy spread and phase space filamentation degrade the x-ray FEL performance [1–7]. The microbunching instability experiments recently carried out at the LCLS [8] provides a good opportunity to validate the computational model used in the simulation [9]. In the microbunching measurement at LCLS, the X-band transverse deflecting cavity (XTCAV) diagnostic [10] is located downstream of the undulator before the dump to measure the longitudinal phase space of the electron beam through the entire accelerator. The start-to-end beam dynamics simulations using the real number of electrons were done using a 3D parallel beam dynamics simulation framework IMPACT [11, 12]. It includes a time-dependent 3D space-charge code module IMPACT-T for injector modeling and a position-dependent 3D space-charge code module for linac and beam transport system model. The simulation starts from the generation of photo-electrons at the photocathode following the initial laser pulse distribution and the given initial thermal emittance. The electron macroparticles out of the cathode will be subject to both the external fields from a DC/RF gun and solenoid, and the space-charge/image charge fields from the Coulomb interaction of the particles among themselves. After exiting from the injector, the electron macroparticle will transport through a linear accelerator and beam transport system that includes laser heater, bunch compressors, accelerating RF cavities, harmonic linearizer, and magnetic focusing elements. Besides the 3D space-charge effects, the simulation also includes coherent synchrotron radiation (CSR) effects through a bending magnet,

incoherent synchrotron radiation inside the bending magnet, RF cavity structure wakefield, and resistive wall wakefield. In the simulations, we track the beam down to the XTCAV screen and compare with the measurements. Figure 1 shows the final longitudinal phase space after the XTCAV from the experimental observation and from the simulation with laser heater turned off for the 1 kA study case.

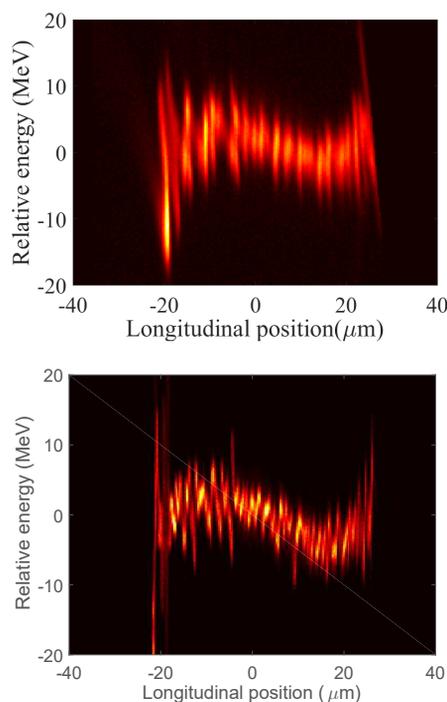


Figure 1: Measurement (top) and simulation (bottom) of the final longitudinal phase space distribution with the laser heater off. Beam current is 1 kA, with bunch charge 180 pC. The bunch head is to the right.

Here, a strong phase space fluctuation due to the microbunching instability can be seen from both the measurement and the simulation. There is no external seeded initial modulation. This large fluctuation arises from the shot-noise inside the beam and is amplified by collective effects, especially space charge effects through the accelerator.

The microbunching instability can be suppressed through Landau damping by increasing the electron beam uncorrelated energy spread before the bunch compressor using the laser heater. Figure 2 shows the final longitudinal phase space after the XTCAV from both the measurement and the simulation with extra 19 keV uncorrelated slice energy spread from the laser heater. The phase space fluctuation is significantly reduced with the use of the laser heater. This is observed in both the measurement and the simulation.

\* jqiang@lbl.gov

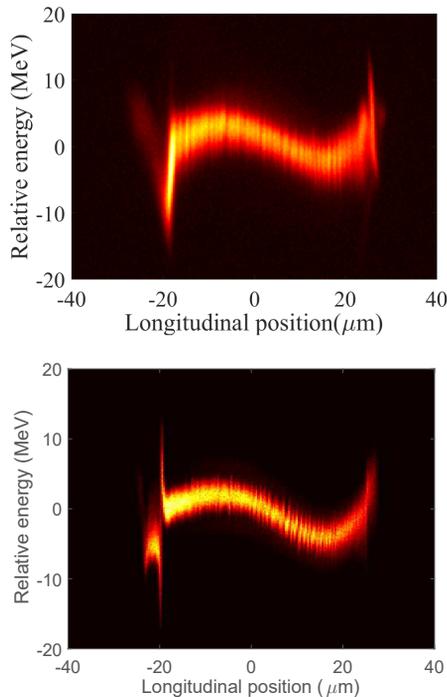


Figure 2: Measurement (top) and simulation (bottom) of the final longitudinal phase space distribution with the laser heater at 19 keV.

The simulation also shows a similar time-energy correlation in the longitudinal phase space to the measurement. The energy dip around the head of the distribution (at  $\sim 15 \mu\text{m}$  in Fig. 2) comes from the effects of resistive wall wakefield in the long, narrow undulator chamber. The dip near the tail of the distribution is due to the longitudinal space-charge and coherent synchrotron radiation effects from the large current spike near the tail of the electron beam.

## GLOBAL OPTIMIZATION OF A NEXT GENERATION LIGHT SOURCE DESIGN

In previous studies, the design optimizations of the injector and the linac were done separately. In recent study, we combined the control parameters in the injector and the linac together into a single group of control parameters through start-to-end simulation for global beam dynamics design optimization [13]. Figure 3 shows a schematic plot of the global optimization including both the injector control parameters and the linac control parameters in the start-to-end beam dynamics optimization. Here, the start-to-end simulation is treated as an objective function in the parallel multi-objective optimizer. The parallel optimizer will call the IMPACT simulation by passing the injector control parameters and the linac control parameters into the objective function. The injector control parameters normally include laser pulse transverse size and length, RF gun amplitude and phase, solenoid strength, buncher and boosting cavity amplitudes and phases. The linac control parameters include linac section 1 cavity amplitude and phase, harmonic linearizer amplitude and phase, bunch compressor 1 bending

angle, linac section 2 cavity amplitude and phase, bunch compressor 2 bending angle, and so on.

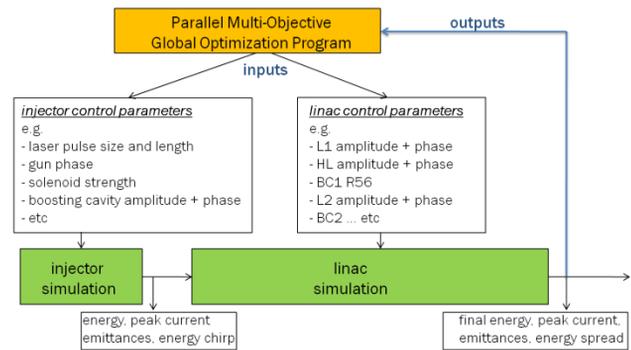


Figure 3: A schematic diagram of the global beam dynamics optimization.

Instead of starting with direct global optimization in the entire control parameter space, we start the optimization with reduced control parameter space that contains only the injector control parameters. The two objective functions, final project transverse emittance and rms bunch length (directly related to peak current) at the exit of the injector are optimized subject to a number of constraints. These constraints are final electron beam energy, beam energy chirp, longitudinal phase space nonlinearity, and so on. After a Pareto optimal front is found for these two objective functions at the exit of the injector, these optimal injector control parameters are combined with some randomly sampled control parameter solutions in the linac. Using the optimal injector control parameters as a partial initial component in the global control parameter solution significantly saves the computational time and speeds up the convergence of the final global solution. During the global beam dynamics optimization, one of the objective (transverse emittance) from the original injector optimization becomes a constraint to the new objective functions. Those solutions at the exit of the injector that can not satisfy this constraint for final start-to-end optimization will be automatically excluded at the beginning of the global optimization. Two objective functions are defined for the global longitudinal beam dynamics optimization. These two functions are fraction of charge and rms energy spread inside a given longitudinal window. The output from the injector such as energy, emittance, and energy spread are used as constraints for the global optimization. Besides the constraint at the exit of the injector, we also put constraints at the final linac output such as energy, peak current etc.

As an application, we applied the above global multi-objective beam dynamics optimization tool to an LCLS-II design optimization with a 20 pC charge. The LCLS-II is a high repetition rate (1 MHz) x-ray FEL that will deliver photons of energy between 200 eV and 5 keV [14, 15]. For the global longitudinal beam dynamics optimization of this accelerator, we have defined 22 control parameters: 12 in the injector, 10 in the linac. Figure 4 shows the Pareto front of the two objective functions from the global opti-

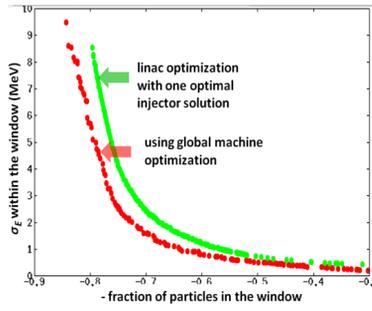


Figure 4: The Pareto front from the global beam dynamics optimization and from the linac only optimization using one optimal injector solution.

mization. These two objective functions are the negative fraction of charge inside and the rms energy spread inside a window between  $-7$  and  $9 \mu\text{m}$ . In this plot, we also show the Pareto front from only the linac optimization using a solution from the injector as an initial distribution. It is seen that the Pareto front from the global optimization is significantly better than that from the linac only optimization. For the same amount of charge inside the window, the global solution has 40% less energy spread in some region. For the same level of the final rms energy spread, the global solution has 15% larger amount of charge. In this simulation besides those constraints for the beam at the exit of the injector, we also put constraints on the final beam energy to be greater than  $3.9 \text{ GeV}$ , final rms energy spread to be less than  $2.5 \text{ MeV}$ , fraction of charge inside the window between  $0.3$  and  $0.9$ . Figure 5 shows the final electron beam current profile from a solution without and with global design optimization. It is seen that the final current profile is significantly improved through the global optimization. This results in more than 50% improvement in the final FEL radiation pulse energy [16].

### A FULLY SYMPLECTIC MODEL FOR SELF-CONSISTENT SPACE-CHARGE SIMULATION

The numerical symplectic integrator is important in long-term tracking simulation in order to preserve phase space structure. In the self-consistent symplectic particle-in-cell (PIC) model [17, 18], macroparticle phase space coordinate advancing through a single step  $\tau$  can be given as:

$$\begin{aligned} \zeta(\tau) &= \mathcal{M}(\tau)\zeta(0) \\ &= \mathcal{M}_1(\tau/2)\mathcal{M}_2(\tau)\mathcal{M}_1(\tau/2)\zeta(0) + O(\tau^3) \end{aligned} \quad (1)$$

where the transfer map  $\mathcal{M}_1$  corresponds to the single particle Hamiltonian including external fields and the transfer map  $\mathcal{M}_2$  corresponds to the space-charge potential from the multi-particle Coulomb interactions. The numerical integrator Eq. 1 will be symplectic if both the transfer map  $\mathcal{M}_1$  and the transfer map  $\mathcal{M}_2$  are symplectic. For a coasting beam inside a rectangular perfectly conducting pipe, the space-charge potential can be obtained from the solution of the

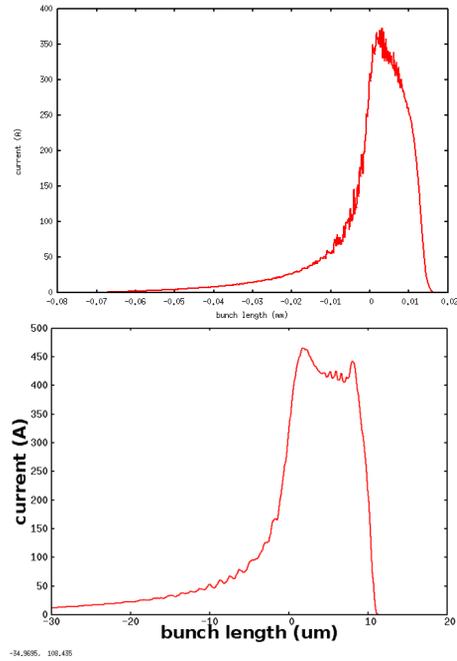


Figure 5: The final electron beam current profile before (top) and after (bottom) global optimization.

Poisson equation using a spectral method. The one-step symplectic transfer map  $\mathcal{M}_2$  of particle  $i$  from the space-charge Hamiltonian is given as:

$$x_i(\tau) = x_i(0) \quad (2)$$

$$y_i(\tau) = y_i(0) \quad (3)$$

$$p_{xi}(\tau) = p_{xi}(0) - \tau 4\pi K \sum_I \sum_J \frac{\partial S(x_I - x_i)}{\partial x_i} \times S(y_J - y_i) \phi(x_I, y_J) \quad (4)$$

$$p_{yi}(\tau) = p_{yi}(0) - \tau 4\pi K \sum_I \sum_J S(x_I - x_i) \times \frac{\partial S(y_J - y_i)}{\partial y_i} \phi(x_I, y_J) \quad (5)$$

where both  $p_{xi}$  and  $p_{yi}$  are normalized by the reference particle momentum  $p_0$ ,  $K = qI/(2\pi\epsilon_0 p_0 v_0^2 \gamma_0^2)$  is the generalized perveance,  $I$  is the beam current,  $\epsilon_0$  is the permittivity of vacuum,  $p_0$  is the momentum of the reference particle,  $v_0$  is the speed of the reference particle,  $\gamma_0$  is the relativistic factor of the reference particle,  $S(x)$  is the unitless shape function (also called deposition function in the PIC model), and the  $\phi$  is given as:

$$\begin{aligned} \phi(x_I, y_J) &= \frac{4}{ab} \sum_{l=1}^{N_l} \sum_{m=1}^{N_m} \frac{1}{\gamma_{lm}^2} \sum_{I'} \sum_{J'} \bar{\rho}(x_{I'}, y_{J'}) \times \\ &\quad \sin(\alpha_I x_{I'}) \sin(\beta_m y_{J'}) \sin(\alpha_I x_I) \sin(\beta_m y_J) \end{aligned} \quad (6)$$

where  $a$  and  $b$  are the horizontal ( $x$ ) and the vertical ( $y$ ) aperture sizes respectively,  $\alpha_l = l\pi/a$ ,  $\beta_m = m\pi/b$ ,  $\gamma_{lm}^2 = \alpha_l^2 + \beta_m^2$ , the integers  $I, J, I'$ , and  $J'$  denote the two dimensional computational grid index, and the summations with

respect to those indices are limited to the range of a few local grid points depending on the specific deposition function. The density related function  $\bar{\rho}(x_{I'}, y_{J'})$  on the grid can be obtained from:

$$\bar{\rho}(x_{I'}, y_{J'}) = \frac{1}{N_p} \sum_{j=1}^{N_p} S(x_{I'} - x_j) S(y_{J'} - y_j), \quad (7)$$

In the PIC literature, compact shape functions are used in the simulation. For example, a quadratic shape function can be written as [19, 20]:

$$S(x_I - x_i) = \begin{cases} \frac{3}{4} - \frac{(x_i - x_I)^2}{\Delta x}, & |x_i - x_I| \leq \Delta x/2 \\ \frac{1}{2} \left( \frac{3}{2} - \frac{|x_i - x_I|}{\Delta x} \right)^2, & \Delta x/2 < |x_i - x_I| \leq 3/2 \Delta x \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$\frac{\partial S(x_I - x_i)}{\partial x_i} = \begin{cases} -2 \frac{(x_i - x_I)}{\Delta x} / \Delta x, & |x_i - x_I| \leq \Delta x/2 \\ \left( -\frac{3}{2} + \frac{(x_i - x_I)}{\Delta x} \right) / \Delta x, & \Delta x/2 < |x_i - x_I| \leq 3/2 \Delta x, x_i > x_I \\ \left( \frac{3}{2} + \frac{(x_i - x_I)}{\Delta x} \right) / \Delta x, & \Delta x/2 < |x_i - x_I| \leq 3/2 \Delta x, x_i \leq x_I \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $\Delta x$  is the mesh size in  $x$  dimension. The same shape function and its derivative can be applied to the  $y$  dimension.

Figure 6 shows the 4D emittance growth  $\left( \frac{\epsilon_x}{\epsilon_{x0}} \frac{\epsilon_y}{\epsilon_{y0}} - 1 \right) \%$  evolution from the symplectic PIC model and those from the nonsymplectic PIC model with the same nominal step size, from the nonsymplectic PIC model with one-half of the nominal step size, and from the nonsymplectic PIC model with one-quarter of the nominal step size. It is seen that as the step size decreases, the emittance growth from the nonsymplectic PIC model converges towards that from the symplectic PIC model.

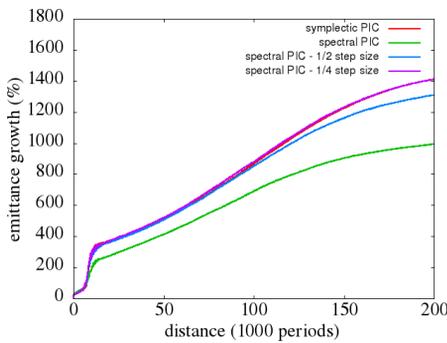


Figure 6: Four dimensional emittance growth evolution from the symplectic PIC model, and the nonsymplectic spectral PIC.

## ANALYSIS AND MITIGATION OF ARTIFICIAL EMITTANCE GROWTH

In the long-term macroparticle space-charge tracking simulation, even with the use of self-consistent symplectic space-charge model, there still exists numerical emittance growth.

SAPLG01

4

The cause of this numerical artificial emittance growth can be understood using a one-dimensional model. Following the spectral method used in the above symplectic PIC model for the space-charge potential, we calculated the sine function expansion mode amplitude from a smooth density distribution function on the grid and from a macroparticle sampled distribution function depositing onto the grid. Here, the amplitude of density mode  $l$  from the sampled macroparticle deposition is given as:

$$\rho^l = \frac{1}{N_p} \frac{2}{N_g \Delta x} \sum_i \sum_I S(x_I - x_i) \sin(\alpha_l x_i) \quad (10)$$

where  $N_p$  is the total number of macroparticles and  $N_g$  is the total number of grid cells. Figure 7 shows the mode amplitude as a function of mode number from the smooth Gaussian function on the grid, from the linear particle deposition, from the quadratic particle deposition, and from the Gaussian kernel particle deposition on the grid using 25,000 macroparticles and 128 grid cells. Here, the Gaussian kernel

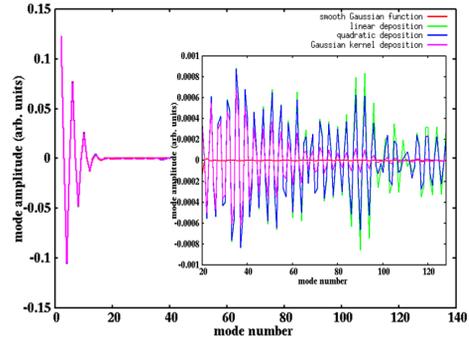


Figure 7: The spectral mode amplitude of a Gaussian distribution as a function of mode number from the smooth Gaussian function on the grid (red), from the linear particle deposition (green), the quadratic particle deposition (blue), and the Gaussian kernel particle deposition on the grid (magenta).

particle deposition shape function is defined as:

$$S(x_I - x_i) = \begin{cases} \exp\left(-\frac{(x_i - x_I)^2}{2\sigma^2}\right); & |x_i - x_I| \leq 3.5\sigma \\ 0; & \text{otherwise} \end{cases} \quad (11)$$

and  $\sigma$  is the chosen as the mesh size. It is seen that for the smooth Gaussian distribution function, with mode number beyond 20, the mode amplitude is nearly zero while the mode amplitude from the macroparticle deposition fluctuates with a magnitude of about  $10^{-4}$ . Those nonzero high frequency modes cause fluctuation in density distribution and induce extra numerical emittance growth. The higher order deposition scheme spreads the macroparticle across multiple grid points and reduces the density fluctuation. However, the Gaussian kernel deposition is computationally much more expensive in comparison to the other two deposition methods.

The above fluctuation of the density mode amplitude from macroparticle deposition can be estimated quantitatively us-

ing the standard deviation (or variance) of the mode amplitude. Given the mode amplitude  $\rho^l$  in Eq. 10, the variance of  $\rho^l$  is given as:

$$\text{var}(\rho^l) = \frac{1}{N_p} \text{var}\left(\frac{2}{N_g \Delta x} \sum_I S(x_I - x_i) \sin(\alpha_l x_i)\right) \quad (12)$$

where

$$\text{var}\left(\frac{2}{N_g \Delta x} \sum_I S(x_I - x_i) \sin(\alpha_l x_i)\right) = \frac{1}{N_p} \left(\frac{2}{N_g \Delta x}\right)^2 \sum_i \left[\sum_I S(x_I - x_i) \sin(\alpha_l x_i)\right]^2 - (\rho^l)^2 \quad (13)$$

Figure 8 shows the mode amplitude standard deviation as a function of mode number for the above Gaussian function by using the linear deposition, the quadratic deposition, and the Gaussian kernel deposition. The mode amplitude standard deviation is small at small mode number and grows quickly to  $10^{-4}$  level and start to decrease after about 10 modes. The standard deviation among the three deposition schemes

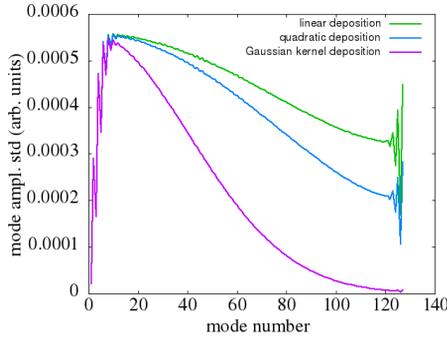


Figure 8: Mode amplitude standard deviation as a function of mode number from the linear particle deposition (green), the quadratic particle deposition (blue), and the Gaussian kernel particle deposition on the grid (magenta) using 25,000 macroparticles and 128 grid cells.

becomes smaller as the order of deposition scheme becomes higher.

The error in the charge density mode amplitude results in error in the solution of space-charge potential and the corresponding force in momentum update in Eqs. 4-5. Assume that the error of force in  $x$  momentum update is  $\delta F$ , after one step  $\tau$ , the emittance growth due to this error will be:

$$\Delta\epsilon \approx (\langle x^2 \rangle \langle x' \delta F \rangle - \langle x x' \rangle \langle x \delta F \rangle) \tau / \epsilon + \frac{1}{2} (\langle x^2 \rangle \langle (\delta F)^2 \rangle - \langle x \delta F \rangle^2) \tau^2 / \epsilon \quad (14)$$

where  $\langle \rangle$  denotes the average with respect to the particle distribution. If  $\delta F$  is a linear function of the position  $x$ , the emittance growth will be zero as expected since the linear force will not change the beam emittance. If  $\delta F$  is a random error force with zero mean and independent of  $x$  and  $x'$ , the emittance growth would be

$$\frac{\Delta\epsilon}{\tau} \approx \frac{1}{2} \langle x^2 \rangle \langle (\delta F)^2 \rangle / \epsilon \quad (15)$$

Assume that this error is due to mode amplitude fluctuation of the finite number of macroparticles sampling, from the above example, we see that  $\langle (\delta F)^2 \rangle \propto 1/N_p$ . This suggests that the numerical emittance growth would decrease as more macroparticles are used. If  $\delta F$  is not a purely random error force (e.g. due to systematic truncation error), the dependence of the emittance growth on the number of macroparticle is more complicated. Figure 9 shows the 4D emittance growth rate from the emittance growth evolution as a function of macroparticle number in a linear FODO lattice and a nonlinear FODO and sextupole lattice. Here, the

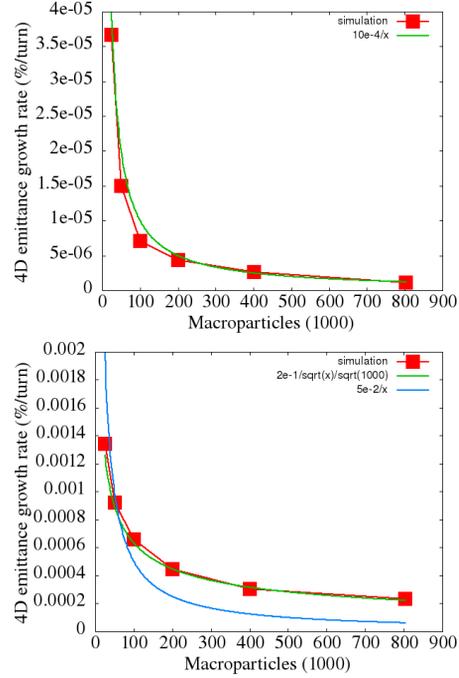


Figure 9: The 4D emittance growth rate as a function of the simulation macroparticle number using a linear FODO lattice (top) and a FODO and sextupole lattice (bottom).

lattice consists of 10 focusing-drift-defocusing-drift (FODO) lattice periods and one sextupole element per turn. The zero current tune of the lattice is 2.417. With 30 A beam current, the corresponding linear space-charge tune shift is 0.113. It is seen that in the linear lattice (no sextupole), the emittance growth rate scales as  $1/N_p$  which is expected from the random sampling errors. In a nonlinear lattice, the emittance growth rate scales close to  $1/\sqrt{N_p}$ . This slower scaling with respect to the macroparticle number  $N_p$  might be due to the interaction between the numerical force error and the nonlinear resonance.

The charge density fluctuation from the macroparticle sampling can be further smoothed out by using a numerical filter in frequency domain besides employing the shape function for particle deposition. Instead of using a standard cut-off method that removes all modes beyond a given mode number (i.e. cut-off frequency), we proposed using an amplitude threshold method to remove unwanted modes. The mode with an amplitude below the threshold value is re-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

moved from the density distribution. The advantage of this method is instead of removing all high frequency modes, it will keep the high frequency modes with large amplitudes. These modes can represent real physics structures inside the beam. The threshold also removes the unphysical low frequency modes associated with the small number of macroparticle sampling. Here, we explored two threshold methods. In the first threshold method, the threshold value is calculated from a given fraction of the maximum amplitude of the density spectral distribution. In the second method, the threshold value is defined as a few standard deviations of the mode amplitude as shown in the one-dimensional Gaussian function example. The mode with an amplitude below the threshold value is regarded as numerical sampling error due to the use of small number of macroparticles and is removed from the density distribution. The advantage of the first method is that the threshold value is readily attainable from the density spectral distribution. The disadvantage of this method is that the threshold fraction is an external supplied hyperparameter. The advantage of the second method is that the threshold value is calculated dynamically through the simulation. The disadvantage of this method is the high computational cost to obtain the standard deviation of each mode. The total computational cost of those standard deviations is proportional to the number of modes multiplied by the number of macroparticles.

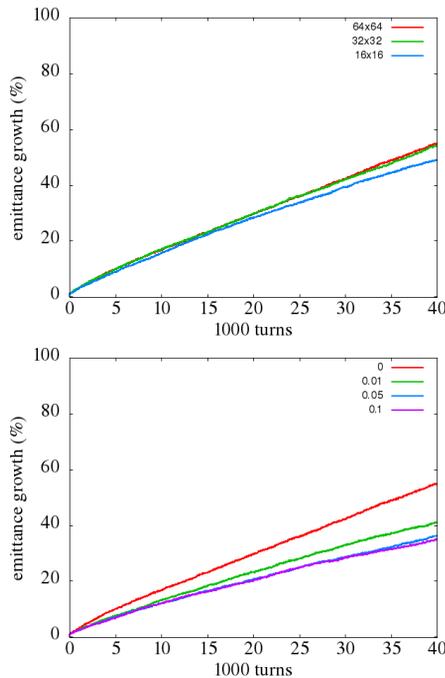


Figure 10: The 4D emittance growth using  $64 \times 64$ ,  $32 \times 32$ ,  $16 \times 16$  modes (top) and with 0 (no filtering) with 0.01, 0.05 and 0.1 threshold filtering (bottom) of charge density distribution using 25k macroparticles in a FODO and sextupole lattice.

We ran the simulation of 30A proton beam transport in the lattice including nonlinear sextupole element. The 4D

SAPLG01

emittance growth evolutions using the brute force cut-off and the threshold filtering are shown in Fig. 10. It is seen that even with  $16 \times 16$  mode cut-off filtering, there still exists significant emittance growth, while a threshold value 0.1 helps significantly lower the emittance growth. Using the four-sigma standard deviation threshold value yields similar emittance growth to the fraction threshold (0.1) as shown in Fig. 11. Those emittance growths include both the physical resonance driven emittance growth and the numerical error driven artificial emittance growth.

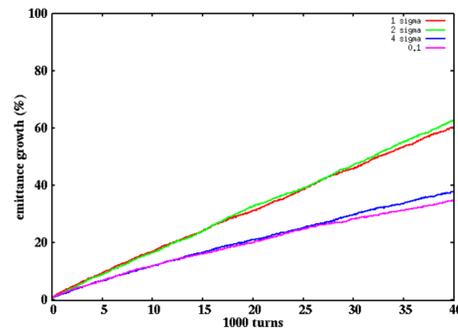


Figure 11: 4D emittance growth with one sigma, two sigma, four sigma standard deviation and 0.1 maximum amplitude threshold filtering of charge density distribution using 25k macroparticles in a FODO and sextupole lattice.

## SUMMARY

In this paper, we have shown that the microbunching instability associated with the high brightness electron beam in a x-ray FEL linac experiment can be well reproduced through the start-to-end simulation using real number of electrons. The global design optimization including both the injector control parameters and the linac control parameters significantly improves the final electron beam longitudinal phase space distribution. The accuracy of simulating a high intensity proton beam can be improved through the use of a fully self-consistent symplectic space-charge model. The artificial numerical emittance growth in the long-term space-charge simulation can be mitigated by using a threshold based numerical filter in frequency domain.

## ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and used computer resources at the National Energy Research Scientific Computing Center.

## REFERENCES

- [1] M. Borland *et al.*, *Nucl. Instrum. Methods Phys. Res.*, Sect. A, vol. 483, pp. 268, 2002.
- [2] E. L. Saldin, E. A. Schneidmiller, and M. V. Yurkov, *Nucl. Instrum. Methods Phys. Res.*, Sect. A, vol. 483, pp. 516, 2002.
- [3] Z. Huang, M. Borland, P. Emma, J. Wu, C. Limborg, G. Stupakov, and J. Welch, *Phys. Rev. ST Accel. Beams*, vol. 7, pp. 074401, 2004.

- [4] M. Venturini, R. Warnock, and A. A. Zholents, *Phys. Rev. ST Accel. Beams*, vol. 10, pp. 054403, 2007.
- [5] J. Qiang, R. D. Ryne, M. Venturini, A. A. Zholents, and I. V. Pogorelov, *Phys. Rev. ST Accel. Beams*, vol. 12, pp. 100702, 2009.
- [6] S. Di Mitri, M. Cornacchia, S. Spampinati, and S. Milton, *Phys. Rev. ST Accel. Beams*, vol. 13, pp. 010702, 2010.
- [7] J. Qiang, C. E. Mitchell, M. Venturini, *Phys. Rev. Lett.*, vol. 111, pp. 054801, 2013.
- [8] D. Ratner, C. Behrens, Y. Ding, Z. Huang, A. Marinelli, T. Maxwell, and F. Zhou, *Phys. Rev. ST Accel. Beams*, vol. 18, pp. 030704, 2015.
- [9] J. Qiang, Y. Ding, P. Emma, Z. Huang, D. Ratner, T. O. Raubenheimer, and F. Zhou, *Phys. Rev. Accel. Beams*, vol. 20, pp. 054402, 2017.
- [10] C. Behrens *et al.*, *Nature Communications*, vol. 5, pp. 3762, 2014.
- [11] J. Qiang, R. Ryne, S. Habib, V. Decyk, *J. Comp. Phys.*, vol. 163, pp. 434, 2000.
- [12] J. Qiang, S. Lidia, R. D. Ryne, and C. Limborg-Deprey, *Phys. Rev. ST Accel. Beams*, vol. 9, pp. 044204, 2006.
- [13] J. Qiang, "Start-to-end beam dynamics optimization of x-ray FEL light source accelerators," in *Proc. NAPAC'16*, Chicago, IL, USA, Oct 2016. doi:10.18429/JACoW-NAPAC2016-WEA3I002
- [14] T. O. Raubenheimer, "Technical challenges of the LCLS-II CW X-ray FEL," in *Proc. IPAC'15*, Richmond, VA, USA, May 2015. doi:10.18429/JACoW-IPAC2015-WEYC1
- [15] P. Emma *et al.*, "Linear accelerator design for the LCLS-II FEL facility," in *Proc. FEL'14*, Basel, Switzerland, THP025, Aug. 2014.
- [16] G. Marcus and J. Qiang, "LCLS-II SCRF start-to-end simulations and global optimization as of September 2016," LCLS-II TN-1704, 2017.
- [17] J. Qiang, *Phys. Rev. Accel. Beams*, vol. 20, pp. 014203, 2017.
- [18] J. Qiang, *Phys. Rev. Accel. Beams*, vol. 21, pp. 054201, 2018.
- [19] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Adam Hilger, New York, 1988.
- [20] C. K. Birdsall and A. B. Langdon, *Plasma Physics Via Computer Simulation*, Taylor and Francis, New York, 2005.

# GENETIC ALGORITHM ENHANCED BY MACHINE LEARNING IN DYNAMIC APERTURE OPTIMIZATION

Yongjun Li,\* Weixing Cheng, Li Hua Yu, Robert Rainer,  
 Brookhaven National Laboratory, Upton, New York 11973

## Abstract

With the aid of machine learning techniques, the genetic algorithm has been enhanced and applied to the multi-objective optimization problem presented by the dynamic aperture of the NSLS-II storage ring.

## INTRODUCTION

Population-based optimization techniques, such as evolutionary (genetic) [1–16] and particle swarm [17–19] algorithms, have become popular in modern accelerator design. Optimization of a nonlinear lattice’s dynamic aperture usually has multiple objectives, such as the area and the profile of the dynamic aperture, energy acceptance, beam lifetime [1, 3], and nonlinear driving terms (NDT) [4] etc. Dynamic aperture and energy acceptance can be evaluated through direct single-particle tracking simulations. NDTs can be extracted analytically from the one-turn-map for a given nonlinear lattice configuration [20–23]. Recent studies have found that the spread from a constant of the action obtained with the square matrix method [24–27] represents a kind of nonlinearity measure of a lattice, which can be treated as an optimization objective as well. Another optimization objective, which is deduced from the square matrix method and used in this paper, is the spread of linear action  $J_{x,y}$  from a constant. The spread is numerically computed from simulated turn-by-turn data [28, 29]. Based on the number of objectives presented in this application, multi-objective genetic algorithm (MOGA) [30] is a suitable optimization tool to compromise among these objectives simultaneously.

A general model for multi-objective optimization is:

- given a set of free variables  $x_n$  within the range  $x_n \in [x_n^L, x_n^U]$ ,  $n \in [1, N]$ ;
- subject to some constraints  $c_j(x_n) \geq 0$ ,  $j \in [1, J]$ , and  $e_k(x_n) = 0$ ,  $k \in [1, K]$ ;
- simultaneously minimize a set of objective functions  $f_m(x_n)$ ,  $m \in [2, M]$ .

Here  $x_n^L$  and  $x_n^U$  are the lower and the upper boundaries of the  $n^{\text{th}}$  free variables.  $N, J, K$  and  $M$  are non-negative integers. Note for simplicity, clarity, and without loss of generality, all constraints are lower bounds, and all objectives are minimized.

A genetic algorithm (GA) is a type of evolutionary algorithm. It can be used to solve both constrained and unconstrained optimization problems based on a natural selection

process [30]. Each candidate has a set of free variables which it inherits from its parents and is mutated at random corresponding to a certain probability. Each candidate’s free variables  $x_n$  can be regarded as an  $N$ -dimensional vector  $\mathbf{x}$ . Their ranges  $[x_n^L, x_n^U]$  define a volume of an  $N$ -dimensional “search space”. The evolution is an iterative process. The new population from each iteration is referred to as a “generation”. The process generally starts with a population that is randomly generated and the fitness of the individuals is evaluated. Individuals with greater fitness are randomly selected, and their genomes are modified to form the next generation. The average fitness of each generation therefore increases with each iteration of the algorithm. The goal of multi-objective optimization (MO) is to optimize functions simultaneously. These functions are sometimes related and their objectives may conflict. In these events, trade-offs are considered among the objectives. In non-trivial MO problems the objectives conflict such that none can be improved without degrading others in value and are referred to as non-dominated or “Pareto optimal”. In these cases a non-dominated sorting algorithm can be used to judge if one candidate is better than another [30]. In the absence of constraints or preferences, however, all Pareto optimal candidates are equally valid and given the same rank. If constraints are provided, the rank of each individual accounts for the constraints, and qualified candidates are guaranteed to dominate unqualified ones. Each qualified candidate has  $M$  fitness values  $f_m$ , which compose another  $M$ -dimensional “fitness space”. The combination of multi-objective, non-dominated sorting with employment of the genetic algorithm forms the basis of the “MOGA” method. MOGA has some limitations in its application to modern storage ring optimization. In general, the application of MOGA on dynamic aperture optimization can be driven by either direct particle tracking, or analytical calculation of nonlinear characterization. It is time-consuming to evaluate the fitness quantitatively, as seen with the calculation of a large-scale storage ring’s dynamic aperture using the symplectic integrator [31].

Although there is no a priori reason why the genetic evolution process needs external intervention, examples without it such as the evolution of biological life on earth or planetary formation in the solar system, were only possible after billions of years [32]. One reason why natural evolution is comparatively slow is that the percentage of elite candidates among the whole population is low. A brute force method for speeding up evolution is to narrow down the search ranges around good candidates found early in the evolution process. This decreases diversity, however, and could lead to a trapping in local minima. An effective intervention step would be able to significantly speed up the evolution in the desired

\* Email: yli@bnl.gov

direction. To do so, some machine learning techniques are introduced to traditional MOGA methods to intervene on the natural process.

## MOGA ENHANCED BY MACHINE LEARNING

During the evolution process, MOGA produces a large data pool. It is possible to reuse the data with machine learning techniques to intervene on the evolution process. Here an intervention method is introduced which is schematically illustrated in Fig. 1. It includes the classification of the search space (unsupervised learning), sorting based on the average fitness and repopulation of potential elite candidates (supervised learning). Starting with randomly distributed individuals, the initial population is allowed to produce descendants via the traditional genetic algorithm. Once all candidates satisfy some desired constraints, sufficient data is accumulated to intervene on the evolution process using machine learning techniques. For each following generation, all populations are classified into different clusters in the search space based on a parameter  $D$ ,

$$D = \sqrt{\sum_{n=1}^N (x_{1,n} - x_{2,n})^2}, \quad (1)$$

which represents the “Euclidean distance” between two candidates  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in the search space. The classification was performed with the K-means algorithm [33] as shown in the subplot (b) of Fig. 1.

After classifying candidates into different clusters, a statistical analysis is carried out on each cluster to evaluate their average or weighted fitness  $F$ , which reads as

$$F = \sum_{m=1}^M w_m f_m(x_n). \quad (2)$$

Here  $w_m$  is the weight on the  $m^{\text{th}}$  fitness value of  $f_m$ . As mentioned previously, our optimization has multiple objectives. Within each generation, most of the candidates belong to the same rank on the Pareto front. Although they are equally good (they exhibit no dominance) and a lot of candidates have one or two good fitness values, the rest have poor fitness. They can survive through many generations unless a constraint is imposed. These types of candidates, however, often have poor trade-offs with conflicting objectives. Weighted fitness  $F$  as a measure for implementing machine learning is therefore introduced. If all weights  $w_m$  are chosen to be  $1/M$ ,  $F$  becomes the average fitness.

The weighted fitness of individuals in each cluster are then evaluated and sorted as illustrated in the subplot (c) in Fig. 1. A few of the better clusters are then selected and labeled with the “elite” status. Some arbitrary number of new candidates (for example, 20% of the total population) are repopulated uniformly and randomly within the narrow “elite range” of these elite clusters within the search space. Since these newly populated candidates share some similarities in

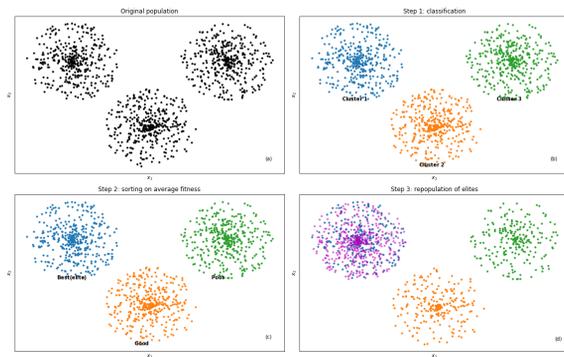


Figure 1: Schematic illustration of intervention using machine learning techniques. Here, a search space of two free variables is assumed. The distribution of the original population is shown in the subplot (a). The candidates are classified into three colored clusters with the K-means algorithm in the subplot (b). The average fitness of each cluster has been evaluated, sorted, and given a status labeled with “Best (elite)”, “Good”, and “Poor” respectively in the subplot (c). In the subplot (d), some potential competitive candidates (marked as the magenta dots) are repopulated inside the range of the “Best (elite)” cluster and then are used to replace the same amount of candidates from the original data pool. After the replacement, the post-population densities of the “Good” and the “Poor” cluster become low. In reality, there may not exist obvious boundaries to separate each cluster and cluster classification is not unique either.

the search space with the elite candidates thus far, they are expected to be more competitive in regard to survivability. From the original population, the same amount of candidates are randomly selected, to be replaced by the newly populated candidates. The average fitness within each generation should therefore increase respectively. This could potentially improve the probability of producing more competitive descendants favored by the optimization goals. While the next generation undergoes the same intervention, the elite range for the following repopulation of descendants will be dynamically re-defined by its own elite clusters. Note that the average fitness is used to define the elite range for repopulation. These repopulated candidates are not guaranteed the “privilege” of being “winners” in each generation. The final candidates still need to be selected through the non-dominated sorting. Considering that general fitness could have different scales in each dimension, they may need to be normalized within a similar range, usually  $\in [0, 1]$ , prior to averaging them [34].

Thus far the proportion of the replacement at each generation is set to a constant value. This is referred to as the Static Replacement Method (SRM). For the SRM, the proportion of replacement is arbitrary, but it is necessary to maintain diversity among the candidates to avoid traps at local minima. When the search space is too large, the distances  $D$  between candidates within the same cluster are far. In this case, it is likely that intervention would mislead evolution because the

expectation on the “elite range” may not be accurate. An optional dynamic replacement method can be used to mitigate this issue. To judge how likely an “elite range” can produce competitive candidates, a supervised machine learning technique is adopted. First, the candidates of an elite cluster are divided into a training set (usually around 90-95% of its population) and a testing set (the residual 5-10%). With the training set data, a learning model (hypothesis)  $H$  using the K-nearest neighboring (KNN) regression algorithm [33] is created. The model is used to predict the testing set’s fitness (prediction). A comparison of the prediction and each individual’s evaluated fitness value can determine the accuracy of the prediction. The comparison is quantitatively measured by a parameter “discrepancy”  $S$  in the fitness space,

$$S = \frac{1}{M} \sqrt{\sum_{m=1}^M \frac{|f_m - h_m|^2}{f_m^2}}. \quad (3)$$

Here,  $h_m$  is the  $m^{\text{th}}$  fitness value predicted from the learning model  $H$  and  $f_m$  is the actual fitness value. In this case,  $f$  is evaluated from a lattice characterization code.  $S = 0$  means they are exactly same. A large  $S$  indicates a large discrepancy between the hypothesis model and the actual value. Based on the average discrepancy of the testing set, the replacement proportion for the population can be dynamically adjusted on a generation basis.

## MOGA APPLICATION AT NSLS-II

The NSLS-II storage ring lattice [35] is used as an example to demonstrate the application of this method. The goal is to optimize the dynamic aperture of the operational lattice. The linear chromaticity is corrected to +2 by chromatic sextupoles. The free “tuning knobs” are six families of harmonic sextupoles with fixed polarities.

The spreads of the linear actions  $J_{x,y}$  computed from turn-by-turn particle tracking simulation are chosen as the optimization objectives. The linear action  $J_u$  is defined as

$$J_{u,i} = \beta_u u_i^2 + 2\alpha_u u_i p_{u,i} + \gamma_u p_{u,i}^2 = \bar{u}_i^2 + \bar{p}_{u,i}^2, \quad (4)$$

where  $u_i = (x, y)_i$  and  $p_{u,i} = p_{(x,y),i}$  are the turn-by-turn coordinates in the horizontal and vertical planes respectively.  $\bar{u} = \frac{1}{\sqrt{\beta_u}} u$ ,  $\bar{p}_i = \frac{1}{\sqrt{\beta_u}} (\alpha_u u + \beta_u p_u)$  are a pair of normalized canonically conjugated coordinates, and  $\alpha$  and  $\beta$  are the linear lattice optics Twiss parameters at the observation point. In the presence of nonlinear magnets, the linear actions have some spread from constants, as illustrated in Fig. 2. Typically the spread gradually increases with betatron oscillation amplitude. In order to obtain a sufficient dynamic aperture, control of the nonlinearity of motion for particles starting from different initial conditions (amplitudes) is needed. Here, five sets of initial conditions are chosen as shown in Fig. 3. The objectives are ten spreads of actions under different sextupole settings (each initial condition has both  $\Delta J_{x,rms}/J_x$  and  $\Delta J_{y,rms}/J_y$ ). For each candidate, the constraint is that all five particles can survive

for multiple turns. All objectives outlined thus far are required to be equally important to ensure that there are no “holes” (particle loss) inside the dynamic aperture.

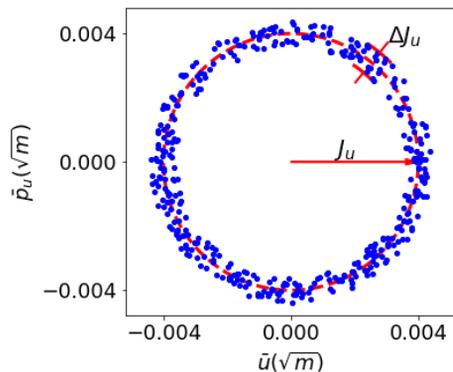


Figure 2: The root means squared (rms) spread of action from a constant is used as an optimization objective. The dashed circle represents a constant linear action at different angles. The dots are the normalized turn-by-turn coordinates.

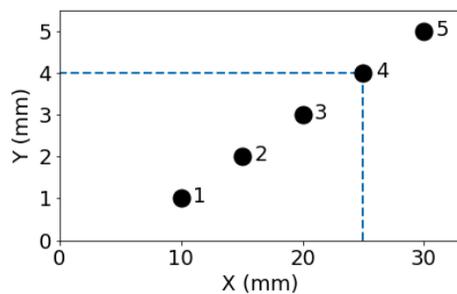


Figure 3: Five initial particle coordinates in the x-y plane with their conjugate momenta  $p_{x,y} = 0$  used for tracking. The turn-by-turn data are used to evaluate the spread of their linear actions. The dashed line is the size of the desired dynamic aperture. The 5<sup>th</sup> particle is chosen beyond the desired dynamic aperture in order to obtain a safe margin. The choice of the initial coordinates is not arbitrary. It may depend on the local optics functions, and physical aperture, etc.

To begin with, a random distribution is chosen in which the entire population is uniformly distributed within ranges limited by field saturation of the magnets and power supply capacity. In the NSLS-II ring, the search space at each sextupole dimension is  $K_2 \in [0, \pm 40] \text{ m}^{-3}$  (Here  $\pm$  is chosen depending on its polarity). Initially a population total of 5,000 is cast. For the first several generations, many candidates cannot survive under 5 initial conditions for dozens of turns. Therefore, the initial population evolves under the initial constraint of self-survival. After the evolution of 6-7 generations, all candidates can survive, but with very poor average fitness (see Fig. 4). Thus far sufficient data may have

already been accumulated to allow the optimizer to learn from the history.

The K-means algorithm is then applied (using an unsupervised learning technique) to classify the total population into  $N = 100$  clusters in the search space and each individual cluster's average fitness is evaluated and sorted. The top three elite clusters are selected, based on their average fitness, to define an "elite" search range. Within this range, 20% of the total population is uniformly repopulated by random candidates. After this intervention (repopulation), the optimizer enters the next iteration.

With intervention, a fast convergence in the average fitness has been observed during evolution. Fig. 4 compares the evolution of the average fitness of MOGA with and without machine learning. Without machine learning, the improvement of fitness relies heavily on random crossover and mutation and global evolution can sometimes stop, or even regress. With the implementation of machine learning, however, the fitness convergence becomes not only faster, but much more steady. More importantly, the amount of competitive candidates is significantly increased, which allows for analysis of the distribution of optimal candidates in the search space.

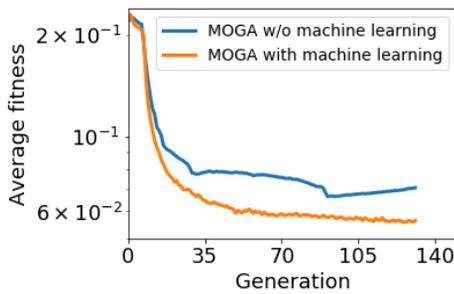


Figure 4: Comparison of the evolution of average fitness with and without machine learning for 135 generations. Without machine learning, the evolution process can sometimes stop, or even regress. On the other hand, the fitness convergence becomes faster and steadier with the introduction of machine learning.

With each generation, all candidates are re-classified. The elite ranges for repopulation also vary as shown in Fig. 5. The ranges are observed to fluctuate, but gradually converge during the evolution. For some free variables, the ranges converged quickly to a small range of optimal values. For example, the *SL1* sextupole's elite range shifts toward zero (limited by its polarity). This sextupole can therefore either be removed from the lattice, or have its polarity changed to see if machine performance can be further improved.

In the final generation's population, most of the candidates are found on the Pareto front. Among them, many have good average fitness. They are reclassified in the search space to study their distribution. These candidates appear to belong to many distinct groups. Each group is like an isolated island

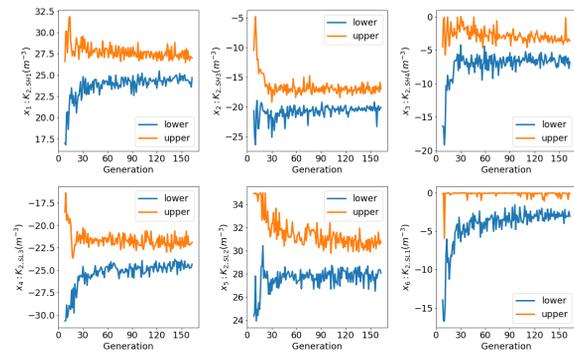


Figure 5: Variation of the ranges for generating new elite population in six-dimensional search space (sextupole's  $K_2$ ) along the evolution. The elite ranges fluctuate, but gradually converge toward much narrower ranges.

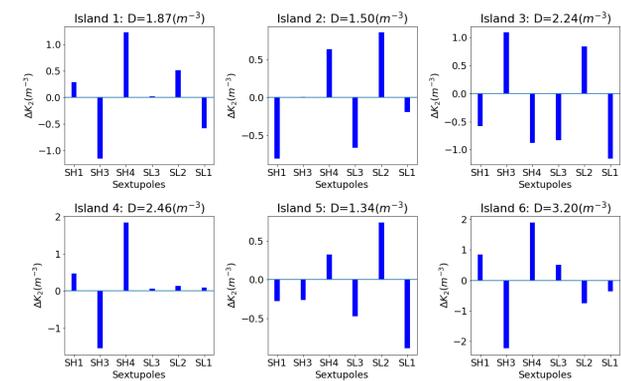


Figure 6: Relative distances of six neighbors from one elite candidate, which is used as the reference here in the search space. The reference sextupole settings are given as  $K_{2,SH1} = 26.20891 \text{ m}^{-3}$ ,  $K_{2,SH3} = -17.87664 \text{ m}^{-3}$ ,  $K_{2,SH4} = -6.39466 \text{ m}^{-3}$ ,  $K_{2,SL3} = -22.42607 \text{ m}^{-3}$ ,  $K_{2,SL2} = 28.54735 \text{ m}^{-3}$ ,  $K_{2,SL1} = -0.22496 \text{ m}^{-3}$ .

in the search space. The island volumes, defined as

$$V = \prod_{n=1}^N (x_n^u - x_n^l), \quad (5)$$

are quite different. Here,  $N$  is the number of dimensions of the search space, and  $x_n^u$  and  $x_n^l$  are their upper and lower boundaries in the  $n^{\text{th}}$  dimension. In general, optimal candidates in large islands are more robust and therefore less impacted by errors than candidates in small islands because average fitness in large islands is less sensitive to the variation in search parameters. In Fig. 6, one island's coordinates are chosen as the origin to illustrate the relative distance to the six neighboring islands. All candidates in these islands yield decent dynamic apertures, but the sextupole settings are quite different.

The following paragraphs describe the detailed tracking results with the simulation code "ELEGANT" [36] and the experimental observations. From many optimal candidates obtained thus far, one solution is chosen and used as the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

origin in Fig. 6, to carry out machine studies. Fitness is determined in regard to the spreads of linear actions through numerical simulations. Tracking 5 particles with different initial conditions is carried out. Their normalized conjugate position-momentum coordinates are shown in Fig. 7. Here, 5 initial conditions are used that differ from those used in the optimization setting (see Fig. 3). The maximum rms spread with the initial condition  $x = 20$  mm and  $y = 3$  mm (the outer ring in the plot) is around 3%, indicating that the motion is quite regular.

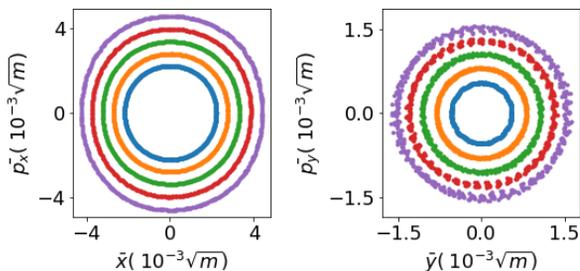


Figure 7: Simulated multi-turn trajectories in the phase space (left: the horizontal plane, right: the vertical plane) for five different initial conditions. The maximum spread for the initial condition  $x = 20$  mm,  $p_x = 0$  mrad and  $y = 3$  mm,  $p_y = 0$  mrad is around 3%.

A frequency map analysis has been carried out for both the on-momentum dynamic aperture (Fig. 8 and Fig. 9), and the off-momentum acceptance (Fig. 10). The dynamic aperture has small nonlinear diffusion [37]: up to 35 mm in the horizontal plane, and 13 mm in the vertical plane as shown in Fig. 8. In the meantime, this particular candidate has large tune-shift-with-amplitude coefficients, which can trap many resonance lines into a very thin stop-band width [38,39] (see Fig. 9). The robustness of this candidate has been confirmed by including the realistic NSLS-II magnet errors.

In this example, after the on-momentum dynamic aperture is optimized, the energy acceptance appears to be sufficient in the view of beam lifetime (see Fig. 10). The same observation holds for other optimal candidates. For the NSLS-II storage ring, it would appear that the two objectives, dynamic aperture and energy acceptance, may not conflict with each other. Should the dynamic aperture and energy acceptance conflict as optimization objectives in other synchrotrons, it is possible to include some off-momentum particle's actions as the optimization objectives.

After testing several evolved candidates on the NSLS-II storage ring, located on different islands within the search space, all yield sufficient dynamic aperture and energy acceptance, and therefore sufficient beam lifetime, for nominal operating conditions. A brief discussion of one particular experimental study period follows. During this time, beam was brought to third order tune resonance  $3\nu_x = 100$  with the same lattice used for the tracking simulation with interesting results.

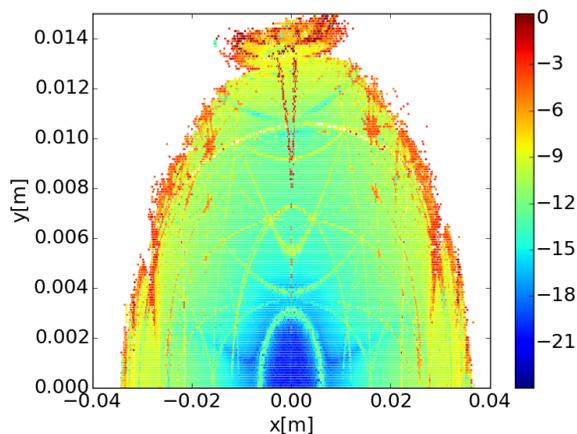


Figure 8: Dynamic aperture for on-momentum particles. The color represents the tune diffusion obtained by turn-by-turn tracking simulation. Diffusion [37] is defined as the difference of tunes  $\Delta\nu$  extracted from the different time durations  $\text{Diff} = \log_{10} \sqrt{\Delta\nu_x^2 + \Delta\nu_y^2}$ . A cool color means the motion is less chaotic and vice versa.

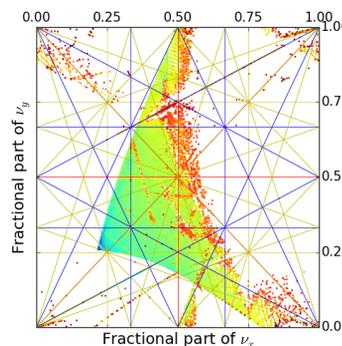


Figure 9: Frequency map corresponding to the on-momentum dynamic aperture in the  $x - y$  planes. A large tune-shift-with-amplitude is observed in this lattice. The third order resonance line can be crossed stably (without obvious diffusion).

The simulated frequency map of the on-momentum dynamic aperture in Fig. 9 indicates that the third-order resonance  $3\nu_x = 100$  was safely covered within the dynamic aperture, with no obvious diffusion (nonlinearity) observed in the tune space. The turn-by-turn particle tracking simulation further shows that the third order resonance has a very narrow stop bandwidth, which can “trap” particles once their trajectories are located inside the islands in the phase space (Fig. 11). During the study period, the machine's horizontal tune  $\nu_x$  was set to 33.332. A short bunch train of 25 buckets was displaced to a particular amplitude using a pulse magnet (pinger). The amplitude of displacement chosen was  $\approx 0.4$  mm, measured at the center of the straight section where  $\beta_x = 21$  m. This particular amplitude allowed the beam horizontal fractional tune to approach as close as

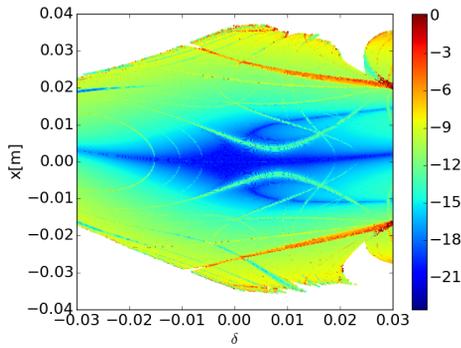


Figure 10: Frequency map of energy (momentum) acceptance in the  $x - \delta$  planes. Here  $\delta = \frac{\Delta p}{p}$  is the particles' relative momentum deviation.

possible to  $1/3 \approx 0.3333 \dots$  (the right subplot in Fig. 12). The beam turn-by-turn trajectories were then observed to be trapped in three isolated islands in the phase space (the left subplot in Fig. 12). The beam began to circulate around the ring on a closed, stable orbit with the periodicity  $1/3$  (see Fig. 13). A similar study was demonstrated and reported on in [40,41].

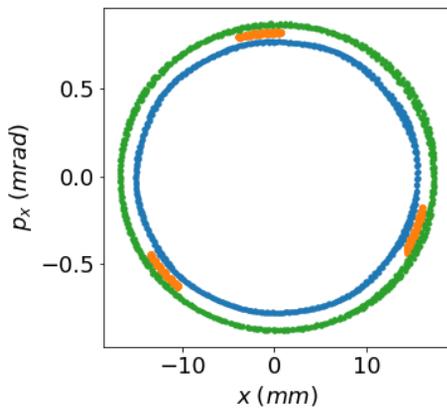


Figure 11: Simulated turn-by-turn trajectories in phase space. A large tune-shift-with-amplitude coefficient shifts the tune away from the third order resonance quickly when the betatron amplitude is slightly off. The stop-band width for this lattice is narrow, which means particle motion is stable even if its tune sits on the resonance.

The closed orbit with a  $1/3$  periodicity repeats itself every 3 turns as illustrated in Fig. 13. It has some potentially interesting applications in dynamics and time-of-flight experiments [42]. For example, using a bunch-by-bunch excitation technique [43,44], selected bunches can be displaced in this closed orbit while keeping the rest of the bunches in the original central orbit. Thus each synchrotron radiation port can deliver up to four distinct x-ray beams. The x-ray beams can have different horizontal positions and angles, and particularly different, distinct time structures. This technique

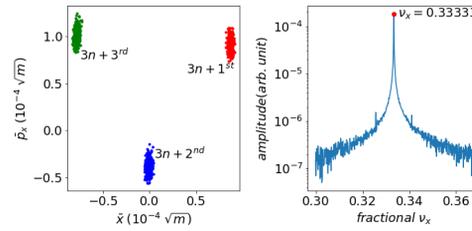


Figure 12: Left: measured beam turn-by-turn trajectories in the phase space with two neighboring beam position monitors (BPM). Three isolated islands are observed with a phase advance of  $\frac{2\pi}{3}$  in-between as expected. Right: the FFT spectrum confirms that the beam remains stable on the third order resonance.

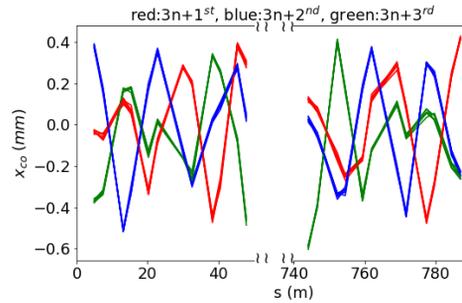


Figure 13: Measured closed orbit at  $\nu_x = 1/3$  with BPM turn-by-turn data. The periodicity of the closed orbit became  $1/3$  rather than 1. In other words, the closed orbit repeated itself every 3 turns.

and its implications, however, are beyond the scope of this paper and require further development and study.

## SUMMARY

The evolution process of the genetic algorithm is significantly sped up when enhanced by machine learning and applied to the NSLS-II storage ring's dynamic aperture. Intervention via machine learning not only speeds up evolution, but increases the number of elite candidates in the data pool. The quality of some optimal candidates obtained with this technique have been confirmed experimentally on the NSLS-II ring and by simulation. This technique can be applied to other population-based optimization problems such as particle swarm algorithms. Extending it to an online mode would be a next logical step and would be driven by a real storage ring's TbT data [45,46].

## ACKNOWLEDGEMENTS

We would like to thank our NSLS-II colleagues for supporting this study. One of authors (YL) would like to thank Dr. Michael Borland, Dr. Yipeng Sun, Dr. Xiaobiao Huang, and Prof. Alexander Wu Chao for the fruitful discussion. This work was supported by Department of Energy Contract No. DE-SC0012704.

## REFERENCES

- [1] L. Yang *et al.*, *Phys. Rev. ST Accel. Beams* 14, 054001 (2011)
- [2] L. Yang *et al.*, *NIM-A* 609, 50-57 (2009)
- [3] M. Borland *et al.*, *Conf. Proc.* C110328, 2354-2356 (2011)
- [4] Y. Li *et al.*, *Int. J. Mod. Phys. A* 31, 1644019 (2016)
- [5] C. Sun *et al.*, *Phys. Rev. ST Accel. Beams* 15, 054001 (2012)
- [6] L. Wang *et al.*, *Conf. Proc.* C1205201, 1380-1382 (2012)
- [7] M. Ehrlichman, *Phys. Rev. Accel. Beams* 19, 044001 (2016)
- [8] K. Tian *et al.*, *Phys. Rev. ST Accel. Beams* 17, 020703 (2014)
- [9] W. Gao *et al.*, *Phys. Rev. ST Accel. Beams* 14, 094001 (2011)
- [10] C. Gulliford *et al.*, *Phys. Rev. Accel. Beams* 19, 093402 (2016)
- [11] C. Gulliford *et al.*, *Phys. Rev. Accel. Beams* 20, 033401 (2017)
- [12] A. He *et al.*, *Phys. Rev. ST Accel. Beams* 18, 014201 (2015)
- [13] I. Bazarov *et al.*, *Phys. Rev. ST Accel. Beams* 8, 034202 (2005)
- [14] R. Bartolini *et al.*, *Phys. Rev. ST Accel. Beams* 15, 030701 (2012)
- [15] A. Hofler *et al.*, *Phys. Rev. ST Accel. Beams* 16, 010101 (2013)
- [16] B Terzic *et al.*, *Phys. Rev. ST Accel. Beams* 17, 101003 (2014)
- [17] X. Huang *et al.*, *NIM-A* 757, 48-53 (2014)
- [18] Y. Jiao *et al.*, *Chin. Phys.* C41, 027001 (2017)
- [19] J. Li, *Private communication* (2017)
- [20] A. Dragt, <https://www.physics.umd.edu/dsat/dsatliemethods.html> (2011)
- [21] E. Forest, *Beam Dynamics: A New Attitude and Framework*, Hardwood Academic/CRC Press (1998)
- [22] M. Berz, *Modern map methods in particle beam physics*, Academic Press (1999)
- [23] C. Wang, *ANL/APS/LS* 330 (2012)
- [24] L. Yu, *Phys. Rev. Accel. Beams* 20, 034001 (2017)
- [25] Y. Liet *et al.*, *arXiv:1706.02195* (2017)
- [26] Y. Liet *et al.*, *NAPAC'2016 TUPOB54* (2017)
- [27] Y. Liet *et al.*, *IPAC'2017 WEPIK122* (2017)
- [28] M. Borland and Y. Sun, *Private communication* (2017)
- [29] Y. Sun *et al.*, *NAPAC'2016 WEPOB15* (2017)
- [30] K. Debet *et al.*, *Trans. Evol. Comp* 6, 182-197 (2002)
- [31] H. Yoshida, *Phys. Lett.* A150, 262-268 (1990)
- [32] D. Goldberg *et al.*, *Machine learning* 3, 95-99 (1988)
- [33] F. Pedregosa *et al.*, *J. Machine Learning Res.* 12, 2825-2830 (2011)
- [34] A. Yaser *et al.*, *Learning from data*, AMLBook (2012)
- [35] BNL, <https://www.bnl.gov/ns1s2/project/PDR/> (2013)
- [36] M. Borland, *ANL/APS/LS* 287 (2000)
- [37] D. Robin *et al.*, *Phys. Rev. Lett.* 85, 558-561 (2000)
- [38] A. Chao, *Private communication* (2017)
- [39] A. Chao *et al.*, *NIM* 121, 129-138 (1974)
- [40] P. Goslawski *et al.*, *IPAC'2017, WEPIK057* (2017)
- [41] A. Huschauer *et al.*, *Phys. Rev. Accel. Beams* 20, 061001, (2017)
- [42] C. Sun *et al.*, *Phys. Rev. Lett.* 109, 264801 (2012)
- [43] W. Cheng *et al.*, *J. Phys. Conf. Ser.* 874, 012082 (2017)
- [44] Y. Li *et al.*, *Phys. Rev. Accel. Beams* 20, 112802 (2017)
- [45] X. Huang *et al.*, *Phys. Rev. ST Accel. Beams* 18, 084001 (2015)
- [46] X. Huang, *Private communication* (2017)

# OPTIMIZATION OF HEAVY-ION SYNCHROTRONS USING NATURE-INSPIRED ALGORITHMS AND MACHINE LEARNING

S. Appel\*, W. Geithner, S. Reimann, M. Sapinski, R. Singh, D. M. Vilsmeier  
GSI, Darmstadt, Germany

## Abstract

The application of machine learning and nature-inspired optimization methods, like for example genetic algorithms (GA) and particle swarm optimization (PSO) can be found in various scientific/technical areas. In recent years, those approaches are finding application in accelerator physics to a greater extent. In this report, nature-inspired optimization as well as the machine learning will be shortly introduced and their application to the accelerator facility at GSI/FAIR will be presented. For the heavy-ion synchrotron SIS18 at GSI, the multi-objective GA/PSO optimization resulted in a significant improvement of multi-turn injection performance and subsequent transmission for intense beams. An automated injection optimization with genetic algorithms at the CRYRING@ESR ion storage ring has been performed. The usage of machine learning for a beam diagnostic application, where reconstruction of space-charge distorted beam profiles from ionization profile monitors is performed, will also be shown. First results and the experience gained will be presented.

## INTRODUCTION

FAIR—the Facility for Antiproton and Ion Research will provide antiproton and ion beams of unprecedented intensities as well as qualities to drive forefront heavy ion and antimatter research [1]. The multi-turn injection (MTI) into heavy-ion synchrotron SIS18 is one of the bottlenecks for

providing unprecedented intensities. The loss-induced vacuum degradation and associated life-time reduction for intermediate charge state ions is one of the key intensity limiting factors for SIS18 [2]. Beam loss during injection can trigger the pressure bump instability. An optimized injection can relax the dynamic vacuum problem, but is also crucial to reach the synchrotron intensity limit by a large multiplication of the injected current [3].

The complexity of the FAIR facility demands a high level of automation to keep anticipated manpower requirements within acceptable levels, as shown in [4]. An example of complexity is the High Energy Beam Transport System of FAIR which forms a complex system connecting among other things seven storage rings and experiment caves and has a total length of 2350 metres [5]. An automatized machine based optimization would improve the time for optimization and control of HEBT.

In the frame of the Swedish in-kind contribution to the FAIR project the storage ring CRYRING@ESR is planned to be used for experiments with low-energy ions and antiprotons. The ring is already installed in the existing GSI target hall and commissioning has started in 2015 [6–8]. Since CRYRING@ESR has its own local injector it can be used stand-alone for testing novel technical developments like automatized configuration of beam line devices. A semi-automatized optimization has been already preformed at the CRYRING in Sweden [9]. Figure 1 shows the CRYRING@ESR and its local injector. Over the second transfer line the CRYRING@ESR can also receive beams from the experimental storage ring ESR.

\* s.appel@gsi.de

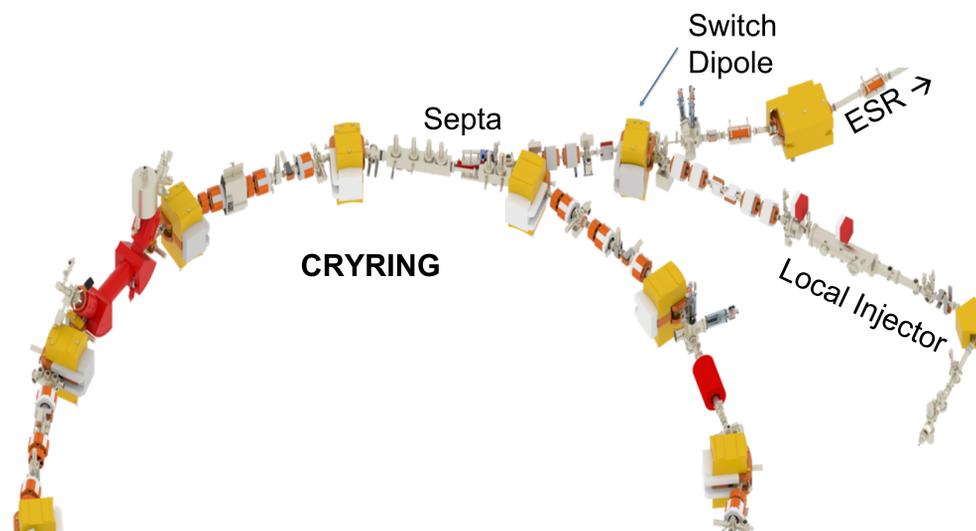


Figure 1: CRYRING@ESR injection from the local injector has been online optimized with an evolutionary algorithm.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

For the optimization and control of synchrotrons the knowledge of beam parameters is a key ingredient. Ionization profile monitors play an important role in non-destructive measurements of the transverse beam profile. They make use of residual gas ionization by the particle beam and collect the ionization products via appropriate guiding fields. However, for the foreseen intensities at heavy-ion synchrotron SIS100 for some beams a profile distortion is expected to be visible. Here the application of machine learning allows the reconstructing of the beam profiles with simulation supported training.

## NATURE-INSPIRED OPTIMIZATION

Nature-inspired optimization algorithms often perform well approximating solutions to all types of problems because they ideally do not make any assumption about the underlying fitness landscape. The *fitness* determines the quality of the solution and determines the probability of its survival for the next optimization step. The fitness is evaluated by an objective function, a simulation code or a real running system. In many real-life problems, multi-quantities have to be optimized. In addition, these quantities can be contradicting and there is more than one equally valid solution. These solutions form a so-called Pareto front (PA front) in the solution space. A solution is Pareto optimal if it is not dominated by any other solution. By using a non-dominated selection algorithm one tries to find solutions near the optimal Pareto set.

### Evolutionary algorithms

An evolutionary algorithm (EA) is inspired by biological evolution, such as *reproduction*, *mutation*, *recombination*, and *selection*. Genetic algorithms (GA) is the most popular type of EA. In GA terminology, a solution vector is called an *individual* and represents a set of variables; one variable is a *gene*. A group of individuals form a *population*, the following child populations are counted in *generations*. The first population is created randomly. The *crossover* operator exchanges variables between two individuals - the parents - to discover with their offspring promising areas in the solution space (*exploration*). For the optimization within a promising area, the mutation operator changes randomly the characteristics of individuals on the gene level (*exploitation*). Reproduction of individuals for the next generation involves selection. During optimization the most promising individuals are chosen to create the next generation. By allowing individuals with poor fitness to take part in the creation process the population is prevented to be dominated by a single individual. The most popular techniques for a single-objective optimization are proportional selection, ranking and tournament selection [10, 11].

### Particle swarm optimization

The initial inspiration for the Particle Swarm optimization (PSO) came from the “graceful but unpredictable choreography of a bird flock” and is a example of alternative algo-

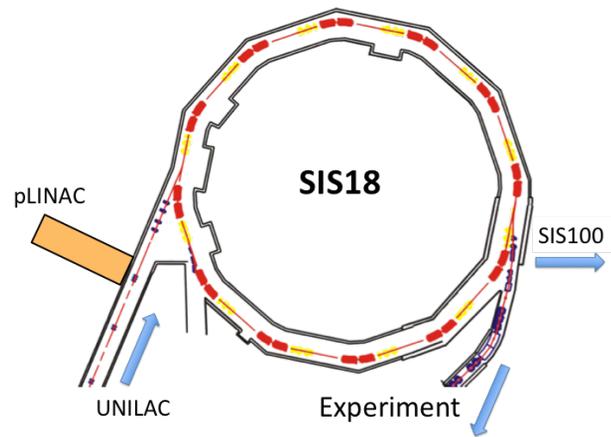


Figure 2: The heavy-ion synchrotron SIS18 and its injectors.

rihms. The key to the swarm success lies in *social influence* and *learning*. Each individual’s behavior is influenced by its own *personal experience* and the *social standard* [11]. Within a swarm, each individual refers to a point in the variable space. It is updated by adding a velocity depending on the personal experience and the socially swarm influenced. The “nostalgia” in the individual tends to return to a place it encountered in the past that best fulfilled the objectives reflected by the personal best *pbest*. Simultaneous, the individuals seek to attain publicized knowledge or social norms, reflected by the best position ever for the entire swarm *gbest*. The movements of the swarm a guide by improved positions, which are updated during the optimization. Including in addition stochastic elements in the algorithm allows to search widely and hopefully finding a satisfactory solution. PSO has shown faster convergence than GA optimization [11].

## INJECTION OPTIMIZATION

SIS18 (Figure 2) will serve as a booster for SIS100 in the FAIR facility to provide ion beams of unprecedented intensities and qualities. An optimized interface between injector linacs and synchrotron is mandatory to achieve this goal. The new FAIR proton linac (pLINAC) will provide the high intensity primary proton beam for the production of antiprotons. The existing GSI heavy ion linac (UNILAC) is able to deliver world record uranium beam intensities for injection into the SIS18, but it is not suitable for FAIR operation. Therefore an upgrade program is planned to replace the post-stripper section. An evolutionary algorithm based optimization of the multi-turn injection (MTI) of the SIS18 has been performed to define the interface parameters for UNILAC and pLINAC. The goal of the optimization is to stack the beamlets injected from the injector in the horizontal phase space until the synchrotron intensity limit is reached. Thereby injection losses on the septum or acceptance have to be minimized to prevent a synchrotron performance reduction to due loss induced vacuum degradation [3]. However, the required MTI brilliance should be in a reachable value frame for the injector linac. As MTI has to fulfill Liouville’s theorem, four bumper magnets create a time variable closed orbit bump such that

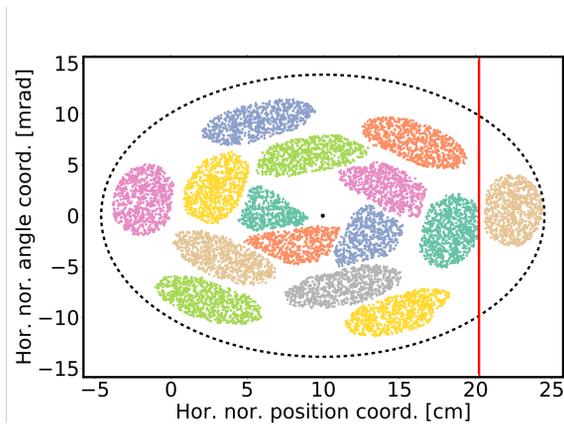


Figure 3: Snapshot of a MTI simulation with loss. The red line indicates the septum and dashed line the acceptance.

the injection septum deflects the next incoming beamlet into available horizontal phase space close to the formerly injected beamlets. For effective adaptation to the free phase space, for instance, an exponential bump reduction can be chosen. During the nature-inspired optimization the parameters on which the MTI depends are altered in consideration of the limiting technical and physical conditions to find an excellent MTI performance. The MTI performance depends on injector emittance and current, position and angular of the incoming beam, the closed orbit at the septum, horizontal tune, miss-match of the incoming beam and the orbit bump reduction. For the optimization the Distributed Evolutionary Algorithms in Python (DEAP) [12] together with pyORBIT has been used. The SIS18 MTI model has been implemented in the particle tracking code pyORBIT—the Python implementation of ORBIT (Objective Ring Beam Injection and Tracking) code—and was carefully validated against experiments [13–15]. Figure 3 shows a snapshot of a MTI simulation with loss in normalized coordinates. The loss areas—inner and outside of the septum as well as the acceptance—are visible. The inner beamlets lost particles at septum earlier during the injection process and therefore not overlap. The injected beams are spirally arranged. The first injected beams are sitting in the center of the spiral next due to the closed orbit indicated by the black dotted. Figure 4 illustrates the evolution of the injection loss obtained from the GA for different numbers of injected turns. The GA finds a better set of parameters than the previous simulation studies (indicated by the dashed lines [14]). The fact that a longer injection time leads to higher losses also holds for the GA optimization if the available acceptance is filled. However, especially in these cases GA discovers a much better solution. The dependence of the gain factor on the injection loss is of particular interest due to the vacuum degradation problem. In order to define the relationship between both, the gain factor has been included as an optimization objective, i.e. to find a 2D Pareto front of both. Figure 5 shows that multi-objectives genetic algorithms (MOGA) finds a much better set of parameters for an improved MTI performance

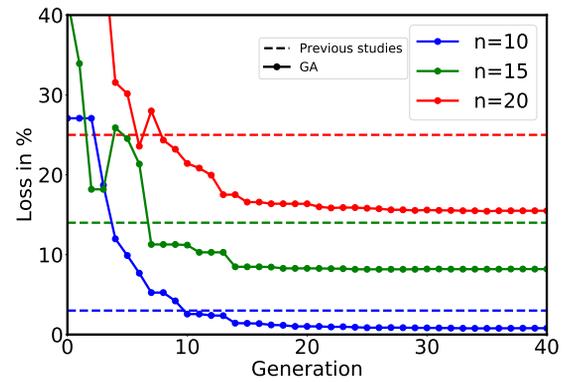


Figure 4: The evolution of loss for injected emittance of 7 mm-mrad. GA found a much better injection parameter setting for a low loss injection than the previous simulation studies (dashed lines).

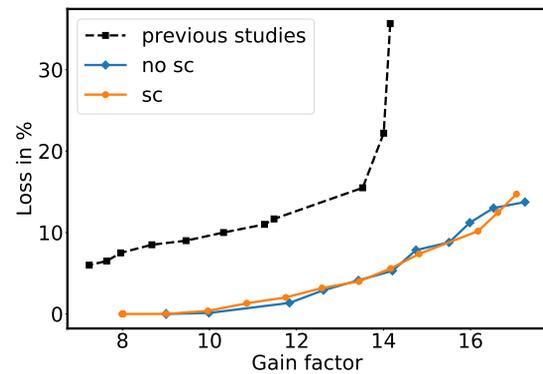


Figure 5: The PA front for gain factor and MTI loss. GA found a much better PA front than the previous studies.

than the previous simulation studies [14]. The influence of space charge on the MTI performance optimization with MOGA is significant even if the discovered PA fronts are similar: The discovered MTI parameters are different with space charge. For the layout of the injector upgrade and the new proton injector is crucial to know the injection dependence on emittance. The demands on the injector could be relaxed if a sufficient MTI performance with a large injection emittance can be discovered. Previous MTI optimization studies [14, 16] clearly demonstrate that the horizontal emittance of the incoming beam has a significant impact on MTI performance. The smaller the injected emittance is, the better the MTI performance gets, which is contradicting to relaxation of the injector demands. A reduction of the horizontal emittance can be achieved e.g. by horizontal collimation [16] or by a round-to-flat transformation [15]. Figure 6 shows in accordance with MTI model and previous studies the trade-off between the objectives over a wide range of parameter variations, which can be summarized as follows: no loss means small injected emittance and low gain factor; a high gain factor implies small emittance with

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

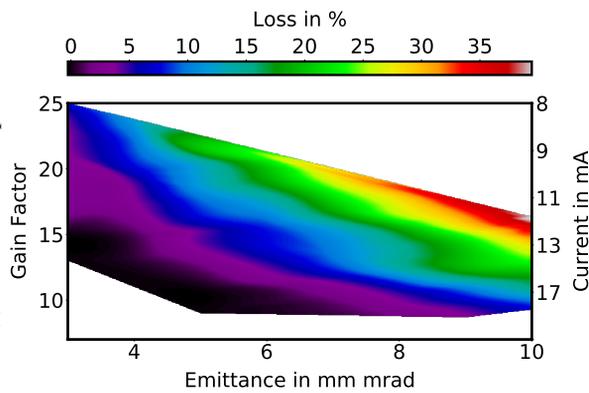


Figure 6: The 3D Pareto front for a simultaneously GA optimization of gain factor, loss and emittance. On the right axes the require injector current is shown to reach the space charge limit in the SIS18.

medium loss; and large emittance means very large loss and small gain factors. This trade-off is a direct consequence of Liouville’s theorem. The obtained results for single and double objective optimization are located also on the 3D PA front. Optimization with multiobjective particle swarm (MOPSA) shown similar result with fast convergence. A 3D Pareto front for proton injector has generated shown in [17]. Outcome of this optimization and heuristic analyze of the MTI demonstrate, that a low-loss injection to fill SIS18 up to the space charge limit for several emittance over many turns for various proton currents could be achieved.

## ONLINE INJECTION OPTIMIZATION

The multi-turn injection of the CRYRING@ESR from the local injector has been online optimized with a genetic algorithm. The aim of the automatized optimization was to maximize the beam current stored in the CRYRING@ESR. The beam current has been measured with the Schottky diagnose in the CRYRING. An end-user application exploiting the genetic algorithm framework Jenetics [18] to optimize unknown beamline settings through the Java based FAIR control system has been implemented [19]. Jenetics is an end-user ready software library implementing an genetic, evolutionary algorithm, written in modern day Java. Therefore the choice to use Jenetics was obvious although faster algorithm are known. The Jenetics algorithm allows independent variation of the merging dipole magnet and the quadrupoles strengths in the transfer line as well the septa, steerer strengths, and the closed orbit defined by the ring dipoles. The result of the successful evolutionary algorithm optimization performance is presented in Figures 7. Shown are two cases of converged genetic scans for the recombination probability of 0.5 and 0.8. The population size was 50 and the offspring fraction 0.5. The tournament size of 15 has been chosen rather large to reach a fast convergence. For large tournament size, weak individuals have less chance of being selected. The first population is created

randomly forming a range around 10–15% of known good values (e.g. from earlier manual settings or beam optics calculations). The performance of the ion source, especially unstable plasma conditions play a crucial role, as it introduces non-deterministic transmission fluctuations which cannot be coped with by the algorithm without further measures. Therefore for each genetic scan step an averaging over ten measurements has been performed. Both scans reached after about 1.5 hours optimization time previous achieved transmission. At present, the time-domain performance is limited by the FAIR control system. Hence, removing performance bottlenecks in the FAIR control system code stack would be a key to fully enable this method’s power.

## MACHINE LEARNING

A principal characteristic of Machine Learning (ML) is to implicitly deduce a set of rules from given data, mapping specific input to output, relieving the user from this tedious task. As such ML is especially suited for problems whose solutions require either a lot of manual fine-tuning or involve long lists of (potentially unknown) rules. Relevant for the presented problem is the later case, where supervised machine learning consisting of regression models is used to predict continuous variables from the given data. Supervised ML covers many different algorithms with varying complexity, from linear approximations like Linear Regression (LR) up to “biologically inspired” Artificial Neural Networks (ANN) [20].

### Linear Regression

Linear regression is a linear approach modelling the relationship between the scalar dependent variable and one or more explanatory variables. In linear regression, the relationship is modeled using linear predictor functions whose unknown model parameters are estimated from the data. The least squares approach is often used for fitting linear regression models.

### Artificial Neural Networks

Artificial neural networks (ANN) are computing systems vaguely inspired by the biological neural networks found in animal brains. The most basic form of ANN typically utilized in supervised learning problems is a fully-connected feed-forward Multi-Layer Perceptron (MLP). It is a specific ANN architecture which is represented by consecutive layers of nodes where all nodes of two consecutive layers are connected to each other. Each node sums all its weighted inputs and transforms the result using an activation function. The activation function should be non-linear in order to represent non-linearities in the data and it must be differentiable in order to comply with the fitting procedure. Weights are usually randomly initialized and then iteratively updated during the fitting procedure in order to minimize the selected loss function.

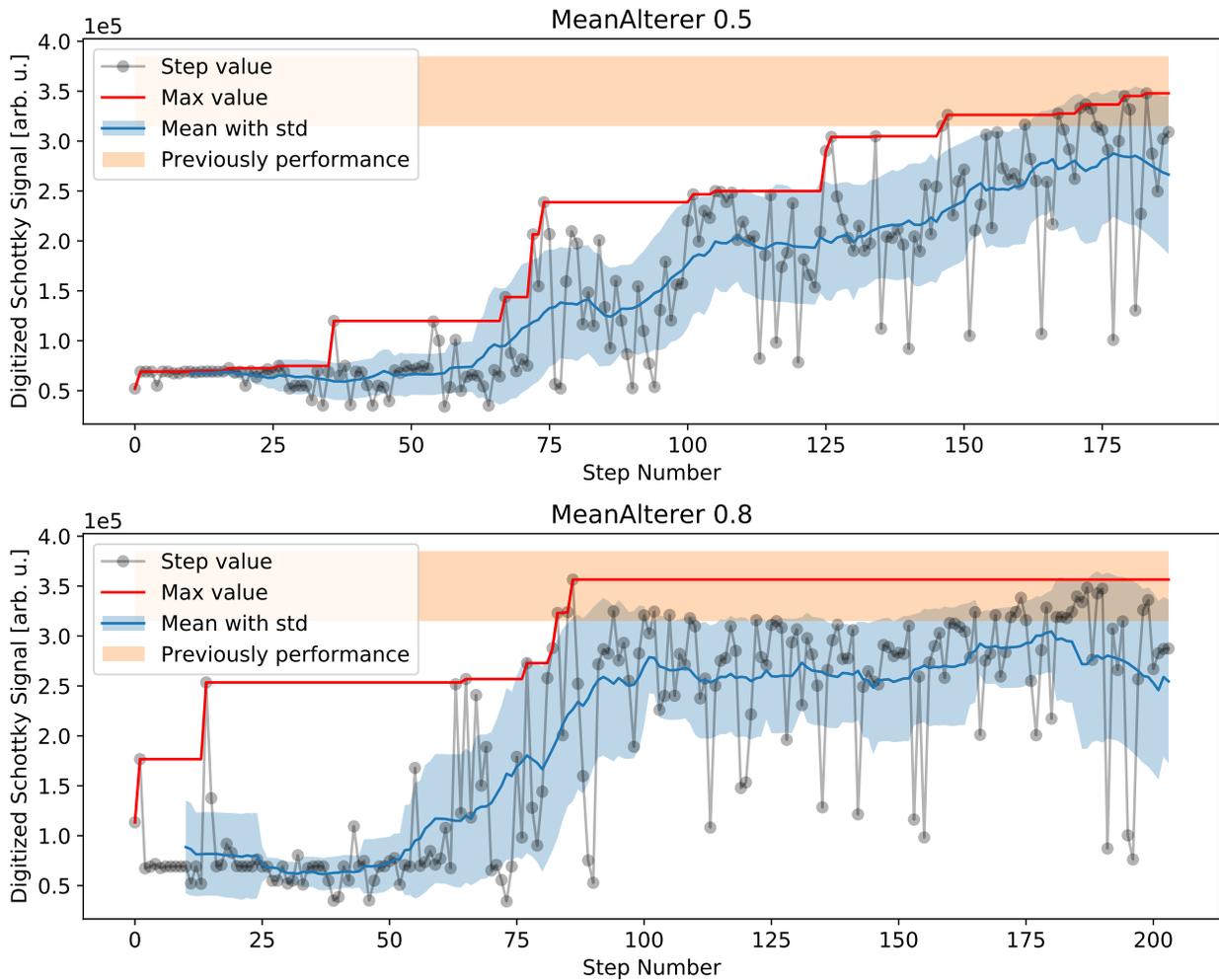


Figure 7: Converged genetic scan driving ten parameters for two different recombination probability. The goal of the optimization has been to maximize the CRYRING@ESR MTI performance. The scans reached the final value after four generations and reached previous good transmission after 89 (upper scan) and 97 minutes (below scan). For each optimization steps an averaging over ten measurements has been performed.

## IPM PROFILE RECONSTRUCTION

The principle of IPMs is the following; the primary beam ionizes the residual gas and the ionized particles (ions or electrons) are extracted via electric fields, sometimes in conjunction with magnetic fields to confine the movement of ionized particles in the plane transverse to the electric field [21]. In the ideal case the ionized particles would move on a straight path towards the detector and the profile of the extracted particles reflects the transverse profile of the primary beam. The electromagnetic fields of the primary beam can affect the trajectory of particle movement towards the detector, see Figure 8. As a consequence the beam profile can be significantly deformed compared to the unimpaird wire scanner measurements. Several attempts have been made to correct or describe such effects parametrically, but no satisfactory analytic procedure was found. At that point a machine learning based approach reliant on good simula-

tion model of the IPM along with space charge effects was performed.

The Virtual-IPM simulation tool was used for simulating the movement of electrons inside the IPM region for a typical LHC case [22], where the beam electric field leads to major distortion. The simulated profiles were binned corresponding to the resolution of an acquisition system based on hybrid-pixel detector [23]. Together with the bunch length and the bunch intensity this data were used for fitting various ML models. Even the simple linear regression model showed very promising results for the beam width reconstruction [22]. The complex artificial neural networks can reconstructed the whole beam profiles as shown in Figure 9 [24].

## CONCLUSION AND OUTLOOK

A fast beam dynamics simulation model has been developed and used together with a multi-objective genetic

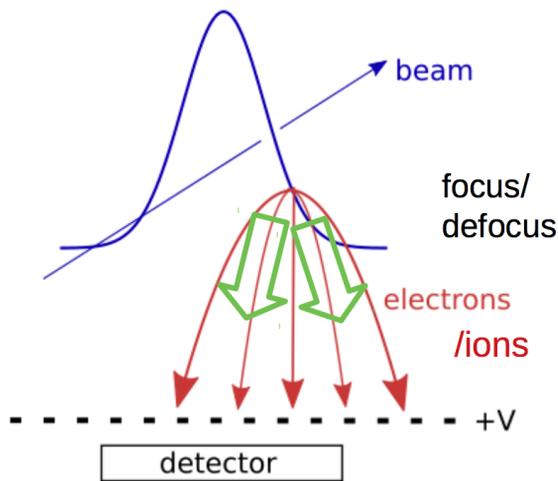


Figure 8: In the ideal case the ionized particles would move on straight lines towards the detector. However the electromagnetic field of the beam can influence the particles movement as shown.

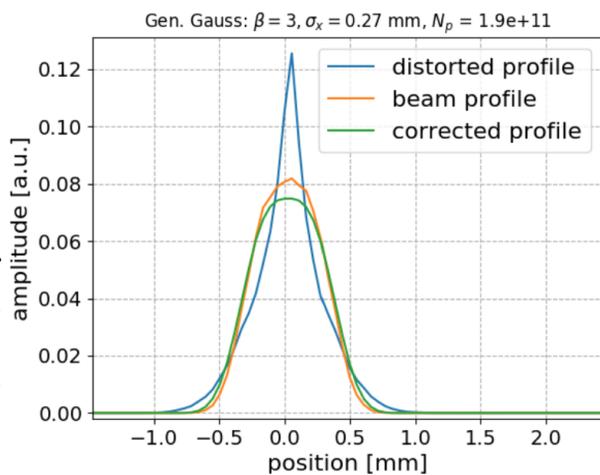


Figure 9: Simulation of profile distortion due to space charge using Virtual IPM together with the ANN corrected profile.

algorithm to optimize the multi-turn injection into SIS18. A loss-free or low-loss injection for several emittance over many turns were identified. Space charge results in a similar PA front, but with different injection settings. With the optimized multi-turn injection a range of injector brilliance could be defined. This crucial information gives more flexibility for the layout of the SIS18 injectors.

An online optimization of multi-turn injection into the storage ring CRYRING@ESR has been presented. After 1.5 hours of optimization time previous transmission could be reached. The nature-inspired optimization has potential to reduce the manpower requirements and variations of quality performance due to the manual procedure. Looking forward, the algorithm shall be applied to SIS18.

A novel method for resolving IPM profile distortion under the influence of magnetic guiding fields based on machine

learning has been presented. The first investigations, using simulated data, yield promising results. Next steps include estimation of influence of error sources on predictions, optimization of model selection and application of the method to measured data. The method has a potential to extend usability and reduce cost of IPMs for high brightness beams. The application of machine learning to time-domain signals like the longitudinal Schottky signals is under investigation.

## REFERENCES

- [1] P. Spiller *et al.*, *FAIR Technical Design Report*, 2008.
- [2] E. Mustafin *et al.*, “A theory of the beam loss-induced vacuum instability applied to the heavy-ion synchrotron SIS18”, *Nucl. Instr. and Meth. A*, vol. 510, pp. 199–205, 2003.
- [3] S. Appel *et al.*, “Injection optimization in a heavy-ion synchrotron using genetic algorithms”, *Nucl. Instrum. Methods A*, vol. 852, pp. 73–79, 2017.
- [4] S. Reimann *et al.*, “Building an Operation Team for FAIR Nearly from Scratch”, presentation at WAO10, Shanghai, China, Sep 2016.
- [5] S. Ratschow *et al.*, “The High Energy Beam Transport System for FAIR”, in *Proc. EPAC’08*, Genoa, Italy, paper THPP104, pp. 3608–3610, Jun 2008.
- [6] F. Herfurth *et al.*, “Commissioning of the low energy storage ring facility CRYRING@ESR”, in *Proc. COOL’17*, Bonn, Germany, Sep 2017. doi:10.18429/JACoW-COOL2017-THM13
- [7] W. Geithner *et al.*, “Status and outlook of the CRYRING@ESR project”, *Hyperfine Interact.*, vol. 238, 2017. doi:10.1007/s10751-016-1383-5
- [8] O. Gorda *et al.*, “Ion-optical design of CRYRING@ESR”, *Phys. Scr. T*, vol. 166, pp. 14043, 2015.
- [9] A. Källberg and A. Simonsson, “Beam Steering with Image Processing in the CRYRING Injection Beamline”, in *Proc. DIPAC’99*, Cheshster, UK, paper PS09, pp. 118–119, May 1999.
- [10] A. Konak, D. W. Coit and A. E. Smith, “Multi-objective optimization using genetic algorithms: A tutorial”, *Reliab. Eng. Syst. Saf.*, vol. 91, no. 9, pp. 992–1007, 2006.
- [11] X. Pang and L. Rybarczyk, “Multi-objective particle swarm and genetic algorithm for the optimization of the LANSCE linac operation”, *Nucl. Instrum. Methods A*, vol. 741, pp. 124–129, 2013.
- [12] F. Fortin and D. Rainville, “DEAP: Evolutionary algorithms made easy”, *J. Mach. Learn. Res.*, vol. 13, pp. 2171–2175, 2012. <https://github.com/DEAP/deap>
- [13] A. Shishlo *et al.*, “The Particle Accelerator Simulation Code PyORBIT”, *Procedia Comput. Sci.*, vol. 51, pp. 1272–1281, 2015. doi:10.1016/j.procs.2015.05.312, <https://github.com/PyORBIT-Collaboration>
- [14] S. Appel and O. Boine-Frankenheim, “Multi-turn injection into a heavy-ion synchrotron in the presence of space charge”, 2014. arXiv:1403.5972v1
- [15] S. Appel *et al.*, “Injection optimization through generation of flat ion beams”, *Nucl. Instr. and Meth. A*, vol. 866, pp. 36–39, 2017.

- [16] Y. El-Hayek *et al.*, “Initial beam loss and control of dynamic vacuum effects in SIS18”, in *Proc. IPAC’13*, Shanghai, China, May 2013, paper MOPFI010, pp. 300–302.
- [17] C. Kleffner *et al.*, “Status of the FAIR Proton Linac”, in *Proc. LINAC’18*, Beijing, China, Sep 2018. doi:10.18429/JACoW-LINAC2018-THP0046
- [18] F. Wilhelmstötter, “Jenetics advanced Genetic Algorithm”, <http://jenetics.io>
- [19] W. Geithner *et al.*, “Genetic Algorithms for Machine Optimization in the FAIR Control System Environment”, in *Proc. IPAC’18*, Vancouver, Canada, May 2018. doi:10.18429/JACoW-IPAC2018-THPML028
- [20] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O’Reilly, 2017.
- [21] R. Singh *et al.*, “Simulation Supported Profile Reconstruction With Machine Learning”, in *Proc. IBIC’17*, Grand Rapids, MI, USA, Aug. 2017. doi:10.18429/JACoW-IBIC2017-WEPCC06
- [22] D. Vilsmeier *et al.*, “Reconstructing Space-Charge Distorted IPM Profiles With Machine Learning Algorithms”, in *Proc. IPAC’18*, Vancouver, BC, Canada, Apr-May 2018, pp. 2099–2102. doi:10.18429/JACoW-IPAC2018-WEPAK008
- [23] S. Levasseur *et al.*, “Time-Resolved Transverse Beam Profile Measurements with a Rest Gas Ionisation Profile Monitor Based on Hybrid Pixel Detectors”, in *Proc. IPAC’18*, Vancouver, BC, Canada, Apr-May 2018, pp. 2361–2364. doi:10.18429/JACoW-IPAC2018-WEPAL075
- [24] M. Sapinski *et al.*, “Application of Machine Learning for the IPM-Based Profile Reconstruction”, in *Proc. HB’18*, Daejeon, Korea, June 2018, pp. 410–415. doi:10.18429/JACoW-HB2018-THA2WE02

# COMPARISON OF MODEL-BASED AND HEURISTIC OPTIMIZATION ALGORITHMS APPLIED TO PHOTOINJECTORS USING LIBENSEMBLE

N. Neveu\*, L. Spentzouris, Illinois Institute of Technology, Chicago, USA  
J. Larson, S. Hudson, Argonne National Laboratory, Lemont, USA

## Abstract

Genetic algorithms are commonly used in the accelerator community and often require significant computational resources and manual adjustment of hyperparameters. Model-based methods can be significantly more efficient in their use of computational resources, but are often labeled as unreliable for the nonlinear or nonsmooth problems that can be found in accelerator physics. We investigate the behavior of both approaches using a photoinjector operated in the space-charge-dominated regime. All model-based optimization runs were coordinated and managed by `libEnsemble`, a Python library at Argonne National Laboratory.

## ARGONNE WAKEFIELD ACCELERATOR FACILITY INTRODUCTION

The Argonne Wakefield Accelerator (AWA) facility houses two photoinjector beamlines. Ongoing research efforts at the AWA includes emittance exchange photocathode studies and two-beam acceleration experiments [1], the latter of which motivates this work. Figure 1 shows the layout of the AWA bunker during two-beam accelerator experiments. The high-charge beamline, often referred to as the drive beam, is being modeled in this work.

## CODE AND RESOURCES

The particle-in-cell code OPAL [2] is used to simulate the high charge beam line at the AWA. OPAL is an open-source parallel code with two version, `OPAL-t` and `OPAL-cycl`. The former was used for this work, the latter version is used for modeling cyclotrons. OPAL can also simulate 3D-space charge, 1D coherent synchrotron radiation, and wakefield effects. Note that the optimization methods being compared in this study are applicable to any beam-dynamics code—Parmela [3], ASTRA [4], GPT [5], all of which have been used by AWA group members—not just `OPAL-t`.

All simulations were run on the Bebop cluster maintained by the Laboratory Computing Resource Center [6] at Argonne National Laboratory. Bebop machine consist of 664 Broadwell nodes, and 352 Knights Landing (KNL) nodes. The simulations presented here were performed on KNL nodes, due to short queue times and readily available resources.

\* nneveu@hawk.iit.edu

## OPTIMIZATION METHODS

### *Heuristic Method: Genetic Algorithm*

Genetic Algorithms (GA) are a popular choice for optimizing simulations in the accelerator physics community. They have been used with success on several types of accelerator physics problems that address challenges facing both linear and circular machines [7]. It is not disputed that in these solutions were found with benefit to many facilities. However, it is also well-known that GAs are computationally intensive. They can require hundreds of thousands of core hours depending on the problem being solved. Performing optimization with fewer calls to the simulation would not only save time, but would also enable facilities to accomplish more design work without access to large core hour allocations.

### *Model Based Methods: APOSMM + BOBYQA*

Model-based derivative free methods are increasingly popular in mathematics and other scientific domains, but have not been widely used in the accelerator physics community. This may be due to the assumption that these methods may become within local minima within a bounded search space. In a sense, this is true if the algorithm is always started with the same initial conditions and weights. However, if the algorithm is started multiple times with various initial conditions, this behavior may be mitigated.

In order to achieve a multistart approach, the asynchronously parallel optimization solver for finding multiple minima (APOSMM) [8] was used. This algorithm maintains a history of all previously evaluated points, and uses this information when deciding starting points for local optimization runs. APOSMM also allows concurrent local optimization runs while honoring the amount of resources available. For details on how local optimization points are determined, see [8, Section 3].

In this paper, we use the bounded optimization by quadratic approximation (BOBYQA) [9] local optimization method. After a set of simulation evaluations are finish, the beam parameters (i.e. objectives) at the desired location are fed to BOBYQA. The algorithm then builds and minimizes a quadratic model of the objectives in order to pick the next point to evaluate.

Both APOSMM and BOBYQA implementations used are open source, freely available, and written in Python.

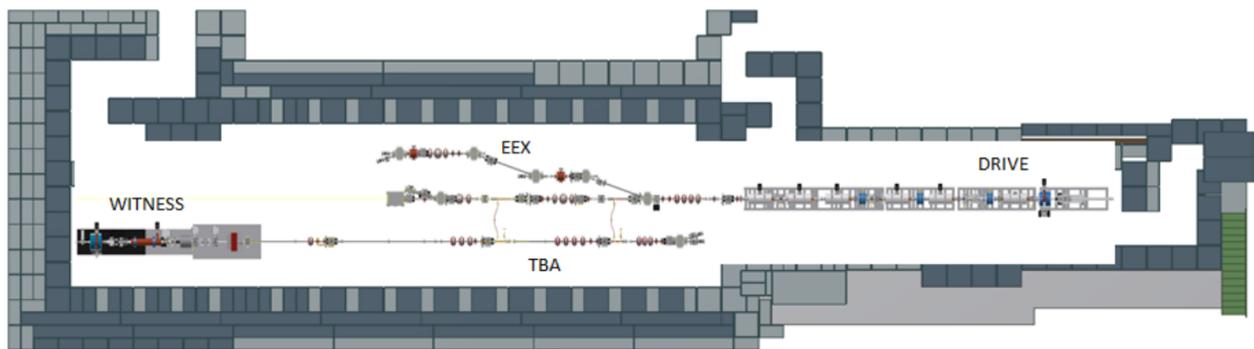


Figure 1: Layout of the AWA bunker. The drive beam with high charge is located on the right side of the image and the witness beam with lower charge and energy is shown on the left side of the image.

## LIBENSEMBLE

`libEnsemble` is a library developed at Argonne for managing ensemble-like collections of computations. This library is capable of coordinating concurrent simulation evaluations and optimization algorithms. In the past, using a local optimization method in a multistart fashion required many serial runs, or much time spent on the users end to decide initial starting points and manage resources. `libEnsemble` in combination with APOSMM manages allocations and resources, so that the user can focus on optimization work and results rather than parallel programming and resource management. Key features of `libEnsemble` include:

- Automatically manages the asynchronous evaluation of calculations and, if desired, the optimization of outputs
- Helpful developers
- Open source Python code, on GitHub.
- Can run on laptops, clusters, and HPC systems

As an example of large resource management, `libEnsemble` has been used to coordinate the evaluation of 1,600 concurrent OPAL evaluations on 200 KNL nodes, each with 64 cores. `libEnsemble` provides many useful features for common accelerator physics use cases:

- Can gracefully kill simulation runs that loose particles *before* the end of simulation (thereby saving significant computational resources in studies with many failures)
- Gracefully kills runs that become unresponsive.
- Saves specified data into a NumPy array for easy access and storage
- Evaluates objectives based on specific beam criteria and  $z$  location.
- Allows for OPAL-t instances using parallel resources

This library has the potential to significantly simplify the use of model-based and multistart algorithms for optimization

Table 1: Parameter Bounds for Linac Optimization

Variable	Range	Unit
Solenoid Strength	$300 \leq S_1 \leq 550$	amps
Solenoid Strength	$180 \leq S_2 \leq 280$	amps
Phase of Gun	$-20 \leq \phi_g \leq 0$	degrees
Cavity Phase	$-20 \leq \phi_L \leq 20$	degrees

problems. In addition, this can save time spent on programming, since parallelization and allocation of resources for concurrent runs is managed by `libEnsemble` instead of user scripts.

## PHOTOINJECTOR OPTIMIZATION

The beamline simulated in OPAL consisted of the gun, two solenoids, and six linac cavities, as shown in Fig. 2. The charge of interest was 40 nC, therefore 3D space charge forces were calculated at all times, while the rf-field maps were 2D. The laser radius was set at 9 mm. Cavity gradients were set to achieve 65 MeV, with a small spread depending on the phase in each cavity. Nine design variables (shown

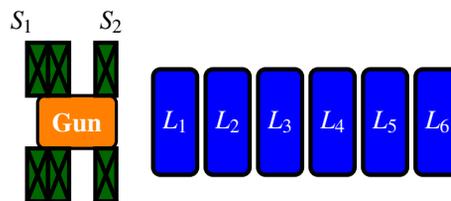


Figure 2: Simulation model used in OPAL and for optimization runs. The charge of interest was 40 nC.

in Table 1) were adjusted during the local optimization runs. Note  $\phi_L$  is a vector containing the six cavity phases. Three objectives were chosen:  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$ . In this case, due to the 2D field maps,  $\sigma_x = \sigma_y$ . The three objectives reduce to two objectives, one representing the transverse beamsize, and one representing the longitudinal beamsize. Therefore the beamsize is optimized at the end of the linac.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

First, `libEnsemble` was used to generate a 1,000 point random sample. These runs were then given by `libEnsemble` to APOSMM. BOBYQA was used as the local optimization method with a multistart approach. The objective was the sum of  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$ . All starting points and subsequent points were chosen by APOSMM. Runs were initiated by `libEnsemble` and each OPAL- $\tau$  simulation was run on four cores. `libEnsemble` maintained objective values and summary information needed by APOSMM after each simulation completed. A limit of 600 evaluations was set for the entire `libEnsemble` runs.

For comparison, we used the NSGA-II [10] implementation within OPAL. Examples of how to use the built-in optimizer can be found on the OPAL wiki<sup>1</sup>. The optimization problem was defined within an OPAL- $\tau$  input file. The results of all simulations are shown in Fig. 3.

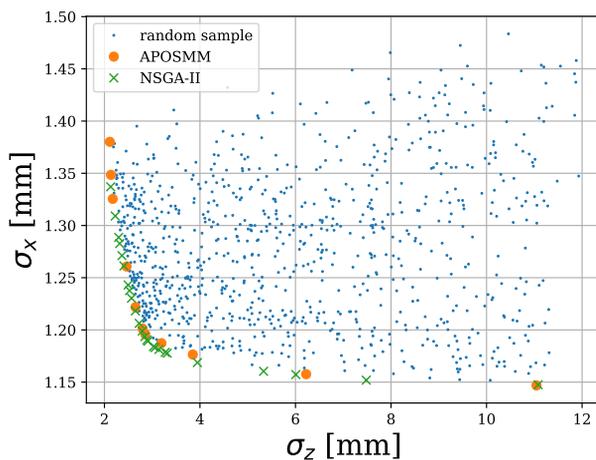


Figure 3: Comparison of GA and APOSMM optimization results. Note only  $\sigma_x$  is plotted, because  $\sigma_x$  and  $\sigma_y$  are equal. This is a result of using 2D symmetric field maps.

## CONCLUSION

There is no significant difference in the Pareto fronts generated by NSGA-II, APOSMM, and the random sample. The GA performed 32,967 simulation evaluations to complete 200 generations. APOSMM completed 600 evaluations with a seed of 1,000 random points. Based on this case alone, it is not clear yet what the time to solution difference is possible for more complicated problems. This was a test case, where the design variables are bounded to regions of acceptable solutions. (The bounds were chosen based experienced gain in prior work.) This point is illustrated by the Pareto front of the 1,000 point random sample. It closely resembles the two optimization methods, which indicates the problem is well bounded. For this optimization problem, running either

<sup>1</sup> <https://gitlab.psi.ch/OPAL/Manual-2.0/wikis/optimiser>

NSGA-II or APOSMM provided only marginal improvement of the solutions. This suggest a random sample may be sufficient in estimating Pareto fronts for some problems. A significant amount of computational resources could be saved in such cases; if a heuristic method is not used.

## FUTURE WORK

Extension of this work is ongoing and will be published. That work includes optimization problems that are not well bounded or defined by a random sample, like the case shown here.

## ACKNOWLEDGMENTS

We gratefully acknowledge the computing resources provided on Bebop, a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357 and grant number DE-SC0015479.

## REFERENCES

- [1] J. Shao *et al.*, “Recent Progress towards Dielectric Short Pulse Two-Beam Acceleration,” in *Proc. IPAC’18*, Vancouver, Canada, Apr-May 2018, doi:10.18429/JACoW-IPAC2018-TUYGBE3
- [2] A. Adelmann *et al.*, “The OPAL (Object Oriented Parallel Accelerator Library) framework,” PSI, Zurich, Switzerland, Rep. PSI-PR-08-02, 2008-2017.
- [3] L. M. Young, “PARMELA,” Los Alamos National Laboratory, Los Alamos, NM, USA, Rep. LA-UR-96-1835 (Revised April 22, 2003).
- [4] ASTRA Manual, April 2014, [http://www.desy.de/~mpyflo/Astra\\_manual/Astra-Manual\\_V3.2.pdf](http://www.desy.de/~mpyflo/Astra_manual/Astra-Manual_V3.2.pdf)
- [5] General Particle Tracker, <http://www.pulsar.nl/gpt>
- [6] Laboratory Computing Resource Center, <https://www.lcrc.anl.gov>
- [7] A. S. Hofler *et al.*, “Innovative applications of genetic algorithms to problems in accelerator physics,” *Phys. Rev. ST Accel. Beams*, vol. 16, p. 010101, 2013.
- [8] J. Larson and S.M. Wild, “Asynchronously parallel optimization solver for finding multiple minima,” *Math. Prog. Comp.*, vol. 10, p. 1-30, Feb. 2018.
- [9] M. Powell, “The BOBYQA algorithm for bound constrained optimization without derivatives,” University of Cambridge, U.K., Rep. NA2009/06, Oct. 2009.
- [10] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,” *IEEE Trans. on Evolutionary Comp.*, vol. 6, no. 2, pp. 182-197, Apr 2002.



Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## Klystron Simulation Codes

Several codes exist for simulating klystrons. AJDISK [6] is a fast 1D code from SLAC National Accelerator Laboratory that can simulate the high efficiency klystron designs being considered in 20 seconds to 3 minutes depending on the desired accuracy. Since AJDISK designs can be iterated quickly it is used as the primary workhorse for global optimization using the genetic algorithm. Tesla [7] is a 2.5D code from the Naval Research Laboratory (NRL) that can model the radial variation in the beam. The code is more accurate than AJDISK but takes approximately 10 to 30 minutes per iteration depending on the required accuracy. Therefore AJDISK was used for the large global optimization that required thousands of iterations and the results were used as the starting point for a more accurate local optimization which only required hundreds of iterations using Tesla. In summary, AJDISK is used to find the approximate location of the global minima using global optimization and Tesla is used to more accurately determine the global minima using a local optimization routine in the vicinity of the global minima found using AJDISK. KlyC [8] is a new 2D code from CERN for simulating klystrons. The code was not used in our optimization efforts but it was used to benchmark the AJDISK and Tesla results as shown in Figure 2. KlyC has a similar execution time to AJDISK. We plan to use KlyC more as we become familiar with it.

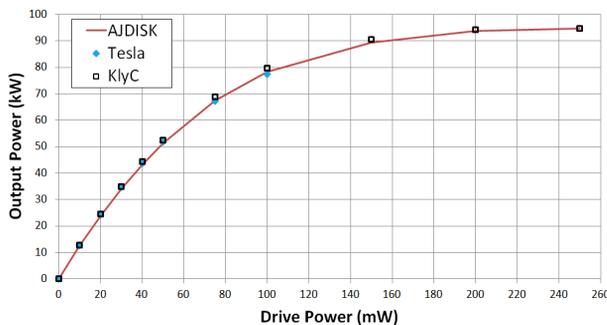


Figure 2: Comparison of different klystron simulation codes.

## OPTIMIZATION ALGORITHMS

### Single Objective Genetic Algorithm

The Single Objective Genetic Algorithm (SOGA) is a powerful global optimization routine available within the Dakota optimization library. It is the primary optimization method used to optimize the BAC and COM klystron designs.

Global optimization routines are ideal for exploring large design spaces that have more than one local minima. Several optimizations were run to optimize the BAC and COM designs. The largest optimization was for the BAC klystron, consisting of 13 variables: 11 cavity frequencies, the location of the output cavity, and the external Q of the output cavity.

The single objective genetic algorithm consists of several steps. The optimization starts with a set number of solutions, known as a “population”, generated by selecting input variables (“genes”) scattered throughout the parameter space. Our optimizations had a population size of 50.

“Parents” within the population are “selected” based on “fitness” which is determined for the klystron optimization based on which “parents” have the highest efficiency. A weighted penalty to fitness is also applied for reflected and slow electrons. Once parents have been selected for “reproduction”, “children” for the next “generation” are produced by copying some of the parents “chromosome” (the string of input variables) up to a “crossover” point. Some genes, or input variables, can also be “mutated” (randomly altered). New generations are then iterated until the population converges on a solution that produces the best answer, or fitness within some convergence criteria.

### Asynchronous Pattern Search

The SOGA method in Dakota typically requires thousands of simulations to converge. The time the optimization takes is therefore directly dependent on the time each simulation takes to execute. To facilitate faster optimization, we run 1D AJDISK which is fast but slightly less accurate than 2D Tesla or KlyC simulations. This approach gives a solution that AJDISK believes is the global minima but is in error by the difference between the 2D solution and 1D AJDISK solution. To remedy this the 1D solution is used as the starting point of a local optimization using Tesla to find the true global minima.

The Asynchronous Pattern Search (APS) in Dakota is used for local optimization. For a two variable optimization a crosshair pattern/template can be imagined where the centre of the crosshair is the starting point solution and the four periphery points of the crosshair are solutions evaluated at some plus/minus offset from the centre. The optimization proceeds by selecting the new best solution. If the best solution is one of the periphery points the crosshair is re-centered at that solution. If the best solution is the centre of the crosshair, the crosshair stencil is reduced in size such that the new periphery solutions will have a smaller offset from the centre solution. These steps are iterated until the input variables change by less than a fixed tolerance defined by the user.

## LARGE SCALE HPC OPTIMIZATION

GSB in combination with Dakota is capable of running very large optimizations. The klystron optimizations in this paper were all run on smaller machines but it is possible to run even larger problems using High Performance Computing (HPC) super computers. As a short diversion we provide a case study of an HPC optimized depressed collector to show the capabilities of GSB. The depressed collector in this study was simulated using MICHELLE [9], a 3D geometry, and more than a dozen variables. The collector was optimized simultaneously against 7 input power levels. On a single CPU it was calculated that the optimization would have taken more than 4 years. On the supercomputer the optimization took less than 24 hours. A scatter plot of the best results is shown in Figure 3. The goal of the optimization was to maximize efficiency while minimizing particles returned to the circuit. The results have been normalized to the best case design that existed before the optimization was run.

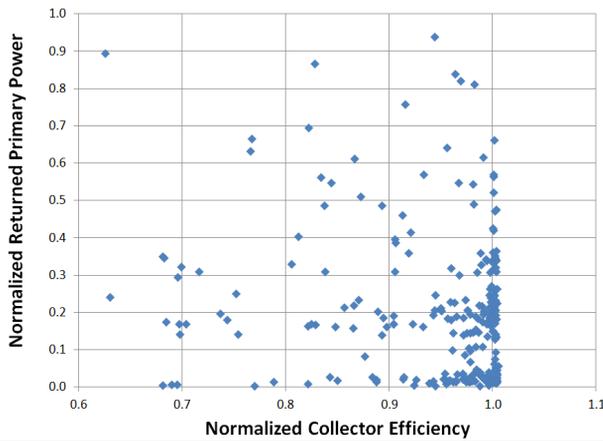


Figure 3: Optimization of a depressed collector.

## HIGH EFFICIENCY KLYSTRON DESIGN

The BAC and COM techniques were used as the basis of our optimizations. Here we discuss an overview of the two techniques.

### Core Oscillation Method

The Core Oscillation Method (COM) is a method of tuning cavity position and klystron cavity frequencies to maximize efficiency. This design approach generates a klystron that is longer than a BAC tuned klystron but uses fewer cavities. The cavities in this method are spaced approximately a half plasma wavelength apart as shown in the Aplegate diagram in Figure 4. Since the space charge forces are stronger in the core (the middle of the bunch), electrons furthest from the core continue to move towards the centre of the bunch even after the core bunch begins de-bunching. This can be seen in Figure 4. By repeating/oscillating the core through several bunching and de-bunching cycles the outermost electrons can eventually join the core bunch. This leads to a long klystron design with high efficiency. The klystron cavity frequencies are tuned high with respect to the operating frequency and can be optimized manually or by using Dakota.

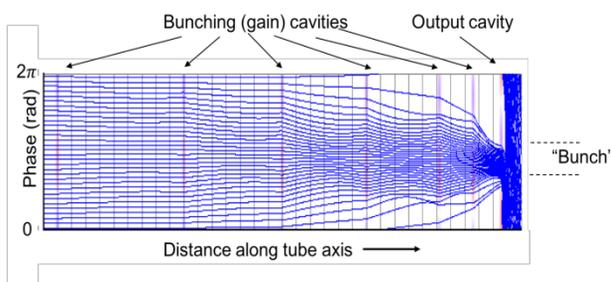


Figure 4: Core Oscillation Method based klystron design.

### Bunch-Align-Compress Method

The Bunch-Align-Compress (BAC) method can be used to generate a klystron of similar efficiency to that achieved using COM in a shorter distance at the expense of using more cavities. The BAC method effectively replaces a single core oscillation in the COM approach with a triplet of cavities as shown in Figure 5. In the simplest case the BAC approach therefore needs three times as many cavities as

the COM technique. The bunching cavity in the BAC technique is analogous to the bunching cavity in the COM approach. The alignment cavity aligns electron velocities in the core such that the bunching cavity can bunch electrons much more quickly without any electrons overtaking each other. The collection cavity operates at the second harmonic to sweep particles in the anti-bunch into the main bunch which is analogous to what the outer electrons do in the COM method but in a much shorter distance. The three cavities in the triplet can be spaced very closely together. The tuning of each cavity in the triplet is a function of several factors including the spacing between the cavities. The complication of optimizing so many cavity frequencies is ideally suited for the SOGA optimization technique.

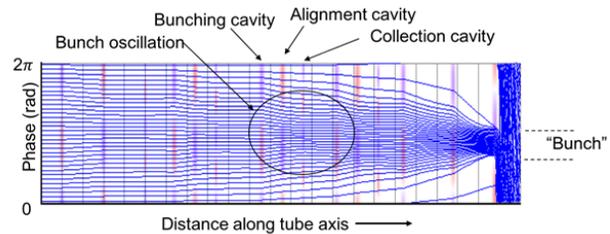


Figure 5: Bunch Align Compress based klystron design.

## KLYSTRON OPTIMIZATION AND RESULTS

Both the COM and BAC klystron designs were optimized to maximize efficiency and minimize reflected particles. The COM design was optimized first using the SOGA method. Each core oscillation was then manually replaced with a BAC triplet of cavities one core oscillation at a time. The BAC design generated in this way was used as the starting point for SOGA optimization of the BAC design.

### COM Optimization

The COM design was conducted first. The input and output cavity frequencies were fixed to the operating frequency. One cavity was added at a time and the cavity position and frequency were decided visually by the engineer such that cavity spacing was on the order of a half plasma wavelength and that a core oscillation could be observed. The number of core oscillations was increased until the efficiency reached approximately 85%. This design was used as the starting point to the SOGA optimization. The cavity to cavity spacing was allowed to vary for each core oscillation by a fixed amount. The cavity frequencies, except for the input were allowed to vary by a fixed amount. Finally the external Q of the output cavity was allowed to vary by a fixed amount.

### BAC Optimization

The BAC design was based on the optimized COM design by replacing each core oscillation in the COM design with a cavity triplet one at a time. The cavity frequencies of the triplet were manually optimized to produce the same or similar results to those achieved in the COM design. This design was the starting point for the BAC based SOGA optimization. The SOGA optimization left the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

cavity positions fixed and optimized the cavity frequencies within some range of their starting values. The external Q and position of the output cavity were also optimized at the same time as the cavity frequencies. The range over which values was allowed to vary was fairly broad with the expectation that the SOGA method would find a true global minima.

### Results

Several SOGA runs were completed. The best result of each run is shown in Figure 6. These results show the expected trend that efficiency increases with circuit length and that the BAC design is shorter than the COM design. Both designs achieved greater than the target efficiency of 85%.

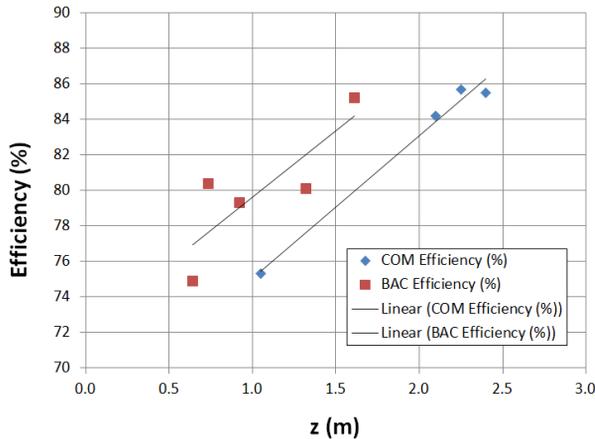


Figure 6: Various SOGA optimizations for both BAC and COM based klystron designs.

The optimized results based on AJDISK were translated for use in Tesla. The first curve in Figure 7 shows the output power reported by Tesla using the optimal AJDISK results. The second curve in Figure 7 shows the results after a local optimization was run using Tesla directly assuming the AJDISK result as a starting point.

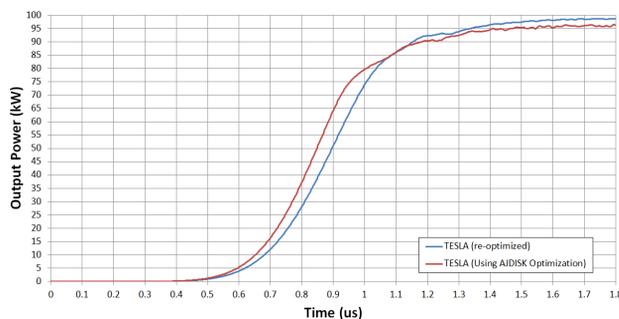


Figure 7: Klystron output power reported by Tesla using AJDISK based SOGA optimization results and Tesla based optimization results.

Next the gun was modelled by Calabazas Creek Research (CCR) as shown in Figure 8. The magnetic field was ramped in the output structure to confine the strong bunching forces in the beam as shown in Figure 9. Using these more realistic conditions for the beam the design was run through final optimization.

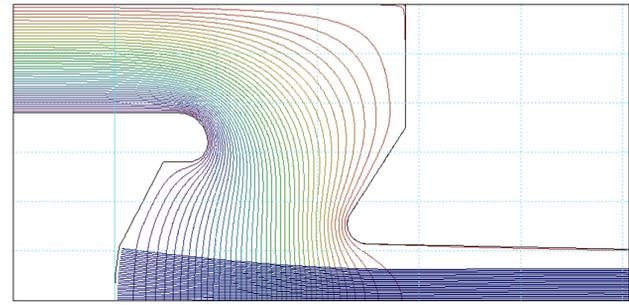


Figure 8: Gun design.

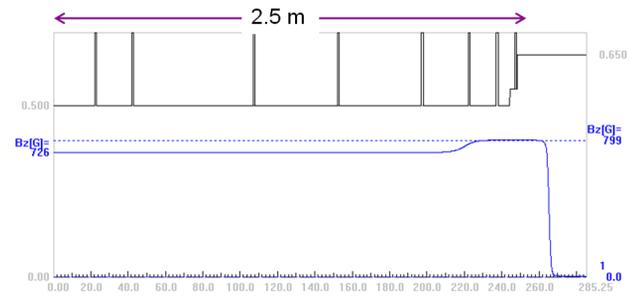


Figure 9: COM based klystron geometry and ramped magnetic field.

The final BAC optimized design achieved 84% efficiency and the final COM optimized design achieved 87% efficiency. To meet the design objective of 85% efficiency, CCR designed a depressed collector for the BAC klystron. The design is shown in Figure 10 and increased the overall BAC based klystron efficiency to 87%. The final BAC design consisted of 15 cavities and the final COM design consisted of 8 cavities.

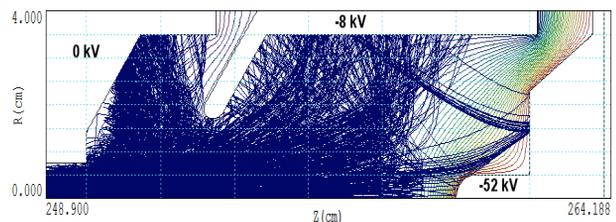


Figure 10: Depressed collector design for BAC based klystron.

The COM based klystron will be used as the basis for the Phase II SBIR. The klystron length is longer than the BAC klystron but simpler from a manufacturing standpoint due to the fact that the design only requires 8 cavities. The efficiency from the optimization study was also higher for the COM design so a depressed collector will likely not be needed, further simplifying the manufacturing and reducing costs.

The final phase space of the COM design is shown in Figures 11 and 12 and a plot of the output power in Tesla is shown in Figure 13. A ripple is observed in the output power plot of Figure 13. The ripple is most likely due to reflected electrons as shown in Figure 12. This effect will be further studied and mitigated in the Phase II SBIR.

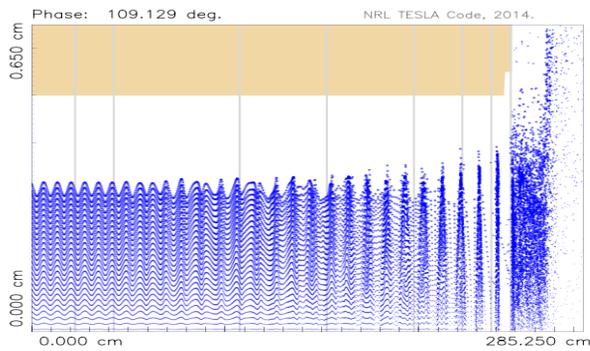


Figure 11: Particle phasespace in Tesla for the COM klystron design.



Figure 12: Velocity versus distance in Tesla for the COM klystron design.

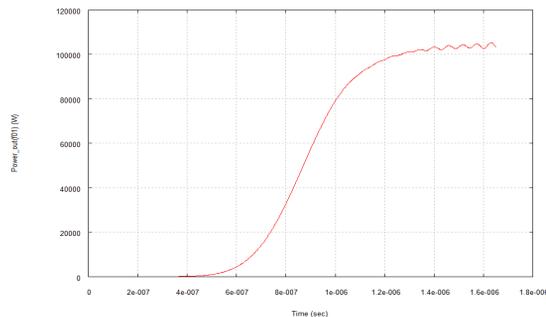


Figure 13: Output power versus time for the final COM based klystron design.

## CONCLUSION

COM and BAC methods achieved 87% and 84% klystron efficiency respectively. The BAC klystron efficiency was increased to 87% with a depressed collector. The SOGA method under Dakota using Galaxy Simulation Builder was an integral part of achieving the high efficiency results. We empirically found that the ordering of the triplet achieved higher efficiency when ordered BCA as opposed to BAC.

## REFERENCES

[1] A. Beunas, G. Faillon, and S. Choroba, "A high power long pulse high efficiency multi beam klystron," presented at the 5th MDK Workshop, Geneva, Switzerland, Apr. 2001.

[2] I. A. Guzilov, "BAC method of increasing the efficiency in klystrons." Vacuum Electron Sources Conference (IVESC), 2014 Tenth International, IEEE, 2014.

[3] A. Yu. Bajkov, D. M. Petrov "Problems of creation powerful and super-power klystrons with efficiency up to 90%", in *Proc. Int. University Conf. Electronics and Radio physics of Ultra-high Frequencies*, St. Petersburg, May 1999, pp. 5–8.

[4] B. Adams *et al.*, "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.0 User's Manual," Sandia Technical Report SAND2010-2183, 2015.

[5] Stellar Science Ltd Co. Galaxy Simulation Builder (GSB) User Guide, Version 6.6. High Power Electromagnetic Division, Air Force Research Lab, Kirtland, NM, 2017.

[6] A. J. Jensen *et al.* "Sheet beam klystron simulations using AJ-DISK." Vacuum Electronics Conference, 2006 held Jointly with 2006 IEEE International Vacuum Electron Sources., IEEE International. IEEE, 2006.

[7] I. A. Chernyavskiy, A. N. Vlasov, T. M. Antonsen, Jr., S. J. Cooke, B. Levush, and K. T. Nguyen, "Simulation of Klystrons with Slow and Reflected Electrons Using Large-Signal Code TESLA," in *IEEE Transactions on Electron Devices*, Vol. 54, No. 6, pp. 1555-1561, JUNE 2007.

[8] CERN. <https://indico.cern.ch/event/656491/contributions/2938743/>

[9] J. Petillo *et al.*, "The MICHELLE three-dimensional electron gun and collector modelling tool: Theory and design," *IEEE Trans. Plasma Sci.*, Vol. 30, No. 3, 2002.

# MODE-ANALYSIS METHODS FOR THE STUDY OF COLLECTIVE INSTABILITIES IN ELECTRON-STORAGE RINGS\*

M. Venturini<sup>†</sup>, LBNL, 94720 Berkeley, CA, USA

## Abstract

We review recent progress on the application of mode analysis to the study of collective instabilities in electron storage rings with Higher Harmonic RF Cavities (HHCs). The focus is on transverse instabilities in the presence of a dominant resistive-wall impedance, a problem of particular relevance to the new generation of diffraction-limited light sources. The secular equation is solved after applying a regularizing transformation, a key step to obtain numerically accurate solutions. We provide a demonstration that with vanishing chromaticity and in the absence of radiation damping the beam motion is always unstable. This is in contrast to the classical Transverse-Mode-Coupling Instability (TMCI) without HHCs, which is known to exhibit a well defined instability threshold.

## INTRODUCTION

A narrow vacuum chamber to accommodate strong magnets or high-performance Insertion Devices (ID) and use of bunch-lengthening Higher-Harmonic Cavities (HHCs) to reduce intrabeam scattering are two distinctive features of the new generation of storage-ring light sources. This paper concerns itself with the HHC effect on the transverse collective instabilities induced by the Resistive Wall (RW) impedance, which in the new machines is a major, if not the largely dominant, source of transverse impedance due to the small chamber aperture.

HHCs achieve bunch lengthening by introducing an amplitude dependence in the synchrotron oscillation frequency and therefore altering the linear character of the longitudinal motion. The resulting frequency spread is commonly associated with the expectation of a beneficial impact on the beam stability, as alluded by the often-encountered ‘Landau cavities’ designation. The reality, however, is more nuanced. While HHCs have the potential to reduce or eliminate certain instabilities through the Landau damping mechanism, whether they actually do depends on a number of other factors. In fact, the presence of HHCs can under some circumstances degrade beam stability. This is known although not widely acknowledged for longitudinal multi-bunch instabilities [1–3]. The main point to be made here is that such a degradation can be realized in the transverse plane as well. This paper illustrates the main results reported in [4], to which we refer for the more technical details.

The focus is on developing a mode-analysis theory in the presence of HHCs applicable to single-bunch instabilities at vanishing chromaticities. We base the analysis on the

familiar DC-conductivity, RW impedance model for a vacuum chamber with uniform circular cross section of radius  $b$ , length  $L$ , and conductivity  $\sigma_c$  (cgs units):

$$Z_y(k) = \frac{\text{sign}(k) - i}{\sqrt{|k|}} \frac{L}{b^3} \sqrt{\frac{2}{\pi c \sigma_c}}, \quad (1)$$

with wake-function  $W_y(z) = -2L\sqrt{c}/(\pi b^3 \sqrt{\sigma_c |z|})$ , for  $z \leq 0$  (and vanishing otherwise).

## THE CLASSICAL TMCI (NO HHCS)

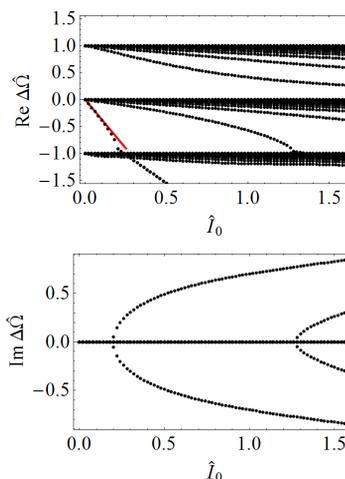


Figure 1: Classical TMCI in the absence of HHCs: real (top) and imaginary (bottom) parts of the mode complex-number frequency shift  $\Delta\Omega = (\Omega - \omega_y)/\omega_{s0}$  over a bunch-current range. The red line in the top picture is the tunes shift for the rigid dipole mode as given by Eq. (7).

In the absence of HHCs the longitudinal motion is linear and at zero chromaticities the beam is susceptible to the Transverse-Mode Coupling Instability (TMCI). The characteristic signature of the instability is the convergence of the dipole ( $m = 0$ ) and head-tail ( $m = -1$ ) azimuthal-mode oscillation frequencies at the critical bunch current [5]. The starting point for the analysis is the linearized Vlasov equation for the perturbation

$$g_1(r, \varphi; t) = e^{-i\Omega t} \sum_{m=-\infty}^{\infty} R_m(r; \Omega) e^{im\varphi}, \quad (2)$$

written as a superposition of azimuthal modes with radial functions  $R_m$  and depending on the longitudinal-motion amplitude/angle coordinates  $(r, \varphi)$ . The perturbation  $g_1$  has a physical interpretation as the transverse (say vertical) offset of the electrons contained in the infinitesimal

\* Work supported by the US Department of Energy Contract No. DE-AC02-05CH11231.

<sup>†</sup> mventurini@lbl.gov

phase-space area  $\Delta r \Delta \varphi$  centered at  $(r, \varphi)$ . Mode analysis entails solving an eigenvalue problem in the form of a system of integral equations for the unknown eigenvectors  $R_m(\rho) \equiv R_m(\rho \sigma_{z0}; \Omega)$  of the form

$$(\Delta \hat{\Omega} - m)R_m(\rho) + i \hat{I}_0 e^{-\rho^2/2} \times \sum_{m'=-\infty}^{\infty} \int_0^{\infty} R_{m'}(\rho') \mathcal{G}_{m,m'}(\rho, \rho') \rho' d\rho' = 0, \quad (3)$$

where we have scaled the radial coordinate by the rms natural bunch length at equilibrium  $\rho = r/\sigma_{z0}$ , and

$$\hat{I}_0 = \frac{Nr_c c}{(2\pi)^{3/2} \gamma v_{s0} b^3 \sqrt{c \sigma_c \sigma_{z0}}} \frac{\beta_{y,u} L_u}{2\pi} \quad (4)$$

is a dimensionless current parameter depending on the bunch population  $N$ , relativistic factor  $\gamma$ , synchrotron tune  $\nu_{s0} = \omega_{s0}/\omega_0$ , undulator length  $L_u$ , where the relevant source of RW is localized, and betatron function  $\beta_{y,u}$  at the undulator. The sought eigenvalue is the complex-number frequency shift  $\Delta \hat{\Omega} = (\Omega - \omega_y)/\omega_{s0}$  from the betatron oscillation frequency  $\omega_y = \nu_y \omega_0$ , scaled by the synchrotron-oscillation frequency  $\omega_{s0}$ . In Eq. (3) the kernel involves the Bessel functions  $J_m$  and has the form

$$\mathcal{G}_{m,m'} = c_{m,m'} \int_0^{\infty} \frac{dk}{\sqrt{k}} J_{|m|}(k\rho) J_{|m'|}(k\rho'), \quad (5)$$

with coefficients  $c_{m,m'} = i^{(m-m')} \{ [1 - (-1)^{m+m'}] - i [1 + (-1)^{m+m'}] \} \times [\text{sign}(m)]^m \times [\text{sign}(m')]^{m'}$ .

The conventional approach to solving the eigenvalue problem is to discretize Eq. (3) by expanding  $R_m(\rho)$  over an orthonormal polynomial basis. Since the bunch equilibrium is gaussian, a natural and efficient choice for this problem is to use Gauss-Legendre polynomials which yield fairly accurate results upon retaining only a few (possibly just one) radial-mode components for the relevant azimuthal modes  $|m| \leq m_{\max} = 1$ .

Alternatively, and for this problem less efficiently, one can introduce a discretization where  $R_m(\rho)$  is represented as a step-wise function on a grid with  $n_{\max}$  grid points. The problem is reduced to finding the roots of the secular equation

$$\text{det}[\Delta \hat{\Omega} - \mathbf{B}] = 0, \quad (6)$$

where  $\mathbf{B}$  is a  $[(2m_{\max} + 1)n_{\max}]^2$ -dimension square matrix. The eigenvalue-analysis result obtained with a uniform  $n_{\max} = 40$  grid and  $m_{\max} = 1$  is shown in Fig. 1. A finite bunch current removes the degeneracy of the radial modes and as its value increases the (real) frequency of one radial-mode component after the other (all having azimuthal mode number  $m = 0$ ) is seen to cross with those relative to the head-tail mode  $m = -1$ , at which point the imaginary part of  $\Delta \hat{\Omega}$  becomes positive signaling instability. The lowest-current crossing involves the  $m = 0$  mode with  $R_0(\rho) \sim e^{-\rho^2/2}$  radial component (rigid dipole) and occurs at  $\hat{I}_0 \approx 0.197$ . To good approximation the current dependence of the real-part of the frequency shift is given by (red line in the top picture of Fig. 1)

$$\text{Re } \Delta \hat{\Omega} = -\Gamma(1/4) \hat{I}_0, \quad (7)$$

where  $\Gamma(1/4) \approx 3.63$  is Euler's Gamma function.

For a practical illustration loosely based on parameters from the ALS-U design studies [6], assume that RW is the only relevant source of transverse impedance and that it is dominated by aggressively narrow ID vacuum chambers of  $b = 3$  mm radius, Table 1. There are 10 straight sections available for IDs and we conservatively assume that the vacuum chamber is identically narrow in all of them. Finally, assuming copper material for the vacuum chamber ( $\sigma_c = 5.3 \times 10^{17} \text{ s}^{-1}$  in cgs units, or  $5.9 \times 10^7 \text{ } \Omega^{-1} \text{ m}^{-1}$  in MKS units), we find a critical  $N_{c0} = 3.3 \times 10^{10}$  bunch population for the instability threshold, equivalent to 8.1 mA single-bunch current, vs. a design  $I_b = 1.76$  mA.

Table 1: Beam/Machine Parameters Loosely Based on ALS-U

Ring circumference		196.5 m
Beam energy		2 GeV
Design bunch current	$I_b$	1.76 mA
Vertical tune	$\nu_y$	20.368
Momentum compaction		$2.79 \times 10^{-4}$
Natural energy spread		$0.835 \times 10^{-3}$
Energy loss per turn		182 keV
Vertical damping time	$\tau_y$	14.4 ms
Main rf cavity voltage		0.76 MV
Main rf cavity frequency		500 MHz
Harmonic rf cavity frequency		1.5 GHz
Rms bunch length (no HHCs)	$\sigma_{z0}$	3.2 mm
Linear synchr. tune (no HHCs)	$\nu_{s0}$	$2.3 \times 10^{-3}$
Rms bunch length with HHCs	$\sigma_z$	13 mm
Avg. synchr. tune with HHCs	$\langle \nu_s \rangle$	$0.44 \times 10^{-3}$
Total ID length	$L_u$	40 m
ID vacuum chamber radius	$b$	3 mm
Avg. beta function along IDs	$\beta_{y,u}$	3 m

## STABILITY ANALYSIS WITH HHCS

Some simplifying assumptions are made to represent the single-particle longitudinal motion in the presence of HHCs. The first is to approximate the total RF potential combining main and harmonic cavities as a purely-quartic polynomial function of the particle longitudinal-coordinate  $z$ , yielding an exactly linear dependence of the synchrotron-oscillation frequency on the oscillation amplitude  $r$ . This is a very good approximation in the regime where the HHCs are tuned for 'optimal' (*i.e.* maximally flat) bunch lengthening. The second approximation is to write  $z = r \cos \varphi$ , as for an harmonic oscillator. Somewhat surprisingly, for a purely quartic potential this is a fairly good approximation, entailing only a few % error [7]. With these approximations the system of integral equations becomes

$$(\Delta \hat{\Omega} - m\rho)R_m(\rho) + i \hat{I} e^{-h_1 \rho^4} \times \sum_{m'=-\infty}^{\infty} \int_0^{\infty} R_{m'}(\rho') \mathcal{G}_{m,m'}(\rho, \rho') \rho'^2 d\rho' = 0, \quad (8)$$

where now the radial coordinate  $\rho = r/\sigma_z$  is scaled by the length  $\sigma_z$  of the bunch stretched by the HHCs, the frequency

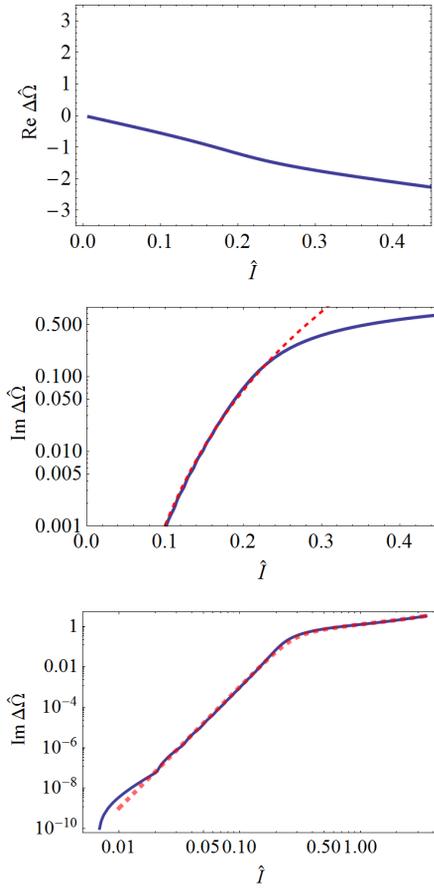


Figure 2: Stability analysis in the presence of HHCs after applying the regularizing transformation to the eigenvalue problem. Real (top) and imaginary (middle and bottom) parts of the root with largest imaginary part of the secular equation (10), as functions of the current parameter  $\hat{I}$ . The bottom picture contains the same data as the middle picture but on a double-log scale. In the limit of small  $\hat{I}$  the numerical solution is consistent with the power law  $\text{Im } \Delta\hat{\Omega} = (2^{5/3}\hat{I})^6$ , red dashed curve in the middle picture. Overall, the numerical solution is reasonably well fitted by Eq. (11), red dashed curve in the bottom picture. Calculation done with  $n_{\max} = 40$ ,  $m_{\max} = 1$ , and  $\rho_{\max} = 3$ .

shift  $\Delta\hat{\Omega} = (\Omega - \omega_y)/(h_2\langle\omega_s\rangle)$  is scaled by the synchrotron-oscillation frequency averaged over the bunch  $\langle\omega_s\rangle$ , and the dimensionless current parameter reads

$$\hat{I} = \frac{Nr_c c}{\pi^{5/2}\gamma\langle v_s\rangle b^3\sqrt{c\sigma_c\sigma_z}} \frac{\beta_{y,u}L_u}{2\pi}. \quad (9)$$

exhibiting  $\langle v_s\rangle$ , the average of the synchrotron-oscillation tune over the bunch, in place of  $v_{s0}$  appearing in Eq. (4). The quantities  $h_1$  and  $h_2$  are numerical coefficients,  $h_1 = 2\pi^2/\Gamma(1/4)^4 \approx 0.114$  and  $h_2 = 2^{3/4}\pi^{3/2}/\Gamma(1/4)^2 \approx 0.712$ .

As for the integral equation, the main difference from Eq. (3) is the appearance of the  $\rho$  dependence in the factor multiplying  $R_m(\rho)$  in the first term. This term is familiar from the analysis of Landau damping in plasma waves or longitudinal instabilities, raising a flag that care should be taken

to handle the singularity occurring when the above factor vanishes for certain values of  $\rho$ . Because of the singularity, the eigenfunctions of Eq. (8) are in general not ordinary functions but distributions in the sense of Dirac [8, 9] and finite-dimension approximations of the problem are not guaranteed to converge [10].

The proper way to proceed is to 'sweep the singularity under the (integral) rug' by introducing a simple change of variable of the unknown function [11],  $R_m(\rho) \rightarrow S_m(\rho) = (\Delta\hat{\Omega} - m\rho)R_m(\rho)e^{h_1\rho^4}$ , yielding the transformed integral equations

$$S_m(\rho) + i\hat{I} \sum_{m'=-\infty}^{\infty} \int_0^{\infty} \frac{S_{m'}(\rho')e^{-h_1\rho'^4}}{\Delta\hat{\Omega} - m'\rho'} \mathcal{G}_{m,m'}(\rho, \rho')\rho'^2 d\rho' = 0.$$

These equations can now be safely discretized with the prescription that the integration path in  $\rho$  should be deformed to go under the pole if  $\Delta\hat{\Omega}$  is real or has negative imaginary part. Since we are primarily interested in establishing the condition for instability ( $\text{Im } \Delta\hat{\Omega} > 0$ ) we can do without the path deformation, provided that we take numerical care to insure the necessary accuracy when the imaginary part of  $\Delta\hat{\Omega}$  is positive but small. An effective integration strategy is to approximate the numerator in the integral by a piece-wise linear or quadratic polynomial, in which case the integral can be carried out analytically.

Upon discretization, the above equation is reduced to the form  $[\mathbf{1} + \mathbf{B}(\Delta\hat{\Omega})]\vec{S} = 0$ , where,  $\mathbf{B}(\Delta\hat{\Omega})$  is now a  $\Delta\hat{\Omega}$ -dependent,  $[(2m_{\max} + 1)n_{\max}]^2$ -dim matrix. Unlike Eq. (6), the resulting secular equation

$$\det[\mathbf{1} + \mathbf{B}(\Delta\hat{\Omega})] = 0 \quad (10)$$

is a transcendental (vs. polynomial) equation in the frequency shift  $\Delta\hat{\Omega}$  and in principle more difficult to solve. In practice, we found that a Newton-method search appropriately initiated never failed to converge. The result of our numerical analysis is shown Fig. 2, reporting real and imaginary parts of the frequency shift of the most unstable mode in a calculation using  $n_{\max} = 40$  radial grid points and  $m_{\max} = 1$ . The main result of this analysis is that transverse single-bunch motion in the presence of the RW impedance is unstable at any current.

Over a large current range the imaginary part of the frequency of the most unstable mode is well fitted by the function (red dashed line in the bottom picture of Fig. 2)

$$\text{Im } \Delta\hat{\Omega} = \frac{(2^{5/3}\hat{I})^6}{1 + 0.55 \times (4\hat{I})^5 [1 + \tanh(\hat{I}/2)]}. \quad (11)$$

It is tempting to make the conjecture that  $\text{Im } \Delta\hat{\Omega} = (2^{5/3}\hat{I})^6$  may be the exact asymptotic limit for  $\hat{I} \rightarrow 0$ . It is seen to track the numerical data quite accurately for  $\hat{I} \lesssim 0.2$ .

Having argued that for proper numerical treatment of the problem it is important to introduce a regularizing transformation, it is nonetheless instructive to naively apply the discretization method employed when HHCs are absent. Effectively, this is equivalent to studying a modified physics model where the unperturbed beam distribution in phase space consists of a set of  $n_{\max}$  equally spaced, concentric,

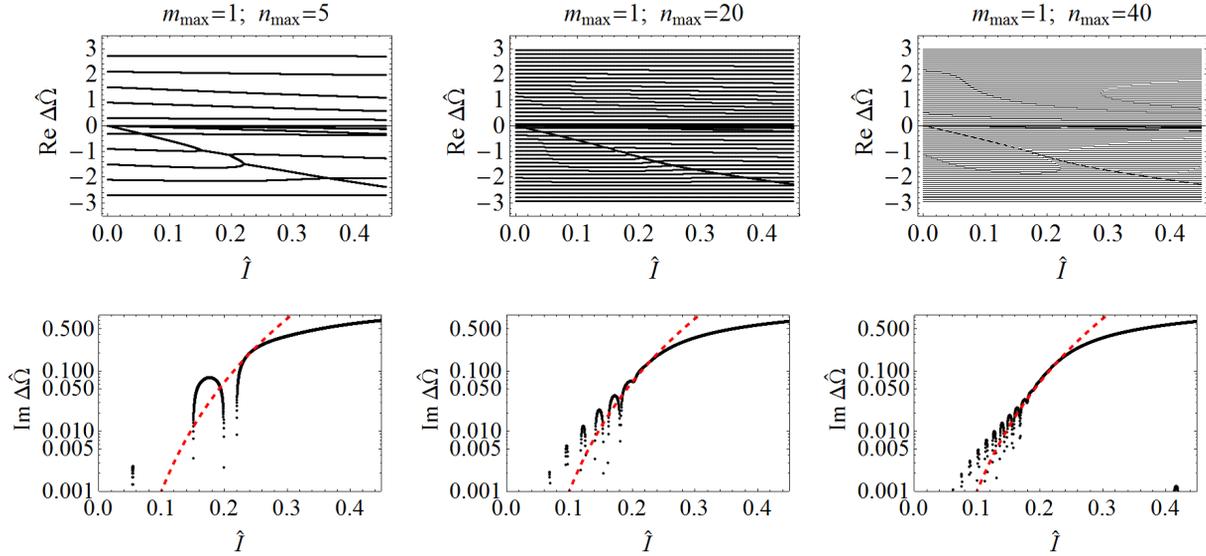


Figure 3: Stability analysis in the presence of HHCs using the conventional eigenvalue-method without the regularizing transformation. The top (bottom) pictures show the real (imaginary) parts of the modes complex-number frequency shifts  $\Delta\hat{\Omega} = (\Omega - \omega_y)/(h_2\langle\omega_s\rangle)$  as functions of the current parameter  $\hat{I}$ , for increasingly finer (left to right) grids in the radial variable  $\rho$ , as indicated. The bottom pictures are in log scale and report only the frequencies with positive imaginary part (unstable modes). Particularly at small  $\hat{I}$ , convergence to what we believe is the exact asymptotic solution of the infinite-dimension problem  $\text{Im } \Delta\hat{\Omega} = (2^{5/3} \hat{I})^6$ , valid for  $\hat{I} \lesssim 0.2$ , red dashed curve, appears slow if not outright questionable.

invariant shells. Results are shown in Fig. 3. While it is apparent that these pictures do not extrapolate well into the continuum limit, they provide valuable insight and suggest that the basic mechanism of mode coupling is still at play. First, notice that all the radial modes relative to azimuthal mode  $m = 0$  are degenerate but not those relative to  $m \neq 0$ , even at vanishing current. This is in contrast to the longitudinal linear-motion case (no HHCs), where at zero current all radial modes for any  $m$  are degenerate. The reason, of course, is related to the fact that particles on different radial shells have different winding (synchrotron oscillation) frequencies. The emergence of instability is triggered by the convergence of one of the  $m = 0$  radial modes frequency with that of one of the  $m = -1$  radial modes. In analogy to the linear case, the offensive  $m = 0$  radial mode has the form of the bunch equilibrium  $R_0(\rho) \sim e^{-h_1\rho^4}$  (rigid dipole). The difference with the linear case is that coupling can now occur at arbitrarily low current as we allow for a finer and finer resolution of the radial beam distribution. For currents less than  $\hat{I} \sim 0.25$ , regions of instability appear interleaved with regions of stability, with the extent of the latter progressively reduced when we increase the number of grid points  $n_{\text{max}}$ . The  $\sim 0.25$  edge corresponds to the radial extension ( $\rho \sim 1$ ) of the beam distribution (outer shells become quickly underpopulated for  $\rho > 1$  and do not contribute to the coupling).

## THE TAKE-HOME RESULT

In electron storage rings radiation damping will eventually prevail if the bunch current is not too high. The condition  $\text{Im } \Omega = \tau_y^{-1}$ , where  $\tau_y$  is the vertical radiation damping time, defines the critical current parameter  $\hat{I} = \hat{I}_c$  as follows:

$\text{Im } \Omega = h_2\langle\omega_s\rangle\text{Im } \Delta\hat{\Omega} = h_2\langle\omega_s\rangle(2^{5/3}\hat{I}_c)^6 = \tau_y^{-1}$ , having restricted our analysis to the regime where the  $\text{Im } \Delta\hat{\Omega} \propto \hat{I}^6$  power law applies. We have

$$\hat{I}_c = \frac{2^{-5/3}}{(h_2\tau_y\langle\omega_s\rangle)^{1/6}} \approx 0.245 \times \left(\frac{T_0}{\tau_y\langle v_s\rangle}\right)^{1/6}. \quad (12)$$

More expressively, we can relate  $N_c$ , the critical bunch population in the presence of HHCs, and  $N_{c0}$ , the critical bunch population in the absence of HHCs, when all the relevant machine parameters are kept unchanged while the HHCs are turned on and off. Combining Eqs. (4), (9) and (12) gives

$$N_c = N_{c0} \times \frac{\pi}{8 \times 2^{1/6}\hat{I}_{c0}} \left(\frac{1}{\tau_y h_2\langle\omega_s\rangle}\right)^{1/6} \frac{\langle v_s\rangle}{v_{s0}} \left(\frac{\sigma_z}{\sigma_{z0}}\right)^{1/2}, \quad (13)$$

where  $\hat{I}_{c0} \approx 0.197$  is the critical current parameter for the onset of the TMC-Instability in the linear case.

Making use of the relationship between synchrotron tunes and bunch lengths with and without HHC for the specific case of third-harmonic cavities, see [4], we obtain the final result

$$N_c \approx 1.15 \times N_{c0} \left(\frac{T_0}{\tau_y v_{s0}}\right)^{1/6} \left(\frac{\sigma_{z0}}{\sigma_z}\right)^{1/3}. \quad (14)$$

Note that the quantity elevated to the 1/6 power now depends on  $v_{s0}$  not  $\langle v_s\rangle$ . Using the machine parameters from the ALS-U example (Table 1), we find a critical current  $\hat{I}_c \approx 0.168 < 0.2$  placing the system in the regime of the validity of the  $\text{Im } \hat{\Omega} \propto \hat{I}^6$  scaling, see Fig 2. Finally, from Eq. (14), we conclude  $N_c/N_{c0} \approx 0.37$ , corresponding to  $I_b = 3$  mA, *i.e.* the instability threshold with HHCs is less than 40% of that without. More in detail,  $[T_0/(\tau_y v_{s0})]^{1/6} \approx 0.52$  and

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

$(\sigma_{z0}/\sigma_z)^{1/3} \approx 4^{-1/3} \approx 0.62$ . A macroparticle simulation with *elegant* [12] confirms the  $\sim I_b^6$  scaling, Fig. 4, and overall is reasonably close to the theory.

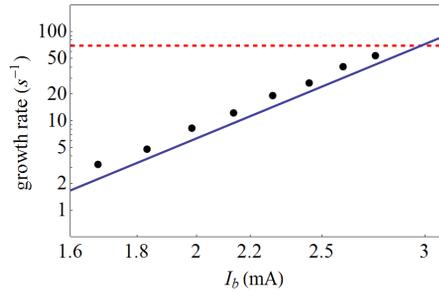


Figure 4: The TMC-Instability growth rate in the presence of HHCs vs. bunch current from macroparticle simulations (dots) tracks reasonably well the theory (solid line). The simulation does not include radiation damping but for reference the expected radiation damping rate (red dashed line) is also reported. ALS-U parameters as in Table 1.

## CONCLUSION

In summary, we have provided a demonstration that, in the absence of radiation damping the transverse motion at vanishing chromaticities is always unstable, regardless of bunch current, with growth rate varying from a  $\text{Im } \Omega \sim I_b^6$  dependence at small bunch current  $I_b$  to  $\text{Im } \Omega \sim I_b$  for larger  $I_b$ , the former being more likely to be encountered in the physical systems of interest. Because of the strong 6<sup>th</sup>-power dependence, macroparticle-simulations results could be easily misinterpreted as indicating the existence of a current threshold if the simulation time is not sufficiently long [13]. Finally, we caution that the formulas in the last section are strictly dependent on the RW nature of the assumed impedance model. Work to analyze impedances of different form is left to future studies. The study presented here is for vanishing-chromaticities. Finite chromaticities have a known stabilizing effects. Interestingly, macroparticle simulation work indicates that their stabilizing effect is magnified not reduced by the presence of HHC, see [14] for multi-bunch and Fig. 5 for a single-bunch study. Extension of the theory to multi-bunch instabilities and finite chromaticities will be addressed elsewhere.

## REFERENCES

- [1] Y-C. Chao and K-Y Ng, “Analytical and numerical evaluation of Landau cavities in the Fermilab Booster”, in *Proc. EPAC’88*, Rome, Italy, Jun 1988, p. 666–668, 1988.
- [2] J. Jacob, O. Naumann, and W. Beinhauer, “Effects of a harmonic cavity at the ESRF”, in *Proc. PAC’99*, New York, NY, USA, Mar-Apr 1999, p. 1647–1649, 1999.
- [3] M. Venturini, “Passive higher-harmonic RF cavities with general settings and multi-bunch instabilities in electron storage rings”, submitted to *Phys. Rev. Accel. Beams*, 2018.

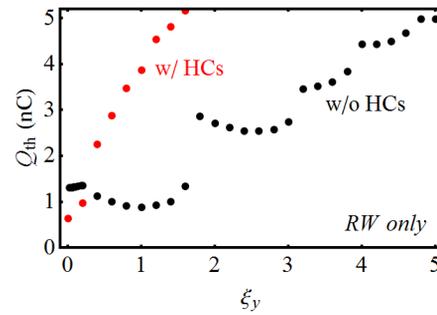


Figure 5: Single-bunch instability threshold vs. chromaticity with and without HHCs (ALS-U studies). *Elegant* [12] simulation. The presence of HHCs considerably enhances the stabilizing effect of a finite positive chromaticity. Radiation damping included.

- [4] M. Venturini, “Harmonic cavities and the transverse mode-coupling instability driven by a resistive wall”, *Phys. Rev. Accel. Beams*, vol. 21, pp. 024402, 2018. doi:10.1103/PhysRevAccelBeams.21.024402
- [5] A. Chao, *Physics of collective beam instabilities in high energy accelerators*, John Wiley & Sons, Inc., 1992.
- [6] C. Steier *et al.*, “Status of the conceptual design of ALS-U”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi:10.18429/JACoW-IPAC2017-WEPAB104
- [7] S. Krinsky and J. M. Wang, “Longitudinal instabilities of bunched beams subject to a non-harmonic RF potential”, *Part. Accel.*, vol. 17, pp. 109, 1985. doi:10.1103/PhysRevSTAB.4.114401
- [8] N. G. van Kampen, “On the theory of stationary waves in plasmas”, *Physica*, vol 21, pp. 949, 1955. doi:10.1016/S0031-8914(55)93068-8
- [9] K. Case, “Plasma oscillations”, *Annals of Physics*, vol. 7, pp. 349, 1959. doi:10.1016/0003-4916(59)90029-6
- [10] G. R. Bart and R. Warnock, “Linear integral equations of the third kind”, *SIAM Journal of Analysis*, vol 4, pp. 609, 1973. doi:10.1137/0504053
- [11] R. Warnock, G. Stupakov, M. Venturini, and J. Ellison, “Linear Vlasov analysis for stability of a bunched beam”, in *Proc. EPAC’04*, Lucerne, Switzerland, Jul 2004, paper WEPLT159, p. 2215–2217, 2004.
- [12] M. Borland, “elegant: A Flexible SDDS-compliant code for accelerator simulation”, Argonne National Laboratory Advanced Photon Source Report No. LS-287 (2000).
- [13] S. Krinsky, “Simulation of transverse instabilities in the NSLS II storage ring”, BNL Informal Report, BNL-75019-2005-IR (2005).
- [14] F. J. Cullinan, R. Nagaoka, G. Skripka, and P. F. Tavares, “Transverse coupled-bunch instability thresholds in the presence of a harmonic-cavity-flattened RF potential”, *Phys. Rev. Accel. Beams*, vol 19, pp. 124401, 2016. doi:10.1103/PhysRevAccelBeams.19.124401

# HOM-MITIGATION FOR FUTURE SPS 33-CELL 200 MHz ACCELERATING STRUCTURES

P. Kramer<sup>1\*</sup>, C. Vollinger, CERN, Geneva, Switzerland

<sup>1</sup>also at Institute of High Frequency Technology (IHF), RWTH Aachen University, Germany

## Abstract

The CERN SPS 200 MHz travelling wave (TW) accelerating structures pose an intensity limitation for the planned High Luminosity (HL-) LHC upgrade. Higher-order modes (HOMs) around 630 MHz have been identified as one of the main sources of longitudinal multi-bunch instabilities. Improved mitigation of these HOMs with respect to today's HOM-damping scheme is therefore an essential part of the LHC injectors upgrade (LIU) project.

The basic principles of HOM-couplers in cavities and today's damping scheme are reviewed, before illustrating the numerous requirements an improved damping scheme for the future 33-cell structures must fulfil. These are, amongst others, the mitigation of HOMs situated in the lower part of the structure where there are no access ports for extraction, a sufficient overall damping performance and an acceptable influence on the fundamental accelerating passband (FPB). Different approaches tackling these challenges are investigated and their performance, advantages and pitfalls are evaluated by ACE3P and CST electromagnetic (EM) field solver suites.

## INTRODUCTION

The Super Proton Synchrotron (SPS) at CERN relies on a 200 MHz multi-cell travelling-wave structure (TWS) for particle acceleration.

Beam dynamic simulations showed that to achieve stable beams for future HL-LHC intensities an additional mitigation by a further factor three of already heavily damped HOMs around 630 MHz is necessary in these structures [1]. A general overview of the SPS accelerating structure and the corresponding longitudinal and transverse HOM damping schemes in use today were presented in [2, 3]. For beam loading reasons, shorter 33-cell structures will be employed after the LIU upgrade together with the 44-cell structures in use today. The longitudinal damping scheme deployed today on 44- and 55-cell structures is insufficient for the HOMs around 630 MHz with future beam intensities. This equally applies if this damping scheme is used on the 33-cell structures (shown as black HOM-couplers in Fig. 1).

\* patrick.kramer@cern.ch

As outlined in [3], these couplers were optimized on a single 11-cell section featuring HOMs with an integer multiple of a  $\pi/11$  phase advance per cell. One such spare section is shown in Fig. 2. On 33-cells, HOMs with phase advances



Figure 2: 11-cell section of the accelerating TWS.

that are not allowed on 11-cells exist and the performance of the HOM-coupler on these modes has to be verified.

To achieve the required additional damping by a factor three of the HOMs around 630 MHz this contribution describes the systematic improvement of the existing damping scheme. The resulting beam impedance is thereby calculated by time domain wakefield simulations. Due to the importance of the damping upgrade for the future operation of the SPS, confirmation of the results by two different solver types was desired. The finite-difference wakefield solver of the CST suite [4] and the finite-element time domain solver (T3P) of the ACE3P [5] suite were used for this purpose.

The first section of this work details the model set-up and the simulation settings used for the two solvers. The HOM-mitigation strategy then comprises the following steps. First, additional couplers are placed in cells with strong electric field of the most dangerous modes present in the 33-cell configuration. These modes feature a high geometry factor R/Q. As a second step, the HOM-coupler is optimized to reach close-to-critical coupling to the HOMs in the relevant frequency range. Sufficient damping can however not be achieved merely by adding HOM-couplers in the available access ports at the top of each cell (Fig. 2). This is due to the fact that the top/ bottom symmetry of the structure is violated and as a result the EM fields of some modes are partially pushed towards the lower half of the cavity where no dedicated access ports are available for HOM-damping (for more details see [3] and compare Fig. 8a later in the text). One particular example for this effect is the high-Q  $17\pi/33$  mode. Several means of damping modes in the lower half

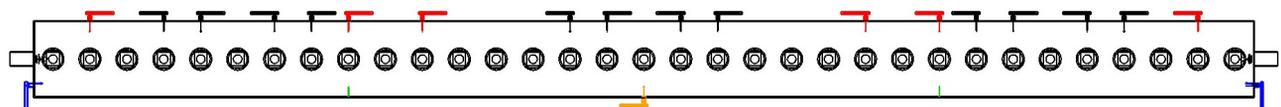


Figure 1: Several HOM-mitigation options shown together in a 3-section model. Black: existing longitudinal damping scheme. Red: additional couplers. Blue: end-plate couplers. Orange: VPP coupler. Green: mitigation by resonant posts.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

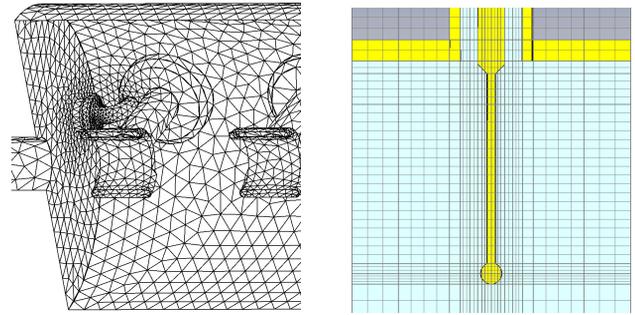
of the structure are therefore studied as a third step. Finally, a new principle of HOM mitigation by introducing slight structural changes via resonant posts was developed and is demonstrated. This way, machining of cavity parts that could pose a potential integrity risk could be avoided. The task of finding a satisfactory mitigation scheme is aggravated even further by the fact that the FPB may merely be influenced to a very small extent by couplers and posts and in addition some access ports are needed for transverse impedance damping.

## WAKEFIELD CALCULATIONS

Although particle acceleration is performed via a travelling wave FPB, the HOMs up to 1.3 GHz are of standing wave type. The fundamental power couplers (FPCs) couple merely very narrowband to some frequencies in the 630 MHz range. Due to modifications in the TWS's amplifiers, the FPCs situated in the end-cells of the accelerating structure (Fig. 1) are currently redesigned and their final configuration is not yet decided. With this comes uncertainty in how much the EM fields are perturbed by the presence of the FPCs and uncertainty in the amount of their coupling to and consequently damping of the HOMs. In addition, the impact of the final matching networks attached to the FPCs for the travelling-wave FPB on the HOMs remains undefined to date. The process of finding a sufficient damping scheme was therefore carried out by assuming the worst case of no coupling of the FPCs to the HOMs. The FPCs must be included in the model to obtain the appropriate field profiles due to their presence in the end-cells. However, they are short-circuited in the plane of the end-plates which then results in full reflection of any HOM power picked up by the FPCs, see Fig. 3a. To obtain the beam impedance for the FPB of course another wakefield simulation has to be conducted with the appropriate matching networks for the travelling-wave condition attached [6].

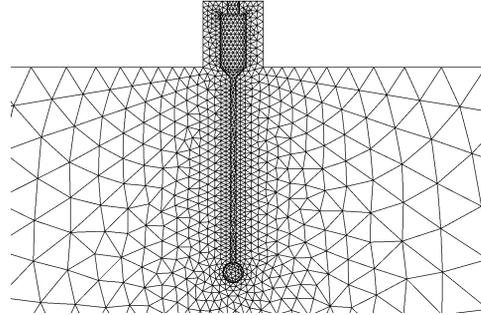
At present, the HOMs of the structure are only of interest up to a frequency of 1 GHz. In T3P the tank volume can then be meshed quite coarse (Fig. 3a) as curved tetrahedral elements are used that are however not visualized in Fig. 3. The electric pick-ups have to be meshed densely and the superiority of a tetrahedral over a hexahedral mesh in modelling the fine structures and making a smooth transition to the large tank volume is evident when comparing the meshes shown in Figs. 3b and 3c. Mesh and time step convergence studies were conducted and led to the final simulation settings given in Table 1. A magnetic boundary condition at the vertical symmetry plane ( $x=0$ ) was used. For details about the calculation methods used in T3P refer to [7].

The results for the impedance of HOMs in the harmful 630 MHz range of a 33-cell structure with today's damping scheme are shown in Fig. 4a. In addition to a discrepancy in frequency, the results obtained from the two solvers also differ significantly in at least one of the impedance peaks. The calculated longitudinal impedance of around 100 kΩ for the 33-cell structures is too high for HL-LHC beam



(a) Shorted FPC elbows and coarse meshing of tank volume for T3P.

(b) Meshing of electric pick-up in CST.



(c) Meshing of electric pick-up for T3P.

Figure 3: Discretization of the computation domain with tetrahedra for T3P and hexahedra in CST.

Table 1: Simulation Settings in T3P and CST Respectively

Parameter	T3P	CST
$f_{\max}$ [GHz]	1	1
bunch $\sigma_z$ [cm]	11.5	10
# mesh [Mio.]	1.2 - 1.4 tets	16 - 21 hex
$\Delta t$ [ps]	24	$\sim \approx 1.3$
wake [km]	2/1.5	3
basis order	2	-
linear solver	MUMPS	-

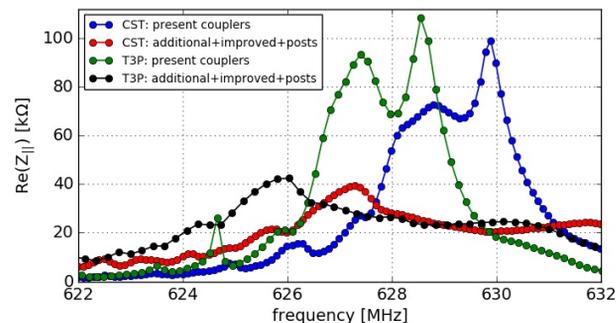
intensities. Several methods to mitigate this impedance are outlined in the following.

### STEP 1: ADDITIONAL HOM-COUPLERS

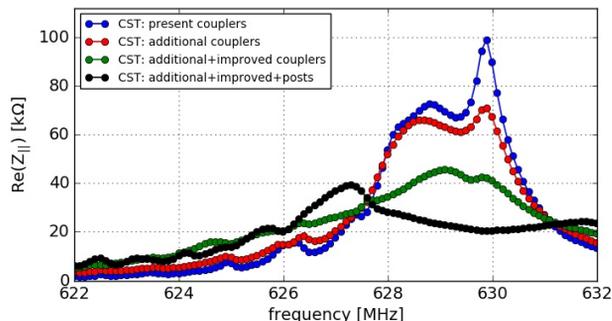
To further damp the two high-R/Q HOMs with  $14\pi/33$  and  $15\pi/33$  phase advance that exist in the 630 MHz frequency band, additional couplers should be put in cells where these two modes feature strong electric field. In this respect, the eigenvector solutions of a coupled resonator chain, as for example outlined in [8], can be used to obtain relative electric field strengths in the centre of each cell:

$$\xi(m, j) = \sqrt{\frac{2 - \delta_{mN}}{N}} \sin \left[ m\pi \left( \frac{2j - 1}{2N} \right) \right] \quad (1)$$

in cell  $j = 1, \dots, N$  for mode  $m = 1, \dots, N$ ,  $N$  being the total number of cells and  $\delta_{mN}$  the Kronecker delta. The eigenvalue solutions of the equation system generated by the simple resonator chain with nearest neighbour coupling are unable



(a) Comparison of CST and T3P impedance results.



(b) Damping improvement due to the different mitigation techniques.

Figure 4: Simulated longitudinal beam impedance in the 630 MHz range for the different HOM mitigation schemes calculated by the two EM solvers.

to predict mode frequencies or model the passband of the HOMs. However, when comparing to 3D EM field solver solutions the eigenvectors proved to predict the relative electric field amplitudes very well even for cases with a high number of HOM-couplers. Figure 5 shows the normalized mode amplitudes for the three most problematic modes in a 33-cell structure. For better readability and symmetry reasons, the amplitudes are plotted for the first 17 cells only. It

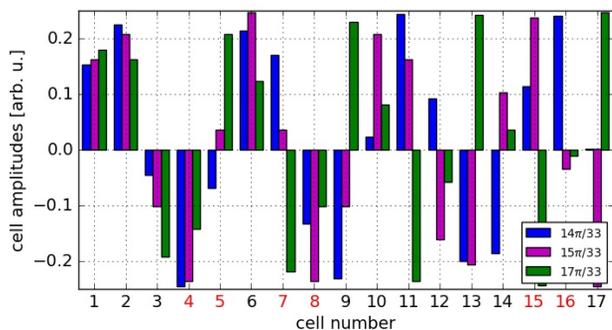


Figure 5: Mode amplitudes following Eq. (1) in the centre of cells for the three most problematic HOMs. Only the  $14\pi/33$ -mode is antisymmetric to cell 17. Red labels mark cells already occupied by today's damping scheme.

is obvious that some of the HOM-couplers in the present damping scheme can contribute well to the damping of the high-R/Q modes. Ideally, additional couplers would now be put in cells with strong electric field of both high-R/Q

modes. As few couplers as possible should be used to avoid that the energy of the HOMs is stored mostly in the lower half of the cavity. This effect will also be taken care of in steps three and four below. Unfortunately the obvious choice of cell six can not be used, as five consecutive couplers push some modes in-between cells, i.e. an area inaccessible for the couplers. After carefully analysing Fig. 5 and confirming with only a few full model simulations, overall optimal results are found by adding six additional couplers drawn in red in Fig. 1. Table 2 shows characteristics of the three modes of concern for this case of additional couplers as obtained from Eigenmode simulations. Comparing the values

Table 2: Characteristics of the Three Most Deteriorating HOMs in a 33-cell Structure for the Cases Illustrated in Fig. 1

	$f$ [MHz]	$Q$	$R/Q$ [ $\Omega$ ]	$R$ [k $\Omega$ ]	$\phi$ [rad]
additional	627.7	8450	7.4	62.5	$17\pi/33$
top-couplers	629.3	281	85	23.9	$15\pi/33$
	630.4	195	123	24.0	$14\pi/33$
end-plate coupler	627.6	2183	9.5	20.7	$17\pi/33$
	629.4	360	75	27.0	$15\pi/33$
	630.4	199	116	23.1	$14\pi/33$
pedestal coupler	627.7	2070	10.4	21.5	$17\pi/33$
	629.2	250	86	21.5	$15\pi/33$
	630.3	195	124	24.2	$14\pi/33$
VPP coupler	627.7	3642	1.5	5.5	$17\pi/33$
	629.3	262	79	20.7	$15\pi/33$
	630.3	194	123	23.9	$14\pi/33$
VPP mitigation	627.7	3683	1.5	5.5	$17\pi/33$
	628.2	271	55	14.9	$15\pi/33$
	630.0	233	94	21.9	$14\pi/33$

with those given in [3] for today's damping scheme, we draw two conclusions. First, the  $Q$  of the  $17\pi/33$ -mode is vastly increased and second, the impedance of the two high-R/Q modes is essentially halved despite the fact that their quality factors are already quite low in today's configuration. The wakefield result for this case is shown in Fig. 4b.

## STEP 2: HOM-COUPLER IMPROVEMENT

The existing 630 MHz coupler was deployed in the first year of SPS operation (1976) and shows very good damping performance [3]. If possible, the coupler must however be further optimized to achieve the required additional damping. The fundamental coupler theory is therefore briefly reviewed on a single-cell cavity in the following. In the circuit of Fig. 6 the cavity is modelled by the parallel  $RLC$  resonator and the electric pick-up by the capacitance  $C_e$ . The parasitic stray capacitance of the probe-tip due to fringe fields is modelled by the capacitance  $C_s$  to ground. For a purely resistive load it can easily be shown that the maximum obtainable damping is limited by the stray capacitance  $C_s$ , as it partially shunts the displacement current picked up by the probe [9]. This can

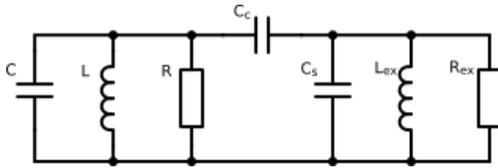


Figure 6: Electric coupling of a resistive ( $L_{ex} = 0$ ) or complex load to a single-cell cavity.

be circumvented at a single frequency by compensating the stray capacitance with an inductance  $L_{ex}$ , thereby forming a second resonant circuit. As cavity and coupler now behave as two electrically coupled resonant circuits, the achievable damping is limited by the load  $R_{ex} = R_{crit}$  for which critical coupling occurs [10]. For  $R_{ex} < R_{crit}$  the cavity resonance is split in two impedance peaks.

Due to the dimensions of the electric probe of the existing HOM-coupler with a length of  $l \approx 105 \text{ mm} \approx \lambda/4$  the probe and therefore the coupler are already resonant around 630 MHz and above described resonant effects occur even without the deployment of a complex load. Consequently, a very good performance of the coaxial coupler can be achieved even if it is merely terminated with its characteristic impedance of  $R_{ex} = 50 \Omega$ . The choice of coupler and load impedance is however not optimal as cavity and coupler are not close-to-critically coupled as shown in Table 3. Eigenmode simulations of a single cell with periodic

Table 3: Coupling to the Two High-R/Q Modes on a Single Cell (SC)

	$5\pi/11$		$14\pi/33$	
	f [MHz]	R [ $\Omega$ ]	f [MHz]	R [ $\Omega$ ]
SC	628.9	340k	629.6	330k
SC+50 $\Omega$ -coupler	628.7	956	629.4	1074
SC+23 $\Omega$ -coupler	624.1	270	626.0	295
	633.1	556	635.6	214

boundaries and loaded by one HOM-coupler show under-critically coupled behaviour for an impedance of  $50 \Omega$  (as the frequency is merely slightly perturbed by the presence of the coupler). An impedance of  $23 \Omega$  however leads to the splitting of the cavity mode into two impedance peaks which are roughly leveled in amplitude for the  $14\pi/33$ -mode. The impedance of both high-R/Q modes is reduced and spread over a wider frequency range, see Table 3.

Due to the large number of couplers deployed today (72) and foreseen for the upgrade ( $\approx 120$ ) it is actually desirable to avoid a change of the coupler geometry. It is therefore investigated if improved damping can be achieved by deploying complex loads on the coupler output ports which are transformed by its  $50 \Omega$  transmission line to  $23 \Omega$  at the base of the probe. This implementation however will have a lower bandwidth than couplers with a  $23 \Omega$  characteristic impedance. A previous test consisted of equipping the

output ports of the couplers with complex conjugate matching networks in simulation for maximum power transfer of the HOMs to the loads. As the input impedances looking into the load ports of the couplers at the frequencies of the two high-R/Q HOMs lie clustered around  $(85-j35) \Omega$  on the Smith chart, the same matching network was applied on all couplers for simplicity. Note that this matching will in any case not be exact since cross-talk between the different couplers is not yet considered. The optimum is however considered to be flat and this rough test already resulted in a significant improvement of damping, see Fig. 4b. The effect that the impedance is increased below 627 MHz is likely due to the more narrowband behaviour of the complex loads. Compared to the  $23 \Omega$  suggested by the critical-coupling approach the impedance seen by the HOMs at the base of each probe is the complex load transformed by the 13 cm long transmission line of the coupler, which results in  $(24-j5) \Omega$ . So for this specific case complex conjugate matching and critical-coupling result roughly in the same complex load.

The final design and implementation of the complex loads are still under investigation and their verification by RF measurements is scheduled. It seems however unlikely that the future damping requirement can be met with additional and improved couplers alone. This is due to the high R/Qs of the HOMs involved but also due to the loss of top/bottom symmetry in the EM field patterns - problems which will be addressed in the next two sections.

### STEP 3: $17\pi/33$ -MODE MITIGATION

The  $17\pi/33$  mode is not being pushed into the lower part of the structure to such extreme extents with the FPCs in place. As the final redesign of the FPCs is not yet known, this mode is nevertheless used as an extreme case to study possibilities of damping modes that have most of their energy stored in the lower half of the structure.

#### End-plate Coupler

Only two access ports are available in the lower half of the cavity and they are situated on the end-plates, Fig. 1. These ports are already in use nowadays for damping of transverse impedance modes. As constraints on transverse impedance are less severe, one might however consider to use these ports for longitudinal damping instead. Taking again Fig. 5 into account, the EM field amplitude of the  $17\pi/33$ -mode is considerable in the end-cells. However, its electric field profile shown in Fig. 7a does not favour electric coupling. A probe reaching deep inside the end-cell close to the drift-tube appears infeasible since also the FPB features a strong electric field in this location. A probe shape as shown in Fig. 7b had therefore to be developed that is able to pull the field into the coaxial coupler. The existing filter part of the 630 MHz HOM-coupler was used during simulations due to its already good performance. Optimization of the probe dimensions is computationally expensive as due to the position of the coupler on the end-plates no single-cell approach can be used. Instead, at least half a 33-cell structure

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

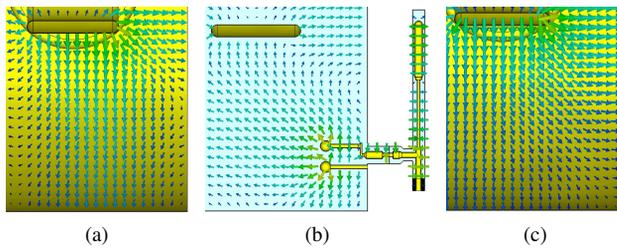


Figure 7: Electric field at end-plates of the  $17\pi/33$ -HOM (a,b) without and with coupler and of the FPB (c).

needs to be simulated by use of an H-symmetry plane in cell 17 to obtain the  $17\pi/33$ -mode. Table 2 exhibits the damping performance of the end-plate couplers on the high-Q mode and the slight detrimental effects on the  $15\pi/33$ -mode when used together with the additional top-couplers.

The end-plate coupler was found to induce a significant frequency shift of almost +100 kHz onto the FPB in standing mode. The reason for this is that the electric field of the FPB favours coupling to it from the end-plate ports, Fig. 7c. While the accelerating structure might be broadband enough to allow such a frequency shift, the notch filter of the coupler would need to handle an undesirable amount of power from the FPB.

### Pedestal Coupler

The electric field of the  $17\pi/33$ -mode is maximum in cell 17 as shown in Fig. 8a on the cross-section of this cell (cf. also Fig. 5). Another option to couple to this mode is in the areas surrounding the pedestals and an effective probe shape is shown in Fig. 8b. The bend vastly increases its coupling efficiency as the electric field is then orthogonally oriented on the probe tip. As observed in Fig. 8c the field profile of the FPB is not at all in favour of coupling to this probe shape and indeed no influence was observed in simulations. Deployment of this coupler would require

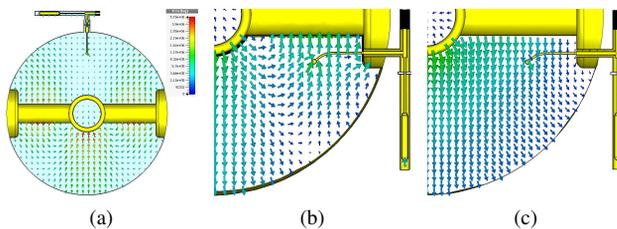


Figure 8: Electric field in cell 17 of (a)  $17\pi/33$ -mode (b) with pedestal coupler and (c) of accelerating mode.

machining or manufacturing a new drift-tube assembly per pedestal coupler (four 33-cell structures will be used after the LIU). Table 2 highlights the damping performance of a single pedestal coupler in cell 17 as a supplement to the existing and additional top-port couplers. The coupler can be conveniently optimized in simulations with an infinite periodic single-cell approach when the top-port coupler is also incorporated to create the  $17\pi/33$ -mode in the bottom of the cell.

### VPP Coupler

Three vacuum pumping ports (VPPs) are available underneath each 11-cell section (cf. Fig. 2) featuring a honeycomb grid in the inner copper layer with a hole diameter of 18.5 mm as shown in Fig. 9b. Not all pumping ports

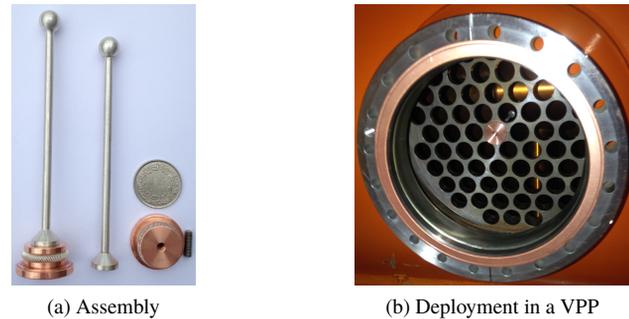


Figure 9: HOM-coupler pick-ups used as preliminary resonant posts.

are in use today nor is this so far planned for the future. Therefore, unused VPPs could be used as access ports for HOM-couplers in the lower half of the cavity. The use of one HOM-coupler in the centre cell that features strong electric field of the  $17\pi/33$ -mode is highlighted in Fig. 1 and its good performance, when used together with the additional top-port couplers, can be verified in Table 2. The coupler in the bottom of the cavity is highly effective for the mitigation of the  $17\pi/33$ -mode as not only its Q is damped but also its R/Q is vastly decreased, cf. Table 2. One can interpret that the electric field is pushed back up towards the top-couplers and the EM fields of the  $17\pi/33$ -mode thereby become much less favourable for storing energy. If necessary, it would be feasible to re-machine the copper grid so that it can host the HOM-coupler as the transition from air to vacuum is done at the level of the VPP flanges. Obviously, better results could be achieved by using more than one HOM-coupler in the lower half of the cavity. Regarding the influence of this mitigation option on the FPB, one needs to observe the overall induced frequency shift due to the total number of couplers placed inside the cavity.

### STEP 4: CHANGE OF STRUCTURE

The observations made with the VPP coupler immediately led to the question if sufficient mitigation of the  $17\pi/33$ -mode could be achieved by primarily targeting its R/Q. This requires a change of the TWS that mitigates the HOM but leaves the FPB untouched. Such a method would also be very valuable in the mitigation of the two high-R/Q modes as conventional methods of damping their Qs come to their limits. The probe of the 630 MHz coupler is already known to be resonant in the targeted frequency range and placing it directly in the honeycomb grid with good RF contact to the outer envelope as shown in Fig. 9 introduces a third resonator in the cell of the VPP. The effect of such a resonant post is most easily demonstrated by loading a long piece of the TWS envelope with a single drift-tube and stem assem-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Figure 10a shows the typical HOM field profile with strong axial electric field of the stem resonator which is not supported by the bare waveguide. With the resonant post in

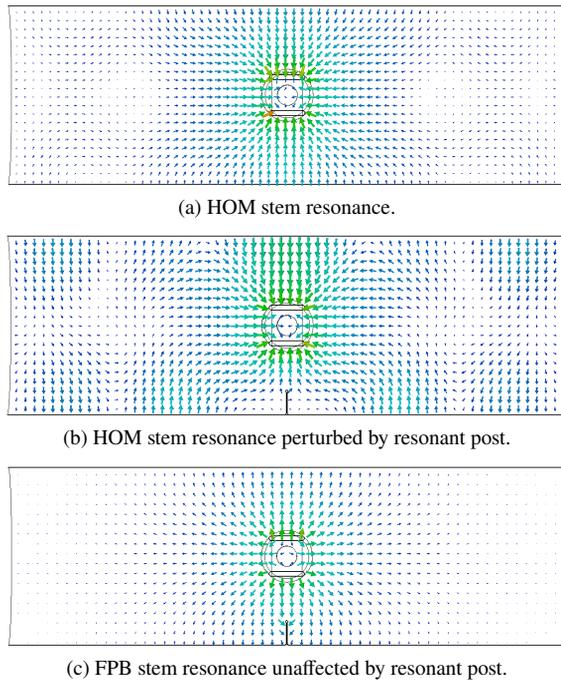


Figure 10: Electric field profiles in a long piece of circular envelope loaded by a single drift-tube and stem assembly (field magnitudes colour coded on a logarithmic scale).

place, Fig. 10b, energy can not be stored in the lower part of the cell. This will avoid that EM fields, e.g. those of the  $17\pi/33$ -mode, are pushed into the lower half of the structure making them inaccessible to the HOM-couplers. Thereby, the resonant posts act as a counter weight to the numerous couplers on top of the structure making these even more effective. In addition, the R/Q of the stem resonance calculated over the length of the cell is reduced from  $18\ \Omega$  for the unperturbed case to  $13.5\ \Omega$  for the case with the resonant post in place (a reduction by 25%). The effects described above can be explained by the fact that the displacement current picked up by the resonant post is shorted to the outer envelope of the cavity. For the case of multi-cell structures also the coupling mechanism to the neighbouring cells is partially suppressed in this manner.

The resonant posts are merely effective in cells with strong electric field and VPPs are only available in cells 3, 6 and 9 of each 11-cell section, Fig. 2. Therefore, Fig. 5 is again consulted to appropriately choose the cells or VPPs respectively on a 3-section TWS for deployment of resonant posts, with the goal to mitigate the three problematic HOMs. The impact of two resonant posts in cells 9 and 25 as marked in Fig. 1 on these modes is shown in the last row of Table 2. Not only is the  $17\pi/33$ -mode heavily mitigated but also the R/Q of the two other HOMs is significantly reduced. The achieved damping for the combination of additional and improved couplers and three resonant posts in cells 9, 17

and 25 as calculated by the CST and T3P solvers is shown in Fig. 4a. The  $14\pi/33$ -mode at 630 MHz is now heavily damped in this configuration whereas part of the impedance is shifted towards lower frequencies. It is conceivable that the complex load deployed on the HOM-coupler outputs is quite narrowband and that the couplers are therefore less effective at 626 MHz than in the range 628-630 MHz. In the case that the  $23\ \Omega$  characteristic impedance setting is used, this problem will not occur. As already mentioned, the complex load will be investigated further to achieve a more uniform damping over the 630 MHz frequency range without having to change the HOM-couplers themselves. It can however in general be noted that the HOM impedance is not only damped but also distributed to lower frequencies (this is not the case towards higher frequencies) by the new mitigation techniques, Fig. 4b.

The field profile of the accelerating FPB under the influence of a resonant post remains basically unchanged, see Fig. 10c. A frequency shift on the order of 100 kHz has however to be expected (the 200 MHz TWS is considered very broadband). It is planned to measure the frequency shift of the travelling wave FPB and the degradation of field flatness due to resonant posts in three cells on a 3-section structure in the lab as soon as it is available. It has to be evaluated if a potentially slight reduction in accelerating voltage can be accepted for a significantly improved HOM damping. First studies for optimizing the shape of the resonant electric pick-ups for their new purpose suggest that also slightly shorter posts with possibly even less impact on the FPB might be effective enough to mitigate the HOMs. The above described effects of the resonant posts on the HOMs could already be confirmed by RF lab measurements on a single section [6].

## CONCLUSION

This work showed a systematic, step-by-step improvement of today's damping scheme for longitudinal beam coupling impedance employed on the SPS 200 MHz TWSs. The complex loads for the HOM-couplers will undergo further optimization and can not only be employed on the 33- but also the 44-cell accelerating structures. The HOM-coupler placed in VPPs and the pedestal-coupler both showed good performance. Best HOM mitigation is achieved by the placement of resonant posts in selected cells, which is cost effective and requires no machining of cavity parts. This solution could even be employed without compromising the original purpose of the VPPs as the installation of a whole HOM-coupler with filter geometry is not required. The impact on the travelling-wave FPB and the accelerating voltage of all couplers, but especially the resonant posts, has however to be studied in further detail. Final optimization of damping performance will be done once one of the options is chosen for implementation.

## ACKNOWLEDGEMENTS

The authors thank F. Caspers, A. Farricker and B. Popovic from BE-RF for fruitful discussions and their input. Furthermore, this work is sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (BMBF) and used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

## REFERENCES

- [1] E. Shaposhnikova *et al.*, “Removing Known SPS Intensity Limitations for High Luminosity LHC Goals,” in *Proc. IPAC2016*, Busan, Korea, May 2016, pp. 989–991, 2016. doi:10.18429/JACoW-IPAC2016-MOP0Y058
- [2] G. Dôme, “The SPS acceleration system travelling wave drift-tube structure for the CERN SPS,” no. CERN-SPS-ARF-77-11, p. 26 p, May 1977. <https://cds.cern.ch/record/319440>
- [3] P. Kramer and C. Vollinger, “Proceedings of the ICFA Mini-Workshop on Impedances and Beam Instabilities,” in *CERN Yellow Reports: Conference Proceedings*, ser. CERN-2018-003-CP, V. Brancolini, G. Rumolo and S. Petracca, Ed., vol. 1/2018. Benevento, Italy: CERN, Geneva, 2017, pp. 57–62. doi:10.23732/CYRCP-2018-001
- [4] CST AG, Darmstadt, Germany, <http://www.cst.com/>
- [5] ACE3P, [https://portal.slac.stanford.edu/sites/ard\\_public/acd/](https://portal.slac.stanford.edu/sites/ard_public/acd/)
- [6] P. Kramer, Ph.D. dissertation, to be published.
- [7] A. Candel, A. Kabel, L. Lee, Z. Li, G. Ng, G. Schussman, and K. Ko, “State of the art in electromagnetic modeling for the Compact Linear Collider,” in *Proc. J. Phys.*, vol. 180, pp. 012004, 2009. doi:10.1088/1742-6596/180/1/012004
- [8] H. Padamsee, T. Hays, and J. Knobloch, *RF Superconductivity for Accelerators*, 2nd ed. Wiley, 2008.
- [9] E. Haebel, “Couplers for cavities,” in *CAS, CERN Accelerator School, Superconductivity in Particle Accelerators*, no. CERN 96-03, 1996, pp. 231–264.
- [10] W. Schminke, “Strong damping of single-cell accelerator cavities by resonant loops,” CERN, Geneva, Tech. Rep. SPS/ARF/WS/gs/81-68, 1981.

# SYMPLECTIC AND SELF-CONSISTENT ALGORITHMS FOR PARTICLE ACCELERATOR SIMULATION

Thomas Planche\*, Paul M. Jung, TRIUMF, Vancouver, Canada

## Abstract

This paper is a review of algorithms, applicable to particle accelerator simulation, which share the following two characteristics: (1) they preserve to machine precision the symplectic geometry of the particle dynamics, and (2) they track the evolution of the self-field consistently with the evolution of the charge distribution. This review includes, but is not limited to, algorithms using a particle-in-cell discretization scheme. At the end of this review we discuss to possibility to derived algorithms from an electrostatic Hamiltonian.

## INTRODUCTION

The conventional approach to simulating charged particle dynamics is to start from equations of motion, such as Newton's law and Poisson's equation, and solve them approximately using some standard ordinary or partial differential equation solver. The truncation errors often lead to non-physical artifacts, such as the non-conservation of phase space volume, or the violation of conservation laws resulting from symmetries of the system (Noether's theorem). By contrast, symplectic integrators produce exactly stationary solutions to an approximate action. Solutions are exact to machine precision. Approximations are made up front, when choosing the approximate Lagrangian or Hamiltonian, and the corresponding approximate discrete action. Once the physical description of the system is chosen, there is no more arbitrariness in the arcane of the algorithm.

In accelerator physics, symplectic integrators are primarily used to study long-term stability of orbits in storage rings [1]. But their properties, the first of which is the lack of arbitrariness after the choice of physics, make them desirable for the study of all conservative processes in particle accelerators.

Self-consistent algorithms are, on the other hand, essential to study betatron resonances, and by extension dynamical aperture, in the presence of space-charge forces [2].

## NOTATIONS

Throughout this paper a bold character always denotes a vector, a vector field, or a matrix. As is the convention in classical field theory, we use a dot to denote a *partial* derivative with respect to time  $t$ . All formulas are given in SI units, and we use  $c$ ,  $\epsilon_0$ , and  $\mu_0$  to denote respectively: the speed of light, the vacuum permittivity, and permeability.

\* tplanche@triumf.ca

## FROM A SINGLE-PARTICLE HAMILTONIAN

A first class of symplectic and self-consistent algorithms may be derived from a single-particle Hamiltonian. For the sake of demonstration we consider a set of (macro-)particles whose dynamics is governed by the following Hamiltonian:

$$H(\mathbf{x}, \mathbf{P}; t) = \frac{\mathbf{P}^2}{2m} + q\phi(\mathbf{x}) + q\psi(\mathbf{x}), \quad (1)$$

where  $m$  is the mass of the particle,  $q$  is its charge,  $\mathbf{x}$  and  $\mathbf{P}$  are its coordinate vector and associated canonically conjugated momentum. The space-charge force derives from the self-potential  $\phi$ . The external focusing forces derive from the scalar potential  $\psi$ . Since this Hamiltonian has no explicit dependence on the independent variable  $t$ , and is the sum of terms depending on either position or momentum alone, the particle motion can be numerically integrated using a concatenation of jolt maps [3]:

$$\left( I - \frac{\Delta t}{2} : \frac{\mathbf{P}^2}{2m} : \right) \left( I - \Delta t q : \phi + \psi : \right) \left( I - \frac{\Delta t}{2} : \frac{\mathbf{P}^2}{2m} : \right). \quad (2)$$

This approximate map, accurate to second order in  $\Delta t$ , is symplectic if  $\phi$  and  $\psi$  are functions of class  $C^2$ . This is shown by proving that the Jacobian matrix of the map is symplectic [4]. Higher order integrators may be derived from this second order one using Yoshida's method [5].

With this approach the numerical method for solving the equation of motion for the self-potential – namely Poisson's equation – is not obtained from a least action principle. This leads to a certain level of arbitrariness in the way the self-potential is to be computed.

## FROM A DISCRETIZED LAGRANGIAN

Let us now consider methods based on variational integrators derived from a Lagrangian. We will see that with these methods all the dynamics – the evolution of the particle distribution as well as the evolution of the self-field – are obtained from Hamilton's principle of stationary action.

### Low's Lagrangian

To illustrate this approach we start from the Lagrangian for non-relativistic collisionless plasma proposed by Low [6]. In the electrostatic limit, where the self-field derives solely from a scalar potential, it writes:

$$L(\mathbf{x}, \dot{\mathbf{x}}, \phi; t) = \int f(\mathbf{x}_0, \dot{\mathbf{x}}_0) L_P(\mathbf{x}(\mathbf{x}_0, \dot{\mathbf{x}}_0, t), \dot{\mathbf{x}}(\mathbf{x}_0, \dot{\mathbf{x}}_0, t); t) d\mathbf{x}_0 d\dot{\mathbf{x}}_0 \quad (3) + \frac{\epsilon_0}{2} \int |\nabla \phi(\bar{\mathbf{x}}, t)|^2 d\bar{\mathbf{x}},$$

where  $L_P$  has the form of a single-particle Lagrangian:

$$L_P(\mathbf{x}, \dot{\mathbf{x}}; t) = \frac{m}{2} |\dot{\mathbf{x}}|^2 - q\phi(\mathbf{x}, t). \quad (4)$$

$\mathbf{x}$  and  $\dot{\mathbf{x}}$  are vector fields that map the initial coordinates  $(\mathbf{x}_0, \dot{\mathbf{x}}_0)$  to the corresponding coordinates at time  $t$ .  $f$  is the initial plasma density function.  $\bar{\mathbf{x}}$  is a dummy integration variable.

Note that the single particle Lagrangian can be made more general by adding external fields term, and by replacing the non-relativistic kinetic energy term by  $-m\sqrt{1 - |\dot{\mathbf{x}}|^2/c^2}$ . For the sake of simplicity, and without much loss of generality, we choose to put aside these refinements.

### Discretized Lagrangian

Let's now discretize our system, i.e. approximate the continuous system by one with a finite number of degrees of freedom. The choice of the discretization scheme, for both the phase-space distribution  $f$  and real-space potential  $\phi$ , is arbitrary. For the sake of illustration we choose the following particular particle-in-cell (PIC) scheme:

$$f(\mathbf{x}_0, \dot{\mathbf{x}}_0) = \sum_i w^i \delta(\mathbf{x}_0 - \mathbf{x}_0^i) \delta(\dot{\mathbf{x}}_0 - \dot{\mathbf{x}}_0^i), \quad (5)$$

$$\phi(\mathbf{x}, t) = \sum_j \phi^j(t) K(\mathbf{x} - \mathbf{x}^j), \quad (6)$$

where  $w^i$  is the weight of the  $i^{\text{th}}$  macro-particle,  $\mathbf{x}_0^i$  and  $\dot{\mathbf{x}}_0^i$  are its initial coordinates.  $\mathbf{x}^j$  is the position of  $j^{\text{th}}$  node of the PIC grid, and  $\phi^j(t)$  is the potential assigned to this node.  $\delta$  is the Dirac function.  $K$  is an interpolation kernel function of class  $C^2$  which satisfies the requirement of norming [7]:

$$\sum_j K(\mathbf{x} - \mathbf{x}^j) = 1, \quad (7)$$

for all  $\mathbf{x}$ . The choice of the kernel function is arbitrary. It is usually chosen among positively defined even functions. A noticeable example of kernel function is the Gaussian wavelet used in COSY INFINITY (see section on 'General Particle Optical Elements' in [8]). Suitable kernel functions may also be constructed out of piecewise polynomials [9]. Note that this discretization scheme is similar, although not identical, to the one used in Ref. [10].

Combining Eqs. (3) and (5) leads to the discretized Lagrangian,  $L_D(\mathbf{x}, \dot{\mathbf{x}}, \phi; t) =$

$$\begin{aligned} & \frac{m}{2} \sum_i w^i |\dot{\mathbf{x}}^i(t)|^2 - q \sum_i \sum_j \phi^j w^i K(\mathbf{x}^i(t) - \mathbf{x}^j) \\ & + \frac{\epsilon_0}{2} \int \left( \sum_j \phi^j \nabla K(\bar{\mathbf{x}} - \mathbf{x}^j) \right)^2 d\bar{\mathbf{x}}, \end{aligned} \quad (8)$$

where  $\mathbf{x}^i(t) = \mathbf{x}(\mathbf{x}_0^i, \dot{\mathbf{x}}_0^i, t)$ .

### Discretized Action and Equations of Motion

The action  $S = \int L_D dt$  can be approximated to first order using a Riemann sum:

$$S \approx S_D = \Delta t \sum_n L_D(\mathbf{x}_n, \frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\Delta t}, \phi_n; t), \quad (9)$$

where the subscript  $n$  denotes an evaluation at time  $t = n \Delta t$ . Minimization of the action follows from the discrete Euler-Lagrange equation (see section 1.1.2 of [11]) which leads to the following equations of motion:

$$m \frac{\mathbf{x}_{n+1}^i - 2\mathbf{x}_n^i + \mathbf{x}_{n-1}^i}{\Delta t} = -q \sum_j \phi_n^j \nabla K(\mathbf{x}_n^i - \mathbf{x}^j), \quad (10)$$

$$\sum_k \phi_n^k \mathcal{M}^{jk} = -\frac{q}{\epsilon_0} \rho_n^j, \quad (11)$$

where:

$$\mathcal{M}^{jk} = \int K(\bar{\mathbf{x}} - \mathbf{x}^j) \nabla^2 K(\bar{\mathbf{x}} - \mathbf{x}^k) d\bar{\mathbf{x}}, \quad (12)$$

and

$$\rho_n^j = \sum_i w^i K(\mathbf{x}_n^i - \mathbf{x}^j). \quad (13)$$

Equation 10 is a discrete form of Newton's equation with the Lorentz force. To solve it numerically one may split this second-order equation into two first order equations [12]. Equation 11 is obtained after integrating by parts the outcome from the Euler-Lagrange equation, and dropping the boundary term. It is a discrete form of Poisson's equation. It defines a linear relation between all  $\phi_n^j$  and  $\rho_n^j$  and can be solved by inverting the square matrix  $\mathcal{M}$ .

Equations 10 and 11 constitute a complete numerical integration scheme. Numerical integration leads to an exact (to machine precision) solution of an approximate action, which makes it a variational integrator. Variational integrators are symplectic integrators [11]. As a matter of fact this particular one is a symplectic Euler.

Higher-order variational integrators can be obtained using higher-order approximations of  $S$ . A second order variational integrator using a spectral discretization of  $\phi$  has been tested in one, two, and three-dimensional, and has demonstrated excellent long-term stability [12].

## FROM A DISCRETIZED HAMILTONIAN

Symplectic integrators are more commonly obtained from a Hamiltonian [1]. Unfortunately the electrostatic Low Lagrangian in Eq. (3) is degenerate: it contains no explicit dependence on  $\phi$ . This makes the application of a Legendre transformation to this Lagrangian, if not impossible, at least beyond the abilities of the authors.

In this section we will discuss two ways to overcome this issue and obtain a Hamiltonian from different versions of Low's Lagrangian.

### Electromagnetic Hamiltonian

Let us choose to use the Weyl gauge (also referred to as temporal gauge) and set  $\phi = 0$ . Low's original Lagrangian

becomes  $L(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{A}, \dot{\mathbf{A}}; t) =$

$$\int f(\mathbf{x}_0, \dot{\mathbf{x}}_0) \left( \frac{m}{2} |\dot{\mathbf{x}}(\mathbf{x}_0, \dot{\mathbf{x}}_0, t)|^2 + q \dot{\mathbf{x}} \cdot \mathbf{A}(\mathbf{x}, t) \right) d\mathbf{x}_0 d\dot{\mathbf{x}}_0 + \frac{\epsilon_0}{2} \int |\dot{\mathbf{A}}(\bar{\mathbf{x}}, t)|^2 - |\mathbf{c} \nabla \times \mathbf{A}(\bar{\mathbf{x}}, t)|^2 d\bar{\mathbf{x}}. \quad (14)$$

### Discretization and Legendre Transformation

Let's use a PIC discretization scheme identical to Eq. (5), only replacing  $\phi$  by  $\mathbf{A}$ . For compactness we write the discretized Lagrangian in matrix form:

$$L_D = \frac{m}{2} \dot{\mathbf{x}}^\top \mathbf{w} \dot{\mathbf{x}} + q \dot{\mathbf{x}}^\top \mathbf{w} \mathbf{K} \mathbf{A} + \frac{\epsilon_0}{2} \dot{\mathbf{A}}^\top \mathcal{K} \dot{\mathbf{A}} - \frac{1}{2\mu_0} \mathbf{A}^\top \mathcal{K}_\times \mathbf{A} \quad (15)$$

where  $\mathbf{x}$  and  $\mathbf{A}$  are now vectors with components  $x^i$  and  $A^j$  respectively;  $\mathbf{x}$  contains  $N_p$  elements (the number of macroparticles), and  $\mathbf{A}$  contains  $N_g$  elements (the number of grid nodes).  $\mathbf{w}$  is a diagonal matrix with components  $w^i$ . The components of the other matrices are:

$$K^{ij} = K(\mathbf{x}^i(t) - \mathbf{x}^j) \quad (16)$$

$$\mathcal{K}^{jk} = \int K(\bar{\mathbf{x}} - \mathbf{x}^j) K(\bar{\mathbf{x}} - \mathbf{x}^k) d\bar{\mathbf{x}} \quad (17)$$

$$\mathcal{K}_\times^{jk} = \int [\nabla K(\bar{\mathbf{x}} - \mathbf{x}^j)]^\top_x [\nabla K(\bar{\mathbf{x}} - \mathbf{x}^k)]_x d\bar{\mathbf{x}}, \quad (18)$$

where  $i$  and  $l$  go from 1 to  $N_p$ , while  $j$  and  $k$  go from 1 to  $N_g$ . Superscript  $\top$  refers to the transpose operation.  $[\ ]_\times$  denotes a skew matrix used to express the cross product as a matrix multiplication ( $[\mathbf{a}]_\times \mathbf{b} = \mathbf{a} \times \mathbf{b}$ ).

The Legendre transformation writes:

$$H_D = \dot{\mathbf{x}}^\top \mathbf{P} + \dot{\mathbf{A}}^\top \mathbf{Y} - L_D, \quad (19)$$

The components of the canonical momentum vectors are  $\mathbf{P}^i = \frac{\partial L_D}{\partial \dot{x}^i}$  and  $\mathbf{Y}^j = \frac{\partial L_D}{\partial \dot{A}^j}$ , which are explicitly as:

$$\mathbf{P} = m\mathbf{w}\dot{\mathbf{x}} + q\mathbf{w}\mathbf{K}\mathbf{A} \quad (20)$$

$$\mathbf{Y} = \epsilon_0 \mathcal{K} \dot{\mathbf{A}}. \quad (21)$$

The discretized Hamiltonian becomes:

$$H_D = \frac{1}{2m} (\mathbf{P} - q\mathbf{w}\mathbf{K}\mathbf{A})^\top \mathbf{w}^{-1} (\mathbf{P} - q\mathbf{w}\mathbf{K}\mathbf{A}) + \frac{1}{2\epsilon_0} \mathbf{Y}^\top \mathcal{K}^{-1} \mathbf{Y} + \frac{1}{2\mu_0} \mathbf{A}^\top \mathcal{K}_\times \mathbf{A}, \quad (22)$$

and the associated canonical Poisson bracket writes  $\{F, G\} =$

$$\sum_i \frac{\partial F}{\partial x^i} \frac{\partial G}{\partial P^i} - \frac{\partial F}{\partial P^i} \frac{\partial G}{\partial x^i} + \sum_j \frac{\partial F}{\partial A^j} \frac{\partial G}{\partial Y^j} - \frac{\partial F}{\partial Y^j} \frac{\partial G}{\partial A^j}. \quad (23)$$

Since  $H_D$  is the sum of exactly solvable terms (for an explicit solution of the  $(\mathbf{P} - q\mathbf{A})^2$ -like term, see [13]), one can

build a second order symplectic integrator by concatenating maps [3].

A similar symplectic integrator derived from the Morrison-Marsden-Weinstein electromagnetic Hamiltonian [14–16] has been tested [9]. Note that a corresponding variational integrator had previously been tested [17].

### Electrostatic Hamiltonian

In most accelerator physics problems particles do not move at relativistic speeds with respect to each other. In such a case a scalar potential is sufficient to describe the self-field. Keeping track of the three components of a vector potential is wasteful.

We have already discussed the fact that an electrostatic Hamiltonian cannot be obtained from Eq. (3). In this section we show that it is however possible to obtain an electrostatic Hamiltonian after changing the independent variable.

The action  $S$  associated with Eq. (3) writes  $S =$

$$\iint f \left( \frac{m}{2} |\dot{\mathbf{x}}|^2 - q\phi(\mathbf{x}, t) \right) d\mathbf{x}_0 d\dot{\mathbf{x}}_0 dt + \frac{\epsilon_0}{2} \iint |\nabla \phi|^2 d\bar{\mathbf{x}} d\bar{t}. \quad (24)$$

We proceed in the first integral to a change of variable using the substitution function:

$$g(x_0, y_0, t_0, x'_0, y'_0, t'_0, z) = (x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0, t), \quad (25)$$

where primes ' denote partial derivative with respect to  $z$ . The determinant of the Jacobian matrix  $\det(D_g) = t'/(t_0^5)$ , where  $t'_0 = \frac{\partial t}{\partial z}|_{z=0}$ . Similarly we proceed in the second integral to the change of variable given by:

$$h(\bar{x}, \bar{y}, \bar{t}, z) = (\bar{x}, \bar{y}, \bar{z}, t), \quad (26)$$

The determinant of the Jacobian of  $h$  is 1. The action becomes  $S =$

$$\int \left[ \int \hat{f} \left( m \frac{x'^2 + y'^2 + 1}{2t'} - q t' \hat{\phi} \right) dx_0 dy_0 dt_0 dx'_0 dy'_0 dt'_0 + \frac{\epsilon_0}{2} \int |\nabla \hat{\phi}|^2 d\bar{x} d\bar{y} d\bar{t} \right] dz = \int L_z(x, y, t, \hat{\phi}, x', y', t', \hat{\phi}'; z) dz, \quad (27)$$

where:

$$\hat{f}(x_0, y_0, t_0, x'_0, y'_0, t'_0, z) = f(x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0, t)/t_0^5, \quad (28)$$

and

$$\hat{\phi}(x_0, y_0, t_0, x'_0, y'_0, t'_0, z) = \phi(x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0, t). \quad (29)$$

To simplify notations we drop the hats.

One notices that  $L_z$  depends explicitly on  $\phi'$ ,  $x'$ ,  $y'$ , and  $t'$ , enabling us to define the following canonical momentum

densities:

$$\Pi = \epsilon_0 \phi' \quad (30)$$

$$P_x = m f \frac{x'}{t'} \quad (31)$$

$$P_y = m f \frac{y'}{t'} \quad (32)$$

$$-E = -m f \frac{x'^2 + y'^2 + 1}{2t'^2} - q f \phi \quad (33)$$

which in turn enables us to perform a Legendre transformation and obtain the following continuous electrostatic Hamiltonian  $H_z =$

$$\int \sqrt{2mf(E - qf\phi) - P_x^2 - P_y^2} dx_0 dy_0 dt_0 dx'_0 dy'_0 dt'_0 + \frac{1}{2} \int \left( \frac{\Pi^2}{\epsilon_0} - \epsilon_0 |\nabla_{\perp} \phi|^2 \right) d\bar{x} d\bar{y} d\bar{t}, \quad (34)$$

where  $|\nabla_{\perp} \phi|^2 = (\partial_x \phi)^2 + (\partial_y \phi)^2$ . An attempt to implement an algorithm based on a discrete relativistic version of this Hamiltonian is presented Ref. [18].

## CONCLUSION

A variety of symplectic and self-consistent multi-particle algorithms have been developed by both the accelerator physics and the plasma physics community. They are superior to most algorithm derived from equations of motion as they guaranty that the symplectic nature of the particle dynamics is conserved to machine precision. Some of them guaranty the conservation of the symplectic nature of the self-field dynamics as well. The algorithms discussed in this paper were all derived from a collision-less picture, and as such are unable to model non-Hamiltonian processes such as intra-beam scattering.

## ACKNOWLEDGEMENT

The authors would like to thank Rick Baartman and Kyle Gao for their corrections and comments on the original manuscript.

## REFERENCES

[1] E. Forest, "Geometric integration for particle accelerators," *Journal of Physics A: Mathematical and General*, vol. 39, no. 19, pp. 5321, 2006. <http://stacks.iop.org/0305-4470/39/i=19/a=S03>

[2] R. Baartman, "Betatron resonances with space charge," in *AIP Conference Proceedings*, AIP, vol. 448, pp. 56–72, 1998.

[3] E. Forest and R. D. Ruth, "Fourth-order symplectic integration", *Physica D: Nonlinear Phenomena*, vol. 43, no. 1, pp. 105–117, 1990. doi:10.1016/0167-2789(90)90019-L

[4] J. Qiang, *Phys. Rev. Accel. Beams*, vol. 20, pp. 014203, 2017. doi:10.1103/PhysRevAccelBeams.20.014203

[5] H. Yoshida, "Construction of higher order symplectic integrators," *Physics Letters A*, vol. 150, no. 5, pp. 262–268, 1990. doi:10.1016/0375-9601(90)90092-3

[6] F. E. Low, "A lagrangian formulation of the boltzmann-vlasov equation for plasmas," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 248, no. 1253, pp. 282–287, 1958. doi:10.1098/rspa.1958.0244

[7] Y. N. Grigoryev, V. A. Vshivkov, and M. P. Fedoruk, *Numerical "Particle-in-Cell" Methods: Theory and Applications*, Walter de Gruyter, 2012.

[8] M. Berz and K. Makino, *COSY INFINITY version 8.1. user's guide and reference manual*, Department of Physics and Astronomy, MSUHEP-20704, Michigan State University.

[9] H. Qin, J. Liu, J. Xiao, R. Zhang, Y. He, Y. Wang, Y. Sun, J. W. Burby, L. Ellison, and Y. Zhou, "Canonical symplectic particle-in-cell method for long-term large-scale simulations of the vlasov-maxwell equations," *Nuclear Fusion*, vol. 56, no. 1, pp. 014001, 2016. <http://stacks.iop.org/0029-5515/56/i=1/a=014001>

[10] B. A. Shadwick, A. B. Stamm, and E. G. Evstatiev, "Variational formulation of macro-particle plasma simulation algorithms," *Physics of Plasmas*, vol. 21, no. 5, pp. 055708, 2014.

[11] J. E. Marsden and M. West, "Discrete mechanics and variational integrators," *Acta Numerica*, vol. 10, pp. 357–514, 2001. doi:10.1017/S096249290100006X

[12] S. D. Webb, "A spectral canonical electrostatic algorithm," *Plasma Physics and Controlled Fusion*, vol. 58, no. 3, pp. 034007, 2016. <http://stacks.iop.org/0741-3335/58/i=3/a=034007>

[13] Y. K. Wu, E. Forest, and D. S. Robin, "Explicit symplectic integrator for s-dependent static magnetic field," *Phys. Rev. E*, vol. 68, pp. 046502, 2003. doi:10.1103/PhysRevE.68.046502

[14] P. J. Morrison, "The maxwell-vlasov equations as a continuous hamiltonian system," *Physics Letters A*, vol. 80, no. 5, pp. 383–386, 1980. doi:10.1016/0375-9601(80)90776-8

[15] A. Weinstein and P. J. Morrison, "Comments on: The Maxwell-Vlasov equations as a continuous hamiltonian system," *Physics Letters A*, vol. 86, no. 4, pp. 235–236, 1981.

[16] J. E. Marsden and A. Weinstein, "The hamiltonian structure of the maxwell-vlasov equations," *Physica D: Nonlinear Phenomena*, vol. 4, no. 3, pp. 394–406, 1982. doi:10.1016/0167-2789(82)90043-4

[17] H. Qin and X. Guan, "Variational symplectic integrator for long-time simulations of the guiding-center motion of charged particles in general magnetic fields," *Phys. Rev. Lett.*, vol. 100, pp. 035006, 2008. doi:10.1103/PhysRevLett.100.035006

[18] P. M. Jung and T. Planche, "S-based multi-particle spectral simulation of an electron gun," presented at ICAP'18, Key West, FL, Oct 2018, paper TUPAG13, this conference.

# POLARIZED PROTON BEAMS FROM LASER-INDUCED PLASMAS

A. Hützen<sup>1</sup>, M. Büscher\*<sup>1</sup>, I. Engin

Peter Grünberg Institut (PGI-6), Forschungszentrum Jülich,  
Wilhelm-Johnen-Str. 1, 52425 Jülich, Germany

J. Thomas, A. Pukhov

Institut für Theoretische Physik I, Heinrich-Heine-Universität Düsseldorf,  
Universitätsstr. 1, 40225 Düsseldorf, Germany

J. Böker, R. Gebel, A. Lehrach<sup>2</sup>

Institut für Kernphysik (IKP-4), Forschungszentrum Jülich,  
Wilhelm-Johnen-Str. 1, 52425 Jülich, Germany

R. Engels

Institut für Kernphysik (IKP-2), Forschungszentrum Jülich,  
Wilhelm-Johnen-Str. 1, 52425 Jülich, Germany

T. P. Rakitzis<sup>3</sup>, D. Sofikitis<sup>3</sup>

Department of Physics, University of Crete, 71003 Heraklion-Crete, Greece

<sup>1</sup>also at Institut für Laser- und Plasmaphysik, Heinrich-Heine-Universität Düsseldorf,  
Universitätsstr. 1, 40225 Düsseldorf, Germany

<sup>2</sup>also at JARA-FAME und III. Physikalisches Institut B, RWTH Aachen,  
Otto-Blumenthal-Str., 52074 Aachen, Germany

<sup>3</sup>also at Institute of Electronic Structure and Laser, Foundation for Research and Technology-Hellas,  
71110 Heraklion-Crete, Greece

## Abstract

We report on the concept of an innovative laser-driven plasma accelerator for polarized proton (or deuteron) beams with a kinetic energy up to several GeV. In order to model the motion of the particle spins in the plasmas, these have been implemented as an additional degree of freedom into the Particle-in-Cell simulation code VLPL. Our first simulations for nuclear polarized Hydrogen targets show that, for typical cases, the spin directions remain invariant during the acceleration process. For the experimental realization, a polarized HCl gas-jet target is under construction where the degree of proton polarization is determined with a Lamb-shift polarimeter. The final experiments, aiming at the first observation of a polarized particle beam from laser-generated plasmas, will be carried out at the 10 PW laser system SULF at SIOM/Shanghai.

## INTRODUCTION

The field of laser-induced relativistic plasmas and, in particular, of laser-driven particle acceleration, has undergone impressive progress in recent years. Despite many advances in the understanding of fundamental physical phenomena, one unexplored issue is how the particle (in particular hadron) spins are influenced by the huge magnetic fields inherently present in the plasmas [1–4].

Several mechanisms can potentially lead to a sizeable degree of polarization of laser-accelerated particle beams: first, a genuine polarization build-up from an unpolarized target by the laser-plasma fields themselves and, second, polarization preservation of pre-aligned spins during the acceleration despite of these fields. The work of our group aims at the first scenario using a novel dynamically polarized Hydrogen target.

Two effects are currently discussed to build up a nuclear polarization in the plasma: either the polarization is generated due to a spin-flip according to the Sokolov-Ternov effect, induced by the magnetic fields of the incoming laser pulse. Besides that, the spatial separation of spin states by the magnetic-field gradient, *i.e.* the Stern-Gerlach effect, may result in the generation of polarization for different beam trajectories [5].

In addition to these two mechanisms, all particle spins precess around the laser or plasma magnetic fields as characterized by the Thomas-Bargmann-Michel-Telegdi (T-BMT) equation describing the spin motion in arbitrary electric and magnetic fields in the relativistic regime.

The first and up to now only experiment measuring the polarization of laser-accelerated protons has been performed at the ARCTurus laser facility at Heinrich-Heine University Düsseldorf [2]. Figure 1 schematically depicts the setup: for the measurements a 100-TW class Ti:Sa laser system with a typical pulse duration of 25 fs and a repetition rate of 10 Hz

\* m.buescher@fz-juelich.de

was used producing an intensity of several  $10^{20}$   $\text{Wcm}^{-2}$ . After impinging the laser pulse under a  $45^\circ$  angle on an (unpolarized) gold foil of  $3 \mu\text{m}$  thickness, protons are accelerated according to the well-known Target Normal Sheath Acceleration (TNSA) mechanism [1] to an energy of typically a few MeV. They are heading towards a stack of Radio-Chromic-Film (RCF) detectors where the flux of protons is monitored. In order to measure the polarization of the proton bunches, the spin dependence of elastic proton scattering off nuclei is employed. At the particular proton energy Silicon is a suitable scattering material since it has a sizeable analyzing power and the scattering cross sections are known.

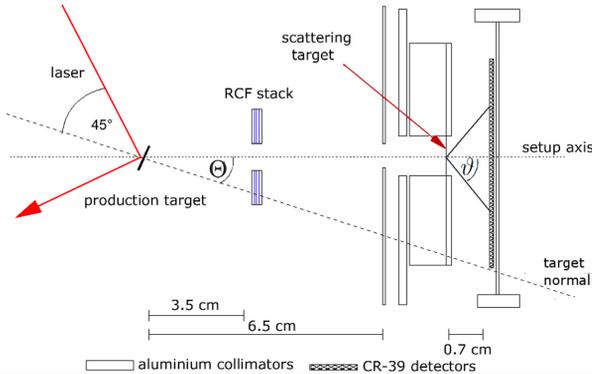


Figure 1: Schematic setup for the first proton polarization measurement [2].

To estimate the magnitude of possible polarizing magnetic fields in case of the Düsseldorf experiment, Particle-In-Cell (PIC) simulations were carried out with the fully relativistic 2D code EPOCH [6]. A  $B$ -field strength of  $\sim 10^4$  T and gradients of around  $10^{10}$   $\text{Tm}^{-1}$  are to be expected. Although these values are much higher compared to those at conventional accelerators, they are yet too small to align the proton spins and, thus, should not lead to a measurable proton polarization [2]. This prediction matches the experimental finding of zero proton polarization.

An important conclusion from this experiment thus is that for measuring a sizesable proton polarization both, a stronger laser pulse with an intensity of about  $10^{23}$   $\text{Wcm}^{-2}$  and an extended gas instead of a thin foil target would be required. Such a scenario has been theoretically considered (without taking into account spin effects) in a paper by Shen *et al.* [7]. Due to the larger target size, the interaction time between the laser accelerated protons and the  $B$ -field is increased. It amounts to approximately 3.3 ps and is much larger than the typical time scale (about 0.1 ps) for spin motion given by the Larmor frequency and, thus, a spin manipulation seems possible.

## PARTICLE SPIN DYNAMICS

We have implemented particle-spin effects into the 3D PIC simulation code VLPL (Virtual Laser Plasma Lab) in order to make theoretical predictions about the degree of proton-spin polarization from a laser-driven plasma accelerator [8,

9]. These calculations consider all relevant effects that may lead to the polarization of proton beams [10].

The Sokolov-Ternov effect is, for example, employed in classical accelerators to polarize the stored electron beams where the typical polarization build-up times are minutes or longer. This effect can, therefore, be neglected in the case of laser-induced acceleration. We refer to our forthcoming publication Ref. [10] for a more quantitative derivation.

Our assessment for the Stern-Gerlach force [10] shows that non-relativistic proton beams with opposite spins are separated by not more than  $\Delta_p \approx 9.3 \cdot 10^{-7} \lambda_L$  with the laser wavelength  $\lambda_L$ . Moreover, the field strengths is of the order of  $E \approx B \approx 10^5$  T and the field gradients  $\nabla|\mathbf{B}| \approx 10^5$  T/R with the laser radius  $R$ , typically  $\lambda_L/R = 1/10$  and a characteristic separation time would be  $t = 100 \omega_L^{-1}$ , where  $\omega_L$  is the laser frequency. Thus, the force on the given length scale is too weak and the Stern-Gerlach effect does not have to be taken into account for further simulation work on proton-spin tracking.

For charged particles the spin precession in arbitrary electric and magnetic fields is given by the T-BMT equation in CGS units:

$$\frac{ds}{dt} = -\frac{e}{m_p c} \left[ \left( a_p + \frac{1}{\gamma} \right) \mathbf{B} - \frac{a_p \gamma}{\gamma + 1} \left( \frac{\mathbf{v}}{c} \cdot \mathbf{B} \right) \frac{\mathbf{v}}{c} - \left( a_p + \frac{1}{1 + \gamma} \right) \frac{\mathbf{v}}{c} \times \mathbf{E} \right] \times \mathbf{s} = -\vec{\Omega} \times \mathbf{s} . \quad (1)$$

Here  $\mathbf{s}$  is the proton spin in the rest frame of the proton,  $e$  is the elementary charge,  $m_p$  the proton mass,  $c$  the speed of light, the dimensionless anomalous magnetic moment of the proton  $a_p = \frac{g_p - 2}{2} = 1.8$  with the  $g$ -factor of the free proton  $g_p$ ,  $\gamma$  the Lorentz factor,  $\mathbf{v}$  the particle velocity,  $\mathbf{B}$  the magnetic field, and  $\mathbf{E}$  the electric field, both in the laboratory frame. Since  $\vec{\Omega}$  always has a component perpendicular to  $\mathbf{s}$ , the single spins in a polarized particle ensemble precess with the frequency  $\omega_s = |\vec{\Omega}|$ . For protons with an energy in the range of a few GeV,  $\gamma \approx 1$  and  $1 \gtrsim \mathbf{v}/c$ , so that:

$$\omega_s < \frac{e}{m_p c} \sqrt{(a_p + 1)^2 \mathbf{B}^2 + \left( \frac{a_p}{2} \right)^2 \mathbf{B}^2 + \left( a_p + \frac{1}{2} \right)^2 \mathbf{E}^2} . \quad (2)$$

Under the assumption  $|\mathbf{B}| \approx |\mathbf{E}| \approx F$  this simplifies to:

$$\omega_s < \frac{e}{m_p c} F \sqrt{\frac{9}{4} a_p^2 + 3 a_p + \frac{5}{4}} . \quad (3)$$

As a consequence, a conservation of the polarization of the system is expected for times

$$t \ll \frac{2\pi}{\omega_s} \approx \frac{2\pi}{3.7 \frac{e}{m_p c} F} \quad (4)$$

for  $a_p = 1.8$ . For the typical field strengths in our first simulations (*cf.* Fig. 2) of  $F = 5.11 \cdot 10^{12}$   $\text{V/m} = 17.0 \cdot 10^3$  T, the preservation of the spin directions is estimated for times  $t < 1$  ps. This time is sufficiently long taking into account

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

that the simulation time is  $t_{\text{sim}} = 0.13 \text{ ps} \ll 1 \text{ ps}$ , so the polarization is maintained during the entire simulation according to the T-BMT equation.

## PARTICLE-IN-CELL SIMULATIONS

In order to reproduce the results of our experiment presented above and to verify the quantitative estimates of Ref. [2], 3D simulations with the VLPL code including spin tracking have been carried out on the supercomputer JURECA [11].

It is important to note that to simulate the plasma behavior, a PIC code first defines a three-dimensional Cartesian grid which fills the simulated volume where the plasma evolves over the simulated time. Not each physical particle is treated individually but they are substituted by so-called PIC particles. This is why the continuous spin vector of a PIC particle represents the mean spin of all substituted particles. Thus, not the spin of each single particle is simulated but the polarization  $P$  of every PIC particle. The sum of spin vectors of different PIC particles within a specific volume (polarization cell) corresponds to the local polarization of the ensemble [10, 12].

Figure 2 shows preliminary simulation results for proton-spin tracking with VLPL. An acceleration of the proton sheath due to the TNSA mechanism is evident. From the simulated strength of the magnetic field behind the target (acting on the accelerated protons) we estimate a proton polarization preservation for at least 0.18 ps. This is much longer than the time needed to accelerate the protons.

Thus, the VLPL simulations demonstrate polarization conservation according to the T-BMT equation when acceleration by the TNSA mechanism takes place [12]. In other words, a compact target would be needed in which the nuclear spins are already aligned at the time of irradiation with the accelerating laser. Unfortunately, solid foil targets suitable for laser acceleration with the TNSA mechanism are not available so far and their experimental realization would be extremely challenging. In solid targets used for classical accelerator experiments Hydrogen nuclear polarization mostly results from a static polarization, *e.g.*, in frozen spin targets [13].

The only gaseous target material with a density sufficient for laser acceleration are hyper-polarized  $^3\text{He}$  [14, 15] or Xe which are, of course, not suitable for proton acceleration. For Hydrogen until now only polarized atomic beam sources based on the Stern-Gerlach principle [16] are available, which have the disadvantage of a much too low particle density. In order to provide a (dynamically) polarized Hydrogen gas target for laser-plasma applications, a new approach is needed.

## EXPERIMENTAL REALIZATION

For the experimental realization of our new concept for a dynamically polarized ion source, three major components are required: a suitable laser system to provide the target polarization, a vacuum interaction chamber including a gas

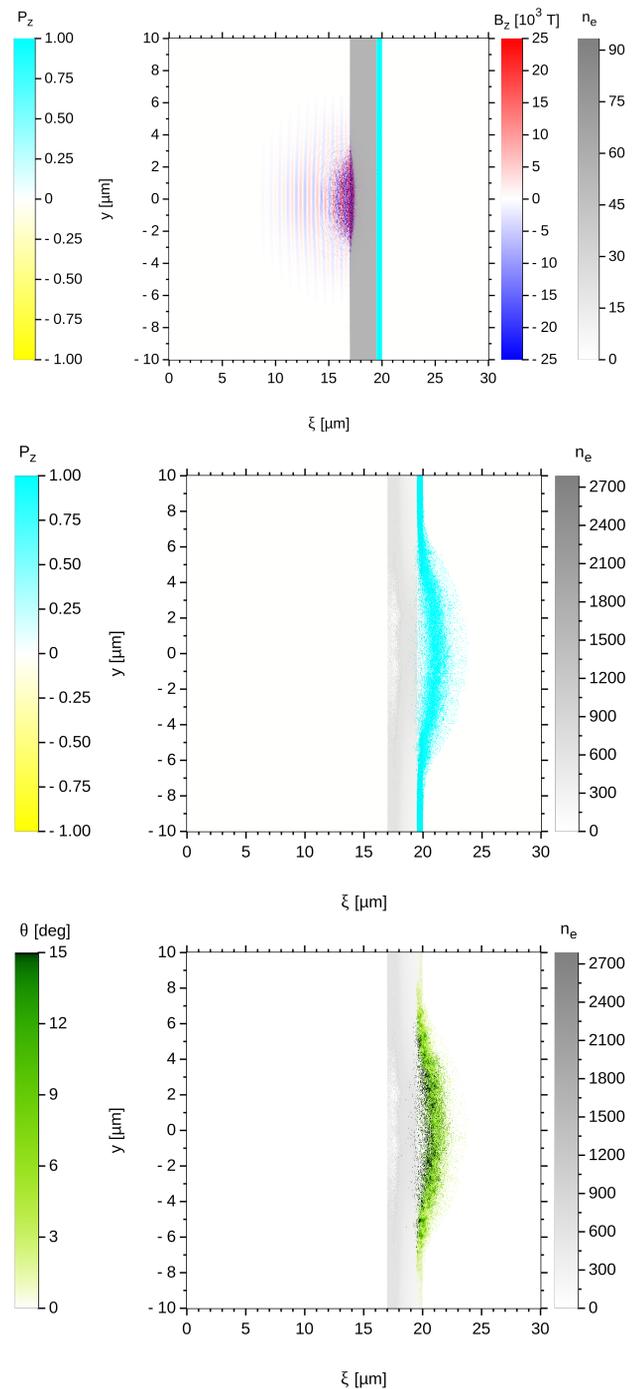


Figure 2: 3D VLPL simulation showing the conservation of proton polarization from an Aluminum foil target ( $2.5 \mu\text{m}$ ,  $35 n_{\text{cr}}$ ) covered with a fully polarized proton layer ( $0.5 \mu\text{m}$ ,  $117 n_{\text{cr}}$ ) at simulation times 32.5 fs (top) and 130 fs (center, bottom). The upper two figures show the degree of proton polarization, while in the lower their spin-rotation angle (relative to the initial value  $\sigma = (0, 0, 1)$ ) is depicted. The of Gaussian-shape laser pulse ( $\lambda_L = 800 \text{ nm}$ , normalized laser amplitude  $a_0 = 12$ , 25 fs duration,  $5 \mu\text{m}$  focal spot size) enters the simulation box from the left. The grid cell size is  $h_x = 0.025 \mu\text{m}$  and  $h_y = h_z = 0.05 \mu\text{m}$ ,  $n_e$  represents the electron density.

jet at the interaction point with the accelerating laser, and a polarimeter. The schematic view of this setup is depicted in Fig. 3.

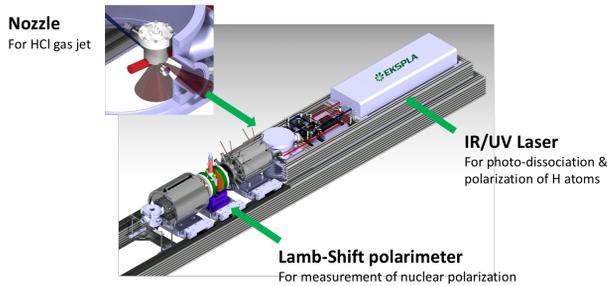


Figure 3: Schematic view of the setup for the proton polarization measurement using a polarized Hydrogen gas target.

As material of the gas target, Hydrogen halides are a viable option [17, 18]. A Hydrogen chloride (HCl) target is preferred in this case due to the rather high polarizability and the easy availability. The HCl gas is injected into the interaction chamber through a standard gas nozzle with a high-speed short-pulse piezo valve that can be operated at 5 bar inlet-gas pressure to produce a gas density in the range of  $\sim 10^{19} \text{ cm}^{-3}$ . Few millimeters below the nozzle, the interaction between gas and laser beams takes place.

The polarizing laser system is a pulsed Ni:YAG laser from EKSPLA [19]. Its peculiarity is the quasi-simultaneous output of the fundamental wavelength at 1064 nm and the fifth harmonic (213 nm). The repetition rate of the laser system is 5 Hz and the pulses are of 170 ps duration which is sufficiently short with regard to the transfer time of the electron spin polarization to the nucleus due to hyperfine interaction ( $\sim 1 \text{ ns}$ ) [17]. The linearly polarized 1064 nm beam with a pulse energy of 100 mJ is focused with an intensity of  $\sim 10^{11} \text{ Wcm}^{-2}$  into the interaction chamber to align the HCl bonds (*cf.* Fig. 4). By this, the signal intensity is increased by an enhancement factor  $x \approx 2$  assuming an interaction parameter of  $\Delta\omega = 49$  and, thus,  $\langle\langle \cos^2\theta \rangle\rangle = 6/7$  since the polarizability interaction is governed by a  $\cos^2\theta$  potential with the angle  $\theta$  between the molecular axis and the electric field distribution [20].

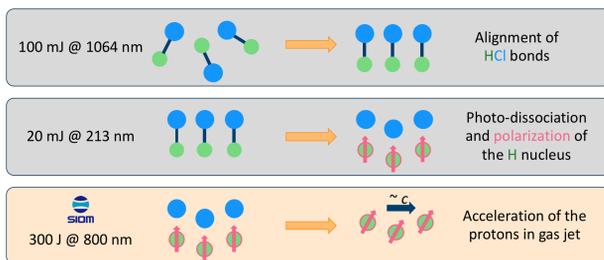


Figure 4: Schematic overview of the production of polarized proton beams.

At the same time, but under a  $90^\circ$  angle, the circularly polarized fifth harmonic with an energy of 20 mJ is focused into the vacuum chamber to an intensity of  $\sim 10^{12} \text{ Wcm}^{-2}$  to interact with the aligned HCl molecules. These are subsequently photo-dissociated by UV excitation via the  $A^1\Pi_1$  state, which has a total electronic angular-momentum projection of  $\Omega = +1$  along the bond axis. Hence, the resulting H and  $\text{Cl}(^2P_{3/2})$  photofragments conserve this +1 projection of the laser photons, producing H and  $\text{Cl}(^2P_{3/2})$  atoms each with the projections of approximately  $m_s = +1/2$  (so that they sum up to +1), and thus the H-atom electron spin is approximately  $m_s = +1/2$  [21]. In a weak magnetic field (Zeeman region), all H atoms are in a coherent superposition of the total angular momentum states  $|F, m_F\rangle$  with the coupling  $\mathbf{F} = \mathbf{S} + \mathbf{I}$  of the electron spin  $\mathbf{S}$  and the nuclear spin  $\mathbf{I}$ . When the electron spin is fixed due to the polarization of the incident laser beam, *e.g.*,  $m_s = +1/2$ , then only the spin combinations  $|m_s = +1/2, m_I = +1/2\rangle$  and  $|+1/2, -1/2\rangle$  can be found in the free Hydrogen atoms. The hyperfine state  $|+1/2, +1/2\rangle = |F = 1, m_F = +1\rangle$  is an eigenstate and will stay unchanged in time. Since the states  $|-1/2, +1/2\rangle$  and  $|+1/2, -1/2\rangle$  are not eigenstates, they are linear combinations of the  $|F = 1, m_F = 0\rangle$  and  $|F = 0, m_F = 0\rangle$  eigenstates, which have different energies. Therefore, atoms produced in the  $|+1/2, -1/2\rangle$  state will oscillate to  $|-1/2, +1/2\rangle$  and back. If now the electron-polarized Hydrogen atoms are produced during a very short time  $t < 1 \text{ ns}$ , they will oscillate in phase. Therefore, after 0.35 ns only the spin combinations  $|+1/2, +1/2\rangle$  and  $|-1/2, +1/2\rangle$  can be found. As a consequence the electron polarization of the Hydrogen atoms, produced by the laser beam, is transferred into a nuclear polarization. If now the Hydrogen atoms are quickly ionized and accelerated, the out-coming protons will remain polarized, even if they undergo spin precession according to the T-BMT equation [17].

Using a Lamb-Shift polarimeter the polarization of an atomic Hydrogen ensemble can be measured in a multi-step process [22, 23]. One important condition is that the atomic beam can be efficiently converted into metastable atoms in the  $2S_{1/2}$  state by ionization with an electron-impact ionizer and a charge reversal in cesium vapor. With a spin filter, individual hyperfine sub-states are selected by applying a static magnetic field, an electric quench field and a high-frequency transition. By varying the resonance condition when changing the magnetic field, single hyperfine components can be detected. Finally, the transition into the ground state within the quenching process is verified by Lyman- $\alpha$  radiation emitted at 121.5 nm. The intensity of the individual hyperfine components allows to measure their occupation number and, therefore, calculate the polarization of incoming protons and, in combination with an ionizer, even for Hydrogen atoms. The entire setup, including laser system, interaction chamber and Lamb-Shift polarimeter, is realized over a length of less than 5 m as a table-top experiment.

As indicated in Fig. 4, the final experiments with the polarized gas target, aiming at the first observation of a polarized proton beam from laser-induced plasmas, will be performed

at the 10 PW laser system SULF (Shanghai Superintense-Ultrafast Lasers Facility) at SIOM (Shanghai Institute of Optics and Fine Mechanics). Theoretical calculations indicate the acceleration of protons out of a gas jet to the GeV energy range for this laser system operated at 300 J energy, 30 fs pulse energy and a focused intensity of  $> 10^{22} \text{ Wcm}^{-2}$  at 1 shot/min. It is predicted that the protons are accelerated in a so-called electron bubble-channel structure [24]. Compared to a traditional electron bubble the plasma density is much higher than the critical density of a relativistic laser pulse for a plasma containing mainly heavy ions such as chlorine. This results from reflections of the laser light from the highly compressed electron layers and self-focusing. As a consequence, only the acceleration of protons but not of heavy ions is expected. Moreover, it has been shown theoretically that in this acceleration scheme protons, which are trapped in the bubble region of the wake field, can be efficiently accelerated in the front of the bubble, while electrons are mostly accelerated at its rear [24]. After the acceleration process the proton polarization will be determined by a detector similar to that one used for our Düsseldorf experiment, but with different scattering foil material to account for the higher proton energies.

Regarding a source for laser-accelerated polarized  $^3\text{He}^{2+}$  ions from a pre-polarized  $^3\text{He}$  gas target an experiment is planned to be conducted at PHELIX, GSI Darmstadt [14, 15, 25, 26].

## DISCUSSION AND CONCLUSION

To summarize, the T-BMT equation, describing the spin precession in electromagnetic fields, has been implemented into the VLPL PIC code to simulate the semi-classical spin motion during laser-plasma interactions. One crucial result of our simulations is that a target containing polarized Hydrogen nuclei is needed for producing polarized relativistic proton beams. A corresponding gas-jet target, based on dynamic polarization of HCl molecules, is now being built at Forschungszentrum Jülich. From a simultaneous interaction of the fundamental wavelength of a Nd:YAG laser and its fifth harmonic with HCl gas, nuclear polarized H atoms are created. Their nuclear polarization will be measured and tuned with a Lamb-shift polarimeter. First measurements, aiming at the demonstration of the feasibility of the target concept, are scheduled for fall 2018. The ultimate experiment will take place in 2019 at the 10 PW SULF facility to observe a up to GeV polarized proton beam from laser-generated plasma for the first time.

## ACKNOWLEDGEMENTS

We thank our colleagues B. F. Shen, L. Ji, J. Xu and L. Zhang from Shanghai Institute of Optics and Fine Mechanics for the various fruitful discussions and their expertise that greatly assisted our research. This work has been carried out in the framework of the JuSPARC (Jülich Short-Pulse Particle and Radiation Center) project and has been supported by the ATHENA (Accelerator Technology HELmholtz iNfrAs-

tructure) consortium. We further acknowledge the computing resources on grant VSR-JPGI61 on the supercomputer JURECA.

## REFERENCES

- [1] A. Macchi, M. Borghesi, and M. Passoni, *Rev. Mod. Phys.*, vol. 85, pp. 751, 2013. doi:10.1103/RevModPhys.85.751
- [2] N. Raab, M. Büscher, M. Cerchez, R. Engels, I. Engin, P. Gibbon, P. Greven, A. Holler, A. Karmakar, A. Lehrach, R. Maier, M. Swantusch, M. Toncian, T. Toncian, and O. Willi, *Physics of Plasma*, vol. 21, pp. 023104, 2014. doi:10.1063/1.4865096
- [3] M. Wen, H. Bauke, and C. H. Keitel, "Dynamical spin effects in ultra-relativistic laser pulses", 2014. arXiv:1406.3659
- [4] J. Vieira, C.-K. Huang, W. B. Mori, and L. O. Silva, *Phys. Rev. ST Accel. Beams*, vol. 14, pp. 071303, 2011. doi:10.1103/PhysRevSTAB.14.071303
- [5] B. M. Garraway, and S. Stenholm, *Contemporary Physics*, vol. 43, pp. 3, 2002. doi:10.1080/00107510110102119
- [6] T. D. Arber, K. Bennett, C. S. Brady, A. Lawrence-Douglas, M. G. Ramsay, N. J. Sircombe, P. Gillies, R. G. Evans, H. Schmitz, A. R. Bell, and C. P. Ridgers, *Plasma Phys. Control. Fusion*, vol. 57, pp. 113001, 2015. doi:10.1088/0741-3335/57/11/113001
- [7] B. F. Shen, X. Zhang, Z. Sheng, M. Y. Yu, and J. Cary, *Phys. Rev. ST Accel. Beams*, vol. 12, pp. 121301, 2009. doi:10.1103/PhysRevSTAB.12.121301
- [8] A. Pukhov, in *Proceedings, CAS - CERN Accelerator School: Plasma Wake Acceleration*, vol. 1, 2016. doi:10.5170/CERN-2016-001.181
- [9] J. Vieira, R. A. Fonseca, and L. O. Silva in *Proceedings, CAS - CERN Accelerator School: Plasma Wake Acceleration*, vol. 1, 2016. doi:10.5170/CERN-2016-001.79
- [10] J. Thomas, A. Hützen, A. Pukhov, A. Lehrach, and M. Büscher, "Tracking of Particle Spins with PIC Codes", publication in preparation.
- [11] Jülich Supercomputing Centre, *Journal of large-scale research facilities A*, vol. 62, pp. 2, 2016. doi:10.17815/jlsrf-2-121
- [12] P. S. Farago, *Rep. Prog. Phys.* vol. 34, pp. 1055, 1971. doi:10.1016/S0065-2539(08)61007-0
- [13] C. D. Keith, J. Brock, C. Carlin, S. A. Comer, D. Kashy, J. McAndrew, D. G. Meekins, E. Pasyuk, J. J. Pierce, and M. L. Seely, "The Jefferson Lab Frozen Spin Target," 2012, arXiv:1204.1250
- [14] I. Engin, M. Büscher, O. Deppert, L. Di Lucchio, R. Engels, S. Frydrych, P. Gibbon, A. Kleinschmidt, A. Lehrach, M. Roth, F. Schlüter, K. Strathmann, and F. Wagner, in *Proceedings of Science (PSTP2015)*, vol. 002.
- [15] H. Soltner, M. Büscher, P. Burgmer, I. Engin, B. Nau-schütt, S. Maier, and H. Glückler, *IEEE T. Appl. Supercon.*, vol. 26, no. 4, 2016. doi:10.1109/TASC.2016.2535293
- [16] A. Nass C. Baumgarten, B. Braun, G. Ciullo, G. Court, P. F. Dalpiaz, A. Golendukhin, G. Graw, W. Haerberli, M. Hennoch, R. Hertenberger, N. Koch, H. Kolster, P. Lenisa, H. Marukyan, M. Raithel, D. Reggiani, K. Rith, M. C. Simani,

- E. Steffens, J. Stewart, P. Tait, and T. Wise, *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 505, pp. 633, 2003. doi: 10.1016/S0168-9002(03)00986-0
- [17] D. Sofikitis, P. Glodic, G. Koumarianou, H. Jiang, L. Bougas, P. C. Samartzis, A. Andreev, and T. P. Rakitzis, *Phys. Rev. Lett.*, vol. 118, pp. 233401, 2017. doi:10.1103/PhysRevLett.118.233401
- [18] D. Sofikitis, L. Rubio-Lago, L. Bougas, A. J. Alexander, and T. P. Rakitzis, *J. Chem. Phys.*, vol. 129, pp. 144302, 2008. doi:10.1063/1.2989803
- [19] EKSPLA, "SL330 series - SBS Compressed Picosecond Nd:YAG Lasers", <http://ekspla.com/product/picosecond-high-energy-ndyag-lasers-sl330-series/>, July 26, 2018.
- [20] B. Friedrich and D. Herschbach, *J. Phys. Chem.*, vol. 99, pp. 42, 1995. doi:10.1021/j100042a051
- [21] T. P. Rakitzis, *ChemPhysChem*, vol. 5, pp. 1489, 2004. doi:10.1002/cphc.200400108
- [22] R. Engels, R. Emmerich, J. Ley, G. Tenckhoff, H. Paetz gen. Schieck, M. Mikirtychiants, F. Rathmann, H. Seyfarth, and A. Vassiliev, *Rev. Sci. Instrum.*, vol. 74, pp. 4607, 2003. doi:10.1063/1.1619550
- [23] R. Engels, E. Emmerichi, K. Grigoryev, J. Ley, M. Mikirtychiants, H. Paetz gen. Schieck, F. Rathmann, J. Sarkad, H. Seyfarth, G. Temckhoff, and V. Vasilyev, *Rev. Sci. Instrum.*, vol. 76, pp. 053305, 2005. doi:10.1063/1.1898923
- [24] B. F. Shen, Y. Li, M. Y. Yu, and J. Cary, *Phys. Rev. E*, vol. 76, pp. 055402, 2007. doi:10.1103/PhysRevE.76.055402
- [25] I. Engin, "Towards Polarization Measurements of Laser accelerated Helium-3 Ions", Ph. D. thesis, Heinrich-Heine-University Düsseldorf (2015).
- [26] I. Engin, M. Büscher, P. Burgmer, K. Dahlhoff, R. W. Engels, P. Fedorets, H. Feilbach, U. Giesen, H. Glückler, F. Klehr, G. Kukhalashvili, A. Lehrach, T. Leipold, W. Lesmeister, S. Maier, B.T. Nauschütt, J. Pfenning, M. Schmitt, H. Soltner, K. Strathmann, E. Wiebe, S. Wolf, *Springer Proceedings in Physics*, vol. 187, 2016. doi:10.1007/978-3-319-39471-8

# SIMULATIONS OF COHERENT ELECTRON COOLING WITH FREE ELECTRON LASER AMPLIFIER AND PLASMA-CASCADE MICRO-BUNCHING AMPLIFIER

J. Ma<sup>1</sup>, G. Wang<sup>1</sup>, V. N. Litvinenko<sup>1,2</sup>

<sup>1</sup> Brookhaven National Laboratory, Upton, New York 11973, USA

<sup>2</sup> Stony Brook University, Stony Brook, New York 11794, USA

## Abstract

SPACE is a parallel, relativistic 3D electromagnetic Particle-in-Cell (PIC) code used for simulations of beam dynamics and interactions. An electrostatic module has been developed with the implementation of Adaptive Particle-in-Cloud method. Simulations performed by SPACE are capable of various beam distribution, different types of boundary conditions and flexible beam line, as well as sufficient data processing routines for data analysis and visualization. Code SPACE has been used in the simulation studies of coherent electron cooling experiment based on two types of amplifiers, the free electron laser (FEL) amplifier and the plasma-cascade micro-bunching amplifier.

## COHERENT ELECTRON COOLING

Coherent electron cooling (CeC) [1, 2, 3] is a novel and promising technique for rapidly cooling high-intensity, high-energy hadron beams. A general CeC scheme consists of three sections: modulator, amplifier and kicker. In the modulator, the ion beam co-propagates with electron beam and each ion imprints a density wake on the electron distribution through Coulomb force. In the amplifier, the density modulation induced by ions is amplified by orders. In the kicker, the electron beam with amplified signal interacts with ion beam, giving coherent energy kick to ions towards their central energy, which consequently leads to cooling of the ion beam.

Figure 1 [3] shows the schematic of CeC using high gain free electron laser (FEL) as the amplifier, which is related with the CeC experiment in the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory (BNL).

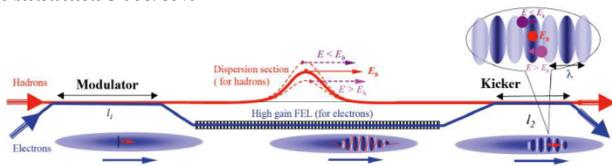


Figure 1: Schematic of coherent electron cooler based on high gain free electron laser.

Figure 2 [4] illustrates the layout of a CeC with a plasma-cascade amplifier (PCA). In the PCA, we use solenoids to control the transverse size of electron beam and make use of the exponential instability of longitudinal plasma oscillations to amplify the initial modulation. A CeC with a PCA does not require bending of ion beam.

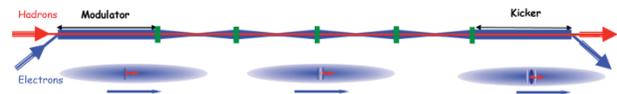


Figure 2: Layout of coherent electron cooler with a plasma-cascade amplifier.

## SIMULATION TOOL

Our main simulation tool is code SPACE [5]. SPACE is a parallel, relativistic, 3D electromagnetic Particle-in-Cell (PIC) code developed for the simulations of relativistic particle beams, beam-plasma interaction, and plasma chemistry. Benchmark test has been performed for SPACE with several accelerator physics codes including MAD-X, ELEGANT and Impact-T, and a good agreement has been achieved. SPACE has been used for the study of plasma dynamics in a dense gas filled RF cavity [6] and the study of mitigation effect by beam induced plasma [7].

Electrostatic module contained in code SPACE has been mainly used in our study, as the particle interaction is essentially electrostatic in the co-moving frame. This code module includes two different approaches. The first one is the traditional PIC method for the Poisson-Vlasov equation, which uses uniform Cartesian mesh, linear charge deposition scheme and fast Fourier transform (FFT) solver. This approach is precise and effective for particles with uniform distribution and computational domain with pure periodic boundary condition. The second approach is a new adaptive Particle-in-Cloud (AP-Cloud) method [8]. This method, based on real particle distribution, generates an adaptively chosen set of computational particles as the mesh, and uses the weighed least squares method for approximation of differential and integral operators. AP-Cloud method is beneficial for particles with non-uniform distribution and computational domain with irregular geometry and mixed type of boundary conditions, such as open boundary condition in the transverse directions and periodic in the longitudinal direction. Both approaches have passed series of verification tests and have been compared in our study. AP-Cloud method produced higher accuracy for electron beam with Gaussian distribution and computational domain with mixed boundary conditions, which are used in CeC simulations, so we have used AP-Cloud method in this study.

SPACE contains various data processing routines and provides sufficient output for data analysis and

visualization. The optional output files from code SPACE includes full 6-D particle distribution, integrated 3-D density distribution in given subset of the computational domain, projected 2-D density distribution on given plane, 1-D density and velocity distribution along longitudinal and transverse directions, particle distribution in phase space and frequency domain, beam parameters including transverse size and emittance.

GENESIS [9] is a three dimensional, time dependent code, developed for high gain FEL simulations. This code has been used in the simulations of FEL amplifier in our study.

## SIMULATIONS OF CEC WITH FEL AMPLIFIER

### Algorithm and Verification

In modulator, the first section of CeC, ions induce modulations in electron beam by attracting surrounding electrons. The relative modulation of electrons due to their interactions with ions is orders of magnitudes smaller than unity, therefore we can co-propagate a single ion with electron beam in modulator simulations and use super position principle to get the modulations by all ions. One of the difficulties using a single ion is the detection of the density and velocity modulation in electron beam, as the signal is too weak compared to the shot noise in electrons. The following algorithm has been used to extract the modulation signal induced by a single ion. We perform two simulations with identical initial distribution of electron, one simulation operates with electron beam only while the other simulation includes a single ion. At the exit of modulator, we can take difference in the electron distribution from the two simulations, to obtain the influence from the single ion. Figure 3 illustrates the typical signal-to-noise ratio in modulator simulations and clearly shows that the shot noise has been eliminated when we extract the modulation signal. Similar algorithm has been used in simulating the FEL amplification process with shot noise [10].

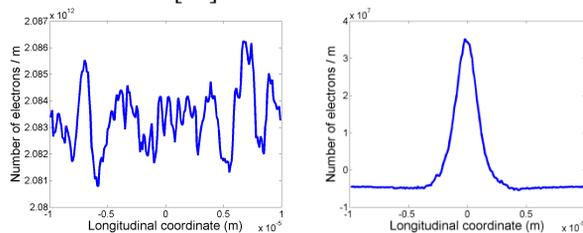


Figure 3: Comparison between shot noise in electrons (left) and modulation by a single ion (right) in longitudinal density distribution of electrons.

We have justified the super position principle in our simulations, as is shown in Figure 4. The blue solid line in Figure 4 shows the summation of density modulations by two ions from separate simulations, and the red dash line is the resulting density modulation when we put the two ions in the same run of simulation. These two modulation signals achieve a good agreement. Therefore, we can use

single ion in our simulations studies of CeC and follow super position principle to get the modulations induced by all ions in a beam.

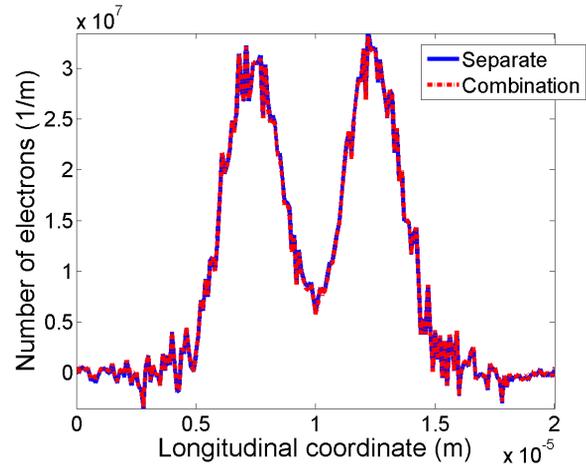


Figure 4: Comparison of density modulations induced by two ions from separate simulations (blue solid line) and from the same simulation (red dash line).

We have verified modulator simulation results through the comparison with theory. Analytical solution to the modulation problem exists for a moving ion co-propagating with an infinite electron beam with uniform spatial distribution [11]. Our simulation results have achieved a good agreement with the analytical solution [12], as is shown in Figure 5 [13].

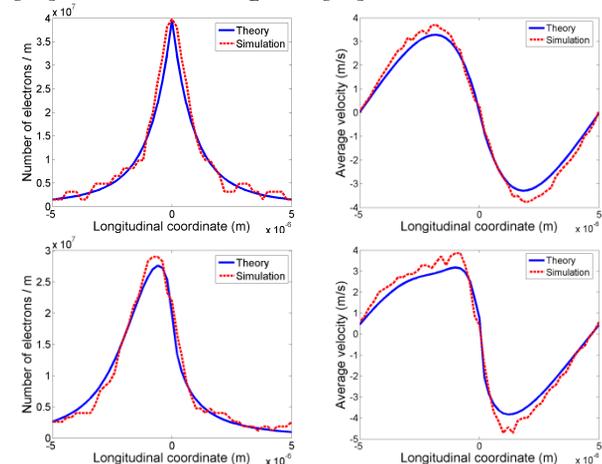


Figure 5: Comparison between theory and numerical simulations in density (left) and velocity (right) modulations by a single ion with reference energy (top) and off-reference energy (bottom) with respect to uniform electron cloud.

### Modulator

We list main parameters of electron and ion beams in Table 1. These parameters are related with the operations of CeC experiment at BNL RHIC, and have been used in numerical simulations to predict the cooling time.

Figure 6 shows the dynamics of  $\beta$  functions of electron beam in the modulator. Quadrupoles are used to focus the electrons in the beam line and to match the transverse beam size at the exit of modulator to obtain high gain in

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

FEL amplifier, the second section of CeC. Electron beam dynamics in a quadrupole beam line using code SPACE have been benchmarked with other accelerator simulations tools, including MAD-X, ELEGANT and Impact-T [13, 14].

Table 1: Parameters of Electron and Ion Beams

	Electron	Ion, Au <sup>+79</sup>
Beam energy	$\gamma=28.5$	$\gamma=28.5$
Peak current	75 A	
Normalized emittance	$8 \pi$ mm mrad	$2 \pi$ mm mrad
R.M.S. energy spread	$1e-3$	$3e-4$

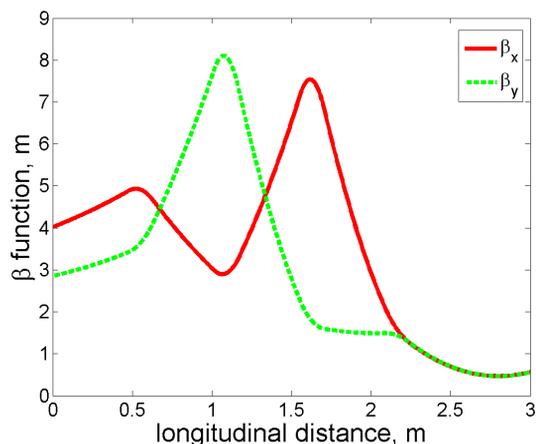


Figure 6: Evolution of electron beam  $\beta$  function in modulator section.

Figure 7 shows that the longitudinal density modulation gradually builds up when the ion co-propagates with the electron beam in modulator.

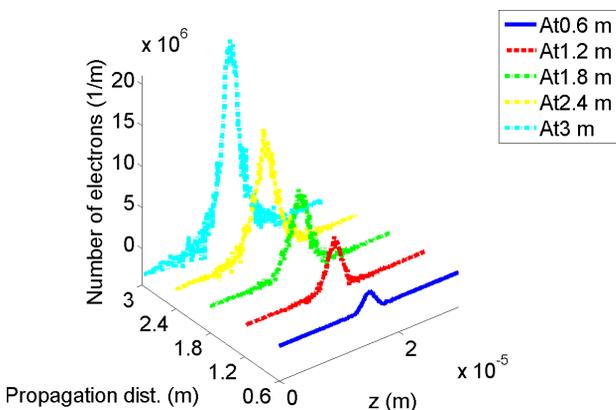


Figure 7: Longitudinal density modulation at several propagation distances in modulator.

We use the parameter bunching factor to quantify the longitudinal modulation. Bunching factor is also used in code GENESIS to obtain the gain in FEL simulations, and is defined in Equation (1) [9],

$$b \equiv \frac{1}{N_\lambda} \sum_{k=1}^{N_\lambda} e^{i \frac{2\pi}{\lambda_{opt}} z_k}, \quad -\frac{\lambda_{opt}}{2} \leq z_k \leq \frac{\lambda_{opt}}{2} \quad (1)$$

where  $\lambda_{opt}$  is the FEL optical wavelength, the summation is over a slice of  $\lambda_{opt}$  wide, centered at the ion's location, and  $N_\lambda$  is the total number of electrons within that slice.

The longitudinal density modulation shown in Figure 7 is induced by a single ion with reference energy and zero transverse offset with respect to the center of the electron beam, and is expected to achieve maximum bunching factor at the exit of modulator. Bunching factor reduces when we start with a single ion with transverse offsets. Figure 8 shows that the modulation becomes weaker when the ion is further away from the center of the electron beam in transverse plane. The dependence of bunching factor on transverse offsets is not symmetric in horizontal direction and vertical direction, as the transverse beam size is not symmetric in the quadrupole beam line, which is shown in Figure 6.

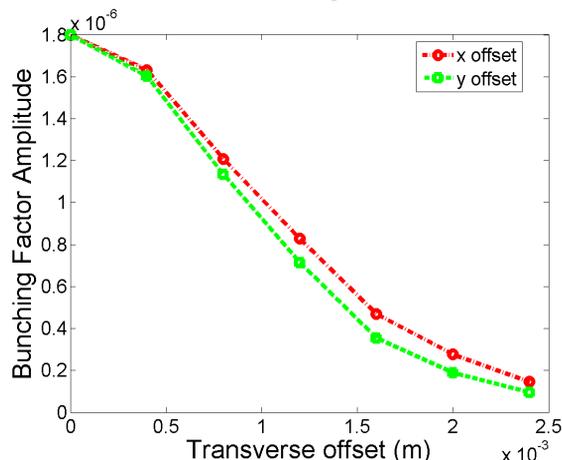


Figure 8: Amplitude of bunching factor induced by ions with various transverse offsets with respect to the center of the Gaussian electron beam.

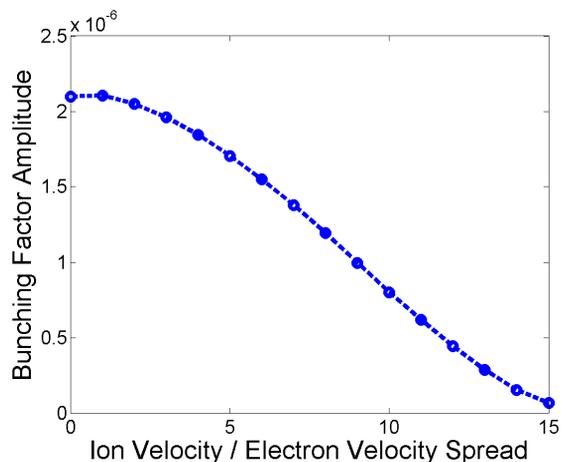


Figure 9: Amplitude of bunching factor induced by ions with various longitudinal velocities with respect to ions the electron beam group velocity. Ion velocity is in the unit of electrons' velocity spread.

Dependence of bunching factor on the energy difference between ions and electrons has also been studied [12], as is shown in Figure 9. Detailed studies have been performed to simulate the modulation process using ions with various combinations of off-reference energies, transverse offsets and transverse velocities [15].

### FEL Amplifier

We have exported the full 6-D particle distribution at the exit of modulator from code SPACE and imported it into code GENESIS to simulate the second section of CeC, FEL amplifier. GENESIS separate particles into longitudinal slices, and the length of each slice is FEL optical wavelength, which is about  $30 \mu\text{m}$  for our settings.

The FEL device installed at BNL RHIC has three sections of wigglers separated by drift space. It is difficult to achieve an envelope with constant transverse beam size over the three-section wigglers. Instead, we have designed an envelope with oscillating beam size and minimized the overall variation, which is shown in Figure 10.

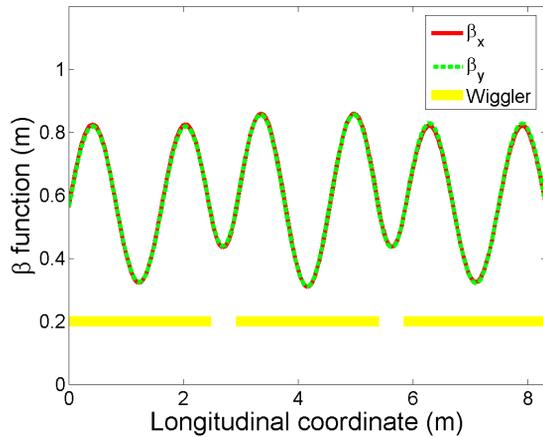


Figure 10: Location of three-section wigglers in FEL amplifier and  $\beta$  function evolution with minimum overall variation.

In FEL section, both the shot noise and the modulation signal are amplified. We need to maximum the gain from FEL to reduce the cooling time, and we should avoid saturation to preserve the correlation between the amplified signal and the ion which induces the initial modulation.

Figure 11 displays the growth of bunching factor of shot noise in FEL and shows that saturation is not reached at the end of FEL section. Figure 12 gives comparison of bunching factor between the initial density modulation at the entrance of FEL section and the final amplified signal at the exit of FEL. This signal has been extracted from the shot noise using the similar algorithm applied in modulator simulations. The comparison in Figure 12 clearly shows the widening of the initial modulation and a gain of 210 in bunching factor over the FEL section, which is sufficient for cooling.

Diffusion rate for CeC has been obtained using the final bunching factor shown in Figure 12, and parameters relevant to CeC experiment at BNL RHIC [16].

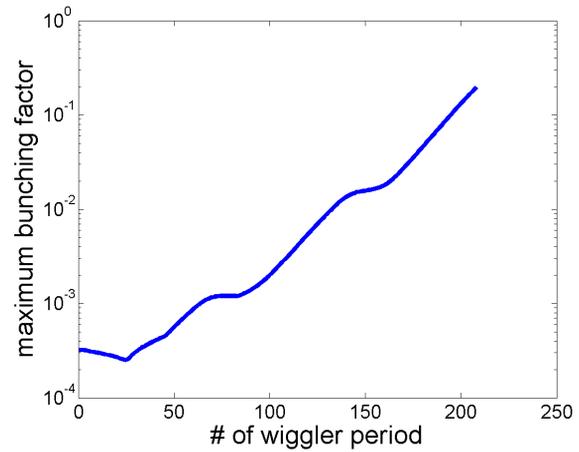


Figure 11: Evolution of bunching factor amplitude of shot noise in three-section FEL.

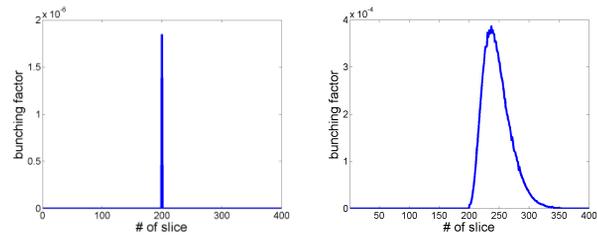


Figure 12: Bunching factor amplitude of modulation signal at the entrance (left) and exit (right) of FEL section.

### Kicker

In the kicker, ions interact with amplified modulation signals carried in electron beam and receive energy kick towards reference energy, which results in the cooling of ion beam. We take the output from GENESIS into SPACE for kicker simulations. Quadrupole setting in the kicker is symmetric with that in modulator. Electron beam envelope in the kicker is shown in Figure 13.

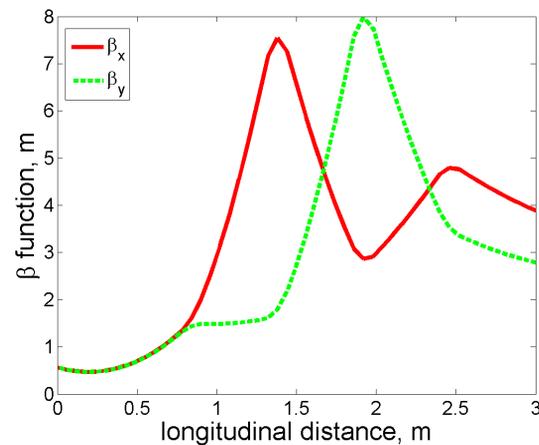


Figure 13: Evolution of electron beam  $\beta$  function in kicker section.

Figure 14 gives a close look at the amplified density modulation in the kicker section, and the coherent longitudinal velocity kick it gives to ions towards the reference energy. Red dot in Figure 14 represents ions

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

with reference energy, and should receive zero energy kick. Green dots indicate ions with energy spread  $5.7e-4$  and this is the upper limit for CeC. Using ion beam with larger energy spread will result in inefficient cooling or even anti-cooling. Yellow dots are ions with energy spread  $3e-4$ , which is the value for CeC experiment listed in Table 1, and this is within the regime of efficient cooling.

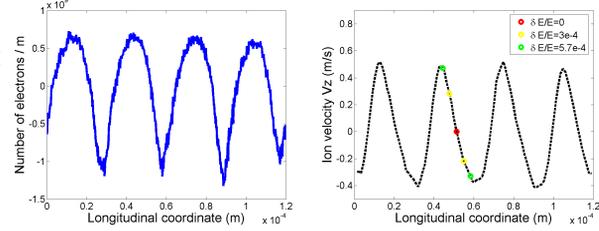


Figure 14: Amplified longitudinal density modulation in the kicker section (left) and the velocity kick it gives to ions at various longitudinal locations in a single pass through CeC system (right). Dots indicate specific ions with typical energy deviation.

We have traced ions with off-reference energies and the coherent kick they received in the kicker section, as is shown in Figure 15. As a result, we can predict the cooling time with only coherent kick included. A more realistic estimation of cooling time should include random kicks from surrounding ions and electrons.

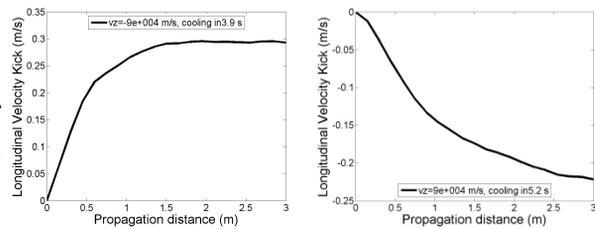


Figure 15: Coherent longitudinal velocity kick to a single ion with lower energy (left) and higher energy (right) with energy spread  $3e-4$  in the kicker section and cooling time. Values of velocities are in co-moving frame.

## SIMULATIONS OF PLASMA-CASCADE MICRO-BUNCHING AMPLIFIER

In the plasma-cascade micro-bunching amplifier, we have used strong solenoids to over-focus the electron beam and made use of the fast-growing plasma instability to boost the initial modulation, which acts as the amplifier for CeC. Designed parameters of PCA for RHIC and eRHIC at BNL are given in [4], and we present simulation studies of PCA using relevant settings.

Transverse beam size evolution in a four-cell PCA from numerical simulation is given in Figure 16. We have used very strong solenoid fields between cells to compress the beam and to make the plasma instability happen and grow. Figure 17 shows the gain we can get from the PCA. In Figure 17, we used Fast Fourier transform (FFT) to convert the longitudinal density distribution into frequency domain and calculate the gain in PCA section.

The maximum gain is about 120, which is sufficient for CeC, and it appears at around 30 THz.

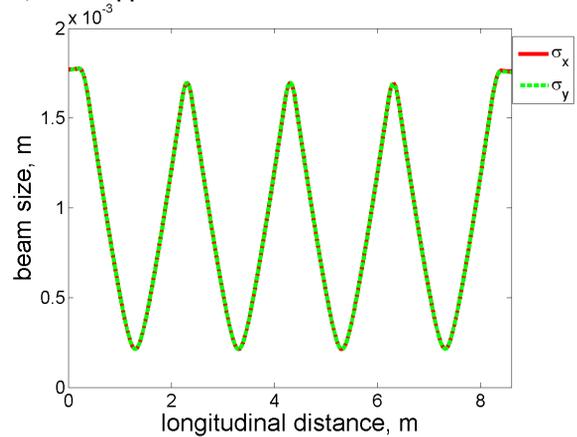


Figure 16: Evolution of transverse beam size in a four-cell plasma cascade amplifier.

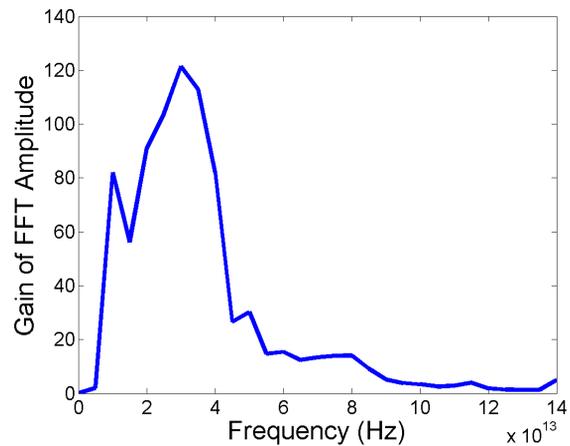


Figure 17: Gain of density modulation for various frequencies at the exit of PCA, calculated in frequency domain.

We have studied the evolution of signal over the PCA section by adding an initial modulation on top of the shot noise, and we have used the similar algorithm in modulator simulations to extract this signal from shot noise in PCA simulations. As is shown in Figure 18, we introduced an initial density modulation at 25 THz and observed the increase of the signal in PCA. Note that Figure 18 only presents the density modulation, and the sharp drops in Figure 18 indicate that density modulation has been transformed into velocity modulation.

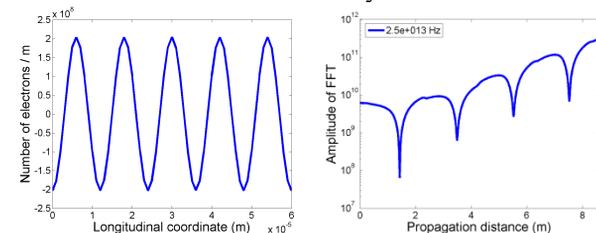


Figure 18: Initial density modulation (left) at the entrance of PCA and the amplification of density modulation (right) within the PCA section at frequency 25 THz.

Similarly, we added an initial velocity modulation to electron beam at the entrance of PCA at 25 THz and had the amplification in signal shown in Figure 19. Note that in the evolution plot of Figure 19, only density modulation is included. So the initial signal is small as we introduced pure velocity modulation, which has not converted into density modulation yet at the entrance of PCA.

A more realistic simulation will use a modulation signal induced by a real ion, which includes both density and velocity modulations and has a wider bandwidth instead of a specific frequency.

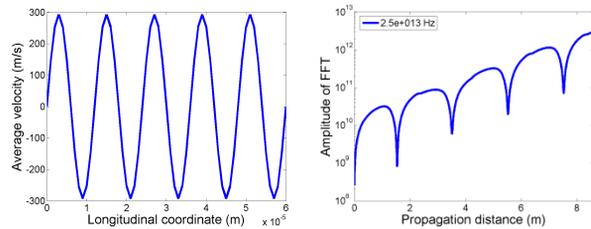


Figure 19: Initial velocity modulation (left) at the entrance of PCA and the amplification of density modulation (right) within the PCA section at frequency 25 THz.

Detailed beam dynamics in PCA section have been investigated, and we present the 2-D density distribution of modulation signal at several locations along the PCA beam line in Figure 20. Electrons at the transverse edge of the beam fall behind the central electrons, as they experience stronger solenoid focusing which introduces larger transverse motions to them and reduces their longitudinal velocities. An increase in density modulation is observed in Figure 20.

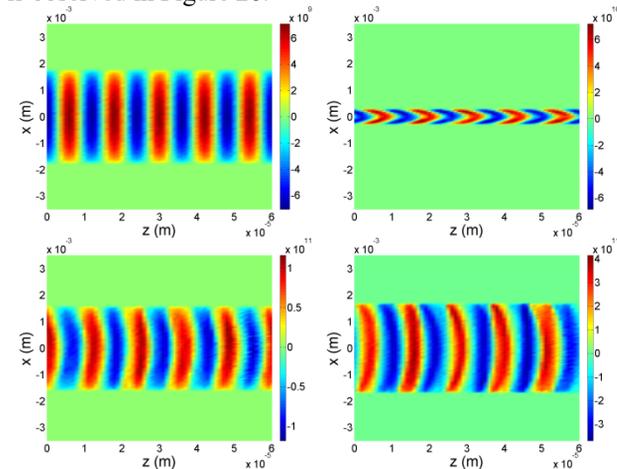


Figure 20: 2-D density modulation in the electron beam at the entrance of PCA (top left), middle of second cell (top right), end of third cell (bottom left) and exit of PCA (bottom right). X-axis is along the horizontal direction and z-axis is along the longitudinal direction.

We have quantified the delay of electrons at the beam edge compared with the central electrons, and present the result in Figure 21. The phase difference between central particle and edge particle is 45 degree, which is obtained from the fitting in Figure 21.

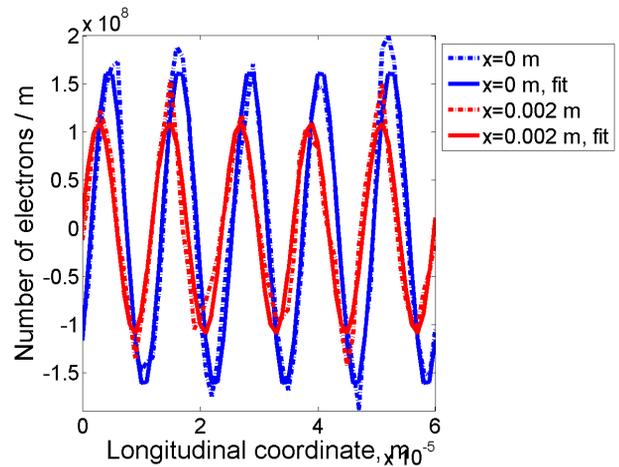


Figure 21: Comparison of density modulations between central electrons (blue lines) and electrons at transverse edge of the beam (red lines) at the exit of PCA.

## CONCLUSION

We present the simulation studies of CeC process using code SPACE and GENESIS.

We have successfully eliminated the shot noise and extracted the modulation signals in simulations using a single ion. Super position principle is justified in SPACE simulations. Simulation results have been verified through the comparison with analytical solutions to the modulator problem.

Start-to-end simulations have been performed for ions passing through the CeC system with FEL amplifier. We have studied the dependence of modulation process on various transverse offsets and off-reference energies of ions, and predicted the cooling time.

We have explored the use of plasma-cascade amplifier, which replaces the FEL amplifier in CeC. The gain and corresponding frequency in PCA have been obtained through numerical simulations. Detailed beam dynamics through the PCA have been analysed.

SPACE will be used in the further study of CeC to provide strong support to the design and operation of CeC experiment at BNL RHIC.

## REFERENCES

- [1] V. N. Litvinenko and Y. S. Derbenev, Free Electron Lasers and High-Energy Electron Cooling, in *Proc. FEL 2007*, Budker INP, Novosibirsk, Russia, BNL-79509-2007-CP.
- [2] V. N. Litvinenko, Coherent Electron Cooling, in *Proc. PAC 2009*, Vancouver, BC, Canada, paper FR1GRI01.
- [3] V.N. Litvinenko, Y.S. Derbenev, Coherent Electron Cooling, *Phys. Rev. Lett.* 102, 114801, 2009.
- [4] V. N. Litvinenko *et al.*, Plasma-Cascade Micro-Bunching Amplifier and Coherent Electron Cooling of A Hadron Beams, arXiv:1802.08677, 2018.
- [5] K. Yu *et al.*, Space Code for Beam-Plasma Interaction, in *Proc. IPAC 2015*, Richmond, VA, USA, MOPMN012.
- [6] K. Yu *et al.*, Simulation of Beam-Induced Plasma in Gas Filled Cavities, *Proc. IPAC 2015*, Richmond, VA, USA, paper MOPMN013.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

- [7] J. Ma *et al.*, Simulation of Beam-Induced Plasma for the Mitigation of Beam-Beam Effects, in *Proc. IPAC 2015*, Richmond, VA, USA, paper MOPMN015.
- [8] X. Wang *et al.*, Adaptive Particle-in-Cloud method for optimal solutions to Vlasov-Poisson equation, *J. Comput. Phys.*, 316 (2016), 682-699.
- [9] S. Reiche, *Introduction on Genesis 1.3*, 2002, [http://genesis.web.psi.ch/download/documentation/gegenes\\_talk.pdf](http://genesis.web.psi.ch/download/documentation/gegenes_talk.pdf)
- [10] Y. Jing *et al.*, Model Independent Description of Amplification and Saturation Using Green's Function, arXiv:1505.04735, (2015).
- [11] G. Wang, M. Blaskiewicz, Dynamics of ion shielding in an anisotropic electron plasma, *Phys. Rev. E*, 78 (2008) 026413.
- [12] J. Ma *et al.*, Simulation Studies of Modulator for Coherent Electron Cooling, submitted to Physical Review Accelerators and Beams.
- [13] J. Ma *et al.*, Simulations of Modulator for Coherent Electron Cooling, *Proc. IPAC 2018*, Vancouver, BC, Canada, paper THPAF005.
- [14] J. Ma *et al.*, Modulator Simulations for Coherent Electron Cooling, *Proc. NAPAC 2016*, Chicago, IL, USA, paper WEPOA55.
- [15] J. Ma, Numerical Algorithms for Vlasov-Poisson Equation and Applications to Coherent Electron Cooling, Ph.D. Thesis, State University of New York at Stony Brook, ProQuest Dissertations Publishing, 2017.
- [16] J. Ma *et al.*, Simulation of Cooling Rate and Diffusion for Coherent Electron Cooling Experiment, *Proc. IPAC 2018*, Vancouver, BC, Canada, paper THPAF006.

# HIGH-FIDELITY THREE-DIMENSIONAL SIMULATIONS OF THERMIONIC ENERGY CONVERTERS\*

N. M. Cook<sup>†</sup>, J. P. Edelen, C. C. Hall, M. Keilman, P. Moeller, R. Nagler  
 RadiaSoft LLC, Boulder, CO, USA  
 J.-L. Vay, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

## Abstract

Thermionic energy converters (TEC) are a class of thermoelectric devices, which promise improvements to the efficiency and cost of both small- and large-scale electricity generation. A TEC is comprised of a narrowly-separated thermionic emitter and an anode. Simple structures are often space-charge limited as operating temperatures produce currents exceeding the Child-Langmuir limit. We present results from 3D simulations of these devices using the particle-in-cell code Warp, developed at Lawrence Berkeley National Lab. We demonstrate improvements to the Warp code permitting high fidelity simulations of complex device geometries. These improvements include modeling of non-conformal geometries using mesh refinement and cut-cells with a dielectric solver. We also consider self-consistent effects to model Schottky emission near the space-charge limit for arrays of shaped emitters. The efficiency of these devices is computed by modeling distinct loss channels, including kinetic losses, radiative losses, and dielectric charging. We demonstrate many of these features within an open-source, browser-based interface for running 3D electrostatic simulations with Warp, including design and analysis tools, as well as streamlined submission to HPC centers.

## INTRODUCTION

Thermionic energy converters (TECs) generate electrical power from external heat sources using thermionic emission. By driving electrons across a narrow vacuum gap connected to an external load, electric power is created. For modest gap distances however, the thermionic current quickly exceeds the Child-Langmuir limit, reducing the peak achievable device power. To overcome space charge limitations an accelerating grid is used to compensate the negative potential generated by the beam space charge. The efficiency of such a device is theoretically limited only by the difference in temperature between the hot emitter and cold collector. However, the presence of a grid, along with realistic material properties of the device, serve to reduce their efficiency. Sophisticated simulations are needed to properly capture these dynamics.

Previous efforts to address these needs led to the development of an efficiency model and self-consistent simulation procedure for evaluating TEC designs [1]. In this paper we provide a brief review of this model and its implementa-

tion using the Warp particle-in-cell framework [2]. We then illustrate the value of this model in optimizing devices using a simple case study in grid placement and transparency. Lastly, we discuss the improvements being made to the electrostatic solver within the Warp to further improve vacuum nano-electronic device modeling.

## EFFICIENCY MODEL

Energy conversion in a TEC is limited by a set of discrete loss channels, including kinetic, thermal, radiative, and resistive losses. Proper evaluation necessitates the tracking and quantification of each loss channel. For the Warp simulations discussed in these studies, we've adopted a model that is well-established in literature [3] and has been applied in recent experimental studies [4].

The model identifies four main loss mechanism of power from the TEC system: power carried by electrons leaving the emitter  $P_{ec}$ , net radiative power from the emitter  $P_R$ , conductive heat loss in the attached circuit  $P_{ew}$ , and finally power lost from holding the voltage on the grid  $P_{grid}$ . We note that the simulations assume periodic boundaries, and so the current and corresponding power quantities are normalized by area. The conversion efficiency of the device is the ratio of the net electrical power generated divided by the net thermal power exhausted. If the electrical power that is generated from circuit load is  $P_{load}$ , the efficiency  $\eta$  is:

$$\eta = \frac{P_{load} - P_{grid}}{P_{ec} + P_R + P_{ew}} . \quad (1)$$

The net power transmitted from emitter to collector is:

$$P_{ec} = J_e (\phi_e + 2k_B T_e) - J_c (\phi_e + 2k_B T_c) . \quad (2)$$

Here,  $J_e$ , the current leaving the emitter, is known exactly from the simulation. The second term accounts for return current back to the emitter. The emitter and collector work functions are  $\phi_e$  and  $\phi_c$ . The collector is assumed to be held at a low temperature ( $T_c < 500^\circ$  K), thus we analytically compute the return current  $J_c$ .

The radiative heat loss is based on an analytic calculation for infinite parallel plates with some shielding from the grid. This is calculated as:

$$P_R = \epsilon \sigma_{sb} (T_e^4 - T_c^4) , \quad (3)$$

where  $\epsilon$  is an effective emissivity,  $\sigma_{sb}$  is the Stefan-Boltzmann constant, and  $T_e$  and  $T_c$  are the emitter and collector temperature respectively.

\* This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC0017162.

<sup>†</sup> ncook@radiasoft.net

The power dissipated from the circuit is calculated as:

$$P_{ew} = 0.5 \left( \frac{L}{\rho_{ew}} (T_{em} - T_{env})^2 - \rho_{ew} (J_{ec} - tJ_c)^2 \right), \quad (4)$$

where  $J_{ec}$  is the current density from the emitter that reaches the collector, known exactly from the simulation. The resistivities  $\rho_{ew}$  and  $\rho_{cw}$  account for the emitter and collector side wiring respectively and are calculated as a function of temperature.

The power lost in the gate is calculated based on the grid voltage  $V_{grid}$ , the current density striking the grid  $J_{grid}$  and the estimated return current, taking into account the geometric transparency,  $t$ , of the grid:

$$P_{grid} = V_{grid} (J_{grid} + (1 - t)J_c). \quad (5)$$

Finally, the power generated by the TEC is simply calculated from net current at the collector  $J_{ec}$  and the load voltage  $V_{load}$ :

$$P_{load} = J_{ec} V_{load}. \quad (6)$$

### Simulation Procedure

The efficiency model described by Eq. (1) is only valid for steady-state operation. Reaching steady-state requires that the transient dynamics dissipate prior to collecting electron current information. In particular, fluctuations in current near the space-charge limit along with corresponding external circuit feedback must first subside.

In a TEC, current drawn through a load prompts a subsequent decrease in the effective voltage across the gap. An external circuit model permits these time-varying adjustments to maintain an accurate, self-consistent current value as the device reaches steady-state. We have implemented such a model, using Python hooks into Warp to adjust the gap voltage concurrently with the simulation. These adjustments can be made with arbitrary stride, meaning that fast fluctuations driven by noise in the simulation may be period-averaged to obtain an accurate correction. The resistance of the load is pre-computed based on the temperature and work functions of the emitter and collector and estimates for circuit material properties. Further details on this implementation can be found in [1].

We have developed a general procedure for verifying steady-state operation along with segregating measurements from this initial background. The procedure divides a simulation into four phases. The first two phases consist of initial emission of background electrons, during which current and circuit fluctuations are allowed to subside. Next, uniquely tagged “measurement” particles are emitted, from which statistics are collected for computing efficiencies. During the final phase of the simulation, background particles are re-emitted and any remaining “measurement” particles are allowed to complete their trajectories. This procedure is controlled using Python hooks into the Warp simulation, enabling user-defined feedback and consistency checks. Figure 1 provides a graphical schematic of this procedure.

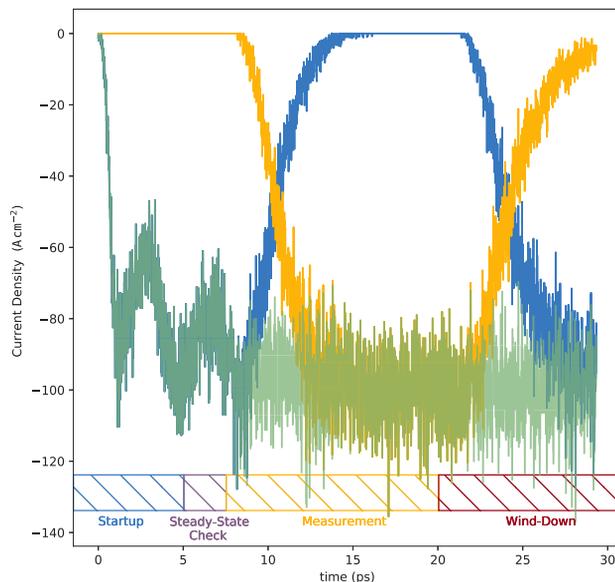


Figure 1: The four phase measurement scheme for validating steady-state and registering measurement particles [1].

## GRID POSITION STUDIES

The need for self-consistent modeling and a comprehensive efficiency model can be seen even when considering a single design parameter. In this case, we consider the role of grid position in a TEC and its influence on device efficiency. We maintain fixed cathode and anode configuration, and vary the position of a grid with fixed geometry. In this case, the grid consists of a 4x4 rectangular lattice of struts, each 5 nm in width and depth. Simulations are performed with varying cathode temperature and grid voltage, and the grid transparency is computed for each run. In each case, grid transparency is only weakly coupled to temperature and to voltage. Figure 2 depicts this dependance on voltage.

We conclude that grid transparency scales roughly linearly with position along the z-axis, peaking when the grid is roughly 80% of the way across the gap. This behavior can be explained by the relative change in transverse momentum of the electrons due to the change in grid position. When the grid is near the cathode, a considerable fraction of the electrons’ momentum is in the transverse plane, thereby increasing the likelihood of impact with the grid. When the grid is moved closer to the anode, the transparency increases due to the effective reduction in transverse momentum. Figure 3 illustrates this relationship.

Implementing the efficiency model quickly identifies a preferred operating point amongst otherwise indistinct conditions. Figure 4 reveals that the efficiency peaks for small voltages before rapidly declining. The reason for this is well understood from the transparency dynamics and the grid loss relationship described by Eq. (5). While the transparency remains relatively unchanged with increasing voltage, the power lost on the grid increases linearly with voltage. Thus, the optimal voltage is the smallest value to permit the full thermionic current to cross the gap.

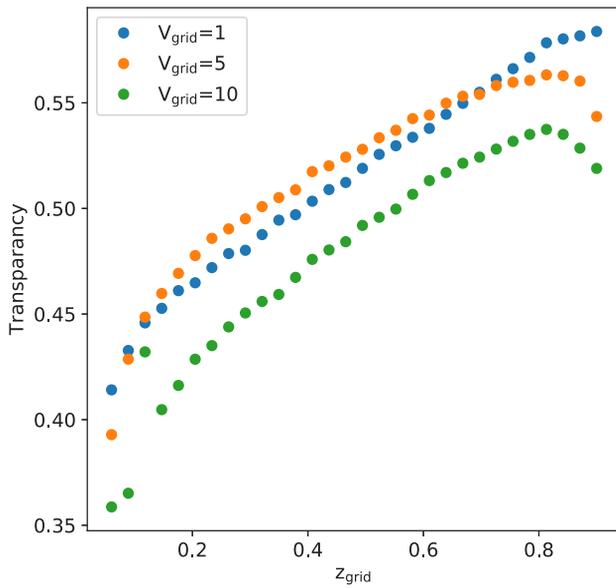


Figure 2: Grid transparency varies only weakly with grid voltage for fixed device geometry.

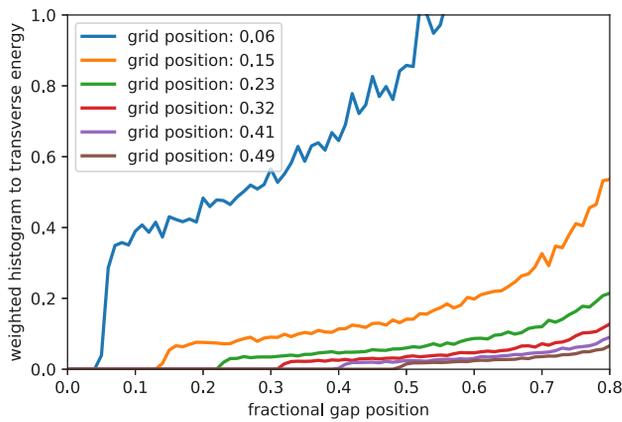


Figure 3: Electron transverse momenta are much larger at the point of grid crossing for grid positions nearer to the cathode. Grid transparency is subsequently reduced.

## IMPROVEMENTS TO THE WARP CODE

Highly resolved, self-consistent simulations of emitted electrons and their interactions with the device geometry are required to properly implement these models. To this end, we employ the Warp particle-in-cell framework, currently in development at Lawrence Berkeley National Laboratory. Warp is a three-dimensional, time-dependent, multi-species PIC framework, which includes a flexible array of solvers for obtaining electrostatic self-fields from Poisson's equations, or full electromagnetic fields from Maxwell's equations [2]. Most TECs operate in a low energy regime for which an electrostatic description will suffice, so we limit our consideration to the electrostatic components of the code.

Here we discuss recent work with the Warp code to improve the modeling of similar classes of vacuum nano-electronic devices. Although the Warp code contains myriad

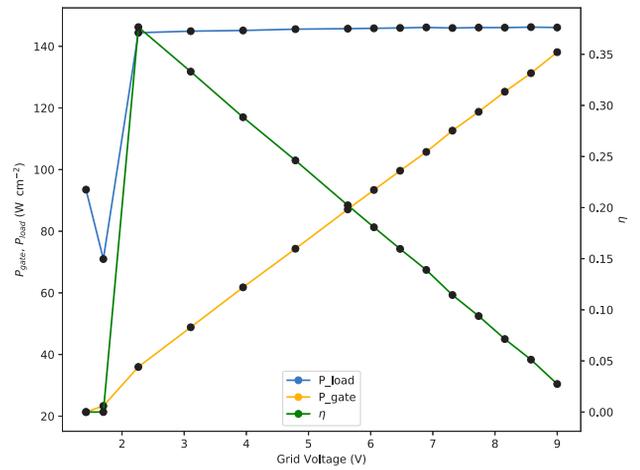


Figure 4: Efficiency,  $\eta$ , as a function of grid voltage, including kinetic losses and other terms in the model. The efficiency quickly peaks for small voltages, as the current crossing the gap and thus the load power  $P_{\text{load}}$  saturates. Additional applied voltage serves only to increase losses on the gate, as represented by the increasing  $P_{\text{gate}}$  curve.

features, further development is required to improve operation in the space-charge limited regime, along with the inclusion of three-dimensional internal dielectric structures.

## Self-consistent Emission Models

In order to reach peak efficiencies, TECs must be operated with a large temperature differential. This is usually accomplished through increasing the emitter temperature, thus generating very high peak currents. The combination of high currents with small device size means that operation is oftentimes space-charge-limited. Therefore, proper device modeling requires an emission model which takes into account beam self-fields as well as external applied fields, even when operating at high temperatures. To this end, we have validated the Schottky emission model in Warp for thermionic emission with large applied fields in the space-charge limit.

For a cathode at temperature  $T$ , with work function  $\phi_e$ , the current density  $j_T$  is given by

$$j_T = AT^2 e^{-(\phi_e - \Delta\phi)/k_B T}, \quad (7)$$

where  $k_B$  is Boltzmann constant, and  $\Delta\phi$  is the effective change in work function of the material, resulting from an external electric field  $E$  according to

$$\Delta W = \sqrt{\frac{e^3 E}{4\pi\epsilon_0}}. \quad (8)$$

This simple model is effective at describing emission behavior in the presence of modest space-charge fields and applied fields up to  $\sim 10^8$  V/m. Warp simulations of a simple diode with varying applied field and cathode temperature show good agreement with this model, as can be seen in Fig. 5. Furthermore, the space-charge forces of the beam are

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

properly accounted for as well, and emission is limited by the Child-Langmuir law in these cases, as shown in Fig. 6.

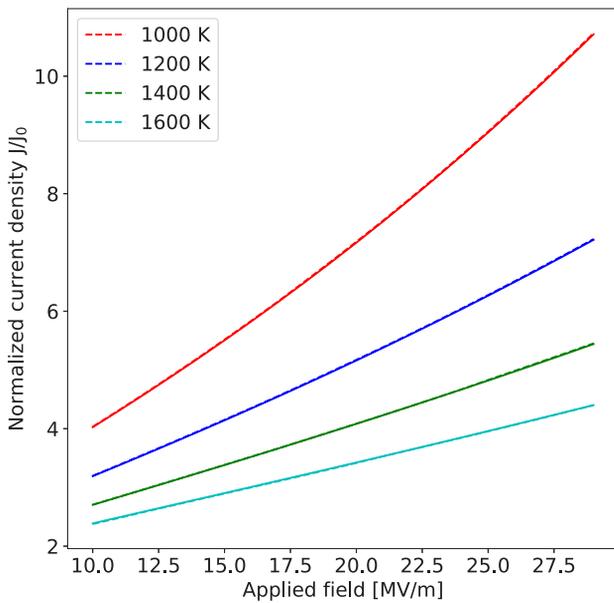


Figure 5: Normalized current density is plotted as a function of applied field on the cathode surface for several cathode temperatures. The simulated current density in Warp shows excellent agreement with the analytic prediction from Schottky emission theory

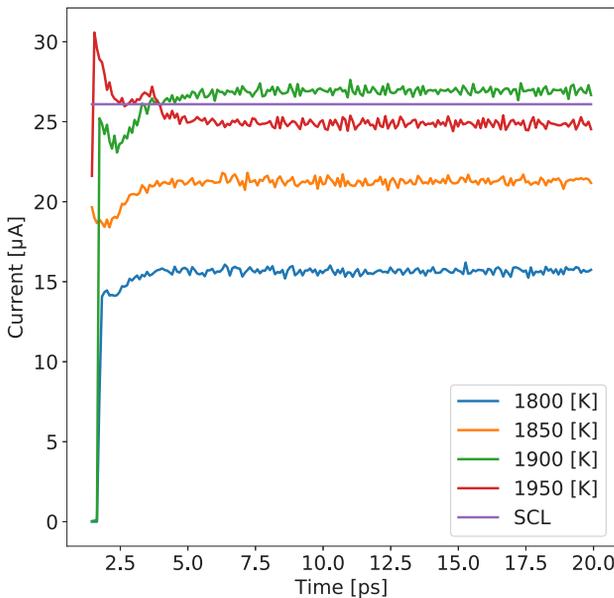


Figure 6: As the space-charge limit is approached with increasing temperature, the simulated emission shows small fluctuations in time before settling to the expected value.

### Dielectric Solver

For the parameter regime under consideration, we use Warp's electrostatic multi-grid solver to obtain solutions to Poisson's equation. The advantage of this solver for TEC

modeling is the ability to incorporate complex boundary conditions using the Shortley-Weller method to obtain subgrid-resolution representations during the solve [5]. As a result, a higher order stencil can be used to obtain accurate representations of the fields near surfaces with minimal additional computational cost. Internal dielectrics may be specified, but the solver has historically only supported 2D geometries.

We have extended this multi-grid dielectric solver to work in 3D, and enabled parallel execution of the solver using Warp's standard MPI decomposition scheme. We have performed benchmarks against analytically solvable systems, for instance a dielectric sphere placed between a cathode and anode held at fixed potential. Figure 7 shows a slice of the solution to the electrostatic potential, along with a lineout comparing the result to theory.

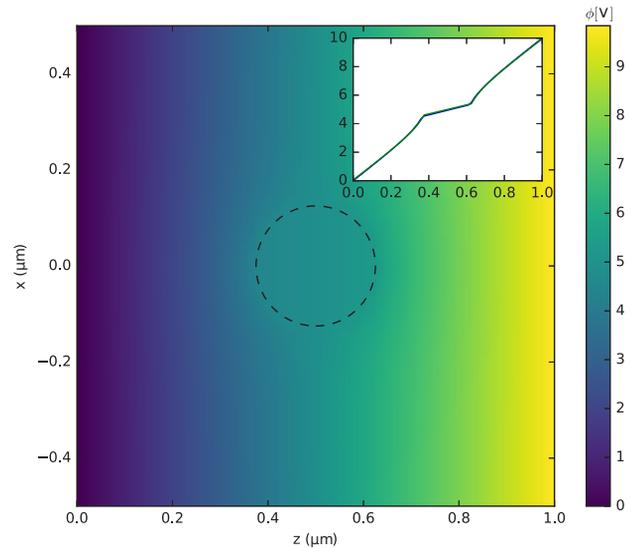


Figure 7: A 2D slice of the simulated potential between two parallel plates with a dielectric sphere placed in-between. Inset, a central lineout shows good agreement with theory.

In addition to extending the available geometries, we have also added the capability of dielectric structures to capture charged particles on their surface. Due to the limited charge mobility, dielectric structures in a TEC may slowly become charged from electron collisions, leading to the development of local field perturbations along the structure's surface. These fields may alter particle trajectories, changing the steady-state operation of the system. Warp has been outfitted with the capability to tag captured particles on dielectric surfaces. These particles remain fixed at the location of their intersection with the surface, and contribute to the surrounding electrostatic potential. Figure 8 shows the effective charging of a slab of dielectric resulting from the impacts of incoming electrons.

## CONCLUSION

We have developed self-consistent tools using the Warp code to model thermionic energy converters. We then implemented and tested a well-established efficiency model for

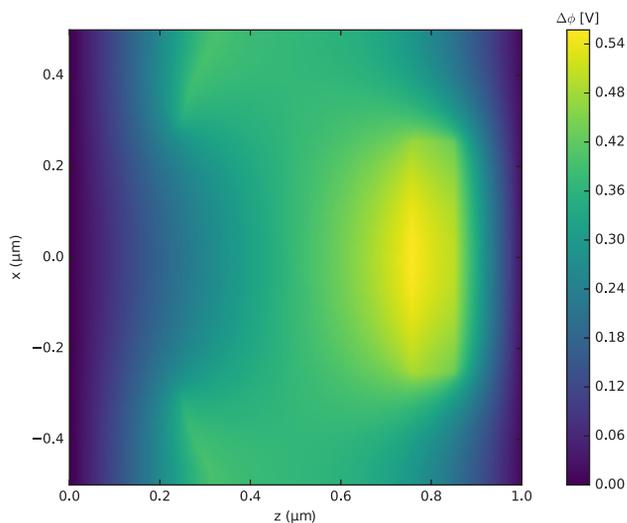


Figure 8: The dielectric particle implementation permits the buildup of surface charge due to impacts by electrons.

computing the individual loss channels and overall efficiency of TEC devices. An external circuit model was also included, which adjusts the potential at the collector to provide realistic feedback. These models have proven effective in identifying design guidelines for basic TEC structures. Further

improvements have been made to Warp to support internal dielectrics, which are necessary for supporting structures in realistic devices, and may alter steady state operation.

## ACKNOWLEDGMENTS

We would like to acknowledge the Warp development team, from the Accelerator Technology & Applied Physics group at Berkeley Lab. We would also like to thank Tony Pan, Eric Clark, Arvind Kannan, and Andrew Koch from Modern Electron, LLC for helpful discussions.

## REFERENCES

- [1] N. M. Cook *et al.*, “Self-Consistent Simulation and Optimization of Space-Charge Limited Thermionic Energy Converters”, in *Proc. IPAC’18*, Vancouver, BC, Canada, Apr-May 2018, doi:10.18429/JACoW-IPAC2018-MOPML060
- [2] J.-L. Vay *et al.*, *Computational Science & Discovery*, vol. 5, pp. 014019, 2012.
- [3] A. Schock, *Journal of Applied Physics*, vol. 32, pp. 1564, 1961.
- [4] S. Meir *et al.*, *Journal of Renewable and Sustainable Energy*, vol. 5, pp. 043127, 2013.
- [5] G. Shortley and R. Weller, “The numerical solution of Laplace’s equation”, *Journal of Applied Physics*, vol. 9, pp. 334-348, 1938. doi:10.1063/1.1710426

# PARTICLE-IN-CELL SIMULATION OF A BUNCHED ELECTRONS BEAM ACCELERATION IN A $TE_{113}$ CYLINDRICAL CAVITY AFFECTED BY A STATIC INHOMOGENEOUS MAGNETIC FIELD\*

E. A. Orozco<sup>†</sup>, Universidad Industrial de Santander, A.A. 678 Bucaramanga, Colombia  
V. E. Vergara, J. D. González, J. R. Beltrán,  
Universidad del Magdalena, A.A.731 Santa Marta, Colombia

## Abstract

The results of the relativistic full electromagnetic Particle-in-cell (PIC) simulation of a bunched electrons beam accelerated in a  $TE_{113}$  cylindrical cavity in the presence of a static inhomogeneous magnetic field are presented. This type of acceleration is known as Spatial AutoResonance Acceleration (SARA). The magnetic field profile is such that it keeps the electrons beam in the acceleration regime along their trajectories. Numerical experiments of bunched electrons beam with the concentrations in the range  $10^8$ – $10^9$   $\text{cm}^{-3}$  in a linear  $TE_{113}$  cylindrical microwave field of a frequency of 2.45 GHz and an amplitude of 15 kV/cm show that it is possible to accelerate the bunched electrons up to energies of 250 keV without serious defocalization effect. A comparison between the data obtained from the full electromagnetic PIC simulations and the results derived from the relativistic Newton-Lorentz equation in a single particle approximation is carried out. This acceleration scheme can be used as a basis to produce hard x-ray.

## INTRODUCTION

The last decades, particle accelerators based on the electron cyclotron resonance (ECR) phenomenon has been extensively studied. Different technological applications based on this phenomenon has been proposed [1–5]. There are different ways to maintain the ECR condition, which use: (i) Transversal electromagnetic (TEM) waves in a homogeneous magnetostatic field [6, 7], (ii) Transversal electric (TE) waves in waveguides placed on inhomogeneous magnetostatic field [8, 9], (iii) TE standing electromagnetic waves in cavities affected by a homogeneous magnetic field growing slowly in time, known as GYRAC [10, 11] or (iv) TE standing electromagnetic waves in cavities affected by an inhomogeneous magnetostatic field, known as SARA [12–16]; among others [17]. In the SARA concept the magnetostatic field is fitted along the resonant cavity axis to keep the ECR acceleration regime as the electrons move in helical trajectories. The SARA concept has been studied both analytically and numerically in cylindrical  $TE_{11p}$  cavities [12–15] as well as in a  $TE_{112}$  rectangular cavities [16]. An X ray source based on the SARA concept has been certificated [18].

In the present paper, the influence of the self-consistent field on the space autoresonance acceleration (SARA) of

bunched electrons beams in the linear  $TE_{113}$  cylindrical cavity is analyzed, by using a full electromagnetic relativistic particle-in-cell code. In our numerical scheme, the simulation is carried out in two stages:

1. Calculation of the  $TE_{113}$  steady-state microwave field before injecting the electrons bunched
2. Self-consistent simulation of the bunched electrons beams in the SARA acceleration.

The cylindrical  $TE_{113}$  cavity, whose radius and length are 4.54 cm and 30 cm respectively, is excited by a 2.45 GHz source. In our numerical model, to excite the  $TE_{113}$  microwave field of 15 kV/cm tension, an input power of 728 kW is injected into the cavity through a  $TE_{10}$  waveguide. The electron's bunched, whose concentrations are in the range  $10^8$ – $10^9$   $\text{cm}^{-3}$ , are described in the framework of the Vlasov-Maxwell equation; which is solved numerically through the particle-in-cell (PIC) method [19].

The obtained results show that it is possible to accelerate bunched electrons up to energies of 260 keV without serious defocalization effect. A comparison between the data obtained from the full electromagnetic PIC simulations and the results derived from the relativistic Newton-Lorentz equation in a single particle approximation [15] is carried out. This acceleration scheme can be used as a basis to produce hard x-ray.

## THEORETICAL FORMALISM AND NUMERICAL METHOD

### *Physical Scheme and Theoretical Formalism*

The electron acceleration in the autoresonance regime by a standing transversal electric microwave field in an inhomogeneous magnetostatic field, known as *Spatial AutoResonance Acceleration (SARA)*, can be realized in the physical system shown in Fig. 1.

The cylindrical cavity 1 is placed inside the current coil set 2 that produces an azimuthally symmetric magnetostatic field whose value at the end where is the electron gun 5 is the corresponding to obtain classical resonance. The magnetostatic field profile has a relation with the used  $TE_{11p}$  ( $p = 1, 2, 3, \dots$ ) mode 4, which is excited through the microwave port 3. The electrons gun 5 injects electrons by one end of the cavity 1 along the magnetostatic field axis, taken as  $z$  axis. The right-hand polarized electric field component of the microwave field accelerates the electrons by electron

\* Work supported by Universidad Industrial de Santander (Colombia) and Universidad del Magdalena (Colombia)

<sup>†</sup> eaorozco@uis.edu.co

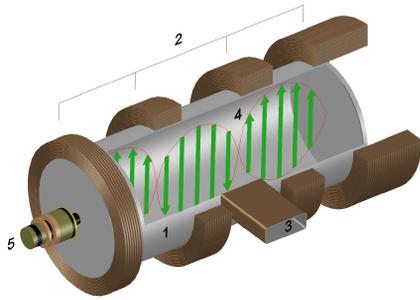


Figure 1: A physical model scheme. 1) cavity, 2) magnetic coils, 3) microwave port, 4) electric field profile; the particular case of  $TE_{113}$  mode, 5) electron gun.

cyclotron resonance (ECR) along their helical paths until they impact the opposite end of the cavity.

For a single particle and by using the paraxial approximation for the fields, the local electron cyclotron frequency  $\omega_c(\vec{r})$  in the SARA concept is given by [12]:

$$\begin{aligned} \omega_c(z)/\omega = & \gamma^{-1} B_z(0, z)/B_0 \\ & + \gamma^{-1} (E_0^c/B_0 c) [1 - \gamma^{-2} + (v_z/c)^2]^{-1/2} \\ & \times |\sin(p\pi z/L_c)| \sin \varphi \end{aligned} \quad (1)$$

where,  $\omega$  is the microwave field frequency;  $\gamma$  is the Lorentz factor;  $B_z(0, z)$  is the magnetostatic field profile along the cavity axis, whose value at the injection point is  $\gamma_0 B_0$ , being  $\gamma_0$  the Lorentz factor in said point and  $B_0 = m_e \omega/e$  is the magnetic field to obtain classical resonance ( $m_e$  and  $e$  are the mass and electric charge of the electron, respectively);  $E_0^c$  is the tension of the right-hand circular polarized component of the electric microwave field;  $c$  is the speed of light;  $v_z$  is the longitudinal component of electron velocity;  $p$  the index of the  $TE_{11p}$  mode;  $z$  the longitudinal coordinate of the electron;  $L_c$  the length of the cavity and finally,  $\varphi$  is the phase-shift between the electron transversal velocity and the electric field component of the microwave field.

A continuous sustenance of the exact resonance is possible only in the particular case of  $p = 1$ , because, if  $p \neq 1$ , the phase-shift  $\varphi$  jumps an angle  $\pi$  in each node of the standing electromagnetic wave. For the exact resonance  $\varphi = \pi$ , equation (1) leads to

$$\omega_c(z) = eB_z(0, z)/\gamma m_e \quad (2)$$

Therefore; to maintain the resonance condition  $\omega = \omega_c$ , the magnetostatic field has to be fitted to compensate the increasing of the relativistic factor as the electrons gain energies along its helical paths. In this case the magnetostatic field grows monotonously (see Fig. 2a).

For the case  $p \neq 1$ , the magnetostatic field grows in a non-monotonous way, which has to be fitted to maintain the phase-shift  $\varphi$  in the range  $\pi < \varphi < 3\pi/2$  (see the case  $p = 2$  in Fig. 2b). This range was named *Acceleration Band* because for these  $\varphi$  values the electromagnetic field can

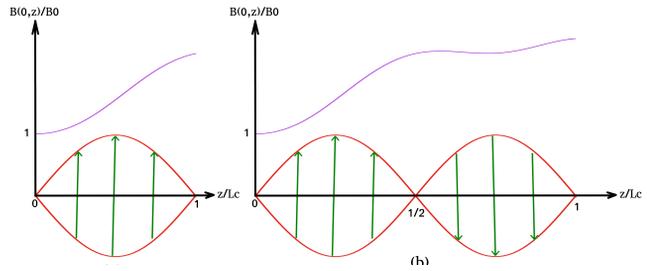


Figure 2: Typical magnetostatic field profiles (purple lines) used in the SARA concept for the modes (a)  $TE_{111}$  and (b)  $TE_{112}$ .

transfers energy to the electrons [12]. It is worth mentioning that in the SARA concept there is present the diamagnetic force, which is one of the important factors limiting the energy which can be achieved in this acceleration mechanism. In order to analyze the influence of the space-charge on the acceleration efficiency, a self-consistent simulation should be considered; being the Particle-in-cell (PIC) the most popular method used for the electrons beam simulations [19].

### Numerical Method

To simulate the proposed system, a numerical scheme based on two sequential stages is used:

1. Calculation of the steady state for the microwave field before to inject the electrons beam
2. Self-consistent simulation of the bunched electrons beams in the SARA acceleration by the  $TE_{113}$  cylindrical microwave field

In our simulations, the perfect electric conductor (PEC) boundary conditions for both the cavity and the waveguide coupled to the microwave port are used. To avoid nonphysical reflections, a perfectly matched layer (PML) in the opposite end of said waveguide is used (see Fig. 3). To simulate the input power, the  $TE_{10}$  mode is excited in a plane adjacent to the PML into the rectangular waveguide (see Fig. 3). In order to calculate the electric and magnetic field on the mesh points we use the Uniaxial perfectly Matched Layer (UPML) method; which solve the Maxwell equations in a finite difference time domain (FDTD) scheme based on a Yee's cell for systems including PML [20, 21].

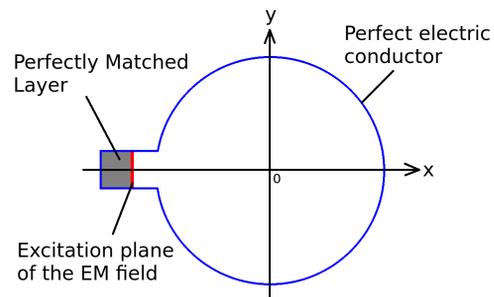


Figure 3: Waveguide-resonant cavity cross section.

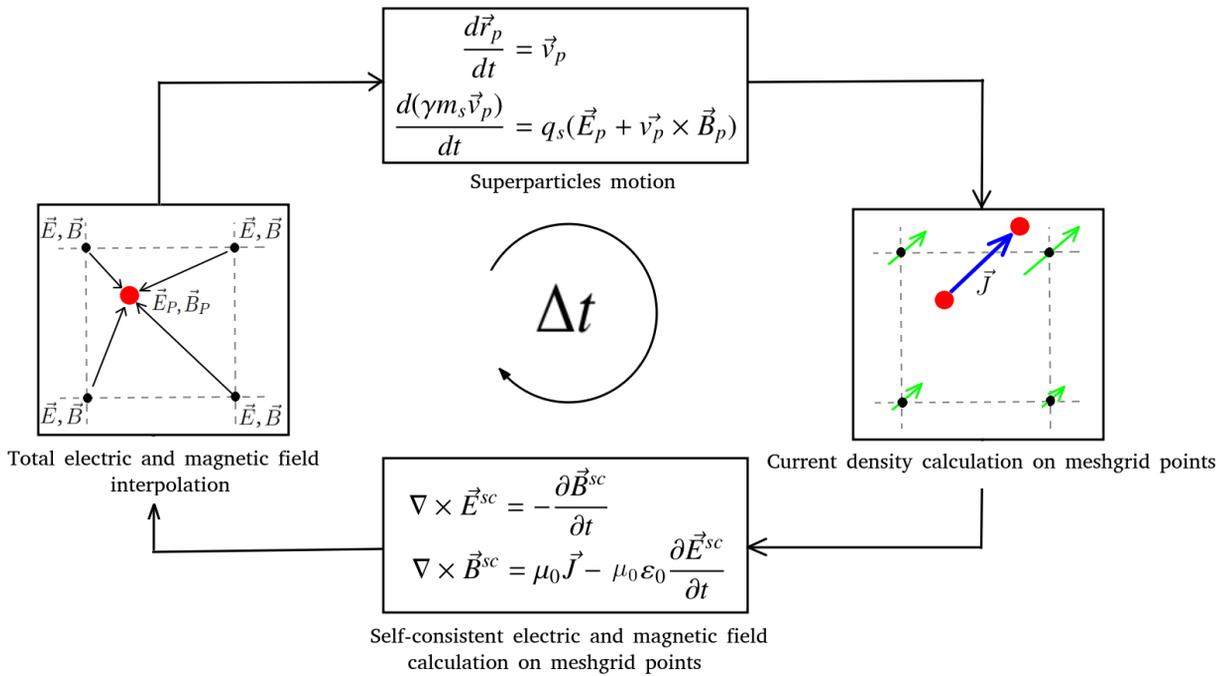


Figure 4: Electromagnetic PIC-algorithm.

For the second stage, the full electromagnetic particle-in-cell (PIC) method is used. In this method, groups of particles close to each other in the phase-space called superparticles (SP), are used to describe the evolution of the distribution function  $f_e(\vec{r}, \vec{v}, t)$  [19].

The electromagnetic PIC-algorithm in a computational cycle is showed in Fig. 4, which involves the following steps:

- (i) Calculation of the current densities in the mesh points (green arrows) from the superparticles (SP) positions and velocities data (blue arrow). In the present work, the current density is calculate using the conservative charge method proposed by Umeda et al in order to fulfill the continuity equation [22].
- (ii) Calculation of the self-consistent field on the meshgrid points from the current density. In the present work,  $\vec{E}^{sc} = \vec{E}^{hf} + \vec{E}^{sg}$ , where  $\vec{E}^{hf}$  is the microwave electric field component and  $\vec{E}^{sg}$  is the self-generated electric field by the electrons bunch. Similarly, the self-consistent magnetic field component,  $\vec{B}^{sc}$ , is defined.
- (iii) Calculation of the total fields,  $\vec{E}_p$  and  $\vec{B}_p$ , acting on the superparticles. These fields are calculated through the interpolation of the total fields on the mesgrid points (see Fig. 4). In the present work,  $\vec{E} = \vec{E}^{sc}$  and  $\vec{B} = \vec{B}^{sc} + \vec{B}^s$ , where  $\vec{B}^s$  is the magnetostatic field showed in Fig. 5.
- (iv) Calculation of new positions and velocities of the SPs through integration of their equations of motion. For this step, the relativistic Newton-Lorentz equation is

solved numerically through the Boris leapfrog procedure.

In our numerical simulations we consider a 2.45 GHz cylindrical cavity, whose radius and length are 4.54 cm and 30 cm respectively. To excite a microwave field of 15 kV/cm tension, an input power of 728 kW is injected into the cavity through a  $TE_{10}$  rectangular waveguide. It is worth mentioning that such high level of the microwave power is because a non-optimized microwave injection system has been used.

The magnetostatic field profile shown in Fig. 5, where  $B_0(= m_e \omega / e) = 0.0875$  T, is generated by four axisymmetric coils whose parameters are given in Table 1, where  $R_i$ ,

Table 1: Magnetic Coil System Parameters

Coil	$R_i$	$R_e$	$L_b$	$J$	$z$
1	6 cm	20 cm	6 cm	1.39 A/mm <sup>2</sup>	-5.75 cm
2	6 cm	20 cm	7.5 cm	1.08 A/mm <sup>2</sup>	8.25 cm
3	6 cm	20 cm	6.9 cm	1.18 A/mm <sup>2</sup>	19.5 cm
4	6 cm	20 cm	6.1 cm	2.07 A/mm <sup>2</sup>	32 cm

$R_e$ ,  $L_b$ , and  $z$  are the internal radius, the external radius, the width of each coil, and the positions of the coils, respectively, and  $J$  is the coil current density.

In order to analyze the influence of the space charge on the spatial autoresonance acceleration, simulations with spherical electrons bunches were carried out and divided in two cases (see Table 2).

The simulations are considered finished when the electrons impact with the cavity.

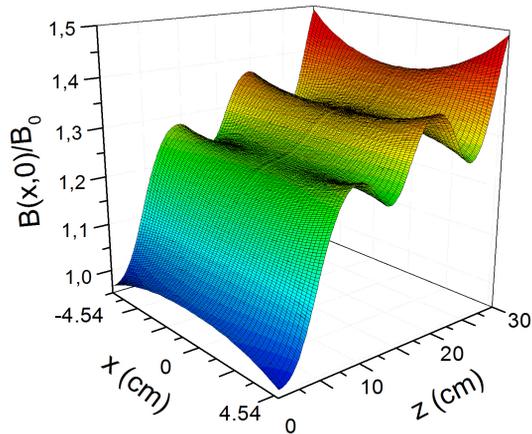


Figure 5: The profile of the magnetostatic field in the  $y = 0$  plane.

Table 2: Parameters of the Simulations

	case 1	case 2
Beam parameters		
Electron Bunch Radius	0.5 cm	0.5 cm
Electron concentration	$n_e = 10^8 \text{ cm}^{-3}$	$n_e = 10^9 \text{ cm}^{-3}$
Injection energy	30 keV	32 keV
Simulation parameters		
$\Delta x$	0.07 cm	0.07 cm
$\Delta y$	0.07 cm	0.07 cm
$\Delta z$	0.3 cm	0.3 cm
$\Delta t$	1.58 ps	1.58 ps
PiC merging factor	$2 \times 10^4$	$2 \times 10^5$

## RESULTS AND DISCUSSION

Figure 6 shows the obtained steady-state electric field distribution in the cross section  $z = L_c/2$  (see Fig. 6a), the longitudinal plane  $y = 0$  (see Fig. 6b) and the longitudinal plane  $x = 0$  (see Fig. 6c) for the first stage of the simulation. This graphics show good agreement with the obtained from the well known analytical expressions of the linear polarized  $TE_{113}$  mode whose amplitude is  $E_0^l = 15 \text{ kV/cm}$ . The electrons interacts effectively only with the right-hand polarized electric field component of the microwave field, which has an amplitude  $E_0^c = E_0^l/2$ . Figure 7 shows the time evolution of the phase-shift between the electrons transversal velocities and the right-hand circular polarized component of the electric microwave field, for the case 1 (see Table 2) and for the single-particle approximation studied in [15]. In this graph, the *Acceleration band* is shown in blue color.

It can be noted that at the injection point there are present all the possible values for the phase-shift. This happens because the electric field has a node in such a point while the self-consistent field pushes outward the electrons in all radial directions. The phase-shift acquire mostly the values

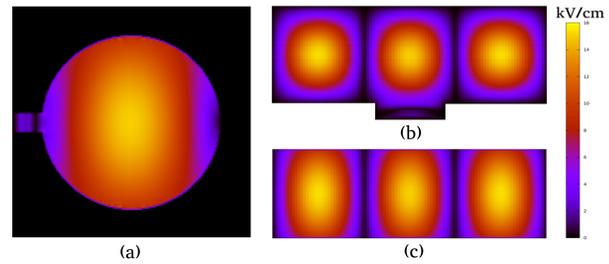


Figure 6: Steady-state electric field distribution in (a) the cross section  $z = L_c/2$ , (b) the longitudinal plane  $y = 0$  and (c) the longitudinal plane  $x = 0$ .

around of the value  $\pi/2$  due to the deviation produced by the magnetic field component of the microwave field [12]. Then a fast phase-focalization occurs by the microwave field at the position  $z \approx 5 \text{ cm}$ , where all the electrons are close to the exact resonance,  $\varphi = \pi$ . We can see that the phase shift

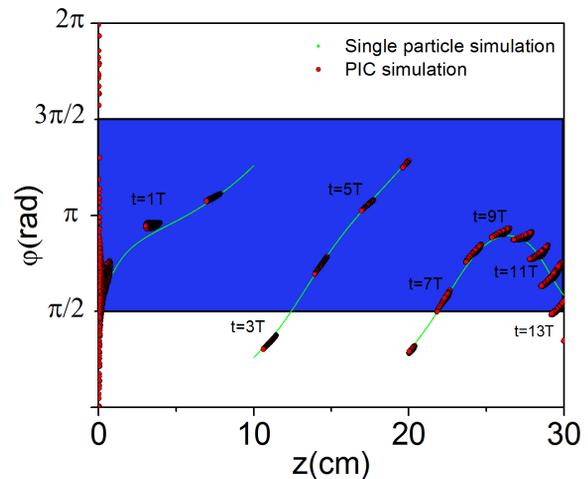


Figure 7: Time evolution of the phase-shift between the electrons transversal velocities and the right-hand circular polarized component of the electric microwave field. Red circles correspond to the case of  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched and the green line for the single particle approximation.

$\varphi$  jumps an angle  $\pi$  at the planes  $z = 10 \text{ cm}$  and  $z = 20 \text{ cm}$ , where the  $TE_{113}$  microwave electric field has nodes (see Fig. 6). These jumps don't remove significantly the phase-shift from the *Acceleration band*  $\pi/2 < \varphi < 3\pi/2$ ; therefore the electrons energies grow monotonously, except for the regions  $10 \text{ cm} \lesssim z \lesssim 12.5 \text{ cm}$  and  $20 \text{ cm} \lesssim z \lesssim 22.5 \text{ cm}$  where the  $\varphi$  values are outside of the *Acceleration band* (see Fig. 8). From this figure we can see that there is not any significant difference between the energy evolution of both the electrons bunched and the single electron. It can be noted in Fig. 7, Fig. 8 and Fig. 9 that its projections onto the  $z$ -axis are intervals of the width about of 1 cm, the diameter of the electrons bunched. For this case, the self-consistent field

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

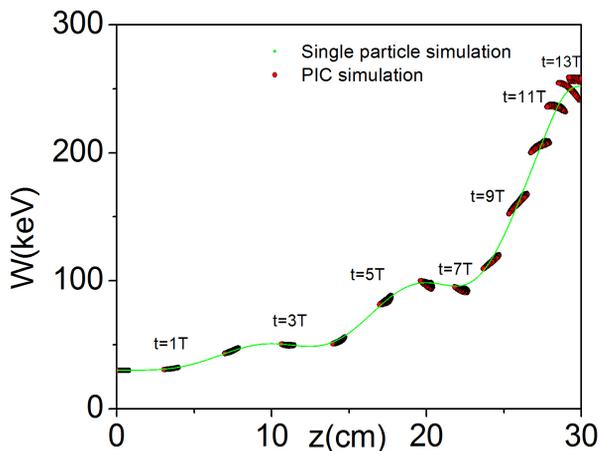


Figure 8: Time evolution of the energy for the  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched (red circles) and for the single particle approximation (green line).

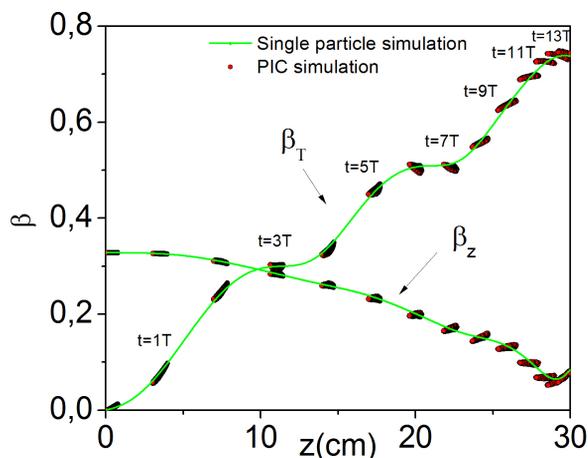


Figure 9: Time evolution of the transversal ( $\beta_T = v_T/c$ ) and longitudinal ( $\beta_z = v_z/c$ ) velocities for the  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched (red circles) and for the single particle approximation (green line).

does not significantly affect the electron-beam focalization (see Fig. 7) and the energy spread of the electrons impacting on the wall is found not greater than 8% (see Fig. 8).

The evolutions of the transversal and longitudinal velocity components are shown in Fig. 9. It can be noted the diamagnetic force effect on the longitudinal velocity, which decrease until the value  $0.08 c$  when the electrons impact to the cavity. In such position the transversal velocity of the electrons is maximum.

For case 2 (see Table 2), the bunch evolution is found similar to described in case 1; except for times greater than 5 microwaves period (see Figs. 8 and 11). This effect is

attributed to the self-generated electric field,  $\vec{E}^{sg}$ ; which is 10 times more intense than for the case 1. The self-generated electric field tends to expand the electrons bunch in all radial directions; however, in the transversal plane, the combined effect of the microwave electric field component  $\vec{E}^{hf}$  and the magnetostatic field contribute to the transversal confinement of the electrons bunch. On the contrary, in the axial direction, there is not any confinement mechanism for the electrons bunch; which causes its widening in such direction. It can be appreciated in Figs. 8 and 11 that its projections onto the  $z$ -axis are intervals of the width about of 3 cm and 4 cm, for the instants  $t = 7$  and  $t = 9$  microwaves periods, respectively. Such widening causes a spread in the velocities

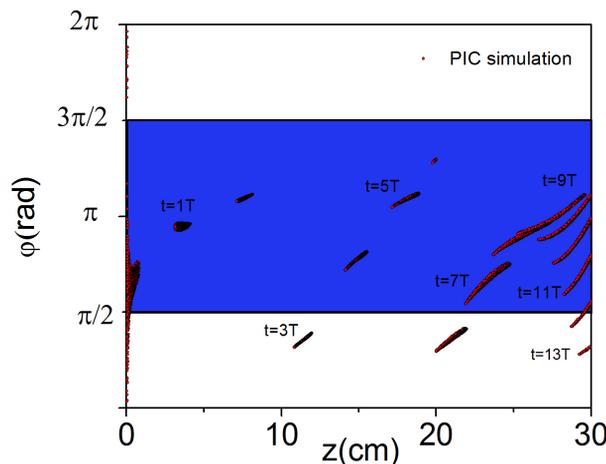


Figure 10: Time evolution of the phase-shift between the electrons transversal velocities and the right-hand circular polarized component of the electric microwave field for the  $n_e = 10^9 \text{ cm}^{-3}$  electrons bunched.

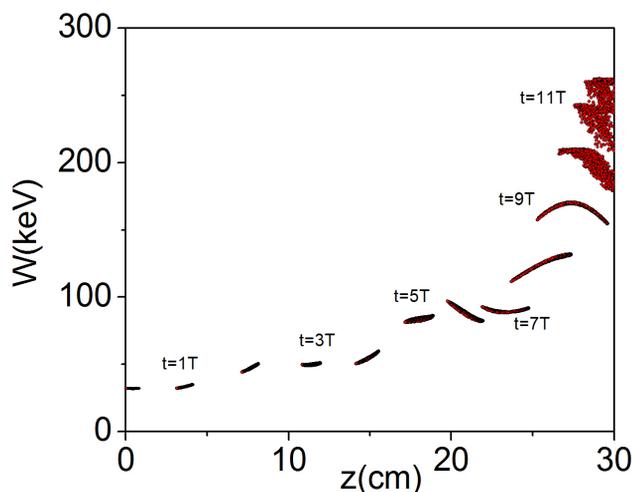


Figure 11: Time evolution of the energy for the  $n_e = 10^9$  electrons bunched.

of the electrons move in different acceleration conditions (see Fig. 10 for  $t \geq 9$  microwaves period). The energy spread observed in Fig. 11 is caused by this effect.

We can note that the spread in the velocity is found similar for both, the longitudinal and transverse velocity (see Fig. 12).

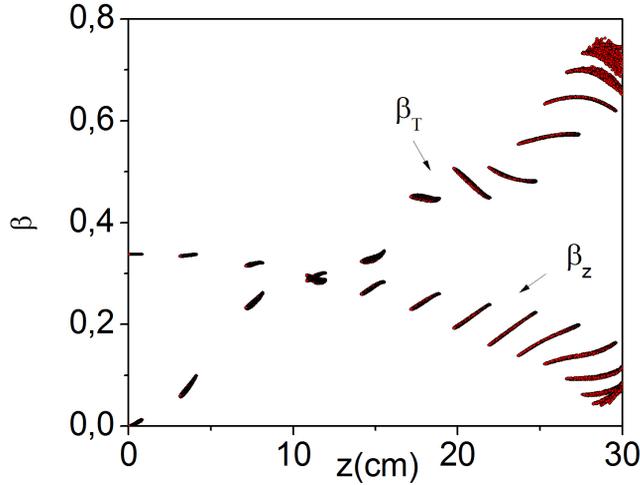


Figure 12: Time evolution of the transversal ( $\beta_T = v_T/c$ ) and longitudinal ( $\beta_z = v_z/c$ ) velocities for the  $n_e = 10^9$  electrons bunched.

Figure 13 shows the energy spectrum of the electrons that impact with the opposite wall of the cavity at the position  $z = L_c$  obtained from our numerical simulations for the two cases considered (see Table 2). We can note that the energy spread for the  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched is not greater than 8%.

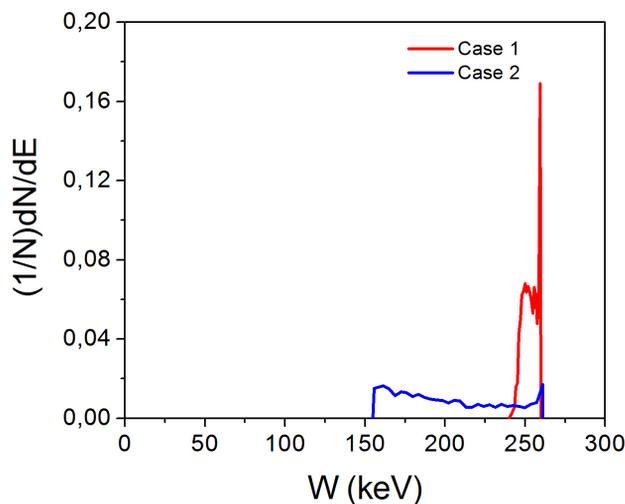


Figure 13: Numerical predictions of the energy spectrum for the electrons impacting on the cavity wall,  $z_{\text{wall}} = L_c$ , for the  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched (red line) and for the  $n_e = 10^9 \text{ cm}^{-3}$  electrons bunched (blue line).

## CONCLUSION

The realized numerical experiment shows that electrons bunched can be accelerated up to energies of 250 keV in spatial autoresonance acceleration conditions by using a  $TE_{113}$  mode. It was shown that for the  $n_e = 10^8 \text{ cm}^{-3}$  electrons bunched there is not present serious defocalization effect. For the  $n_e = 10^9 \text{ cm}^{-3}$  electrons bunched, the self generated electric field spread it in longitudinal direction, which affect the acceleration regime. However, this effect can be reduced by using a continuous electron beam in the injection process.

## REFERENCES

- [1] H. Gardner, T. Ohkawa, A. Howald, A. Leonard, L. Peranich, and J. D'Aoust, "An inexpensive x-ray source based on an electron cyclotron", *Review of Scientific Instruments*, vol 61, no 2, pp. 724-727, 1990.
- [2] R. Baskaran and T. Selvakumar, "Studies on enhancement of x-ray flux in the compact electron cyclotron resonance plasma x-ray source", *Review of Scientific Instruments*, vol 71, no 2, pp. 1203-1205, 2000.
- [3] R. Symons, J. Hirshfield, and C. Wang, "Multi-stage cavity cyclotron resonance accelerator", U.S. Patent and Trademark Office No. 6,914,396, Jul. 31, 2005.
- [4] T. Inoue, T. Hattori, S. Sugimoto, and K. Sasai, "Design study of electron cyclotron resonance-ion plasma accelerator for heavy ion cancer therapy", *Review of Scientific Instruments*, vol 85, no 2, p. 02A958, 2014.
- [5] V. Andreev, D. Chuprov, V. Ilgisonis, A. Novitsky, and A. Umnov, "Gyromagnetic autoresonance plasma bunches in a magnetic mirror", *Physics of Plasmas*, vol 24 no 9, p. 093518, 2017.
- [6] A. Kolomenskii and A. Lebedev, "Self-Resonant Particle Motion in a Plane Electromagnetic Wave", *Doklady Akademii Nauk*, vol 145, no 6, pp. 1259-1261, 1962.
- [7] V. Davydovskii, "Possibility of Resonance Acceleration of Charged Particles by Electromagnetic Waves in a Constant Magnetic Field", *Sov. Phys. JETP*, vol 16, pp. 629-630, 1963.
- [8] R. Shpitalnik, C. Cohen, F. Dothan, and L. Friedland, "Autoresonance microwave accelerator", *Journal of applied physics*, vol 70 no 3, pp. 1101-1106, 1991.
- [9] L. Friedland, "Spatial autoresonance cyclotron acceleration", *Physics of plasmas*, vol 1, no 2, pp. 421-428, 1994.
- [10] K. Golovanivsky, "Autoresonant acceleration of electrons at nonlinear ECR in a magnetic field which is smoothly growing in time", *Physica Scripta*, vol 22, no 2, pp. 126-133, 1980.
- [11] O. Gal, "GYRAC: a compact, cyclic electron accelerator", *IEEE Transactions on Plasma Science*, vol 17, no 4, pp. 622-629, 1989.
- [12] V. Dugar-Zhabon and E. Orozco, "Cyclotron spatial autoresonance acceleration model", *Physical Review Special Topics-Accelerators and Beams*, vol 12, no 4, p. 041301, 2009.
- [13] V. Dugar-Jabon, E. Orozco, and A. Umnov, "Modeling of electron cyclotron resonance acceleration in a stationary inhomogeneous magnetic field", *Physical Review Special Topics-Accelerators and Beams*, vol 11, no 4, p. 041302, 2008.

- Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.
- [14] V. Dugar-Zhabon, J. González, and E. Orozco, “3D electromagnetic simulation of spatial autoresonance acceleration of electron beams”, in *Journal of Physics: Conference Series*, vol 687, no 1, p. 012077, 2016.
- [15] V. Vergara, J. González, J. Beltrán, and E. Orozco, “Electrons acceleration in a  $TE_{113}$  cylindrical cavity affected by a static inhomogeneous magnetic field”, in *Journal of Physics: Conference Series*, vol 935, no 1, p. 012076, 2017.
- [16] V. Dugar-Zhabon, E. Orozco, and A. Herrera, “Self-consistent simulation of an electron beam for a new autoresonant x-ray generator based on  $TE_{102}$  rectangular mode”, in *Journal of Physics: Conference Series*, vol 687, no 1, p. 012076, 2016.
- [17] V. Milant’ev, “Cyclotron autoresonance—50 years since its discovery”, *Physics-Uspekhi*, vol 56, no 8, pp. 823-832, 2013.
- [18] V. Dugar-Zhabon and E. Orozco, “Compact self-resonant X-ray source”, U.S. Patent and Trademark Office No. 9,666,403, May 30, 2017.
- [19] R. Hockney and J. Eastwood, *Computer simulation using particles*. Bristol: Hilger, 1988.
- [20] A. Taflove and S. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method in Electromagnetics*. Norwood, MA: Artech House, 2000.
- [21] K. Yee, “Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media”, *IEEE Transactions on antennas and propagation*, vol 14, no 3, pp. 302-307, 1966.
- [22] T. Umeda, Y. Omura, T. Tominaga, and H. Matsumoto, “A new charge conservation method in electromagnetic particle-in-cell simulations”, *Computer Physics Communications*, vol 156, no 1, pp. 73-85, 2003.

# SPARSE GRID PARTICLE-IN-CELL SCHEME FOR NOISE REDUCTION IN BEAM SIMULATIONS

L. F. Ricketson\*, Lawrence Livermore National Laboratory, Livermore, CA, USA

A. J. Cerfon, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

## Abstract

We demonstrate that the sparse grid combination technique, a scheme originally designed for grid based solvers of high-dimensional partial differential equations, can be effectively applied to reduce the noise of Particle-in-Cell (PIC) simulations. This is because the sparse grids used in the combination technique have large cells relative to a comparable regular grid, which, for a fixed overall number of particles, increases the number of particles per cell, and thus improves statistical resolution. In other words, sparse grids can accelerate not only the computation of the electromagnetic fields, but also the particle operations, which typically dominate the computation and storage requirements.

## INTRODUCTION

In charged beams in particle accelerators, the Coulomb collision frequency is much smaller than the other frequencies of interest, even at the highest achievable beam intensities. A kinetic description of the beam is therefore required, in which one solves for the beam distribution function giving the number of particles in an infinitesimal six-dimensional phase space volume. Because of the high dimensionality of the phase space volume, intense beam simulations are very computationally intensive; even a modest grid resolution for each of the six dimensions pushes the limits of today's largest supercomputers.

To circumvent this difficulty, particle based approaches to the problem have been widely adopted, usually in the form of the Particle-in-Cell (PIC) algorithm [1]. The PIC method has the advantages of being conceptually intuitive, being well-suited for massive parallelization, and only requiring discretization of configuration space. Detailed PIC simulations of intense charged beams are routinely run and relied upon to explain experimental results and to design new accelerators [2, 3]. Even so, the accuracy of these PIC simulations remains limited due to the probabilistic nature of the PIC scheme, which requires a large number of particles to be simulated in order to reduce statistical noise. We will indeed show in this article that the slow decay of this noise with the number of simulated particles implies that for a given target accuracy, standard PIC simulations may even be more computationally intensive than grid based simulations.

In this work, we present a new algorithm which addresses this unsatisfying state of affairs by reducing the noise in PIC simulations. The algorithm is based on the sparse grid combination technique, a method originally intended to address the poor scaling of grid based PDE solvers with dimension [4]. We will explain how to apply the combination

technique in a PIC setting, and demonstrate its promise by using our algorithm to solve standard problems in plasma and beam physics. The structure of the article is as follows. In the first Section, we compare the asymptotic run time complexity of a standard grid based kinetic solver and a standard PIC solver, and arrive at the conclusion that noise reduction strategies need to be implemented for the PIC approach to be desirable from a complexity point of view. In the second Section, we briefly present the sparse grid combination technique in the simple yet enlightening context of linear interpolation. The combination technique is the central idea motivating our new scheme. In the third Section, we explain how the combination technique can indeed be favorably applied in the context of PIC solvers, and numerically demonstrate the significant reduction in noise from doing so in the fourth Section. The fifth Section focuses on the limitations of our new sparse PIC scheme in its current, somewhat naive implementation, and highlights directions for further improvement. We provide a brief summary of our work in the last Section.

## THE CURSE OF DIMENSIONALITY VS THE CURSE OF NOISE

It is well known that grid-based solvers for the kinetic equations describing beam evolution scale badly with the dimensionality of the problem. The computational complexity  $\kappa$  of a grid-based code can be expressed, in the best case scenario, as

$$\kappa \sim \frac{h^{-d}}{\Delta t},$$

where  $h$  is the grid size,  $\Delta t$  is the time step, and  $d$  is the dimension of the problem. If we consider a typical solver which would be second-order accurate in space and time, the numerical error  $\varepsilon$  scales like  $h^2$  and  $\Delta t^2$ , so numerical error  $\varepsilon$  and run-time complexity  $\kappa$  can be related through the scaling

$$\kappa \sim \varepsilon^{-\frac{d+1}{2}}.$$

The exponential dependence of  $\kappa$  on  $d$  is a major reason why continuum kinetic simulations are computationally intensive. It is often referred to as the *curse of dimensionality*.

Particle based algorithms such as the PIC algorithm address the curse of dimensionality by approximating the distribution function in terms of macro-particles which evolve in configuration space, which is at most three-dimensional. If  $N_p$  is the number of particles used in the simulation, we can write

$$\kappa \sim \frac{dN_p}{\Delta t}$$

\* ricketson1@llnl.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

At this point, it appears that the dependence of  $\kappa$  on dimension is much more favorable. However, the proper measure is to consider the complexity for a certain level of numerical error  $\varepsilon$ . The price one pays by adopting a particle based approach is in its inherent statistical noise, which only decreases with the square root of the number of simulated particles. Specifically, in a standard PIC scheme, if  $N_c$  is the number of cells used and  $d_X$  the number of dimensions of the configuration space,  $\varepsilon \sim (N_p/N_c)^{-1/2} \sim N_p^{-(1/2)} h^{-d_X/2}$ ,  $\varepsilon \sim h^2$ , and  $\varepsilon \sim \Delta t^2$ . Using these scalings in the formula for  $\kappa$ , one finds

$$\kappa \sim d\varepsilon^{-(2.5 + \frac{d_X}{2})}.$$

One notes that because of the slow convergence of the numerical error with the number of simulated particles, the scaling of  $\kappa$  is not favorable for the PIC scheme for small values of  $d$  and  $d_X$ . Even for  $d = 6$  and  $d_X = 3$ , which correspond to the largest values one needs to consider in beam physics,  $\kappa$  depends more strongly on  $\varepsilon$  in a standard PIC scheme than in a standard continuum scheme:  $\varepsilon^{-4}$  vs  $\varepsilon^{-7/2}$ . This is what we may call the *curse of noise*.

In the remainder of this article, we will present and analyze the performance of a numerical scheme we propose to address the curse of noise in PIC simulations. The algorithm is directly inspired from a numerical method known as the sparse grid combination technique, which was originally invented to tackle the curse of dimensionality of grid based solvers for partial differential equations [4]. As we will show, it can also be an efficient technique to address the curse of noise in particle based solvers.

## THE SPARSE GRID COMBINATION TECHNIQUE: ILLUSTRATION WITH INTERPOLATION

In this section, we present the fundamental idea underpinning the sparse grid combination technique by focusing on its application in a key step of the PIC scheme, namely the interpolation from values on a grid to points off the grid. For the simplicity of the presentation, we will consider a two-dimensional situation. However, the same ideas are applicable - and in fact more valuable - in three-dimensional situations.

We consider a standard case in which we know the values of the function  $u(x, y)$  on a Cartesian grid discretizing the square  $[-1, 1] \times [-1, 1]$ , with grid width and height  $h_x$  and  $h_y$ , and we wish to approximate  $u$  on the entire domain via bilinear interpolation. The error between the exact function  $u$  and its approximation  $u$  at a particular point off the grid is given by the *exact* formula [4]

$$u(x, y) - u(x, y) = C_1(h_x)h_x^2 + C_2(h_y)h_y^2 + C_3(h_x, h_y)h_x^2h_y^2. \quad (1)$$

where the  $C_i$  above are *functions* with a uniform upper bound. In the absence of additional information about  $u$ , one typically chooses  $h_x = h_y = h$  and finds an error of  $O(h^2)$ . Furthermore, the computational complexity  $\kappa$  of the optimal scheme scales linearly in the number of grid points, so that

$\kappa = O(h^{-2})$ . We conclude that  $\kappa$  scales with the numerical error  $\varepsilon$  according to  $\kappa \sim \varepsilon^{-1}$ . In arbitrary dimension  $D$ , the same reasoning leads to the optimal complexity  $\kappa \sim \varepsilon^{-D/2}$ , where we observe an exponential dependence on  $D$  - this is precisely the curse of dimensionality.

The combination technique is designed to reduce the strong dependence of  $\kappa$  on  $D$  by using cancellation across different "sparser" grids. Suppose the desired resolution is  $h_N = 2^{-N}$  for some positive integer  $N$ . Let  $h_x^i = 2^{-i}$ ,  $h_y^j = 2^{-j}$  and  $u_{i,j}$  be the approximation of  $u$  on the corresponding grid. Then, consider the quantity  $u_N$  defined by

$$u_N = \sum_{i+j=N+1} u_{i,j} - \sum_{i+j=N} u_{i,j}. \quad (2)$$

In each of the sums,  $i$  and  $j$  are strictly positive integers. The combination (2) is depicted graphically in Figure 1.

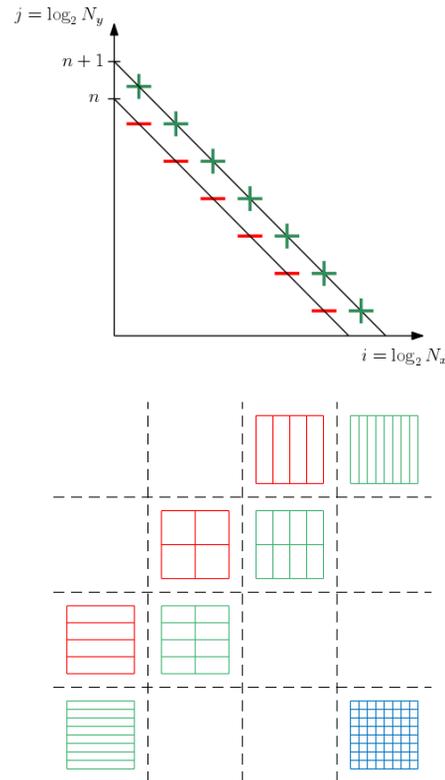


Figure 1: A graphical depiction of the combination of grids used in (2). (Top) The green '+' signs represent grids that give a positive contribution, while red '-' signs are subtracted. Cancellation arises from pairing neighboring grids along vertical and horizontal axes. (Bottom) An illustration of the sparse grids giving a positive contribution (in green), and the sparse grids giving a negative contribution (in red), for the case  $N = 4$ . The blue grid is the equivalent full grid one would use in a standard PIC simulation

By considering Figure 1 and the error formula (1), we see that a great deal of cancellation occurs in computing the error corresponding to  $u_N$ . Specifically, for any particular  $i$  be-

tween 1 and  $N-1$ , a grid with horizontal spacing  $h_x^i$  appears exactly once in each of the two sums in (2). For those two grids, the term  $C_1(h_x)h_x^2$  that appears in (1) cancels *exactly*, because it is independent of  $h_y$ . The only contribution from the  $O(h_x^2)$  term thus comes from the grid with  $h_x = 2^{-N}$ . Analogous reasoning for the y-direction leads to the result

$$u - u_N = C_1(h_N)h_N^2 + C_2(h_N)h_N^2 \quad (3)$$

$$+ h_N^2 \left\{ \frac{1}{4} \sum_{i+j=N+1} C_3(h_x^i, h_y^j) - \sum_{i+j=N} C_3(h_x^i, h_y^j) \right\}, \quad (4)$$

where we have used the fact that  $h_x^i h_y^j = h_N/2$  when  $i+j = N+1$  and  $h_x^i h_y^j = h_N$  when  $i+j = N$ . The expression in braces contains  $2N-1$  terms which are all uniformly bounded by constants, so we find that

$$|u - u_N| = O(Nh_N^2) = O(h_N^2 \log h_N). \quad (5)$$

In other words,  $u_N$  approximates  $u$  *nearly as well* as an approximate solution using  $h_x = h_y = 2^{-N}$ . This is illustrated in Figure 2, in which we plot the maximum error when interpolating the function  $u(x, y) = \sin(2\pi x) \cos(3\pi y)$  at 50 randomly located off-grid points in the square  $[-1, 1] \times [-1, 1]$ . For each value of  $N$ , the interpolation error for the 50 points are plotted, with black dots for the sparse grids technique, and red stars for standard linear interpolation. The results confirm our asymptotic scalings. We however note a drawback of the sparse grid combination technique, which is explicit in this figure and which we will return to later in this article: the leading term in the sparse grid error depends on  $C_3$ , which is proportional to the fourth-order mixed derivative  $u_{xxyy}$ . This is the reason the constant for the purple curve in Figure 2 is larger than the constant for the yellow curve, and highlights the fact that the sparse grid technique requires functions to be quite smooth to be more efficient than the standard scheme.

Now, to understand why this technique leads to a favorable run time complexity, observe that each grid used in the combination technique has  $O(h_N^{-1})$  grid points, and there are  $O(N)$  grids, implying the scaling  $\kappa = O(h_N^{-1} \log h_N)$ , which we can write in terms of the error  $\varepsilon$ :  $\kappa \sim \varepsilon^{-1/2} |\log \varepsilon|^2$ . For three dimensional problems, this scaling generalizes to [5]

$$\kappa \sim \varepsilon^{-1/2} |\log \varepsilon|^4. \quad (6)$$

We see that with sparse grids, the dimensionality of the problem affects the complexity of the algorithm only through a weak logarithmic dependence. At least asymptotically, one can thus achieve the same accuracy considerably faster with the combination technique than with a single regular grid.

## COMBINING SPARSE GRIDS WITH PIC

Sparse grids are not limited to interpolation, and can also be applied to another expensive stage in a standard PIC scheme, in which we assign a charge density to each macroparticle, through the introduction of a shape functions  $S$ . To see why this is so, note that the numerical error of approximating the true particle density  $\rho$  with the approximate

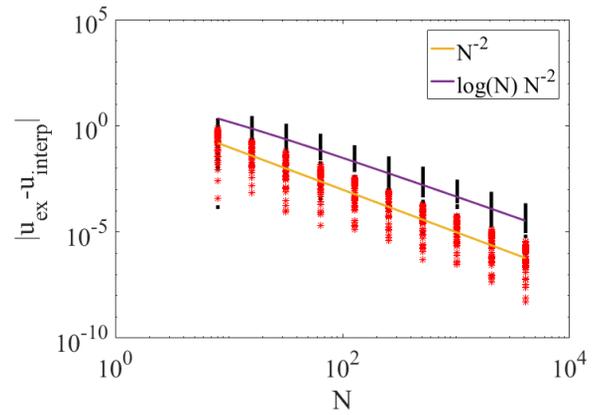


Figure 2: Numerical error as a function of the number of grid points  $N$  for the linear interpolation of the  $u(x, y) = \sin(2\pi x) \cos(3\pi y)$  at randomly located off-grid points in the square  $[-1, 1] \times [-1, 1]$ . The black dots correspond to the error obtained with the sparse grids combination technique, and the red stars correspond to the error obtained with standard interpolation. The asymptotic scalings  $O(N^{-2})$  and  $O(N^{-2} \log N)$  are shown with continuous lines for comparison.

density  $\rho$  can be written as [5]

$$\rho(\mathbf{x}_k) - \rho(\mathbf{x}_k) = C_1(h_x)h_x^2 + C_2(h_y)h_y^2 + C_3(h_x, h_y)h_x^2 h_y^2 + \xi_k, \quad (7)$$

where  $\xi_k$  is a random variable with  $\mathbb{E}[\xi_k] = 0$  and  $\text{Var}[\xi_k] \approx \frac{4Q\rho(\mathbf{x}_k)}{9} \frac{1}{h_x h_y N_p}$ , with  $Q$  the total charge of the beam. The first three terms on the right-hand side of Eq. (7) can be interpreted as the grid-based error of the scheme, while the last term is the particle sampling error. The grid based error has exactly the same form as in Eq. (1), so applying the combination technique to the evaluation of  $\rho$  will yield the same benefits as before for this contribution to the total error. Furthermore, we have shown in [5] that the sparse grids combination technique also eliminates the curse of dimensionality for the particle sampling error, which scales as  $\varepsilon_{\text{samp}} \sim |\log h_N|^{D-1} (N_p h_N)^{-1/2}$  in a sparse grids implementation of a problem with spatial dimension  $D$ . These results prompt us to consider the following sparse grids modification to the standard PIC algorithm:

- (i) Push particles exactly as in standard PIC.
- (ii) Assign to each particle a sequence of shape functions  $S_{i,j}(\mathbf{x} - \mathbf{x}_p) = \tau(2^i(x - x_p))\tau(2^j(y - y_p))/2^{i+j}$ , where  $\tau$  is the ‘‘hat’’ function such that  $\tau(x) = 1 - |x|$  if  $|x| \leq 1$  and  $\tau(x) = 0$  otherwise, and approximate the overall charge density via

$$\rho \approx \varrho = \sum_{i+j=n+1} \varrho_{i,j} - \sum_{i+j=n} \varrho_{i,j}, \quad (8)$$

where  $\varrho_{i,j}$  is defined at grid points  $\mathbf{x}_{k,\ell} = (k2^{-i}, \ell2^{-j})$  by

$$\varrho_{i,j}(\mathbf{x}_{k,\ell}) = \frac{Q}{N_p} \sum_p S_{i,j}(\mathbf{x}_p - \mathbf{x}_{k,\ell}) \quad (9)$$

and extended to the entire domain using bilinear interpolation.

- (iii) Use a grid-based Poisson solver to compute  $\varphi_{i,j}$  and  $\mathbf{E}_{i,j}$  by solving  $-\nabla^2 \varphi_{i,j} = \rho_{i,j}$
- (iv) Evaluate  $\mathbf{E}$  at the particle positions  $\mathbf{x}_p$  via

$$\mathbf{E}(\mathbf{x}_p) = \sum_{i+j=n+1} \mathbf{E}_{i,j}(\mathbf{x}_p) - \sum_{i+j=n} \mathbf{E}_{i,j}(\mathbf{x}_p). \quad (10)$$

Use this to repeat step i.

The run time complexity  $\kappa$  in terms of the error  $\varepsilon$  of this sparse-PIC algorithm is  $\kappa \sim D \frac{N_p}{\Delta t} \sim D \varepsilon^{-3} |\log \varepsilon|^{3(D-1)}$  and depends weakly on the dimension of the problem [5]. The improvement over standard PIC can be intuitively understood in the following way. The figure of merit for the statistical error in a PIC scheme is the number of particles *per cell*  $P_c = N_p/N_c$ , where  $N_c$  is the number of cells. Since  $N_c$  scales inversely with the cell volume,  $N_c \sim h^{-3}$  on a regular grid in 3-D. However, on a 3-D sparse grid,  $N_c \sim (4h)^{-1}$ . We thus achieve *many* more particles per cell, even with the total particle number fixed, by using a hierarchy of sparse grids. Observe furthermore that in a typical PIC scheme, the cell size has to be smaller than the Debye length in all dimensions. In contrast, in sparse PIC only few grids have to resolve the Debye length, and when they do, they only do so in one cell direction. In the next section, we will show that these theoretical results are confirmed in practice, with significant speed up as compared to a standard PIC scheme.

In addition to the stark algorithmic advantages described above, sparse PIC also holds the promise of tremendous benefit in massively parallel implementations. On modern distributed memory architectures, PIC's spatial grid is typically domain-decomposed across many compute nodes. Thus, at each time step, particle data must be moved onto the node corresponding to its location on the grid. This creates considerable communication and load-balancing overhead, both of which play increasing roles in determining overall computation time as architectures advance. The sparse combination grids being promoted here, however, require *far* less memory to store than their full-grid counterparts - for example, a  $2048 \times 2048 \times 2048$  full grid of double precision floats requires 64 gigabytes of storage, while the analogous sparse combination grid requires only 1.44 megabytes. With sparse grids, it thus becomes trivially cheap to replicate the *entire* spatial grid on every compute node. By doing so, particles may remain on a single node for the entirety of the simulation, eliminating data motion and making load balancing for particle operations trivial - one simply initializes each node with exactly the same number of particles!

## NUMERICAL EXAMPLE

To illustrate noise reduction through the sparse grids combination technique, we consider the time evolution of a Gaussian electron distribution in a periodic simulation domain. Two things are expected to happen: 1) the electron distribution is subject to periodic Langmuir oscillations at the

plasma frequency; 2) the amplitude of the density distribution decreases due to nonlinear Landau damping. This is indeed what we observed. The measure of interest for this article is the level of statistical noise in the density distribution after one Langmuir oscillation, which can be visualized in Figure 3. This figure demonstrates the remarkable savings in terms of computer time and memory usage sparse grids bring about: with 100 times fewer particles, the sparse-PIC scheme yields comparable accuracy as the standard PIC scheme.

Since sparse grids are designed to combat the curse of dimensionality, the savings are even more significant for 3-dimensional simulations, as we have shown for Landau damping in three dimensions in [5]. We chose not to plot the results here, because they are not as visually striking. However, we will highlight the central conclusions. We found that sparse PIC consistently uses less memory than standard PIC for a given target accuracy, often by an order of magnitude. For a target error in the electron density, sparse PIC also consistently uses less computation time than standard PIC; for the electric field, which is less affected by sampling noise, the computation times can be comparable.

## A WORD OF CAUTION

This article summarizes very recent progress for a project that is still in its infancy. We therefore would like to emphasize drawbacks of the sparse grids technique as presented here, which one will need to address in order to obtain a scheme which is indeed superior to the standard PIC scheme in most situations of physical interest. First, it is clear from the presentation in the third Section ("The Sparse Grid Combination Technique: Illustration with Interpolation") that the combination technique requires a structured grid. This may not be too stringent a constraint for the simulation of the acceleration phase in cyclotrons, but may be more challenging in other situations, such as axial injection into the cyclotron [3]. Furthermore, as we also highlighted in the third Section, the combination technique is not well suited for functions which have fine structure in all directions, which can be understood from the fact that the sparse grids never have fine resolution in all directions. Put differently, sparse grids perform best when the structure of the solution is aligned with the grid in a tensor product-like structure [6]. We have verified this empirically in [5] by considering a physical problem for which the solution is more efficiently represented in cylindrical coordinates. We found that in this situation the performance of the standard PIC scheme was superior to that of the sparse PIC scheme in the simple Cartesian implementation presented in this article.

Because our work is still at a very early stage, we do not see these drawbacks as condemning the sparse PIC algorithm to only be used for the very specific cases considered here. Instead, it is a motivation to improve the young version of our scheme we presented in this article, in order to make it more robust and versatile. A promising avenue for improvement is to rely on higher order interpolation and higher order shape

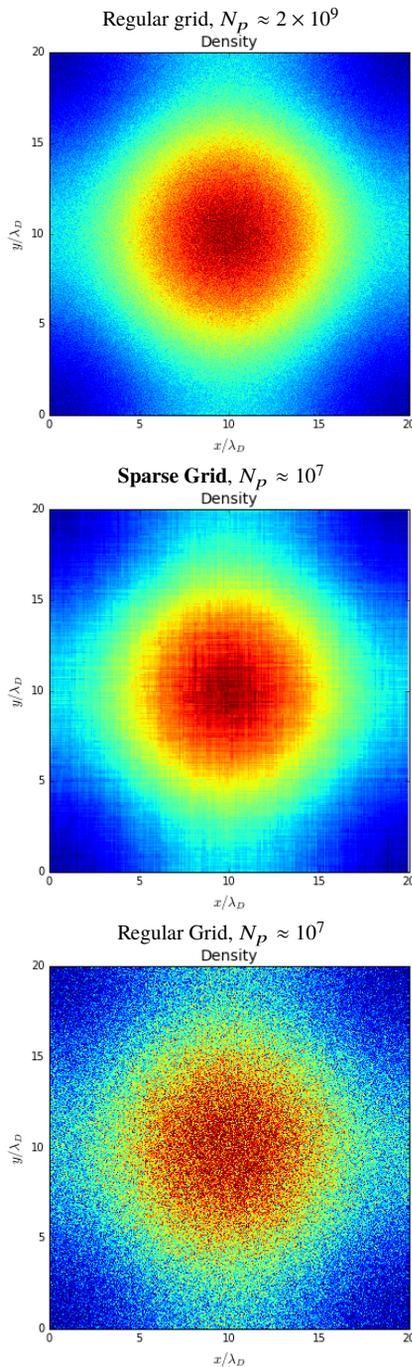


Figure 3: Time snapshots of the electron density from three simulations of the same nonlinear Landau damping problem. All have  $1024 \times 1024$  effective resolution. Compared to the top figure, the sparse grid solution (middle) has comparable statistical resolution but runs 30 times faster. Compared to the bottom figure, the sparse scheme runs in comparable time, but with considerably improved statistics.

functions, in order to reduce the grid-based error due to the sparse grids combination technique. Combining the sparse grids combination technique with adaptive mesh refinement

[7] may also have a strong potential, as it would allow us to better resolve the fine scale structures the sparse grids are missing. Finally, one could numerically construct optimized coordinate systems to be used by the sparse-PIC solver as the solution evolves, which would be designed to optimally align with the solution at each time step.

## CONCLUSION

We have presented a new strategy for reducing statistical noise in PIC simulations based on the sparse grid combination technique, a numerical method which had previously only been considered for grid based solvers. We found that our algorithm could lead to major savings in memory and computation time because the number of particles required to reach a certain level of accuracy is drastically reduced as compared to the requirements for the standard PIC scheme. This is because the sparse grids used in our numerical scheme all have larger cells than in a standard PIC scheme, thus increasing the number of particle per cells for a given total number of particles. And by virtue of the sparse grid combination technique, the price we pay for this is only a slight increase in the grid-based error.

Our straightforward algorithm underperforms in situations in which the solution is far from aligning with the directions of the grid and has fine structure in all dimensions. We are currently considering improvements to our scheme to tackle these issues.

## ACKNOWLEDGEMENTS

The authors would like to thank the organizers of ICAP'18 and the scientific program committee for inviting one of us and giving us the opportunity to present our results.

## REFERENCES

- [1] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Hilger, New York, 1988.
- [2] Y. J. Bi, A. Adelman, R. Dölling, M. Humbel, W. Joho, M. Seidel, and T. J. Zhang, *Phys. Rev. ST Accel. Beams*, vol. 14, pp. 054402, 2011.
- [3] D. Winklehner, A. Adelman, A. Gsell, T. Kaman, and D. Campo, *Phys. Rev. Accel. Beams*, vol. 20, pp. 124201, 2017.
- [4] M. Griebel, M. Schneider, and C. Zenger, "A combination technique for the solution of sparse grid problems", in *Iterative Methods in Linear Algebra*, ed. R. Bequwens and P. de Groen, Amsterdam: Elsevier, pp. 263–81, 1990.
- [5] L. F. Ricketson and A. J. Cerfon, *Plasma Phys. Control. Fusion* vol. 59, pp. 024002, 2017.
- [6] L. N. Trefethen, "Cubature, Approximation, and Isotropy in the Hypercube", *SIAM Review*, vol. 59, no. 3, pp. 469–491, 2017. doi:10.1137/16M1066312
- [7] H.-J. Bungartz and M. Griebel, "Sparse grids", *Acta Numerica*, vol. 13, pp. pp. 147–269, 2004. doi:10.1017/S0962492904000182

# RECONSTRUCTION OF PARTICLE DISTRIBUTIONS AT RFQ EXIT AT SNS BEAM TEST FACILITY\*

Z. Zhang<sup>†</sup>, S. Cousineau<sup>1</sup>, Department of Physics and Astronomy, University of Tennessee, Knoxville, USA

A. Aleksandrov, Oak Ridge National Laboratory, Oak Ridge, USA

<sup>1</sup> also at Oak Ridge National Laboratory, Oak Ridge, USA

## Abstract

Fluctuations of beam parameters and uncertainties of quadrupole gradients during measurements have effects on the reconstruction of initial particle distributions. To evaluate these effects, the concept of a distribution discrepancy is proposed. Results suggest effects of fluctuations of beam parameters are small, while uncertainties of quadrupole gradients are the main factors that affect the reconstructed distributions. By comparing the measured distributions with distributions produced by tracking the reconstructed initial distributions, it is proved that the real or quasi-real (closest to real) initial distribution can be obtained as long as the minimum distribution discrepancy is found.

## INTRODUCTION

The Beam Test Facility (BTF) at SNS consists of a 65 kV H<sup>-</sup> ion source, a 2.5 MeV RFQ, a beam line with advanced transverse and longitudinal beam diagnostic devices and a 6kW beam dump (as shown in Fig. 1). One of the main goals of the BTF is to provide a platform for conducting R&D for novel accelerator physics and technological concepts related to high intensity hadron beam generation, acceleration, manipulation and measurement [1]. Of particular importance is to conduct the first direct 6D phase space measurement of a hadron beam [2] which will be used to high intensity beam halo study [3].

Reconstruction of particle distributions from 6D phase space measurement is not trivial. Therefore, reconstruction of a 2D distribution without considering coupling between horizontal, vertical and longitudinal planes can be carried out as the first step, which can help validate the approach of reconstruction of particle distributions and obtaining of the real initial distribution which may be affected by beam parameters and quadrupole gradients, gaining experiences for the eventual reconstruction of 6D phase space particle distributions.

There are usually two methods to reconstruct the initial particle distributions by PIC simulation codes, one is fitting the RMS beam sizes of the measured distributions, the other is a tomography-like technique [4]. At SNS, a more direct method based on emittance data in both transverse directions and the backward tracking ability of PyORBIT is used [5,6], and a dedicated PIC back-tracking simulation code based on PyORBIT has been developed. The simulation code can transform measured particle

distributions into bunches for backward tracking, create backward lattice to track the bunch from measurement location to entrance of the lattice. In the work presented here, the lattice is the section of the BTF from the RFQ exit to the first slit (Slit 1) which is used to measure the transverse particle distributions. There are four quadrupoles (Q1, Q2, Q3 and Q4) in the lattice, which can be seen in Fig. 1. This paper mainly focuses on the investigation of influences of beam parameters and quadrupole gradients on the reconstructed particle distributions and how to obtain the real initial distributions.

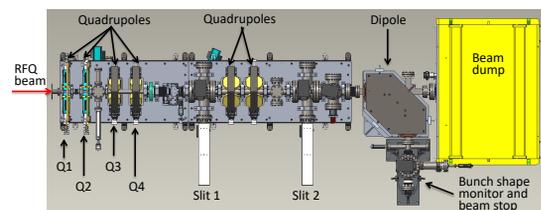


Figure 1: Layout of beam test facility at SNS.

## VALIDATION OF BACK-TRACKING SIMULATION CODE

The back-tracking simulation code needs to be verified that it can accurately backward track a distribution before it is applied to the distribution reconstruction.

First, an ideal Gaussian distribution is used to validate the back-tracking code. An ideal Gaussian distribution is generated at the RFQ exit and tracked to the slit 1 by the forward-tracking code. Then the back-tracking code tracks the distribution at slit 1 backward to the RFQ exit with the same quadrupole settings, and the final distribution is compared with the initial distribution. Specifically, corresponding particle coordinates are compared between the two distributions, and the maximum particle coordinates discrepancy in phase space is used as a measure of comparison. The parameters of the initial distributions are  $\alpha_x = -1.99$ ,  $\beta_x = 20$  mm/mrad,  $\epsilon_x = 0.16$  mm·mrad,  $\alpha_y = -1.99$ ,  $\beta_y = 20$  mm/mrad,  $\epsilon_y = 0.16$  mm·mrad, respectively, and the particle number is ten thousand.

Figure 2 displays the initial Gaussian distributions (black) and the distributions produced by back tracking (red). It illustrates the two distributions are coinciding completely in both  $x$ -plane and  $y$ -plane. Detailed comparison results show that the maximum particle coordinates discrepancies in  $x$  and  $y$  planes are ( $5.65 \times 10^{-5}$  mm,  $5.05 \times 10^{-4}$  mrad), ( $5.0 \times 10^{-5}$  mm,  $5.0 \times 10^{-4}$  mrad) for 0 mA and ( $6.37 \times 10^{-5}$  mm,  $8.21 \times 10^{-4}$  mrad), ( $5.31 \times 10^{-5}$  mm,  $8.6 \times 10^{-4}$  mrad) for 50 mA, which prove that

\* Work supported by NSF Accelerator Science grant 1535312

<sup>†</sup> email address: zzhang87@utk.edu

the back-tracking simulation code can reconstruct ideal distributions accurately.

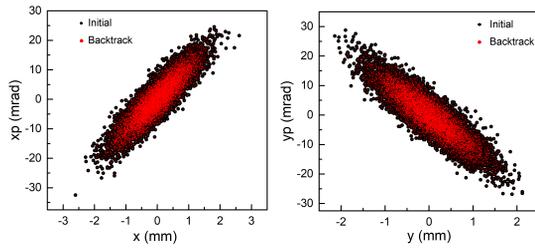


Figure 2: Initial particle distribution and distribution produced by back tracking.

Next, the back-tracking code is tested with measured distributions. If the distributions, produced by forward tracking the reconstructed distributions which are generated by backward tracking the measured distributions, are the same as the measured ones, the back-tracking code is validated. In contrast to the ideal distribution which is generated in both the transverse and longitudinal phase spaces by the forward-tracking code, the measured distributions have no particle distribution in the longitudinal phase space. Thus, a longitudinal distribution is generated by forward tracking the RFQ output distributions and assigned to the measured distributions to form a complete input distribution for the back-tracking code. By comparing the measured distributions with the distributions produced by forward tracking the reconstructed distributions, the maximum particle coordinates discrepancies are  $(3.47 \times 10^{-5} \text{ mm}, 1.10 \times 10^{-4} \text{ mrad})$  and  $(7.89 \times 10^{-5} \text{ mm}, 8.8 \times 10^{-5} \text{ mrad})$  in  $x$ -plane and  $y$ -plane, respectively. This result certifies that the back-tracking simulation code is reliable for measured distributions, too.

## RECONSTRUCTION OF INITIAL PARTICLE DISTRIBUTION AT RFQ EXIT

### Distribution Discrepancy

Due to the uncertainties of quadrupole gradients and fluctuations of beam current during measurements, the reconstructed initial distribution by one measured distribution may deviate from the real initial distribution. Therefore, the real initial distribution needs to be confirmed by two or more measured distributions produced by different quadrupole settings. In order to investigate the influences of beam parameters and quadrupole gradients on the reconstructed distributions and to obtain the real initial distribution, the concept of distribution discrepancy is proposed.

Assume there are two particle distributions in  $x$ - $xp$  phase space (red distribution and blue distribution, as plotted in Fig. 3), and their particles are divided by the same grids into different squares according to their coordinates. Then the distribution discrepancy, denoted by  $\text{DistD}$ , is defined by the following formula:

$$\text{DistD} = \sum_{i,j} \left| \frac{N_{i,j}^r}{TN^r} - \frac{N_{i,j}^b}{TN^b} \right|,$$

Where  $i$  and  $j$  are the grid number in  $x$  and  $xp$  direction,  $r$  and  $b$  mean the red distribution and blue distribution,  $N_{i,j}^r$  and  $N_{i,j}^b$  stand for the particle number of red distribution and blue distribution in the  $(i, j)$  square,  $TN^r$  and  $TN^b$  are total particle number in the red distribution and blue distribution, respectively. Big grid numbers produce high position and angle resolutions. Here grid numbers  $100 \times 100$  are chosen in  $x$ - $xp$  and  $y$ - $yp$  phase spaces, which means the position and divergency angle resolutions are about  $0.05 \text{ mm}$  and  $0.4 \text{ mrad}$ . If the two distributions are reconstructed by two different quadrupole settings, a small distribution discrepancy means they are not only close to each other, but also all close to a specific distribution which is the real initial distribution. While it is difficult to find the real initial distribution which requires the distribution discrepancy to be zero, a quasi-real initial distribution which is the closest to the real initial distribution can be obtained as long as the minimum distribution discrepancy is found.

In order to enhance the reconstruction accuracy, four measured distributions of  $20 \text{ mA}$  produced by four different quadrupole settings are used to find the real or quasi-real initial distributions at the BTF RFQ exit. In this way, every two reconstructed distributions are compared with each other which means there are total six distribution discrepancies. The quasi-real initial distribution is obtained when the maximum distribution discrepancy ( $\text{max\_DistD}$ ) among the six reaches the minimum value.

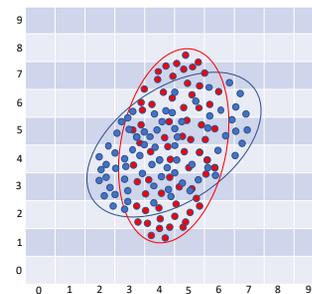


Figure 3: Particle distributions comparison.

### Effects of Assigned Longitudinal Beam Distribution and Beam Current

The assigned longitudinal beam distribution is produced by the forward-tracking code with the  $20 \text{ mA}$  RFQ output distributions. Therefore, a study of effects of the assigned longitudinal beam distribution on distribution discrepancies is conducted by studying the effects of the parameters of the RFQ output longitudinal distribution. The Twiss parameters and emittance of the RFQ output longitudinal distribution are  $\alpha_z \approx 0$ ,  $\beta_z \approx 0.6 \text{ mm/mrad}$  and  $\varepsilon_z \approx 0.2 \text{ mm-mrad}$ , respectively, according to the RFQ design result [7], and a range around these values (as shown in Fig. 4) is used in the study. The top two and the bottom left plots in Fig. 4 display the relationship between  $\text{max\_DistD}$  and the parameters of the longitudinal distribution at the RFQ exit. They demonstrate that the effects of the assigned longitudinal beam distribution on distribution discrepancies are very small, and consequently,

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

the designed RFQ output longitudinal distribution can be used to generate the initial distributions at RFQ exit.

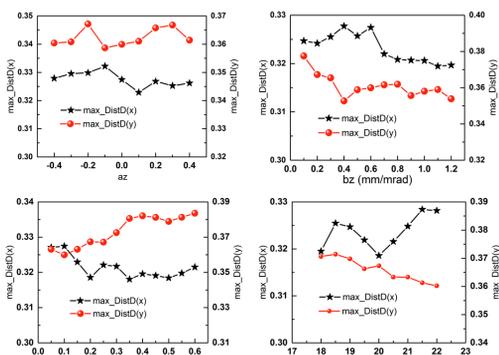


Figure 4: Effects of assigned longitudinal distribution and beam current on distribution discrepancies.

The bottom right plot in Fig. 4 shows fluctuation of beam current has a small effect on distribution discrepancies, too. Thus, 20 mA can be used for the reconstruction of initial distributions.

### Effects of Quadrupole Gradients

The quadrupole gradients used for the study are obtained from the quadrupoles control system by comparing the current readings from the power supplies with the known relationship between gradients and currents [8]. The uncertainty of the current readings causes uncertainty in quadrupole gradients. Figure 5 illustrates the relationship between  $\max\_DistD$  and gradient variations of the fourth quadrupole during the fourth measurement, in which the setting gradient is 6.875 T/m according to the current reading. It can be seen from Fig. 5 that a change from  $-5\%$  to  $5\%$  of the gradient causes about 27% and 45% variations of distribution discrepancies in  $x$ - $xp$  phase space ( $\max\_DistD(x)$ ) and  $y$ - $yp$  phase space ( $\max\_DistD(y)$ ), which means the uncertainty of quadrupole gradients have a great impact on the initial particle distributions at RFQ exit. Figure 5 shows the effects of gradient uncertainty of only one quadrupole during one measurement. There are total four measurements and each quadrupole of every measurement has influence on the distribution discrepancies. Therefore, all the influences have to be investigated to obtain the initial distributions. Due to the large amount of gradient combinations a script was written to handle the calculations and to analyze the results.

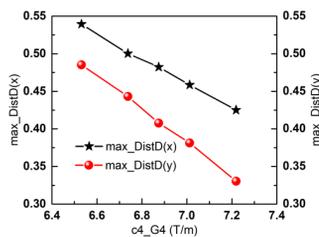


Figure 5: Relationship between  $\max\_DistD$  and gradient uncertainty ( $c4\_G4$  means gradient of the fourth quadrupole during the fourth measurement).

### Generation of Initial Particle Distributions

It has been found that the minimum  $\max\_DistD(x)$  and  $\max\_DistD(y)$  do not occur at the same gradient setting combination during data analysis, in other words,  $\max\_DistD(y)$  is 0.460 when  $\max\_DistD(x)$  is the minimum value of 0.256, while  $\max\_DistD(x)$  is 0.430 when  $\max\_DistD(y)$  is the minimum value of 0.261. The initial particle distributions should take place when both  $\max\_DistD(x)$  and  $\max\_DistD(y)$  have relative small values with the same quadrupole settings. Hence, to facilitate finding the best pair of  $\max\_DistD(x)$  and  $\max\_DistD(y)$  the average and difference values are calculated, and finally  $\max\_DistD(x) = 0.281$  and  $\max\_DistD(y) = 0.270$  are found to be the best result. Figure 6 displays the initial distributions when  $\max\_DistD(x) = 0.281$  and  $\max\_DistD(y) = 0.270$ , which are the combined results of the four separate reconstructed initial distributions and are considered to be very close to the real initial distributions.

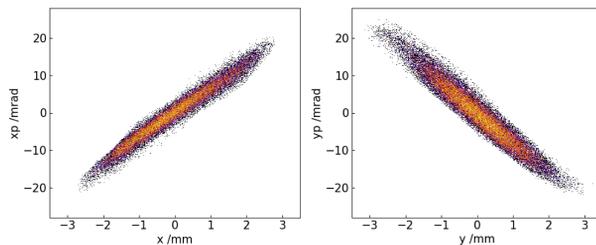


Figure 6: Generated particle distributions at RFQ exit.

The particle distributions in Fig. 6 are forward tracked to the first slit where the tracked distributions are compared with the measured ones, and Fig. 7 displays one of the comparison results. In Fig. 7, the red plots are the measured distributions and the black plots are forward tracked distributions. It demonstrates the measured distributions and the tracked distributions agree well with each other except for a small area with low particle density in both  $x$ - $xp$  and  $y$ - $yp$  phase spaces in the tracked distributions. The Twiss parameters and emittances of the measured and tracked distributions are also nearly the same (as listed in Table 1). Considering the measurement errors, these results suggest the reconstructed initial particle distributions in Fig. 6 are reliable and using the minimum distribution discrepancy to obtain the real or quasi-real initial distributions is reasonable.

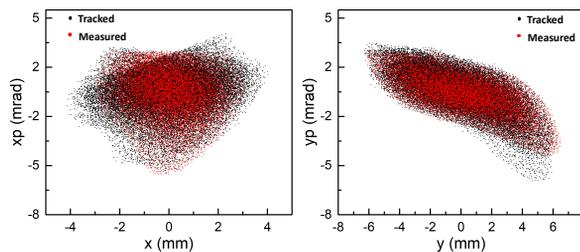


Figure 7: Comparisons between measured distributions and forward tracked distributions which are produced by using the generated initial distributions as input.

Table 1: Twiss Parameters and Emittances of the Measured and Tracked Distributions

Distribution	Measured	Tracked
$\alpha_x$	0.11	-0.10
$\beta_x$ (mm/mrad)	0.80	1.09
$\varepsilon_x$ (mm·mrad)	1.71	1.81
$\alpha_y$	0.91	0.95
$\beta_y$ (mm/mrad)	3.10	2.79
$\varepsilon_y$ (mm·mrad)	2.27	2.25

The above results were obtained by using the “Hard Edge” magnetic fields of the quadrupoles, meanwhile, using the measured field distributions of the quadrupoles could produce the same results.

## CONCLUSION

In order to evaluate the effects of fluctuations of beam and uncertainties of quadrupole gradients on the reconstruction of initial particle distributions the concept of distribution discrepancy is proposed. Studies using this concept suggest the effects of beam parameters are very small. It has been found that variations of quadrupole gradients influence the initial distributions greatly, therefore, all the possible combinations of quadrupole gradients have to be studied. The combination which produces small distribution discrepancies in both  $x$ - $xp$  and  $y$ - $yp$  phase spaces is considered to be able to generate the initial particle distributions which are very close to the real ones. Distributions produced by forward tracking the reconstructed initial distributions are compared with the measured ones, and results show they agree well with each other, which suggests using the minimum distribution discrepancy to obtain the real or quasi-real initial distributions is reliable.

## ACKNOWLEDGEMENTS

This work has been supported by NSF Accelerator Science grant 1535312.

The authors would like to thank A. Shishlo and A. Zhukov at SNS for their help during data analysis and scripts optimization.

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

## REFERENCES

- [1] A. Aleksandrov, B. Cathey, S. Cousineau, *et al.*, “Commissioning of the new SNS RFQ and 2.5 MeV beam test facility”, in *Proc. IPAC2017*, Copenhagen, Denmark, 2017, pp. 2438–2440.
- [2] B. Cathey, S. Cousineau, A. Aleksandrov, *et al.*, “First six dimensional phase space measurement of an accelerator beam”, *Physical Review Letters* 121, 064804 (2018).
- [3] S. Cousineau, “Upgrade of the SNS Beam Test Facility for Studies of Beam Halo Formation at Low Energies”, in *Proc. Space Charge 2017*, Darmstadt, 2017.
- [4] V. A. Dimov, J. B. Lallement, A. M. Lombardi, *et al.*, “Emittance reconstruction techniques in presence of space charge applied during the LINAC4 beam commissioning”, in *Proc. HB2016*, Malmö, Sweden, 2016, pp. 433–438.
- [5] A. Shishlo, S. Cousineau, J. Holmes, *et al.*, *Procedia computer science*, vol. 5, 2015, pp. 1272–1281.
- [6] A. Shishlo, A. Aleksandrov, A. Zhukov, *et al.*, “Model based optics studies in the MEBT sections of SNS”, in *Proc. IPAC2017*, Copenhagen, Denmark, 2017, pp. 814–816.
- [7] A. Ratti *et al.*, “The design of a high current, high duty factor RFQ for the SNS,” *Proc. European Particle Accelerator Conference*, Vienna, Austria, June 2000, 495–497.
- [8] A. Aleksandrov and J. Staples, “3-D calculations of field distribution in the MEBT quad and effect of the fringe field on beam dynamics”, in *Proc. PAC’01*, Chicago, USA, 2001, pp. 1746–1748.

# SPACE CHARGE AND TRANSVERSE INSTABILITIES AT THE CERN SPS AND LHC

E. Métral<sup>†</sup>, D. Amorim, G. Arduini, H. Bartosik, H. Burkhardt, E. Benedetto, K. Li, A. Oeftiger,  
 D. Quattraro, G. Rumolo, B. Salvant, C. Zannini, CERN, Geneva, Switzerland

## Abstract

At the CERN accelerator complex, it seems that only the highest energy machine in the sequence, the LHC, with space charge (SC) parameter close to one, sees the predicted beneficial effect of SC on transverse coherent instabilities. In the other circular machines of the LHC injector chain (PSB, PS and SPS), where the SC parameter is much bigger than one, SC does not seem to play a major (stabilising) role, and it is maybe the opposite in the SPS. All the measurements and simulations performed so far in both the SPS and LHC will be reviewed and analysed in detail.

## INTRODUCTION

In the PSB, transverse instabilities (which still need to be fully characterized) are observed without damper during the ramp, where space charge could potentially play a role but no important change of instability onset was observed along the cycle when changing the bunch length (and shape) for constant intensity.

In the PS, a Head-Tail (HT) instability with six nodes is predicted at injection without space charge and observed with natural chromaticities and in the absence of Landau octupoles, linear coupling and damper.

In the SPS, a fast vertical single-bunch instability is observed at injection above a certain threshold (depending on the slip factor), with a travelling-wave pattern along the bunch. Several features are close to the ones from the predicted Transverse Mode-Coupling Instability (TMCI) between modes - 2 and - 3 without SC (for  $Q' \sim 0$ ).

Finally, in the LHC, the predicted HT instability with one node for a chromaticity of about five units, with neither Landau octupoles nor damper, is observed only above a certain energy, as confirmed by simulations with space charge. Furthermore, the intensity threshold for the TMCI at injection for a chromaticity close to zero (which has not been measured yet as it is much higher than the current LHC intensities) is predicted to be significantly increased by space charge according to simulations.

Considering the case of a TMCI with zero chromaticity, a two-particle approach would conclude that both SC and/or a reactive transverse damper (ReaD) would affect TMCI in a similar way and could suppress it (see Fig. 1).

Using a two-mode approach (instead of the previous two-particle approach), a similar result would be obtained in the “short-bunch” regime (i.e. TMCI between modes 0 and - 1, such as in the LHC) as both a ReaD and SC are expected to be beneficial: the ReaD would shift the mode 0 up and SC would shift the mode - 1 down, but in both

cases the coupling would therefore occur at higher intensities. However, the situation is more involved for the “long-bunch” regime (i.e. TMCI between higher-order modes, such as in the SPS). As the ReaD modifies only the (main) mode 0 and not the others (where the mode-coupling occurs), it is expected to have no effect for the main mode-coupling (as confirmed in Fig. 2, using the Vlasov solver GALACTIC [3]). As concerns SC, it modifies all the modes except 0, and the result is still in discussion and the subject of this paper, which is structured as follows: the first section is devoted to the many SPS studies, while the LHC results will be discussed in the second section before concluding and discussing the next steps.

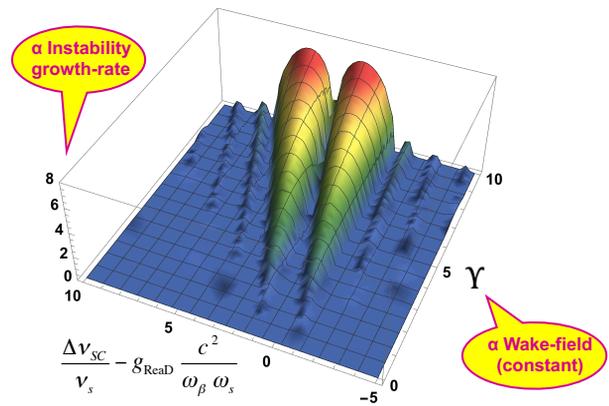


Figure 1: Two-particle approach for the TMCI following Ref. [1] but adding a reactive transverse damper (ReaD). This combines the results from Ref. [1] (with SC only) and Ref. [2] (with reactive damper only).

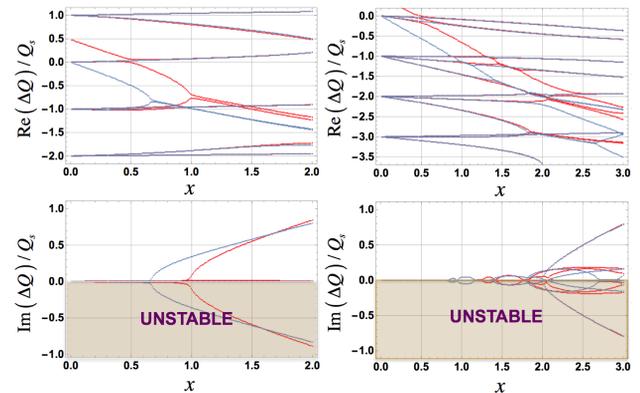


Figure 2: Usual TMCI plots for the LHC (left) and SPS (right) assuming a Broad-Band resonator impedance (with  $Q' = 0$ ), without / with ReaD (50 turns) in blue / red [3].

<sup>†</sup> Elias.Metral@cern.ch

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## SPS

A fast vertical single-bunch instability with protons ( $p^+$ ) was observed at the SPS injection in 2003 using a longitudinal emittance of  $\sim 0.2$  eVs, i.e. much smaller than the nominal one of 0.35 eVs, to probe the transverse single-bunch limit of the machine (see Fig. 3) [4].

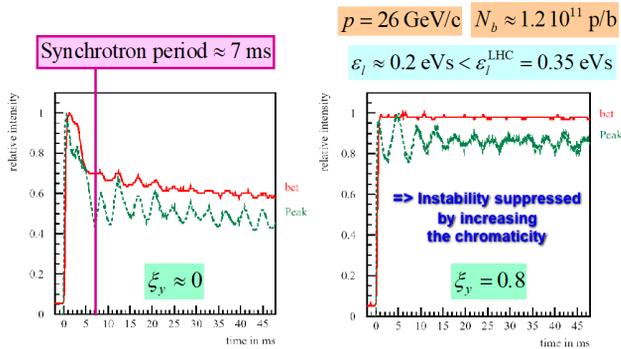


Figure 3: (Left) observation of a fast (compared to the synchrotron period) vertical single-bunch instability with protons ( $p^+$ ) at the SPS injection in 2003. (Right) stabilisation by increasing the chromaticity. bct stands for beam current transformer, which measures the total intensity, whereas Peak measures a bunch length dependent bunch intensity.

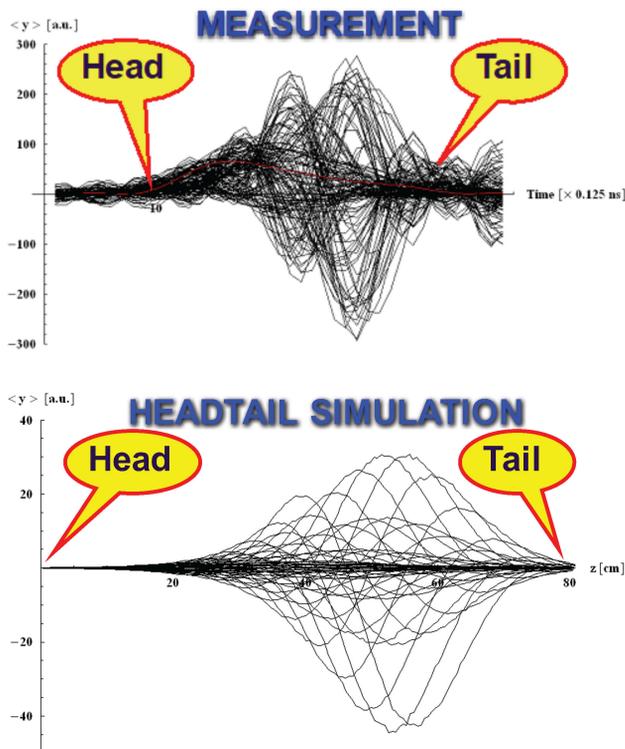


Figure 4: Comparison of the intra-bunch motion between measurements (upper) and HEADTAIL [5] simulations (lower) using a single bunch with  $1.2 \cdot 10^{11}$  p/b, an rms bunch length of 0.7 ns and zero chromaticity, interacting with a Broad-Band resonator (with a resonance frequency of 1 GHz, a quality factor of 1 and a shunt impedance of  $20 \text{ M}\Omega/\text{m}$ ).

This instability featured a travelling-wave pattern along the bunch (with a frequency close to 1 GHz), which was in relatively good agreement with HEADTAIL simulations using a Broad-Band resonator model (with a resonance frequency of 1 GHz, a quality factor of 1 and a shunt impedance of  $20 \text{ M}\Omega/\text{m}$ ), as can be seen in Fig. 4. A travelling-wave pattern along the bunch should be the sign of a TMCI as the coupling between two HT modes (which are standing-wave patterns) generates a travelling wave, as can be seen in Fig. 5 [6]. This was confirmed in Fig. 6 for the case of the SPS, using a Broad-Band resonator with a shunt impedance of  $10 \text{ M}\Omega/\text{m}$ . A TMCI between modes  $-2$  and  $-3$  (for the main mode-coupling, i.e. after some mode-coupling decoupling due to the “long-bunch” regime) is predicted in the absence of SC. A similar result was predicted with the full impedance model which was developed in parallel [9,10].

The next question was: why do we observe “what looks like a TMCI (with a travelling-wave along the bunch)” whereas space charge should suppress it, according to some past theoretical analyses, with the pioneer work of M. Blaskiewicz in 1998 [11] followed by several other analyses [12-15]? Can we observe the coupling of the (negative or positive) modes? How do measurements compare to HEADTAIL simulations? According to Ref. [11], the negative modes should rapidly disappear for a strong SC parameter, defined as the ratio between the space charge tune spread (for a KV distribution or half the tune spread for a Gaussian distribution) and the synchrotron tune (see Fig. 7). So do we still see the negative modes predicted without SC (or do we see the positive ones or another mechanism taking place)?

First simulations with the combined effect of an impedance and SC, using a 3<sup>rd</sup> order symplectic integrator for the equation of motion, taking into account non-linear SC forces coming from a Gaussian shaped bunch revealed a minor beneficial effect of SC, raising the intensity threshold by  $\sim 5\text{-}10\%$ , as shown in Fig. 8. It is more difficult to see what happens exactly to the modes but it could still be compatible with a mode-coupling between modes  $-2$  and  $-3$ , as the main activity appears at about the same position as without SC.

Direct measurements of the modes in the SPS were tried and resulted in Fig. 9. Here again, it is difficult to conclude but it could still be compatible with a mode-coupling between modes  $-2$  and  $-3$ , as the main activity appears at about the same position as without SC.

An indirect measurement of mode-coupling (in addition to the travelling-wave pattern resulting from mode-coupling between two HT modes with standing-wave patterns) in the “long-bunch” regime consists of measuring the beam stability vs. increasing bunch intensity, as the bunch should be first stable until mode-coupling (of the low-order modes) and then stable again after decoupling before becoming very unstable at the main mode-coupling. This is what was predicted from HEADTAIL simulations using both a Broad-Band resonator model and a more realistic impedance model of the SPS and this is what was measured, as reported in Fig. 10.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

This observation was another argument to state that mode-coupling was taking place even if the direct measurement of mode-coupling was still missing...

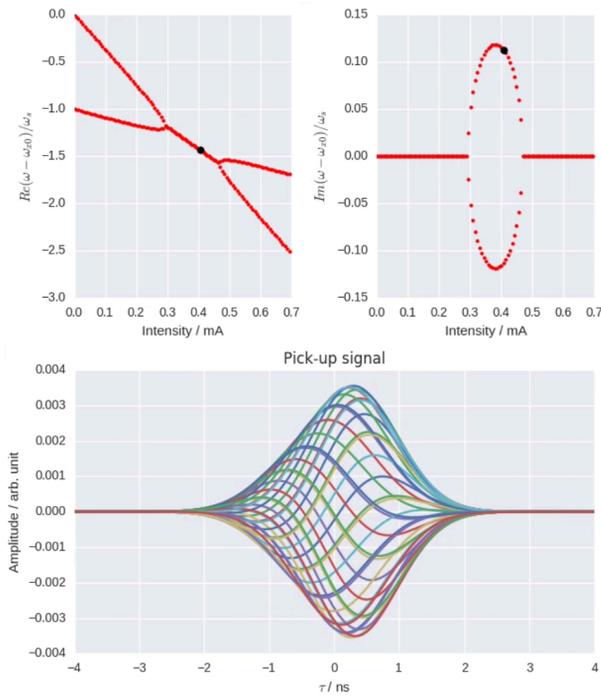


Figure 5: The coupling of two HT modes (standing-wave patterns) generates a travelling-wave pattern. Example from the DELPHI Vlasov solver for a coupling between modes 0 and -1 [6].

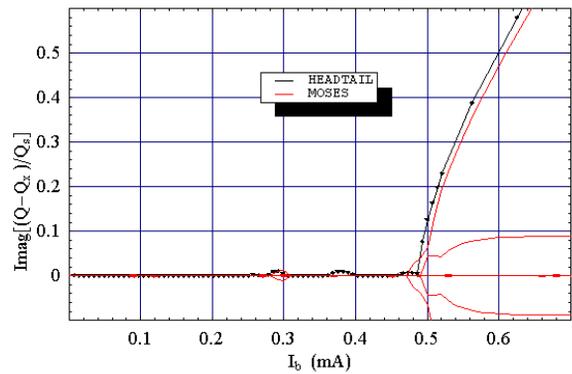
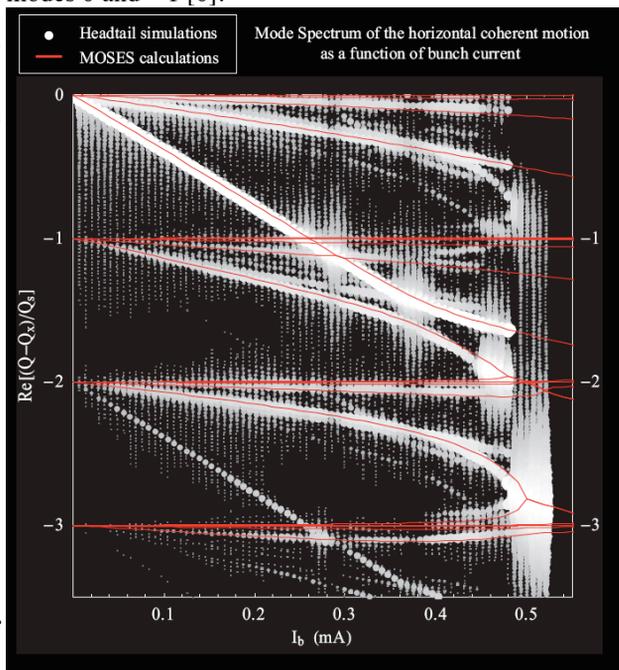


Figure 6: Comparison between MOSES [7] and HEADTAIL simulations (using SUSSIX [8] to process the output data) using the parameters of Table B.3 (for MOSES) and B.4 (for HEADTAIL) of Ref. [9].

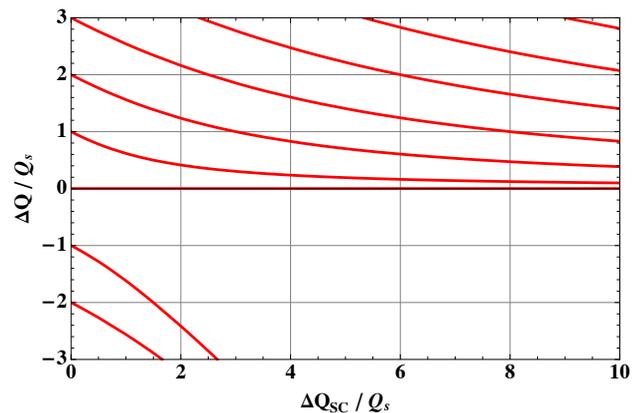


Figure 7: Mode-frequency shifts vs. the SC parameter for the case of the Air-Bag Square well (or ABS) model [11].

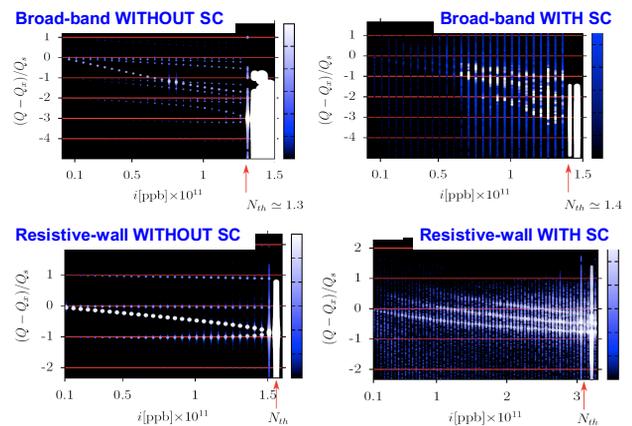


Figure 8: Simulations with both impedance and SC, revealing a minor beneficial effect from SC for the SPS case (upper) [16].

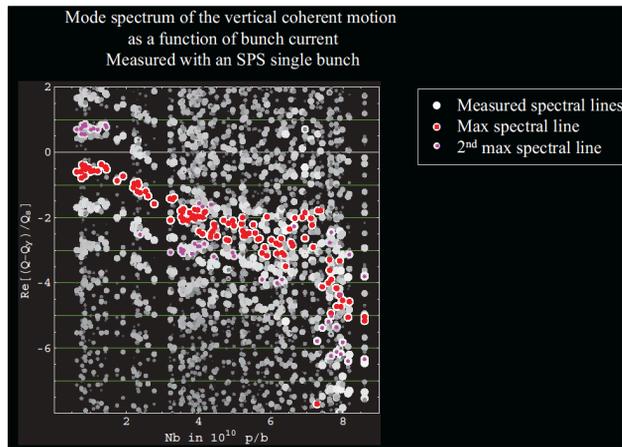


Figure 9: Direct measurements of the modes in the SPS [9].

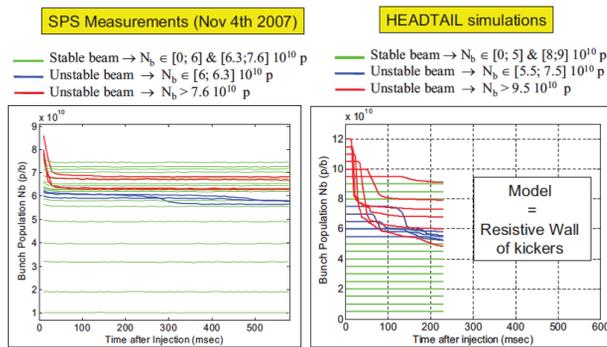


Figure 10: Indirect measurement of mode-coupling by trying to identify in the SPS the regions of stable and unstable bunch intensities (see Fig. 6.26 in Ref. [9]). These measurements were performed for a longitudinal emittance of 0.16 eVs, and rms bunch length of 0.7 ns and a chromaticity as close to zero as possible.

As i) the SPS instability seemed to be relatively well described by TMCI using a Broad-Band resonator (without SC) and ii) in this case (“long-bunch” regime) a simple formula exists within a numerical factor 2 as mentioned in Refs. [17-18] (see Eq. (1)), which was recently checked in Ref. [19], it was proposed to modify the optics to increase the slip factor [20]: the “Q20” optics (named like this as the integer part of the tune is 20) replaced the previous “Q26” optics (where the integer part of the tune was 26). Within the TMCI formalism, the simple formula can be derived using a two-mode approach (considering the two most critical modes overlapping the maximum of the real part of the impedance): bunch stability is reached when the head and the tail are swapped sufficiently rapidly (due to synchrotron oscillations) compared to the instability rise-time, i.e. when the synchrotron period divided by  $\pi$  is equal to the instability rise-time derived from this simple model (sm):  $T_s = \pi \tau_{TMCI}^{sm}$ . This leads to the following stability criterion for the threshold number of protons  $N_{b,th}$ , which can be written in two forms (e.g. in the vertical plane)

$$N_{b,th} = \frac{4\pi^3 f_s Q_{y0} E \tau_b^2}{e c} \times \frac{f_r}{|Z_y|} \quad (1)$$

$$N_{b,th} = \frac{8\pi Q_{y0} |\eta| \epsilon_l}{e \beta^2 c} \times \frac{f_r}{|Z_y|}$$

where  $f_s$  is the synchrotron frequency,  $Q_{y0}$  the unperturbed (low-intensity) vertical tune,  $E$  the total energy,  $\tau_b$  the full ( $4\sigma$ ) bunch length in s,  $e$  the elementary charge,  $c$  the speed of light,  $f_r$  the resonance frequency of the Broad-Band ( $Q = 1$ ) resonator,  $Z_y$  the shunt impedance,  $\beta$  the relativistic velocity factor,  $\epsilon_l$  the longitudinal emittance and  $\eta$  the slip factor, given by

$$\eta = -\frac{df_{rev}/f_{rev}}{dp/p} = \alpha_p - \frac{1}{\gamma^2} = \frac{1}{\gamma_i^2} - \frac{1}{\gamma^2}$$

It is interesting to note that i) within the framework of this model the simple formula giving the instability rise-time well above the TMCI threshold (which was checked with MOSES and HEADTAIL, within the same factor 2 as before for the intensity threshold [21]) can be written as  $\tau_{TMCI}^{sm} = (T_s / \pi) \times (N_{b,th} / N_b)$  and ii) in the second form of Eq. (1), the notion of synchrotron oscillations disappears. This equation is the same as for coasting beams, but written with peak values, where the Landau damping is provided by the momentum spread. What is important is the product of the longitudinal emittance and the slip factor, i.e. the distance to transition. As the longitudinal emittance should be kept at 0.35 eVs for the beams to be sent from the SPS to the LHC, the only parameter on which one can act is the product of the vertical tune times the slip factor. For machines made of simple FODO cells it can be shown that the slip factor is approximately given by the horizontal tune ( $\gamma_t \approx Q_{x0}$ ), which means that if one wants to modify  $\gamma_t$ , one should modify the horizontal tune. The SPS slip factor as a function of the horizontal tune is depicted in Fig. 11. For the Q26 optics,  $\gamma_t \approx 22.8$ , whereas for the Q20 optics,  $\gamma_t \approx 18$ . This means that for the Q20 optics, the product of the slip factor times the horizontal tune gives  $1.80 \cdot 10^{-3} \times 20.13 \approx 0.0362$ , whereas for the Q26 optics it gives  $0.62 \cdot 10^{-3} \times 26.13 \approx 0.0162$  (considering for this case a non-integer part of the tune of 0.13, which can be slightly different in practice but this does not change the picture). Therefore, according to Eq. (1) the intensity threshold should be increased by the factor  $0.0362 / 0.0162 \approx 2.2$ . This is in good agreement with measurements, as can be seen in Fig. 12, where an intensity increase of a factor 4.0 / 1.6  $\approx 2.5$  was observed.

These results are also in good agreement with HEADTAIL simulations from 2014 for different optics (adding also the case of the Q22 optics, which is a better optics for

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

some RF considerations), with full impedance model but still without SC (see Fig. 13).

A good agreement was also achieved between measurements and simulations looking at different longitudinal emittances, using the full impedance model but still without SC (see Fig. 14), even if the Q26 was maybe a bit more critical in measurements than in simulations.

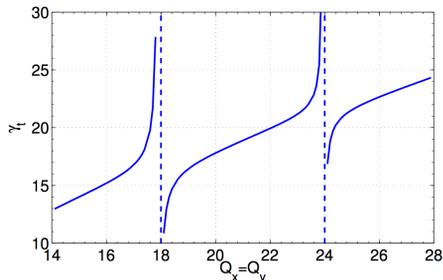


Figure 11: SPS slip factor as a function of the horizontal tune [20].

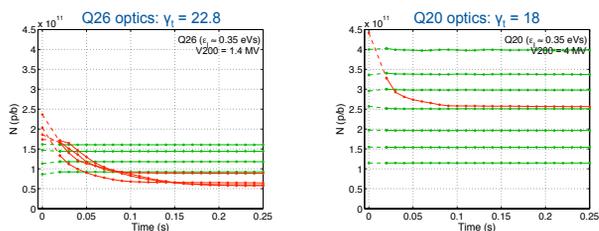


Figure 12: Comparison of the measured bunch intensity thresholds between the Q26 and Q20 optics. The longitudinal emittance is 0.35 eVs, i.e. about two times larger than in Fig. 10, which explains why the intensity threshold with Q26 is about two times larger than in Fig. 10.

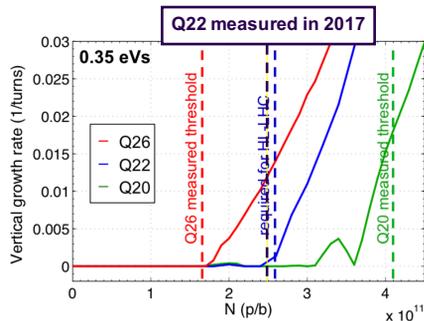


Figure 13: Comparison of the bunch intensity thresholds between measurements and HEADTAIL simulations.

Finally, a good agreement was also reached between measurements and pyHEADTAIL [22] simulations with SC this time for the Q20 optics, still considering the Broad-Band resonator model, as the intensity threshold was found close to the no-SC case (see Fig. 15). However, a detailed analysis of the modes involved seems to reveal different modes involved at the start of the instability: without SC, a mode-coupling between azimuthal modes  $-2$  and  $-3$  (with radial mode 0) is observed while

with SC, a mode-coupling between azimuthal modes 1 and 2 (with radial mode 1) seems to be found [23].

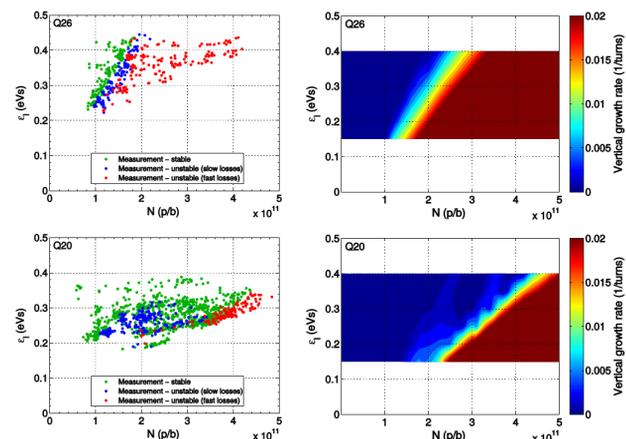


Figure 14: Comparison of the bunch intensity thresholds between measurements (left) and HEADTAIL simulations (right) looking at different longitudinal emittances, using the full impedance model but still without SC.

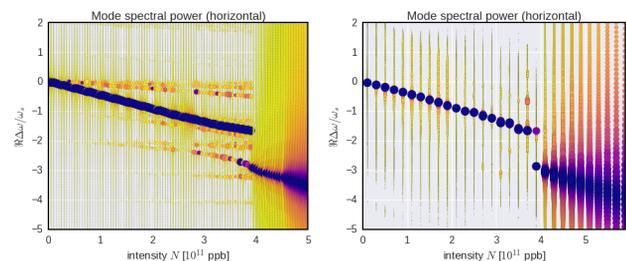


Figure 15: pyHEADTAIL simulations with the Q20 optics comparing the cases without SC (left) and with SC (right).

The comparison of the intra-bunch motions in Fig. 16 between measurements and HEADTAIL simulations without SC reveals also a good agreement in particular for the Q20 optics. For the Q26 optics, the measured intra-bunch motion seems to be more towards the tail than in the simulations, as recently pointed out by A. Burov, who discovered a new destabilising effect of SC, which could be responsible for this effect [24]. This disagreement was not present in the first studies (without SC) of Fig. 4, which means that both impedance and SC could have a similar effect, which needs to be disentangled. The Q20 optics seems less subject to this effect but it is true that the Q26 optics has a much higher SC parameter ( $\sim 27$ ) compared to Q20 ( $\sim 5$ ).

New pyHEADTAIL simulations without and with SC were performed for different shunt impedances of Broad-Band resonators and the very interesting results are depicted in Fig. 17. It is seen that in the absence of SC, a higher shunt impedance leads to an intra-bunch motion pushed towards the tail. The effect of SC seems three-fold: i) its pushes the intra-bunch motion even more towards the tail; ii) it increases the frequency of oscillation and iii) it increases or reduces the oscillation amplitude.

The (simple) 2-mode approach (with a mode-coupling between two consecutive modes  $m$  and  $m + 1$ ), which was used in the past in the case of the (very) “long-bunch” bunch regime to reveal almost no effect of SC on TMCI [25,26], can be extended also to the general case, leading to the intensity threshold of Eq. (2)

$$Q_s \left[ \sqrt{q_{sc}^2 + (q + 1)^2} - \sqrt{q_{sc}^2 + q^2} \right] = 2 \left| \Delta Q_{q,q+1}^{s,y} \right| \quad (2)$$

where  $q_{sc} = \Delta Q_{sc} / (2 Q_s)$  and  $q = |m| + 2k$  (with  $0 \leq k \leq +\infty$  defining the radial mode number). This means that the same intensity threshold as the no-SC case is obtained, i.e. it is the same as Eq. (1), except that  $Q_s$  is now multiplied by the term  $\sqrt{q_{sc}^2 + (q + 1)^2} - \sqrt{q_{sc}^2 + q^2}$ , which is equal to 1 when  $q \gg q_{sc}$  and to 0 when  $q_{sc} \gg q$ . In particular, the same scaling with respect to the other parameters is obtained and therefore the same mitigation measure should be applied.

Based on these new results, another measurement campaign is planned to try and disentangle between the impedance and SC effects by varying the SC tune spread. However, it is worth emphasizing that a solution has been already found in practice for this instability in the SPS and that it is not a performance limitation anymore.

## LHC

Using the impedance model of the High-Luminosity (HL-) LHC at injection and considering the case of zero chromaticity, it was found with pyHEADTAIL simulations that the TMCI between modes 0 and -1 without SC is suppressed over the intensity range studied [26].

For the chromaticity  $Q' = +5$ , a HT instability with one node ( $m = -1$ ) is observed without SC whereas it is completely suppressed with SC [26]. Studying the effect of energy during the ramp, which reduces the SC tune spread (by increasing the transverse emittances at injection energy), the instability re-appears at  $\sim 2$  TeV. This energy is the energy at which the first transverse single-bunch instability was observed in the LHC during the first ramp performed in 2010 with neither Landau octupoles nor transverse damper [27].

## CONCLUSION

A beneficial effect of SC is predicted in the (HL-) LHC (working in the “short-bunch” regime) for both the HT instability and TMCI. SC simulation with pyHEADTAIL gives an explanation of the first single-bunch HT instability observed in the LHC in 2010 with neither Landau octupoles nor transverse damper. This might be good to re-do a controlled experiment to confirm it. Furthermore, SC simulation also predicts that SC increases significantly the TMCI intensity threshold ( $Q' = 0$ ) at (HL-) LHC injection. This could not be studied at the moment as the TMCI is currently out of reach in the LHC.

As concerns the SPS (working in the “long-bunch” regime), several past measurements were close to the case without SC. The intensity threshold was increased considerably in practice by increasing the slip factor (based on

theoretical analysis without SC) and this is working very well: Q20 optics has replaced Q26 optics in the SPS for all the beams to be sent to the LHC. However, a recent theoretical analysis by A. Burov [24] predicts a detrimental effect of SC (even below the TMCI intensity threshold without SC), which was confirmed by recent SC simulations with Q26. The (simple) 2-mode approach was also extended to the general case and the same intensity threshold as the no-SC case is obtained, except that the synchrotron tune is now reduced by SC. However, the same scaling as without SC is obtained and therefore the same mitigation measure should be applied. A new measurement campaign is planned to analyze all this in detail.

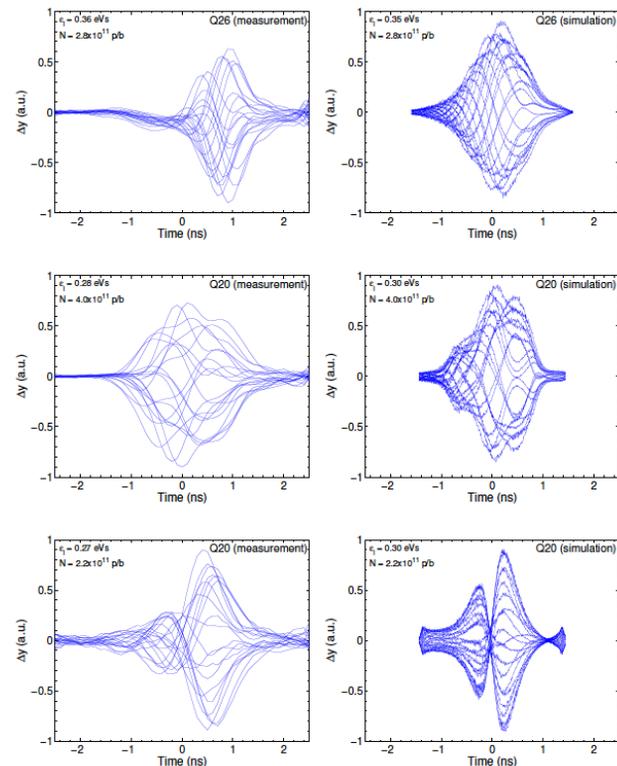


Figure 16: Comparison of intra-bunch motions between measurements (left) and simulations with HEADTAIL (right) for different cases with Q26 and Q20 optics.

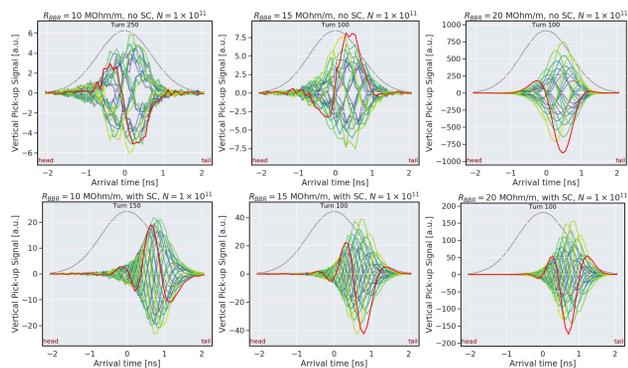


Figure 17: Intra-bunch motion from pyHEADTAIL simulations without and with SC for different shunt impedances of Broad-Band resonator impedances.

## ACKNOWLEDGEMENTS

Many thanks to A. Burov for all the very interesting and inspiring discussions about this subject for many years.

## REFERENCES

- [1] Y.H. Chin, A.W. Chao and M. Blaskiewicz, "Two particle model for studying the effects of space-charge force on strong head-tail instabilities", *Phys. Rev. Accel. Beams*, vol. 19, pp. 014201, 2016.
- [2] A. Burov, "Efficiency of feedbacks for suppression of transverse instabilities of bunched beams", *Phys. Rev. Accel. Beams*, vol. 19, pp. 084402, 2016.
- [3] E. Métral *et al.*, "Destabilising effect of the LHC transverse damper", in *Proc. IPAC'18*, Vancouver, BC, Canada, Apr-May 2018, paper THPAF048, pp. 3076-3079. doi:10.18429/JACoW-IPAC2018-THPAF048
- [4] G. Arduini, H. Burkhardt, and E. Métral, "Observation of a fast single bunch instability on protons in the SPS", CERN AB-Note-2003-093 (MD), Dec. 2003.
- [5] G. Rumolo and F. Zimmermann, "Practical user guide for HEADTAIL", CERN SL-Note-2002-036-AP, 2002.
- [6] D. Amorim, "DELPHI pickup signal with mode coupling", CERN internal HSC meeting, 12/06/2017, [https://indico.cern.ch/event/643949/contributions/2622544/attachments/1474370/2283035/2017-06-08\\_DELPHI\\_Eigenvectors\\_results\\_split\\_modes.pptx](https://indico.cern.ch/event/643949/contributions/2622544/attachments/1474370/2283035/2017-06-08_DELPHI_Eigenvectors_results_split_modes.pptx)
- [7] Y.H. Chin, "User's guide for new MOSES Version 2.0". CERN/LEP-TH/88-05, 1988.
- [8] R. Bartolini and F. Schmidt, "Computer Code for Frequency Analysis of Non-Linear Betatron Motion", CERN SL-Note-98-017-AP, 1998.
- [9] B. Salvant, "Impedance model of the CERN SPS and aspects of LHC single-bunch stability", Ph.D. thesis 4585, Ecole Polytechnique, Lausanne, Switzerland 2010.
- [10] C. Zannini, "Electromagnetic simulation of CERN accelerator components and experimental applications", Ph.D. thesis 5737, Ecole Polytechnique, Lausanne, Switzerland (2013).
- [11] M. Blaskiewicz, "Fast head-tail instability with space charge", *Phys. Rev. ST Accel. Beams*, vol. 1, pp. 044201, 1998.
- [12] A. Burov, "Head-tail modes for strong space charge", *Phys. Rev. ST Accel. Beams*, vol. 12, pp. 044202, 2009.
- [13] V. Balbekov, "Transverse mode coupling instability threshold with space charge and different wakefields", *Phys. Rev. Accel. Beams*, vol. 20, pp. 034401, 2017.
- [14] V. Balbekov, "Threshold of transverse mode coupling instability with arbitrary space charge", *Phys. Rev. Accel. Beams*, vol. 20, pp. 114401, 2017.
- [15] T. Zolkin, A. Burov, and B. Pandey, "TMCI and Space Charge II", (2017), arXiv:1711.11110 [physics.acc-ph].
- [16] D. Quatraro and G. Rumolo, "Effects of direct space charge on the transverse mode-coupling instability", in *Proc. IPAC'10*, Kyoto, Japan, May 2010, paper TUPD046, pp. 2027-2029.
- [17] E. Métral, "Stability criteria for high-intensity single-bunch beams in synchrotrons", in *Proc. EPAC'02*, Paris, France, June 2002.
- [18] E. Métral, "Effect of bunch length, chromaticity, and linear coupling on the transverse mode-coupling instability due to the electron cloud", in *Proc. ECLLOUD'02*, CERN, Geneva, Switzerland, April 2002, p. 211.
- [19] A. Burov and T. Zolkin, "TMCI with resonator wakes", arXiv:1806.07521v2 [physics.acc-ph].
- [20] H. Bartosik, "Beam dynamics and optics studies for the LHC injectors upgrade", CERN-THESIS-2013-257.
- [21] E. Benedetto and E. Métral, "Instability rise-time far above the TMCI threshold: comparison between simple theory, MOSES and HEADTAIL", CERN/GSI beam dynamics and collective effects collaboration meeting, 18-19/02/2009, GSI, Darmstadt, Germany. <http://www-linux.gsi.de/~boine/CERN-GSI-2009/benedetto.ppt>
- [22] E. Métral *et al.*, "Beam instabilities in hadron synchrotrons", *IEEE Trans. on Nucl. Sci.*, vol. 63, No. 2, April 2016, pp. 1001-1050.
- [23] A. Oeftiger, "SPS transverse mode-coupling instability with space charge", CERN internal space charge working group meeting, 12 June 2018, [https://indico.cern.ch/event/735602/contributions/3033902/attachments/1666224/2671281/oeftiger\\_TMCI\\_SC.pdf](https://indico.cern.ch/event/735602/contributions/3033902/attachments/1666224/2671281/oeftiger_TMCI_SC.pdf)
- [24] A. Burov, "Convective instabilities of bunched beams with space charge", arXiv:1807.04887v4 [physics.acc-ph].
- [25] E. Métral, "Why direct space charge was/is believed to have only a small effect on the TMCI intensity threshold in the SPS?", CERN internal HSC meeting, 11/07/2016, [https://espace.cern.ch/be-dep-workspace/abp/HSC/Meetings/TMCIwithSC\\_SPSwithLongBunch\\_SimpleModel\\_EM\\_11-07-2016.pdf](https://espace.cern.ch/be-dep-workspace/abp/HSC/Meetings/TMCIwithSC_SPSwithLongBunch_SimpleModel_EM_11-07-2016.pdf)
- [26] E. Métral, "Effect of space charge on the CERN LHC and SPS transverse instabilities: simulations vs. measurements", Space Charge 2017 Workshop, Darmstadt, Germany, 2017.
- [27] E. Métral *et al.*, "Stabilization of the LHC single-bunch transverse instability at high-energy by Landau octupoles", in *Proc. IPAC'11*, San Sebastián, Spain, Sep. 2011, paper MOPS074, pp. 775-777.

# CHALLENGES IN EXTRACTING PSEUDO-MULTIPOLES FROM MAGNETIC MEASUREMENTS

S. Russenschuck, G. Caiafa, L. Fiscarelli, M. Liebsch, C. Petrone, P. Rogacki  
CERN, Geneva, Switzerland

## Abstract

Extracting the coefficients of Fourier-Bessel series, known as pseudo-multipoles or generalized gradients, from magnetic measurements of accelerator magnets involves technical and mathematical challenges. First, a novel design of a short, rotating-coil magnetometer is required that does not intercept any axial field component of the magnet. Moreover, displacing short magnetometers, step-by-step along the magnet axis, yields a convolution of the local multipole field errors and the sensitivity (test function) of the induction coil. The deconvolution must then contend with the low signal-to-noise ratio of the measurands, which are integrated voltages corresponding to spatial flux distributions. Finally, the compensation schemes, as implemented on long coils used for measuring the integrated field harmonics, cannot be applied to short magnetometers. All this requires careful design of experiment to derive the optimal length of the induction coil, the step size of the scan, and the highest order of pseudo-multipoles in the field reconstruction. This paper presents the theory of the measurement method, the data acquisition and deconvolution, and the design and production of a saddle-shaped, rotating-coil magnetometer.

## INTRODUCTION

The magnetic measurement section within the magnet group of CERN's technology department is responsible for the qualification of all superconducting and normal conducting magnets in CERN's accelerator complex. To supplement the long rotating-coil magnetometers and stretched-wire systems (the section's workhorses for magnetic measurements) we have recently developed moving induction-coil arrays, axial and transversal rotating-coil scanners [1], and induction-coil transducers for solenoidal magnets. Applications of these tools require, however, a sophisticated post-processing step based on the regularity conditions of electromagnetic fields. To this end, the magnet bores can be considered as trivial domains, i.e., simply connected and source-free with piecewise smooth, closed and consistently oriented boundaries. Calculating the transversal field harmonics as a function of the coordinate in the magnet's axial direction, for example, by using the numerical field calculation program ROXIE [2], or measuring these harmonics with a very short, rotating-coil scanner, allows the extraction of the coefficients of Fourier-Bessel series, known as pseudo-multipoles [3] or generalized gradients [4].

However, the raw measurement data from the field transducers are induced voltages that are integrated using a digital integrator, triggered by an angular encoder. Developing these signals into Fourier series results in convoluted

functions of the spacial flux distribution, because strictly speaking, point-like measurements of the magnetic flux density are not possible.<sup>1</sup> Before such signals can be used as boundary data for harmonic analysis or boundary-element methods (BEM), a deconvolution is required.

A careful design of experiment is required, considering a low signal-to-noise ratio of the measurand, the sensitivity of the induction coil with respect to transversal harmonics, the step-size of the longitudinal scan, and the compensation schemes for the main-field component. In this paper we present the design and production of the transversal-field scanner and the challenges in applying the pseudo-multipole theory to measurement data.

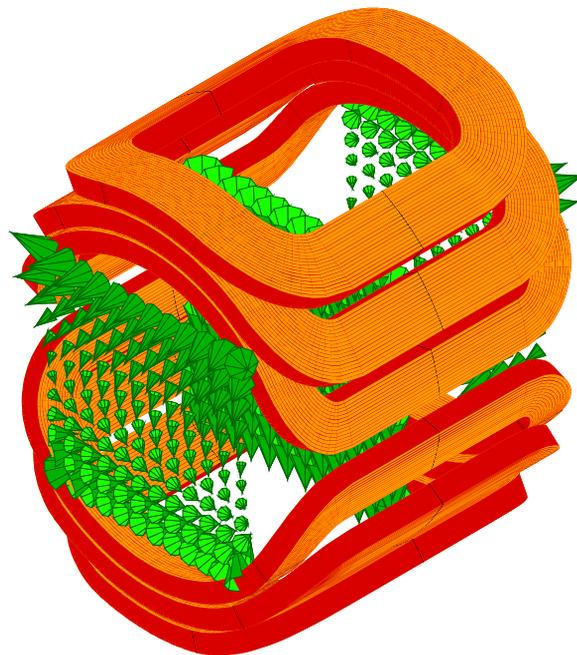


Figure 1: Representation of the magnetic flux density in 3 different planes of an orbit corrector for the ELENA project. Notice the large  $z$ -component in the end fields. Computed with the CERN field computation program ROXIE [2].

## PSEUDO MULTIPOLES

The local field distribution in short magnets, such as the one shown in Fig. 1, cannot be expressed by the usual field harmonics (Fourier series) for the integrated fields because they do not constitute a complete orthogonal basis of the

<sup>1</sup> Transversal Hall sensors come close but their active area (the Hall plate) typically has a diameter of 2-3 mm.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

transversal field distribution in the magnet ends. In other words, the field distribution in the magnet extremities is not holomorphic and therefore does not obey the  $r^{n-1}$  scaling laws derived for the integrated fields. Following Erdelyi [5] and Caspi [6], and using a combination of a Fourier series in  $\varphi$  and a power series in  $r$  around the axis yields

$$\phi_m = \sum_{n=1}^{\infty} r^n (\tilde{C}_n(r, z) \sin n\varphi + \tilde{D}_n(z) \cos n\varphi), \quad (1)$$

where

$$\tilde{C}_n(r, z) := C_{n,n}(z) - \frac{C_{n,n}^{(2)}(z)}{4(n+1)} r^2 + \frac{C_{n,n}^{(4)}(z)}{32(n+1)(n+2)} r^4 - \dots \quad (2)$$

In the interest of brevity, the similar expressions for the skew components  $\tilde{D}_n(z)$  have been omitted here. To reduce the burden on notation we will henceforth, and without loss of generality, assume ideal magnets without skew field component, i.e.,  $\mathcal{D}_{n,n}(z) = 0$ .

In three dimensions, the single harmonic component (index  $n$  to describe the  $\varphi$  dependence) will contain pseudo multipoles accounting for the transverse field components no longer exhibiting a pure  $r^{n-1}$  dependence on the radius. In the straight section of the magnet the partial derivatives with respect to  $z$  vanish and the equations will become identical with the well known 2D equations, as expected. The field components at any radius within the bore of the magnet are then given by  $B_r = -\mu_0 \frac{\partial \phi_m}{\partial r}$ ,  $B_\varphi = -\mu_0 \frac{1}{r} \frac{\partial \phi_m}{\partial \varphi}$ , and  $B_z = -\mu_0 \frac{\partial \phi_m}{\partial z}$ . Consequently,

$$\begin{aligned} B_r &= -\mu_0 \sum_{n=1}^{\infty} r^{n-1} \tilde{C}_n(r, z) \sin n\varphi, \\ B_\varphi &= -\mu_0 \sum_{n=1}^{\infty} n r^{n-1} \tilde{C}_n(r, z) \cos n\varphi, \\ B_z &= -\mu_0 \sum_{n=1}^{\infty} r^n \frac{\partial \tilde{C}_n(r, z)}{\partial z} \sin n\varphi, \end{aligned} \quad (3)$$

where

$$\tilde{C}_n(r, z) := n C_{n,n}(z) - \frac{(n+2)C_{n,n}^{(2)}(z)}{4(n+1)} r^2 + \frac{(n+4)C_{n,n}^{(4)}(z)}{32(n+1)(n+2)} r^4 - \dots \quad (4)$$

A dipole field ( $n = 1$ ) rolling off at the magnet's extremity gives rise to a pseudo sextupole and higher-order (odd and only odd) pseudo multipoles, while a quadrupole field gives rise to higher-order, even pseudo multipoles. Pseudo-multipole terms have the same angular dependence as the leading terms  $C_{n,n}(z)$ .

It is common practice to speak of *feed down* when lower-order multipoles are generated from higher-order multipoles by axis misalignment of the magnet or the measurement shaft. We might equally speak of *feed-up* when higher-order multipoles are generated due to field variations of the lower-order multipoles along  $z$ . Since the pseudo terms are even

derivatives of the leading terms, their  $z$  integrals over the entire magnet will be zero, as to be expected.

It is important to note that the coefficients  $C_{n,n}(z)$  are still unknown at this stage. This is the reason for choosing calligraphic characters in the typesetting. Although the  $C_{n,n}(z)$  are the leading terms of the series expansion, they are not identical to the  $B_n(z)$  components in the Fourier expansion of the (measured or calculated) transverse field. Fortunately, because all the pseudo-multipole terms can be calculated from the leading terms, the problem is reduced to extracting the  $C_{n,n}(z)$  from measured or calculated data on the domain boundary. It is worth mentioning, that even if these terms were confused, the reconstructed field would still obey the 3D Laplace equation. A powerful *Maxwellification* so to say.

## THE TRANSVERSAL FIELD SCANNER

Measuring the transversal field harmonics with a short rotating coil (measurement radius  $r_0$ ) yields a convoluted function of the multipole field components  $B_n(r_0, z)$ . Using this data we must solve the differential equations

$$\begin{aligned} B_n(r_0, z) &= -\mu_0 r_0^{n-1} \tilde{C}_n(r_0, z) = \\ &= -\mu_0 r_0^{n-1} \left( n C_{n,n}(z) - \frac{(n+2)C_{n,n}^{(2)}(z)}{4(n+1)} r_0^2 + \right. \\ &\quad \left. \frac{(n+4)C_{n,n}^{(4)}(z)}{32(n+1)(n+2)} r_0^4 - \dots \right). \end{aligned} \quad (5)$$

for  $n = 1, 3, 5, \dots, N$ . Applying a Fourier transform to the functions  $B_n(r_0, z)$  and  $C_{n,n}(z)$  it follows from Eq. (5):

$$\mathcal{F}\{C_{n,n}(z)\} = \frac{-\mathcal{F}\{B_n(r_0, z)\}}{\mu_0 r_0^{n-1} U_n^K}, \quad (6)$$

where

$$U_n^K := \left( n - \frac{(n+2)(i\omega)^2}{4(n+1)} r_0^2 + \frac{(n+4)(i\omega)^4}{32(n+1)(n+2)} r_0^4 - \dots \right)$$

is the feed-up term of order  $n$ , up to the highest  $z$ -derivative  $K$ . Let  $\tilde{B}_n$  denote the measured function given by the field harmonic  $B_n$  convoluted by the  $z$ -dependent coil-sensitivity factors  $s_n$  of the measurement coil.<sup>2</sup> In this case, the Fourier transform of  $C_{n,n}(z)$  yields

$$\mathcal{F}\{C_{n,n}(z)\} = \frac{-\mathcal{F}\{\tilde{B}_n(r_0, z)\}}{\mathcal{F}\{s_n(r_0, z)\}} \frac{1}{\mu_0 r_0^{n-1} U_n^K}. \quad (7)$$

If we were able to produce an infinitely short induction coil,  $s_n$  would become a delta function in  $z$  and its Fourier transform would become one.

The Fourier transform of the leading term  $\mathcal{F}\{C_{n,n}(z)\}$  can be extracted from the Fourier transform of the measured data  $\tilde{B}_n(r_0, z)$ . The final step then consists in recovering the function  $C_{n,n}(z)$  from the spectrum by means of the Fourier integral

$$C_{n,n}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}\{C_{n,n}(\omega)\} e^{i\omega z} d\omega. \quad (8)$$

<sup>2</sup> The calculation of the sensitivity factors is deferred to the next sections.

## THE SENSOR DESIGN

The shafts used for the field measurements are usually an assembly of induction coils of different radii which are series-connected to compensate for the induced voltage signal for the main dipole or quadrupole field component. The size and arrangement of these coils is based on the  $r^{n-1}$  scaling laws derived from the 2D field solution. The flux linkage through the induction coil can be calculated from the Stokes theorem:  $\Phi(\varphi) = N \int_{\mathcal{A}} \mathbf{B} \cdot d\mathbf{a} = N \int_{\mathcal{A}} \text{curl } \mathbf{A} \cdot d\mathbf{a}$ , and the  $z$ -component of the magnetic vector potential can be expressed as

$$A_z(r_c, z) = \sum_{n=1}^{\infty} \frac{r_c}{n} (B_n(r_c, z) \cos n\varphi), \quad (9)$$

assuming that there are no skew components in the field. The flux linkage through the induction coil can then be calculated from

$$\Phi(\varphi) = \int_{z_0-\ell/2}^{z_0+\ell/2} \left( \sum_{n=1}^{\infty} s_n^{\text{tan}} (B_n(r_c, z) \sin n\varphi) \right) dz, \quad (10)$$

where  $s_n^{\text{tan}} = \frac{2N}{n} r_c \sin\left(\frac{n\delta}{2}\right)$ , is called the coil-sensitivity function, and  $\delta$  is the coil's opening angle that may be a function of  $z$ . The coil radius is denoted  $r_c$ . The physical unit of the sensitivity is  $[s_n^{\text{tan}}] = 1 \text{ m}$ . If the coil is short and used to map the field in the magnet end region, a number of challenges arise: Because of the absence of a simple scaling law between the multipole coefficients at  $r_1$  and  $r_2$ , these cannot be disentangled from the measurements, when the signals are compensated on the analogue side. The induction coil must be saddle-shaped, i.e., its radius on the shaft must be constant, because otherwise, a voltage is induced in the induction-coil ends when it is rotated around its axis; see Fig. 2.

Because the induction coil is shorter than the magnet and the size of its coil ends cannot be neglected with respect to its straight section, the sensitivity  $s_n^{\text{tan}}$  will become a function of  $z$ . This, in turn, becomes the test function of the convolution of the harmonic content along  $z$ .

A solution for these problems is the nesting of induction coils of the same radius and choosing the number of turns and their opening angles to compensate for the main dipole field. The shaft design is based on PCB technology, manufactured as a flexible stack of two double-layer PCBs, bent around a precision machined shaft; see Figs. 3 and 4. Spring-loaded roller bearings allow the displacement and centering of the shaft within an aluminum tube, which is inserted in the magnet bore. A clamping mechanism is used for fixing the flexible PCB on the shaft in order to reduce the tolerances on the coil's radius and alignment.

The main coil in the center has a small opening angle to provide sensitivity to the higher-order field components. Some constraints apply due to the manufacturing process of the PCB. A compromise had to be found between the maximum number of turns, and thus the coil's sensitivity, and the lengths of the coil ends.

The compensation coil should be sensitive to the dipole component only. This is achieved by a shell-type coil with

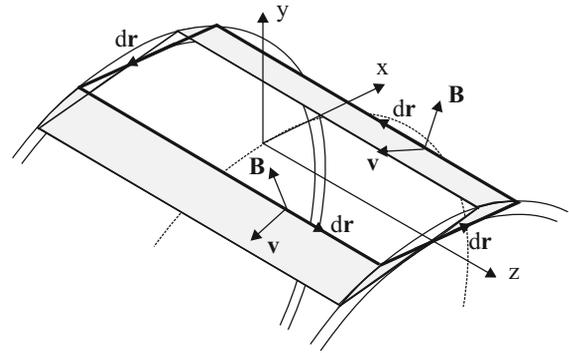


Figure 2: A (classical) tangential coil rotated by a certain angle traces out two patches of the cylindrical mandrel but also the surfaces between the chord and the apex, which intercepts the  $B_z$  field component. Such a coil must therefore be made long enough so that it covers the entire fringe-field region of the magnet and it is thus guaranteed that the  $B_z$  components are zero at the coil ends.



Figure 3: Image of the sensor showing the flexible PCB coil and clamping mechanism. The ends of the nested induction coils are all on the same radius and therefore trace out no surface when rotated. Therefore no voltage is induced by the axial field component.

about 60 degree opening angle. This opening angle and the spacing between turns has been optimized to fine-tune the dipole sensitivity of the compensation coil.

For short coils, the end-effects must be considered; they will lead to a different geometric-mean length and magnetic length of the coil depending on the multipole order. The graphs of the compensated sensitivity functions in the end region of the induction coils are given in Fig. 5.

An uncertainty analysis yielded a maximum allowable error in the track positions of about  $30 \mu\text{m}$  and in the radius positioning of about  $\pm 200 \mu\text{m}$ .<sup>3</sup> Metrological measurements were performed using a coordinate measuring machine (CMM). The uncertainty in the spanned surface, due to the tolerances in the track positioning in the PCB pro-

<sup>3</sup> Although the measurement radius can be calibrated in a reference quadrupole magnet, a radius deviating from nominal (19 mm), will lead to a lower compensation ratio because of a change in arc length.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

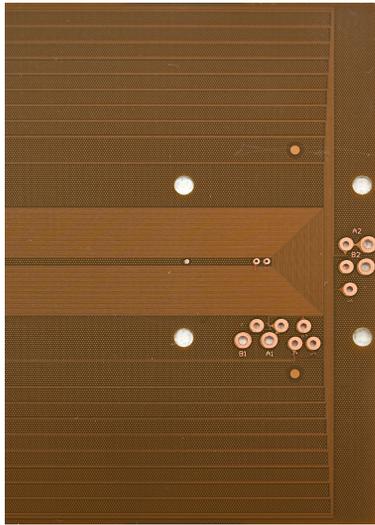


Figure 4: Image of the PCB coil. Notice the additional tracks (dashed) and copper patches designed to minimize the concentration of corrosive acid between the tracks of the compensation coil.

duction, and the uncertainty of the coil calibration in the reference magnets are both on the order of one unit in  $10^4$ . While the presence of open circuits is obvious, short circuits between turns are difficult to detect with resistance measurements, because of the varying thickness of the tracks. For that reason we still need to validate the sensor in the straight section of a long reference dipole magnet.

The outer diameter of the shaft (and its bearings) is guided in a tube of 50 mm inner diameter. In this way the transducer fits into the aperture of the section's reference dipole.<sup>4</sup> The mole design reduces the space needed for a displacement system; the shaft can either be positioned using a cableway or extension tube, or it can be mounted on an arm of a displacement stage (mapper).

## MEASUREMENT UNCERTAINTY

In order to study the deconvolution and the design of experiments, we must first discuss the highest achievable accuracy to date, which is by a transversal-field scanner of the classical design with radial coils in PCB technology. In the center of the reference dipole magnet there is no axial field component and therefore the results will yield a meaningful lower bound for the measurement uncertainty. An uncompensated measurement, using only the main coil, results in a precision of about one unit in  $10^5$ . With a compensation ratio of the main component exceeding 3000, the precision is increased to one unit in  $10^7$ . From this result, the known surface of the induction coil, and the maximum flux density in the magnet, we can conclude that the minimum flux

<sup>4</sup> Although magnets like the one shown in Fig. 1 constitute the ultimate application of the saddle-shaped coil magnetometers, the metrological characterization was done in the reference dipole and quadrupole magnets featuring long and homogenous straight sections that can be used for cross-calibration of the different coil transducers and Hall sensors.

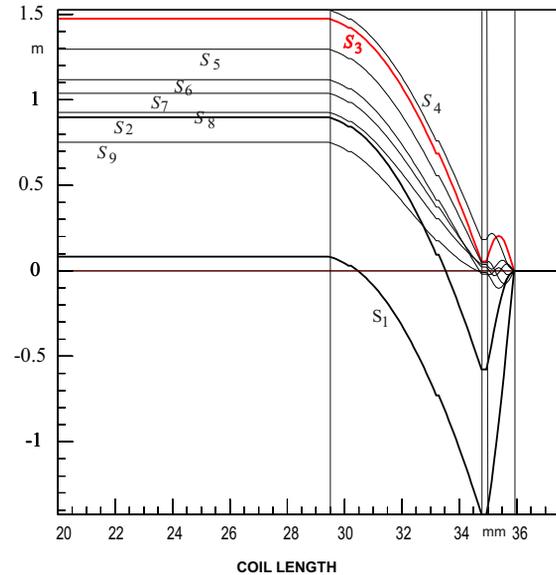


Figure 5: Sensitivity functions  $s_n^{\tan}(z)$  along the axis of the rotating-coil magnetometer (from the center to the coil-end region). The straight section has been shortened by 20 mm to better visualize the roll-off. Notice the compensation for the main (dipole) field sensitivity  $s_1$ ; also notice the different shapes of the roll-off, depending on the multipole order. Because the compensation coil is longer than the central coil (coil ends indicated by the vertical lines), the cross-section of the central coil is designed to compensate for the end effects, i.e., the overshoot into the negative values of  $s_1$ .

linkage in the induction coil must be larger than  $10^{-8} \mu\text{Vs}$ . Unfortunately, a compensation ratio of more than 1000 is not achievable with the nested saddle-shaped coils. The design value is about 630, limited by the different lengths of the nested coils. The completed transducer achieves a compensation ratio of no more than 27 because of the tolerance on the measurement-coil radius.

Another uncertainty results from positioning errors (longitudinal position and instability of rotation) during the scanning and measurement process. While the standard deviation of the harmonics, extracted from the compensated signal, is reduced by two-orders of magnitude, one would expect more because of the high compensation ratio of 3000. This is a result of the coil eccentricity and rotational instability.

For the longitudinal positioning error we obtain about  $20 \mu\text{m}$  per meter distance from the laser tracker and about  $60 \mu\text{m}$  for the alignment, resulting in a total uncertainty of about 0.1 mm. In the case of integral measurements of a long magnet, obtained by combining the results of multiple position scans, this error stays below  $10^{-4}$  as random errors will average out. For local measurements in the magnet's fringe-field region we must expect uncertainties on the order of 1–2% for a 100 mm long coil. This is severe because feed-down corrections for axis misalignment cannot be done as in the magnet straight section.

We are therefore placing special attention on accurate longitudinal positioning of the shaft, the stabilizing of the rotation, and the centering within the magnet bore.

## DECONVOLUTION OF THE MEASURED SIGNALS

The challenge is now to find a suitable order  $n$  of the pseudo-multipoles  $C_{n,n}$  and the highest order  $m$  of derivatives  $C^{(m)}$ , in order to minimize the reconstruction uncertainty of the local magnetic field. The uncertainty of the method will also depend on the step size chosen for the displacement of the transducer and on the signal-to-noise level of the measurand.

Let  $\tilde{B}_n(r_0, z_k)$  denote the measured, transversal field component of order  $n$ , sampled at positions  $z_k, k = 1, \dots, K$  along the magnet axis, affected by noise  $n(z_k)$ . In the following, we omit the notation of the radial dependencies. The noisy, convoluted signal is then given by

$$\tilde{B}_n[k] = (s_n * B_n)[k] + n[k]. \quad (11)$$

To save on notation we write

$$\tilde{B}_n(f) = s_n(f)B_n(f) + n(f) \quad (12)$$

instead of  $\mathcal{F}\{\tilde{B}[k]\} = \mathcal{F}\{s_n[k]\}\mathcal{F}\{B_n[k]\} + \mathcal{F}\{n[k]\}$  for the corresponding equation in the frequency domain. To reconstruct the transversal field harmonics we apply a discrete filter  $g[k]$ :

$$\hat{B}_n[k] = g[k] * \tilde{B}_n[k], \quad \hat{B}_n(f) = g(f)\tilde{B}_n(f), \quad (13)$$

where the hat denotes the reconstructed (estimated) multipole-field distribution  $\hat{B}_n(z_k)$ ; see Fig. 6.

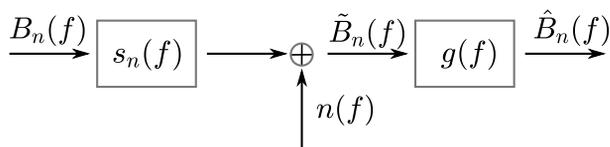


Figure 6: Signal path in the frequency domain. We distinguish the estimated spectrum  $\hat{B}_n(f)$ , the measured spectrum  $\tilde{B}_n(f)$  and the true harmonic  $B_n(f)$ .

A straightforward method of deconvolution uses the inverse of the sensitivity spectrum  $g(f) = 1/s_n(f)$ , which results in an amplification of the noise  $n(f)$  for frequencies where the spectrum  $s_n(f)$  has small values. An alternative approach is to use a filter that minimizes the expected mean-squared error in the frequency domain:

$$\begin{aligned} \mathbb{E}[e^2] &= \mathbb{E}\left[|B_n(f) - \hat{B}_n(f)|^2\right] \\ &= \mathbb{E}\left[|B_n(f) - g(f)\left(s_n(f)B_n(f) + n(f)\right)|^2\right]. \end{aligned} \quad (14)$$

Assuming that the noise is uncorrelated and has zero mean, a minimum of the mean-squared error, Eq. (14), can be found by the Wiener-Kolmogorov filter

$$g_w(f) = \frac{1}{s_n(f)} \frac{|s_n(f)|^2}{|s_n(f)|^2 + \frac{\mathbb{E}[n(f)]^2}{\mathbb{E}[B_n(f)]^2}}, \quad (15)$$

where the second expression in the denominator is the inverse of the expected signal-to-noise ratio  $SNR(f) = \mathbb{E}[B_n(f)]^2 / \mathbb{E}[N(f)]^2$ . Filtering with  $g_w(f)$  will therefore damp frequencies with low  $SNR(f)$ . The value of  $g_w(f)$  will approach the inverse of the sensitivity  $s_n(f)^{-1}$  for frequencies with a high  $SNR(f)$ .

In Fig. 7 the deconvolution was applied to a noisy  $\tilde{B}_3$ , which was generated from simulations with added Gaussian random noise. As expected, the convolution by the inverse

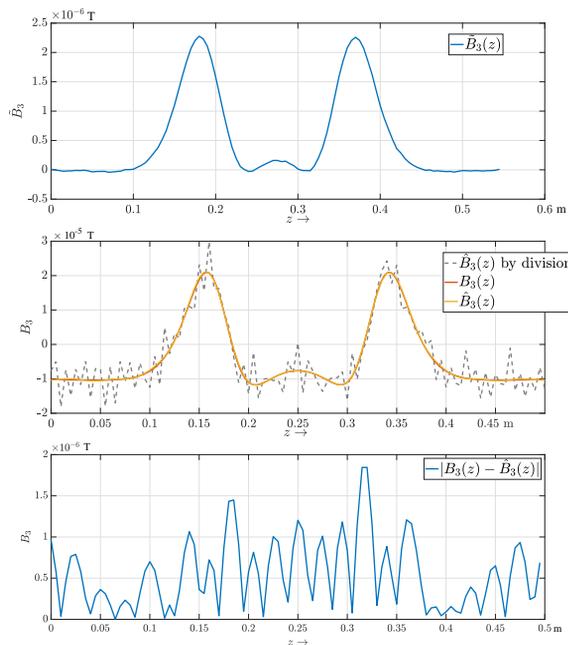


Figure 7: Top: Simulated  $\tilde{B}_3$  along the axis of the magnet shown in Fig. 1 with added Gaussian random noise, signal-to-noise ratio of  $10^{-4}$ . Center: The deconvolution using the inverse of the sensitivity spectrum and a filter  $g(f)$  given by Eq. (15) compared to the simulated  $B_3(z)$ . Bottom: Absolute error  $|B_3(z) - \hat{B}_3(z)|$ .

spectrum of  $s_n(f)$  leads to an amplification of the noise. The absolute error between  $B_3(z)$  and the Wiener deconvolution lies below 2.5% of the maximum of  $B_3(z)$ .

## DESIGN OF EXPERIMENT

The most sensitive parameters for the deconvolution of the measurand are the coil length and the sampling step-size. In order to be sensitive in all relevant frequencies of  $B_n(f)$ , the spectrum of the sensitivities impulse response  $s(f)$ , i.e., the frequency spectrum of the induction coil's sensitivity function, should be non-vanishing in the frequency band

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

imposed by the highest spatial frequency in the measurand. To optimize the coil length (or estimate whether or not an available transducer has high enough sensitivity) we consider the sensitivity function as a rectangular pulse in the spacial domain. Its spectrum will have the shape of a  $(\sin x)/x$  function, containing zeros at the frequencies  $f_k = k/l_s$ , for  $k = \{0, 1, \dots, K\}$ , where  $l_s$  is the length of the pulse, i.e., the hard-edge model of the induction coil. An infinitely short and highly sensitive coil, corresponding to a Dirac-shaped impulse response, is technically not possible. The maximum sensitivity is limited by the number of turns in one layer of the flexible PCB. As shown above, a minimum flux linkage of  $\Phi_{\min} = 10^{-8}$  Vs in the main induction coil is required. Denoting the minimum required accuracy of field harmonic by  $B_{\min}$ , the minimum length of the induction coil can be estimated to  $l_{s,\min} = \Phi_{\min}/(h_s B_{\min})$ , where  $h_s$  is the height of the coil's impulse response in the hard-edge model. Therefore the optimum coil length  $l_s$  is given by

$$\frac{\Phi_{\min}}{h_s B_{\min}} < l_s < \frac{1}{f_{n,\max}}, \quad (16)$$

where  $f_{n,\max}$  denotes the expected (computed) highest frequency in  $B_n(f)$ . In other words, the coil must be short enough to resolve the highest spatial frequency, but long enough to accommodate enough turns for a sufficiently high sensitivity  $S_n$ . In this case, the shape of the sensitivity function is not critical for the Wiener deconvolution. Simulations show that a smooth roll-off in the coil sensitivity is even preferable; for the magnetic field distribution shown in Fig. 7 the optimal ratio between the coil's straight section  $l_1$  and its overall length  $l_s$  is about 0.5. This yields sufficient flexibility for the coil design. A higher sampling rate will lead to a better resolution of the multipole-field distribution, as the maximum meaningful frequency will increase by the Nyquist sampling theorem. This implies that the maximum step-size be

$$\Delta z \leq \frac{1}{2f_{n,\max}}. \quad (17)$$

The minimum step-size is, however, constraint by the positioning accuracy of 0.1 mm. Resolving frequencies of  $s(f)$  higher than  $f_{n,\max}$  yields no improvement.

## CONCLUSION

The theory of pseudo-multipoles is known from the literature. The extraction of the leading term in the Fourier-Bessel series requires the solution of a differential equation by means of a discrete Fourier transform. This yields a natural way to deconvolute the measured distribution of the

multipole content. We have studied and presented the limitations of point-like measurements with Hall sensor stacks and thus motivated the measurement technique using induction-coil sensors. These require a novel design employing saddle-shaped, iso-perimetric coils in order to avoid interception of the axial field component.

The compensation of the main signal cannot be accomplished with the classical arrangement of tangential (or radial) induction coils at different radii, because no easy scaling law exists.

The study of the Wiener filter for the signal deconvolution allows a design of experiment based on the optimal number of coefficients and required  $z$ -derivatives of the leading terms, and the step size in the field-scanning process. Unfortunately, the signal-to-noise ratio is nowhere near the values obtained with standard rotating coil sensors.

Nevertheless, and to our knowledge for the first time, it is possible to extract the transversal field components (and only those) from measurements in the coil-end regions. A reasonable approach will therefore be to validate the design and construction of accelerator magnets using such sensors and to gauge the numerical (FEM, BEM) models for the use in beam-tracking studies. These can then be performed using simulated field distributions that are sufficiently smooth to extract higher-order pseudo-multipoles.

With the presented methodology and hardware it will also be possible to better characterize fast-ramping magnets and magnets with strong hysteresis effects, where  $10^{-4}$  predictive models do not exist.

## REFERENCES

- [1] Arpaia, P., Buzio, M., De Matteis, E., Russenschuck, S., A rotating coil transducer for magnetic field mapping, *JINST*, 2015.
- [2] Russenschuck, S., *Field Computation for Accelerator Magnets*, Wiley-VCH, 2010.
- [3] Berz, M., Joh, K., Nolen, J. A., Sherrill, B. M., Zeller, A. F., Reconstructive correction of aberrations in nuclear particle spectrographs, *Phys. Rev. C*, 537, 1993.
- [4] Dragt, A. J., *Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics*, University of Maryland, 2011.
- [5] Erdelyi, B., Berz, M., Lindemann, M., Differential algebra based magnetic field computations and accurate fringe field maps, American Physical Society, April Meeting, 2000, Long Beach, CA.
- [6] Caspi, S., Helm, M., Laslett, L. J., Brady, V. O., An approach to 3D magnetic field calculation using numerical and differential algebra methods, LBL-32624, 1992.

# MUON BACKGROUND STUDIES FOR BEAM DUMP OPERATION OF THE K12 BEAM LINE AT CERN

M. Rosenthal\*, D. Banerjee, J. Bernhard, M. Brugger, N. Charitonidis, M. van Dijk, B. Döbrich,  
L. Gatignon, A. Gerbershagen, E. Montbarbon, B. Rae

European Organization for Nuclear Research (CERN), 1211 Geneva 23, Switzerland

T. Spadaro

Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati (INFN/LFN), Frascati, Italy

## Abstract

In the scope of the Physics Beyond Colliders study at CERN a future operation of the NA62 experiment in beam dump mode is discussed, enabling the search for dark sector particles, e.g. heavy neutral leptons, dark photons and axions. For this purpose, the 400 GeV/c primary proton beam, extracted from the SPS, will be dumped on a massive dump collimator located in the front end of the K12 beam line. Muons originating from interactions and decays form a potential background for this kind of experiment. To reduce this background, magnetic sweeping within the beam line is employed. In this contribution, the muon production and transport has been investigated with the simulation framework G4beamline. The high computational expense of the muon production has been reduced by implementing sampling methods and parametrizations to estimate the amount of high-energy muons and efficiently study optimizations of the magnetic field configuration. These methods have been benchmarked with measured data, showing a good qualitative agreement. Finally, first studies to reduce the muon background by adapting the magnetic field configuration are presented, promising a potential background reduction by a factor four.

## INTRODUCTION

The North Area at the Super Proton Synchrotron (SPS) at CERN has a long history of fixed target experiments and R&D studies. Extracted from the SPS, a 400 GeV/c proton beam is directed via transfer lines to two experimental halls (EHN1, EHN2) and an underground cavern (ECN3) located at the CERN Prévessin site. In ECN3, a high-intensity secondary hadron beam, that has been created at a beryllium target, is transported via the K12 beam line towards the NA62 experiment [1]. This experiment aims to measure the branching ratio of the very rare decay  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ . First results of this operation have been presented recently [2]. Besides the main measurement program, a future proposal for NA62 suggests the search for dark sector particles such as heavy neutral leptons, dark photons and axions in “dedicated dump runs” [3]. For this purpose, the beryllium target will be removed and the primary proton beam will be dumped on a 3.2 m long massive dump collimator (TAX) in order to create the hypothetical dark sector particles. The decay of these particles into Standard Model particles, e.g. muons in

the final state, might be observable by the NA62 experiment. The implementation of this proposal is studied by the the Conventional Beams Working Group (CBWG) within the Physics Beyond Colliders (PBC) framework. Muons directly produced in the primary interactions within the TAX pose an crucial background for this kind of experiment, e.g. through random spatiotemporal track combinations of muons and antimuons. Studies to understand the trajectories of this muon background are essential to further optimize the magnetic sweeping of the K12 beam line, when operated in beam dump mode. To investigate and reduce this muon background, the optimisation of the magnetic sweeping along the K12 beam line is performed. Monte Carlo studies based on the program G4beamline [4] have been combined with analytical parametrisations of the muon distributions to reduce the computational demands. In this contribution, benchmarking results with recorded data as well as first results from the optimization studies are shown.

## THE K12 BEAM LINE MODEL

The simulation of production and transport of the muon background is computationally highly expensive and requires the precise knowledge of the magnetic field maps in the entire K12 beam line. A G4beamline model, including a simplified model of the NA62 experiment, has been developed to investigate the particle production, transport and decay of the particles in the beam line. It is illustrated in Fig. 1. The 400 GeV/c protons with a nominal intensity of  $3 \times 10^{12}$  protons per burst impinge on a 400 mm-long beryllium target (T10), which corresponds to about one nuclear

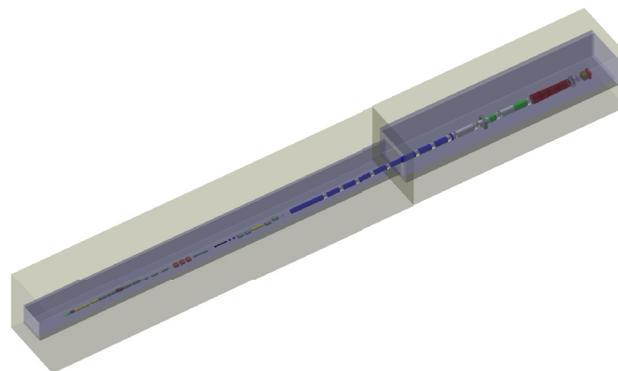


Figure 1: The G4Beamline model of the K12 beam line and NA62 experiment.

\* marcel.rosenthal@cern.ch

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

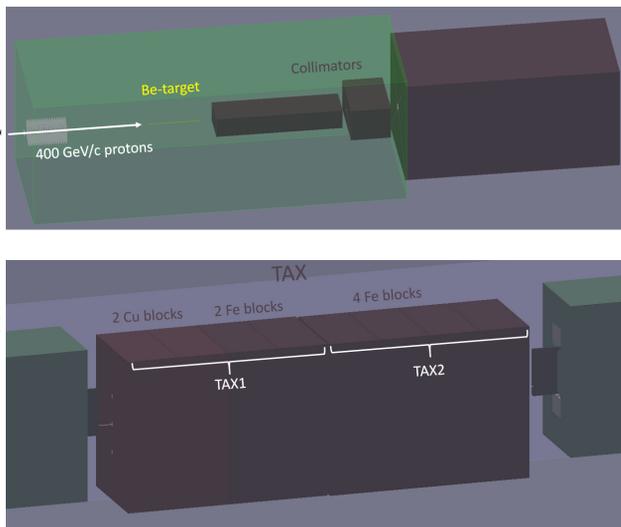


Figure 2: Model of the T10 target region (up) and TAX (down). The 400 GeV/c proton beam enters from the left in the upper figure. The green volume illustrates an air volume, in which an upstream monitor, the T10 target and two collimators are located. The TAX consists of 2 copper and 2+4 iron blocks with bores for the beam passage.

interaction length. The target is followed by a set of collimators and a small-aperture, radiation-hard quadrupole triplet defining the maximum acceptance of the secondary beam, mostly consisting of pions, protons and kaons. To maximize the number of signal events in the fiducial decay volume of the NA62 experiment, the beam line optics is tuned for a transfer of positively charged particles with a momentum of 75 GeV/c [5]. Behind the quadrupole triplet, a momentum selection is performed, mainly by a 3.2 m long massive dump collimator (TAX) centered in a four bend achromat. The model of the target region and the TAX are shown in Fig. 2. The first pair of bends of the achromat induces a vertical deflection to the charged particle beam, which amounts to  $-110$  mm for the 75 GeV/c-beam fraction, allowing it to pass the 1-cm diameter bores in the TAX and being returned to the central axis by the second pair of bends. Other momentum slices, especially the non-interacting primary 400 GeV/c protons, are dumped on the TAX with different vertical offsets, i.e.  $-20.625$  mm for the primary protons. Muons mainly created in pion and kaon decays upstream of the TAX are not stopped and form a halo around the selected hadron beam. Various magnetic elements, i.e. muon sweeping dipole magnets with a field-free region for the hadron beam passage and scraping magnets with a toroidal field around the axis are employed to reduce this halo before the beam enters the decay volume of the NA62 experiment at about 102 m downstream the production target. The detailed study of muon halo requires the simulation of the magnetic fields within the magnet apertures but also in the iron yokes in the entire beam line. For the current version of the model these field distributions have been extracted from existing measurements and simulations previously im-

plemented in the simulation tool HALO [6]. Within the experiment, several detector components are employed to identify the particles leaving the decay volume, i.e. four straw spectrometer stations for track reconstruction and momentum identification, two charged particle hodoscopes and a muon veto system at the end of the line. A more detailed description of the beam line and the various sub-detectors can be found in Ref. [1].

### Nominal Configuration

The current implementation of geometry has been validated with a simulation based on nominal beam optics for the current NA62 operation. Figure 3 illustrates the distribution of the pion momentum in front of the decay volume. It is centered around 75 GeV/c with a momentum spread of 1.15%. The latter can be further adjusted by additional momentum-selection collimators within the beam line. Similar shapes are observed for kaons and protons. The corresponding particle rates (Table 1) obtained with the FTFP\_BERT physics list of Geant4 were compared to the reference values estimated in Ref. [1] using the HALO software together with the Atherton formula for particle production [7]. Both estimates yield pion and kaon rates in very good agreement within a few percent, while a slightly enhanced proton rate

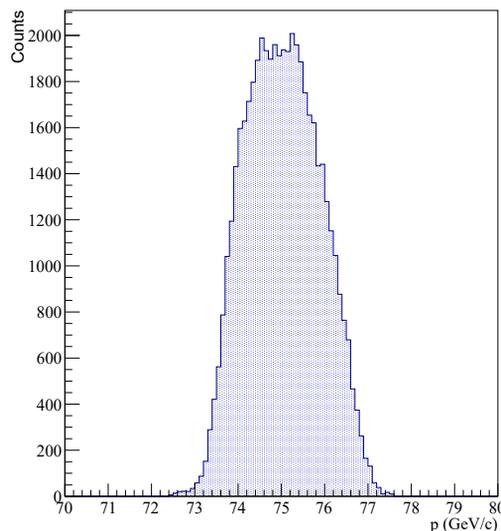


Figure 3: Simulated pion momentum distribution at  $z = 102$  m for  $10^8$  protons on target. The nominal beam is centered around 75 GeV/c with a momentum spread of 1.15%.

Table 1: Fluxes at 102 m After Target per  $1.1 \times 10^{12}$  Incident Protons per Second

Particle	Reference [1] (MHz)	G4beamline (MHz)
Pions	525	547
Protons	173	308
Kaons	45	45

is observed in the Geant4 based simulation with respect to the reference estimate.

### Beam Dump Configuration

In dedicated beam dump configuration, the beryllium target is removed from the beam line to reduce the muon halo background originating from the production and decay of pions and kaons upstream of the TAX. The residual material budget in the target region is reduced to about 1% compared to the nominal setup including the beryllium target [8]. The remaining material is modeled by the vacuum windows, beam instrumentation elements and the surrounding air in the target region, as shown in Fig. 2. After passing this region, the primary proton beam passes the quadrupole triplet and is dumped directly on the TAX. The two independently moveable parts of the TAX (TAX1+TAX2) are vertically displaced with respect to each other to prevent the passage for any particle coming from the upstream direction of the beam line. The muon background investigated in this study is composed from the contributions of upstream decays and the production within the TAX.

## MUON SIMULATION ALGORITHMS

The simulated muon spectrum and spatial distribution obtained directly after the TAX is depicted in Fig. 4. It shows the concentration around the proton impact point, that is vertically shifted due to the first two bends of the first achromat. The radial symmetry is slightly broken due to the existing tungsten inserts around the bores, which are located above and below the proton impact point. Furthermore, a rate decrease with increasing momentum of several orders of magnitude is observable, drastically limiting the statistics

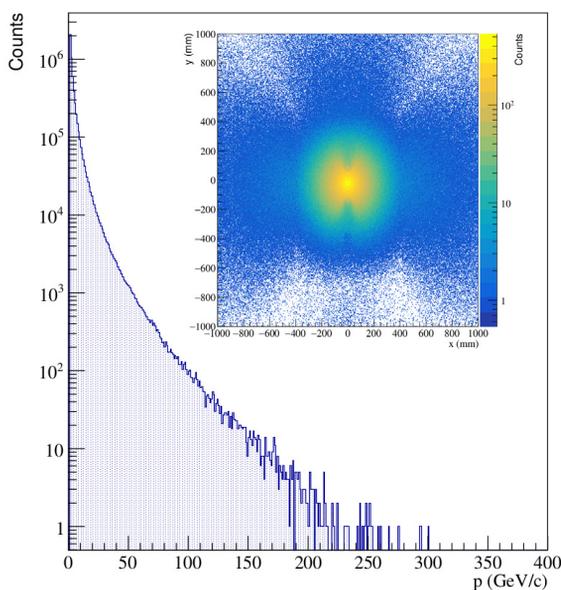


Figure 4: Simulated muon spatial distribution and spectrum after the TAX obtained for  $10^9$  incident protons.

for high-energy muons. Since the dump simulation of the 400 GeV/c protons is computationally highly expensive (currently  $\sim 10$  protons per second for an illustrated sample of  $10^9$  incident protons), an increase of the number of incident protons for the study of multiple magnetic configurations is not reasonable. Instead, two methods based on the recorded particle information directly after the TAX have been employed to reduce the computational effort and increase the statistics for higher momenta.

### Method 1: Sampling

In the first method, the recorded particles behind the TAX are binned according to position, momentum and type. From the number of particles in each multi-dimensional bin a probability distribution is deduced to generate new and individually-sized samples representing the particle distribution after the TAX. These samples are employed to study the particle transport starting after the TAX removing the computationally expensive simulations of the interacting protons in the TAX. The limiting factor of this method is the underlying particle sample. Due to the low statistics for higher momenta as shown in Fig. 4, the calculated probability distribution is not representing the full phase-space for these particles. Thus, certain momenta appear only at similar locations or are entirely missing.

### Method 2: Parametrization

A second method has been implemented in order to reduce the variance for higher momenta. For that purpose, the muons and antimuons created in and recorded directly after the TAX are first sorted in bins of different longitudinal momenta. The horizontal and vertical phase space of each bin can be parametrized by a two dimensional Gaussian distributions in each bin, respectively. This is shown exemplary in Fig. 5 for a longitudinal momentum of  $p_z = 25$  GeV/c and the horizontal phase space. The evolution of the parameters describing the Gaussian distributions can be illustrated with respect to the longitudinal momentum, as depicted exemplary for the horizontal width of the antimuon distribution in Fig. 5. This evolution is further parametrized by analytical functions up to a longitudinal momentum of 80 GeV/c and extrapolated for momenta up to 350 GeV/c to estimate the high-momenta muon distribution, where the statistics of the underlying sample is limited. To additionally increase the statistics of the high-momenta muons, the new samples are created uniformly distributed in momentum, but a weighting factor representing the probability of appearance is assigned. This allows for a re-normalization of the simulated muons recorded in the NA62 detectors, after the tracking through the K12 beam line elements downstream of the TAX is performed. Bases on this method, samples representing the major contributions of the muon background are created.

### Comparison

To validate the algorithms, the generated samples have been compared with each other. Figure 6 depicts the momentum spectrum and vertical distribution of antimuons

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

directly after the TAX obtained with both methods and with similar number of particles in the samples. The momentum spectrum is very well represented in both methods, but for the second method the variations at higher momenta could be drastically reduced. Simultaneously, also the vertical distribution, the core below 200 mm is in very good agreement between both methods, while larger relative differences are only observable at bigger distances. A similar effect is observable in the horizontal distribution. Overall, a satisfactory

agreement between both methods could be shown. Due to the reduced computational expenses and a better representation of the high-momenta muons, the second method is used in the muon background studies described in the following.

### COMPARISON TO MEASURED RATES

Based on the previously described method, a comparison to measured data from the NA62 experiment has been performed. For this purpose, recorded data taken in short runs in beam dump mode has been used. The trigger selection

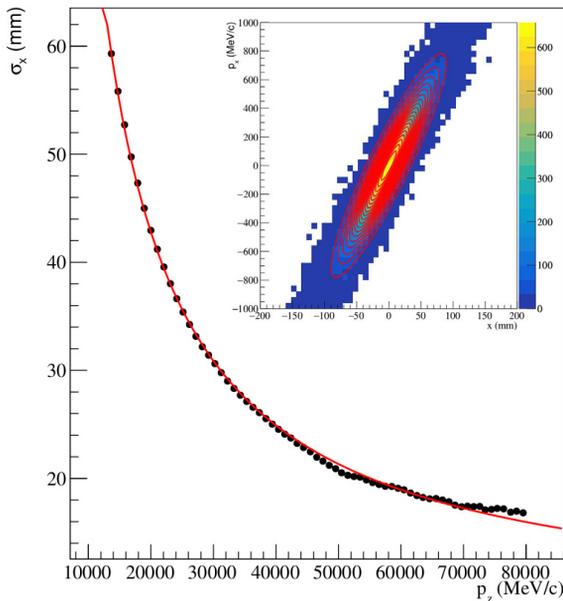


Figure 5: Parametrization of the transversal width  $\sigma_x$  for different longitudinal momenta  $p_z$ .

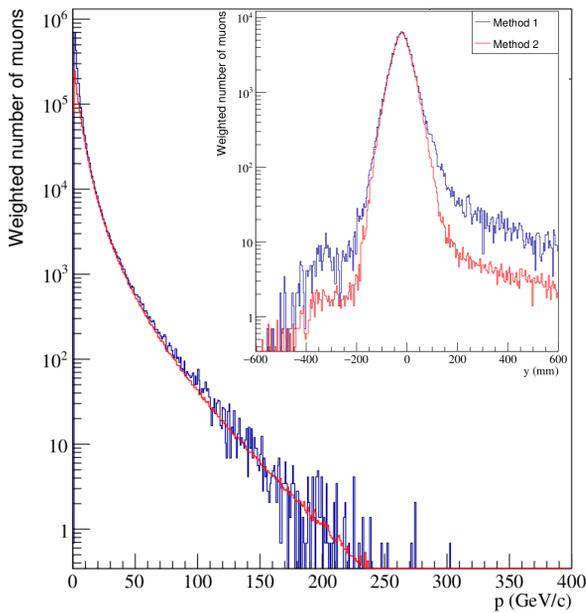


Figure 6: Simulated antimuon momentum spectrum and vertical distribution after the TAX for both methods. The distributions are normalized to  $10^9$  incident protons.

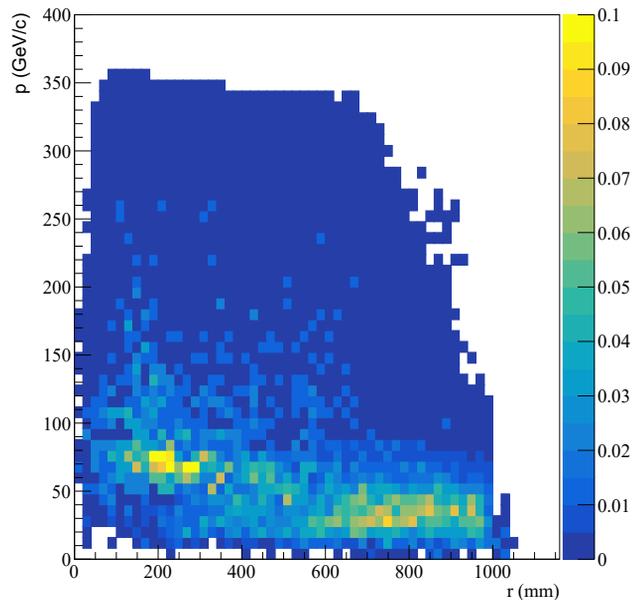
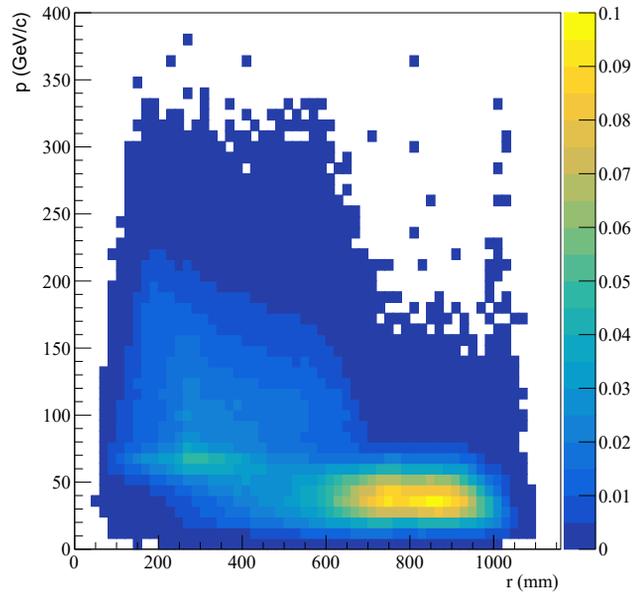


Figure 7: Comparison of measured (up) and simulated distributions (down) of positively charged muons in beam dump configuration reconstructed at 180 m, close to the first straw chamber of NA62. The color scale represents the number of tracks per  $10^9$  incident protons and bin. The measured data is downscaled by a factor five.

required at least one hit in the charged particle hodoscope (CHOD) of NA62. The charged particle tracks were reconstructed using the information of the straw detector chambers, yielding the momentum, the charge, and the position at the first straw chamber. These information were used to compare the spatial distributions and spectra of positive and negative tracks between recorded data and simulation. Since the simulation currently focuses on the study of muons, a time-associated hit in the muon veto detector MUV3 was requested for the reconstructed tracks. Figure 7 depicts the radius  $r = \sqrt{x^2 + y^2}$  and momentum of the positive charged tracks obtained in measurement and simulation at 180 m, close the first straw chamber of NA62. Both distributions feature two accumulations in similar regions. In separate studies of the muon distribution originating in upstream decays, the accumulation at about 75 GeV/c and a radius of 250 mm could be associated to the muon production before the TAX, while the large accumulation at lower momenta and larger radii at about 800 mm originates from processes occurring within the TAX. The good qualitative agreement validates the implementation of the magnetic configuration in the G4beamline model. Quantitatively, the calculated muon rates in the data sample are a factor five larger than observed in the present simulation, showing an enhanced contribution from muons originating from the TAX region in the recorded data. This excess is investigated in present studies.

## FIRST RESULTS OF OPTIMIZATION STUDIES

To further reduce the muon background in beam dump operation, optimization studies of the magnetic layout of the beam line have been performed. For this purpose, the muon distribution at the MUV3 detector at the end of the NA62 experiment  $z \approx 247$  m has been investigated. The simulated muon momentum spectrum at the MUV3 detector plane using this configuration is illustrated in Fig. 8. A similar spectrum for muons and antimuons is observed in the present nominal beam dump configuration. The extrapolation boundary for higher momenta leads to the hard cut at 350 GeV/c. Figure 9 depicts the spatial distribution on the MUV3 detector plane in initial configuration.

The entire sensitive region of the MUV3 detector is populated by muons. The accumulation in the vicinity of the beam pipe is dominated by a contribution of muons with a slightly lower momentum than 75 GeV/c. This is related to the beam optics optimization for 75 GeV/c in the nominal configuration, leading to a favored transport of positively charged muons in this momentum region. Energy losses slightly reduce the final momentum by a few GeV/c.

First studies aim for a reduction of the muon background at the MUV3 plane applying only a minimal amount of changes to the existing beam line. To eliminate the 75 GeV/c-component observed from the muons in upstream decays, the first quadrupole triplet is optimized for the transport of the 400 GeV/c protons and the quadrupoles downstream of

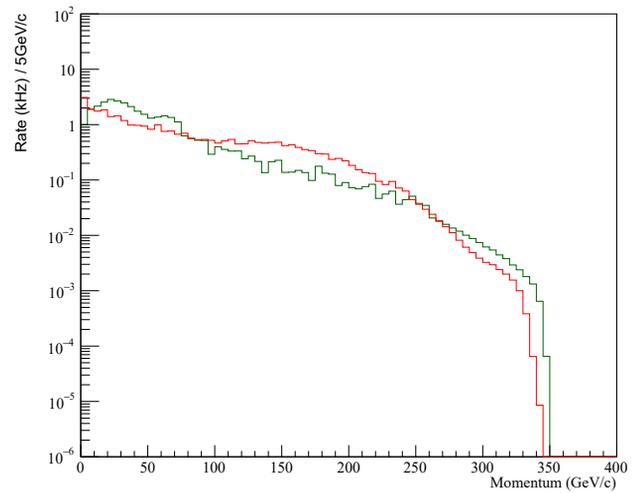


Figure 8: Simulated momentum spectrum of muons (red) and antimuons (green) at the MUV3 detector for the initial configuration in beam dump mode. The rate is normalized to  $1.1 \times 10^{12}$  protons per eff. second. A cutoff at 350 GeV/c is applied.

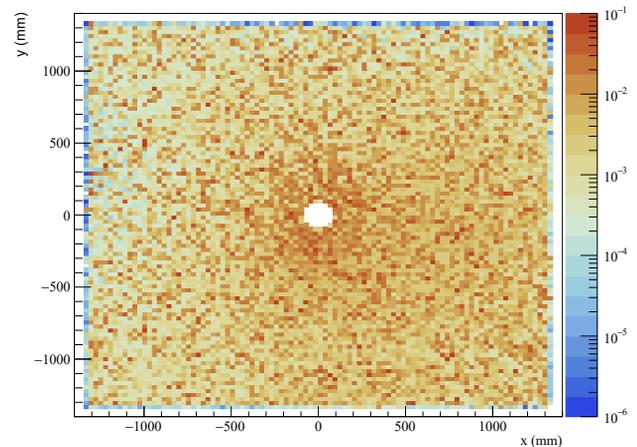


Figure 9: Simulated muon spatial distribution at the MUV3 detector for the initial configuration in beam dump mode. The color scale represents the number of hits per  $10^9$  incident protons and bin.

the TAX are turned off. Additionally, a deflection close to the production location in the TAX is preferable to further improve the muon sweeping. For that purpose, the magnetic setup of the first achromat surrounding the TAX is investigated. The first two bending magnets are turned off, letting the proton beam impinge centrally on the TAX. The magnetic field of the second two dipoles is scanned to study the influence on the remaining muon rate with respect to the initial nominal configuration in beam dump mode. The results are shown in Fig. 10. The highest rate is observed for disabled sweeping within the first achromat. In this case, the muon flux is increased by a factor five. Operating the magnets with same strengths but opposite sign shows a general enhancement of the muon rates, since the angular deflection of the consecutive magnets cancels each other. For same sign, the muon rates are reduced by increasing the

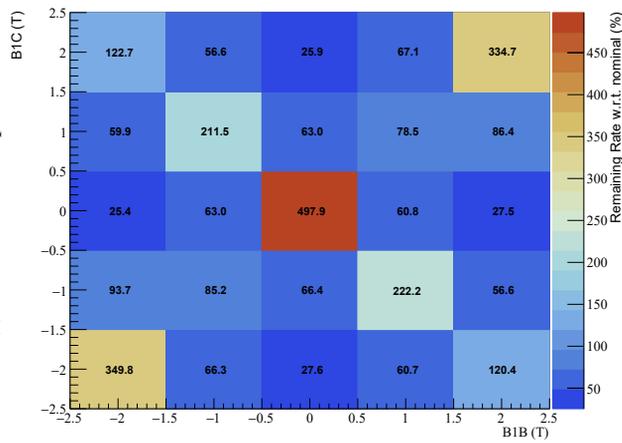


Figure 10: Remaining muon rate with respect to the initial configuration in beam dump mode. The two bending magnets after the TAX (B1B,B1C) have been scanned in steps of 1 T field strength in the center of the gap.

field strengths up to a certain point. At this point, the deflection for low momenta muons in the first bend is sufficiently strong to let them reach the iron yokes of the second bend. Thus, the return field of the second bend deflects them back towards the sensitive part of the muon veto detector. This significantly enhances the rate due to the large amount of low-momenta muons. For maximum field of the the first magnet at  $-1.82$  T, an optimum for the second magnet is found at about  $-0.3$  T. The spatial distribution for this particular setting is depicted in Fig. 11.

The accumulation close to the beam pipe disappeared due to the change of the magnetic configuration. The observable horizontal deflection arises from the three muon sweepers with vertical magnetic field, while the left-right asymmetry is related to the optimized sweeping of positive muons in the nominal setup using a muon scraper with toroidal field that is located downstream in the K12 beam line. Overall, the total simulated muon flux is reduced by a factor four in this setting compared to the nominal dump configuration. Taking

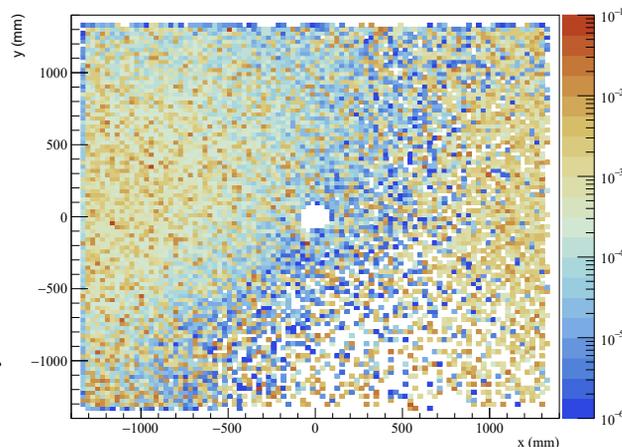


Figure 11: Simulated muon spatial distribution at the MUV3 detector for the optimized configuration of the first achromat. The color scale represents the number of hits per  $10^9$  incident protons and bin.

only muons above a threshold of  $15$  GeV/c into account, the reduction can be further maximized by operating the two last bending magnets of the first achromat at their maximum field of  $-1.82$  T. In this configuration the simulated muon rate above  $15$  GeV/c is reduced by a factor 20, but the rate of low-momenta muons is enhanced simultaneously, as explained above.

## CONCLUSION & OUTLOOK

The study of the muon background in beam dump operation of the NA62 experiment required the development of a new model of the K12 beam line that transports the beam to this experiment. Using the available magnetic field maps, this model has been implemented in the G4beamline software framework, which allows for investigation of the production, transport and decay of particles. The production of the initial muon population created by the dumped  $400$  GeV/c proton beam is strongly suppressed and hence requires a substantial amount of computational power. Consequently, methods reducing the computational time by estimates of the distribution at higher momenta have been successfully implemented. Benchmarking of the simulated muon background distribution with already measured data showed a promising qualitative agreement. A further reduction by a factor four could be achieved by small changes to the beam line, mainly by the optimization of the first two bending magnets behind the TAX. These results will be further used to quantify the benefit of bigger upgrades of the K12 beam line for an operation of the NA62 experiment in beam dump configuration.

## REFERENCES

- [1] NA62 Collaboration, “The Beam and detector of the NA62 experiment at CERN”, *JINST*, vol. 12, no. 05, p. P05025, 2017.
- [2] R. Marchevski on behalf of the NA62 Collaboration, “ $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ : first NA62 results”, *Moriond EW Conference*, La Thuile, Italy, 10–17 March 2018.
- [3] B. Döbrich on behalf of the NA62 Collaboration, “Searches for very weakly-coupled particles beyond the Standard Model with NA62”, in *Proc. 13th Patras Workshop on Axions, WIMPs and WISPs*, p. 145–148, 2017.
- [4] T. J. Roberts, K. Beard, D. Huang, S. Ahmed, D. M. Kaplan, and L. Spentzouris, “G4beamline Particle Tracking in Matter-dominated Beam Lines”, in *Proc. EPAC’08*, paper WEPP120, 2008.
- [5] NA62 Collaboration, “NA62 Technical Design Report”, December 2010.
- [6] Ch. Iselin, “HALO: a computer program to calculate muon halo”, *CERN Yellow reports*, 1974.
- [7] H. W. Atherton *et al.*, “Precise measurements of particle production by  $400$  GeV/c protons on beryllium targets”, *CERN Yellow reports*, 1980.
- [8] L. Gatignon and N. Doble, “Comparisons T10 Target IN versus Target OUT with K12 TAX Open or Closed”, *Internal Note*, 2016.

# SIMULATION CHALLENGES FOR eRHIC BEAM-BEAM STUDY\*

Yun Luo, BNL, Upton, NY, USA

Yue Hao, FRIB, East Lansing, MI, USA

Ji Qiang, LBNL, Berkeley, CA, USA

Yves Raymond Roblin, JLab, Newport News, VA, USA

## Abstract

The 2015 Nuclear Science Advisory Committee Long Range Plan identified the need for an electron-ion collider (EIC) facility as a gluon microscope with capabilities beyond those of any existing accelerator complex. To reach the required high energy, high luminosity, and high polarization, the eRHIC design, based on the existing heavy ion and polarized proton collider RHIC, adopts a very small  $\beta$ -function at the interaction points, a high collision repetition rate, and a novel hadron cooling scheme. A full crossing angle of 22 mrad and crab cavities for both electron and proton rings are required. In this article, we will present the high priority R&D items related to the beam-beam interaction studies for the current eRHIC design, the simulation challenges, and our plans and methods to address them.

## INTRODUCTION

The key EIC machine parameters identified in the 2015 Long Range Plan [1] are: 1) polarized (70%) electrons, protons, and light nuclei, 2) ion beams from deuterons to the heaviest stable nuclei, 3) variable center of mass energies  $\sim 20\text{--}100$  GeV, upgradable to  $\sim 140$  GeV, 4) high collision luminosity  $\sim 10^{33}\text{--}10^{34}$   $\text{cm}^{-2}\text{sec}^{-1}$ , and possibly have more than one interaction region. To reach such a high luminosity, both designs of eRHIC at Brookhaven National Laboratory (BNL) and JLEIC at Thomas Jefferson National Accelerator Facility (JLab) aimed to increasing the bunch intensities, reducing the beam sizes at the interaction points (IPs), and increasing the collision frequency, while keeping achievable maximum beam-beam parameters for involved beams [2, 3].

The relative priorities of R&D activities for a next generation EIC were published in the 2016 NP Community EIC Accelerator R&D panel report [4]. The panel evaluated the R&D items needed for each of the current EIC design concepts under considerations by the community. Beam-beam interaction have been identified as one of the most important challenges needed to be addressed to reduce the overall design risk.

We join the expertise from BNL, JLAB, Lawrence Berkeley National Laboratory (LBNL), and Michigan State University (MSU) to address 4 challenging items related to the EIC beam-beam interaction in the two EIC ring-ring designs, namely, 1) beam dynamics study and numerical simulation of crabbed collision with crab cavities, 2) quantitative understanding of the damping decrement to the beam-beam performance, 3) impacts on protons with electron bunch

swap-out in eRHIC ring-ring design, and 4) impacts on beam dynamics with gear-changing beam-beam interaction in JLEIC design.

To address the above critical items related to EIC beam-beam interaction, we propose new beam-beam simulation algorithms and methods to the existing strong-strong beam-beam simulation codes, together with a deep physics understanding of the involved beam dynamics. At the completion of this proposal, we should have a clear understanding of the beam-beam interaction in the next generation EIC designs and be able to provide robust counter-measures to possible beam-beam interaction related beam lifetime reduction, beam emittance growth, beam instabilities, and luminosity degradation. This work would significantly mitigate the technical risks associated with the EIC accelerator designs.

In this article, we will only focus on the simulation challenges related to the eRHIC beam-beam study, or the first three R&D items listed above. They are the common challenges to the eRHIC and JLEIC designs. JLEIC design also have another challenge: impacts on beam dynamics with gear-changing beam-beam interaction, which will not be discussed here.

## eRHIC DESIGN PARAMETERS

For the present eRHIC design, the maximum beam-beam parameters for the electron and proton beams are  $\xi_e = 0.1$  and  $\xi_p = 0.015$ , respectively. The choice of the beam-beam parameter of  $\xi_e = 0.1$  for the electron beam is based on the successful operational experience of KEKB, where it was achieved with a transverse radiation damping time of 4000 turns. The choice of the beam-beam parameter for the proton ring is based on the successful operational experience of RHIC polarized proton runs, where a beam-beam parameter of  $\xi_p = 0.015$  was routinely achieved.

To avoid long-range collisions, a crossing-angle collision scheme is adopted. For the present design, the proton and electron beams collide with a total horizontal crossing angle of 22 mrad. Such a crossing angle scheme is also required by the experiment to avoid separator dipoles in or near the detector, thus minimizing the background in the interaction region (IR). To compensate the luminosity loss by the crossing angle collision, crab cavities are to be used to tilt the proton and electron bunches such that they collide head-on at the IP. Table 1 shows key beam-beam interaction related parameters of the current eRHIC design. Without cooling, the design luminosity is  $4.4 \times 10^{33} \text{cm}^{-2}\text{s}^{-1}$ . With cooling in the proton ring, it is  $1.05 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ .

\* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Table 1: Machine and Beam Parameters for eRHIC Design

Parameter	Unit	Proton ring	Electron ring
Circumference	m	3833.8451	
Energy	GeV	275	10
Bunch Intensity	$10^{11}$	1.05	3.0
Working point	-	(29.31, 30.305)	(51.08, 48.06)
synchro. tune	-	0.01	0.069
$\beta_{x,y}^*$	cm	(90,5.9)	(63, 10.4)
rms emittance	nm	(13.9,8.5)	(20,4.9)
Bunch length	cm	7	1.9
Energy spread	$10^{-4}$	6.6	5.5
Crossing angle	mrad	22	

To compensate the geometric luminosity loss due to the crossing angle, crab cavities are to be installed to tilt the proton and electron bunches by 11 mrad in the  $x - z$  plane at IPs so that the two beams collide head-on. The crab cavities provide a horizontal deflecting force to the particles in a bunch. Ideally, the deflecting electric field should be proportional to the longitudinal position of particles. For the local crabbing scheme, the horizontal betatron phase advances between the crab cavities and IP are  $\pi/2$ . The voltage of the crab cavity is

$$\hat{V}_{RF} = -\frac{cE_s}{4\pi f_{RF}\sqrt{\beta_x^*\beta_{cc}}}\theta_c. \quad (1)$$

Here  $c$  is the speed of light,  $E_s$  is the particle energy in eV,  $f_{RF}$  is the crab cavity frequency, and  $\theta_c$  is the full crossing angle.  $\beta_x^*$  and  $\beta_{cc}$  are the horizontal  $\beta$  functions at the IP and the crab cavity, respectively.

A higher frequency of crab cavities requires a lower crab cavity voltage. However, due to the sine wave shape of the crab cavity voltage, particles in the bunch tail may not be perfectly crabbed. In the following, we assume 338 MHz for the crab cavities in both proton and electron rings. The final choice of the crab cavity frequency is not made yet.

With crabbed collision between the electron and proton bunches, we focus on the emittance growth and luminosity degradation. For this purpose, we combine strong-strong and weak-strong beam-beam simulation methods. The strong-strong beam-beam simulation is used to reveal any possible coherent beam-beam instability in a few electron damping periods. If there is no clear coherent beam-beam motion from the strong-strong beam-beam simulation, then a weak-strong beam-beam simulation is to be used to evaluate the long-term stability of the protons. In the weak-strong simulation, the equilibrium electron beam sizes from the strong-strong simulation are used.

## SIMULATION CHALLENGES

### *Dynamics Study and Numerical Simulation of Crabbing Collision with Crab Cavities*

For collision with a crossing angle and crab cavities, when the bunch length is comparable with the wavelength of the

crab cavity, the sinusoidal form of the crab-cavity voltage may lead to the transverse deviation of particles at the head and tail as the function of the longitudinal position of the particles. As an example, Figure 1 shows the proton and electron bunch profiles in the  $x - z$  plane in the head-on collision frame.

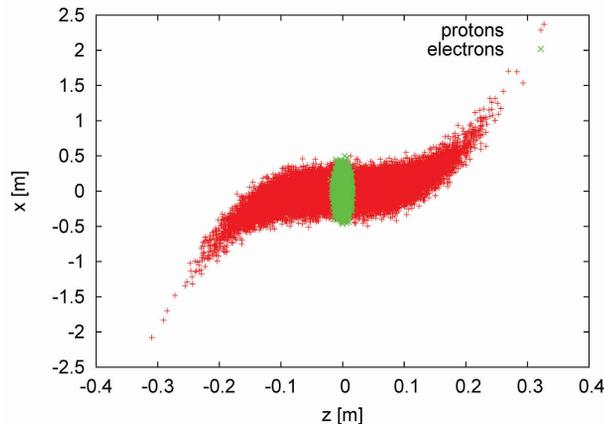


Figure 1: Electron and proton bunch profiles in the head-on frame.

In 2017, supported by the NP proposal award (PI: Yue Hao and Ji Qiang), a special synchro-betatron resonance, which coupled through beam-beam interaction, was found due to the imperfect crab kick, using a strong-strong beam-beam simulation code [5]. The resonance raises from the beam-beam induced coupling between the transverse motion of the electron beam and the synchrotron motion of the proton beam, and causes luminosity reduction of  $\sim 1\%$  per second from the simulation, which depends on the frequency of the crab cavity and the proton synchrotron tunes as shown in Figure 2.

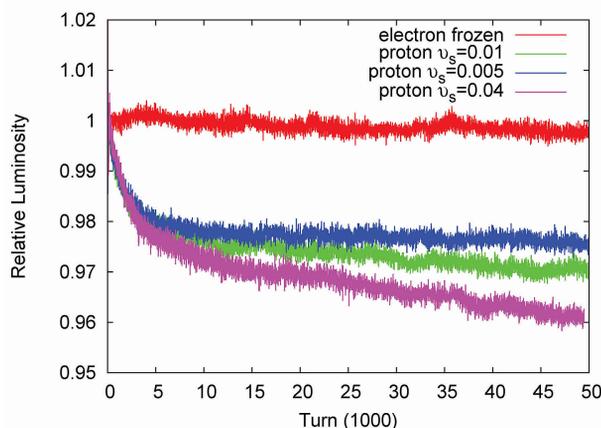


Figure 2: Luminosity degradation as function of proton synchrotron tune.

As we know, numerical noise in the self-consistent strong-strong beam-beam simulation can cause artificial emittance growth and may block the true physics driven emittance growth. Currently, the computational method used

in those simulations is based on a particle-in-cell method with Green's function to calculate the beam-beam force self-consistently.

To verify the small emittance growth observed from the strong-strong simulations, the most challenging task is to separate the beam degradation due to the nonlinear resonance from the artificial emittance growth induced by the numerical noise in the strong-strong beam-beam simulation code. The numerical noise reduction is an essential step for the further understanding of the EIC crab crossing scheme.

At present, the only crab crossing scheme is accomplished by KEKB [6]. The beam-beam induced synchro-betatron resonance can be suppressed by the synchrotron radiation damping of both colliding beams. The situation is quite different in an EIC, since the ion beam does not have fast damping. Therefore, to achieve a more reliable prediction, it is desired to develop special codes and/or simulation methods, which exclude or largely reduce the artificial numerical noise in the beam-beam simulation.

### Quantitative Understanding of the Damping Decrement to the Beam-beam Performance

To reach the beam-beam parameter 0.1 for the electron rings of eRHIC and JLEIC, based on the experience at KEKB, it requires a radiation damping decrement of 1/4000, or a radiation damping time of 4000 turns, in the transverse plane. To achieve the same radiation damping decrement at the low electron beam energies, super-bends are being considered for the electron ring lattice design in eRHIC. The purpose of these complicated three-segment super-bends is to be able to radiate additional synchrotron radiation energy at low electron energies to increase the radiation damping rate.

Since the connection between the damping decrement and the achievable beam-beam parameter is empirical, we carried out beam-beam simulations to study the beam-beam performance with different radiation damping decrements with strong-strong beam-beam simulation codes [7]. Figure 3 shows the evolution of the horizontal beam size of the electron beam with different radiation damping times.

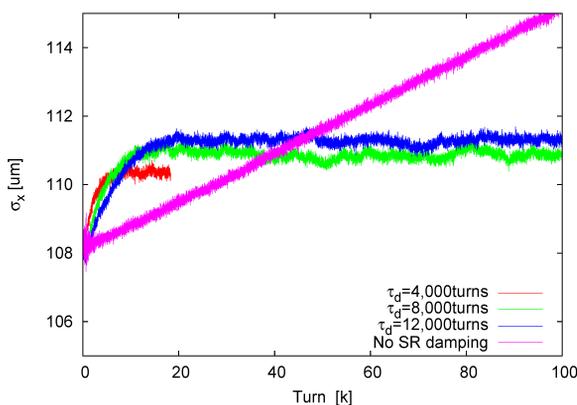


Figure 3: Equilibrium electron horizontal beam size as a function of the radiation damping time.

In these simulation studies, we did not observe coherent beam-beam motion with the different damping times as shown in Fig. 3. Simulation results show that with a longer damping time, it takes a longer time to reach the electron equilibrium beam size. However, there are not significant differences in equilibrium beam sizes and luminosities even when the radiation damping time is up to 12,000 turns, or 3 times the design value.

Lepton beams can tolerate beam-beam tune shift parameters  $\sim 0.1$  that are about ten times larger than corresponding values for collisions between hadron beams. The common understanding of these facts is the presence of radiation damping in lepton beams and the absence of damping in the hadron beam. It is of great importance for EIC running with low electron energies. Therefore, further investigations with dedicated simulation methodology and computer codes are required to study the effects of damping decrement to the beam-beam performance, and establish the connections between the damping decrement and the maximum beam-beam parameter at various collision energies for the current EIC ring-ring designs.

### Impacts on Protons with Electron Bunch Swap-out in eRHIC Ring-ring Design

In the current eRHIC ring-ring design, a rapid cycling synchrotron (RCS) is chosen as the baseline injector to the main electron storage ring. The RCS will be accommodated in the existing RHIC tunnel. It will be capable of accelerating the electron beam from a few hundred MeV up to 18 GeV and maintaining the electron polarization during acceleration.

The required electron bunch intensity of up to 50 nC in the eRHIC electron storage ring exceeds the capabilities of the electron gun, and such a high bunch intensity would also lead to instabilities at an injection energy in the RCS. These limitations necessitate accumulation of electrons in the electron storage ring.

To minimize detector background during the injection process, an accumulation in the longitudinal phase space is being proposed. After one electron bunch in the electron storage ring is kicked off, it will be replaced with 5 electron bunches from the RCS. The bunch intensity from the RCS is about 10 nC. The time interval between the injected RCS bunches is 1 second, or 7800 turns. To maintain high electron polarization in the electron storage ring, we will replace one electron bunch in 1 second and replace all electron bunches in 5 minutes.

With zero dispersion throughout the detector and the upstream beamline, the newly injected bunches travel on the same closed orbit in the region as the stored beam. However, the beam-beam effect of the injected electron bunches from the RCS on the stored proton bunch needs to be studied. The beam-beam parameter for the corresponding proton bunch changes during the electron bunch replacement.

A weak-strong study simulation code was developed to study the proton bunch emittance blow-up during the electron bunch replacement [8]. In the code, the proton bunch

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

is represented by macro-particles and the electron bunches are represented by rigid charge distribution. The 4-d beam-beam kick is used. The effect of radiation damping is simply included by adjusting the position and the energy deviation of the rigid electron bunches.

Figure 4 shows the calculated horizontal and vertical emittance evolution over the course of 100 electron bunch replacements from the above weak-strong code. Since each bunch is replaced every 5 minutes, the time for 100 bunch replacements is about 9 hours. From the plot, the emittance growth from the beam-beam interaction during the electron bunch replacement is less than 4%/hour.

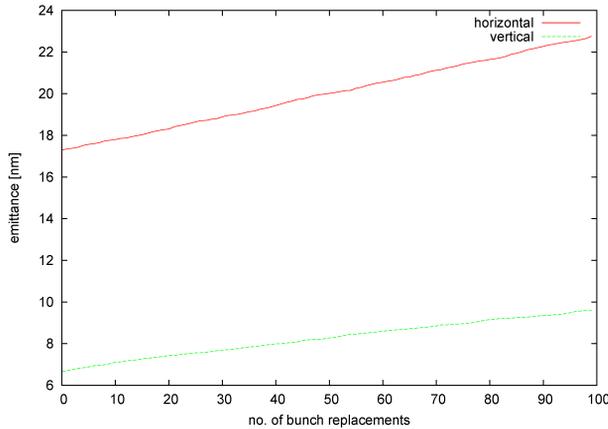


Figure 4: The simulated emittance evolution of the proton bunch during 100 electron bunch replacement.

The above 4-d weak-strong simulation to study the electron bunch replacement in the eRHIC ring-ring design is not self-consistent. The injected electron bunch may not have a 4-d Gaussian charge distribution. During the period of the electron bunch passing through the proton bunch, its beam size can be altered by the beam-beam force too. And the electron bunch does not always collide with the proton bunch at IP. A self-consistent 6-d strong-strong beam-beam simulation code is needed to study the beam-beam effects during the electron bunch replacement.

## PROPOSED RESEARCH AND METHODS

Both strong-strong and weak-strong beam-beam simulation codes are to be used to address the above simulation challenges in the eRHIC beam-beam studies. We choose Dr. Qiang's code BeamBeam3D [9] for the strong-strong beam-beam simulations, and Dr. Luo's code SimTrack [10] for the weak-strong beam-beam simulations. To meet the needs for the required EIC beam-beam simulations, we will make several modifications to these existing codes.

### Beam Dynamics Study and Numerical Simulation of Crabbed Collision With Crab Cavities

In the most of existing beam-beam strong-strong beam-beam simulation codes, the particle-in-cell and Green's function methods are used to solve the 2-dimensional Poisson

equation to obtain the electro-magnetic fields from one slice of one bunch. To reduce the numerical noises in the strong-strong beam-beam simulations, we propose to use a spectral method that uses a finite number of global basis functions to approximate the charge density distribution. Such a spectral method helps smooth the numerical noise associated with a finite small number of macro-particles (in comparison to the real number of particles in a bunch) and mitigate the numerical noise driven emittance growth.

Figure 5 compares the emittance growth evolution by using the standard Green's function method and the spectral method in a single slice beam-beam force model [11]. It is seen that the spectral method yields much less numerical noise driven emittance growth than Green's function method. This example shown here is the nominal LHC parameters without crossing angle and with a single interaction point. For those parameters, it is expected that there should be little emittance growth under the stable operational condition. In our plan, we would like to extend the above spectral method to multi-slice beam-beam interaction model. We also plan to implement this method on the massive parallel computers using a hybrid parallel programming model.

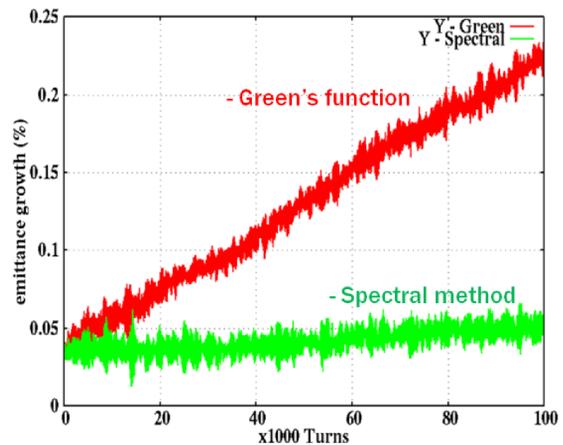


Figure 5: Comparison of calculated emittance growth with Green's function and spectral method.

With the new code development, we will try to find the scaling behavior of the luminosity degradation due to the synchro-betatron resonance as function of the beam-beam parameters of both beams, as well as the crab cavity frequency and the crossing angle. we also will evaluate if the non-zero dispersion function or the non- $\pi/2$  phase advance at the location of crab cavities will lead to beam quality degradation. The effects of the noises in the voltage and phase of crab cavities will be evaluated too.

### Quantitative Understanding of the Damping Decrement to the Beam-beam Performance

To fully understand the effects of synchrotron damping time on the beam-beam performance, the lattice non-linearity should be included into the strong-strong beam-beam simulation. The equilibrium emittances are decided

by the ratio of the radiation damping and the nonlinear lattice caused diffusion. Both the beam-beam and the lattice nonlinearities generate diffusion. The beam-beam force decreases like  $1/r$  while the nonlinear magnetic force increases like polynomials with the particle amplitude. The simulation shows that without the lattice nonlinearities, the diffusion solely due to beam-beam interaction is weak.

For most of the existing strong-strong beam-beam simulation codes, the ring is simply represented by a  $6 \times 6$  linear matrix to save the computing time involved in the beam-beam interaction calculation. However, from single particle element-by-element weak-strong beam-beam simulations, we learned that the interplay between the beam-beam interaction and the lattice non-linearities plays a crucial role to the dynamic aperture.

To include the lattice non-linearities without time-consuming element-by-element particle tracking in the strong-strong beam-beam simulation code, we propose the following methods: 1) Replace the linear ring map by a nonlinear map to up to a certain order. The nonlinear map should be symplectic in order to avoid artificial diffusion. To shorten the tracking time, a low order map, for example up to the 5th order, is planned during the first test. A higher order map can be implemented later as necessary. 2) Implement high order nonlinear field errors in the interaction region. According to the RHIC experiences, these high order field errors play an important role in the dynamic aperture reduction. 3) In the longitudinal plane, we plan to use the real RF cavities instead of linear synchrotron oscillation.

With the proposed lattice nonlinear in the strong-strong beam simulation codes, we will be able to study the effects of the damping decrement to the beam-beam performance. We will establish the connection between the damping decrement and the maximum beam-beam parameter for eRHIC.

### *Impacts on Protons with Electron bunch Swap-out in eRHIC Ring-ring Design*

Instead of early 4-d weak-strong simulation method, we propose to use the self-consistent 6-d strong-strong beam-beam code BeamBeam3D to simulate the electron bunch replacement in the eRHIC ring-ring design. To be suitable for this study, some modifications to BeamBeam3D are needed.

In the eRHIC ring-ring design, one electron bunch in the storage ring will be knocked out every 5 mins. 5 RCS bunches with a smaller bunch charge will be injected in the same bucket in the longitudinal phase space. The interval between these 5 injections is 1 second. In the code, we will first simulate the interaction between an electron bunch and a proton bunch up to several electron damping times to reach the equilibrium. After the 5 RCS bunches are injected, we also need to continue to track the beam-beam interaction between the newly injected electron bunch and the proton bunches up to a few damping times.

With these code modifications, we will evaluate the emittance growth during the electron bunch replacement. We will record the proton bunch emittance's change during the

electron bunch kick-off, each RCS bunch injection, and the final equilibrium. The emittance blow-up will be compared with the analytical estimate based upon a linear beam-beam force assumption and that from the previous weak-strong beam-beam simulation. We also will study the effects of any errors or noises during the electron bunch replacement, for example, the injection jitters, the fluctuation in the RCS bunch intensities, and so on.

## SUMMARY

In this article, we have presented the high priority R&D items related to the beam-beam interaction for the current eRHIC design. To mitigate the technical risks associated with the EIC accelerator designs, we joined beam-beam simulation experts from 3 laboratories and 1 university. We outlined the new beam-beam simulation algorithms and methods to the existing strong-strong beam-beam simulation codes. At the completion of this proposal, we should have a clear understanding of the beam-beam interaction in the next generation EIC designs and be able to provide robust counter-measures to possible beam-beam interaction related beam lifetime reduction, beam emittance growth, beam instabilities, and luminosity degradation.

## REFERENCES

- [1] *The 2015 Long Range Plan for Nuclear Science*, The 2015 Nuclear Science Advisory Committee, 2015.
- [2] V. Ptitsyn, "Progress in eRHIC design", presented in the EIC collaboration meeting, Oct. 1-3, 2017, Brookhaven National Laboratory, NY, USA.
- [3] Y. Zhang, "Progress in JLEIC design", presented in the EIC collaboration meeting, Oct. 1-3, 2017, Brookhaven National Laboratory, NY, USA.
- [4] *Report of Community Review of EIC Accelerator R&D for the Office of Nuclear Physics*, published on Feb. 13, 2017.
- [5] Y. Hao, "Dynamics with crab crossing for EIC", presented in the EIC collaboration meeting, Oct. 1-3, 2017, Brookhaven National Laboratory, NY, USA.
- [6] K. Ohmi, M. Tawada, Y. Cai, S. Kamada, K. Oide, and J. Qiang, "Luminosity limit due to beam-beam interaction with or without crossing angle", *PRST-AB*, vol. 7, pp. 104401, 2004. doi:10.1103/PhysRevSTAB.7.104401
- [7] Y. Luo, "Additional beam-beam challenges", presented in the EIC collaboration meeting, Oct. 1-3, 2017, Brookhaven National Laboratory, NY, USA.
- [8] C. Montag, unpublished, 2017.
- [9] J. Qiang, M. Furman, and R. Ryne, "Strong-strong beam-beam simulation using a Green function approach", *PRST-AB*, vol. 5, pp. 104402, 2002. doi:10.1103/PhysRevSTAB.5.104402
- [10] Y. Luo, "SimTrack: A compact c++ code for particle orbit and spin tracking in accelerators", *Nuclear Instruments and Methods in Physics Research Section, A*, vol. 801, pp. 95-103, 2015. doi:10.1016/j.nima.2015.08.014
- [11] J. Qiang, unpublished, 2018.

# BEAM DYNAMICS SIMULATIONS OF MEDICAL CYCLOTRONS AND BEAM TRANSFER LINES AT IBA

J. Van de Walle<sup>†</sup>, E. Forton, V. Nuttens, W. Kleeven, J. Mandrillon, E. Van der Kraaij,  
Ion Beam Applications, 1347 Louvain-la-Neuve, Belgium

## Abstract

At the Belgian company Ion Beam Applications (IBA), several in-house developed computational tools are used to simulate beam dynamics from a range of proton and electron accelerators. The main beam dynamics simulation tool is the “Advanced Orbit Code” (AOC), which integrates the equations of motion in any 2D or 3D magnetic field map with superimposed time variable or fixed electric fields. CLORIBA is the in-house closed orbit code for cyclotrons, which provides the tunes and isochronicity conditions for the isochronous cyclotrons. A tool developed especially for the superconducting synchro-cyclotron (S2C2) is the phase space motion code, which tracks energy, RF phase and orbit centre coordinates in the synchro-cyclotron. The calculation tools are described briefly and some examples are given of their applicability on the IBA accelerators.

## INTRODUCTION

The computational tools at IBA are oriented towards three main categories of accelerators. The first types are isochronous cyclotrons, which operate at a fixed RF frequency. The three most common isochronous cyclotrons are the 18 MeV proton cyclotron called the Cyclone® KIUBE, used for production of radioactive F-18 [1], the 70 MeV proton cyclotron called the Cyclone® 70, mainly used for production of radiopharmaceuticals other than F-18 (for ex. Sr-82) and the 230 MeV proton cyclotron the Cyclone® 230. The latter is used in proton therapy systems. The second type of accelerator is the superconducting synchro cyclotron (the S2C2) [2], which is so far the only superconducting cyclotron at IBA and operates at a variable RF frequency (from 90 to 60 MHz). This accelerator delivers a pulsed (1 kHz) 230 MeV protons beam with pulse lengths of 10  $\mu$ s. The last type of accelerator which will be covered is the rhodotron electron accelerator, which is a special arrangement of magnets around an accelerating cavity. This accelerator typically delivers a 10 MeV electron beam [3].

## CALCULATION TOOLS

The main beam tracking code used at IBA is the “Advanced Orbit Code” (AOC) [4]. This code was originally developed by W. Kleeven and solves the equations of motion for a range of particles (protons, electrons, etc...) relevant to IBA accelerators. The independent integration variable is the RF phase advance ( $\tau$ ), which is related to time in the following way:

$$\tau = \omega_0 \int_0^t f(t') dt',$$

where  $\omega_0$  is the angular RF frequency and  $f(t')$  is an arbitrary function of time. For isochronous cyclotrons  $f(t')=1$ . The differential equations are solved with a 4<sup>th</sup> order Runge Kutta integrator with variable step size. As input AOC can handle 2 or 3D static magnetic field maps, 3D potential maps on which a RF frequency is applied and static electric potential maps. In case of a 2D static magnetic field map, the magnetic field can be expanded around the median plane up to 3<sup>rd</sup> order. AOC is mainly used in studies related to extracted beam emittances, resonance studies and to optimize magnetic designs.

At IBA an in-house closed orbit code called CLORIBA was developed. This code is available in both C++ and python (pyCLORIBA). The code uses the established algorithm developed by Gordon [5] to determine the tunes and phase slip.

A last computational code is called “phase space motion” and was developed especially for the S2C2. This code tracks the energy, RF phase, orbit centre coordinates and the vertical beam motion of protons in a synchro-cyclotron. It uses a 4<sup>th</sup> order Runge-Kutta integrator with adaptive step size and takes as input the harmonic components of the magnetic field map, the frequency sweeps as a function of time and the voltage profile as a function of time.

The utilization of these three calculation tools will be illustrated in the following paragraphs.

## ISOCHRONOUS CYCLOTRONS

### Closed Orbit Program

The first step after the mechanical and electrical assembly of an isochronous cyclotron is a magnetic mapping to ensure the isochronicity of the magnetic field, so that a constant RF frequency can be applied and the beam is accelerated up to full energy without beam losses. These measured 2D maps serve as input to CLORIBA. A python script was developed which calculates the tune curves, the phase slip and the needed magnetic shimming which needs to be performed to make the cyclotron isochronous. An example 2D magnetic map from the Cyclone® KIUBE with the closed orbits on top of it is shown in Figure 1. The yellow regions in the centre of the poles are the regions where the “pole insertions” are placed. These are removable pieces of iron which can easily be machined to obtain isochronicity of the machine. The amount of machining of these pole pieces is directly calculated in pyCLORIBA based on the closed orbit analysis and magnetic perturbation maps calculated with the OPERA3D software.

**Example: Cyclone® KIUBE**

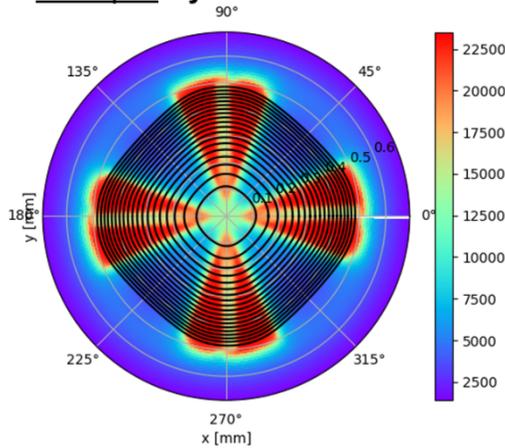


Figure 1: (Upper) measured 2D field map of the Cyclone KIUBE with closed orbits shown in black. (lower) top view on lower half of the Cyclone® KIUBE.

*Extracted Emittance Calculation*

AOC is often used to calculate the extracted beam emittance from the cyclotrons. As an example, a typical layout of the Cyclone® 70 cyclotron with associated beam lines is shown in Figure 2.

In this figure, two closed orbits at 60 and 70 MeV are plotted together with the reference extracted beam path calculated with AOC. Since this cyclotron accelerates H- ions, the extraction happens by stripping. The stripper position is optimized in AOC to have the extracted beam directed towards the switching magnet which is shown in the bottom part of the figure. From the front of the switching magnet, envelope calculations of the horizontal and vertical beam are performed with the TRANSPORT code [6]. Crucial for these envelope calculations are the input emittances at this position. Figure 3 shows the extracted beam emittance from the Cyclone® 70 at the level of the switching magnet as calculated with AOC. The calculated emittance is compared with measured emittances. The measurement was performed in one of the beam transfer lines with the variable quadrupole method. The comparison is fair for the vertical emittance, whereas the measured horizontal emittance is much larger and can probably be attributed to a degraded stripper foil at the moment of the measurement, which is influencing to a large extent the horizontal quality of the extracted beam.

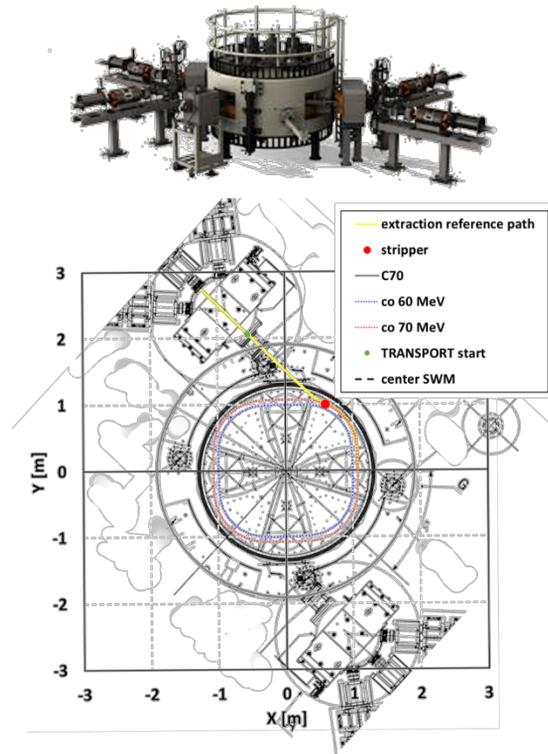


Figure 2: (Upper) 3D view on the Cyclone® 70 cyclotron with 1 switching magnet and 3 transfer beamlines at each side. (lower) top view on the cyclotron with in blue and red the closed orbits at 60 and 70 MeV. The red dot indicates the stripper position, the yellow line is the simulated extraction path and the small green dot in front of the switching magnet is where the optics calculations with TRANSPORT start.

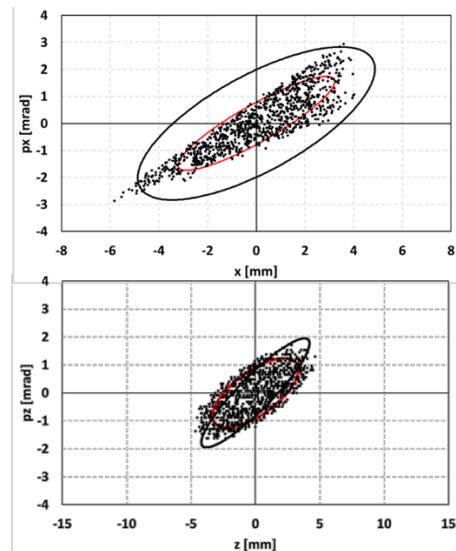


Figure 3: Horizontal (top) and vertical (bottom) emittances out of the Cyclone® 70. The red ellipse is fitted to the calculated distributions and the black ellipse is the measured emittance.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## THE SYNCHRO-CYCLOTRON (S2C2)

### Main Characteristics

The S2C2 [2] (shown in Figure 4) is a superconducting synchro-cyclotron with a central field of 5.7 T, accelerating protons up to 230 MeV. The RF frequency sweeps from around 90 MHz at injection to 60 MHz at extraction over a period of 500  $\mu$ s. During another 500  $\mu$ s the RF frequency sweeps the other way around, leading to a repetition rate of 1 kHz. The beam pulses have a temporal length of about 10  $\mu$ s. In order to ensure longitudinal phase stability, a small acceleration (dee) voltage is applied of maximum 10 kV. The intensity of the beam pulses is regulated by changing the dee voltage from 7 to 10 kV.



Figure 4: Side view of the S2C2

### Simulation Challenges

Because of the small energy gain per turn in a synchro cyclotron, the number of turns from injection to extraction is about 40000. To simulate the beam dynamics from the injection to extraction, a full detailed tracking in AOC would be very time consuming (e.g. a full tracking of 1 proton takes about 20 minutes). Therefore the simulation was split into three parts. The injection part, where losses in the central region and the phase acceptance in the RF bucket are determined is simulated in detail with AOC, typically up to 5 MeV. The acceleration part, from 5 MeV up to close to the extraction energy (typically 225 MeV) is performed with a code called 'phase space motion'. This code primarily tracks the energy, RF phase and the orbit centre coordinates. From these parameters, the RF bucket and the emittance can be evaluated at high energy (225 MeV). In the tracking the following equations are integrated:

$$\begin{aligned} \frac{dE}{dt} &= eF_{RF}V_{RF} \sin(\varphi), \\ \frac{d\varphi}{dt} &= 2\pi(F_{RF} - F_p), \\ \frac{dx_c}{dt}, \frac{dy_c}{dt} & \end{aligned}$$

where  $F_{RF}$  and  $V_{RF}$  are the RF frequency and voltage,  $\varphi$  is the RF phase,  $F_p$  is the particle revolution frequency (depending on its energy and the average magnetic field) and  $x_c, y_c$  are the orbit centre coordinates. These equations are

integrated with time steps of 0.1  $\mu$ s, which makes the integration much faster compared to a detailed tracking with AOC (about 5000 steps instead of several million).

Finally, the extraction process is simulated again in detail with AOC, where the input parameters for the beam are taken from the output beam parameters in the phase space motion code.

In the following paragraph, the phase space motion code is described a bit more in detail.

### The Orbit Centre Coordinates

Figure 5 shows the orbit centre coordinates of the closed orbits in the S2C2. The orbit centres of the accelerated protons will oscillate around these orbit centres.

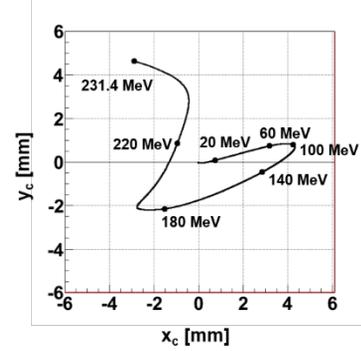


Figure 5: Closed orbit centre evolution in the S2C2.

The equations of motion for the accelerated proton orbit centre coordinates are derived from the following Hamiltonian (see [7]):

$$\begin{aligned} H(x_c, y_c) &= \frac{1}{2}(v_r - 1)(x_c^2 + y_c^2) + \frac{r}{2}(A_1x_c + B_1y_c) \\ &+ (D_3x_c + D_4y_c)(x_c^2 + y_c^2) \\ &+ \frac{1}{4}\left(A_2 + \frac{1}{2}A_2'\right)(x_c^2 - y_c^2) \\ &+ \frac{1}{2}\left(B_2 + \frac{1}{2}B_2'\right)x_cy_c \\ &+ \frac{1}{48r}\left(D_1(4x_c^3 - 3x_c(x_c^2 + y_c^2))\right. \\ &\left.+ D_2(3y_c(x_c^2 + y_c^2) - 4y_c^3)\right) + \vartheta, \end{aligned}$$

where the factor  $A, B, C$  and  $D$  are related to the harmonics of the magnetic field. The sub-indices are related to the harmonic number and the ' indicates a radial derivative. Details on these parameters can be found in [7]. A more intuitive picture of this complex Hamiltonian is illustrated in Figure 6, where the Hamiltonian is plotted as a potential energy surface. As can be seen at 220 MeV, far below the extraction energy of 230 MeV the orbit centre coordinates are trapped in a potential well and the orbits are stable, whereas closer to the extraction energy (at 229.6 MeV), orbit centre coordinates can become unstable if they are far from the global minimum. This global minimum coincides with the orbit centre of the closed orbit, which is shown with the black dot. The small difference between the black dot and the global minimum in the coloured surface is due

to the fact that the closed orbit calculation takes into account all harmonics, whereas the Hamiltonian only up to the 3<sup>rd</sup> harmonic was taken into account. The main conclusion from these plots is that when the accelerated orbits are too much off-centred (far from the closed orbit centre), the beam can be extracted at a lower energy.

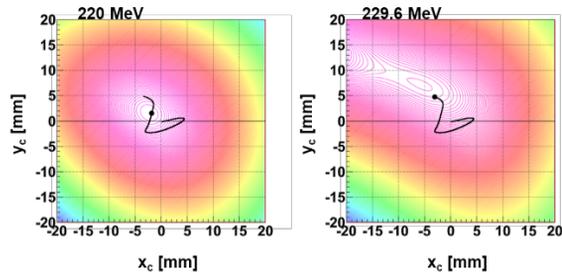


Figure 6: Illustration of the Hamiltonian describing the orbit centre motion in terms of a potential energy surface for 2 energies: 220 MeV (far from extraction: stable orbit centre) and close to extraction at 229.6 MeV (orbit centres can become unstable for large off-centrings with respect to the closed orbit)

### Experimental Observations

Some experimental observations were made when the S2C2 was attached to the proton gantry, which is a rotatable beam line taking the proton beam from the accelerator up to the patient (see [8] for details). The gantry is a large spectrometer, where the dispersion is maximized near the end of the beam line and an energy selection can be made with a momentum slit. The proton gantry is an extremely sensitive equipment which shows small fluctuations in beam energy and intensity.

A first experimental observation which was made on the prototype S2C2 is related to small energy changes of the proton beam when the dee voltage is changed. Figure 7 shows the beam optics in the proton gantry at 230 MeV from the exit of the S2C2 up to isocenter. The green dashed lines are quadrupoles, the red shaded areas are dipoles and the blue lines are slits. The proton trajectories are color coded according to their energy. As can be seen, the dispersion is maximized on the last slit. At that place in the gantry we have a dispersion of about 140 keV/mm.

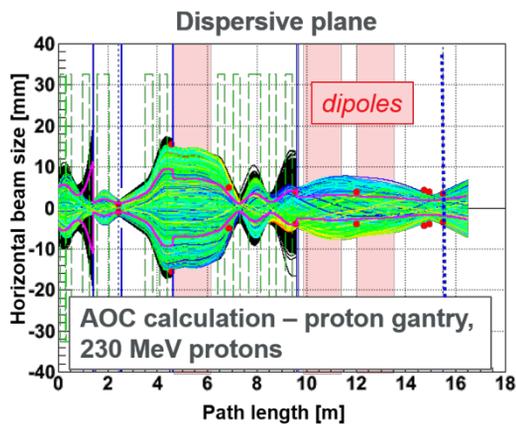


Figure 7: Proton trajectories in the dispersive plane of the proton gantry. See text for details.

The measured beam profiles at the level of the momentum slit for two different dee voltage settings is shown in Figure 8.

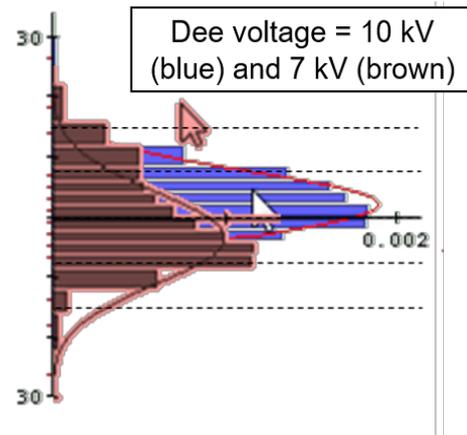


Figure 8: Beam profiles on the momentum slit in the dispersive plane. An energy shift is present for different dee voltages.

It is clearly seen that a small energy shift is present of about 200 keV. This effect could be explained by an inaccurate centering of the proton source at the centre of the S2C2. From simulations with AOC and phase space motion, it was seen that for different dee voltages, the orbit centering at injection depends on the dee voltage when the source is not precisely positioned. Since it is clear from the right part of Figure 6, different off-centrings for different dee voltages can lead to a different extracted beam energy (since the orbit centres become unstable at a different moment of the acceleration) When the source is accurately positioned, this effect is largely suppressed.

A second observation is shown in Figure 9. The upper figure shows the frequency sweep near the extraction frequency as a function of time. The lower figure shows the “normal” situation, where the beam pulse is extracted at the extraction frequency. The middle figure shows extracted beam pulses which are observed before and after this extraction frequency. These observations were made with a highly sensitive diamond detector.

The origin of these beam pulses was found from the phase space motion code. Figure 10 shows the origin of the problem. It concerns protons which are lost from stable RF bucket very close to the extraction frequency. This can be seen in the middle figure, where a proton is lost at 228 MeV. The phase space motion code can track this proton for another RF period. Fast energy oscillations are observed and at times when the proton revolution frequency coincides with the RF frequency, a short acceleration (or deceleration) can be observed before the fast energy oscillations set in again. Another resonance effect is seen in the orbit centre coordinates at a RF frequency  $f_s = f_p \pm (\nu_r - 1)f_p$  where  $f_p$  is the protons revolution frequency and  $\nu_r$  is the horizontal tune. This resonance condition can be understood from the structure of the equation of motion of the orbit centre coordinate:

$$\frac{dx_c}{d\theta} = (\nu_r - 1)y_c + (\dots) + \beta_2 x_c + (\dots),$$

where the term  $\beta_2 x_c$  is responsible for this resonance:  $\beta_2$  oscillates with the same frequency as the energy:  $(f_s - f_p)$  and  $x_c$  oscillates with the frequency  $(\nu_r - 1)f_p$ .

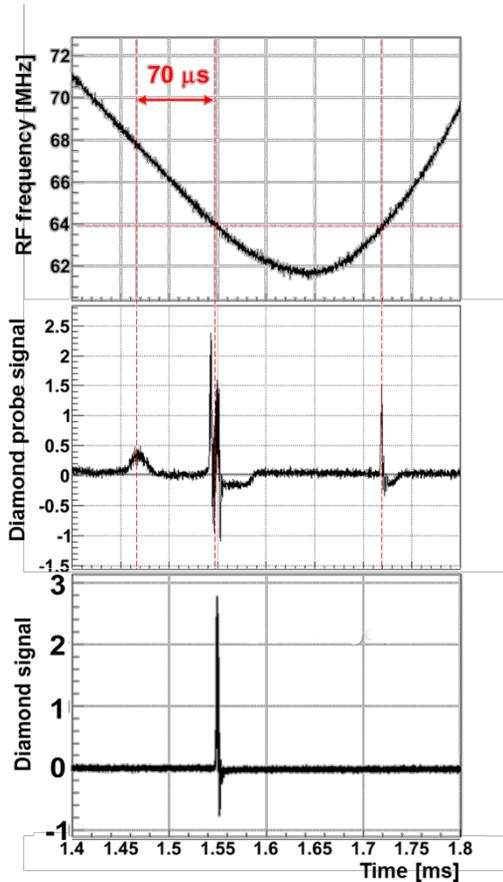


Figure 9: (Upper) detail of the RF frequency sweep near the extraction time (middle) observed extracted beam pulses in the cases beam losses are present inside the S2C2 close to the extraction energy. (lower) extracted beam pulse when the beam is accelerated up to the end without losses near the end.

From the bottom part of Figure 10 it can be seen that when the resonance in the orbit centre coordinates is large and the proton is close to the extraction energy, the orbit off centering can become so large that the orbit centre coordinate enters into the unstable region and the proton will leave the S2C2.

### THE RHODOTRON (TT50)

Figure 11 shows a simulated layout of the TT50 rhodotron. This electron accelerator brings electrons up to 10 MeV. The electrons are accelerated by a central cavity and are bend back towards the cavity after each passage by magnets which are located in the circumference of this cavity. There are 9 magnets in total with field level from 1000 to 5000 Gauss. Since the most powerful magnet is located

right next to the electron gun, where the electron has a low energy, any fringe field from this magnet can influence the trajectory of the electron beam and potentially induce important losses. Since the TT50 rhodotron deals with beam currents up to 10 mA, beam losses have to be avoided. At the same time, the TT50 magnets were made from permanent magnets, which are fixed in strength. Other rhodotrons use electromagnets, where beam tuning can more easily be done. This contributed to the uncertainty on the beam dynamics and stability during the design phase of the TT50. Therefore, the AOC code was updated to incorporate full 3D fields (from the permanent magnet bend dipoles) which can be positioned in 3D space as needed. As such a full model was created of the TT50 and all magnetic effects and possible perturbations (in magnet position, strength, tilt, etc...) could be simulated. Full details on this work can be found in [3].

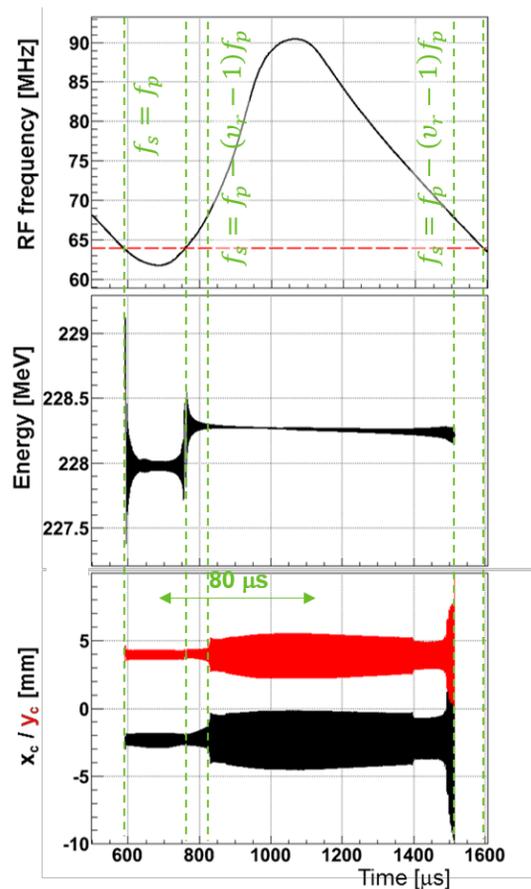


Figure 10: (Upper) 1 period of the frequency sweep (1 ms) with the “synchronous” extraction frequency ( $f_s$ ) indicated. (middle) fast energy oscillations of a proton which is lost (desynchronized) close to the extraction frequency. (lower) orbit centre coordinates of the desynchronized proton. Resonances are observed at specific frequencies, coinciding with the observed beam pulses.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

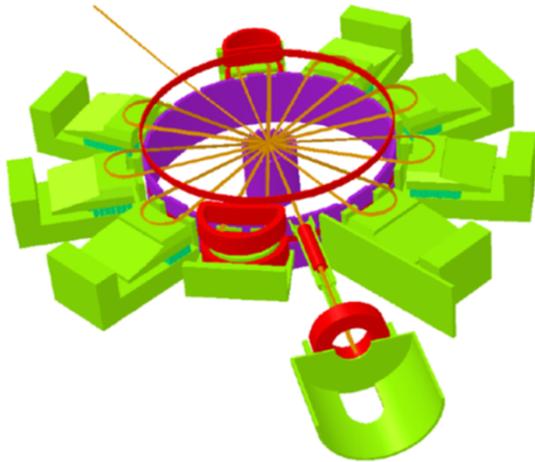


Figure 11: View on the simulated TT50 rhodotron: 9 magnets surrounding a central cavity, the electron gun and several correction coils. The orange line is the simulated electron beam in AOC.

### BEAM LINE SIMULATIONS WITH AOC

Thanks to the developments on the AOC code related to the TT50, it has become possible to construct full beamline models in AOC and to track particles in a sequence of full 3D maps of quadrupoles, dipoles, steering magnets, etc... As such, the proton gantry was simulated in AOC. Figure 7 shows the proton trajectories at 230 MeV from the S2C2 exit up to isocenter. A benchmark simulation for the AOC model of the proton gantry consists in the scanned beam pattern at isocenter. In this case, the proton beam is scanned by two perpendicular scanning magnets in between the last two dipoles. As such, a pattern is built at isocenter over an area of about  $15 \times 20 \text{ cm}^2$ . Due to imperfections in the dipole field of the last bend magnet (which has a large aperture), the scanning pattern at isocenter is slightly distorted. This distortion is not clearly visible in the pattern shown in Figure 12, but careful analysis shows deviations from the required position of up to 2 mm. This effect is known as the “pillow effect” and can be compensated by small corrections on the scanning magnets. The simulation of this pillow effect is a stringent test for the accuracy of the tracking code and the 3D magnetic field map of the last dipole.

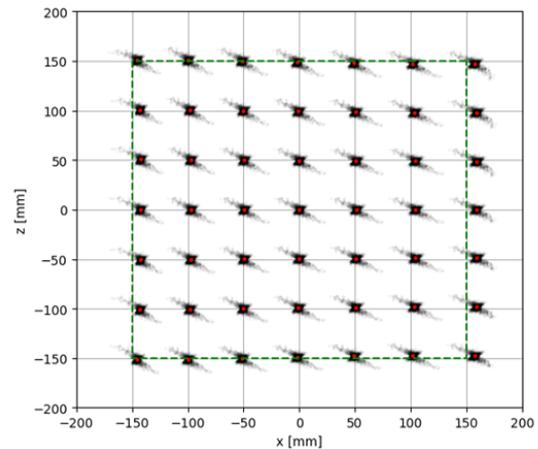


Figure 12: The “pillow” effect of the scanned beam at the isocenter, simulated with AOC.

### REFERENCES

- [1] S. Zarembo *et al.*, in *Proc. of Cyclotrons'16*, Zurich, Switzerland, Sept. 2016.  
doi:10.18429/JACoW-Cyclotrons2016-TUP03
- [2] J. Van de Walle *et al.*, “A Diamond Detector Test Bench to Assess the S2C2 Beam Characteristics”, in *Proc. of Cyclotrons'16*, Zurich, Switzerland, Sept. 2016.  
doi:10.18429/JACoW-Cyclotrons2016-MOP19
- [3] W. Kleeven, in *Proc. of IPAC'18*, Vancouver, BC, Canada, Apr. 2018.  
doi:10.18429/JACoW-IPAC2018-MOPML020
- [4] W. Kleeven *et al.* “Extraction System Design for the New IBA Cyclotron for PET Radioisotope Production”, in *Proc. of IPAC'16*, Busan, Korea, May 2016.  
doi:10.18429/JACoW-IPAC2016-TUPOY002
- [5] M. M. Gordon, *Part. Acc.* 16 (1984), p. 39-62
- [6] U. Rohrer *et al.*,  
[http://aea.web.psi.ch/Urs\\_Rohrer/MyWeb/trans.htm](http://aea.web.psi.ch/Urs_Rohrer/MyWeb/trans.htm)
- [7] H. L. Hagedoorn and N. F. Verster, “Orbits in an AVF cyclotron”, *Nucl. Instr. Meth.*, vol. 18 -19, pp. 201-228, 1962.
- [8] E. Pearson *et al.*, “Magnet Developments and Commissioning for the IBA Compact Gantry”, *IEEE Trans. on Applied Superconductivity*, Vol. 24, 2014.

# DESIGN STUDY OF A FAST KICKER MAGNET APPLIED TO THE BEAMLINE OF A PROTON THERAPY FACILITY

Wenjie Han<sup>†</sup>, Xu Liu, Bin Qin

Institute of Applied Electromagnetic Engineering  
 Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

## Abstract

A proton therapy facility based on an isochronous superconducting cyclotron is under development in HUST (Huazhong University of Science and Technology). A fast kicker magnet will be installed in the upstream of the degrader to perform the beam switch function by kicking the proton beam to the downstream beam stop. The rising and falling time of the kicker is about 100  $\mu$ s, and the maximum repetition rate is 500 Hz. This paper introduces simulation and optimization of the eddy current and dynamic magnetic field of the fast kicker, by using FEM code OPERA-3D. For kicker materials, laminated steel and soft ferrite are compared and the MnZn ferrite is chosen. Designing considerations includes the eddy current effect, field hysteresis, and mechanical structure of the kicker will also be introduced.

## INTRODUCTION

HUST proton therapy facility (HUST-PTF) is based on an isochronous superconducting cyclotron and spot scanning technique. Two 360-degree rotation gantry treatment rooms and one fixed beamline treatment station will be constructed at first stage. A detail description of the facility parameters can be found in Ref. [1]. During the treatment process of the pencil beam spot scanning, the proton beam is applied to the patient for only a few milliseconds, and then kicked away. After repositioning and/or readjustment of the beam energy, the beam is directed back to the patient [2]. A fast kicker magnet will be installed in the upstream of the degrader, to perform the ‘beam off’ function by kicking the proton beam to the downstream beam stop.

This paper mainly compares two material schemes for the kicker magnet yoke and analyses the eddy current and field hysteresis effect of kicker magnets. The design of mechanical structure is also introduced.

## PHYSICAL SPECIFICATIONS

The kicker magnet system is one of the key components of spot scanning technique. The layout of the kicker system is shown in Fig 1. In HUST-PTF, vertical kicker scheme is adopted. The main parameters of the kicker magnet are listed in Table 1. The kicker magnet is located at 1.24 m before the degrader. There is a quadrupole (Q3) between the kicker magnet and the degrader, whose defocusing direction is the same to the kicker deflected direct

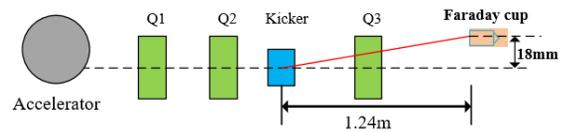


Figure 1: Layout of kicker system.

ion. The proton beam is deflected by the kicker magnet, then passing through a drift, the quadrupole between the kicker magnet and the degrader will further bend the proton beam to a Faraday cup (FC). According to the simulation of beam trajectory, the gap and pole width of kicker magnet is determined, the distance from beam stop to the center of the kicker magnet is about 1.24 m. The minimum integral field is 0.0252 T·m, deflecting the 250 MeV proton beam an angle about 10.36 mrad. The beam offset at the FC is about 18 mm (7 mm for the beam size, 3 mm for the thickness of FC, 8 mm for the radius of FC). As for the power supply, the kicker magnet is excited by pulse current with a maximum repetition frequency of 500 Hz and a rising/falling time of 100  $\mu$ s. The current ramping speed is up to 5040 kA/s and the magnetic field ramping speed is up to 1010 T/s.

Table 1: Parameters of the Kicker Magnet

Name	Parameter
Deflection angle	10.36 mrad
Magnet gap	50 mm
Integral field	0.0252 T·m
Magnet length	200 mm
Number of coil turn	4 Turns/pole
Field strength	0.101 T
Good field region	$\pm 30$ mm (vertical) $\pm 14$ mm (horizontal)
Coil Induction	44 $\mu$ H
Max repetition Frequency	500 Hz
Rise/fall time	100 $\mu$ s

## MAGNET DESIGN

Kicker magnet applied to HUST-PTF is a window frame type magnet with two bedstead coils. To insure the required rapid change of the magnetic field, eddy currents in the core must be evaluated. Soft ferrite or laminated silicon sheets can be chosen as the material of the magnet cores.

\* Work supported by The National Key Research and Development Program of China, with grant No.2016YFC0105305, and Natural Science Foundation of China with grant No. 11375068

<sup>†</sup> email address: hanwenjie@hust.edu.cn

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

### Simulation and Parameter Setting

To study the eddy current effect in kicker magnet, the 3D transient electromagnetic simulation and steady-state thermal simulation are performed in the Vector Fields Opera 18 simulation code [3]. The procedure of electromagnetic and thermal analysis is shown in Fig 2.

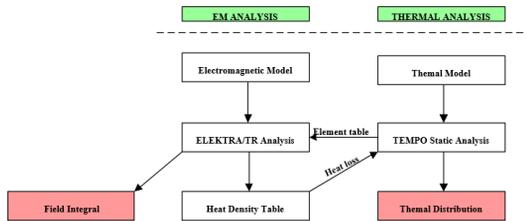


Figure 2: Procedure of electromagnetic and thermal analysis.

The excitation current of coils is a trapezoidal wave with a ramp/down time of 100  $\mu$ s and a steady time of 900  $\mu$ s, Electromagnet simulations with the adaptive time step are carried out on the ELEKTRA/TR program to obtain the field integral and eddy current loss in a period. The time setting is dense at the rise/down time to get a precise simulation result. The coordinates table of centroid in each element generated from the unsolved TEMPO model is then imported into the ELEKTRA/TR program to evaluate the heat loss [4]. A table file of average heat density value over a cycle is calculated by

$$HEAT = \frac{\sum_{i=1}^n heat_i * (t_i - t_{i-1})}{t_n}, \quad (1)$$

where *HEAT* is the average power of each element, *n* is the total number of simulation point, *t<sub>i</sub>* is the simulation setting time of point *i*.

Then the average power of each element is import to the unsolved TEMPO Static Analysis and carried out, the temperature distribution of the kicker magnet will be presented.

In electromagnet simulations, the laminated steel is defined as anisotropic with a packing factor about 0.95, and there is no current along the laminated direction. The MnZn ferrite is regard as one block, it is isotropy with low conductivity. For TEMPO/SS, the magnet is assumed to be natural cooling with the transfer coefficient of 14 W/(K·m<sup>2</sup>) and the ambient temperature 20°C. The detail parameters of laminated steel are shown in Table 2, and the detail parameters of MnZn ferrite is shown in Table 3.

### Laminated Steel Yoke

Generally, laminated silicon steel sheets for iron core are used to reduce the eddy current and the heat loss. However, the laminated magnets are still not free from the eddy current, it will reduce the magnetic field rising speed. and the hysteresis effect is going to be large. Slits in the end laminations of iron core are proven to be an effective method to reduce the eddy current [5]. The laminated steel yoke with slits is considered for the magnet design. The field integral curves of current raising process are shown in Fig. 3. With in 100  $\mu$ s, the field integral can reach 82.7%. With slits in the end of limited laminated steel yoke, the field integral

can only reach 83.8%. The field integral curves of current falling process are shown Fig. 4. When the current is falling down to zero, the remanence is respect to be large. and it takes a long time to eliminate remanence. The temperature distributions are shown in Fig. 5. The maximum temperature decrease from 99.5°C to 71.6°C. This means slits in the end laminations of iron core can reduce eddy heat efficiently, but it does little for the response speed.

Table 2: Parameters of Laminated Steel

Name	Laminated Direction	Other Direction
Conductivity	3.5e+6 S/m	0
transfer coefficient	5.4 W/m/K	368 W/m/K

Table 3: Parameters of MnZn Ferrite

Name	MnZn
Conductivity	0.33 S/m
transfer coefficient	5 W/m/K

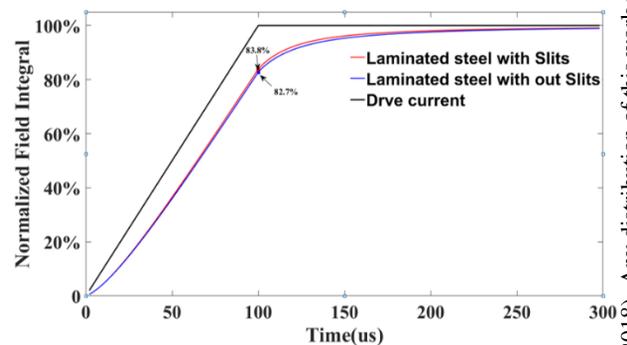


Figure 3: The field integral curves of current raising process.

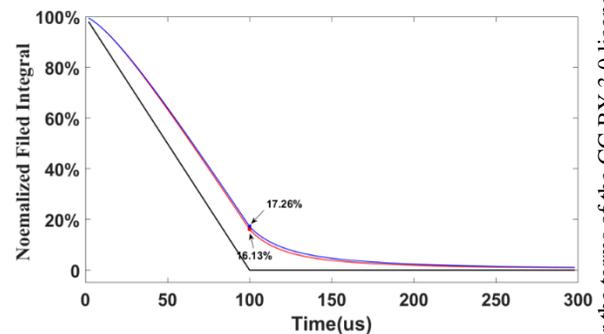
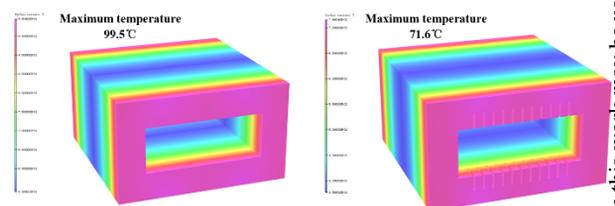


Figure 4: The field integral curves of current falling process.



(a) laminated steel without slits (b) laminated steel with slits

Figure 5: The temperature distribution: (a) laminated steel without slits; (b) laminated steel with slits.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

### MnZn Ferrite Yoke

Due to low electric conductivity, MnZn ferrite usually works in 1 kHz–1 MHz. In low frequency, it has little eddy current, and there's almost no temperature rise, but it is easy to saturate in the corners. The drive current and field integral curve are shown in Figs. 6 and 7. In the rising processing, the integral field can be up to 100% within 100  $\mu$ s. In the falling processing, the remanence is almost zero. As for the temperature rising, the temperature of the yoke is only the ambient temperature about 20°C. The magnetic field distribution is shown in Fig. 8, and there is little saturation region.

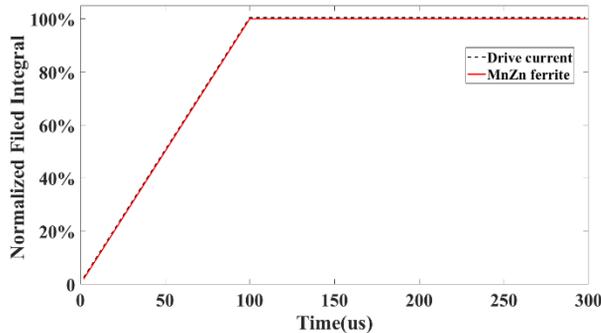


Figure 6: The MnZn ferrite field integral curves of current rising process.

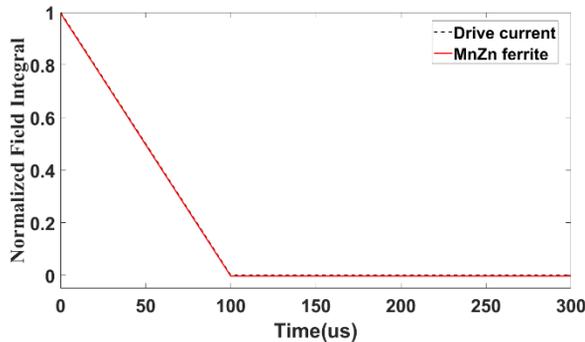


Figure 7: The MnZn ferrite field integral curves of current falling process.

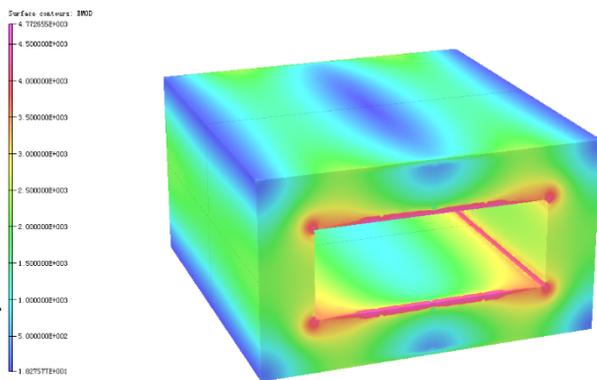


Figure 8: The field distribution of MnZn ferrite.

### Mechanical Structure

The mechanical performance of ferrite is hard and fragile. When the volume is too large, the homogeneity would become worse. In HUST-PTF, the kicker magnet is made out of six blocks attached to steel plates with runaway type ceramic vacuum chamber. The inner aperture is 15 mm for horizontal, 34 mm for vertical. The kicker magnet is fixed by four bolts. see in Fig. 9.

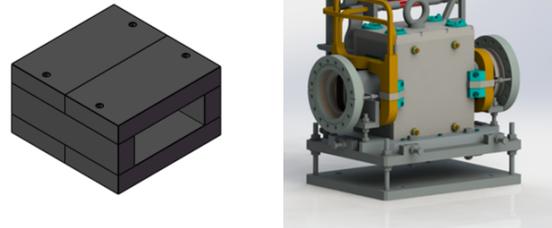


Figure 9: The ferrite yoke of the fast kicker magnet with runaway type ceramic vacuum chamber.

### CONCLUSION

This paper shows the layout of the kicker magnet in HUST-PTF and describes the design consideration of the magnet yoke. The simulation method of eddy current and temperature rise in fast ramp magnet is introduced. Two different materials are compared for the kicker magnet yoke: 1) laminated steel; 2) MnZn ferrite material. The eddy current in laminated steel is large. With slits in the end of laminated steel, the temperature can reach the requirement of the steel, the maximum temperature is 71°C, but the field hysteresis and remanence is expected to be large. Within 100us, the normalized integral field changes from 82.7% to 83.8%, and it takes a long time to eliminate the remanence. As for the MnZn ferrite, the field integral follows the change of the current, and the temperature of the yoke is only the ambient temperature about 20°C, there is no field hysteresis caused by eddy current, and the field saturation region is little. Finally, MnZn ferrite is chosen, and the yoke is made out of 6 blokes, fixed with bolts through embedded holes and the vacuum is runaway type ceramic vacuum chamber.

### REFERENCES

- [1] B. Qin *et al.*, “Design Study of a 250 MeV Superconducting Isochronous Cyclotron for Proton Therapy”, 2014.
- [2] R. Künzi *et al.*, “Kicker Converters for Fast Proton Beam Bending”, EPE-PEMC 2004, Riga, September 2004.
- [3] Opera-3D reference, <http://www.vectorfields.com>
- [4] M. Negrazus *et al.*, “Eddy Current Reduction in Fast Ramped Bending Magnets.” *IEEE Transactions on Applied Superconductivity*, vol. 16, no. 2, pp. 228-230, 2006.
- [5] X. Liu *et al.*, “Eddy current analysis and optimization of fast scanning magnet for a proton therapy system,” *Nuclear Instruments & Methods in Physics Research*, vol. 862, 2017.

# COMPUTATIONAL ACCELERATOR PHYSICS: ON THE ROAD TO EXASCALE

R. D. Ryne\*, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

## Abstract

The first conference in what would become the ICAP series was held in 1988. At that time the most powerful computer in the world was a Cray YMP with 8 processors and a peak performance of 2 gigaflops. Today the fastest computer in the world has more than 2 million cores and a theoretical peak performance of nearly 200 petaflops. Compared to 1988, performance has increased by a factor of 100 million, accompanied by huge advances in memory, networking, big data management and analytics. By the time of the next ICAP in 2021 we will be at the dawn of the Exascale era. In this talk I will describe the advances in Computational Accelerator Physics that brought us to this point and describe what to expect in regard to High Performance Computing in the future. This writeup is based on my presentation at ICAP'18 along with some additional comments that I did not include originally due to time constraints.

## INTRODUCTION AND BRIEF HISTORY

The first conference in what would become the Computational Accelerator Physics series was held 30 years ago in San Diego, California in January 1988. At the time I was 28 years old. The meeting was called the Conference on Linear Accelerator and Beam Optics Codes [1]. I think there are three of us here now who were present for that meeting: Martin Berz, Herman Wollnik, and me. I'll describe the ICAP conference series in a moment, but first want to briefly address the the origins of the field of Computational Accelerator Physics. This summary is based on the paper, "Oh Camelot! A memoir of the MURA years," by F.T. Cole [2].

As is well known, Lawrence invented the first cyclotron in 1930 inspired by the work of Rolf Wideroe on resonance acceleration. In 1940 Donald Kerst built the first betatron, a 2 MeV electron machine. Soon after WWII Edwin McMillan was at Los Alamos waiting to return to Berkeley. According to Cole, McMillan told him that, in a single evening, he worked out the concepts for the sychrocyclotron and the synchrotron. Independently in the Soviet Union Vladimir Veksler did the same. Two proton synchrotrons were built in the early 1950's to go beyond a GeV, the Cosmotron at Brookhaven and the Bevatron at Berkeley.

Along with progress in circular accelerators there were also developments in linear accelerators. Luis Alvarez developed the first proton linac at Berkeley in 1948. Also, developments in radar during WWII led to high frequency, GHz power sources that Hanson and Panofsky used to develop electron linacs at Stanford.

A revolution in accelerator physics took place in 1952 with the invention of strong focusing by Courant, Snyder,

and Livingston. As it turns out, Nick Christopholis had actually filed for a patent on strong focusing in 1950 and it was eventually granted in 1954. John Blewitt (BNL) applied alternating-gradient focusing to high intensity linacs. Also, the concept of Fixed Field Alternating Gradient (FFAG) was invented independently by multiple researchers, including Symon in 1954.

Strong focusing provided a totally new approach to high energy accelerators. A new lab, CERN, was founded after the war. Thanks to Lew Kowarski CERN acquired its first electronic computer in 1958. The CERN PS was commissioned in the Fall of 1959. The 30 GeV AGS at BNL began operation in 1960.

## THE BEGINNING OF COMPUTATIONAL ACCELERATOR PHYSICS

So far I've described some key developments in accelerator physics through the 1950's. The 1950's also brings us to the first digital computations for accelerator modeling. While there was plenty of activity in the field, I would particularly like to mention the work of L. Jackson Laslett. Laslett was a pioneer in using digital computers for orbit calculations and for calculating electromagnetic fields. There are records of Laslett performing his simulations on a computer known as the ILLIAC I, a computer comprised of 2800 vacuum tubes. While working for the Midwestern Universities Research Association Laslett observed and analyzed sensitive dependence on initial conditions – what we now call chaos. He did this in the mid 1950's. His studies actually predate the work of Edward Lorenz who discovered chaos in weather simulations and whose 1962 paper launched chaos theory. Of course mathematicians going back to Poincare had predicted dynamical behavior that we now describe as chaotic dynamics.

I would also like to mention another important event of the 1950's involving scientific computing that included someone who would later become heavily involved in Computational Accelerator Physics. That event was the simulation of the Fermi-Pasta-Ulam (FPU) problem, and the person involved was Mary (Tsingou) Menzel [3]. Mary was the programmer for the FPU problem on the MANIAC computer at Los Alamos National Laboratory (LANL). I met her in the 1980's. By then she was a member of the Accelerator Technology Division at LANL and I was a graduate student who spent my summers there. I remember Mary telling me there were cans of water on top of the computer for cooling!

Along with computational developments, there were also key theoretical developments in the 1950s. Most notably, Kolmogorov published his original paper in 1954, which set the stage for the KAM Theorem. A key consequence,

\* rdryne@lbl.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

relevant to long-term dynamics in circular accelerators, is that, under sufficiently small perturbations of an integrable Hamiltonian system, there remains a set of initial conditions whose orbits are quasi-periodic. Let me also mention that Jim Ellison is in the audience, and Jim's former student Scott Dumas has published a wonderful book on KAM history [4].

As evidence of the growing appreciation of computing, let me mention that the 1967 issue of CERN Courier was devoted to "the electronic computer and its use at CERN" [5]. Since this conference is attended by a large number of programmers I thought it would be interesting to show this quote from that issue:

*The designers of the early computers assumed that programming would be done by small groups of specialists, probably mathematicians, and that it would be undesirable to make the task too easy. For example, von Neumann and Goldstine, who in 1946 proposed what is essentially the modern computer, argued against built-in floating-point arithmetic: "The floating binary point represents an effort to render a thorough mathematical understanding of at least part of the problem unnecessary, and we feel that this is a step in a doubtful direction."*

In 1972 a second issue of the CERN Courier was published that was devoted to computers [6]. The opening article was by Lew Kowarski and titled, "Computers: Why?" It is remarkable how prescient Kowarski was about how computers would be used in the future. In the article he states, "We are only beginning to discover and explore the new ways of acquiring scientific knowledge which have been opened by the advent of computers..." He then goes on to state eight modes of application: (1) numerical mathematics, (2) data processing, (3) symbolic calculations, (4) computer graphics, (5) simulation, (6) file management and retrieval, (7) pattern recognition, and (8) process control.

## EARLY ACCELERATOR CODES

The preceding developments in digital computing and in accelerator theory and dynamical systems theory would lead to the topics that we address in these conference series. The code TRANSPORT came on the scene in the 1960's. It was originally developed by Karl Brown at SLAC [7]. A second-order version was released around 1969. (For those of you familiar with the Berkeley Lab's Computational Research Division, the division head, David Brown, is Karl's son.) Dave Cary, at Fermilab, developed a third-order version TRANSPORT. Ed Heighway, at Los Alamos, developed a version called TRANSOPTR for design optimization.

A breakthrough in single-particle optics came with the invention of Lie Algebraic methods. In the USA this was led by Alex Dragt and his group at the University of Maryland [8] Alex was a originally a theorist in elementary particle physics. He later applied his skills to plasma physics, and with John Finn published the Dragt-Finn factorization theorem [9]. This shows how a Taylor series, as represented in a code like TRANSPORT, can be represented as a factored

product of Lie transformations, as in a code like MaryLie. See Fig. 1.

Alex's involvement in Accelerator Physics came by accident. He was planning a sabbatical in the Plasma Physics Division at Los Alamos in 1978/79 when the division folded. Fortunately Richard Cooper suggested to Alex that he do his sabbatical in the Accelerator Theory group that he headed in the Accelerator Technology Division at Los Alamos. This launched Alex's involvement in Accelerator Physics. And the rest is history...

$$\begin{aligned}
 \zeta^f &= \sum M \zeta^i + \sum \sum T \zeta^i \zeta^i + \sum \sum \sum U \zeta^i \zeta^i \zeta^i + \dots \\
 \mathcal{M} &= e^{f_2} e^{f_3} e^{f_4} \dots \\
 \zeta^f &= M \zeta^i = e^{f_2} (1 + f_3 + \frac{1}{2} f_3^2 + \dots) (1 + f_4 + \dots) \zeta^i
 \end{aligned}$$

Figure 1: Correspondence between a map represented as a Taylor series and a map represented as a factored product of Lie transformations.

Alex and his student Etienne Forest published an article on the equations of motion for the matrix  $M$  and for the polynomials  $f_3, f_4 \dots$ , in the Lie algebraic representation of the transfer map [10]. This opened the door to computing transfer maps for realistic beamline elements, i.e., for elements with fringe fields. The application of this became a portion of my Ph.D. thesis, known as the "genmap" capability in MaryLie. Eventually it was used to model realistic solenoids, dipoles, quadrupoles, and RF cavities [11].

Alex Dragt is retired but still active in the field. For those who would like to learn about Lie Algebraic methods, Alex has written a more than 2500 page book, that is freely available, "Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics" [12].

Though this talk is mainly about beam dynamics it is worth mentioning that the Superfish code, developed by Klaus Halbach and Ron Holsinger, was released in 1976 [13]. Later the Poisson and Superfish codes were maintained and developed by the Los Alamos Accelerator Code Group (LAACG). Though the Poisson and Superfish codes are only 2D they are still widely used for the early stages of accelerator design. The codes PARMILA, PARMELA, PARMTEQ, TRACE, and TRACE3D were all developed at Los Alamos.

In Europe, the first version of the code MAD was developed in the early 1980's [14]. This was led by F. Christoph Iselin along with Jim Niederer and Eberhard Keil. Originally a TRANSPORT-like code, Christoph eventually put large portions of MaryLie inside MAD. It's also worth mentioning that people like Karl Brown, Dave Cary, Christoph Iselin, and others, led an effort to develop a common input format that many of the major beam dynamics codes now use.

So far I have mentioned map-based codes like TRANSPORT, MaryLie, and MAD. Starting in the 1980's a different

approach emerged, based on direct numerical integration of the equations of motion by a symplectic method. This involved many people, but in accelerator physics it began with the third-order integrator of Ruth [15]. Later Forest and Ruth derived a fourth-order integrator using Lie methods [16]. Yoshida showed how to obtain an integrator of order  $2n+2$  by combining integrators of order  $2n$  [17]. The work of Yoshida was later extended by Forest *et al.* [18].

Another major development was the application of Differential Algebraic techniques and automatic differentiation to beam dynamics by Martin Berz [19]. This opened the door to performing Taylor series calculations to arbitrary order, as implemented in the code COSY-INFINITY [20]. Along with all these advances came the development of normal form techniques [21] that have been critical to understanding global properties of periodic transport systems and designing these systems.

The mention of COSY-INFINITY brings me back to the first Computational Accelerator Conference in 1988. The proceedings contain papers related to many of the codes above. And it is interesting to note that a paper by Berz mentions a code “under the tentative name COSY INFINITY.”

As a sign of how much things have changed since 1988, consider this quote from a paper in the proceedings: “The problem shown required 22 seconds on the IBM 3080 and 23 minutes on a machine with 8 MHz clock. . . The PC had the Intel 80287 Math co-processor and 1.1 Mbyte storage. . .”

As a sign of the promise of the future, consider this quote from a paper by Ed Heighway: “. . . the beam transport designer’s world is richer and probably evolving faster than at any time since Karl Brown first put finger to keypunch.” At this point in my presentation I felt obliged to show a picture of a keypunch machine because I thought that some people in the audience might not know what I was talking about.

Let me mention one more thing before leaving that 1988 conference: The proceedings say nothing about parallel computing. But that was about to change.

## 1990’s: PARALLEL COMPUTING ENTERS

The next meeting in the series was in 1990, hosted by Los Alamos. It was called the Conference on Computer Codes and the Linear Accelerator Community [22]. I counted five papers in the proceedings that mentioned parallel computing, although some of those described compatibility with parallel processing, not that they were actually doing it. One that I will mention specifically is, “Wakefield Calculation on Parallel Computers,” by Paul Schoessow. He mentions finite difference codes that run on an Alliant FX/8 and on a Connection Machine CM-2. This is the earliest paper I’m aware of on massively parallel accelerator modeling.

The next conference, called CAP’93, was held in 1993 in Pleasanton, California [23]. It was co-organized by me and Susarla Murty. This was the last CAP conference before the Superconducting Supercollider (SSC) was cancelled, and there were several talks from people associated with the SSC.

The 1990’s saw massively parallel computing emerge as a major new paradigm. My involvement came at the Advanced Computing Laboratory at Los Alamos. There, with Salman Habib and other colleagues, we developed early parallel beam dynamics codes and, eventually, the first version of the IMPACT code [24]. To a large extent this was motivated by a desire to use a large number of macroparticles to simulate very low density beam halos. It also opened the door to performing practical 3D space-charge calculations. We did this mainly on a computer by Thinking Machines Corporation called the Connection Machine 5. At first we used a technique from computational cosmology by Ferrell and Bertschinger to compute space-charge effects. But eventually we adopted parallel particle-in-cell techniques of Paulette Liewer, Victor Decyk, and others [25].

During the 1980’s when I spent my summers at Los Alamos I remember thinking that the space-charge code developers (who were interested in high intensity linacs) and the single-particle optics modelers (who were interested in aberrations, dynamic aperture, fringe fields, etc.) did not interact much. By the mid-1990’s my experience with high-order optics and parallel particle-in-cell methods, along with symplectic integrators, led me to introduce split-operator methods as a means to combine the best of both worlds [26, 27]. See Fig. 2.

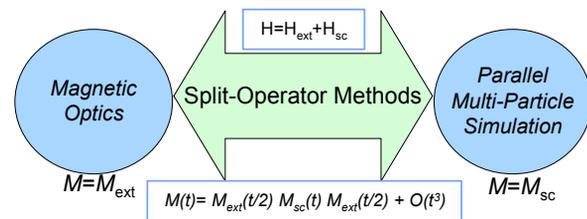


Figure 2: Split-operator method for combining high-order optics with space-charge.

The next conference, called CAP’96, was held in Williamsburg in 1996 [28]. At about this time (1996) NERSC moved to Lawrence Berkeley National Lab. Also, computer systems were shifting away from vector machines to massively parallel machines. I remember being involved in an email exchange about what the next big computer should be at NERSC. A decision was made that it should be a massively parallel Cray T3E. It was called mcurie.

By this time parallel computing was becoming a major activity in the accelerator physics community. In the USA, in 1997, the US Department of Energy launched the DOE Grand Challenge in Computational Accelerator Physics. The first Teraflop computer also came on the scene in 1997.

The first conference to be called the *International Computational Accelerator Physics Conference* was held in Monterey, California in 1998, organized by me and Kwok Ko [29]. Parallel processing is highly evident in these proceedings. Also, Python begins to be seen, mentioned in a paper by Grote, Friedman, and Haber called “New Methods in WARP.”

## 2000's: TERASCALE ERA

In the 2000's there were five ICAP conferences: Darmstadt in 2000, East Lansing in 2002, St. Petersburg in 2004, Chamonix in 2006, and San Francisco in 2009 [30–34]. As is evident, the conference venue had become truly international. During this decade there were major advances both in single-particle beam dynamics codes and in large-scale multi-physics beam dynamics codes. The DOE SciDAC program started in 2002. The IMPACT code suite was fully developed at the Berkeley Lab by J. Qiang [35]. The first version of Synergia was developed at Fermilab by Panagiotis Spentzouris and James Amundson [36]. It combined portions of IMPACT with the Leo Michelotti's mxyzptlk/beamline libraries. Andreas Adelman developed the OPAL library [37]. Bmad was developed by David Sagan and others [38]. The Polymorphic Tracking Code, PTC, was developed by Etienne Forest [39]. The first million particle strong-strong beam-beam simulation was performed in 2004 [40]. The first billion particle linac simulation was performed in 2007 [41]. The ACE3P package was developed at SLAC, led by Kwok Ko, and was able to perform very high accuracy electromagnetic calculations involving extremely complicated 3D structures [42]. The first petaflop computer appeared in 2008.

## 2010's: PETASCALE ERA

Following ICAP 2009 it was decided to have the conference on a three-year cycle. So in the 2010's there have been three conferences, Rostock in 2012, Shanghai in 2015, and this meeting in Key West in 2018 [43–45].

During that time we've seen the meaning of "large scale" grow from tens of thousands of processes in the previous decade to hundreds of thousands in the current decade. And the very largest scale simulations now exceed a million processes.

Big Data emerged as a major paradigm. Though the accelerator community's design needs don't usually involve it, our experiments, like those at places such as LHC, RHIC, and the light sources have helped drive developments in Big Data. It's well known that in fields like Cosmology, observations are quickly analyzed on supercomputers where the results drive the direction of observational resources. But that's happening in Accelerator Science too. For example, there is now a data pipeline between light source experiments and the NERSC supercomputer center.

Multi-level parallelization has grown increasingly dominant during this decade. This includes multiple levels of MPI, or MPI with threads, or MPI across nodes with hardware acceleration on a node. Multi-level MPI, in particular, has provided a relatively easy path to parallel parameter scans and parallel design optimization. Parallel design optimization has become one of the main uses of large-scale modeling, with many people using genetic optimizers.

## PRESENT DAY, INTO THE FUTURE

I'll begin this final section of my talk by describing something that we are doing now, namely, performing 3D simulations of coherent synchrotron radiation.

As I look back I see the 2000's as a kind of Golden Era in space-charge modeling. Over the course of that decade, several multi-physics parallel beam dynamics code emerged that had 3D space-charge capability.

Now we are at the beginning of such an era in radiation modeling. This problem is extremely challenging. Consider that an N-body space-charge calculation requires  $N^2$  operations; an analogous Lienard-Wiechert calculation would require  $N^2$  operations but its difficulty is compounded by the fact that it would include the time-history of all particles. The physics of the problem further complicates the situation because the radiation cone is extremely narrow at high energy. In the past I have described this as being like a large number of flashlights that interact when their narrow light beams collide, taking into account light travel time.

One method for addressing this problem is known as the Lienard-Wiechert Particle-Mesh (LWPM) method [46]. This approach extends the widely used convolution-based method for modeling space-charge, but replaces the Coulomb Green function with the Lienard-Wiechert Green function.

It is well known that the most common method of computing 3D space charge in unbounded systems is to perform an FFT-based discrete convolution of a charge density with a Green function,

$$\phi_{i,j,k} = \frac{\delta_x \delta_y \delta_z}{4\pi\epsilon_0} \sum_{i'=1}^{i'_{max}} \sum_{j'=1}^{j'_{max}} \sum_{k'=1}^{k'_{max}} \rho_{i',j',k'} G_{i-i',j-j',k-k'}, \quad (1)$$

where  $(\delta_x, \delta_y, \delta_z)$  is the grid cell size,  $\rho_{i,j,k}$  is the charge density at the grid points, and  $G_{i-i',j-j',k-k'}$  denotes  $G$  at values of grid point separation. Naively a convolution would scale as  $N^2$ , but because the approach is FFT-based it scales as  $N \log N$ .

In the case of a space-charge modeling code the quantity  $G$  is just the Coulomb Green function for the potential or the fields. The calculation of the  $G$  would then require only a few floating point operations (flops). The model can be made much more robust by using an Integrated Green function (IGF) instead of using the value of the "bare" Green function at the grid points [47]. Even so, the calculation of the IGF requires just a modest number of flops.

The transition to a model that includes both space-charge and radiation begins with the following observation: In space-charge codes the process is usually described as transforming the particles to the bunch frame where the motion is non-relativistic, solving for the potential or field on a grid, and transforming back to the lab frame. But the procedure can also be viewed as using the Heaviside representation of the Green function in the lab frame. In the case of the potential,

$$G_{\phi,heav} = \frac{1}{\gamma^2 r} \frac{1}{(1 - |\boldsymbol{\beta} \times \hat{\mathbf{r}}|^2)^{1/2}}, \quad (2)$$

where  $\hat{\mathbf{r}}$  points from the (instantaneous) position of the charge to the observation point.

The transition to Lienard-Wiechert modeling replaces the Heaviside Green function – which is based on straight line motion at constant velocity – with the full Lienard-Wiechert Green function. For example, in the case of the electric field,

$$\mathbf{G}_{LW} = \left[ \frac{q}{\gamma^2 \kappa^3 R^2} (\hat{\mathbf{n}} - \vec{\beta}) + \frac{q}{\kappa^3 R c} \hat{\mathbf{n}} \times \left\{ (\hat{\mathbf{n}} - \vec{\beta}) \times \frac{\partial \vec{\beta}}{\partial t} \right\} \right]_{\text{ret}} \quad (3)$$

Here,  $\hat{\mathbf{n}} = \mathbf{R}/|R|$  is a unit vector pointing from the retarded emission point to the observation point,  $\vec{\beta} = \mathbf{v}/c$ ,  $c$  is the speed of light,  $\gamma = 1/\sqrt{1 - \beta^2}$ , and  $\kappa = 1 - \hat{\mathbf{n}} \cdot \vec{\beta}$ .

But because it is preferable to use an IGF, one should embed this capability within a 3D quadrature package. So, to compute the Green function in a LWPM code, one calls a Lienard-Wiechert solver potentially millions of times to compute the Green function. This would be absolutely impossible in a serial code. But in a parallel code it is quite effective. It is even a good fit to current architectures because it involves a huge number of flops but basically no data movement to compute the Green function.

Previous studies have shown that the LWPM method agrees well with brute-force Lienard-Wiechert summation for the case of steady-state dipole radiation and for the case of a bunched beam inside a wiggler magnet [48]. Recently we have looked at the dipole example in a regime where the Coulomb field and the radiation field are comparable. Figure 3 shows the transverse electric field. This example corresponds to a 40 MeV electron bunch in a 0.16 T magnetic field. The bunch is Gaussian with an rms bunch size of 100 micron in  $x$ ,  $y$ , and  $z$ . In this case the summation used 1 billion simulation particles. The summation and convolution results are in excellent agreement.

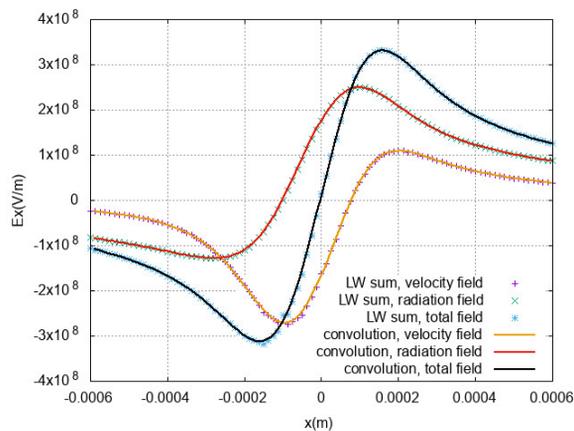


Figure 3:  $E_x$  vs  $x$  for the steady-state dipole test problem, plotted along the  $x$ -axis going through the bunch center. The LW velocity field, LW radiation field, and total field are shown. Results are shown for the LW summation over 1B particles and for the convolution-based method.

A 2D plot of the magnitude of the total transverse field is shown in Fig. 4. Note that magnitude is slightly larger

for positive  $x$ , and there is a slight tilt in the dark band with respect to the line  $x = 0$ . These features would not be present in a space-charge code, i.e., a code based on the Heaviside approximation.

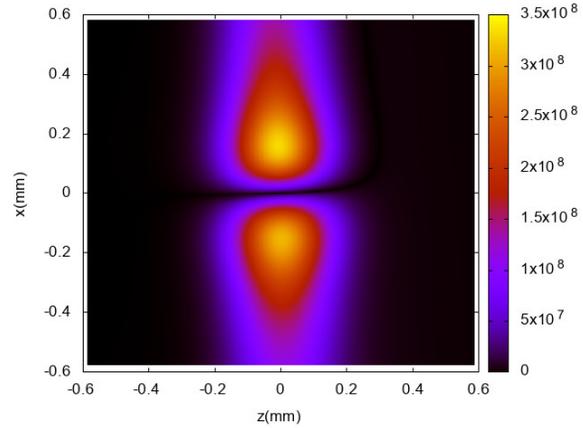


Figure 4:  $|E_{x,total}|$  in the midplane  $(x, 0, z)$  for the steady-state dipole test problem. The combined field does not show the significant asymmetry that was present in  $|E_{x,vel}|$  and  $|E_{x,rad}|$  separately, but the field is still slightly larger in magnitude at positive  $x$ . Also, there is a slight tilt visible in the dark band with respect to the line  $x = 0$ .

To conclude this section I will discuss plans for exascale simulation in the 2020's. But first consider the following: At the time of the first Computational Accelerator Physics Conference in 1988, the fastest computer in the world was the Cray Y-MP. It had 8 vector processors running at 167 MHz and a performance of around 2 Gflops. Ten years later, at the time of ICAP'98 in Monterey, the teraflop barrier had recently been broken (in 1997) by the Intel ASCI Red computer built under the Accelerated Strategic Computing Initiative. Ten years after that, a year before ICAP'09 in San Francisco, the IBM Roadrunner computer at Los Alamos broke the petaflop barrier. And immediately people were thinking about the next big advance, as is clear from this ComputerWorld headline on June 9, 2008: "All hail Roadrunner's petaflop record; now, what about the exaflop?"

Now we're at ICAP 2018. Here in the USA there's a project called the Exascale Computing Project (ECP) [49]. It covers many scientific fields. One of the fields is advanced particle accelerator design, particularly plasma accelerator design [50]. Plasma accelerators have the potential to greatly reduce the size and cost of accelerators, with profound consequences for science and society. They may also provide a novel and economically viable path to the high-energy frontier through a plasma-based collider.

This is a case where large-scale modeling serves multiple purposes: First, it is a tool of discovery that allows us to explore the complex physical processes occurring in plasma accelerators. In some cases these processes may be extremely difficult or impossible to access experimentally, or it may be very expensive and time-consuming to do so. Exploration of plasma accelerators via large-scale

simulation will lead to insights that would otherwise be inaccessible. Second, it allows us to examine the feasibility of advanced concepts like plasma-based colliders. At present such simulations are too slow for rapid and thorough exploration of the parameter space. Under ECP, through advanced hardware and advanced algorithms able to run effectively on that hardware, such simulations will be possible in the early 2020's, which is the same time that exascale systems will become available. Lastly, large-scale simulation allows us to optimize the design of advanced accelerator concepts, and to develop designs that reduce cost and risk.

As we march toward the exascale era the computing environment is changing. Users requiring the most massive resources will soon have a fraction of an exaflop at their fingertips. Importantly, medium-scale users will also have increased computer power. To get a major boost in performance we will have to write code for heterogeneous hardware (CPUs, GPUs, etc.), using a mix of computer programming methods. In addition software libraries are already being written with a view toward exascale. For example, under ECP the Center for Particle Applications is developing a number of software libraries including libraries for parallel FFTs [51, 52]. Already FFTs can be performed on hundreds of thousands of cores for problem sizes up to 10,000<sup>3</sup>.

To conclude this section I will mention an example from another field, Computational Cosmology. Under the ExaSky ECP project [53], a simulation was performed on 1.5 million cores of the Sequoia computer at a sustained performance of nearly 14 petaflops, and used 3.6 trillion simulation particles. Such high resolution simulations are needed to make comparisons with high precision experimental measurements. The simulations are used to solve inverse problems to determine several key cosmological parameters like the amount of dark matter, parameters of primordial fluctuations, etc.

## CONCLUSION

I will conclude my talk by quoting something that I presented 10 years ago at the 2008 European Particle Accelerator Conference [54]. It is found in the Proceedings of the 1971 International Conference on High Energy Physics [55]. In response to a talk by Viktor Weisskopf, Lew Kowarski (who I mentioned previously) made a comment that was recorded in the Proceedings. Weisskopf had described the emergence of “a new type of physicist... the machine physicists...” In the question and answer session Kowarski spoke. According to the proceedings he said,

*“Early experimentalists worked with their hands: Galileo’s legendary tossing of stones from the Tower of Pisa, or the alchemists mixing by hand the ingredients in their mixing bowls. In a similar way the theoreticians manipulated their numerical quantities and symbols by their unaided brain-power. Then came the machines to extend the experimenter’s manual skill and to open whole new worlds of things to be handled in ways nobody could predict or even imagine before they really got going. Now we are at the be-*

*ginning of a new kind of extension by machine: the computer comes to supplement the theoretician’s brain. We cannot foresee what this fourth kind of creativity in physics will bring...”*

This comment was made nearly 50 years ago when the fastest computer in the world was the CDC 7600 with a performance of about 10 Mflops. Sometime in the 2020's we will have exascale resources that have 100 trillion times the computing power that Kowarski knew 1971. Such a mind-boggling increase in computing power would have been almost unimaginable in 1971, and validates Kowarski's comment that “we cannot foresee what this fourth kind of creativity in physics will bring.”

At conferences like this ICAP conference we share our experiences of what this fourth kind of creativity has brought to our field. More than ever, advanced computational modeling is enabling major advances and discoveries in Accelerator Physics. Opportunities abound in concepts like laser, plasma, and dielectric accelerators, in new approaches like integrable optics, in accelerator control and operation, in concepts for future colliders and future light sources, and in applications of accelerators.

## ACKNOWLEDGEMENTS

The author thanks Salman Habib, John Shalf, and Jean-Luc Vay for helpful discussions. The simulation results presented here used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

## REFERENCES

- [1] *Linear Accelerator and Beam Optics Codes*, Charles R. Emmerich, Ed., AIP Conference Proceedings, vol. 177, 1988.
- [2] F. T. Cole, “Oh, Camelot! A Memoir of the MURA Years”, in *Proc. Cycl. Conf.*, April 11, 1994. <http://jacow.org/c01/cyc2001/extra/Cole.pdf>
- [3] T Dauxois, “Fermi, Pasta, Ulam, and a mysterious lady,” *Physics Today*, vol. 61, pp. 55. <https://arxiv.org/pdf/0801.1590.pdf>
- [4] H. Scott Dumas, *The KAM Story: A Friendly Introduction to the Content, History, and Significance of Classical Kolmogorov-Arnold-Moser Theory*, World Scientific Publishing Company, 2014.
- [5] *CERN Courier*, vol. 7, no. 9, September 1967.
- [6] *CERN Courier*, vol. 12, no. 3, March 1972.
- [7] K. L. Brown, “First and Second Order Matrix Theory for Design of Beam Transport Systems,” Stanford Linear Accelerator Center, report SLAC-75, 1967.
- [8] Alex J. Dragt *et al.*, “Lie Algebraic Treatment of Linear and Nonlinear Beam Dynamics,” *Ann. Rev. Nucl. Part. Sci.*, vol. 38, pp. 455-496, 1988.
- [9] A. J. Dragt and J. M. Finn, “Lie series and invariant functions for analytic symplectic maps,” *J. Math. Phys.*, vol. 17, pp. 2215, 1976.

- [10] A. Dragt and E. Forest, "Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods," *J. Math. Phys.*, vol. 24, pp. 2734, 1983.
- [11] R. D. Ryne and A. J. Dragt, "Numerical Computation of Transfer Maps using Lie Algebraic Methods," in *Proc. PAC'87*, Washington, DC, USA.
- [12] A. J. Dragt, *Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics*, <https://www.physics.umd.edu/dsat/>.
- [13] <https://laacg.lanl.gov/>.
- [14] F. Christoph Iselin, "The MAD Program," in *Proc HEACC83*, Batavia, IL, USA.
- [15] R. D. Ruth, "A canonical integration technique," *IEEE Trans. Nucl. Sci.*, NS-30(4), August 1983.
- [16] E. Forest and R. D. Ruth, "Fourth-order symplectic integration," *Physica D*, vol. 43, no. 1, pp. 105-117, 1989.
- [17] H. Yoshida, "Construction of higher order symplectic integrators," *Phys. Lett. A*, vol. 150, no. 5-7, pp. 262-268, 1990.
- [18] E. Forest, J. Bengtsson, and M. F. Reusch, "Application of the Yoshida-Ruth techniques to implicit integration and multi-map explicit integration," *Phys. Lett. A*, vol. 158, no. 3, pp. 99-101, 1991.
- [19] M. Berz, "Differential Algebraic Description of Beam Dynamics to Very High Orders," *Particle Accelerators*, vol. 24, pp. 109-124, 1989.
- [20] <http://cosyinfinity.org/>.
- [21] E. Forest, M. Berz, and J. Irwin, "Normal form methods for complicated periodic systems: A complete solution using Differential algebra and Lie operators," *Particles Accelerators*, vol. 24, pp. 91, 1989.
- [22] *Proceedings of the Conference on computer Codes and the Linear Accelerator Community*, Richard K. Cooper, Ed., LANL Report No. LA-11857-C, 1990, [https://inis.iaea.org/collection/NCLCollectionStore/\\_Public/21/094/21094506.pdf](https://inis.iaea.org/collection/NCLCollectionStore/_Public/21/094/21094506.pdf)
- [23] *Proc. CAP'93*, Robert Ryne, Ed., AIP Conference Proceedings, vol. 297, 1993.
- [24] R. D. Ryne *et al.*, "The US DOE Grand Challenge in Computational Accelerator Physics," in *Proc. Linac'98*, Chicago, USA, 1998.
- [25] P. C. Liewer and V. K. Decyk, "A general concurrent algorithm for plasma particle-in-cell simulation codes," *J. Comput. Phys.*, vol. 85, pp. 302, 1989.
- [26] R. D. Ryne, S. Habib, and T. P. Wangler, "Halos of Intense Proton Beams," in *Proc. PAC'95*, Dallas, TX, USA, 1995.
- [27] R. D. Ryne and S. Habib, "Beam Dynamics Calculations and Particle Tracking using Massively Parallel Processors," in *Proc. LHC'95*, Montreux, Switzerland, 1995.
- [28] *1996 Computational Accelerator Physics Conference*, J. Bisognano and A. Mondelli, Eds., AIP Conference Proceedings, vol. 391, 1997. <http://inspirehep.net/record/458928>
- [29] *Proc. ICAP'98*, K. Ko and R. D. Ryne, Eds., [http://www.slac.stanford.edu/econf/C980914/papers/ICAP98\\_eConf.pdf](http://www.slac.stanford.edu/econf/C980914/papers/ICAP98_eConf.pdf)
- [30] <http://prst-ab.aps.org/speced/ICAP2000/>.
- [31] <http://www.bt.pa.msu.edu/ICAP02/main.html>
- [32] *Proc ICAP'04*, St. Petersburg, Russia, June 29-July 2, 2004 D. Ovsyannikov, M. Berz, K. Makino, Eds., *Nucl. Instrum. Meth. A558*, 2006, pp.1-365.
- [33] <http://jacow.org/ICAP06/>.
- [34] <http://jacow.org/ICAP2009/>.
- [35] J. Qiang *et al.*, "An object-oriented parallel particle-in-cell code for beam dynamics simulation in linear accelerators," *J. Comput. Phys.*, vol. 163, no. 2, pp. 434-451, 2000.
- [36] J. Amundson, P. Spentzouris, J. Qiang, and R. Ryne, "Syn-ergica: An accelerator modeling tool with 3-D space charge," *J. Comput. Phys.*, vol. 211, no. 1, pp. 229-248, 2006.
- [37] A. Adelmann *et al.*, "The object-oriented parallel accelerator library (OPAL), design, implementation, and application," in *Proc. ICAP'09*, <http://jacow.org/ICAP2009/papers/we3iopk01.pdf>
- [38] J. Urban, L. Fields, D. Sagan, "Linear Accelerator Simulations with Bmad," in *Proc. PAC'05*, Knoxville, TN, USA.
- [39] E. Forest, Y. Nogiwa, and F. Schmidt, "The FPP Documentation," in *Proc. ICAP'06*, Chamonix, Switzerland.
- [40] J. Qiang, M. A. Furman, and R. D. Ryne, "Strong-strong beam-beam simulation using a Green function approach," *Phys. Rev. ST Accel. Beams*, vol. 5, pp. 104402, 2002.
- [41] J. Qiang, I. Pogorelov, and R. D. Ryne, "Parallel Beam Dynamics Simulation Tools for Future Light Source Linac Modeling," in *Proc. PAC'07*, Albuquerque, NM, USA.
- [42] [https://portal.slac.stanford.edu/sites/ard\\_public/acd/Pages/Default.aspx](https://portal.slac.stanford.edu/sites/ard_public/acd/Pages/Default.aspx)
- [43] <http://jacow.org/ICAP2012/>.
- [44] <http://jacow.org/ICAP2015/>.
- [45] <http://jacow.org/ICAP2018/>.
- [46] R. D. Ryne, J. Qiang, C. E. Mitchell, and B. E. Carlsten, "Self-Consistent Modeling using a Lienard-Wiechert Particle-Mesh Method," in *Proc. IPAC'18*, Vancouver, BC, Canada.
- [47] J. Qiang, S. Lidia, R. D. Ryne, and C. Limborg-Deprey, "3D quasistatic model for high brightness beam dynamics simulation," *Phys. Rev. ST Accel. Beams*, vol. 9, pp. 044204, 2006.
- [48] R. D. Ryne *et al.*, "Using a Lienard-Wiechert Solver to Study Coherent Synchrotron Radiation Effects," in *Proc. FEL'13*, New York, NY, USA.
- [49] <http://www.exascaleproject.org/>.
- [50] <http://www.exascaleproject.org/project/warpx-exascale-modeling-of-advanced-particle-accelerators/>.
- [51] <http://www.exascaleproject.org/project/copa-co-design-center-particle-applications/>.
- [52] S. Plimpton, "fftMPI, a distributed-memory parallel FFT library," <https://fftmpl.sandia.gov/>.
- [53] <http://www.exascaleproject.org/project/exasky-computing-sky-extreme-scales/>.
- [54] R. D. Ryne, "Advanced Computing Tools and Models for Accelerator Physics," in *Proc. EPAC'08*, Genoa, Italy.
- [55] V. F. Weisskopf, "The Past and the Future of High Energy Physics", in *Proc. 8th International Conference on High-Energy Accelerators*, M. Hildred Blewett, ed., 1971.

# CHALLENGES IN SIMULATING BEAM DYNAMICS OF DIELECTRIC LASER ACCELERATION

U. Niedermayer<sup>1,2\*</sup>,

A. Adelman<sup>7</sup>, R. Aßmann<sup>4</sup>, S. Bettoni<sup>7</sup>, D. S. Black<sup>2</sup>, O. Boine-Frankenheim<sup>1</sup>, P. N. Broaddus<sup>2</sup>, R. L. Byer<sup>2</sup>, M. Calvi<sup>7</sup>, H. Cankaya<sup>4,11,15</sup>, A. Ceballos<sup>2</sup>, D. Cesar<sup>10</sup>, B. Cowan<sup>9</sup>, M. Dehler<sup>7</sup>, H. Deng<sup>2</sup>, U. Dorda<sup>4</sup>, T. Egenolf<sup>1</sup>, R. J. England<sup>3</sup>, M. Fakhari<sup>4</sup>, A. Fallahi<sup>11</sup>, S. Fan<sup>2</sup>, E. Ferrari<sup>5,7</sup>, F. Frei<sup>7</sup>, T. Feurer<sup>12</sup>, J. Harris<sup>2</sup>, I. Hartl<sup>4</sup>, D. Hauenstein<sup>7</sup>, B. Hermann<sup>7,12</sup>, N. Hiller<sup>7</sup>, T. Hirano<sup>2</sup>, P. Hommelhoff<sup>6</sup>, Y.-C. Huang<sup>13</sup>, Z. Huang<sup>2</sup>, T. W. Hughes<sup>2</sup>, J. Illmer<sup>6</sup>, R. Ischebeck<sup>7</sup>, Y. Jiang<sup>2</sup>, F. Kärtner<sup>4,11,15</sup>, W. Kurozumi<sup>4,15</sup>, T. Langenstien<sup>2</sup>, Y. J. Lee<sup>8</sup>, K. Leedle<sup>2</sup>, F. Lemery<sup>4</sup>, A. Li<sup>6</sup>, C. Lombosi<sup>7</sup>, B. Marchetti<sup>4</sup>, F. Mayet<sup>4,15</sup>, Y. Miao<sup>2</sup>, A. Mittelbach<sup>6</sup>, P. Musumeci<sup>10</sup>, B. Naranjo<sup>10</sup>, A. Pigott<sup>2</sup>, E. Prat<sup>7</sup>, M. Qi<sup>8</sup>, S. Reiche<sup>7</sup>, L. Rivkin<sup>5,7</sup>, J. Rosenzweig<sup>10</sup>, N. Saprà<sup>2</sup>, N. Schönenberger<sup>6</sup>, X. Shen<sup>10</sup>, R. Shiloh<sup>6</sup>, E. Skär<sup>1</sup>, E. Simakov<sup>14</sup>, O. Solgaard<sup>2</sup>, L. Su<sup>2</sup>, A. Tafel<sup>6</sup>, S. Tan<sup>2</sup>, J. Vuckovic<sup>2</sup>, H. Xuan<sup>4,15</sup>, K. Yang<sup>2</sup>, P. Yousefi<sup>6</sup>, Z. Zhao<sup>2</sup>, J. Zhu<sup>4</sup>

<sup>1</sup> Technische Universität Darmstadt, 64289 Darmstadt, Germany

<sup>2</sup> Stanford University, Stanford, CA 94305, USA

<sup>3</sup> SLAC National Accelerator Laboratory, Menlo Park, CA 94025

<sup>4</sup> Deutsches Elektronen-Synchrotron, D-22607 Hamburg, Germany

<sup>5</sup> École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

<sup>6</sup> Friedrich-Alexander Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

<sup>7</sup> Paul Scherrer Institut CH-5232 Villigen, Switzerland

<sup>8</sup> Purdue University, West Lafayette, IN 47907, USA

<sup>9</sup> Tech-X Corporation, Boulder, CO 80303, USA

<sup>10</sup> University of California, Los Angeles, CA 90095, USA

<sup>11</sup> Center for Free-Electron Laser Science, 22607 Hamburg, Germany

<sup>12</sup> Universität Bern, Switzerland

<sup>13</sup> Nat. Tsing Hua University, Taiwan

<sup>14</sup> Los Alamos National Laboratory, USA

<sup>15</sup> Universität Hamburg, 22761 Hamburg, Germany

## Abstract

Dielectric Laser Acceleration (DLA) achieves the highest gradients among structure-based electron accelerators. The use of dielectrics increases the breakdown field limit, and thus the achievable gradient, by a factor of at least 10 in comparison to metals. Experimental demonstrations of DLA in 2013 led to the Accelerator on a Chip International Program (ACHIP), funded by the Gordon and Betty Moore Foundation. In ACHIP, our main goal is to build an accelerator on a silicon chip, which can accelerate electrons from below 100 keV to above 1 MeV with a gradient of at least 100 MeV/m. For stable acceleration on the chip, magnet-only focusing techniques are insufficient to compensate the strong acceleration defocusing. Thus spatial harmonic and Alternating Phase Focusing (APF) laser-based focusing techniques have been developed. We have also developed the simplified symplectic tracking code DLATRACK6D, which makes use of the periodicity and applies only one kick per DLA cell, which is calculated by the Fourier coefficient of the synchronous spatial harmonic. Due to coupling, the

Fourier coefficients of neighboring cells are not entirely independent and a field flatness optimization (similarly as in multi-cell cavities) needs to be performed. The simulation of the entire accelerator on a chip by a Particle In Cell (PIC) code is possible, but impractical for optimization purposes. Finally, we have also outlined the treatment of wake field effects in attosecond bunches in the grating within DLATRACK6D, where the wake function is computed by an external solver.

## INTRODUCTION

The Accelerator on a Chip International Program (ACHIP) [1], funded by the Gordon and Betty Moore Foundation in the period between 2015 and 2020, aims to explore Dielectric Laser Acceleration (DLA). This nascent acceleration scheme provides the highest gradients among structure-based (non-plasma, non-vacuum, etc.) electron accelerators and thus allows reduction of the size of high energy electron accelerators significantly. The principle of DLA relies on the inverse Smith-Purcell (or the inverse Cerenkov effect) and was first proposed in 1962 [2, 3]. In 2013, the accel-

\* niedermayer@temf.tu-darmstadt.de

eration of relativistic electrons was first demonstrated at SLAC with a gradient of more than 250 MeV/m in a SiO<sub>2</sub> double grating structure driven by a 800 nm Ti:Sapphire laser [4]. In the same setup, the gradient was later increased to 690 MeV/m [5]. Also in 2013, strongly sub-relativistic electrons (27.7 keV) were accelerated by the group at FAU Erlangen with a gradient of 25 MeV/m using a single grating structure at the third spatial harmonic [6]. The group at Stanford University used a silicon dual pillar structure to accelerate 96 keV electrons with a gradient of more than 200 MeV/m [7] and a similar experiment at 30 keV with few-cycle laser pulses was done at FAU Erlangen [8]. An example of such a dual pillar structure is shown in Fig. 1. These schemes all utilize laterally incident lasers with polar-

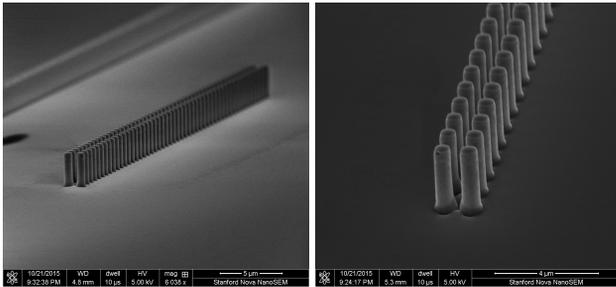


Figure 1: Scanning Electron Microscope (SEM) pictures of a dual pillar acceleration structure.

ization in the electron beam direction, thus the accelerating near field is a standing wave. It is also possible to use longitudinally coupled (traveling wave) structures, see [9] for details and a general overview.

The goal of ACHIP is to build an accelerator as sketched in Fig. 2, which can accelerate electrons from electrostatic sources (<100 keV) to above 1 MeV. A second goal is to make use of the accelerator by exploring the options of DLA based deflection (see e.g. [10]), which can potentially lead to laser driven undulators [11–13].

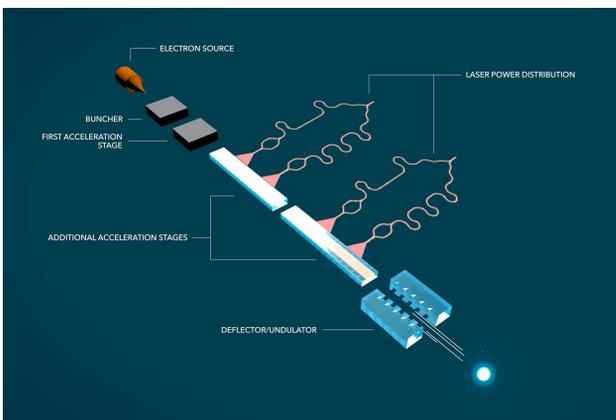


Figure 2: Sketch of the goals of the ACHIP collaboration.

Different materials have been investigated for DLA [14]. In order to achieve the highest gradient, the material-specific damage threshold fluence

$$F_{\text{dam}} > F = \frac{P\Delta t}{A}, \quad (1)$$

where  $P/A$  is the laser intensity and  $\Delta t$  is the pulse length, is approached, but must not be exceeded. Note that the dependence on the laser wavelength and pulse length can be strongly nonlinear, see e.g. [15] for an in-depth discussion and more empirical data. Moreover, if the laser travels a longer distance through the material, the nonlinear phase shift also needs to be considered [16]. In general, a shorter pulse allows for a higher gradient at the same fluence. Moreover, a high band gap material as e.g. SiO<sub>2</sub> will have a higher damage threshold and a low band gap material such as Silicon has a lower damage threshold but a higher refractive index.

For the sub-relativistic experiments in the ACHIP collaboration we mostly use  $\lambda_0 = 2 \mu\text{m}$  femtosecond laser pulses, generated by Optical Parametric Amplifiers (OPA) or by novel Tm or Ho-Tm fiber laser amplifiers currently under development. The electron source needs to provide ultra-low emittance, particularly at low energy. For a lossless sub-100-keV injection into a DLA operating at  $2 \mu\text{m}$ , geometric emittances smaller than 0.1 nm are required [17]. Different emitters are available to produce these low emittances, e.g. [18], or see [19] for an overview. At these emittances, the achievable charge is quite small at the moment. We hope to achieve higher average charge in the future by increasing both the repetition rate and the single microbunch charge.

## DIFFERENT MEANS OF LASER COUPLING

All the DLA experiments performed so far have used free space laser coupling. At longer interaction length, it is necessary to provide symmetric fields, such that there is no coherent deflection force. There are different means to obtain symmetric fields in the acceleration channel, the simplest is to illuminate the structure symmetrically from both sides with equal phase and polarization. If this is impractical from the optics point of view, the fields of a single side drive laser can also be symmetrized by using a Bragg mirror on the chip [20, 21].

The laser fluence on each DLA cell can be reduced by shortening the pulse and tilting the pulse front [22, 23], such that it remains synchronous with a few electron bunches over a distance (or duration) much longer than a single DLA cell is illuminated, see Fig. 3. Practically, the pulse front tilt

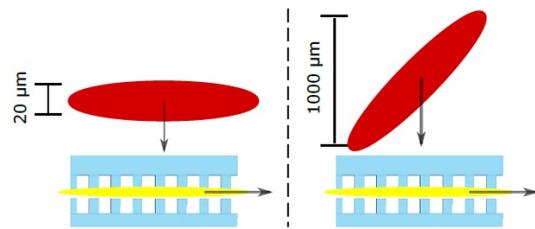


Figure 3: Flat vs. tilted laser pulse. At roughly the same pulse length for each grating cell, the interaction length is significantly increased. Picture adapted from [22].

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

can be achieved, for example, by a prism or by a reflection grating with unequal incidence and reflection angles [22,23].

Another option to increase the interaction length in the accelerator structure is an on-chip waveguide system [24], see Fig. 4. This supplies different parts of the accelerator struc-

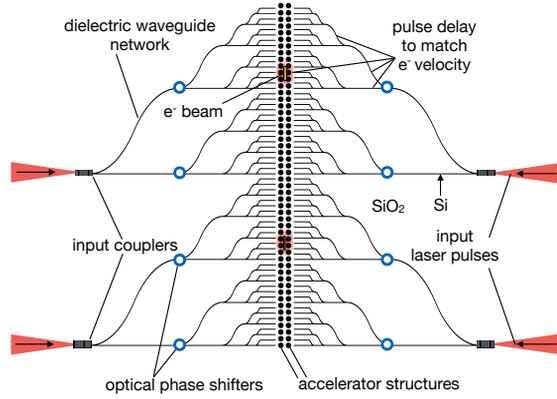


Figure 4: On-chip waveguide laser power delivery system. Picture adapted from [24].

ture with the appropriate phase and group delay. Moreover, a different, higher damage threshold and lower refractive index material can be used to convey higher laser fluence in the waveguides, which can be split in to many waveguides before coupling to the high refractive index accelerator structure. Using this technique, it is possible to produce a similar illumination pattern as would be obtained by pulse front tilt in free space.

## FIELD COMPUTATION FOR A SINGLE DLA CELL

The laser field computation of a single DLA cell using periodic boundary conditions is not a challenge, since its electrical length  $L/\lambda$  is on the order of one. It can be simulated by various techniques such as Finite Difference / Finite Integration Time Domain (FD/FI TD) codes [25, 26], Finite Difference Frequency Domain (FDFD) codes [27], or Finite Element Frequency Domain (FEFD) codes [26, 28]. These can be combined with various optimization techniques, in order to find structures with highest gradient, lowest field inside the material, or highest bandwidth. Of course, these optimization goals compete, such that an optimum can only be found in the sense of a Pareto-front. Simple DLA structures can also be designed from a physical point of view, such that maximum electric field modulation at the synchronous harmonic is achieved, which results in Bragg cavity structures [28, 29]. A more mathematical approach is rather to use adjoint methods to perform large-scale, gradient-based optimization of the full permittivity distribution [30]. However, these methods tend to generate non-intuitive device geometries and sometimes require additional constraints to create fabricable structures. Moreover, adjoint methods have also been used for other parts of the integrated DLA, such as grating couplers [31].

## BEAM DYNAMICS SIMULATIONS IN DLATRACK6D

We will summarize and slightly add to DLATRACK6D, the one kick per cell tracking approach originally conceived in [29]. The kicks are sufficiently described by one complex coefficient per DLA cell, where the longitudinal and transverse dependencies are derived analytically. Although the derivation holds true only for strict periodicity, small deviations can be accepted within reasonable error. Also fringe fields are not included, even in practice they should be reduced as much as possible. The effect of fringe fields is however strongly dependent on the quality factor of the structure, usually determined by the available bandwidth.

Starting from the longitudinal energy gain, the kicks in all directions are computed and then used for symplectic tracking. The energy gain of an electron with charge  $q = -e$  can be written as function of the time domain electric field  $E_z$ , its Frequency Domain (FD) phasor  $\underline{E}_z$ , or by means of spatial Fourier series in periodic DLA structures

$$\Delta W(x, y, s) = q \int_{-\lambda_{gz}/2}^{\lambda_{gz}/2} E_z(x, y, z; t = (z + s)/v) dz \quad (2)$$

$$= q \lambda_{gz} \text{Re} \left\{ e^{2\pi i \frac{s}{\beta \lambda_0}} \underline{e}_m(x, y) \right\}. \quad (3)$$

Here the spatial Fourier coefficient is computed as

$$\underline{e}_m(x, y) = \frac{1}{\lambda_{gz}} \int_{-\lambda_{gz}/2}^{\lambda_{gz}/2} \underline{E}_z(x, y, z) e^{im \frac{2\pi}{\lambda_{gz}} z} dz. \quad (4)$$

The above relation holds only if the Wideroe condition  $\lambda_{gz} = m\beta\lambda_0$  is fulfilled, where  $\lambda_{gz}$  is the grating period and  $\beta$  is the velocity in units of  $c$ . In the following we will restrict the arbitrary integer spatial harmonic  $m$  to  $m = 1$ , which usually has the strongest amplitude  $\underline{e}_1$ .

The transverse kicks can be obtained by exploiting the known transverse dependency of  $e_1(x, y)$  on the transverse coordinates. From the Panofsky-Wenzel theorem [32], conveniently written as

$$\nabla' \times \Delta \vec{p}(x, y, s) = 0, \quad (5)$$

where the relative gradient is  $\nabla' = (\partial_x, \partial_y, -\partial_s)$ , we obtain under the synchronicity condition

$$\Delta \vec{p}_\perp(x, y, s) = -\frac{\lambda_{gz}^2}{2\pi} q \frac{1}{\beta c} \text{Im} \left\{ e^{2\pi i \frac{s}{\beta \lambda_0}} \nabla_\perp \underline{e}_1(x, y) \right\}. \quad (6)$$

For a symmetric laser illumination the transverse dependencies can be written as

$$\underline{e}_1(x, y) = \underline{e}_1(0, 0) \cosh(ik_y y) e^{ik_x x}, \quad (7)$$

where  $k_x$  includes the option of a tilt of the grating or the laser incidence,  $k_z = 2\pi/(\beta\lambda_0)$ , and  $k_y = \pm\sqrt{(2\pi/\lambda_0)^2 - k_z^2 - k_x^2}$ . For a grating tilt angle  $\alpha$  (see Fig. 5) we obtain [29]  $k_x = k_z \tan(\alpha)$  and for a laser tilt angle  $\vartheta$  (keeping the polarization parallel to the electron beam) we obtain  $k_x = k_z \sin(\vartheta)$ . After some manipulations, we finally

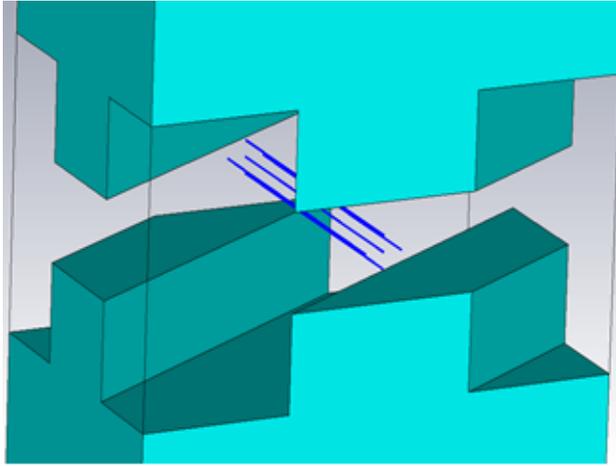


Figure 5: Unit cell of a tilted grating (tilt angle  $\alpha$ ). The blue lines indicate possible electron trajectories, the laser is incident from top and bottom, with polarization in electron beam direction and possible tilt angle  $\vartheta$ .

obtain the kicks [29]

$$\Delta x' = -\frac{q\lambda_0}{p_{z0}c} \tan(\alpha) \cosh(ik_y y) \text{Re} \{ \underline{e}_1 e^{i\varphi + ik_x x} \} \quad (8a)$$

$$\Delta y' = \frac{-ik_y \lambda_0^2 q\beta}{2\pi p_{z0}c} \sinh(ik_y y) \text{Im} \{ \underline{e}_1 e^{i\varphi + ik_x x} \} \quad (8b)$$

$$\Delta\delta = \frac{q\lambda_{gz}}{\gamma m_e c^2} \text{Re} \left\{ \underline{e}_1 \left( \cosh(ik_y y) e^{i\varphi + ik_x x} - e^{i\varphi_s} \right) \right\}. \quad (8c)$$

In case of an anti-symmetric illumination, i.e. a  $\pi$  phase shift between the two laser beams, the transverse dependence is

$$\underline{e}_1(x, y) = \underline{e}_1(0, 0) \sinh(ik_y y) e^{ik_x x} \quad (9)$$

and the kicks are accordingly

$$\Delta x' = -\frac{q\lambda_0}{p_{z0}c} \tan(\alpha) \sinh(ik_y y) \text{Re} \{ \underline{e}_1 e^{i\varphi + ik_x x} \} \quad (10a)$$

$$\Delta y' = \frac{-ik_y \lambda_0^2 q\beta}{2\pi p_{z0}c} \cosh(ik_y y) \text{Im} \{ \underline{e}_1 e^{i\varphi + ik_x x} \} \quad (10b)$$

$$\Delta\delta = \frac{q\lambda_{gz}}{\gamma m_e c^2} \text{Re} \{ \underline{e}_1 \sinh(ik_y y) e^{i\varphi + ik_x x} \}. \quad (10c)$$

Note that in the sinh-mode, the energy gain of the synchronous particle is always zero, since the longitudinal electric field vanishes in center of the channel. The sinh-mode can be used as a diagnostic, in order to convert a temporal profile into an angle distribution profile [33], which is usually referred to as beam streaking.

The symplectic one-kick-per-cell tracking is independent of the realization of the kick functions and reads

$$\begin{pmatrix} x \\ x' \\ y \\ y' \\ \varphi \\ \delta \end{pmatrix}^{(n+1)} = \begin{pmatrix} x \\ Ax' + \Delta x' \\ y \\ Ay' + \Delta y' \\ \varphi \\ \delta + \Delta\delta(\varphi_s) \end{pmatrix}^{(n)} + \begin{pmatrix} \lambda_{gz} x' \\ 0 \\ \lambda_{gz} y' \\ 0 \\ -\frac{2\pi}{\beta^2 \gamma^2} \delta \\ 0 \end{pmatrix}^{(n+1)}, \quad (11)$$

with the tracking variables in paraxial approximation

$$\begin{aligned} x' &= \frac{p_x}{p_{z0}}, & \Delta x' &= \frac{\Delta p_x(x, y, \varphi)}{p_{z0}}, \\ y' &= \frac{p_y}{p_{z0}}, & \Delta y' &= \frac{\Delta p_y(x, y, \varphi)}{p_{z0}}, \\ \varphi &= 2\pi \frac{s}{\lambda_{gz}}, & \delta &= \frac{W - W_0}{W_0}, \\ \Delta\delta &= \frac{\Delta W(x, y, \varphi) - \Delta W(0, 0, \varphi_s)}{W_0}, \end{aligned} \quad (12)$$

where  $W_0 = \gamma m_e c^2$  and  $p_{z0} = \beta \gamma m_e c$ . The adiabatic damping of the transverse emittance is described by

$$A^{(n)} = \frac{(\beta\gamma)^{(n+1)}}{(\beta\gamma)^{(n)}} = 1 + \left[ \frac{\lambda_0 q \text{Re} \{ e^{i\varphi_s} \underline{e}_1 \}}{\beta \gamma m_e c^2} \right]^{(n)}. \quad (13)$$

While keeping the synchronicity condition by appropriately chirping the structure, the acceleration ramp can be written as

$$W(N) = W(0) + q \sum_{n=1}^N \lambda_g^{(n)} \text{Re} \{ e_1^{(n)} e^{i\varphi_s^{(n)}} \}. \quad (14)$$

We note that chirping the structure length while maintaining constant phase  $\arg(\underline{e}_1)$  is possible by correcting the phase drift with another parameter in the structure design [20].

DLAtrack6D is written in Matlab [34], it is based on a phase space structure that allows for vectorized updates in each DLA cell. The code, together with a brief manual, will be made available to the community soon.

## ELECTRON BEAM FOCUSING

Optical near field accelerators cannot rely on magnetic focusing only, since the small scale of the near field requires sub-micron beam sizes which in turn would require magnetic field strengths unachievable by conventional quadrupole magnets [35]. Thus a laser-based focusing scheme is required to make DLA scalable.

Two different options have been proposed for focusing with the phase dependent transverse laser fields. In 2012 the group at UCLA has proposed spatial harmonic focusing [36]. Stability of the electron beam could be predicted by means of retracting ponderomotive forces due to non-synchronous harmonics, while the synchronous harmonic serves for acceleration. However, the beam envelope at given emittance could not be determined more accurately than in the smooth approximation. Moreover, the focusing harmonic needs to be quite strong (much stronger than the accelerating harmonic), which puts a constraint on the choice of materials and pulse length due to the damage threshold fluence. In particular, this scheme has been implemented with SiO<sub>2</sub> structures at relativistic energies [37].

Recently, Niedermayer et al. have proposed Alternating Phase Focusing for DLA [17]. Here, we work only with a single spatial harmonic, i.e.  $\underline{e}_1$ , but its phase can be changed by means of fractional period drift elements. The hereby generated longitudinal/transversal alternating focusing gradients

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

can be integrated in the Courant-Snyder sense. Thus a scalable scheme is obtained, where about half (dependent on the synchronous phase) of the synchronous harmonic is translated to acceleration gradient. The required pre-bunching on the optical scale can be obtained with the same scheme, see [17]. Due to the exact integration of the lattice (and thus a precise determination of the beam envelope) and the efficient translation of incident field to acceleration gradient, this scheme is particularly suited for sub-relativistic DLAs working with high refractive index Silicon structures that have a rather low damage threshold. An outline of such a structure is depicted in Fig. 6.

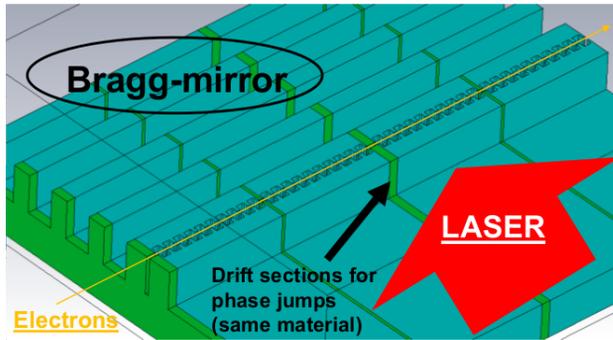


Figure 6: Outline of an APF DLA structure, picture adapted from [29].

## FULL SCALE SIMULATION TECHNIQUES

For full scale tracking and PIC simulations we mostly use the codes CST Studio Suite [26] and VSim [25]. A full scale PIC simulation is however numerically quite heavy, thus we prefer to do this only for finished designs and rely on DLATRACK6D for design studies. In CST, we have the convenient option to calculate the fields in either in TD or FD and store them as frequency domain phasors, i.e. one complex number per mesh edge. Note that storing in FD does not necessary mean computation in FD; in fact, FD simulations become intractable as the simulation size becomes large, so

instead, TD computation can be performed along with an on-the-fly Fourier Transform.

We use VSim in cases where large scale computing is required. Moreover, including a pulse front tilted laser beam is rather involved in CST at the moment, thus we did this in VSim. The capability of VSim for electron energy loss in materials, to model experiments where part of the electron beam clips the structure, was also added and showed good agreement with experiments [5]. In the near future, we plan to conduct high performance PIC simulations to assess the effects of fringe fields and imperfect field flatness, wake fields, and radiation emission using NERSC cluster time awarded to the ACHIP collaboration partner TechX.

## ONGOING EXPERIMENTS

Simulations are conducted for the design and the evaluation of different ongoing DLA experiments. The simplest one uses the intrinsic phase focusing properties of a DLA structure [38]. In this experiment, the electrons are injected at random phase, which means that they are either focused or defocused, i.e. a cross-shape is formed in transverse phase space (see also [20]). An aperture lets only the focused electrons pass, the defocused ones are lost. This transmission is however still higher than in the absence of the laser, i.e. when there is no focusing at all. This experiment can also be run in deflection (sinh) mode, i.e. the electrons are deflected to the left or right, dependent on their phase. Strong deflection will reduce the transmission through the aperture accordingly, which is readily measurable.

Another ongoing focusing experiment is APF with a Bragg mirror, as outlined in Fig. 7. As discussed in detail in [17], this APF focusing channel transports particles at all phases. A pre-bunching is not required for transport only, however it would be required for acceleration. In the bottom of the figure, the longitudinal electric field is plotted. Different options for filling the half cell spaces for phase jumps have been simulated. In the end, it turned out that the spacers are only helpful in the beginning and the end of the structure and within it is sufficient to leave half a cell empty.

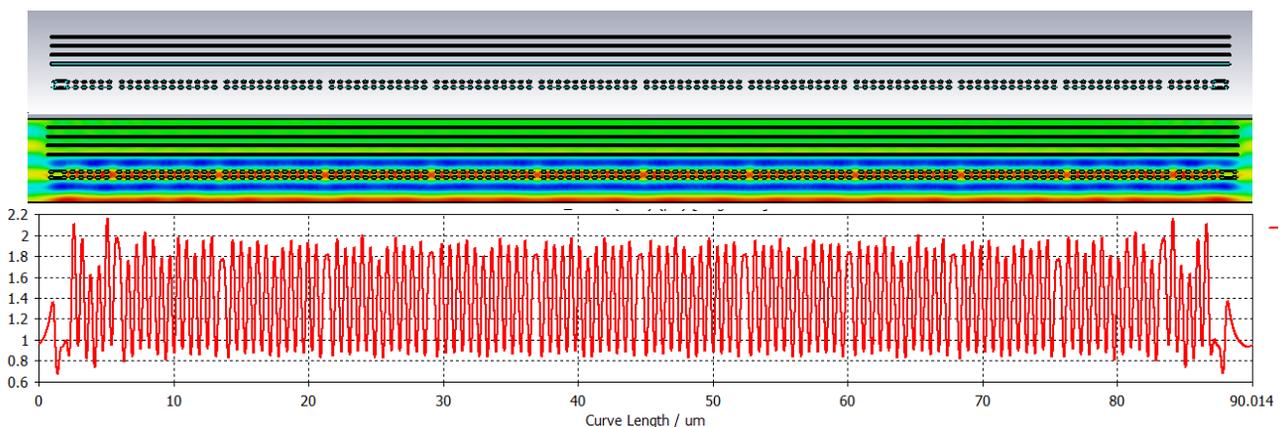


Figure 7: Alternating Phase Focusing single laser beam transport structure (top), normalized longitudinal electric field phasor magnitude with laser from bottom (center) and on axis field flatness plot (bottom), obtained by CST MWS in TD.

Moreover, another crucial optimization is to tune the Bragg mirror distance such that the fields in the channel become symmetric. After the optimized fields have been determined, electron tracking can be done both in DLAttrack6D and full PIC codes. Figure 8 shows how the fraction of electrons transported through the structure is dependent on the laser field strength (DLAttrack6D).

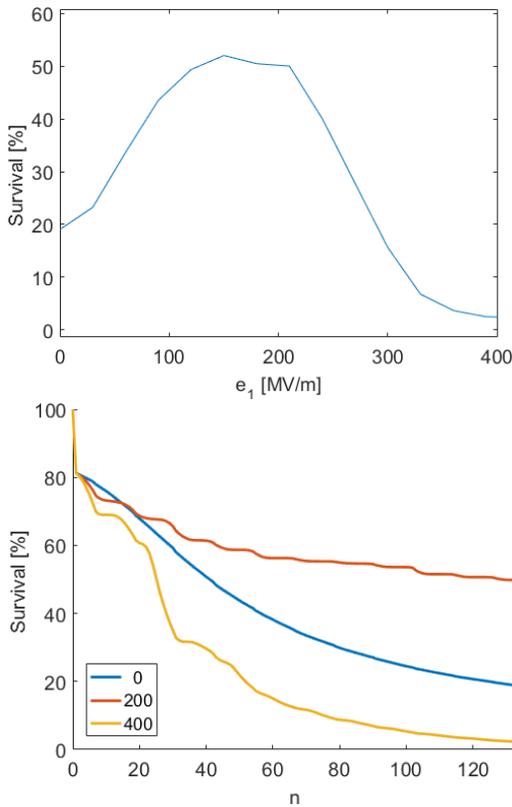


Figure 8: Transmission rate as function of laser field strength (top) and electron loss as function of DLA cell index for  $e_1 = (0, 200, 400)$  MV/m laser field strength (bottom).

The minimal beta-function is obtained at roughly 200 MV/m, where maximum transmission occurs. Beyond that point, the beam is first over-focused, and eventually leaves the area of stability. The reason for the losses at the matched  $e_1$  is the geometric emittance of 0.3nm (Gaussian distribution initially), which is larger than the acceptance of the structure. Once saturated at about 50%, the electrons can be transported over an arbitrary distance, which is limited only by the defocusing in the direction of the pillar height. Roughly the same results are obtained by tracking in the CST PIC solver.

In succession to [33, 39], we plan a two stage buncher-accelerator or buncher-streaker experiment, see Fig. 9. Here we use dual drive from two stages, with independent phase control and independent amplitude control of lasers 1 and 2. The first DLA stage is normally run in cosh-mode and serves as a buncher. The second stage can be either run in cosh-mode for acceleration, or in sinh-mode for streaking. Replacing the first (buncher) structure by an APF-type buncher, see [17], will allow us to obtain short bunches with



Figure 9: Combined buncher and accelerator/streaker DLA.

low energy spread. This gives the opportunity to observe both coherent streaking and coherent acceleration at the same setup on a spectrometer screen.

Additionally to the low energy experiment, there are also high energy experiments planned. Most prominently, we will use the 3.2GeV beam at SwissFEL at PSI to inject into a DLA. Due to the extremely small geometric emittance at such high energy, this injection will be almost lossless. Additionally, at such energy the deflection and acceleration defocusing is rather small. Therefore, we are restricted only by the conventional electron optics (Rayleigh length) and the available laser pulse energy, which can be cast in a tilted pulse. Details of the outlined experiments can be found in [40,41].

Moreover, relativistic energy experiments are also outlined at the SINBAD facility at DESY, where an inverse FEL undulator together with a chicane will be employed for optical-scale bunching of the beam before it is injected into the DLA. Driving the DLA with the same laser as is used for seeding the inverse FEL allows for the precise phase control required for coherent acceleration [42].

The group at UCLA aims for a 2-cm long DLA experiment at about 5 MeV injection energy at the Pegasus facility [43]. Challenges are that focusing of the beam is still required and also a slight chirp needs to be imprinted to account for the slightly sub-relativistic velocity. Creating both the focusing harmonics and the chirp is planned to be achieved by a strictly periodic SiO<sub>2</sub> grating fed by a tilted laser pulse that is modulated by a Spatial Light Modulator, see e.g. [44].

## CURRENT STATUS AND OUTLOOK

We are now able to perform start-to-end simplified simulations of larger DLA chips with DLAttrack6D. For full scale 3D PIC or tracking simulations a cluster computer is required. The experiments performed at the moment can still be simulated well in 2D by available PIC codes. However, it is expected that the structure lengths will soon significantly increase.

One option for efficient large scale PIC for DLA would be a moving window code, which discretizes only the co-moving environment of a few electron micro-bunches. The rest of the structure contains neither electrons nor laser energy, since we strongly restrict the interaction region by applying the pulse front tilt or other means of selective synchronized illumination.

The charge we accelerate in current DLA experiments is mostly rather low. However, at particular high energy experiments as e.g. at PSI, the entire beam is put through the small aperture of the DLA. We expect to see wake field effects here for the first time. Simulations of wake field effects are already in place [45]. We outline to integrate

linear and non-linear wake kicks from precomputed wake functions into DLATRACK6D as well. With this we will be able to properly predict the strength of beam loading effects and longitudinal and transverse beam instabilities in longer DLA structures.

## ACKNOWLEDGMENT

ACHIP is funded by the Gordon and Betty Moore Foundation (Grant No. GBMF4744 to Stanford). U.N. acknowledges also the funding by the German Federal Ministry of Education and Research (Grant No. FKZ:05K16RDB). B.C. (TechX) acknowledges usage of NERSC, a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

## REFERENCES

- [1] ACHIP website: <https://achip.stanford.edu>
- [2] K. Shimoda, *Applied Optics*, vol. 1, pp. 33, 1962.
- [3] A. Lohmann, IBM Technical Note, 169, 1962.
- [4] E. A. Peralta *et al.*, “Demonstration of electron acceleration in a laser-driven dielectric microstructure”, *Nature*, vol. 503, pp. 91, 2013. doi:10.1038/nature12664
- [5] K. P. Wootton *et al.*, *Opt. Lett.*, vol. 41, pp. 2696, 2016.
- [6] J. Breuer and P. Hommelhoff, *Physical Review Letters*, vol. 111, pp. 134803, 2013.
- [7] K. J. Leedle, A. Ceballos, H. Deng, O. Solgaard, R. Fabian Pease, R. L. Byer, and J. S. Harris, *Optics Letters*, vol. 40, pp. 4344, 2015.
- [8] M. Kozak *et al.*, *NIM A*, vol. 865, pp. 84–86, 2016.
- [9] R. J. England *et al.*, *Reviews of Modern Physics*, vol. 86, pp. 34, 2014.
- [10] J. McNeur *et al.*, *Optica*, vol. 5, issue 6, pp. 687-690, 2018. doi:10.1364/OPTICA.5.000687
- [11] T. Plettner and R. L. Byer, *Phys. Rev. STAccel. Beam*, vol. 11, pp. 030704, 2008.
- [12] T. Plettner and R. L. Byer, *Nucl. Instrum. Methods Phys. Res. A*, vol. 593, pp. 63, 2008.
- [13] T. Plettner, R. Byer, C. McGuinness, and P. Hommelhoff, *Phys. Rev. ST Accel. Beams*, vol. 12, pp. 101302, 2009.
- [14] K. Soong, R. L. Byer, E. R. Colby, R. J. England, and E. A. Peralta, *AIP Conference Proceedings*, vol. 1507, pp. 511, 2012.
- [15] P. Pronko *et al.*, *Physical Review B*, vol. 58, pp. 2387, 1998.
- [16] D. Cesar *et al.*, *Nat. Comm. Phys.*, vol. 1, pp. 46, 2018.
- [17] U. Niedermayer, T. Egenolf, O. Boine-Frankenheim, and P. Hommelhoff, accepted for publication in *Phys. Rev. Lett.*, 2018. arXiv:1806.07287
- [18] D. Ehberger, J. Hammer, M. Eisele, M. Krueger, J. Noe, A. Hoge, and P. Hommelhoff, *Physical Review Letters*, vol. 114, pp. 1, 2015.

- [19] A. Feist *et al.*, *Ultramicroscopy*, vol. 176, pp. 63, 2017.
- [20] U. Niedermayer, O. Boine-Frankenheim, and T. Egenolf, *Journal of Physics, Conference Series*, vol. 874, 2017.
- [21] P. Yousefi *et al.* <https://arxiv.org/abs/1801.05234>
- [22] D. Cesar, J. Maxson, P. Musumeci, X. Shen, R. J. England, and K. P. Wootton, SLAC-PUB-17180, 2018.
- [23] M. Kozak *et al.*, *Journal of Applied Physics*, vol. 124, 2018. doi:10.1063/1.5032093
- [24] T. W. Hughes *et al.*, *Physical Review Applied*, vol. 9, pp. 54017, 2018.
- [25] VSim, [www.txcorp.com](http://www.txcorp.com)
- [26] CST Studio Suite 2018, [www.cst.com](http://www.cst.com)
- [27] W. Shin and S. Fan, *Optics Express*, vol. 21, no. 19, pp. 22578–22595, 2013.
- [28] T. Egenolf, O. Boine-Frankenheim, and U. Niedermayer, *Journal of Physics: Conference Series*, vol. 874, 2017. doi:10.1088/1742-6596/874/1/012040
- [29] U. Niedermayer, T. Egenolf, and O. Boine-Frankenheim, *Phys. Rev. AB*, vol. 20, pp. 111302, 2017.
- [30] T. Hughes *et al.*, *Optics Express*, vol. 25, Issue 13, 2017. doi:10.1364/OE.25.015414
- [31] N. Sapra *et al.*, <https://arxiv.org/abs/1808.07630>
- [32] W. K. H. Panofsky and W. A. Wenzel, *Review of Scientific Instruments*, vol. 27, pp. 31, 1956.
- [33] K. J. Leedle, D. S. Black, Y. Miao, K. E. Urbanek, A. Ceballos, H. Deng, J. S. Harris, O. Solgaard, and R. L. Byer, *Optics Letters*, vol. 43, pp. 2181, 2018.
- [34] MathWorks, Matlab, 2016.
- [35] A. Ody, P. Musumeci, J. Maxson, D. Cesar, R. J. England, and K. P. Wootton, *NIM A*, vol. 865, pp. 75, 2017.
- [36] B. Naranjo, A. Valloni, S. Putterman, and J. B. Rosenzweig, *Physical Review Letters*, vol. 109, pp. 1, 2012.
- [37] D. Cesar, J. Maxson, P. Musumeci, X. Shen, R. J. England, K. P. Wootton, and S. Tan, arXiv preprint accepted for publication in *Optics Express*, 1804.00634, 2018.
- [38] D. Black, K. Leedle, Y. Miao, U. Niedermayer, R.L.Byer, and O. Solgaard, unpublished.
- [39] M. Kozak *et al.*, *Nat. Comm.*, vol. 8, pp. 14342, 2017.
- [40] E. Prat *et al.*, *NIM A*, vol. 865, pp. 87–90, 2017.
- [41] E. Ferrari *et al.*, *NIM A*, vol. 907, pp. 244-247, 2018.
- [42] F. Mayet *et al.*, accepted for publication in *NIM-A*, 2018. doi:10.1016/j.nima.2018.01.088
- [43] D. Cesar, P. Musumeci and R. J. England, in *Proc. of AAC2018*, IEEE Conf Series, to appear.
- [44] ThorLabs EXULUS-HD1 Spatial Light Modulator [www.thorlabs.com](http://www.thorlabs.com)
- [45] T. Egenolf, U. Niedermayer, and O. Boine-Frankenheim in *Proc. of AAC2018*, IEEE Conf Series, to appear.

# SPIN DYNAMICS IN MODERN ELECTRON STORAGE RINGS: COMPUTATIONAL AND THEORETICAL ASPECTS

Klaus Heinemann\*, Oleksii Beznosov, James A. Ellison, UNM, Albuquerque, NM, USA  
Desmond P. Barber<sup>1</sup>, DESY, Hamburg, Germany  
Daniel Appelö, University of Colorado, Boulder, CO, USA  
<sup>1</sup> also at UNM, Albuquerque, NM, USA

## INTRODUCTION

In this presentation we describe some numerical and analytical results from our work on the spin polarization in high energy electron storage rings aimed towards the proposed Future Circular Collider (FCC-ee) and the proposed Circular Electron Positron Collider (CEPC). Photon emission in synchrotron radiation imparts a stochastic element (“noise”) into particle motion and there are also damping effects. However, instead of considering single particles it is often convenient to model the stochastic photon emission as a Gaussian white noise process and to then study the evolution of the particle density in phase space with a Fokker-Planck equation.

The noise in trajectories together with the spin-orbit coupling embodied in the Thomas-BMT equation of spin precession [1], can cause spin diffusion and thus depolarization. On the other hand photon emission can lead to a build up of polarization via spin flip. This is the Sokolov-Ternov process [2]. The attainable polarization is the outcome of the balance of the two effects.

So far, analytical estimates of the attainable polarization have been based on the so-called *Derbenev-Kondratenko formulas* [3,4]. In analogy with studies of the trajectories of single particles, that approach leans towards the study of single spins and relies in part on plausible assumptions grounded in deep physical intuition. However, just as with particle motion it would be convenient to have a treatment of the Fokker-Planck (F-P) kind and thereby minimize the reliance on assumptions. But the polarization at a point in phase space cannot be handled in that way since polarization is not a density. Nevertheless a density is available, namely the density in phase space of the spin angular momentum and with this there is a generalization of the F-P equation which we call the Bloch equation. We use that name to reflect the analogy with equations for magnetization in condensed matter [5]. In fact the Bloch equation works with the so-called *polarization density*. This is proportional to the spin angular momentum density per particle in phase space. With this we can calculate the polarization vector of the bunch.

Thus we study the initial value problem of what we call the full Bloch equation (FBE). The FBE takes into account non spin-flip and spin-flip effects due to synchrotron radiation including the spin-diffusion effects and the Sokolov-Ternov effect with its Baier-Katkov generalization. The FBE was introduced by Derbenev and Kondratenko in 1975 [6] as a generalization to the whole phase space (with its noisy tra-

jectories) of the Baier-Katkov-Strakhovenko (BKS) equation which just describes the evolution of polarization by spin flip along a single trajectory [7]. The FBE is a system of three F-P equations coupled by a Thomas-BMT term and the BKS terms but uncoupled within the F-P terms. By neglecting the spin flip terms in the FBE we obtain what we call the reduced Bloch equation (RBE). The RBE approximation is sufficient for computing the physically interesting depolarization time and it shares the terms with the FBE that are challenging to discretize. Thus, here we only consider the discretization of the RBE.

Our approach has three parts. First we approximate the RBE analytically using the method of averaging, resulting in an average RBE which allows us to use large time steps. The minimum length of the time interval of interest is of the order of the orbital damping times. Secondly, the phase space coordinates of the average RBE come in  $d = \{1, 2, 3\}$  pairs of polar-radial coordinates that we discretize using a Fourier-Chebyshev pseudospectral approach. The averaging decouples the parabolic and mode coupling terms allowing for a parallel implementation with only local communication. Thirdly, we further exploit the decoupling by evolving the resulting system of ODEs by an implicit-explicit (ARK) method. Parabolic operators are treated implicitly and can be inverted rapidly due to the decoupling. If each of the  $d$  angle variables is discretized on a grid of  $M$  grid points and if each of the  $d$  radial variables is discretized on a grid of  $N$  grid points then the total number of operations for each time step scales, to leading order, as  $O(N^{dq} M^d)$  where  $1 \leq q \leq 3$ , depending on the algorithms used for the linear solve. For Gaussian elimination  $q = 3$ . Details and more results have been presented in this meeting by O.Beznosov, see [8].

The main issues for very high energy rings like the FCC-ee and CEPC are: (i) Can one get polarization, (ii) what are the theoretical limits of the polarization? We believe that the FBE offers a more complete starting point for very high energy rings than the Derbenev-Kondratenko formulas. See [9] for a recent review of polarization history and phenomenology.

## RBE IN LAB FRAME

In a semiclassical probabilistic description of an electron bunch the spin-orbit dynamics is described by the *spin-1/2 Wigner function* (also called the *Stratonovich function*)  $\rho$

\* Corresponding author: heineman@math.unm.edu

written as

$$\rho(t, z) = \frac{1}{2} (f(t, z)I_{2 \times 2} + \vec{\sigma} \cdot \vec{\eta}(t, z)), \quad (1)$$

where  $f$  is the classical phase-space density normalized by  $\int f(t, z)dz = 1$  and  $\vec{\eta}$  is the polarization density of the bunch and thus proportional to the spin angular momentum density. Here  $z = (r, p)$  where  $r$  and  $p$  are the position and momentum vectors of the phase space and  $t$  is the time. Also,  $\vec{\sigma}$  is the vector of the three Pauli matrices. Thus  $f = Tr[\rho]$  and  $\vec{\eta} = Tr[\rho\vec{\sigma}]$ . The polarization vector  $\vec{P}(t)$  of the bunch is  $\vec{P}(t) = \int \vec{\eta}(t, z)dz$ . Here and in the following we use arrows on spin-related quantities and no arrows on other quantities. Moreover the spin-related quantities will be represented by column matrices. When the particle motion is governed just by a Hamiltonian, as in the case of protons, the phase-space density is conserved along a trajectory so that the polarization density obeys the Thomas-BMT equation along each trajectory. However, if the particles are subject to noise and damping due to synchrotron radiation, the evolution of the density of particles in phase space is more complicated. But as advertised above it can be handled with a F-P formalism. Then by neglecting collective effects and after several other approximations,  $\rho$  evolves via

$$\partial_t f = L_{FP}(t, z)f, \quad (2)$$

$$\begin{aligned} \partial_t \vec{\eta} = & L_{FP}(t, z)\vec{\eta} + \Omega(t, z)\vec{\eta} + G(t, z)\vec{\eta} \\ & + \vec{g}(t, z)f + \vec{L}(t, z)f \end{aligned} \quad (3)$$

where (2) is the F-P equation for the orbital density and (3) is the FBE mentioned above, both in the lab frame, i.e., in cartesian coordinates. The F-P operator  $L_{FP}$  is the linear second-order partial differential operator commonly used for electron synchrotrons and storage rings [10, Section 2.5.4], [11, 12]. The skew-symmetric matrix  $\Omega(t, z)$  in the FBE takes into account the Thomas-BMT spin-precession effect. The terms  $G\vec{\eta}$ ,  $\vec{g}f$  and  $\vec{L}f$  take into account spin flips due to synchrotron radiation. In particular they include the Sokolov-Ternov effect and its Baier-Katkov correction the latter belonging to  $G\vec{\eta}$ . As usual, since it is minuscule compared to all other forces, the Stern-Gerlach effect from the spin onto the orbit is neglected in (2). The explicit forms of  $L_{FP}$ ,  $\Omega$ ,  $G$ ,  $\vec{g}$  and  $\vec{L}$  are given in [6].

If we neglect the spin flip terms in the FBE then (3) simplifies to

$$\partial_t \vec{\eta} = L_{FP}(t, z)\vec{\eta} + \Omega(t, z)\vec{\eta} \quad (4)$$

The RBE (4) just takes care of spin diffusion due to the orbital motion.

The Equations (2) and (3) can be derived from quantum electrodynamics, followed by making the semiclassical approximation of the Foldy-Wouthuysen transformation of the Dirac Hamiltonian and finally by making a Markov approximation [13]. We stress however, that the RBE (4) can be derived purely classically as in [14]. In fact, we show again how to do this at the end of the next section.

## RBE IN THE BEAM FRAME

In the beam frame, i.e., in accelerator coordinates, the RBE (4) becomes

$$\partial_\theta \vec{\eta}_Y = (L_Y + L_{Y,TBMT})\vec{\eta}_Y \quad (5)$$

where  $\theta$  is the accelerator azimuth,

$$L_Y = - \sum_{j=1}^6 \partial_{y_j} \left( \mathcal{A}(\theta)y \right)_j + \frac{1}{2} \omega_Y(\theta) \delta_{y_6}^2,$$

$$L_{Y,TBMT} \vec{\eta}_Y = \Omega_Y(\theta, y) \vec{\eta}_Y$$

and where  $\mathcal{A}(\theta)$  is a  $6 \times 6$  matrix encapsulating radiationless motion and the deterministic effects of synchrotron radiation. Also  $\Omega_Y(\theta, y)$  is the Thomas-BMT term and it is a skew-symmetric  $3 \times 3$  matrix linear in  $y$  and  $\omega_Y(\theta)$  is the magnitude of the noise. Note that  $\mathcal{A}(\theta)$ ,  $\Omega_Y(\theta, y)$  and  $\omega_Y(\theta)$  are  $2\pi$ -periodic in  $\theta$ . Given the beam frame polarization density  $\vec{\eta}_Y$  the beam frame polarization vector  $\vec{P}(\theta)$  of the bunch at azimuth  $\theta$  is

$$\vec{P}(\theta) = \int dy \vec{\eta}_Y(\theta, y) \quad (6)$$

Our central computational focus in this paper is the RBE (5) with  $\vec{P}(\theta)$  being a quantity of interest. To proceed with this it is important that (5) has an underlying system of Langevin equations and thus an underlying F-P equation. In fact the system of Langevin equations is

$$Y' = \mathcal{A}(\theta)Y + \sqrt{\omega_Y(\theta)} e_6 \xi(\theta), \quad (7)$$

$$\vec{S}' = \Omega_Y(\theta, Y) \vec{S} \quad (8)$$

where  $\xi$  is a version of the white noise process,  $e_6 = (0, 0, 0, 0, 0, 1)^T$  and  $\vec{S}$  is the single-particle spin expectation value. Note that (7) can be written as the Ito stochastic differential equation:  $dY = \mathcal{A}(\theta)Y d\theta + \sqrt{\omega_Y(\theta)} e_6 dW$  which is linear in the narrow sense and thus defines a Gaussian process  $Y$  if  $Y(0)$  is Gaussian. In principle (5) could be obtained by transforming (4) and the coefficients  $\mathcal{A}$ ,  $\Omega_Y$  and  $\omega$  from the lab frame to the beam frame, However this is not necessary since (7) and (8) and the  $\mathcal{A}$ ,  $\Omega_Y$  and  $\omega$  can be found in virtually every exposition on spin in high-energy electron storage rings, e.g., [15]. Note that these expositions make some approximations. We use [15] which involves linearizing w.r.t.  $y$  as can be seen in (7) and (8). For (5) see also [14].

The F-P equation for the Gaussian process  $Y$  is

$$\partial_\theta \mathcal{P}_Y = L_Y \mathcal{P}_Y \quad (9)$$

For getting (9) from (7) see [16–18]. With (7) and (8) the evolution equation for the spin-orbit joint probability density  $\mathcal{P}_{YS} = \mathcal{P}_{YS}(\theta, y, \vec{s})$  is the following F-P equation:

$$\partial_\theta \mathcal{P}_{YS} = L_Y \mathcal{P}_{YS} - \sum_{j=1}^3 \partial_{s_j} \left( \left( \Omega_Y(\theta, y) \vec{s} \right)_j \mathcal{P}_{YS} \right) \quad (10)$$

Note that  $\mathcal{P}_Y$  is related to  $\mathcal{P}_{YS}$  by

$$\mathcal{P}_Y(\theta, y) = \int ds \mathcal{P}_{YS}(\theta, y, \vec{s}) \quad (11)$$

where the integral is over  $\mathbb{R}^3$ . Note also that since the spin variable  $\vec{S}$  is normalized,  $\mathcal{P}_{YS}$  is supported on the 2-sphere, i.e., where  $|\vec{s}| = 1$ . Hence  $\mathcal{P}_{YS}(\theta, y, \vec{s})$  is proportional to  $\delta(|\vec{s}| - 1)$ . By integrating (10) over  $\vec{s}$  one recovers (9). The polarization density  $\vec{\eta}_Y$  corresponding to  $\mathcal{P}_{YS}$  is defined by

$$\vec{\eta}_Y(\theta, y) = \int ds \vec{s} \mathcal{P}_{YS}(\theta, y, \vec{s}) \quad (12)$$

The RBE (5) follows from (10) by differentiating (12) w.r.t.  $\theta$ .

### APPROXIMATING THE BEAM FRAME RBE BY THE METHOD OF AVERAGING

Because the coefficients of  $L_Y$  are  $\theta$ -dependent, the RBE (5) is numerically quite complex. So we first approximate it analytically in order to solve it numerically. We will find this approximate RBE by refining the averaging technique presented by Ellison, Mais and Ripken in the Accelerator Handbook [19, Section 2.1.4]. This refinement allows us to use that method of averaging to approximate the system of Langevin Equations (7). We just give a sketch here (a detailed account will be published elsewhere [20]). Note that both [19, Section 2.1.4] and our refinement are restricted to first-order averaging. We first rewrite (7) as

$$Y' = (A(\theta) + \epsilon \delta A(\theta))Y + \sqrt{\epsilon} \sqrt{\omega(\theta)} e_6 \xi(\theta) \quad (13)$$

with  $\sqrt{\epsilon} \sqrt{\omega(\theta)} = \sqrt{\omega_Y(\theta)}$ , where  $A(\theta)$  is the Hamiltonian part of  $\mathcal{A}(\theta)$  and  $\epsilon$  is a perturbation parameter, and where  $\epsilon \delta A(\theta)$  represents the part of  $\mathcal{A}(\theta)$  associated with damping effects due to synchrotron radiation and cavities (see, e.g., [15, eq. 5.3]). The mean  $m_Y$  and covariance matrix  $K_Y$  of  $Y$  satisfy the ODEs

$$m'_Y = (A(\theta) + \epsilon \delta A(\theta))m_Y, \quad (14)$$

$$K'_Y = (A(\theta) + \epsilon \delta A(\theta))K_Y + K_Y(A(\theta) + \epsilon \delta A(\theta))^T + \epsilon \omega(\theta) e_6 e_6^T \quad (15)$$

In (15) the  $\delta A$  terms and the  $\omega$  are balanced at  $O(\epsilon)$  and so can be treated together in first order perturbation theory. This is the reason for the  $\sqrt{\epsilon}$  in (13). However this balance is also physical as the damping and diffusion come from the same source and the cavities replenish the energy loss.

To apply the method of averaging to (14) and (15) we must transform them to a standard form for averaging. We do this by using a fundamental solution matrix  $X$  of the unperturbed  $\epsilon = 0$  part of (13) and (14), i.e.,

$$X' = A(\theta)X \quad (16)$$

We thus transform  $Y$ ,  $m_Y$  and  $K_Y$  into  $U$ ,  $m_U$  and  $K_U$  via

$$Y = X(\theta)U, \quad m_Y = X(\theta)m_U, \quad K_Y = X(\theta)K_U X^T(\theta) \quad (17)$$

and (13), (14) and (15) are transformed to

$$U' = \epsilon \mathcal{D}(\theta)U + \sqrt{\epsilon} \sqrt{\omega(\theta)} X^{-1}(\theta) e_6 \xi(\theta) \quad (18)$$

$$m'_U = \epsilon \mathcal{D}(\theta)m_U, \quad (19)$$

$$K'_U = \epsilon (\mathcal{D}(\theta)K_U + K_U \mathcal{D}^T(\theta)) + \epsilon \mathcal{E}(\theta) \quad (20)$$

Here  $\mathcal{D}(\theta)$  and  $\mathcal{E}(\theta)$  are defined by

$$\mathcal{D}(\theta) = X^{-1}(\theta) \delta A(\theta) X(\theta), \quad (21)$$

$$\mathcal{E}(\theta) = \omega(\theta) X^{-1}(\theta) e_6 e_6^T X^{-T}(\theta) \quad (22)$$

Of course, (18)–(20) carry the same information as (13)–(15).

Now, applying the method of averaging to (19) and (20), we obtain

$$m'_V = \epsilon \bar{\mathcal{D}} m_V, \quad (23)$$

$$K'_V = \epsilon (\bar{\mathcal{D}} K_V + K_V \bar{\mathcal{D}}^T) + \epsilon \bar{\mathcal{E}} \quad (24)$$

where the bar denotes  $\theta$ -averaging, i.e., the operation  $\lim_{T \rightarrow \infty} (1/T) \int_0^T d\theta \dots$ . For physically reasonable  $A$  each fundamental matrix  $X$  is a quasiperiodic function whence  $\mathcal{D}$  and  $\mathcal{E}$  are quasiperiodic functions so that their time averages  $\bar{\mathcal{D}}$  and  $\bar{\mathcal{E}}$  exist. By averaging theory  $|m_U(\theta) - m_V(\theta)| \leq C_1(T)\epsilon$  and  $|K_U(\theta) - K_V(\theta)| \leq C_2(T)\epsilon$  for  $0 \leq \theta \leq T/\epsilon$  where  $T$  is a constant (see also [21–24]) and  $\epsilon$  small. However, we expect to be able to show that these estimates are uniformly valid on  $[0, \infty)$ , since the long time behavior is exact.

The key point now is that every Gaussian process  $V$ , whose mean  $m_V$  and covariance matrix  $K_V$  satisfy the ODEs (23) and (24), satisfies the system of Langevin equations

$$V' = \epsilon \bar{\mathcal{D}} V + \sqrt{\epsilon} \mathcal{B}(\xi_1, \dots, \xi_k)^T \quad (25)$$

Here  $\xi_1, \dots, \xi_k$  are statistically independent versions of the white noise process and  $\mathcal{B}$  is a  $6 \times k$  matrix which satisfies

$$\mathcal{B} \mathcal{B}^T = \bar{\mathcal{E}} \quad (26)$$

with  $k = \text{rank}(\bar{\mathcal{E}})$ . Since  $m_U(\theta) = m_V(\theta) + O(\epsilon)$  and  $K_U(\theta) = K_V(\theta) + O(\epsilon)$  we get  $U(\theta) \approx V(\theta)$ . In particular  $X^{-1}(\theta)V(\theta) \approx Y(\theta)$  (more details will be in [20]). Conversely, the mean vector  $m_V$  and covariance matrix  $K_V$  of every  $V$  in (25) satisfy the ODEs (23) and (24).

It's likely that stochastic averaging techniques [25, and references therein] can be applied directly to (18) giving (25) as an approximation and we are looking into that. However, because (18) is linear and defines a Gaussian process, the theory for getting to (25) from the ODEs for the moments could not be simpler, even though it is indirect.

To include the spin we extend (25) to the spin-orbit system of Langevin equations

$$V' = \epsilon \bar{\mathcal{D}} V + \sqrt{\epsilon} \mathcal{B}(\xi_1, \dots, \xi_k)^T, \quad (27)$$

$$\vec{S}' = \Omega_Y(\theta, X(\theta)V) \vec{S} \quad (28)$$

With (27) and (28) the evolution equation for the spin-orbit probability density  $\mathcal{P}_{VS} = \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s})$  is the following F-P equation:

$$\partial_\theta \mathcal{P}_{VS} = L_V \mathcal{P}_{VS} - \sum_{j=1}^3 \partial_{s_j} \left( \left( \Omega_Y(\theta, X(\theta)\mathbf{v}) \vec{s} \right)_j \mathcal{P}_{VS} \right) \quad (29)$$

where

$$L_V = -\epsilon \sum_{j=1}^6 \partial_{v_j} (\bar{\mathcal{D}}_V)_j + \frac{\epsilon}{2} \sum_{i,j=1}^6 \bar{\mathcal{E}}_{ij} \partial_{v_i} \partial_{v_j} \quad (30)$$

The polarization density  $\vec{\eta}_V$  corresponding to  $\mathcal{P}_{VS}$  is defined by

$$\vec{\eta}_V(\theta, \mathbf{v}) = \int ds \vec{s} \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s}) \quad (31)$$

so that by (29), the RBE is

$$\partial_\theta \vec{\eta}_V = (L_V + L_{V,TBMT}) \vec{\eta}_V \quad (32)$$

where

$$L_{V,TBMT} \vec{\eta}_V = \Omega_Y(\theta, X(\theta)\mathbf{v}) \vec{\eta}_V \quad (33)$$

The coefficients of  $L_V$  are  $\theta$ -independent for every choice of  $X$  and this is necessary for our numerical method. Note that the averaging which leads to (32) affects only the orbital variables. It was justified by using the fact that (27) is linear whence it defines a Gaussian process when the initial condition is Gaussian. This allowed us to apply the averaging approach to the first and second moments rather than the Langevin equation itself. We cannot apply this approach to the combined spin-orbit dynamics in (27)-(28) because (28) has a quadratic nonlinearity. In future work, we will pursue this using stochastic averaging as in [25].

We now need an appropriate  $X$  and we note that

$$X(\theta) = M(\theta)C \quad (34)$$

where  $C$  is an arbitrary invertible  $6 \times 6$  matrix and  $M$  is the principal solution matrix, i.e.,  $M' = A(\theta)M$ ,  $M(0) = I$ . Thus choosing  $X$  boils down to choosing a good  $C$ . As is common for spin physics in electron storage rings we emulate Chao's approach [19, Section 2.1.4], [26,27] and use the eigenvectors of  $M(2\pi)$ . We assume that the unperturbed orbital motion is stable. Thus  $M(2\pi)$  has a full set of linearly independent eigenvectors and the eigenvalues are on the unit circle in the complex plane [28]. We further assume a non-resonant condition on the orbital frequencies. We construct  $C$  as a real matrix using the real and imaginary parts of the eigenvectors in its columns and using the fact that  $M(2\pi)$  is symplectic (since  $A(\theta)$  is a Hamiltonian matrix). It follows that  $\bar{\mathcal{D}}$  has block diagonal form and  $\bar{\mathcal{E}}$  has diagonal form. Then the three degrees of freedom are uncoupled in the

operator  $L_V$  in (30). Explicitly,

$$\bar{\mathcal{D}} = \begin{pmatrix} \mathcal{D}_I & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & \mathcal{D}_{II} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & \mathcal{D}_{III} \end{pmatrix}, \quad (35)$$

$$\mathcal{D}_\alpha = \begin{pmatrix} a_\alpha & b_\alpha \\ -b_\alpha & a_\alpha \end{pmatrix}, (\alpha = I, II, III) \quad (36)$$

and  $\bar{\mathcal{E}} = \text{diag}(\mathcal{E}_I, \mathcal{E}_I, \mathcal{E}_{II}, \mathcal{E}_{II}, \mathcal{E}_{III}, \mathcal{E}_{III})$  with  $a_\alpha \leq 0$  and  $\mathcal{E}_I, \mathcal{E}_{II}, \mathcal{E}_{III} \geq 0$ . Thus the three degrees of freedom are uncoupled in  $L_V$  since, by (30),

$$L_V = L_{V,I} + L_{V,II} + L_{V,III} \quad (37)$$

where each  $L_{V,\alpha}$  is an operator in one degree of freedom and is determined by  $\mathcal{D}_\alpha$  and  $\mathcal{E}_\alpha$  via (30) ( $\alpha = I, II, III$ ).

We now have  $Y(\theta) = X(\theta)U(\theta) \approx Y_\alpha(\theta) := X(\theta)V(\theta)$  and it follows that  $\vec{\eta}_Y$  in (5) is given approximately by

$$\vec{\eta}_Y(\theta, \mathbf{y}) \approx \vec{\eta}_{Y,\alpha}(\theta, \mathbf{y}) = \det(X^{-1}(0)) \vec{\eta}_V(\theta, X^{-1}(\theta)\mathbf{y}) \quad (38)$$

Now (32) and the RBE for  $\vec{\eta}_{Y,\alpha}$  carry the same information. However in general the RBE for  $\vec{\eta}_{Y,\alpha}$  does not have the nice features of (32), e.g., (35), (36) and  $L_V$  being  $\theta$ -independent, which make the latter useful for our numerical method (see below). Hence we discretize (32) rather than the RBE for  $\vec{\eta}_{Y,\alpha}$ .

We finally mention a feature of  $\vec{\eta}_V$  which is helpful for finding an appropriate numerical phase space domain for  $\vec{\eta}_V$ . The orbital probability density  $\mathcal{P}_V$  corresponding to  $\mathcal{P}_{VS}$  is defined by

$$\mathcal{P}_V(\theta, \mathbf{v}) = \int ds \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s}) \quad (39)$$

whence by (31),

$$\begin{aligned} |\vec{\eta}_V(\theta, \mathbf{v})| &= \left| \int ds \vec{s} \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s}) \right| \leq \int ds |\vec{s}| \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s}) \\ &= \int ds \mathcal{P}_{VS}(\theta, \mathbf{v}, \vec{s}) = \mathcal{P}_V(\theta, \mathbf{v}) \end{aligned} \quad (40)$$

so that the numerical phase space domain for  $\vec{\eta}_V$  can be identified with the numerical phase space domain for  $\mathcal{P}_V$ . The latter is easy to find since we generally use exact expressions of  $\mathcal{P}_V$ , e.g., the one for orbital equilibrium.

## TWO-DEGREE-OF-FREEDOM CASE

We now consider a case of two degrees of freedom in a flat ring just with FODO cells and cavities. The case of two degrees of freedom is a natural step towards three degrees of freedom. Moreover the case of a flat ring allows us to use a one-dimensional approach to spin, leading to a linear spin-orbit system, a system to which we can apply our averaging approach. The Gaussian nature of the associated process allows us to analytically solve the average RBE.

In our flat ring model  $\Omega_Y$  has the simple form

$$\Omega_Y(\theta, Y) = -a_Y(\theta)Y\mathcal{J}, \quad \mathcal{J} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where  $Y = (Y_1, Y_2, Y_3, Y_4)^T$  represents the horizontal and longitudinal motions which are uncoupled from the vertical motion in the flat ring model. It is convenient to use spherical coordinates as spin variables, i.e.,  $\tilde{S} = (\cos(\Psi) \sin(\Phi), \sin(\Psi) \sin(\Phi), \cos(\Phi))^T$ . The beam frame system of Langevin equations are then

$$Y' = (A(\theta) + \epsilon \delta A(\theta))Y + \sqrt{\epsilon} \sqrt{\omega(\theta)} (0, 0, 0, 1)^T \xi(\theta), \quad (41)$$

$$\Psi' = a_Y(\theta)Y, \quad (42)$$

$$\Phi' = 0 \quad (43)$$

where the row vector  $a_Y(\theta)$  is  $2\pi$ -periodic in  $\theta$ . To apply the method of averaging to the system (41)-(43) we transform the system to a standard form for averaging. We do this by defining  $\tilde{Y} := (Y_1, Y_2, Y_3, Y_4, \Psi, \Phi)^T$  and by using a fundamental solution matrix  $Z$  of the unperturbed  $\epsilon = 0$  part of (41)-(43), i.e.,

$$Z' = \begin{pmatrix} A(\theta) & 0_{4 \times 2} \\ a_Y(\theta) & 0_{1 \times 2} \\ 0_{1 \times 4} & 0_{1 \times 2} \end{pmatrix} Z \quad (44)$$

By transforming  $\tilde{Y}$  into  $Q$  via

$$\tilde{Y} = Z(\theta)Q \quad (45)$$

one gets the system of Langevin equations

$$Q' = \epsilon \mathcal{D}(\theta)Q + \sqrt{\epsilon} \sqrt{\omega(\theta)} Z^{-1}(\theta) e_4 \xi(\theta) \quad (46)$$

where  $e_4 = (0, 0, 0, 1, 0, 0)^T$  and

$$\mathcal{D}(\theta) = Z^{-1}(\theta) \begin{pmatrix} \delta A(\theta) & 0_{4 \times 2} \\ 0_{2 \times 4} & 0_{2 \times 2} \end{pmatrix} Z(\theta), \quad (47)$$

Thus the mean  $m_Q$  and covariance matrix  $K_Q$  of  $Q$  satisfy the ODEs

$$m'_Q = \epsilon \mathcal{D}(\theta) m_Q, \quad (48)$$

$$K'_Q = \epsilon (\mathcal{D}(\theta) K_Q + K_Q \mathcal{D}^T(\theta)) + \epsilon \mathcal{E}(\theta) \quad (49)$$

where

$$\mathcal{E}(\theta) = \omega(\theta) Z^{-1}(\theta) e_4 e_4^T Z^{-T}(\theta) \quad (50)$$

By averaging (48) and (49) we get the ODEs

$$m'_W = \epsilon \bar{\mathcal{D}} m_W, \quad (51)$$

$$K'_W = \epsilon (\bar{\mathcal{D}} K_W + K_W \bar{\mathcal{D}}^T) + \epsilon \bar{\mathcal{E}} \quad (52)$$

where the bar denotes  $\theta$ -averaging. Since the ODE system (51),(52) is autonomous it can be analytically solved. For physically reasonable choices of the parameters in (41)-(43) each fundamental matrix  $Z$  is a quasiperiodic function whence  $\mathcal{D}$  and  $\mathcal{E}$  are quasiperiodic functions so that their  $\theta$ -averages  $\bar{\mathcal{D}}$  and  $\bar{\mathcal{E}}$  exist. By averaging theory  $m_Q(\theta) = m_W(\theta) + O(\epsilon)$  and  $K_Q(\theta) = K_W(\theta) + O(\epsilon)$  for  $0 \leq \theta \leq T/\epsilon$  where  $T$  is a constant (see also [21–24]). Every Gaussian

process  $W$ , whose mean  $m_W$  and covariance matrix  $K_W$  satisfy the ODEs (51) and (52), satisfies the system of Langevin equations

$$W' = \epsilon \bar{\mathcal{D}} W + \sqrt{\epsilon} \mathcal{B}(\xi_1, \dots, \xi_k)^T \quad (53)$$

Here  $\xi_1, \dots, \xi_k$  are statistically independent versions of the white noise process and where  $\mathcal{B}$  is a  $6 \times k$  matrix which satisfies

$$\mathcal{B} \mathcal{B}^T = \bar{\mathcal{E}} \quad (54)$$

with  $k = \text{rank}(\bar{\mathcal{E}})$ . Since  $m_Q(\theta) = m_W(\theta) + O(\epsilon)$  and  $K_Q(\theta) = K_W(\theta) + O(\epsilon)$  we get  $Q(\theta) \approx W(\theta)$ . In particular  $Z^{-1}(\theta)W(\theta) \approx \tilde{Y}(\theta)$ . Clearly the third component of the spin does not evolve, the spins only evolve in the plane.

As in the case of three degrees of freedom we assume that the unperturbed orbital motion is stable and nonresonant. Thus, as in the case of three degrees of freedom, we can construct a fundamental matrix  $Z$  such that the orbital part of  $\bar{\mathcal{D}}$  has block diagonal form and such that the orbital part of  $\bar{\mathcal{E}}$  has diagonal form, i.e.,

$$\bar{\mathcal{D}} = \begin{pmatrix} \mathcal{D}_I & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & \mathcal{D}_{II} & 0_{2 \times 2} \\ \bar{\mathcal{D}}_{51} & \bar{\mathcal{D}}_{52} & \bar{\mathcal{D}}_{53} & \bar{\mathcal{D}}_{54} & 0_{1 \times 2} \\ 0_{1 \times 2} & 0_{1 \times 2} & 0_{1 \times 2} & 0_{1 \times 2} & 0 \end{pmatrix}, \quad (55)$$

$$\bar{\mathcal{E}} = \begin{pmatrix} \mathcal{E}_I & 0 & 0 & 0 & \bar{\mathcal{E}}_{15} & 0 \\ 0 & \mathcal{E}_I & 0 & 0 & \bar{\mathcal{E}}_{25} & 0 \\ 0 & 0 & \mathcal{E}_{II} & 0 & \bar{\mathcal{E}}_{35} & 0 \\ 0 & 0 & 0 & \mathcal{E}_{II} & \bar{\mathcal{E}}_{45} & 0 \\ \bar{\mathcal{E}}_{15} & \bar{\mathcal{E}}_{25} & \bar{\mathcal{E}}_{35} & \bar{\mathcal{E}}_{45} & \bar{\mathcal{E}}_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (56)$$

where  $\mathcal{D}_I, \mathcal{D}_{II}$  are  $2 \times 2$  matrices of the form (36) and  $\mathcal{E}_I, \mathcal{E}_{II}$  are nonnegative. If  $\mathcal{P}_W = \mathcal{P}_W(\theta, w)$  is a probability density of a Gaussian process associated with (53) then the polarization density  $\vec{\eta}_W$  corresponding to  $\mathcal{P}_W$  is defined by

$$\vec{\eta}_W(\theta, w) = \int dw_5 dw_6 \begin{pmatrix} \cos(w_5) \sin(w_6) \\ \sin(w_5) \sin(w_6) \\ \cos(w_6) \end{pmatrix} \mathcal{P}_W(\theta, w) \quad (57)$$

and satisfies the RBE

$$\begin{aligned} \partial_\theta \vec{\eta}_W &= -\epsilon \sum_{j=1}^2 \partial_{w_j} \left( \left( \mathcal{D}_I(w_1, w_2)^T \right)_j \vec{\eta}_W \right. \\ &\quad \left. - \epsilon \sum_{j=3}^4 \partial_{w_j} \left( \left( \mathcal{D}_{II}(w_3, w_4)^T \right)_j \vec{\eta}_W \right) \right. \\ &\quad \left. + \frac{\epsilon}{2} \mathcal{E}_I \left( \partial_{w_1} \partial_{w_1} + \partial_{w_2} \partial_{w_2} \right) \vec{\eta}_W \right. \\ &\quad \left. + \frac{\epsilon}{2} \mathcal{E}_{II} \left( \partial_{w_3} \partial_{w_3} + \partial_{w_4} \partial_{w_4} \right) \vec{\eta}_W \right. \\ &\quad \left. - \epsilon \sum_{j=1}^4 \bar{\mathcal{D}}_{5j} w_j \mathcal{J} \vec{\eta}_W - \frac{\epsilon}{2} \bar{\mathcal{E}}_{55} \vec{\eta}_W + \epsilon \sum_{j=1}^4 \bar{\mathcal{E}}_{j5} \mathcal{J} \vec{\eta}_W \right) \quad (58) \end{aligned}$$

Since the ODE system (51),(52) can be analytically solved,  $\mathcal{P}_W$  can be computed analytically for every Gaussian process. Then by (57),  $\vec{\eta}_W$  can be computed analytically.

## ONE-DEGREE-OF-FREEDOM CASE

We now consider the case of one degree of freedom using the model studied in [29, 30], which involves only synchrotron motion. The case of one degree of freedom is the first step towards two and three degrees of freedom. The one-degree-of-freedom model here is obtained from the two-degrees-of-freedom flat-ring model of the previous section by setting, in (55) and (56),

$$0 = \mathcal{D}_{II} = \bar{\mathcal{D}}_{52} = \bar{\mathcal{D}}_{53} = \bar{\mathcal{D}}_{54} = \mathcal{E}_{II} = \bar{\mathcal{E}}_{25} = \bar{\mathcal{E}}_{35} = \bar{\mathcal{E}}_{45}, \\ \mathcal{D}_I = -I_{2 \times 2}, \mathcal{E}_I = 1, \bar{\mathcal{E}}_{15} = -\bar{\mathcal{D}}_{51}, \bar{\mathcal{E}}_{55} = (\bar{\mathcal{E}}_{15})^2 \quad (59)$$

One can justify the step from (55) and (56) to (59) as a good approximation by applying the betatron-dispersion formalism to the flat ring model [31]. With (59) the variables  $W_3, W_4, W_6$  are uncoupled so that we are left with the following one-degree-of-freedom model resulting in the following system of Langevin equations for the orbital variables  $W_1, W_2$  and the spin variable  $W_5$ :

$$\begin{pmatrix} W_1' \\ W_2' \\ W_5' \end{pmatrix} = \epsilon \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ g & 0 & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_5 \end{pmatrix} \\ + \sqrt{\frac{\epsilon}{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -g & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

where  $g = \bar{\mathcal{D}}_{51} = -\bar{\mathcal{E}}_{15}$  and  $\xi_1, \xi_2$  are statistically independent versions of the white noise process. Denoting the polarization density for our one-degree-of-freedom model by  $\vec{\eta}_{1D}$ , one can show in analogy to the previous section that it satisfies the RBE

$$\partial_\theta \vec{\eta}_{1D} = \epsilon \left( \partial_{w_1} (w_1 \vec{\eta}_{1D}) + \partial_{w_2} (w_2 \vec{\eta}_{1D}) \right) + \frac{\epsilon}{4} \partial_{w_1} \partial_{w_1} \vec{\eta}_{1D} \\ + \frac{\epsilon}{4} \partial_{w_2} \partial_{w_2} \vec{\eta}_{1D} - \epsilon g w_1 \mathcal{J} \vec{\eta}_{1D} - \frac{\epsilon}{2} g \mathcal{J} \partial_{w_1} \vec{\eta}_{1D} - \frac{\epsilon}{4} g^2 \vec{\eta}_{1D} \quad (60)$$

Note that the analytic solutions of the RBE (60) give strong evidence for the validity of the averaging method since the analytic solutions of (60) can be compared with the solutions of the analytic solutions of the exact RBE in [29, 30].

## SKETCH OF THE NUMERICAL APPROACH

We now briefly sketch our numerical approach to the RBEs (32), (58) and (60). For simplicity we here focus on (32). The numerical computations are performed by using 3 pairs  $(r_\alpha, \varphi_\alpha)$  of polar coordinates, i.e.,  $v_1 = r_I \cos \varphi_I, \dots, v_6 = r_{III} \sin \varphi_{III}$ . The angle variables are Fourier transformed hence the Fourier coefficients are functions of time and the radial variables. We discretize the radial variables by using the pseudospectral method [32, 33] using a Chebychev grid for each radial variable. This results for each Fourier mode in a system of linear first-order ODEs in  $\theta$  which we discretize by using an implicit/explicit  $\theta$ -stepping

scheme. Because of (30), (35) and (36) the Fourier modes are uncoupled in  $L_V \vec{\eta}_V$  so that the only coupling of Fourier modes in (32) comes via  $L_{V, TBMT} \vec{\eta}_V = \Omega_V(\theta, X(\theta)v) \vec{\eta}_V$  and this coupling is local since  $\Omega_V(\theta, X(\theta)v)$  is linear in  $v$ . Thus the parabolic terms are separated from the mode coupling terms, i.e., in the time stepping  $L_V \vec{\eta}_V$  is treated implicitly and  $L_{V, TBMT} \vec{\eta}_V$  is treated explicitly. The implicit time stepping involves a linear solver whose efficiency depends on  $L_V$  being  $\theta$ -independent. Note that the pseudospectral method is a minimal-residue method by which the residual of a PDE is zero at the numerical grid points. Note also that the numerical boundary conditions are periodic in the angle variables and for each radial variable  $r_\alpha$  we impose homogenous Dirichlet boundary conditions at  $r_\alpha = r_{\max}$ . The latter are justified by the inequality (40) and the fact we impose these boundary conditions also on the orbital probability density  $\mathcal{P}_V$ .

## ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0018008.

## REFERENCES

- [1] J. D. Jackson, “*Classical Electrodynamics*”, 3rd edition, Wiley, 1998.
- [2] A. A. Sokolov and I. M. Ternov, “On Polarization and Spin Effects in Synchrotron Radiation Theory”, *Sov. Phys. Dokl.*, vol. 8, no. 12, pp. 1203, 1964.
- [3] Ya. S. Derbenev and A. M. Kondratenko, “Polarization kinetics of particles in storage rings”, *Sov. Phys. JETP*, vol. 37, p. 968, 1973.
- [4] D.P. Barber and G. Ripken, *Radiative Polarization, Computer Algorithms and Spin Matching in Electron Storage Rings, Handbook of Accelerator Physics and Engineering*. Eds. A. W. Chao and M. Tigner, 1st edition, 3rd printing, World Scientific, 2006. See also arXiv:physics/9907034v2
- [5] F. Bloch, *Phys. Rev.*, vol. 70, p. 460, 1946.
- [6] Ya. S. Derbenev and A. M. Kondratenko, “Relaxation and equilibrium state of electrons in storage rings”, *Sov. Phys. Dokl.*, vol. 19, p. 438, 1975.
- [7] V. N. Baier, V. M. Katkov, and V. M. Strakhovenko, “Kinetics of Radiative Polarization”, *Sov. Phys. JETP*, vol. 31, p. 908, 1970.
- [8] O. Beznosov, J. A. Ellison, K. Heinemann, D. P. Barber, and D. Appelö, “Spin dynamics in modern electron storage rings: Computational aspects”, presented at ICAP’18, Key West, FL, USA, Oct 2018, paper MOPAF04, this conference.
- [9] E. Gianfelice, “Self Polarization in Storage Rings”, Invited talk at SPIN 2018, Ferrara. See <https://agenda.infn.it/getFile.py/access?contribId=151&sessionId=11&resId=1&materialId=slides&confId=12464>
- [10] *Handbook of Accelerator Physics and Engineering*, 1st edition, third printing, edited by A. W. Chao, M. Tigner, 2006.
- [11] M. Sands, “The physics of electron storage rings”, SLAC-121, 1970.

- [12] J. Jowett, "Introductory Statistical Mechanics for electron storage rings", SLAC-PUB-4033, 1986.
- [13] K. Heinemann, unpublished notes
- [14] K. Heinemann and D. P. Barber, "Spin transport, spin diffusion and Bloch equations in electron storage rings", *Nucl. Instr. Meth. A*, vol. 463, pp. 62–67, 2001. Erratum-ibid.A469:294.
- [15] D. P. Barber, K. Heinemann, H. Mais, and G. Ripken, "A Fokker-Planck treatment of stochastic particle motion", DESY-91-146, 1991.
- [16] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, Wiley, New York, 1974.
- [17] T. C. Gard, *Introduction to Stochastic Differential Equations*, Dekker, New York, 1988.
- [18] C. W. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences*, Springer, Berlin, 1985.
- [19] *Handbook of Accelerator Physics and Engineering*. Eds. A. W. Chao, K. H. Mess, M. Tigner, F. Zimmermann, 2nd edition, World Scientific, 2013.
- [20] J. A. Ellison *et al.*, "Details of Orbital Eigen-analysis for Electron Storage Rings", in preparation.
- [21] J. A. Ellison and H-Jeng Shih, "The Method of Averaging in Beam Dynamics", in *Accelerator Physics Lectures at the Superconducting Super Collider, AIP Conference Proceedings*, vol. 326, edited by Yiton T. Yan, James P. Naples and Michael J. Syphers, 1995.
- [22] J. A. Ellison, K. A. Heinemann, M. Vogt, and M. Gooden, "Planar undulator motion excited by a fixed traveling wave: Quasiperiodic Averaging, normal forms and the FEL pendulum", *Phys. Rev. ST Accel. Beams*, vol. 16, pp. 090702, 2013.
- An earlier version is on the archive at arXiv:1303.5797 (2013) and published as DESY report 13-061.
- [23] J. A. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, 2nd Edition, Springer, New York, 2007.
- [24] J. Murdock, *Perturbations: Theory and Methods*, SIAM, Philadelphia, 1999.
- [25] R. Cogburn and J.A. Ellison, "A Stochastic Theory of Adiabatic Invariance", *Communications of Mathematical Physics*, vol. 148, pp. 97–126, 1992.
- [26] A. W. Chao, *J. Appl. Phys.*, vol. 50, pp. 595, 1979.
- [27] A. W. Chao, *Nucl. Instr. Meth.*, vol. 180, pp. 29–36, 1981.
- [28] K. R. Meyer, G. R. Hall, and D. Offin, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, 2nd edition, Springer, New York, 2009.
- [29] K. Heinemann, "Some models of spin coherence and decoherence in storage rings", arXiv:physics/9709025, 1997.
- [30] D. P. Barber, M. Böge, K. Heinemann, H. Mais, and G. Ripken, in *Proc. 11th Int. Symp. High Energy Spin Physics*, Bloomington, IN, USA, 1994.
- [31] H. Mais and G. Ripken, "Spin-orbit motion in a storage ring in the presence synchrotron radiation using a dispersion formalism", DESY-86-029, 1986.
- [32] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods. Fundamentals in Single Domains*, Springer, Berlin, 2006.
- [33] B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, Cambridge, 1996.

# REALISTIC MODELING OF THE MUON $g - 2$ EXPERIMENT BEAMLINES AT FERMILAB\*

D. A. Tarazona<sup>†</sup>, M. Berz, K. Makino, Michigan State University, East Lansing, MI, USA  
D. Stratakis, M. J. Syphers<sup>1</sup>, Fermi National Accelerator Laboratory, Batavia, IL, USA  
<sup>1</sup>also at Northern Illinois University, DeKalb, IL, USA

## Abstract

The main goal of the Muon  $g - 2$  Experiment at Fermilab (E989) is to measure the muon anomalous magnetic moment ( $a_\mu$ , also dubbed as the “anomaly”) to unprecedented precision. This new measurement will allow to test the completeness of the Standard Model (SM) and to validate other theoretical models beyond the SM. Simulations of the beamlines from the pion production target to the entrance of the  $g - 2$  Storage Ring (SR) using COSY INFINITY [1] contribute to the understanding and characterization of the muon beam production in relation to the statistical and systematic uncertainties of the E989 measurement. The effect of nonlinearities from fringe fields and high-order contributions on the beam delivery system performance are considered, as well as interactions with the beamline elements apertures, particle decay channels, spin dynamics, and beamline misalignments.

## INTRODUCTION

The most recent measurement of  $a_\mu$  at the Brookhaven National Laboratory Muon  $g - 2$  Experiment (E821) yielded an experimental relative uncertainty of 0.54 ppm, which differs from current SM predictions by about  $3.7\sigma$  [2]. In contrast to E821, the goal of E989 is to deliver a measurement of the anomaly to a precision of 0.14 ppm or less to reach  $> 5\sigma$  discrepancy with the SM and therefore strongly establish evidence for new physics.

For that purpose, the number of recorded muon decays in the  $g - 2$  storage ring at E989 is required to increase by a factor of 20 with respect to E821. The Fermilab Muon Campus E989 beam delivery system (BDS), which is a set of 1 km-long beamlines between the pion-production target and the entrance of the storage ring downstream, is designed to meet the statistical goal and deliver  $(0.5-1.0) \times 10^5$  muons to the storage ring per  $10^{12}$  protons interacting with the pion-production target.

On the other hand, the relative statistical uncertainty in the experimental  $a_\mu$  is inversely proportional to the muon beam polarization (see [2], Eq. (16.6)). Thus, it is worth to study the effect of nonlinearities and perform spin dynamics simulations. In addition, due to the momentum acceptance of about  $\pm 0.5\%$  of the storage ring it is of interest to numerically evaluate the dynamical properties of the muon beam as it is delivered to the muon storage ring.

Motivated by the reasons exposed above, we have developed a model of the E989 beamlines to reproduce numerical

simulations of the muon beam’s statistical performance and dynamical behavior including spin using COSY INFINITY. This program prepares detailed high-order transfer maps calculated with an 8th order Runge-Kutta integrator and Differential-Algebraic (DA) methods to solve the beam optics Ordinary Differential Equations (ODEs) and perform beam tracking. In particular, we present results from tracking of secondary protons, pions, daughter muons from pion decay, and muons produced right at the entrance of the E989 beamlines downstream the pion-production target. Nonlinear effects due to standard fringe fields, up to 4th-order beam dynamics, spin dynamics, beam collimation, and misalignments of the multiple BDS beamline elements are considered.

The paper begins with a brief description of the Muon Campus E989 beam delivery system. Then details of the beam dynamics simulations throughout the E989 beamlines from the production target to the storage ring entrance and beam performance results considering nonlinear effects are discussed.

## E989 BEAM DELIVERY SYSTEM (BDS)

The main purpose of the E989 beamlines considered in simulations, depicted in Figure 1, is to deliver a clean muon beam with momentum  $p_0 = 3094 \text{ MeV}/c$  to the storage ring. Batches of four bunches made of  $10^{12}$  protons each are directed to an Inconel-600 “pion-production” target located at the AP0 target hall, from which positive secondary particles emerge. 30 cm downstream the target, a 232 T/m magnetic gradient produced by a lithium lens focuses the secondaries. Thereafter, a pulsed magnet with a field of 0.53 T selects  $3.115 \text{ GeV}/c \pm 10\%$  positive particles and bends them  $3^\circ$  towards the 50 m long M2 line, which consists of matching quadrupoles followed by eight more quadrupoles and a dipole that horizontally bends the beam  $3^\circ$  to match with the M3 line (230 m long).

The M3 line is made of FODO cells that maintain small beta functions to provide continuity of pions and daughter muons from pion decay. By the end of M3, the secondary beam is mainly composed of protons that do not interact with the target, pions that need more time to decay, and muons. In order for such beam to become a clean muon beam, the M3 line is aligned with the injection leg of the Delivery Ring (DR) for which horizontal bends deviate the beam to the right by about  $18^\circ$  at  $s \sim 160 \text{ m}$  away from the target ( $s$  represents the longitudinal distance). At the end of M3, a series of magnets involving a C-magnet, a pulsed magnetic

\* Fermilab report: FERMILAB-CONF-18-620-APC.

<sup>†</sup> Email: tarazona@msu.edu. ORCID: 0000-0002-7823-7986.

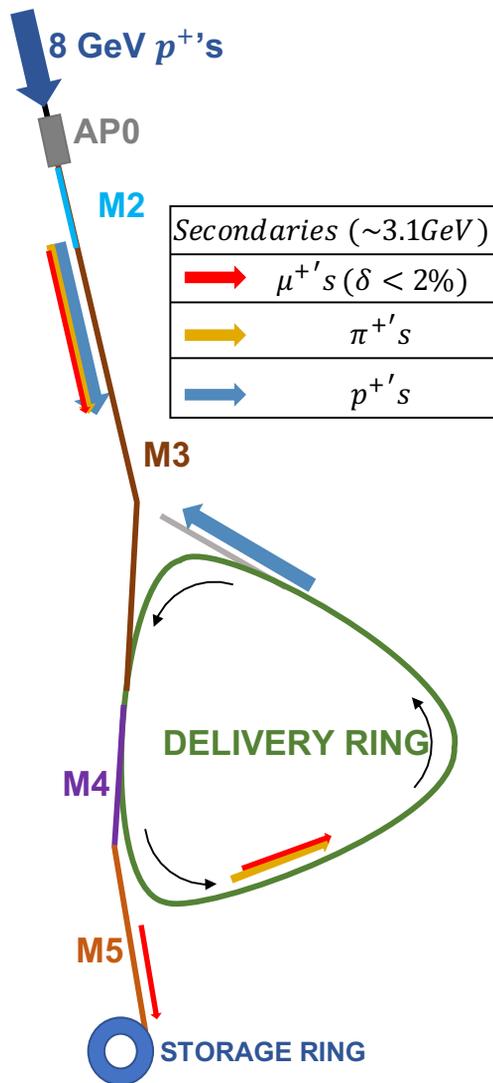


Figure 1: A schematic layout of the BDS.

septum dipole, and kicker modules inject the beam DR after a vertically upward bend of about  $5.7^\circ$ .

Through the 505 m of circumference of the DR, previously used as a debuncher ring and now reconditioned for E989, the remaining pions have enough time to decay into mostly muons as they circulate four times before being extracted into the M4 line. The DR also allows protons to spatially separate from the other lighter particles by a rate of 75 ns per turn [3]; this feature lets a kicker within the DR to safely remove the protons after the fourth turn. The optics functions of the DR using COSY are shown in Fig. 2.

The resulting muon beam is then extracted to the multiple vertical bends and quads of the M4 beamline (30 m), followed by the M5 line (100 m) with a tunable final focus section and ultimately be delivered into the entrance of the storage ring.

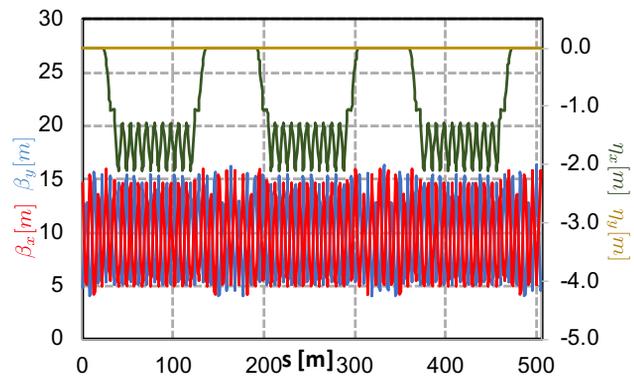


Figure 2: DR optics functions. Fringe fields considered in simulations change the beta functions by less than 3%. The dispersion horizontal function ( $\eta_x$ ) reveals the three-fold symmetry of the ring. Optics functions have served as a channel to validate our BDS model.

### SIMULATION DETAILS AND RESULTS

The following results have served to benchmark other numerical studies at E989 modeled with G4Beamline [3] and Bmad [4]. Moreover, the inclusion into our model of nonlinearities from fringe fields and high-order effects and an analysis covering various misalignment scenarios provide further realism to the characterization of the muon beam that is delivered to the storage ring.

Tracking simulations of the E989 beamlines start with a 6D initial distribution at the exit of the pion-production target from a modeling with MARS [5] of  $10^9$  protons on target. We consider protons, muons, and pions, although other particles emerge from the target as verified by experimental evidence though at significantly smaller ratios [6].

The presented studies consider aperture beam collimation. There are several aperture geometries determined by the multiple purposes of the elements in the E989 beamlines as well as the expected beam size behavior from design. They range from simple squared and circular apertures to more involved star-shaped apertures which can be approximated as an overlaying of ellipses and rectangles in COSY. In simulations, particles are removed from the beam if their spatial transversal coordinates surpass the dimensions of the aperture. This algorithm is repeated every 20 cm or so, depending on the beamline longitudinal size. COSY permits to track 10,000 particles simultaneously, which reduces the computational time. In this manner, the statistical performance of the beam and resulting beam dynamics variables distributions are determined.

Figure 3 shows the number of secondaries throughout the M2 and M3 lines after the production target. At the end of these lines about  $10^{-4}$  protons per proton on target (POT) remain within the secondary beam, which are kicked out downstream at the Delivery Ring. The in-flight decay channels considered in simulations are  $\pi^+ \rightarrow \mu^+ + \nu_\mu$  and  $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$ . The effect of muon decays on the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

overall number of muons throughout the beamlines is not significant due to the short time it takes for the beam to travel from the production target to the entrance of the storage ring, i.e.  $8.1 \mu\text{s}$ . Since the momentum admittance at the storage ring is  $\pm 0.5\%$ , we track muons within specific momentum offsets ( $\delta$ ) as shown in Fig. 3.

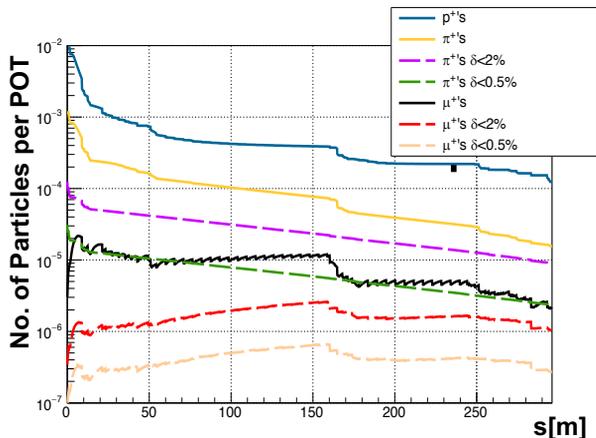


Figure 3: Statistical performance along M2/M3.

The range  $|\delta| < 2\%$  is considered as well in connection with upcoming efforts at E989 to manipulate the beam momentum range with wedge cooling [7]. As shown in Fig. 3, M2/M3 lines maintain the statistics of muons with  $|\delta| < 2\%$  emerging from pion decay. Main muon losses of about 11% and 20% take place at the  $18.5^\circ$  horizontal bend ( $s \sim 160.0 \text{ m}$ ) and along the vertical injection upstream the DR ( $s \sim 280.0 \text{ m}$ ), respectively. The black dot around  $s \sim 235 \text{ m}$  in Fig. 3 depicts the number of total particles per POT from measurements, suggesting reliability on our simulations results. Aperture beam collimation allows to predict the final beam momentum distribution at the entrance of the SR, shown in Fig. 4. Similar simulations were performed along the rest of the E989 beamlines downstream M3 (results are summarized in Table 1).

Table 1: Statistical Performance Along BDS [POT]

	M3 exit	DR exit ( $n = 4$ )	SR entrance
$\mu^+s$	$2.19 \times 10^{-6}$	$1.04 \times 10^{-6}$	$7.48 \times 10^{-7}$
$\mu^+s^*$	$2.72 \times 10^{-7}$	$2.85 \times 10^{-7}$	$1.89 \times 10^{-7}$
$\pi^+s$	$1.55 \times 10^{-5}$	0	0
$p^+s^{**}$	$1.24 \times 10^{-4}$	$6.80 \times 10^{-5}$	$5.86 \times 10^{-5}$

\*Results for  $\delta < 0.5\%$

\*\*Results for the case of proton removal at DR turned off.

Fringe fields map computations are performed at the edges of each beamline element. The longitudinal-dependent tapering of the multipole strengths is modeled by a six parameter Enge function:

$$F(z) = \frac{1}{1 + \exp(a_1 + a_2 \cdot (z/D) + \dots + a_6 \cdot (z/D)^5)}$$

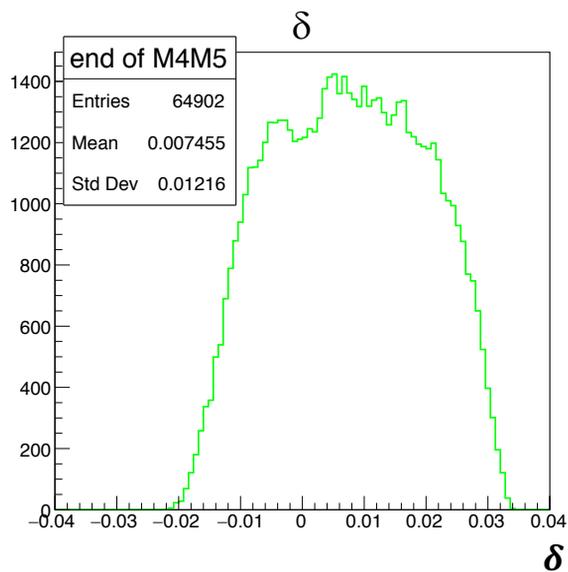


Figure 4: Beam relative momentum distribution at the storage ring entrance.

where  $z$  is the distance perpendicular to the effective field boundary and  $D$  is the full aperture of the particle optical element. The  $a_i$  coefficients are taken by default based on measured data from PEP at SLAC [8]; for such cases, the integrated multipole strengths along the optical axis of each beamline element remain the same as for simulations with hard-edge modeling. Therefore, the following comparisons between numerical results with fringe field effects turned on and off are reliable.

4th-order numerical calculations with and without fringe fields were implemented. Figure 5 shows the differences between the two scenarios along the DR.

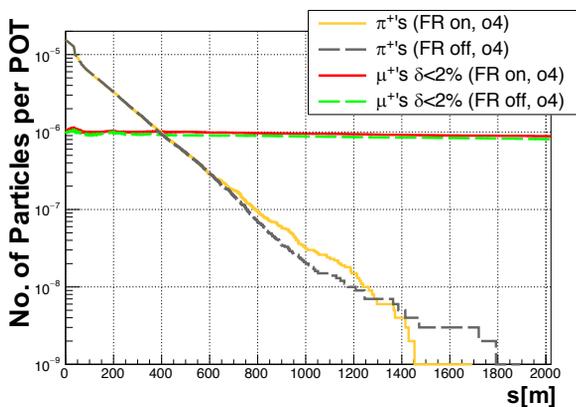


Figure 5: Fringe field effects on beam population at the DR.

After four turns, simulations suggest a favorable contribution due to fringe fields on the muons (i.e. 9.4% increment) and pions population. Fringe fields introduce fields longitudinal to the beam motion, which may be the essential contributor to keep more particles focused in a similar way

to solenoids for low-energy experiments. However, such effect is not sufficiently large to compensate for losses at the beam bends along the Beam Delivery System in a significant way. Specifically, simulation results show an increment of 5.2% more muons at the entrance of the storage ring. Moreover, fringe fields have a larger contribution in the number of surviving muons than in pions; i.e. at the end of the M3 line ( $s = 290$  m)—where most of pions still have not decayed—fringe fields maintained 8.9% more muons on track whereas pions population increment at that location increased by only 0.4%, which is within the  $\sim 2\%$  range of the simulations' statistical error.

On the other hand, fringe fields from the BDS beamline elements were also considered in spin tracking simulations. For the pion decay channel, the resulting muon beam polarization is calculated based on a module that considers the weak interaction process to get the direction of the muon spin vector [9] and tracked with COSY's DA mapping calculation [8]. In Fig. 6, the spin components of a beam made of 64,902 muons at the entrance of the SR are plotted from simulations.

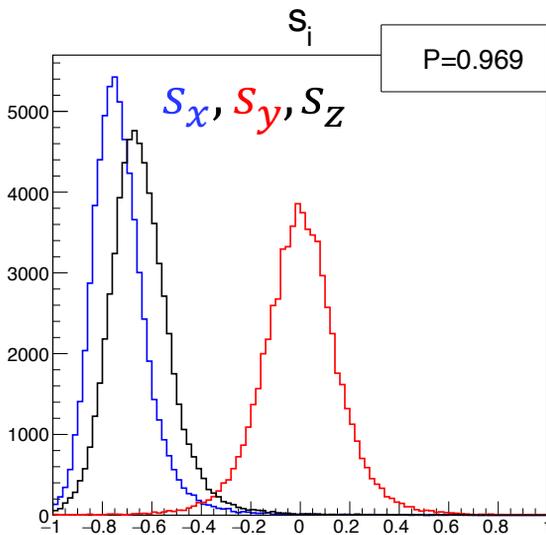


Figure 6: Muon beam spin components at SR entrance.

The components are expressed in terms of the coordinates that describe the relative dynamics around the reference orbit [10]. The resulting polarization is  $P = 0.97$ , in agreement with G4Beamline numerical simulations [3] which do not consider fringe fields. Thus, fringe fields do not interfere with the muon beam polarization. Another spin variable worth to consider is the polarization projection in the horizontal plane with respect to the reference optical axis,  $\varphi_a$  (see Fig. 7). The precession frequency of this phase,  $\omega_a$ , plays an essential role in the final measurement of  $a_\mu^{E989}$  which implies a deep understanding of its evolution as the beam circulates through the SR. In particular, the correlation between  $\varphi_a$  and the Lorentz factor  $\gamma$  of a muon results in a sub-ppm systematic effect of  $\omega_a$ ; high-momentum muons

decay rate is slower than low-momentum muons and consequently  $\omega_a$  shifts as the muon beam decays [11].

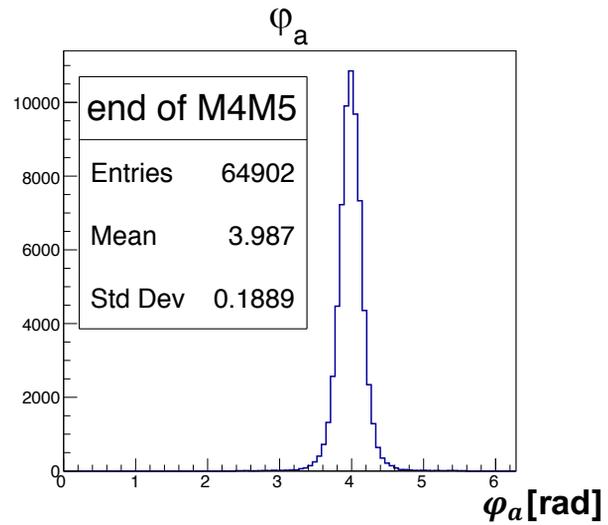


Figure 7:  $\varphi_a$  at SR entrance.

Tracking simulations were performed with and without fringe fields to study the effect of fringe fields on the spin-momentum correlation  $m_\delta = d\langle\varphi_a\rangle/d\gamma$ . For the case of fringe fields turned on, simulations show a correlation  $m_\delta$  equal to  $29.2 \pm 9.4$  mrad after 4 turns around the DR as shown in Fig. 8. On the other hand, similar simulations without fringe fields—conventional hard-edge model—indicate a correlation  $m_\delta = 92.1 \pm 9.8$  mrad. Thus, our numerical studies suggest a significant effect of fringe fields on spin-momentum correlations.

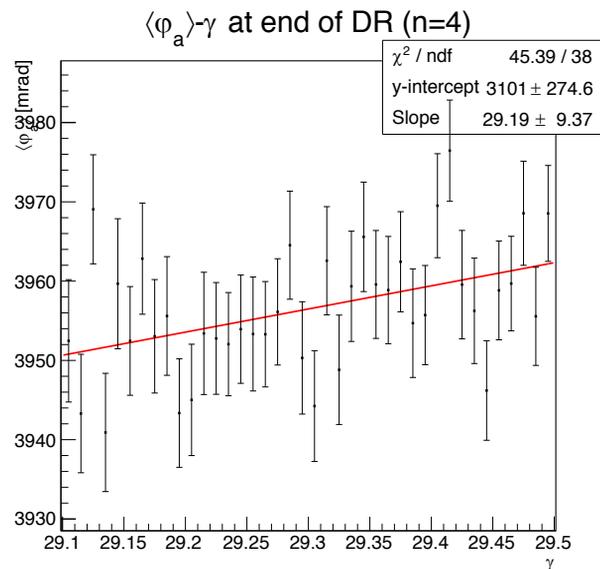


Figure 8:  $\langle\varphi_a\rangle$  vs.  $\gamma$  at DR exit after 4 turns.

Misalignments are introduced in simulations by transforming the transport maps that represent each beamline element.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Transformations follow randomly Gaussian-distributed horizontal and vertical misplacements with standard deviations of 0.25 mm, introducing constant terms to the maps. A total of ten random misalignment configurations of the beam delivery system initialized with different random seeds were considered to procure statistical ranges within which the beam performance would be expected to occupy.

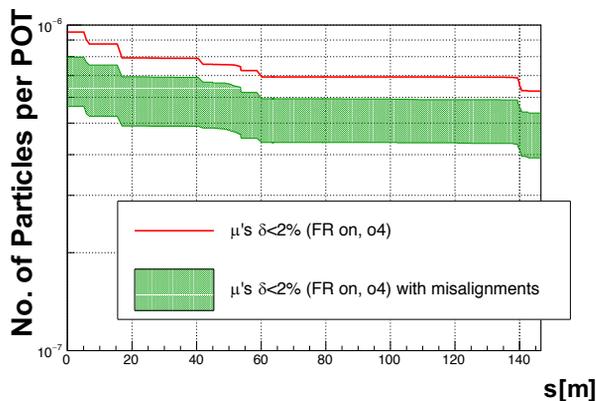


Figure 9: Misalignment effects on beam population at M4/M5 lines.

Figure 9 shows how the misalignments under consideration could decrease the number of muons that make it to the end of M5. However, even though 0.25 mm is a small RMS, the correctors along the beamlines are expected to reduce such detrimental effect.

In addition to the linear description of the beam dynamics, COSY INFINITY allows the computation of higher order effects of the beamline elements [8]. In specific, particle coordinates as defined in COSY are calculated as follows:

$$r_i = \sum_{l_1, l_2, \dots, l_6=0}^{\leq 4} (r_i | x^{l_1} a^{l_2} y^{l_3} b^{l_4} l^{l_5} \delta^{l_6}) x_0^{l_1} a_0^{l_2} y_0^{l_3} b_0^{l_4} l_0^{l_5} \delta_0^{l_6}$$

where the expression in parenthesis indicates the corresponding transport map component, up to fourth-order in our simulations.

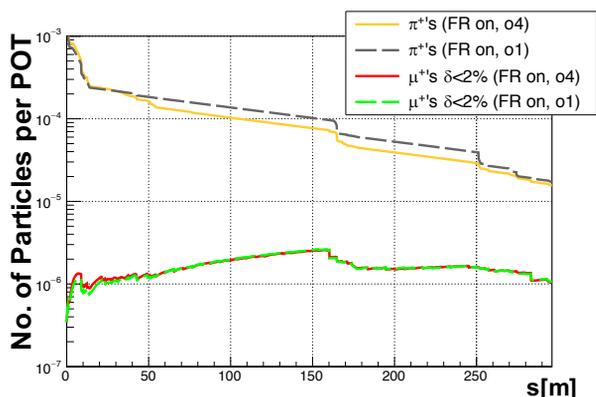


Figure 10: High-order effects on beam population at M2/M3 lines.

As shown in Fig. 10, high-order components seem not to change the statistical performance of the beam.

## CONCLUSION

A detailed model of the Beam Delivery System at E989 has been developed using COSY INFINITY. Realistic features based on DA methods, such as fringe fields, high-order effects, and misalignments are included to describe the statistical and dynamical performance of the secondary beam produced at the pion-production target. Simulation results suggest that fringe fields increase the number of muons that are delivered to the storage ring by  $\sim 5\%$ , whereas the muon beam polarization is unaffected. However, spin-momentum correlations that could add systematic effects to the final measurement of E989 due to differential decays are significantly affected by the fringe fields from the rectangular magnets of the delivery ring. High-order effects do not affect the secondary beam performance and misalignments without correctors were also considered in simulations. Our simulations also served to validate numerical calculations prepared by other members of the Muon  $g-2$  Collaboration and check the performance of the BDS.

## ACKNOWLEDGEMENTS

We are thankful to the Muon  $g-2$  Collaboration at Fermilab, specially W.M. Morse, J.P. Morgan, M. Korostelev, and V. Tishchenko for all the discussions and suggestions. This work was supported by the U.S. Department of Energy under Contract No. DE-FG02-08ER41546. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. We are grateful to the PhD Accelerator Program at Fermilab and also for a Strategic Partnership Grant from the MSU Foundation.

## REFERENCES

- [1] K. Makino and M. Berz, “COSY INFINITY Version 9”, *Nucl. Instr. Meth. A*, vol. 558, pp. 346–350, 2006.
- [2] J. Grange *et al.*, “Muon  $g-2$  Technical Design Report”, Fermi National Accelerator Laboratory, Batavia, IL, USA, FERMILAB-FN-0992-E, Jan. 2015.
- [3] D. Stratakis, M. E. Convery, C. Johnstone, J. Johnstone, J. P. Morgan, D. Still, J. D. Crnkovic, V. Tishchenko, W. M. Morse, and M. J. Syphers, “Accelerator performance analysis of the Fermilab muon campus”, *Phys. Rev. Accel. Beams*, vol. 20, pp. 111003, Nov 2017.
- [4] M. Korostelev *et al.*, “End-to-end beam simulations for the new muon  $g-2$  experiment at Fermilab,” in *Proc. IPAC'16*, Busan Korea, May 2016. doi:10.18429/JACoW-IPAC2016-WEPMW001
- [5] MARS, <http://mars.fnal.gov>
- [6] D. Stratakis, “Recent Results from the Muon Campus runs”,  $g-2$  Experiment Document GM2-doc-13633, Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, USA, 2018.

- [7] D. Stratakis, “Recent Results from the Muon Campus runs”,  $g-2$  Experiment Document GM2-doc-11124, Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, USA, 2018.
- [8] M. Berz and K. Makino, “COSY INFINITY 10.0 Beam Physics Manual”, Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA, Rep. MSUHEP-151103, Oct 2017.
- [9] F. Combley and E. Picasso, The muon ( $g-2$ ) precession experiments: Past present and future, *Phys. Rep.* 14, 1 (1974).
- [10] M. Berz, K. Makino, and W. Wan, *An Introduction to Beam Physics*, CRC Press, Boca Raton, FL, USA 2016.
- [11] J. Crnkovic *et al.*, “Spin Correlations study for the new  $g-2$  Experiment at Fermilab,” in *Proc. IPAC’16*, Busan Korea, May 2016. doi : 10.18429/JACoW-IPAC2016-TUPMR028

# POLARIZATION LIFETIME IN AN ELECTRON STORAGE RING, AN ERGODIC APPROACH IN eRHIC EIC\*

F. Méot, BNL C-AD, Upton, NY, USA

## Abstract

Electron polarization in a storage ring is subject to two long-term effects: Sokolov-Ternov polarization and depolarization by diffusion. Over a long time scale this leads to an equilibrium state and, simulation-wise, can be highly CPU time and memory consuming. Simulations aimed at determining optimal ring storage energy in an electron-ion collider use to track thousand particle bunches, for a long time—yet still short compared to depolarization time scales, due to HPC limitations. Based on considerations of ergodicity of electron bunch dynamics in the presence of synchrotron radiation, tracking a single particle instead is investigated. This allows substantial saving in the required HPC volume, “CPU-time × Memory-allocation”. The concept is illustrated with polarization lifetime and equilibrium polarization simulations at the eRHIC electron-ion collider.

## INTRODUCTION

The eRHIC installation is briefly described in Fig. 1 [1]. The 18 GeV eRHIC electron storage lattice used in the

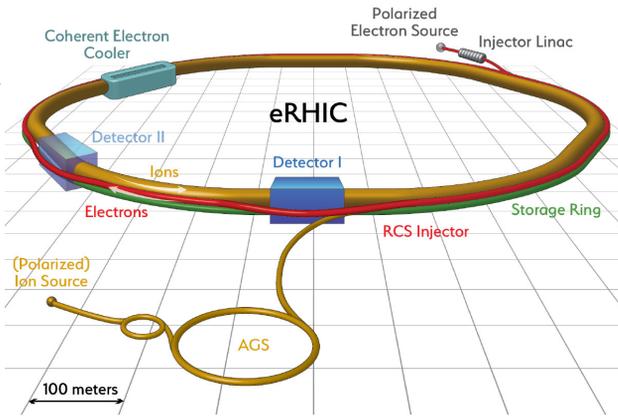


Figure 1: The eRHIC electron-ion collider complex, an 18 GeV–255 GeV/nucleon electron-ion collider installation.

present spin polarization simulations has been provided by S. Tepikian [1], optical parameter values relevant to the present simulations will be introduced in due place. The eRHIC lattice includes a double non-planar rotator system (Fig. 2) at the interaction point (IP), comprised of strong solenoids and series of bends, which allows to locally move the stable spin precession direction  $\vec{n}_0$ , from vertical in the arcs to longitudinal at the IP. In a defect-free ring, this region of off-vertical  $\vec{n}_0$  is a major contribution to spin diffusion.

Bunches are injected in the storage ring with alternately up and down polarization, and replaced every 6 min in order

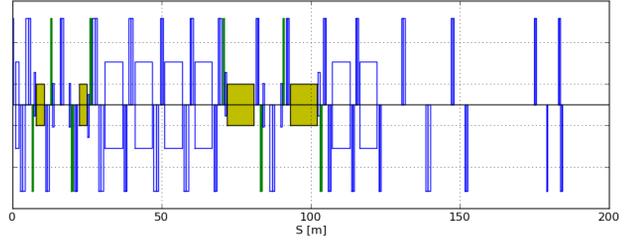


Figure 2: Half of the spin rotator system at eRHIC (the system is symmetric with respect to the IP, at the origin here). In green: solenoids.

to ensure an average polarization of 70% over the hundreds of bunches stored [1]. A proper lattice should maintain bunch depolarization below 20% (absolute) over the 6 min storage. The evolution of the polarization, from  $P_0 = \pm 0.85$  at injection to  $P_{eq}$  at equilibrium (an asymptotic quantity to be determined), satisfies

$$P(t) = P_{eq}(1 - e^{-t/\tau_{eq}}) + P_0 e^{-t/\tau_{eq}}. \quad (1)$$

This results from (i) synchrotron-radiation (SR) self-polarization and (ii) polarization loss by diffusion, with time constant  $\tau_D$ , such that

$$1/\tau_{eq} = 1/\tau_{SP} + 1/\tau_D \quad (2)$$

Sokolov-Ternov (ST) self polarization in a flat ring has a time constant  $\tau_{ST[sec.]} \approx 99 \rho_{[m]}^2 R_{[m]} / E_{[GeV]}^5$  [2], about 30 min at eRHIC at 18 GeV, 10 hrs at 10 GeV, with asymptotic value  $P_{ST} = 92.4\%$ ; the asymptotic self-polarization is taken instead  $P_{SP} = 90\%$  here to account for the non-planar spin rotator, and with time constant  $\tau_{SP}$ , such that [2]

$$P_{eq} = P_{SP} \times \tau_{eq} / \tau_{SP}. \quad (3)$$

The goal in tracking spin motion is (i) to validate a ring design, including preservation of polarization under the effect of defects, corrections, etc. and (ii) to determine an optimal working point  $a\gamma_{ref}$  ( $a = 1.15965 \times 10^{-3}$  is the electron anomalous magnetic moment).

In the following, a method based on single-particle tracking is discussed. First, basic aspects of the stochasticity of particle and spin motions are recalled. Then tracking outcomes are displayed and the single-particle method is discussed.

The numerical simulations discussed in this paper have strongly benefited from NERSC means and environment [3].

## STOCHASTIC MOTION

The dynamical system of a high energy stored electron bunch at equilibrium is ergodic: over a long time interval, trajectories in the system cover all parts of the 6D phase space.

\* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy

Time averages over one or more trajectories are equivalent to phase space averages,

$$\lim_{T \rightarrow \infty} \int_{t_0}^{t_0+T} f(\vec{X}(t)) dt = \int f(\vec{X}) \rho(\vec{X}) d^N \vec{X} \Big|_{\text{time}=t} \quad (4)$$

For all three motions, transverse and longitudinal, the evolution of the bunch emittance with time,  $t$ , satisfies

$$\bar{\epsilon}_n(t) = \epsilon_{n,\text{eq}} \left(1 - e^{-t/\tau_n}\right) + \epsilon_{n,i} e^{-t/\tau_n} \quad (5)$$

( $n$  stands for  $x$ ,  $y$ , or  $l$ ) with  $\epsilon_{n,i}$  and  $\epsilon_{n,\text{eq}}$  respectively the starting and equilibrium emittances,  $\tau_n = \frac{T_{\text{rev}} E_S}{U_S J_n}$  the damping time constant.  $J_{n=x,y,l}$  are the partition numbers,  $J_x + J_y + J_l = 4$ ,  $J_l \approx 2$ . Equation (5) indicates that after a few damping times, the bunch dynamical system can be considered at equilibrium, bunch emittances have reached their asymptotic values. In the following  $\tau_{\text{SR}} = \tau_x \approx \tau_y$  denotes the transverse damping time constant.

At 18 GeV the energy loss amounts to 38.7 MeV/turn (a result from prior tracking of a 2000 particle bunch with Monte Carlo SR), thus the damping time amounts to  $\tau_{\text{SR}} = 18_{\text{GeV}}/38.7_{\text{MeV/turn}} = 465$  turns, 6 ms.

Figures 3 and 4 display the stochastic motion of a single particle over  $10^3 \tau_{\text{SR}}$  and by comparison the instantaneous horizontal and vertical phase spaces of a  $10^3$ -electron bunch observed at  $time = 10^3 \tau_{\text{SR}}$ . In this example, statistical variable values such as *rms* coordinates, emittances, either single particle projected over a long tracking time, or multiparticle at time  $t = 10^3 \tau_{\text{SR}}$ , resulting from both methods, essentially satisfy Eq. (4). Over a sufficiently long time interval, an electron has explored the all 6D phase-space, which is a necessary condition for ergodicity to be satisfied.

Out of equilibrium,  $\vec{X}(t) - \bar{\vec{X}}(t)$  can be taken as the statistical variable, with  $\bar{\vec{X}}(t)$  the average value.  $\bar{\vec{X}}(t)$  can be determined from a fit using the theoretical damping, for instance in the single particle case, see below. Spin motion is not at equilibrium, the polarization decays with time, fast in resonant conditions. Both the decay time constant and the asymptotic polarization are zero on the resonance, as  $\tau_{\text{eq}} \sim \delta^2 \times \tau_{\text{SP}}$ ,  $P_{\text{eq}} \sim \delta \times P_{\text{SP}}$ , with  $\delta = a\gamma_{\text{Res}}$ .  $-a\gamma$  the distance to the resonance [2, p. 125]. Figure 5 displays typical stochastic spin motion in eRHIC storage ring at 18 GeV. In a similar way that  $\tau_{\text{SR}}$  can be obtained from the observation of the damped motion of a single electron far from equilibrium,  $\tau_{\text{D}}$  can be obtained from long enough observation of spin motion.

## POLARIZATION

In order to assess polarization properties of a storage ring depending on its energy setting, spin tracking simulations are performed over an ensemble of  $a\gamma_{\text{ref}}$  rings covering some  $a\Delta\gamma_{\text{ref}} < 1$  range of interest. In these hypothesis, all these rings have the same optics: bend strengths  $1/\rho$ , focusing strengths  $G/B\rho$ , chromaticities, etc., are unchanged, what changes is the energy  $a\gamma_{\text{ref}}$  at which each ring is run.

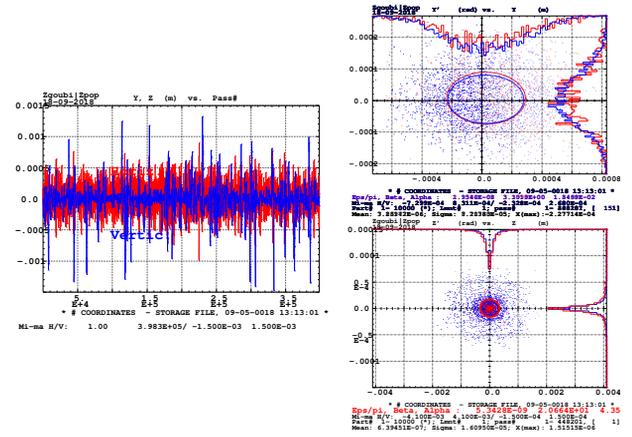


Figure 3: Left: transverse particle excursion, over time interval  $t/\tau_{\text{SR}} : 1 \rightarrow 10^3$  (450,000 turns about). Right: transverse phase spaces, matching ellipses and histograms; blue: projection of the single particle motion of the left plot; red: for comparison, case of a  $10^3$  particle bunch, observed at time  $t = 10^3 \tau_{\text{SR}}$ . Note that the  $\epsilon_y/\epsilon_x$  ratio represents a 27% coupling, of which the source is the spin rotator in IR6 which includes solenoids (Fig. 2).

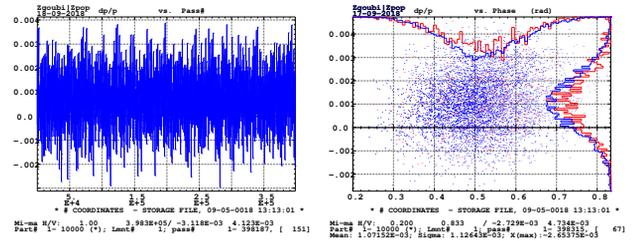


Figure 4: Left: stochastic energy excursion over time interval  $t/\tau_{\text{SR}} : 1 \rightarrow 10^3$ .  $\sigma_{\delta p/p} = 1.14 \times 10^{-3}$ . Right: longitudinal phase space; red: case of a  $10^3$  particle bunch observed at  $t = 10^3 \tau_{\text{SR}}$ ;  $\sigma_{\delta p/p} = 1.13 \times 10^{-3}$ ; blue: projection of the multiturn single particle motion of Fig. 4.

Particles are all launched with their initial spin direction parallel to the local nominal stable spin precession direction  $\vec{n}_0$  (*i.e.*, longitudinal if the origin is taken at IP6, vertical at IP8). Spin tracking only starts after 10 damping times about (5,000 turns) when the bunch has reached its equilibrium emittance. If the motion happens to neighbor a depolarizing resonance, spin will tilt away from the nominal direction toward possibly large angles depending on the strength of/distance to the resonance (in the absence of SR and at constant energy, the spin would steadily rotate around the local tilted  $\vec{n}_\delta$ ). Away from any resonance, the spin is expected to only be subject to slow diffusion.

The single particle “depolarization landscape” is expected to look as shown in Fig. 6, obtained in HERA-e conditions, which include a spin rotator which introduces strong depolarization in the  $a\gamma = \text{integer}$  regions. Single particle tracking here yields Figure 7 (18 GeV) and Fig. 8 (10 GeV), which appears qualitatively similar to DESY simulation outcomes, Fig. 6. This is the outcome of the tracking, over a time

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

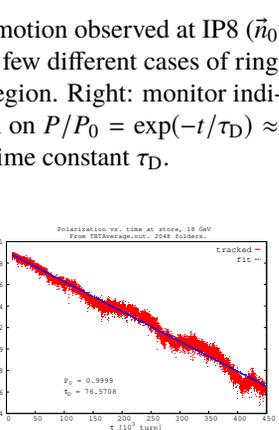
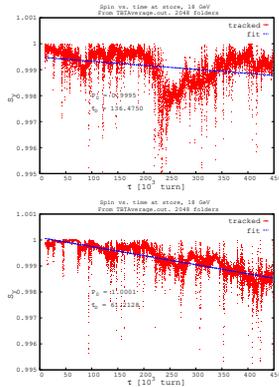
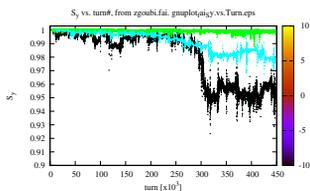


Figure 5: Left: stochastic spin motion observed at IP8 ( $\vec{n}_0$  vertical there), single particle, a few different cases of ring rigidity settings in the 18 GeV region. Right: monitor individual spins, a linear regression on  $P/P_0 = \exp(-t/\tau_D) \approx 1 - t/\tau_D$  provides the diffusion time constant  $\tau_D$ .

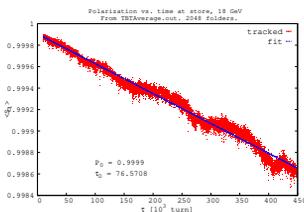
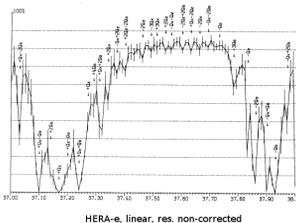


Figure 6: Left: asymptotic polarization at HERA, using SITROS [4]. This graph was produced by tracking bunches of a few hundred polarized electrons, including Monte Carlo SR, over a few SR damping times, for a series of evenly spaced ring rigidity values over an  $a\Delta\gamma = 1$  interval ( $37 \leq a\gamma \leq 38$ ). Each  $a\gamma$  value represents a particular operation rigidity of the ring, however with unchanged optics. Right, obtained from the present eRHIC simulations: turn-by-turn average spin value over the complete bunch population yields the self-polarization time constant, from what the asymptotic polarization, similarly to the representation in the left plot, can be drawn (average over just a few particles is shown here).

interval  $[0, t]$ , of a single particle in each one of 1024 (or 2048) rings, all operated with the same optics but with each its particular operation energy  $a\gamma_{ref}$ . These distributions feature similar topology, at both timings. Zooming in on any reduced  $a\Delta\gamma_{ref}$  interval also shows a similar spin distribution (sort of “fractal”). The energy excursion over that time interval  $[0, t]$  is displayed in Fig. 9. Note a property that will be referred to later: the equilibrium energy spread is  $\sigma_E \approx 10^{-3}E$  at 18 GeV, or an extent  $\sigma_{a\gamma} \approx 0.04$ , thus the beam covers  $\approx 40$  (80) bins of a 1024 (2048) bin  $a\Delta\gamma_{ref} = 1$  interval.

Spin diffusion has to be a slow process for a lattice to be viable, in particular this cannot be the case if, during its energy excursion, a particle neighbors (is within a few resonance strengths from) a depolarizing spin resonance ( $\nu_s \pm l\nu_x \pm m\nu_y \pm n\nu_l = \text{integer}$ ). In such case, the de-

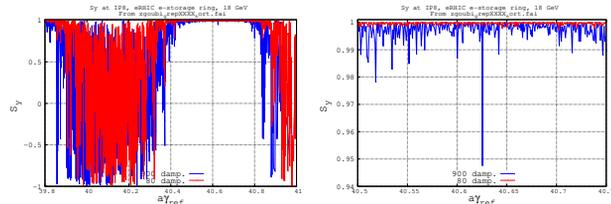


Figure 7: Spin rotation landscape (a photo of the spins,  $S_l(a\gamma_{ref})|_{\text{time}=t}$  at IP8 where  $\vec{n}_0$  is vertical), at either (red)  $t = 80\tau_{SR}$  ( $4 \times 10^4$  turns) or (blue)  $t = 900\tau_{SR}$  ( $4.5 \times 10^5$  turns or 5 s). Left: the 2048 rings tracked cover over a  $a\Delta\gamma = 1.2$  interval, encompassing integer  $a\gamma_{ref}$  values where full spin flip is observed, as expected. Right: a zoom in on a reduced  $40.5 < a\Delta\gamma_{ref} < 40.75$  interval; no strong resonances in that region instead, and spins remain close to  $S_y = 1^-$ .

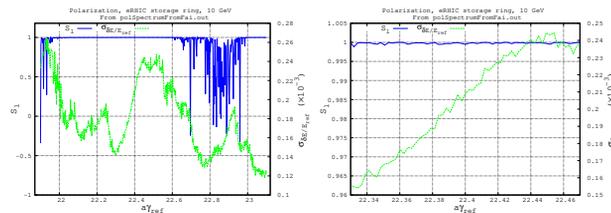


Figure 8: Spin rotation landscape (a photo of the spins,  $S_l(a\gamma_{ref})|_{\text{time}=t}$  at IP6 where  $\vec{n}_0$  is longitudinal), at 160 SR damping times ( $\approx 5 \times 10^5$  turns). The right and left plots differ by the  $a\Delta\gamma$  interval width. Integer  $a\gamma_{ref}$  regions show full spin flip as expected. In the reduced  $22.33 < a\Delta\gamma_{ref} < 22.47$  region no strong resonance is observed, spins remain close to  $S_y = 1^-$ . The right vertical scale in both plots is the *rms* width of the energy interval explored by a particle during the tracking.

polarization is not slow (the orientation of the spin vector changes substantially during the tracking: the vertical spin component  $S_y$  moves towards  $S_y = -1$  in the present representation, Fig. 5). Thus, the working point of concern,  $a\gamma_{ref}$  near resonance, is not optimal. *A contrario*, observing only slow change in the spin vector means absence of harmful resonance in the energy interval that the particle spans due to SR, and potentially a viable working point.

### Diffusion Time Constant

Spin tracking here does not include the self-polarization process, it is assumed that  $\tau_{SP}$  in Eq. (2) is obtained from the lattice. Thus, a bunch keeps depolarizing, due to diffusion, polarization tends toward  $P_{eq} = 0$  ( $\tau_{SP} = \infty$ ) with a time constant  $\tau_D = \tau_{eq}$ .

In order to ensure the required polarization survival (70% on average over the ensemble of bunches in the ring, stored 6 min each)  $\tau_D$  has to be sufficiently long compared to the store duration; this eliminates, for a viable rigidity setting of the ring, the regions  $a\gamma_{ref} < 40.45$  and  $> 40.75$  in Fig. 7 (18 GeV),  $a\gamma_{ref} < 22$  and  $> 22.4$  in Fig. 8 (10 GeV). Finally, with  $\tau_D$  much larger than the time interval covered by the tracking (of the order of seconds at best, whereas  $\tau_D$  has to

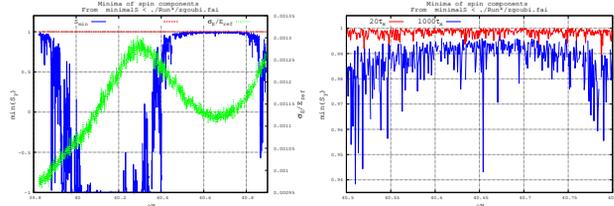


Figure 9: Left, blue curve: depolarization landscape  $S_{y,\min}(a\gamma_{\text{ref}})|_{t \in [0,T]}$ , over  $39.8 < a\gamma_{\text{ref}} < 40.9$  (18 GeV region), observed at IP8 (nominal  $\bar{n}_0$  is vertical there) at time  $T = 10^3 \tau_{\text{SR}}$ . This is the smallest value of the projection of particle spin  $\vec{S}$  on the vertical axis, reached over the time interval  $[0, T]$ . The right vertical scale (green curve) is the *rms* width of the energy interval explored by a particle during the tracking. Right: a zoom in on a reduced  $40.5 < a\delta\gamma_{\text{ref}} < 40.8$  interval, showing  $S_{y,\min}(a\gamma_{\text{ref}})|_{t \in [0,T]}$  at (red)  $T = 20\tau_{\text{SR}}$  (9500 turns) and (blue)  $T = 1000\tau_{\text{SR}}$  (450,000 turns or 6 s).

amount to tens of minutes for a lattice to be viable), one can use

$$P(t)/P_0 = \exp(-\frac{t}{\tau_D}) \approx 1 - t/\tau_D. \quad (6)$$

Single particle spin tracking data are displayed in Fig. 5, a fit of these data provides  $\tau_D$ . From that  $\tau_{\text{eq}}$  can be derived (Eq. (2)) yielding in turn  $P_{\text{eq}}$  (Eq. (3)) and  $P(t = 6 \text{ min})$  (Eq. (1)).

In order to assess the method, in the following for simplicity, and  $P_{\text{eq}}$  being a sub-product, primary tracking outcomes are considered, namely, spin orientation or  $\tau_D$  landscapes.

## A METRIC

Typically, the energy dependence of particle spins over a  $a\Delta\gamma_{\text{ref}}$  interval looks as shown in Figs. 7 and 8. A different criterion to quantify the depolarization could be instead, Fig. 9, the energy dependence of the minimal value that spins reached in the course of the tracking. This minimum may happen earlier in the tracking, as observed in Fig. 5, as spins oscillate around a given tilted local  $\bar{n}_\delta$ , as long as the latter does not change due to photon emission.

In order to allow comparisons between lattices a metric is required. However, Figs. 7 and 8 styles of data do not lend themselves to straightforward comparisons, essentially due to the stochastic aspect. A couple of different possibilities are assessed here instead, based on sliding averaging.

## Spin

A sliding average of the data of Fig. 7, with a small sliding interval  $a\delta\gamma_{\text{ref}}$  (a few particles / bins), namely

$$\langle S_y \rangle_N(a\gamma_{\text{ref}}, \frac{N}{2}) = \frac{1}{N} \sum_i^{i+N-1} S_y(a\gamma_{\text{ref},i}), \quad (7)$$

greatly smooths the fluctuations, as observed in Fig. 10. The local excursion of  $S_y$  over a small  $\delta\gamma_{\text{ref}}$  interval in the optimal region ( $P_f/P_i \approx 0.9983$ ) are grossly below  $\pm 2 \times 10^{-4}$ . This

corresponds to a fluctuation of  $\tau_D$  of less than  $\pm 7$  min around an average  $\approx 60$  min,  $\approx \pm 10\%$  relative, a good first approach. This can be further improved by increasing the number of bins, for instance, once determined that the region  $40.6 < a\gamma_{\text{ref}} < 40.7$  is viable, the computation can be reiterated.

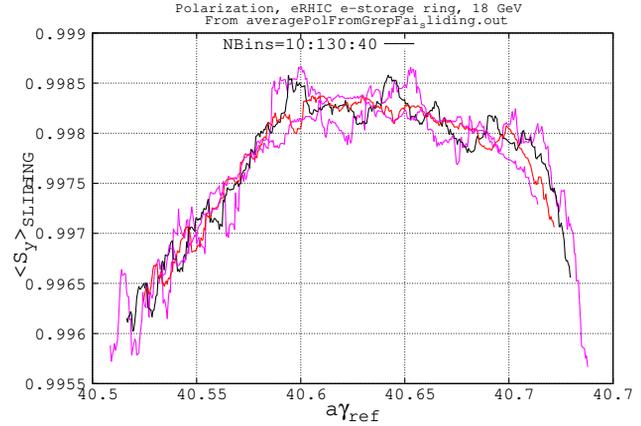


Figure 10: Derived from Fig. 7, by applying a sliding average (Eq. (7)). The four curves differ by the sampling: average over  $N =$  either 10, 50, 90 or 130  $a\delta\gamma_{\text{ref}}$  samples, centered at given  $a\gamma_{\text{ref}}$ . In the present conditions, the distribution converges when increasing the number of samples,  $N$ , i.e., the width of the sliding interval,  $[a\delta\gamma_{\text{ref},i}, a\delta\gamma_{\text{ref},i+N}]$ .

Why allow a sliding average:

1. with the present extent  $a\Delta\gamma_{\text{ref}} \approx 1$  covered in 1024 bins (or 2048), the energy extent of the bunch  $\sigma_{\gamma_{\text{ref}}}/\gamma_{\text{ref}} \approx 10^{-3}$  or  $a\delta\gamma_{\text{ref}} = 0.04$ , covers about 40 (or 80) bins,
2. thus, a set of a few neighboring bins almost belong in the same ring, averaging over a few bins is not so different from averaging over a few particles in the same bin,
3. in any case a possible strong, nearby resonance would cause a dip in the distribution, indicating a non viable  $a\delta\gamma_{\text{ref}} = 1$  region, which has to be avoided.

The strong smoothing effect of a sliding average suggest to apply it directly to the final spin distribution, this is done in Fig. 10, case of a  $\Delta\gamma$ :  $40.4 \rightarrow 40.9$  interval covered in 1024 bins. In the present conditions, this series converges when increasing the width of the sliding interval  $[a\delta\gamma_{\text{ref},i}, a\delta\gamma_{\text{ref},i+N}]$ . At some point however, increasing the sliding interval would cause it to reach  $a\gamma_{\text{ref}}$  regions where the fluctuations change in a sensible manner (as in Fig. 7, left), for instance featuring a different average, or including high amplitude spikes, so abruptly changing the sum of the series, however the eRHIC lattices of interest have to satisfy  $\langle S_y \rangle(a\gamma_{\text{ref}}) \approx 1$ , which prohibits such changes. In the present case of 1024 ring samples over  $\Delta\gamma$ :  $40.4 \rightarrow 40.9$ ,  $N = 40$  appears appropriate; this is a sliding window of full width  $a\Delta\gamma_{\text{ref}} = 40 \times (40.9 - 40.4)/1024 \approx 0.02$ . Figure 10

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

confirms that, with the 40.59–40.63 interval yielding a final  $\tau_D \approx 55\%$  within about  $\pm 5\%$  of convergence values for both 40.60–40.62 and 40.56–40.66 sliding windows. It may also be thought of increasing the bins density in the  $a\gamma_{\text{ref}}$  of interest to obtain a better homogeneous distribution.

### Diffusion Time Constant

The diffusion time constant is derived from single particle motion using a linear regression (Eq. (6)), an illustration can be found in Fig. 5 (right). Applying to each of the 2048 rings over the  $a\Delta\gamma_{\text{ref}}$  interval, yields the scan in Fig. 11. Tracking over several SR damping times is needed to draw  $\tau_D$  from individual spin motion, It can be seen in Fig. 11 that, in the region of  $\tau_D$  values of interest the statistics over 80, 160 and  $10^3 \tau_{\text{SR}}$  superimpose. This indicates that the required tracking time is comparable with that needed to determine  $\tau_{\text{SR}}$  from particle motion, which is a few tens of damping times or less. On the other hand, in case the strong fluctuations of the spin would cause too strong a dependence of the  $\tau_{\text{SR}}$  value (from the fit) on the fit sample, rather than increasing the damping time a possibility is instead to launch a few particles per ring: the smoothing effect is immediate, this can be seen by comparison of the spin motion in Fig. 5 (right) and the averaging over a few particles in Fig. 6 (right). These considerations matter as to the interest of the single particle method, this is discussed in the next section.

Again a sliding average, applied to the data of Fig. 11, greatly smooths the fluctuations, as observed in Fig. 12. The distance between  $\tau_D$  distributions can be derived from Fig. 12 type of data, which are thus usable for comparing polarization performances of different versions of the eRHIC e-storage ring.

## SUMMARY

Assume similar resolution using both methods, “HPC-Hungry” and “Ergodic”, namely, the same number of reference rings, nRings, over the same interval  $a\Delta\gamma$ .

In the present hypotheses (eRHIC lattice, energy, etc.):

- first method: the HPC volume is  $n\text{Rings} \times 10^3$  [particles/bunch]  $\times$  a few  $\tau_{\text{SR}}$ ,
- second method: the HPC volume is  $n\text{Rings} \times$  a few  $\tau_{\text{SR}}$ .

This is a 3 orders of magnitude difference in the HPC volume. On the other hand, greater HPC volume translates in one or the other of, longer queues, longer computing time, more processors, greater volume of I/Os, larger data analysis HPC volume.

It remains to determine how close the single particle method can get to the accuracy of the bunch method (an ongoing work). However it already appears an efficient first approach to the diffusion time constant, in view of qualifying an evolution of a lattice design, the efficiency of error correction and other spin matching schemes. Because it is faster it allows a greater reach (for instance in terms of

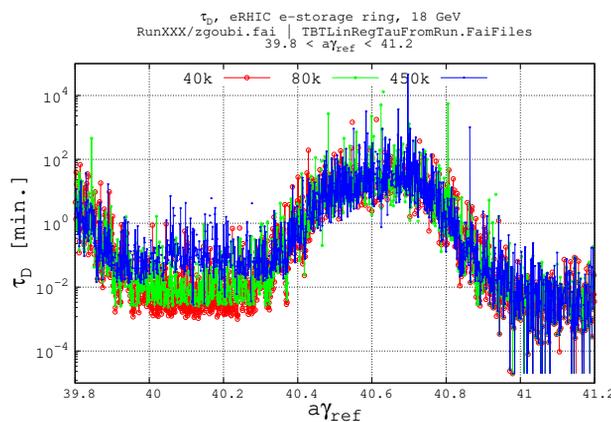


Figure 11: Energy scan of the diffusion time constant in the single particle tracking method. 2048 bins cover  $a\Delta\gamma$ : 39.8–41.2.  $\tau_D$  values interpolated from 80 (red), 160 (green) and  $10^3 \tau_{\text{SR}}$  tracking.

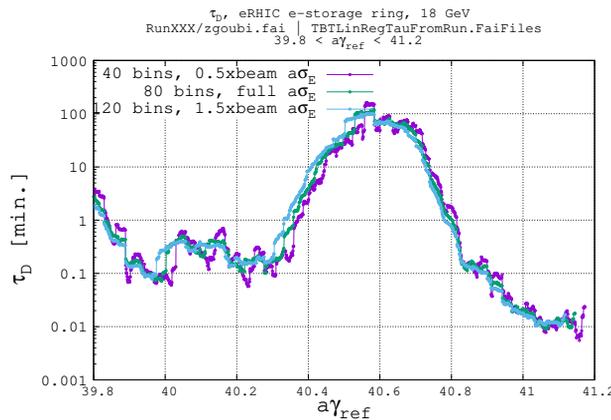


Figure 12: Sliding over energy scan of the diffusion time constant in the single particle tracking method. 2048 bins.

parameter space exploration) in machine simulations and design optimizations.

## CONCLUSION

Obviously these results are very preliminary, they are essentially indications that the HPC volume could be reduced. More simulations are required, for further inspection, comparisons between the two methods, etc. Mathematical background and support is in order.

The simulations discussed here were performed on NERSC [3], using stepwise ray-tracing tools for spin motion accuracy [5, 6]. Electron dynamics and spin diffusion in the presence of Monte Carlo SR is a long installed and, needless to say, thoroughly benchmarked feature of the code [7].

## ACKNOWLEDGEMENTS

This study has benefited from fruitful discussions with D. Abell (Radiasoft), A. Cerfon (CIMS-NYU) and B. Nash (Radiasoft).

## REFERENCES

- [1] eRHIC p-CDR, BNL, 2018.
- [2] S.Y. Lee, *Spin dynamics and snakes in synchrotrons*, World Scientific, 1997.
- [3] NERSC computing, <http://www.nersc.gov/>
- [4] J. Kewisch, "Depolarisation der Elektronenspins in Speicherringen durch nichtlineare Spin-Bahn-Kopplung", DESY 85-109, Okt. 1985.
- [5] [https://zgoubi.sourceforge.io/ZGOUBI\\_DOCS/Zgoubi.pdf](https://zgoubi.sourceforge.io/ZGOUBI_DOCS/Zgoubi.pdf)
- [6] <https://www.osti.gov/scitech/biblio/1062013-zgoubi-users-guide>
- [7] F. Méot, "Simulation of radiation damping in rings, using stepwise ray-tracing methods", *JINST*, vol. 10, pp. T06006, 2015.

# SPIN DYNAMICS IN MODERN ELECTRON STORAGE RINGS: COMPUTATIONAL ASPECTS

Oleksii Beznosov\*, James A. Ellison, Klaus A. Heinemann, UNM, Albuquerque, NM, USA  
 Desmond P. Barber<sup>1</sup>, DESY, Hamburg, Germany  
 Daniel Appelö, University of Colorado, Boulder, CO, USA  
<sup>1</sup> also at UNM, Albuquerque, NM, USA

## INTRODUCTION

In [1] we report on our spin/polarization project for understanding the possibility of polarization for the next generation of the high energy particle (HEP) accelerators, e.g., the Future Circular Collider (FCC) and Circular Electron Positron Collider (CEPC). The physics background and the basic model to compute the polarization is discussed there. The starting point is what we call the full Bloch equation (FBE) in the Lab frame. This model includes synchrotron radiation and the concomitant depolarization from the radiation caused by damping and diffusion as well as Sokolov-Ternov spin-flip polarization effects and its Baier-Katkov generalization. Ignoring spin flip we obtain the reduced Bloch equation (RBE) which we believe contains the most difficult part of the FBE to integrate numerically. We then introduce the 3 degree of freedom (DOF) reduced Bloch equation (RBE) in the beam frame in the first section below. We further discuss the general computational issues and give an estimates for what can be done with current computational techniques. For  $d = \{1, 2, 3\}$  DOF the polarization density has  $(2d + 1)$  independent variables. For simplicity, suppose that each of the space-like variables has been discretized on a grid with  $N$  grid-points, then the computational cost of each time step will scale no better than  $O(N^{2d})$ . The presence of parabolic terms in the governing equations necessitates implicit time stepping and thus solutions of linear systems of equations. For a fully coupled 3 DOF problem this will bring the per time step cost to  $O(N^{6q})$ , with  $1 \leq q \leq 3$ , depending on the algorithms used for the linear solve. However, only algorithms with  $q \approx 1$  are feasible (for Gaussian elimination  $q = 3$ ). Fortunately, as we outline below, the structure of the averaged equations (e.g the parabolic terms are uncoupled from mode coupling terms) allows the efficient parallel implementation. Further, we exploit the decoupling by evolving the resulting ODE system with the additive Runge-Kutta (ARK) method. Described in [2] ARK methods are high order semi-implicit methods that are constructed from a set of consistent Runge-Kutta (RK) methods. In the RBE the parabolic part of the equation is treated with a diagonally implicit RK method (DIRK) and the hyperbolic mode coupling part is treated with an explicit RK (ERK) method which does not require a linear solve. The ODE system in time can be evolved independently for each Fourier mode resulting in a computational cost for each timestep that scales as  $O(N^{3q})$  per mode.

We first summarize the 3 DOF problem and 2 DOF problem from [1]. Then we describe the new algorithm on the

\* Corresponding author: obeznosov@unm.edu

example of 1 DOF model with parameters taken from the Hadron-Electron Ring Accelerator (HERA). Using that the RBE in 1 DOF can be solved exactly we demonstrate the accuracy of the algorithm by comparing the exact polarization to the polarization measured by integrating the numerical solution in space. Further, we present the results showing that achieved accuracy of the algorithm for the polarization density after 1500 turns for varying discretization parameters which allows us to conclude that the algorithm is feasible for the accurate simulation of the 3 DOF model.

## RBE IN 3 DEGREES OF FREEDOM

Consider the system of Langevin equations for the orbital phase space variable  $Y \in \mathbb{R}^6$  and the spin variable  $\vec{S}$  in the beam frame given by

$$Y' = \mathcal{A}(\theta)Y + \sqrt{\varepsilon}\sqrt{\omega(\theta)}e_6\xi(\theta), \quad (1)$$

$$\vec{S}' = \Omega_Y(\theta, Y)\vec{S}. \quad (2)$$

Here  $\theta$  is the accelerator azimuth and  $\xi$  is a version of the white noise process and  $e_6 = (0, 0, 0, 0, 0, 1)^T$ . Also  $\mathcal{A}(\theta)$  is a  $6 \times 6$  matrix encapsulating radiationless motion and the deterministic effects of synchrotron radiation (see, e.g., [3, eq. 5.3]). Moreover  $\Omega_Y(\theta, Y)$  in the Thomas-BMT term is a skew-symmetric  $3 \times 3$  matrix linear in  $Y$  and  $\omega(\theta)$  is real valued. Note also that  $\mathcal{A}(\theta)$ ,  $\Omega_Y(\theta, y)$  and  $\omega(\theta)$  are  $2\pi$ -periodic in  $\theta$ .

The RBE for the polarization density  $\vec{\eta}_Y$  is

$$\partial_\theta \vec{\eta}_Y = (L_Y + L_{Y, TBM T})\vec{\eta}_Y, \quad (3)$$

where

$$L_Y = - \overbrace{\sum_{j=1}^6 \partial_{y_j} \left( \mathcal{A}(\theta)y \right)_j}^{\text{Drift}} + \overbrace{\frac{1}{2} \omega_Y(\theta) \partial_{y_6}^2}^{\text{Diffusion}},$$

$$L_{Y, TBM T} \vec{\eta}_Y = \underbrace{\Omega_Y(\theta, y) \vec{\eta}_Y}_{\text{Spin}}.$$

Our ultimate aim is to understand the solutions of (3). The main quantity of interest is the polarization of the bunch

$$\vec{P}(\theta) = \int \vec{\eta}_Y(\theta, y) dy.$$

However, as noted in the introduction, numerical discretization of (3) will have the an enormous computational cost. To simplify the problem we first use the method of averaging.

We split  $\mathcal{A} = A + \varepsilon\delta A$  to isolate the Hamiltonian part  $A$ . Then by using the fundamental solution matrix  $X(\theta)$  of the unperturbed problem, e.g

$$X' = A(\theta)X, \quad (4)$$

the method averaging  $Y$ -frame transforms to a  $V$ -frame associated with averaged problem posed in terms of the new variable  $V$ . In the  $V$ -frame the polarization density  $\eta_V$  satisfies the RBE

$$\partial_\theta \vec{\eta}_V = (L_V + L_{V,TBMT})\vec{\eta}_V, \quad (5)$$

where

$$L_V = -\varepsilon \overbrace{\sum_{j=1}^6 \partial_{v_j} (\bar{D}v)_j}^{\text{Drift}} + \frac{\varepsilon}{2} \overbrace{\sum_{i,j=1}^6 \bar{E}_{ij} \partial_{v_i} \partial_{v_j}}^{\text{Diffusion}}, \quad (6)$$

$$L_{V,TBMT} \vec{\eta}_V = \underbrace{\Omega_Y(\theta, X(\theta)v)}_{\text{Spin}} \vec{\eta}_V. \quad (7)$$

and

$$\bar{D} = \begin{pmatrix} \mathcal{D}_I & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & \mathcal{D}_{II} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & \mathcal{D}_{III} \end{pmatrix}, \quad (8)$$

$$\mathcal{D}_\alpha = \begin{pmatrix} a_\alpha & b_\alpha \\ -b_\alpha & a_\alpha \end{pmatrix}, \quad (\alpha = I, II, III), \quad (9)$$

with  $\bar{E} = \text{diag}(\mathcal{E}_I, \mathcal{E}_I, \mathcal{E}_{II}, \mathcal{E}_{II}, \mathcal{E}_{III}, \mathcal{E}_{III})$  and  $a_\alpha \leq 0$  and  $\mathcal{E}_I, \mathcal{E}_{II}, \mathcal{E}_{III} \geq 0$ . The RBE in the  $V$ -frame has  $\theta$ -independent uncoupled parabolic operators and that will be exploited by our numerical approach.

## RBE IN 2 DEGREES OF FREEDOM. FLAT RING

We now consider the case of two degrees of freedom in a flat ring with FODO cells and cavities. Moreover the case of a flat ring allows us to use a one-dimensional approach to spin which in turn allows us to average over orbit *and* spin. In our flat ring model  $\Omega_Y$  has the simple form

$$\Omega_Y(\theta, Y) = -a_Y(\theta)Y\mathcal{J}, \quad \mathcal{J} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Here  $Y \in \mathbb{R}^4$  represents the horizontal and longitudinal motions which are uncoupled from the vertical motion in the flat ring model. It is convenient to use spherical coordinates as spin variables (i.e.,  $\vec{S} = (\cos(\Psi) \sin(\Phi), \sin(\Psi) \sin(\Phi), \cos(\Phi))^T$ ). In the  $Y$ -frame system of Langevin equations then becomes

$$Y' = (A(\theta) + \varepsilon\delta A(\theta))Y + \sqrt{\varepsilon\omega(\theta)}(0, 0, 0, 1)^T \xi(\theta), \quad (10)$$

$$\Psi' = a_Y(\theta)Y, \quad (11)$$

$$\Phi' = 0. \quad (12)$$

Here the row vector  $a_Y(\theta)$  is  $2\pi$ -periodic in  $\theta$ .

Following the approach outlined in previous section, we apply the method averaging and transform the current frame to a  $W$ -frame associated with an averaged problem posed in terms of a new variable  $W$ .  $W$  now incorporates both spin and phase space variable. In the  $W$ -frame the polarization density  $\vec{\eta}_W$  satisfies the RBE

$$\begin{aligned} \partial_\theta \vec{\eta}_W = & -\varepsilon \sum_{j=1}^2 \partial_{w_j} \left( \left( \mathcal{D}_I(w_1, w_2)^T \right)_j \vec{\eta}_W \right) \\ & -\varepsilon \sum_{j=3}^4 \partial_{w_j} \left( \left( \mathcal{D}_{II}(w_3, w_4)^T \right)_j \vec{\eta}_W \right) \\ & + \frac{\varepsilon}{2} \mathcal{E}_I \left( \partial_{w_1} \partial_{w_1} + \partial_{w_2} \partial_{w_2} \right) \vec{\eta}_W \\ & + \frac{\varepsilon}{2} \mathcal{E}_{II} \left( \partial_{w_3} \partial_{w_3} + \partial_{w_4} \partial_{w_4} \right) \vec{\eta}_W \\ & -\varepsilon \sum_{j=1}^4 \bar{D}_{5j} w_j \mathcal{J} \vec{\eta}_W - \frac{\varepsilon}{2} \bar{E}_{55} \vec{\eta}_W + \varepsilon \sum_{j=1}^4 \bar{E}_{j5} \mathcal{J} \vec{\eta}_W, \end{aligned} \quad (13)$$

where

$$\bar{D} = \begin{pmatrix} \mathcal{D}_I & 0_{2 \times 2} & 0_{2 \times 2} \\ \bar{D}_{51} & \bar{D}_{52} & \bar{D}_{53} & \bar{D}_{54} & 0_{1 \times 2} \\ 0_{1 \times 2} & 0_{1 \times 2} & 0_{1 \times 2} & 0_{1 \times 2} & 0 \end{pmatrix}, \quad (14)$$

$$\bar{E} = \begin{pmatrix} \mathcal{E}_I & 0 & 0 & 0 & \bar{E}_{15} & 0 \\ 0 & \mathcal{E}_I & 0 & 0 & \bar{E}_{25} & 0 \\ 0 & 0 & \mathcal{E}_{II} & 0 & \bar{E}_{35} & 0 \\ 0 & 0 & 0 & \mathcal{E}_{II} & \bar{E}_{45} & 0 \\ \bar{E}_{15} & \bar{E}_{25} & \bar{E}_{35} & \bar{E}_{45} & \bar{E}_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (15)$$

Here  $\mathcal{D}_I, \mathcal{D}_{II}$  are  $2 \times 2$  matrices of the form (5) and  $\mathcal{E}_I, \mathcal{E}_{II}$  are nonnegative. For Gaussian processes associated with (10)-(12) the polarization density  $\vec{\eta}_W$  can be computed analytically.

## RBE IN 1 DEGREE OF FREEDOM

We now consider the case of one degree of freedom using the model studied in [4,5]. The one degree of freedom model here is obtained from the two degrees of freedom flat ring model of the previous section by setting, in (14) and (15),

$$\begin{aligned} 0 = \mathcal{D}_{II} = \bar{D}_{52} = \bar{D}_{53} = \bar{D}_{54} = \mathcal{E}_{II} = \bar{E}_{25} = \bar{E}_{35} = \bar{E}_{45}, \\ \mathcal{D}_I = -I_{2 \times 2}, \quad \mathcal{E}_I = 1, \quad \bar{E}_{15} = -\bar{D}_{51}, \quad \bar{E}_{55} = (\bar{E}_{15})^2. \end{aligned} \quad (16)$$

One can justify the step from (14) and (15) to (16) as a good approximation by applying the betatron-dispersion formalism to the flat ring model [6]. With (16) the variables  $W_3, W_4, W_6$  are uncoupled so that we are left with the following one degree of freedom system of Langevin equations for

the orbital variables  $W_1, W_2$  and the spin variable  $W_5$ :

$$\begin{pmatrix} W'_1 \\ W'_2 \\ W'_5 \end{pmatrix} = \varepsilon \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ g & 0 & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_5 \end{pmatrix} + \sqrt{\frac{\varepsilon}{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -g & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

where  $g = \bar{\mathcal{D}}_{51} = -\bar{\mathcal{E}}_{15}$  and  $\xi_1, \xi_2$  are statistically independent versions of the white noise process. Denoting the polarization density for our one degree of freedom model by  $\vec{\eta}$ , one can show in analogy to the previous section that it satisfies the RBE

$$\begin{aligned} \partial_\theta \vec{\eta} &= \varepsilon \left( \partial_{w_1} (w_1 \vec{\eta}) + \partial_{w_2} (w_2 \vec{\eta}) \right) + \frac{\varepsilon}{4} \partial_{w_1} \partial_{w_1} \vec{\eta} \\ &+ \frac{\varepsilon}{4} \partial_{w_2} \partial_{w_2} \vec{\eta} - \varepsilon g w_1 \mathcal{J} \vec{\eta} - \frac{\varepsilon}{2} g \mathcal{J} \partial_{w_1} \vec{\eta} - \frac{\varepsilon}{4} g^2 \vec{\eta}, \end{aligned} \quad (17)$$

where  $\varepsilon \approx 0.008$  and  $g \approx 2.07$  for the HERA ring. We present the numerical approach to solve (17) next.

## NUMERICAL APPROACH

We first transform (17) to polar coordinates using

$$\begin{aligned} w_1 &= r \cos \varphi, \quad w_2 = r \sin \varphi, \\ \nabla \cdot (W \eta_l) &= (2 + r \partial_r) \eta_l, \\ \nabla \cdot \nabla \eta_l &= (\partial_r^2 + r^{-1} \partial_r + r^{-2} \partial_\varphi^2) \eta_l, \\ \partial_{w_1} \eta_l &= \left( \cos \varphi \frac{\partial}{\partial r} - \frac{\sin \varphi}{r} \frac{\partial}{\partial \varphi} \right) \eta_l \end{aligned}$$

and the RBE in 1 DOF becomes

$$\begin{aligned} \partial_t \eta_l &= \frac{\varepsilon}{4} \left[ \left( (8 - g^2) + (4r + r^{-1}) \partial_r + \partial_r^2 + r^{-2} \partial_\varphi^2 \right) \eta_l \right. \\ &\left. + 2g J_{lm} \left( 2r \cos \varphi + \cos \varphi \partial_r - r^{-1} \sin \varphi \partial_\varphi \right) \eta_m \right], \\ l, m &= 1, 2, \quad l \neq m, \end{aligned} \quad (18)$$

We pose (18) on a disk  $r \leq r_{\max}$ ,  $\varphi \in [0, 2\pi]$ . The boundary conditions are periodic in  $\varphi$  and we take  $r_{\max}$  large enough to impose homogenous Dirichlet boundary conditions at  $r = r_{\max}$ . Here and in the following we drop arrows and replace  $\theta$  by  $t$ . We seek approximations to  $\eta$  on a Chebyshev grid in  $r$  and a uniform grid in  $\varphi$ ,

$$\begin{aligned} r_i &= -\cos \left( \frac{\pi i}{n_r} \right), \quad i = 0, \dots, n_r, \\ \varphi_j &= j \frac{2\pi}{n_\varphi}, \quad j = 1, \dots, n_\varphi, \end{aligned}$$

and expand it in a Fourier series in the  $\varphi$  direction:

$$\eta(t, r_i, \varphi_j) \approx \sum_{k=-n_\varphi/2+1}^{n_\varphi/2} \hat{\eta}(r_i, k, t) e^{-ik\varphi_j}. \quad (19)$$

For the  $k$ th Fourier mode we determine  $\hat{\eta}(t, r, k)$  from

$$\begin{aligned} \partial_t \hat{\eta}_l &= \frac{\varepsilon}{4} \left[ \left( (8 - g^2) + (4r + r^{-1}) \partial_r + \partial_r^2 - r^{-2} k^2 \right) \hat{\eta}_l \right. \\ &\left. + g J_{lm} \left( (2r + \partial_r) (\hat{\eta}_m^- + \hat{\eta}_m^+) - r^{-1} ((k \hat{\eta}_m)^- - (k \hat{\eta}_m)^+) \right) \right], \end{aligned} \quad (20)$$

where  $l, m = 1, 2, l \neq m$  and

$$\hat{\eta}_l^- = \hat{\eta}_l(t, r, k-1), \quad \hat{\eta}_l^+ = \hat{\eta}_l(t, r, k+1).$$

Now denote by  $\hat{u}_l(t, k)$  the grid function on the  $r$  grid for a fixed mode, i.e.  $\hat{u}_l(t, k) = [\hat{u}_l(t, r_0, k), \dots, \hat{u}_l(t, r_{n_r}, k)]^T$ , describes the  $l$ th component of  $\hat{\eta}$ . Then for each component we have

$$\frac{d\hat{u}_l(t, k)}{dt} = \frac{\varepsilon}{4} \left[ F_l^k(\hat{u}_l) + J_{lm} F_E^k(\hat{u}_m) \right]. \quad (21)$$

Here  $F_l$  and  $F_E$  are linear operators representing the Fokker-Planck operator and spin terms

$$\begin{aligned} F_l(\hat{u}_l) &= \left( (8 - g^2) I + (4R + R^{-1}) D_1 + D_2 - R^{-2} k^2 \right) \hat{u}_l, \\ F_E(\hat{u}_m) &= \left( (2R + D_1) (\hat{u}_m^+ + \hat{u}_m^-) - R^{-1} ((k \hat{u}_m)^- - (k \hat{u}_m)^+) \right). \end{aligned}$$

Here  $I$  is the  $(n_r + 1) \times (n_r + 1)$  identity matrix,  $R = \text{diag}(r_0, \dots, r_{n_r})$ ,  $D_1$  and  $D_2$  are spectral differentiation matrices. The entries of the differentiation matrices are found by the techniques for constructing finite difference approximations of any order of accuracy, for any order of the derivative and on general grids described by Fornberg in [7]. To be precise, the coefficients are computed using a numerically stable recursion relation derived from the Lagrange interpolant associated with the grid points (see also the subroutine `weights.f` provided in [7]). To evolve in time we use a fourth order additive  $N$ -stage Runge-Kutta scheme (ARK), see e.g. [2]. Let  $\hat{u}^\nu(k) = \hat{u}(k, \nu \Delta t)$  then, for each mode, we compute

$$\hat{u}^{\nu+1} = \hat{u}^\nu + \sum_{s=1}^N \gamma_s \mathbf{k}_s, \quad (22)$$

$$\mathbf{k}_s = \frac{\varepsilon \Delta t}{4} \left[ F_l \left( \hat{u}^\nu + \sum_{l=1}^s \alpha_{sl} \mathbf{k}_l \right) + F_E \left( \hat{u}^\nu + \sum_{l=1}^{s-1} \beta_{sl} \mathbf{k}_l \right) \right].$$

Thus at each step we compute  $u^{\nu+1}$  given  $u^\nu$ . The general  $N$ -stage ARK scheme combines  $N$ -stage diagonally implicit Runge-Kutta scheme (DIRK) with  $N$ -stage explicit Runge-Kutta scheme (ERK) of a same order of accuracy. The coefficients  $\alpha_{sl}, \beta_{sl}, \gamma_s$  can be found so that the combined schemes are consistent. For higher number of DOF the algorithm stays the same.

As indicated in the introduction, the cost of the solve in (22) depends on the choice of algorithm but can always be split into an initial cost (e.g.  $LU$ -factorization in addition to the FFT) and a solve cost (e.g. back and forward substitution). Ignoring the startup cost, which can be amortized over many

time steps, the cost per time step is that of  $n_\varphi$  solves of size  $n_r$  for each stage, i.e.  $O(n_\varphi n_r^q)$ .

For 3 DOF the complexity estimate becomes

$$C = O(n_\varphi^3 n_r^{3q}).$$

Assuming that  $n_r = n_\varphi = N$  we find that for  $N = 50$  and  $q = 1, 4/3, 2$  the cost  $C = 1.5 \cdot 10^{10}, 7.8 \cdot 10^{11}$  and  $2 \cdot 10^{15}$ , respectively. As a single modern processing unit may be able to carry out  $O(10^8 - 10^9)$  arithmetic operations per second it appears plausible that an efficient parallel implementation can result in time-per-time step on the order of one to several seconds for  $q = 4/3$ . Note that there are several modern solution techniques, like the Hierarchical Poincaré-Steklov operator technique by Martinsson [8] that can reach  $q \approx 1$  for spectrally accurate discretizations.

## NUMERICAL RESULTS

### Example 1: 1 DOF model

Here the reduced Bloch Equations (17) can be solved exactly [9]. For example if

$$\eta(0, w) = \frac{2}{\pi} \begin{pmatrix} \cos(\psi_0) \\ \sin(\psi_0) \end{pmatrix} e^{-2(w_1^2 + w_2^2)}. \quad (23)$$

then

$$\eta(t, w) = \frac{2}{\pi} e^{\Sigma_2} \begin{pmatrix} \cos(\psi_0 + \Sigma_1 w_1) \\ \sin(\psi_0 + \Sigma_1 w_1) \end{pmatrix} e^{-2(w_1^2 + w_2^2)}, \quad (24)$$

$$\Sigma_1(t) = -g(1 - e^{-\varepsilon t}), \quad \Sigma_2(t) = \frac{g^2}{8}(e^{-2\varepsilon t} - 1),$$

The polarization vector of the bunch at time  $t$  is

$$P(t) = \int_{\mathbb{R}^2} \eta(t, w) dw,$$

and

$$|P(t)| = e^{-\frac{1}{8}\Sigma_1^2(t)} e^{\Sigma_2(t)}. \quad (25)$$

This example was used to verify that the polarization is computed accurately. The polarization  $|P(t)|$  obtained by integrating the numerical solution, see Fig. 1. The result is very close to the exact polarization (25), within the error we expect.

In Fig. 2 the numerical solution  $\eta_1$  snapshots are taken at initial time and after 25, 250 and 1500 turns (10 turns HERA ring correspond to  $t = 4$ ) showing that the solution approaches the equilibrium

$$\eta_{eq}(w) = \frac{2}{\pi} e^{-\frac{g^2}{8}} \begin{pmatrix} \cos(\psi_0 - gw_1) \\ \sin(\psi_0 - gw_2) \end{pmatrix} e^{-2(w_1^2 + w_2^2)}. \quad (26)$$

### Example 2: Spectral convergence for 1 DOF model

To confirm the spectral convergence in  $r$  and  $\varphi$  we evolve (18) with the initial data taken to be the exact solution (24) at  $t = 0$ . It has an equilibrium solution (26) making it a good test case for the numerical method. To be precise, the errors displayed in Fig. 3 are the maximum deviation from the exact solution (24) taken over all grid points and all the variables at  $t = 5/\varepsilon$  which corresponds to 1500 turns. The results clearly show the spectral accuracy of the spatial discretization.

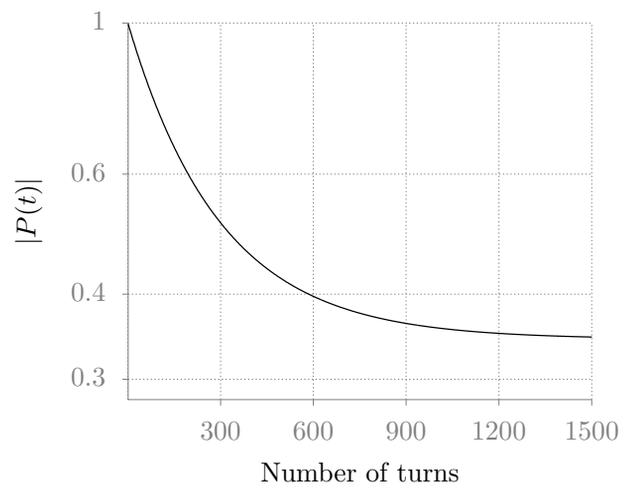


Figure 1: Polarization in 1 DOF model computed from the numerical solution.

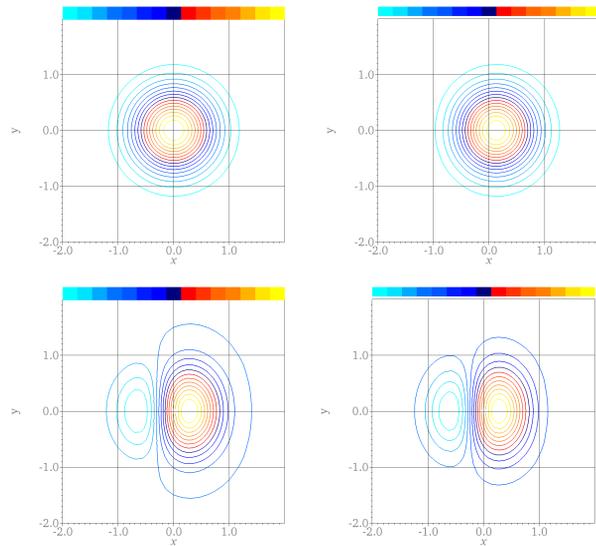


Figure 2: Solution  $\eta_1$  at time  $t = 0, 10, 100, 600$

## DISCUSSION AND NEXT STEPS

We are preparing an extended version of this brief note for an archival journal which will complete the work on the reduced Bloch equation in 2 DOF. An important aspect will be a more detailed discussion of the algorithm. The codes will be made available in a repository. A goal is to make our work easily reproducible. Next we will incorporate the spin flip by considering the full Bloch equations and do a careful study of the depolarization and polarization effects for the simple lattice. This will include depolarization and polarization times and equilibrium. We will then study a more realistic lattice in the 2 DOF case and begin the 3 DOF work, where a parallel algorithm will surely be important/necessary.

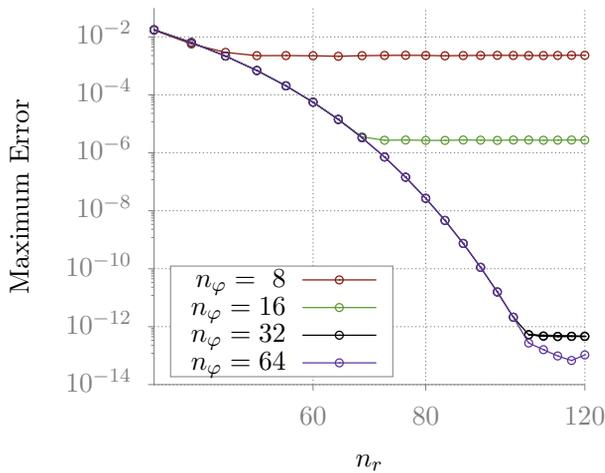


Figure 3: Convergence for the 1 DOF model.

### ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0018008.

### REFERENCES

[1] K. Heinemann, D. Appelö, D.P. Barber, O. Beznosov, J.A. Ellison, “Spin dynamics in modern electron storage rings:

Computational and theoretical aspects”, *Invited talk and paper, ICAP18, Key West, Oct 19–23, 2018.*

[2] C.A. Kennedy, M.H. Carpenter, “Additive Runge-Kutta schemes for convection-diffusion-reaction equations”, *Appl. Numer. Math.*, vol. 44, pp. 139–181, 2003.

[3] D.P. Barber, K. Heinemann, H. Mais, G. Ripken, “A Fokker-Planck treatment of stochastic particle motion”, DESY-91-146, 1991.

[4] K. Heinemann, “Some models of spin coherence and decoherence in storage rings”, arXiv:physics/9709025, 1997.

[5] D.P. Barber, M. Böge, K. Heinemann, H. Mais, G. Ripken, Proc. 11th Int. Symp. High Energy Spin Physics, Bloomington, Indiana (1994).

[6] H. Mais, G. Ripken, “Spin-orbit motion in a storage ring in the presence synchrotron radiation using a dispersion formalism”, DESY-86-029, 1986.

[7] B. Fornberg, “Classroom Note: Calculation of Weights in Finite Difference Formulas”, *SIAM Review*, vol. 40(3), pp. 685–691, 1998.

[8] P.G. Martinsson, “A direct solver for variable coefficient elliptic PDEs discretized via a composite spectral collocation method”, *J. Comput. Phys.*, vol. 242, pp. 460 – 479, 2013.

[9] K. Heinemann, unpublished notes.

# APPROACHES TO OPTIMIZING SPIN TRANSMISSION IN LATTICE DESIGN

V. H. Ranjbar\*, BNL, Upton, USA

## Abstract

We present our experiences in optimizing the proposed Rapid Cycling Synchrotron (RCS) injector for the eRHIC Storage ring and the RHIC 2017 lattice. We have developed a Python code to drive lattice calculations in MADX [1] which are then used to calculate spin resonances using the DEPOL algorithm [2]. This approach has been used to minimize intrinsic spin resonances during the RCS acceleration cycle while controlling lattice parameters such as dispersion and beta functions. This approach has also been used to construct localized imperfection bumps using a spin response matrix and the singular valued decomposition (SVD) algorithm. It has also been used to reduce interfering intrinsic spin resonances during the RHIC acceleration ramp.

## INTRODUCTION

The design of lattices to transport beam with minimal polarization loss requires special attention to the potential sources of depolarizing spin resonances. In the past accelerators were usually optimized initially without consideration of their spin effects. Later careful spin matching and harmonic bumps were applied to reduce the effects of the various spin resonances. These were usually performed by using simple analytical estimates which were then verified using more exact numerical calculations using codes like DEPOL [2]. However the process was slow and decoupled from the optics calculations since for small perturbations the optic changes were assumed to be small. Later the use of full and partial snakes reduced the need to worry as much about the strength of the spin resonances.

However experience with the full snakes in RHIC and partial snakes in the AGS have shown that optimizing the underlying spin resonances can still have an important impact on the performance of the lattice even with snakes. For example the overlap of minor spin resonances during strong spin resonances crossings has been shown to reduce polarization transmission on the RHIC ramp [3].

Also in the design of a future high energy polarized electron injector for the proposed eRHIC facility, snakes are not a viable option due to the large orbit excursion and radiative effects they induce in lighter particles. So a new design has been developed which exploits high periodicity to avoid spin depolarizing resonances [4]. This new design required optimizing for both the beam dynamics and polarization transmission.

We have developed a Python code which we call SOptim [5], to perform these types of optimizations. It uses MADX [1] as an optics calculator and performs its own DEPOL calculations across a specified range of spin resonances.

It then varies the magnet strengths, usually quadrupoles, or vertical corrector magnets to achieve a desired spin resonance structure and optics.

## SOptim CODE STRUCTURE

The core of the code is a function which takes the name of a MADX sequence file, a string containing the names of the magnets to be varied, and a vector of associated magnet strengths. These are read together with the anomalous G factor, emittance and a switch to indicate if an intrinsic or imperfection spin resonance calculations are to be performed.

This function then feeds the sequence and new magnet strengths to MADX and takes back the optics and orbit information which is then used to perform a native spin resonance calculation following the DEPOL algorithm. These are done for a currently hard coded range of spin resonance energies.

The function returns an array of spin resonance strengths, twiss parameters, dispersion values and a flag indicating if the twiss calculation failed or not, as well as a MADX readable string containing the magnet settings used. This string is saved once the optimization is completed so that it can be easily included in future MADX calculations.

The values for the spin resonances and optics values are then combined with various weights to produce an overall penalty function. This penalty function is then iterated upon using Scipy's [6] minimize function which is part of its optimize library. The minimize function can use a variety of optimizers and takes bounds for the magnet settings.

Alternatively it is possible to construct spin response matrix which relates the spin resonance strengths to the magnet settings. One can invert this using a pseudo inverse to either suppress or generate spin resonance 'bumps'.

## RHIC LATTICE OPTIMIZATION

Studies of polarization transport with two orthogonal snakes during the RHIC acceleration ramp revealed the importance of interfering spin resonances [3]. For example during the crossing of the strongest intrinsic spin resonance at  $G\gamma = 422.685$  it was found that the weak spin resonance located at  $G\gamma = 423.325$  plays an important role in determining the total polarization transmission. As is shown in Figs. 1 and 2 it was shown that minimizing this can improve the total spin transmission. We used the SOptim code to reduce these neighboring resonances. The results from spin-orbit 6D tracking are shown in Fig. 3, where the polarization transmission versus emittance is greatly expanded. Similar optimization was applied around the three major spin resonances at  $G\gamma = 422.685$ ,  $382.325$ , and  $260.685$  for both the Blue and Yellow RHIC rings.

\* vranjbar@bnl.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

### Analysis of FY12 lattice Crossing 393+NU resonance

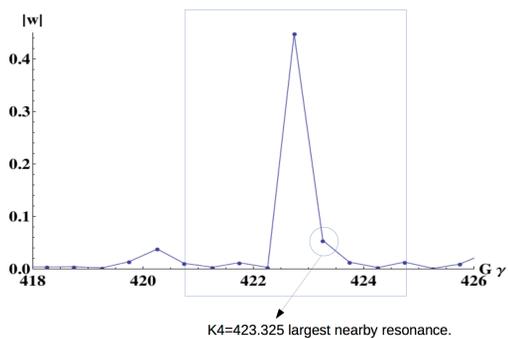
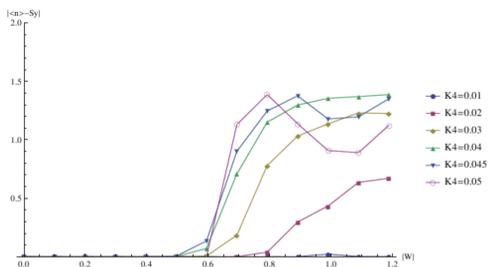


Figure 1: Spin Resonance magnitude  $|w|$  versus beam energy in terms of the anomalous  $G$  factor times relativistic  $\gamma$ .

### K4 seems singular in determining polarization aperture.



Lattice independent tracking with 5 resonance from FY12 varying  $K4=423.325$  strength

Figure 2: Showing the magnitude of net spin loss after accelerating through the primary resonance. This is plotted against the spin resonance magnitude of the primary resonance  $|w|$  (proportional to the action or emittance of the particles). We see the effect of changing the neighboring resonance.

### RCS LATTICE OPTIMIZATION

The proposed eRHIC injector is designed to accelerate polarized electrons from 400 MeV to 18 GeV. A periodicity of 96 and tune of 50 was necessary to avoid all the intrinsic spin resonances in this energy range. However the natural 6-fold symmetry of the existing RHIC tunnel made achieving the 96 periodicity challenging. Our solution was to make the contributions to the spin resonance integral from the straight sections negligible. This was made more difficult since detector bypasses are necessary for such a machine.

We used the SOptim code to drive down spin resonance strengths in the accelerating energy range from  $G\gamma = 0.9$  to 41. The results yielded a lattice with intrinsic spin resonances with strengths so low that the cumulative effects only start to manifest themselves at the 5% loss level above 1000 mm-mrad emittance for the 100 msec proposed ramp-

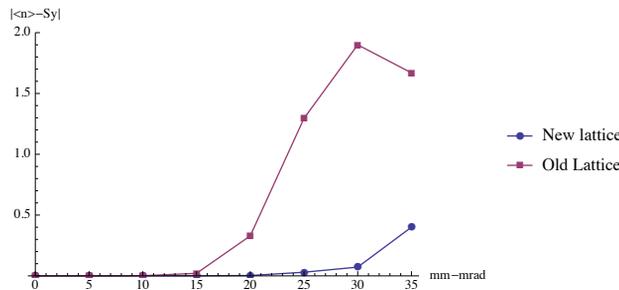


Figure 3: Polarization deviation from invariant spin field direction versus emittance curve after crossing  $G\gamma = 422.685$  spin resonance. Comparing old RHIC 255 GeV lattice (old lattice) to optimized lattice (new lattice).

ing rate. The resultant intrinsic spin resonances are shown in Fig. 4.

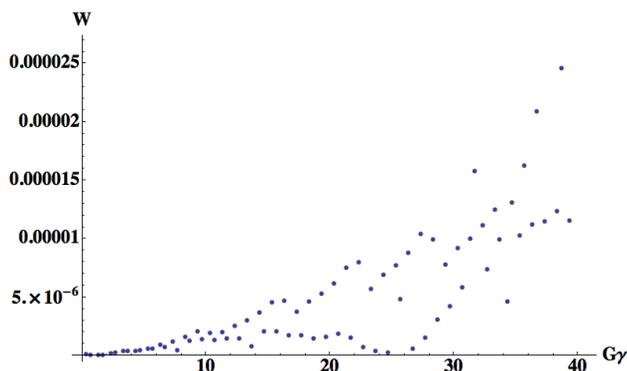


Figure 4: Plot of intrinsic spin resonances versus  $G\gamma$  for a particle of 10 mm-mrad normalized emittance. This is the result of spin optimization code SOptim applied to the RCS lattice after applying symmetry breaking detector bypasses.

### RCS IMPERFECTION BUMPS

Control of the effects of imperfect spin resonances remains an important issue for the design of the RCS. Studies showed that if we can control the RMS vertical orbit distortion to the 0.5 mm level we should be able to achieve above a 95% polarization transmission. In the event this isn't achieved, we proposed using orthogonal imperfection bumps. These bumps target the real and imaginary part of the spin resonance. Before the commissioning of the partial snake in the AGS, imperfection bumps were used extensively to tune out imperfection driven polarization losses. More recently tests with orthogonal bumps using a spin response matrix method have been performed in RHIC [7].

By building a corrector to imperfection spin response matrix ( $\vec{M}_S$ ) using SOptim, one can calculate the necessary corrector strengths to achieve isolated and orthogonal bumps at any point in the ramp. Here we use,

$$\vec{C} \cdot \vec{M}_S = \vec{\epsilon}, \quad (1)$$

where  $\vec{C}$  is a vector containing all the correctors,  $\vec{\epsilon}$  is a vector containing all the real and imaginary parts of the

imperfection spin resonances which are targeted. In our case it is 80 elements long containing the real and imaginary imperfection resonances at  $G\gamma = 1$  to 40. Inverting this non-square matrix using a pseudo inverse one can use it to construct an arbitrary set of imperfection bumps across the whole energy range.

Using this approach we have constructed such bumps representing arbitrary imperfection bumps in the imaginary and real plane for  $G\gamma = 34$  to 40. This is the energy range we expect that imperfections could be strong enough to depolarized the beam greater than 5%. These bumps are plotted in Fig. 5. Here we produced resonance strengths as high as  $\pm 0.005$ , everywhere else has imperfection strengths less than  $10^{-5}$ . A spin resonance strength of 0.005 at 100 to 200 msec ramp rates, represent 10 to 15% spin kick respectively.

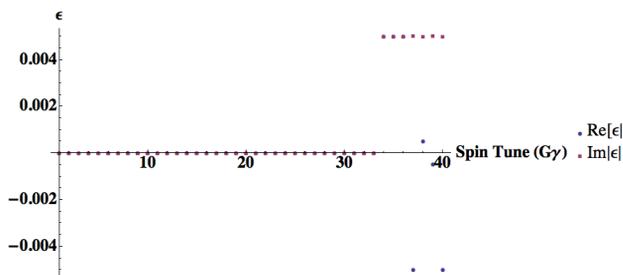


Figure 5: Real and imaginary imperfection resonance strengths versus  $G\gamma$  for a fixed corrector magnet settings constructed using spin response matrix.

The associated corrector magnet strengths are all less than  $\pm 3 \times 10^{-5}$  rad as shown in Fig. 6. With imperfection bumps constructed in this fashion, there is no need to alter the corrector strengths over the acceleration cycle. They only need to ramp with the main dipole current.

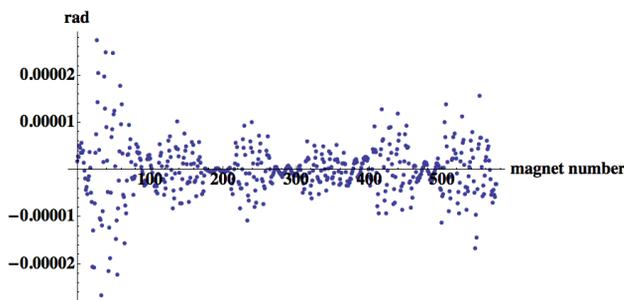


Figure 6: Corrector magnet settings versus magnet number, which resulted in Imperfection resonances shown in Fig. 5.

Thus using these bumps, we should be able to control any foreseeable imperfection spin resonances in the RCS.

## CONCLUSION

The Python code SOptim is tool which can assist in the optimization of hadron and lepton lattices for spin transmission. It accomplishes this by calculating both the spin resonances and optics for a given lattice and employing various optimizing approaches to achieve either global suppression of a range of spin resonances or by constructing arbitrary spin bumps to cancel spin resonances.

## ACKNOWLEDGEMENTS

Work supported by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. This research used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## REFERENCES

- [1] MADX, <http://madx.web.cern.ch/madx/>
- [2] Courant, E. D. and Ruth, R. D., "The Acceleration of Polarized Protons in Circular Accelerators", BNL-51270, (1980).
- [3] V. H. Ranjbar, F. Méot, M. Bai, D. T. Abell and D. Meiser, "Polarization response of RHIC electron lens lattices," *Phys. Rev. Accel. Beams* **19**, no. 10, 101002 (2016). doi:10.1103/PhysRevAccelBeams.19.101002
- [4] V. H. Ranjbar *et al.*, "Spin Resonance Free Electron Ring Injector", *22nd International Spin Symposium Hosted by: University of Illinois and Indiana University*, September 25–30, 2016 at UIUC; BNL-203602-2018-TECH, C-A/AP/599 January 2017.
- [5] SOptim, <https://sourceforge.net/projects/t-bmt-spin-resonance-tracker/>
- [6] Scipy, <https://www.scipy.org>
- [7] M. Bai, H. Huang, A. Marusic, M. G. Minty, V. Ptitsyn and V. H. Ranjbar, "Experimental Effects of Orbit on Polarization Loss in RHIC," *Conf. Proc. C* **1205201**, 3788 (2012).

# PLASMA WAKEFIELD START TO END ACCELERATION SIMULATIONS FROM PHOTOCATHODE TO FEL WITH SIMULATED DENSITY PROFILES\*

A. Marocchino<sup>†</sup>, E. Brentegani, A. Biagioni, E. Chiadroni, M. Ferrario,  
A. Giribono, F. Filippi, R. Pompili, C. Vaccarezza, INFN-LNF, Frascati, Italy  
A. Cianchi, University of Rome Tor Vergata, Rome, Italy  
A. Bacci, A. R. Rossi INFN Milano, Milano, Italy  
V. Petrillo, Università di Milano, Milano, Italy

## Abstract

Plasma Wakefield acceleration is a promising new acceleration technique that profits by a charged bunch, e.g. an electron bunch, to break the neutrality of a plasma channel to produce a wake where a trailing bunch (or witness) is eventually accelerated. The quest to achieve extreme gradient conserving high brightness has prompted to a variety of new approaches and techniques. Most of the proposed schemes are however limited to the only plasma channel, assuming in the vast majority of cases, ideal scenarios (e.g. ideal bi-gaussian bunches and uniform density plasma channels). Realistic start-to-end simulations, from the photo-cathode to FEL via a high gradient, emittance and energy spread preserving plasma section, are mandatory for paving the way towards plasma-based user facilities.

## INTRODUCTION

Plasma-based accelerators (PBAs) represent, today, a new frontier for the development of compact advanced radiation sources and next generation linear colliders. High brightness electron beams are the future goal of such kind of accelerators, a new technology that is envisioned to compete with the RF technology. Since the very first conception [1] great effort (with equivalent progress) has been placed [2–8] to demonstrate the acceleration of electron beams with gradients of the order of several tens of GV/m, produced either by a laser drive (LWFA) or with a particle driven (PWFA). In PWFA the high gradient wakefield is induced by a high-energy charged particle beam (referred to as driver bunch) travelling through a pre-ionised plasma. The background electrons by shielding the charge breakup produced by the driver induce an accelerating field. The second bunch (referred to as trailing bunch or witness) placed on the right phase is so accelerated by the induced electric field [9–11].

The aforementioned publications indicate that either LWFA and PWFA are capable to produce strong accelerating gradients, but do not address the issue of beam acceleration with quality retention. In other words, the capability for a PWFA scheme to accelerate a trailing bunch with an acceptable quality for any scientific applications, such a Free

Electron Laser (FEL) or a future particle-particle collider, or even for medical applications is still an open question.

The work presented in this proceeding tries to address the problem of quality bunch transport numerically for a PWFA scheme. The scheme proposed leverages on the well establish RF know-how to produce and manipulate the bunches to match the strict plasma matching requirements, so that, once injected into the plasma the trailing bunch is boosted to the desired energy. The scheme can thus be decoupled into two parts. A first stage aims in generating two electron bunches (the driver and the trailing bunch) with a given time-separation so that once injected into the plasma they will be at right phase to maximize acceleration. The generation, distance positioning and acceleration is achieved using the COMB [12] technique and an RF in the X-band regime, for a final particle energy of 500 MeV. The plasma with a  $10^{16} \text{ cm}^{-3}$  number density is confined by a millimeter diameter capillary and ionized with plasma discharges, accelerates the particles from 500 MeV to 1 GeV in about 30 cm. The particle extracted from the plasma accelerating section are then injecting into a FEL. The evolution of the beam, from the photo-injector to the FEL via the plasma cell, is performed with a *single* simulation by concatenating several codes used to address the different physics occurring in the different accelerating section. We refer to this single simulation as a *start-to-end* simulation. This work is part of EuPRAXIA [13, 14] (European Plasma Research Accelerator with eXcellence In Applications) European project whose final goal is to use PBA to seed a FEL for physical and biological applications. Specifically EuPRAXIA@SPARC\_LAB [14] is the envisioned EuPRAXIA European Italian-located facility operating at 1–5 GeV for FEL experiments.

## SECTION I: INJECTOR

A comb-like configuration for the electron beam is used to generate simultaneously a 200 pC driver followed by a 30 pC trailing bunch. The comb-like operation foresees the generation of two or more bunches within the same RF accelerating bucket through the so-called laser-comb technique [12, 15] consisting in a train of laser time-spaced pulses illuminating the photo-cathode. The witness arrives earlier than the driver on the photo-cathode and then they are reversed in time at the end of the velocity bunching process, during

\* Work supported by European Union Horizon 2020 research and innovation program, Grant Agreement No. 653782 (EuPRAXIA)

<sup>†</sup> alberto.marocchino@lnf.infn.it

which the longitudinal phase space is rotated. Experimental results have been obtained at SPARC\_LAB where the laser-comb technique is routinely used in order to produce trains of multiple electron bunches [16] for narrow-band THz generation [17], two-color FEL experiments [18, 19] and resonant particle driven PWFA [20]. Computational studies have been dedicated to provide two bunches separated by 0.55 ps [21], which corresponds to  $\lambda_p/2$ , where  $\lambda_p = 330 \mu\text{m}$  for  $n_p = 10^{16} \text{cm}^{-3}$ . Both driver and witness bunches must be compressed down to 50 fs and 10 fs (FWHM), respectively. The process need some fine tuning to avoid emittance growth, naturally occurring because of the witness-driver overlapping during the velocity bunching regime. The photocathode laser has been shaped in order to provide at the cathode a witness and a driver bunches separated by  $\Delta t = 4.8 \text{ps}$ . With this configuration the beam crossing occurs in the second TW accelerating cavity and a fine-tuning of the RF phases suffices to provide the desired 0.55 ps beam separation corresponding to  $\lambda_p/2$ . Gaussian longitudinal distribution with  $\sigma_z = 120 \mu\text{m}$  (RMS) and uniform transverse distribution of radius  $r = 0.35 \text{mm}$  have been assumed for the witness pulse at the cathode. The different intensities between the two pulses permit to generate bunches with different charges: a 30 pC trailing bunch and a 200 pC driver. It is worth to notice by adopting a  $\sigma_D = 0.35 \text{mm}$ , the FWHM witness length does not suffer lengthening, although the minimum RMS witness length is obtained for  $\sigma_D = 0.25 \text{mm}$ . The sliced current at the end of the photo-injector line is plotted in Fig.1.

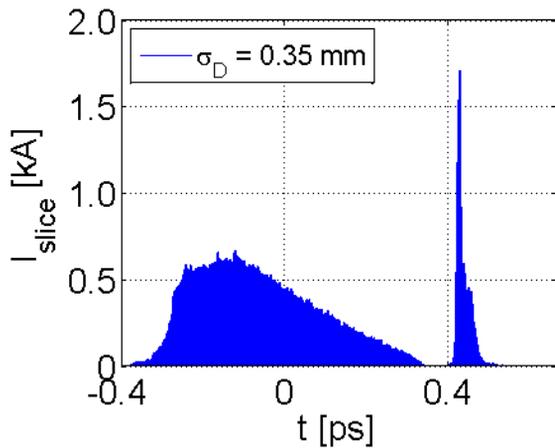


Figure 1: Driver and trailing bunch longitudinal current distribution at the photo-injector exit. The beam is propagating from right to left with the driver arriving earlier than the trailing bunch.

Driver and trailing bunch have been simulated with the TSTEP code and with 30k and 200k macro-particles respectively. In the described configuration the driver arrives 0.58 ps earlier than the witness at the X-band booster. It is worth to notice that the trailing bunch length is about 3  $\mu\text{m}$  FWHM with a normalized transverse emittance of  $\sim 0.7 \text{mm mrad}$ .

## SECTION II: LINAC

The X-band RF linac is used to accelerates the entire beam to 500 MeV [22, 23], energy of the beam entering the plasma, specifically the trailing bunch is prepared for plasma injection with a transverse dimension of 1-2  $\mu\text{m}$  spot size. The comb-like electron beam undergoes deep over-compression in the photo-injector by means of the velocity bunching scheme. The same accelerating gradient of  $E_{acc} \approx 20\text{--}36 \text{MV/m}$  is applied in L1 and L2 linac section respectively, and the final electron beam energy is  $E_{L2exit} \approx 580 \text{MeV}$ , with an energy spread less than 0.1%. The driver and witness bunches are characterized by high charge/low current and low charge/high current, respectively. Moreover, the initial matching conditions for the injection in the X-band linac are quite different for the two bunches, as shown by their transverse phase space at the linac entrance (i.e. injector exit). In this regard, an efficient sharing of the same lattice is achieved by means of a mild transverse focusing that aims to keep the RMS size of the comb beam compatible with the beam stay-clear-aperture through all the X-band accelerator. The same argument applies also to the focusing stage with the permanent quadrupoles at the entrance of the plasma capillary where a residual asymmetry between horizontal and vertical plane for the witness beams is present, see Fig. 2.

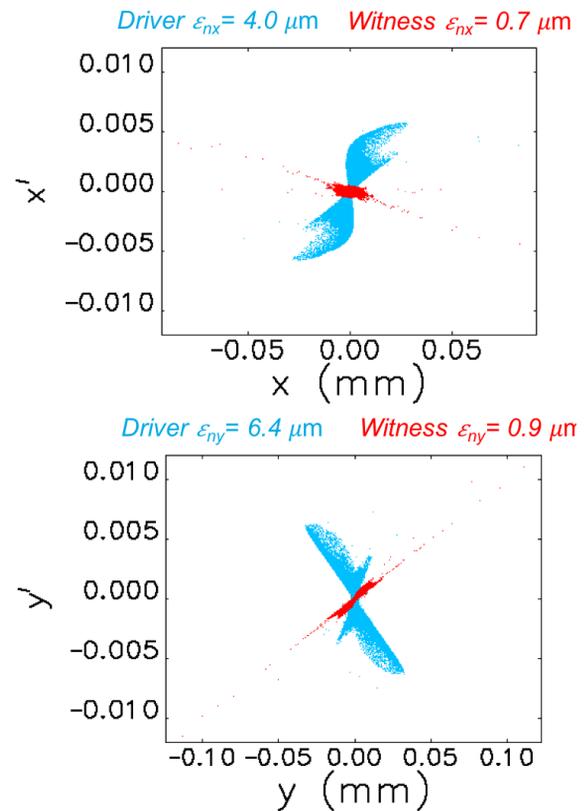


Figure 2: Horizontal and vertical phase space distribution of the PWFA driver (cyan dot) and witness (red dot) beams at the capillary entrance.

### SECTION III: THE PLASMA CELL

Start-to-end simulations use, for the plasma section, the state-of-the-code ARCHITECT [24, 25]. The use of Architect has been dictated by the necessity to run long simulations where a classical particle-in-cell (PIC) approach would have been computationally too expensive. The Architect reduced model, which relies on a fluid background, has been benchmarked numerically against the ALaDyn-PWFA [26] PIC code [25] and versus SPARC\_LAB experimental results [27, 28]. Start-to-end simulations are performed by concatenating the use of codes without any phase space manipulation or remapping. Specifically, for the interface RF X-band with plasma The ELEGANT code is used to track particle up to the plasma entrance, the particle phase-space is then imported into the Architect code for the evolution in the plasma section. Simulations described in this section have been run with a longitudinal resolution of 1  $\mu\text{m}$  and a transverse resolution of 0.4  $\mu\text{m}$ , a mesh that allows to resolve the fine structure with a reasonable computational cost. The advancing time step is of 1.1 fs. The number of particle used to discretize the driver is, on average, 30 particles per cell while the witness is discretized with an average of 100 particle per cell. We recall that the goal is to accelerate the trailing bunch retaining as much as possible its original quality. To limit the energy spread growth, the trailing bunch has been designed with a (as much as possible) *triangular* shape. The triangular shape together with a specific value of peak charge represents the optimized longitudinal density profile that limits energy growth spread during propagation. Driver and witness have been designed to perform the best acceleration in terms of quality and in terms of energy transfer (*transformer ratio*,  $R$ ) [29], parameter that identifies the rate of energy transfer from the driver to the witness. In our case the transformer ratio is estimated in the range 2-3. The foreseen experiment is planned in the so-called weakly-non-linear regime, where despite fields resemble a sawtooth profile linear field sum is still possible. The parameter we used to measure the degree of nonlinearity is the *reduced charge parameter* [30, 31],  $Q = \frac{N_b}{n_0} \kappa_p^3$ , with  $N_b$  the electron bunch number (bunch charge divided by the elementary charge). For this case the driver reduced charge is  $Q \sim 0.8 - 0.9$ .

Nonetheless, if the upstream application is a FEL application that requires a slice current around 2 kA, combining this requirement with the triangular shape, simulations helps identifying that beam loading compensation occurs for a driver-witness distance 184  $\mu\text{m}$  or  $0.55 \times \lambda_p (n_0 = 10^{16})$ , where the electric field experienced by the witness is around 1.1 GV/m. We recall that the accelerating field together with the plasma wavelength depend upon the plasma number density  $n_0$ ,  $\propto n_0^{1/2}$  and  $\propto n_0^{-1/2}$  respectively. The capability to control the density would permit some flexibility and adjustments in the bubble profile and accelerating fields, this to compensate -on site- whenever the distance between driver and witness would oscillate or change for experimental unforeseen reasons. The flat density profile, together with the

required value is achieved with a capillary tube. The capillary tube, confining the ejected gas, permits a high degree of control to which we can rely on for experimental on site optimization. At present simulations consider some density ramps of the order of 0.5 cm, that are experimentally reasonable and whose length is below the betatron wavelength assuring no bunch oscillations within the ramps to increase acceleration robustness [26]. At plasma entrance, the trailing bunch is delivered with a shape that resemble the triangular required shape, transversally the bunch is fairly symmetric in size and in emittance Table 1.

Table 1: PWFA Bunch Parameters at Plasma Entrance and at Plasma Exit (the best slice value is also reported)

Beam	units	Driver IN	Driver OUT	Tr. B. IN	Tr. B. OUT
Charge	pC	200	200	30	30
$\sigma_x$	$\mu\text{m}$	8	6.4	1.47	1.42
$\sigma_y$	$\mu\text{m}$	3.1	10	3.17	1.4
$\sigma_z$	$\mu\text{m}$	52	50	3.85	3.8
$\varepsilon_x$	$\mu\text{m}$	2.56	4.1	0.6	0.96
$\varepsilon_y$	$\mu\text{m}$	4.8	11.4	0.55	1.2
$\sigma_E$	%	0.2	20	0.07	1.1
E	MeV	567	420	575	1030
Best Slice					
current	kA			2	2.0
$\varepsilon_x$	$\mu\text{m}$			0.59	0.57
$\varepsilon_y$	$\mu\text{m}$			0.58	0.62
$\sigma_E$	%			0.011	0.034

The driver is instead of lower brightness quality, but since the driver will undergo depletion -anyway- we can accept a lower quality since its main purpose is to *only* drive the wake. The setting of the RF line in favor of the witness naturally bring the driver on a less optimized point that is delivered at plasma with a lower quality than the trailing bunch. The driver at plasma entrance has the front part that is highly convergent, convergence that will cause a consequent expansion within the plasma channel producing a unique funnel shape, Fig. 3. The driver is capable to induce a weakly nonlinear wake with an effective maximum field that peaks (at bubble closure) around 2.5–3.0 GV/m, as can be retrieved from Fig. 3. The central part of the driver that mostly contributes to generate the wake looses after 40 cm about 150 MeV, the witness gains about 460 MeV. The slice analysis, reported numerically in Table 1 suggests that the witness head and tail undergo a phase space dilution, while the central slice with very high current retain high quality. The witness, at plasma entrance, has the emittance in both planes as well as the energy spread almost uniform along the entire witness length. After the plasma acceleration section, the bunch has lost this homogeneity exhibiting a different slice quality along its length.

The front part and the rear part of the witness are characterized by large emittance and energy spread. While ideally

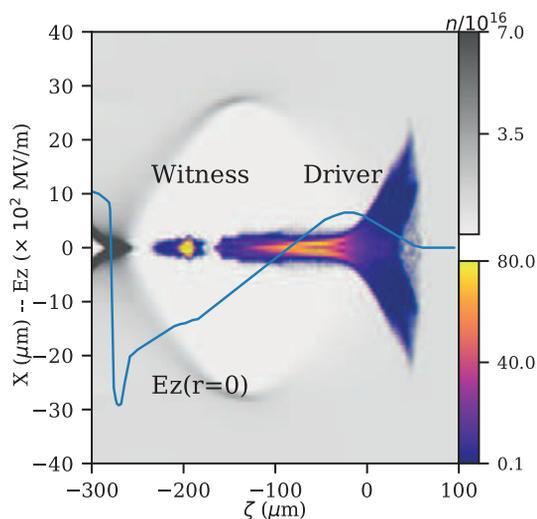


Figure 3: Bunch and background density colormaps after 5 mm within the plasma. The bunch density is plotted with a *plasma* colormap, while the background is plotted with a grey colormap. The longitudinal accelerating electric field, on axis, is over-imposed with a solid blue line. For scale purposes and for sake of clarity the Ez is plotted in hundreds of MV/m.

we wish to conserve quality all along the trailing bunch, the head and tail are characterized by a lower current, condition that allows us some flexibility on these regions since their lasering within the FEL would be negligible. However, and most importantly, the region within the high current bell retains its quality. From Fig. 4, top panel, we notice that under the region of high current the emittance in both plane is almost conserved with little deterioration. The energy spread undergoes some general increase also in the region of maximum current. The slice value, in the region of maximum current, stays below 0.1%. The peak current value corresponds to the transition from a higher value to the lowest one. The best slice characterized by a 2 kA current, has an energy spread as low as 0.034%, an emittance of 0.57 mm–mrad and 0.62 mm–mrad in the *x* and *y* plane respectively.

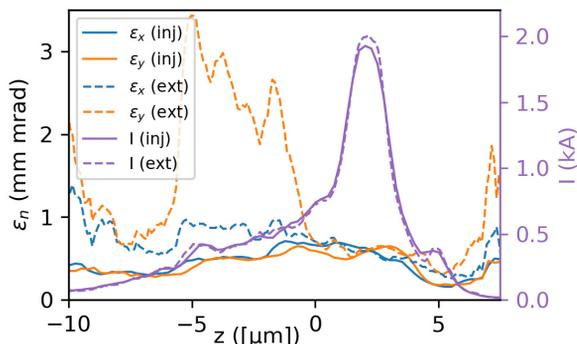


Figure 4: Slice analysis for the witness bunch at plasma input, dashed color, and at plasma exits, solid color.

## The Architect Code

One of the codes used to complete the start-to-end simulations is the ARCHITECT code used to study the bunch dynamics within the plasma. ARCHITECT integrates plasma wakefield acceleration equation by combining a PIC approach with fluid equations. The PIC approach is used to discretise the electron bunches while the background electrons are treated as a cold fluid, electromagnetic fields are evolved accordingly to Maxwell's equations. The set of equations solved can be written, in a compact way, as follows,

$$\begin{aligned}
 \partial_t n_e + \nabla \cdot (\boldsymbol{\beta}_e c n_e) &= 0 \\
 \partial_t \mathbf{p}_e + c \boldsymbol{\beta}_e \cdot \nabla \mathbf{p}_e &= q(\mathbf{E} + c \boldsymbol{\beta}_e \times \mathbf{B}) \\
 \nabla \times \mathbf{E} + \partial_t \mathbf{B} &= 0 \\
 \nabla \times \mathbf{B} - c^{-2} \partial_t \mathbf{E} &= q \mu_0 c (n_e \boldsymbol{\beta}_e + n_b \boldsymbol{\beta}_b) \\
 d_t \mathbf{p}_{\text{particle}} &= q(\mathbf{E} + c \boldsymbol{\beta}_{\text{particle}} \times \mathbf{B}) \\
 d_t \mathbf{x}_{\text{particle}} &= \mathbf{v}_{\text{particle}}
 \end{aligned} \tag{1}$$

where  $\mathbf{E}$  is the electric field,  $\mathbf{B}$  the magnetic field,  $c$  the speed of light,  $\boldsymbol{\beta}_e = \mathbf{u}_e/c$  the relativistic  $\beta$  for the background electrons,  $\boldsymbol{\beta}_b$  the relativistic  $\beta$  for the electron bunch,  $\mathbf{p}_e$  the fluid relativistic momentum for electrons,  $n_e$  is the electron density and  $n_b$  the bunch density. For each single particle of the kinetic bunch(es) we identify a relativistic momentum,  $\mathbf{p}_{\text{particle}}$ , a relativistic beta,  $\boldsymbol{\beta}_{\text{particle}}$ , a velocity,  $\mathbf{v}_{\text{particle}}$ , and a position,  $\mathbf{x}_{\text{particle}}$ . The first and second equations of Eq. (1) are the fluid mass conservation and the fluid momentum conservation respectively. The third and the fourth are Faraday's law and Ampere's law respectively. The last two equations of Eq. (1) are the kinetic compound to the model, the relativistic Newton's law for each single particle composing the bunch(es): the momentum equation and the position-velocity equation. In Eq. (1) the fluid velocity classically written as  $\mathbf{u}_e$  has been written as a function of  $\beta$ . Ions are assumed as a static background.

Dynamic in plasma-based accelerators spans over a large range of timescales, while beams evolve on a timescale on the order of the inverse betatron oscillations, the background plasma evolves on a timescale on the order of the inverse of the plasma frequency. The shortest dominating timescale is the background plasma frequency, by using a fluid description we drastically reduce the computational costs in a variety of mean: from memory occupations to evolution costs since fluid algorithms are less expensive than a 3D3V particle evolution.

The hybrid approach algorithmic strength consists on the combination of mature state-of-the-art numerical techniques both for the kinetic description as well as for the fluid part. The hybrid approach leverage on the wise combination of the different algorithms using different timescales. The interaction between the kinetic and fluid scale is made possible via the bunch current, calculated and weighted on the fluid mesh. The novelty introduced by architect is the resolution of evolution equation in a time explicit domain, there is no quasi-static approximation. Moreover electron bunches are

initialized in vacuum with their self consistent field, we treat the transition from vacuum to plasma.

## SECTION IV: FEL

The electron beam accelerated by plasma are then transported to an undulator and photon production is studied with the use of the GENESIS [32] code. Considering a  $a_w = 0.8$ , the resonant wavelength is  $\lambda = 2.987$  nm. The matching to the undulator leads to  $\sigma_x = 40.6 \mu\text{m}$  and  $\sigma_y = 28.6 \mu\text{m}$  with the quadrupoles 9 cm long and set at 18 T/m. The FEL parameter results to be  $\rho = 2.51 \times 10^{-3}$ , its 3D value is  $\rho_{3D} = 1.86 \times 10^{-3}$  for  $L_{g,3D} = 0.37$  m. Simulations predicts The growth of the radiation, as given by simulations with GENESIS 1.3 [32], is shown in Fig. 5. The saturation length is about 15-25 m with emitted energy  $6.5 \mu\text{J}$  at 30 m, for a photon flux of  $9.76 \times 10^{10}$  per shot. The minimum bandwidth value, achieved at 20 m is 0.3%, while at 30 m saturation effects have increased it at 0.9%. Finally, the nominal case al has been worsened in current, emittance and energy spread by 5% and 10%, we observe a decrease in the emission respectively of 8% and 13%.

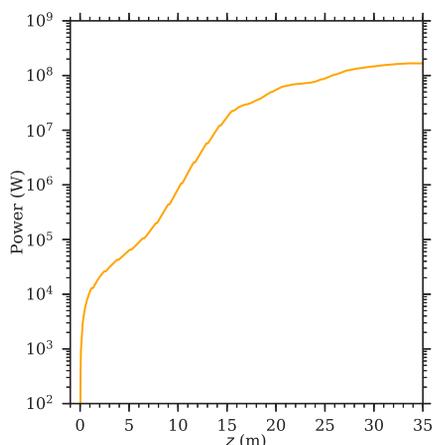


Figure 5: FEL radiation growth along the undulator length.

## CONCLUSION

This work focuses on a start-to-end simulations for a FEL machine, driven by a PWFA accelerator whose bunch is generated and accelerated by a RF photo-injector. The RF photo-injector and the subsequent X-band boosting line had been tuned to deliver to the plasma a driver able to accelerated a trailing bunch from 0.5 GeV to 1 GeV by preserving the quality of its core so to well laser inside a FEL machine. The simulation is possible by combining in cascade a series of codes, each devoted to reproduce the relevant physics in the given section.

## ACKNOWLEDGEMENTS

This work was supported by the European Unions Horizon 2020 research and innovation programme under grant agreement No. 653782. We acknowledge the CINECA

award under the IS CRA initiative, for the availability of high performance computing resources and support.

## REFERENCES

- [1] T. Tajima and J. M. Dawson, "Laser electron accelerator," *Physical Review Letters*, vol. 43, no. 4, pp. 267–270, 1979.
- [2] S. Corde *et al.*, "Multi-gigaelectronvolt acceleration of positrons in a self-loaded plasma wakefield," *Nature*, 2015.
- [3] M. Litos *et al.*, "High-efficiency acceleration of an electron beam in a plasma wakefield accelerator," *Nature*, vol. 515, no. 7525, pp. 92–95, 2014.
- [4] W. P. Leemans *et al.*, "Gev electron beams from a centimetre-scale accelerator," *Nature Physics*, vol. 2, no. 10, pp. 696–699, 2006.
- [5] H. Schwöerer *et al.*, "Laser-plasma acceleration of quasi-monoenergetic protons from microstructured targets," *Nature*, vol. 439, no. 7075, pp. 445–448, Jan. 2006.
- [6] I. Blumenfeld *et al.*, "Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator," *Nature*, vol. 445, no. 7129, pp. 741–744, Feb. 2007.
- [7] P. Muggli *et al.*, "Generation of Trains of Electron Microbunches with Adjustable Subpicosecond Spacing," *Physical Review Letters*, vol. 101, no. 5, p. 054801, Aug. 2008.
- [8] E. Kallos *et al.*, "High-Gradient Plasma-Wakefield Acceleration with Two Subpicosecond Electron Bunches," *Physical Review Letters*, vol. 100, no. 7, p. 074802, Feb. 2008.
- [9] S. Van der Meer, "Improving the power efficiency of the plasma wakefield accelerator," *CERN-PS-85-65-AA. CLIC-Note-3*, 1985.
- [10] T. Katsouleas, "Physical mechanisms in the plasma wake-field accelerator," *Physical Review A - General Physics*, vol. 33, no. 3, pp. 2056–2064, Mar. 1986.
- [11] J. B. Rosenzweig *et al.*, "Acceleration and focusing of electrons in two-dimensional nonlinear plasma wake fields," *Physical Review A (Atomic*, vol. 44, no. 1, pp. R6189–R6192, Nov. 1991.
- [12] M. Ferrario, D. Alesini, A. Bacci, M. Bellaveglia, R. Boni, M. Boscolo, P. Calvani, M. Castellano, E. Chiadroni, A. Cianchi *et al.*, "Laser comb with velocity bunching: Preliminary results at sparc," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 637, no. 1, pp. S43–S46, 2011.
- [13] P. A. Walker *et al.*, "Horizon 2020 EuPRAXIA design study," *Journal of Physics: Conference Series*, vol. 874, p. 012029, Jul. 2017.
- [14] M. Ferrario *et al.*, "Eupraxia@sparc\_lab design study towards a compact fel facility at Inf," *Nucl. Instrum. Meth.*, 2018. <http://www.sciencedirect.com/science/article/pii/S0168900218301414>
- [15] F. Villa *et al.*, "Laser pulse shaping for high gradient accelerators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 829, pp. 446–451, 2016.
- [16] A. Mostacci *et al.*, "Advanced beam manipulation techniques at sparc," in *IPAC 2011-2nd International Particle Accelerator Conference*, 2011, pp. 2877–2881.

- [17] F. Giorgianni *et al.*, “Tailoring of highly intense thz radiation through high brightness electron beams longitudinal manipulation,” *Applied Sciences*, vol. 6, no. 2, p. 56, 2016.
- [18] V. Petrillo, M. Anania, M. Artioli, A. Bacci, M. Bellaveglia, E. Chiadroni, A. Cianchi, F. Ciocci, G. Dattoli, D. Di Giovenale *et al.*, “Observation of time-domain modulation of free-electron-laser pulses by multi-peaked electron-energy spectrum,” *Physical Review Letters*, vol. 111, no. 11, p. 114802, 2013.
- [19] A. Petralia *et al.*, “Two-color radiation generated in a seeded free-electron laser with two electron beams,” *Physical review letters*, vol. 115, no. 1, p. 014801, 2015.
- [20] E. Chiadroni and *at al.*, “Beam manipulation for resonant plasma wakefield acceleration,” *Accept. Nucl. Instrum. Meth.*, vol. 865, pp. 139–143, 2017.
- [21] A. Giribono *et al.*, “Eupraxia@sparc\_lab: the high-brightness rf photo-injector layout proposal,” *Accept. Nucl. Instrum. Meth.*, 2018.
- [22] C. Vaccarezza *et al.*, “Eupraxia@sparc\_lab: Beam dynamics studies for the x-band linac,” *Accept. Nucl. Instrum. Meth.*, 2018.
- [23] M. Diomede *et al.*, “Rf design of the x-band linac for the eupraxia@sparc\_lab project,” in *Proceedings IPAC 2018*, 2018.
- [24] A. Marocchino *et al.*, “Efficient modeling of plasma wakefield acceleration in quasi-non-linear-regimes with the hybrid code Architect,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 829, pp. 386–391, sep 2016.
- [25] F. Massimo, S. Atzeni, and A. Marocchino, “Comparisons of time explicit hybrid kinetic-fluid code Architect for Plasma Wakefield Acceleration with a full PIC code,” *Journal Of Computational Physics*, vol. 327, pp. 841–850, 2016.
- [26] A. Marocchino *et al.*, “Design of high brightness Plasma Wakefield Acceleration experiment at SPARC\_LAB test facility with particle-in-cell simulations,” *Nucl. Instrum. Meth. A*, 2018.
- [27] A. Marocchino *et al.*, “Experimental characterization of the effects induced by passive plasma lens on high brightness electron bunches,” *Applied Physics Letters*, vol. 111, no. 18, p. 184101, 2017.
- [28] R. Pompili, M. Anania, M. Bellaveglia, A. Biagioni, S. Bini, F. Bisesto, E. Brentegani, G. Castorina, E. Chiadroni, A. Cianchi *et al.*, “Experimental characterization of active plasma lensing for electron beams,” *Applied Physics Letters*, vol. 110, no. 10, p. 104101, 2017.
- [29] F. Massimo, A. Marocchino, E. Chiadroni, M. Ferrario, A. Mostacci, P. Musumeci, and L. Palumbo, “Transformer ratio studies for single bunch plasma wakefield acceleration,” *Nuclear Inst. and Methods in Physics Research, A*, vol. 740, no. C, pp. 242–245, mar 2014.
- [30] J. Rosenzweig *et al.*, “Energy loss of a high charge bunched electron beam in plasma: Simulations, scaling, and accelerating wakefields,” *Physical Review Special Topics-Accelerators and Beams*, vol. 7, no. 6, p. 061302, 2004.
- [31] P. Londrillo *et al.*, “Numerical investigation of beam-driven PWFA in quasi-nonlinear regime,” *Nuclear Inst. and Methods in Physics Research, A*, vol. 740, pp. 236–241, 2014.
- [32] S. Reiche, “Genesis 1.3: a fully 3d time-dependent fel simulation code,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 429, no. 1-3, pp. 243–248, 1999.

# EFFICIENT MODELING OF LASER WAKEFIELD ACCELERATION THROUGH THE PIC CODE SMILEI IN CILEX PROJECT

Francesco Massimo\*, Arnaud Beck, Imen Zemzemi, Arnd Specka  
Laboratoire Leprince-Ringuet, 91128 Palaiseau, France

Julien Derouillat, Maison de la Simulation, 91191 Gif-sur-Yvette, France

Mickael Grech, Frédéric Pérez,

Laboratoire d'Utilisation des Lasers Intenses, 91128 Palaiseau Cedex, France

## Abstract

The design of plasma acceleration facilities requires considerable simulation effort for each part of the machine, from the plasma injector and/or accelerator stage(s), to the beam transport stage, from which the accelerated beams will be brought to the users or possibly to another plasma stage. The urgent issues and challenges in simulation of multi-stage acceleration with the Apollon laser of CILEX facility will be addressed. To simulate the beam injection in the second plasma stage, additional physical models have been introduced and tested in the open source Particle in Cell collaborative code SMILEI. The efficient initialisation of arbitrary relativistic particle beam distributions through a Python interface allowing code coupling and the self consistent initialisation of their electromagnetic fields will be presented. The comparison between a full PIC simulation and a simulation with a recently developed envelope model, which allows to drastically reduce the computational time, will be also shown for a test case of laser wakefield acceleration of an externally injected electron beam.

## INTRODUCTION

Laser Wakefield Acceleration (LWFA) is a promising technique to accelerate particles with gradients order of magnitudes higher than those of metallic accelerating cavities [1–3]. A high intensity laser pulse propagating in a plasma and of length of the order of the plasma wavelength can create a cavity empty of electrons in its wake. In this “bubble”, the generated high gradient wakefields are suitable for electron focusing and acceleration. The realization of the PetaWatt laser Apollon in the CILEX (Centre Interdisciplinaire Lumière EXtrême) facility [4] in France will pave the way to innovative LWFA experiments. The use of a second plasma stage of LWFA in the weakly nonlinear regime is considered, implying both experimental and modelization challenges. In this work we present new features in the Particle-in-Cell (PIC) code SMILEI [5] to address the simulation challenges of the project.

The length of the first plasma stage, acting as an electron injector in nonlinear regimes, is of the order of millimeters. The length of this second stage will instead need to be at least of the order of the centimeters in order to accelerate particles at high energies with the less intense fields generated in weakly nonlinear regimes. The standard PIC

technique [6] would be unfeasible for the much longer distances to simulate required by the second plasma stage. A solution to considerably reduce the computation time consists in using an envelope model for the laser pulse [7, 8]. In this approach one only needs to sample the envelope spatio-temporal scales, of the order of the plasma wavelength  $\lambda_p$  and frequency  $\omega_p = 2\pi c/\lambda_p = c/k_p$  instead of the laser wavelength  $\lambda_0$  and frequency  $\omega_0 = 2\pi c/\lambda_0$ . Doing so allows for a coarser, and cheaper, resolution while retaining all the relevant physics. The use of cylindrical symmetry in an envelope model, like in [9] would be unsuited for CILEX, since even a cylindrically symmetric beams exiting from the first plasma stage would be influenced by the intrinsic asymmetry in the focusing elements of the conventional transport line towards the second plasma stage. Thus, we developed a 3D completely parallelized envelope model for the laser-plasma dynamics, first implemented in the PIC code ALaDyn [10] and described in detail in [8]. In this paper, we briefly recall the envelope model's equations and the initialization of arbitrary beam phase distributions with their self-consistent electromagnetic fields, as initial conditions for a simulation (following the procedure described in [11, 12]). Both these features have been implemented in SMILEI. After showing the results of two validation tests of the envelope model against analytical theory, we show an application of these two features in a SMILEI simulation of a second plasma stage of LWFA.

## ENVELOPE MODEL

The hypothesis of the envelope model, i.e. a shape of the laser pulse vector potential  $\mathbf{A}$  given by a slowly varying complex envelope  $\tilde{\mathbf{A}}$  modulated by oscillations at the laser frequency  $\omega_0 = k_0 c$  can be expressed as

$$A(\mathbf{x}, t) = \text{Re}[\tilde{A} e^{ik_0(x-ct)}]. \quad (1)$$

The laser pulse is supposed to propagate in the positive  $x$  direction. Following [8], the envelope hypothesis can be inserted in D'Alembert's Equation for the laser vector potential, obtaining the envelope equation in laboratory coordinates, solved in [8] and in SMILEI:

$$\nabla \tilde{A} + 2ik_0 \left( \partial_x \tilde{A} + \frac{1}{c} \partial_t \tilde{A} \right) - \frac{1}{c^2} \partial_t^2 \tilde{A} = \chi \tilde{A}, \quad (2)$$

where  $\chi$  is the plasma susceptibility, which takes into account the envelope modification due to the presence of the

\* massimo@lir.in2p3.fr

plasma. The susceptibility term is computed as

$$\chi = \sum_p^{N_{\text{particles}}} \frac{q_p^2}{c^2 \epsilon_0 m_p} \frac{n_p}{\bar{\gamma}_p}. \quad (3)$$

where  $q_p$ ,  $m_p$ ,  $\bar{n}_p$ ,  $\bar{\mathbf{p}}_p$  and  $\bar{\gamma}_p$  are respectively the particle  $p$ 's charge, mass, number density, momentum and ponderomotive Lorentz factor, defined as

$$\bar{\gamma}_p = \sqrt{1 + \frac{\|\bar{\mathbf{p}}_p\|^2}{m_p^2 c^2} + \frac{q_p^2}{m_p^2 c^2} \frac{\|\bar{A}(\bar{\mathbf{x}}_p)\|^2}{2}}, \quad (4)$$

where  $\bar{\mathbf{x}}_p$  is the particle  $p$ 's position. The bar over the physical quantities above refers to the fact that they are averaged over the optical cycle. The same notation will be used in the following.

The electromagnetic fields  $\bar{\mathbf{E}}$ ,  $\bar{\mathbf{B}}$  evolve according to the averaged Maxwell's equations:

$$\begin{aligned} \partial_t \bar{\mathbf{B}} &= -\nabla \times \bar{\mathbf{E}} \\ \partial_t \bar{\mathbf{E}} &= c^2 \nabla \times \bar{\mathbf{E}} - c^2 \mu_0 \bar{\mathbf{J}}. \end{aligned} \quad (5)$$

As result of the averaging process [7], the particles averaged equations of motion contain the ponderomotive force term in the momentum equation and become

$$\begin{aligned} \frac{d\bar{\mathbf{x}}_p}{dt} &= \frac{\bar{\mathbf{p}}}{m_p \bar{\gamma}_p} \\ \frac{d\bar{\mathbf{p}}_p}{dt} &= q_p \left[ \bar{\mathbf{E}}(\bar{\mathbf{x}}_p) + \frac{\bar{\mathbf{p}}_p}{m_p \bar{\gamma}_p} \times \bar{\mathbf{B}}(\bar{\mathbf{x}}_p) \right] + \\ &- \frac{q_p^2}{4m_p \bar{\gamma}_p} \nabla \|\bar{A}\|^2(\bar{\mathbf{x}}_p). \end{aligned} \quad (6)$$

As described in [8], the envelope equation Eq. (2) can be discretized through centered finite differences, obtaining an explicit solver scheme which is easily parallelizable. Maxwell's equations Eq. (5) are solved as in standard PIC codes or with more advanced dispersion-free schemes [5, 6], while the particles equations of motion are solved using a modified Boris pusher scheme, described in [8]. Some equations of the envelope model, as the momentum evolution equation Eq. (7) contain the ponderomotive Lorentz factor  $\bar{\gamma}$ , which depends on the envelope  $\bar{A}$  itself. The standard PIC temporal loop [6] must be modified to solve the above equations with the schemes described in [8]. The modified temporal loop of Smilei in envelope mode is shown in Fig. 1. At each time iteration, the electromagnetic force (including the ponderomotive force) acting on each particle is interpolated from the grid. The susceptibility of each particle is then projected on the grid, following Eq. (3). The particles momenta are updated through the use of the force they are subject to, solving Eq. (7) through a modified Boris pusher described in [8]. The envelope equation is then solved, advancing the envelope value in time. The explicit envelope solver scheme is described in detail in [8]. The particles positions are updated solving Eq. (6) as described in [8].

The particles current density is projected then on the grid and the electromagnetic fields are advanced through a Yee scheme [6]. The loop iteration can then be repeated until the end of the simulation.

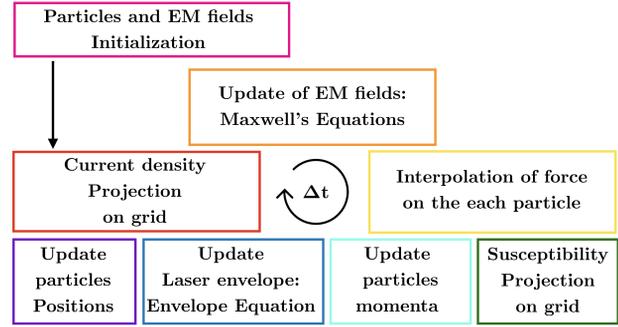


Figure 1: The envelope PIC temporal loop, showing the operations performed at each temporal iteration.

## RELATIVISTIC BEAM INITIALIZATION

Under the hypothesis of monoenergetic phase space distribution, the electromagnetic fields of a relativistic particle beam can be initialized in a simulation through the technique explained in [11, 12], here briefly recalled. This technique has been implemented in SMILEI to perform simulations of LWFA with external injection of an electron beam into a second plasma stage. Once the beam charge density  $\bar{\rho}$  is known, the “relativistic Poisson’s equation”, i.e.

$$\left( \frac{1}{\gamma_0^2} \partial_x^2 + \nabla_\perp^2 \right) \bar{\Phi} = -\frac{\bar{\rho}}{\epsilon_0}, \quad (8)$$

gives  $\bar{\Phi}$  and the beam self-consistent electromagnetic fields can be found through the relations:

$$\bar{\mathbf{E}} = \left( -\frac{1}{\gamma_0^2} \partial_x, -\partial_y, -\partial_z \right) \bar{\Phi} \quad (9)$$

$$\bar{\mathbf{B}} = \beta_0 c \hat{x} \times \bar{\mathbf{E}}. \quad (10)$$

SMILEI allows to easily define an initial beam distribution through its Python input interface, which permits to define ideal bunches as well as load datafiles with beam distributions obtained from transport codes.

## ENVELOPE MODEL BENCHMARKS AGAINST ANALYTICAL THEORY

As first tests for the envelope model feature of SMILEI, this section presents the simulation of vacuum diffraction of a Gaussian laser beam and of the laser wakefield inside a plasma in the linear regime. Figure 2 reports the comparison between the simulated rms waist size of a Gaussian beam with initial waist  $w_0 = 12 \mu\text{m}$  and the vacuum Rayleigh diffraction formula  $w(\bar{x})/w_0 = \sqrt{1 + \bar{x}^2}$ , where  $\bar{x} = x/Z_R$  is the propagation distance divided by the Rayleigh length

$Z_R = \pi w_0^2 / \lambda_0$  [13]. The longitudinal grid cell size, the transverse grid cell size and the timestep are respectively  $\Delta x = 0.69 c / \omega_0$ ,  $\Delta y = \Delta z = 5 c / \omega_0$ ,  $\Delta t = 0.57 1 / \omega_0$ .

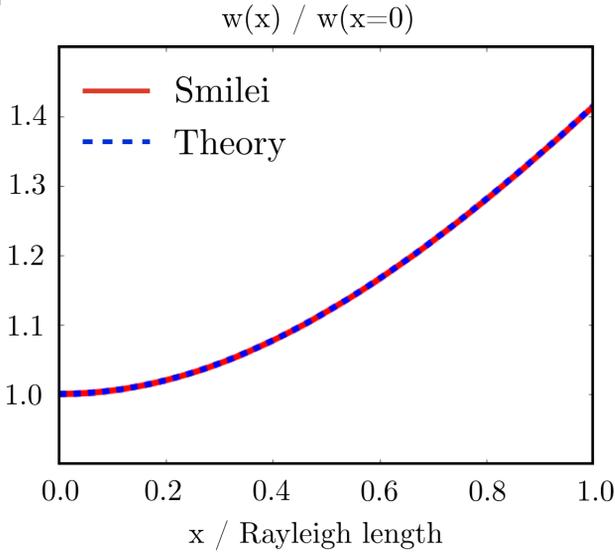


Figure 2: Evolution of the waist size  $w(x)$  of a Gaussian laser pulse simulated through the envelope model against analytical theory.

The second validation test we report (see Fig. 3) is a comparison of the results of the envelope model against the analytical theory for the longitudinal wakefield  $E_x$  of a Gaussian laser pulse in a plasma in the linear regime (Eq. (37a) from [14]):

$$E_x(x, r, t) = \frac{mc^2 k_p^2}{e} \frac{1}{4} \int_z^\infty a(x', r, t) e^{-\frac{r^2}{w_0^2}} \cos[k_p(x-x')] dx', \quad (11)$$

for a laser pulse with envelope  $A(x, r, t) = a(x, t) e^{-r^2/w_0^2}$  and  $w_0 \gg \lambda_p$ . For this benchmark, we chose an initial waist size of  $12 \mu\text{m}$  and a Gaussian longitudinal envelope  $a(x, t = 0)$  with initial FWHM duration in intensity  $\tau_0 = 20 \text{ fs}$  and peak value  $a_0 = 0.01$ , to remain in a linear regime. The plasma density is  $n_0 = 0.0017 n_c$ . The same grid cell size and timesteps as the vacuum diffraction benchmark have been used, and 8 particles per cell sample the plasma.

For both the benchmarks, we observe a very good agreement with the analytical predictions.

## CASE STUDY: EXTERNAL INJECTION IN A SECOND PLASMA STAGE

As a benchmark and as an example of application of both the envelope model and the relativistic beam field initialization, we present the preliminary results of two simulations, one performed with a standard PIC procedure (hereafter called “standard laser simulation”) and one with an envelope model for the laser (hereafter called “envelope simulation”). The physical setup is the external injection of a relativistic electron beam (whose fields have been initialized with the procedure described above) into the plasma wave in the wake

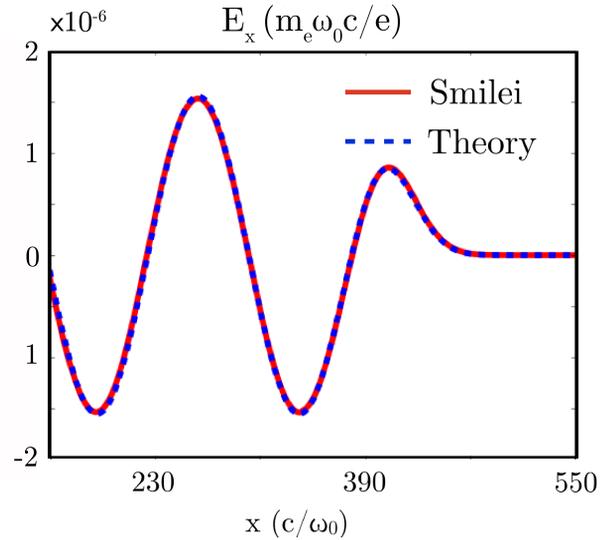


Figure 3: Simulated longitudinal electric field in the wake of a Gaussian laser pulse modeled through the envelope against the analytical linear theory. The laser is propagating towards right.

of a laser in a plasma stage. The laser, plasma and electron beam parameters, briefly recalled in the following, have been chosen from [15]. The driver laser is a Gaussian pulse, linearly polarized in the  $y$  direction, with waist size  $w_0 = 45 \mu\text{m}$ ,  $a_0 = \sqrt{2}$  and initial FWHM duration in intensity  $\tau_0 = 108 \text{ fs}$ , focused at the plasma entrance. The initial laser center position is at a distance  $2c\tau_0$  from the plasma entrance. The plasma has an idealized parabolic density profile along the transverse direction with density  $n_e(r) = n_0 \left(1 + \frac{\Delta n}{n_0} \frac{r^2}{r_0^2}\right)$ , with  $r$  the distance to axis,  $n_0 = 1.5 \cdot 10^{17} \text{ cm}^{-3}$ ,  $\frac{\Delta n}{n_0} = 0.25$ ,  $r_0 = 45 \mu\text{m}$ . The relativistic electron beam, with charge  $30 \text{ pC}$  and normalized emittance  $1 \text{ mm}\cdot\text{mrad}$ , has an initial energy of  $150 \text{ MeV}$  with  $0.5\%$  rms energy spread and is initially positioned at waist at a distance  $3/4\lambda_p$  after the laser pulse. The beam longitudinal and transverse rms sizes are  $\sigma_x = 2 \mu\text{m}$  and  $\sigma_y = \sigma_z = 1.3 \mu\text{m}$ . The grid cell sizes for the standard laser simulation are  $\Delta x_{\text{laser}} = \lambda_0/32$ ,  $\Delta y = \Delta z = \lambda_0$  and the integration timestep is  $\Delta t_{\text{laser}} = 0.95\Delta x/c$ . For the envelope simulation, because the length of the laser pulse envelope is so much longer than a single optical cycle, the longitudinal grid cell size and the integration timestep could be set to  $\Delta x_{\text{envelope}} = 16\Delta x_{\text{laser}}$ ,  $\Delta t_{\text{envelope}} = 0.8\Delta x_{\text{envelope}}$  respectively. The transverse cell length is the same as in the standard laser simulation. In both simulations, the plasma is sampled with 8 particles per cell and the beam is sampled with  $10^6$  particles. Figures 4 and 5 compare the colormaps of the electron density  $n_e$  and of the longitudinal electric field  $E_x$  for the two simulations after a propagation distance of  $3 \text{ mm}$  in the plasma. Figures 6 and 7 compare the same quantities on the axis at the same propagation distance. Apart from a minimum lag of the electron beam behind the laser pulse in the case of the envelope simulation, the results have a very good agreement. The envelope simulation reproduces

the laser and plasma dynamics, as well as the beam loading of the bunch on the plasma wave. Further investigations are necessary to understand the differences in the dynamics of the electron beam in the two simulations. The greater dilution of the electron beam in the envelope simulation may be caused by a different growth rate of the numerical Cherenkov radiation [16], due to the difference mesh cell size and integration timestep compared to the standard laser simulation. The envelope simulation needed 4400 CPU hours, while the standard laser simulation needed a total amount of resources twenty times as large.

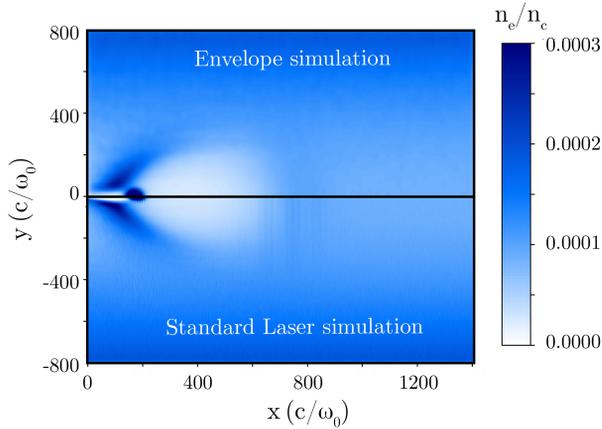


Figure 4: Electron charge density, normalized by the critical density  $n_c$ , after 3 mm of propagation in the plasma.

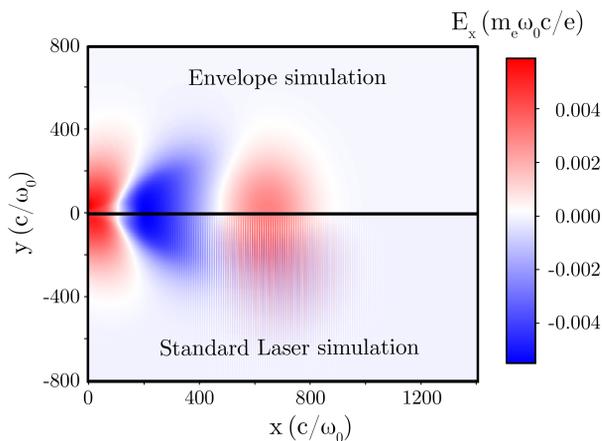


Figure 5: Longitudinal electric field  $E_x$  after 3 mm of propagation in the plasma.

## CONCLUSIONS

A 3D explicit envelope model and a procedure for the initialization of relativistic particle beams electromagnetic fields have been implemented in the PIC code SMILEI. These techniques have been developed to face the modelization challenges of the multi-stage LWFA experiments in the CILEX project. We reported the results benchmarks of the

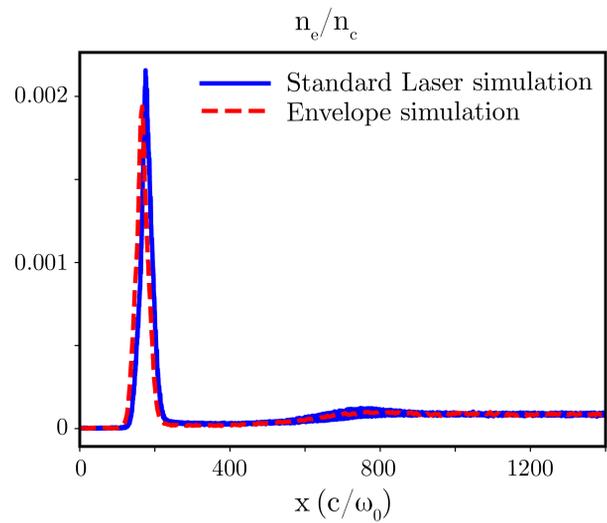


Figure 6: Electron charge density on the propagation axis, normalized by the critical density  $n_c$ , after 3 mm of propagation in the plasma.

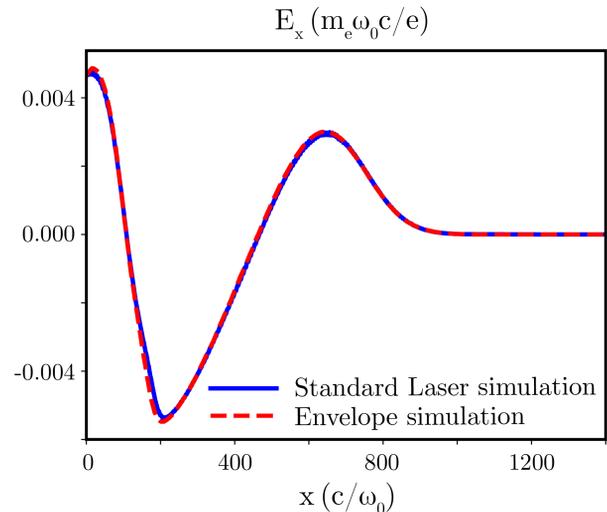


Figure 7: Longitudinal electric field  $E_x$  on the propagation axis after 3 mm of propagation in the plasma.

envelope model against analytical theory for laser vacuum diffraction and linear laser wakefield in the plasma. We presented also the comparison between a standard 3D PIC simulation and a 3D PIC simulation with envelope model for the laser in the case of external injection of an electron beam in a plasma channel. Excellent agreement is found after 3 mm of propagation, with a computational speedup of 20 using the envelope model.

## ACKNOWLEDGEMENTS

The project has received funding from the P2IO LabEx (ANR-10-LABX-0038) in the framework “Investissements d’Avenir” (ANR-11-IDEX-0003-01) managed by the Agence Nationale de la Recherche (ANR, France). This work was granted access to the HPC resources of TGCC/CINES un-

der the allocation 2018-A0030510062 and Grand Challenge “Irene” 2018 project gch0313 made by GENCI. The authors are grateful to the TGCC and CINES engineers for their support. The authors thank engineers of the LLR HPC clusters for resources and help. The authors are grateful to the ALaDyn development team for the help and discussions during the development of the envelope model, in particular D. Terzani, A. Marocchino and S. Sinigardi.

## REFERENCES

- [1] T. Tajima, J.M. Dawson, *Phys. Rev. Lett.* 43 (4), 1979, p. 267.
- [2] E. Esarey, C. B. Schroeder, W. P. Leemans, *Rev. Mod. Phys.* 81, 2009, p. 1229.
- [3] V. Malka, *Phys. Plasmas* 19, 2012, p. 055501.
- [4] B. Cros et al., *Nucl. Instr. and Meth. in Phys. A* 740, 2014, p. 27.
- [5] J. Derouillat, A. Beck, F. Pérez, T. Vinci, M. Chiaramello, A. Grassi, M. Flé, G. Bouchard, I. Plotnikov, N. Aunai, J. Dargent, C. Riconda, M. Grech, *Comput. Phys. Comm.* 222, 2018, p. 351.
- [6] C. K. Birdsall, A. B. Langdon, *Plasma Physics Via Computer Simulation* CRC Press, Boca Raton, 2004.
- [7] B.M. Cowan, D.L. Bruhwiler, E. Cormier-Michel, E. Esarey, C.G.R. Geddes, P. Messmer, K. M. Paul, *Journal of Computational Physics* 230, 2011, p. 61.
- [8] D. Terzani, P. Londrillo, “Yet a faster numerical implementation of envelope model for laser-plasma dynamics,” submitted to *Computer Physics Communications*.
- [9] C. Benedetti, E. Esarey, W. P. Leemans, C. B. Schroeder, in *Proc. ICAP 2012*, 2012.
- [10] S. Sinigardi, A. Marocchino, P. Londrillo, D. Terzani, F. Massimo, F. Mira, A. Sgattoni, *ALaDyn/ALaDyn: ALaDyn v2018.2 (Version v2.1.0)*. Zenodo, 2018, <http://doi.org/10.5281/zenodo.1406920>.
- [11] F. Massimo, A. Marocchino, A.R. Rossi, *Nucl. Instrum. Meth. A* 829, 2016, p. 378.
- [12] J.-L. Vay, *Phys. Plasmas* 15, 2008, p. 056701.
- [13] O. Svelto, *Principles of Lasers* Springer US, 2012.
- [14] R. Lehe, M. Kirchen and I. A. Andriyash, B. B. Godfrey, J.-L. Vay, *Comput. Phys. Comm.* 203, 2015, p. 66.
- [15] X. Li, A. Mosnier, P. A. P. Nghiem, “Design of a 5 GeV laser plasma accelerating module in the quasi-linear regime”, *Nucl. Instrum. Meth. A*, in press, 2018.
- [16] R. Lehe, A. Lifschitz, C. Thauray, V. Malka, X. Davoine, *Phys. Review Special Topics Acc. and Beams* 16, 2013, p. 021301.

# UPGRADE OF MAD-X FOR HL-LHC PROJECT AND FCC STUDIES

L. Deniau\*, A. Latina, T. Persson, I. Shreyber, P. Skowronski,  
H. Burkhardt, R. De Maria, M. Giovannozzi, J.M. Jowett, F. Schmidt, CERN, Meyrin, Switzerland  
T. Gläbke, HIT, Heidelberg, Germany

## Abstract

The design efforts for the High Luminosity upgrade project of the Large Hadron Collider (HL-LHC) and for the FCC-ee studies required significant extensions of the MAD-X code, widely used for designing and simulating particle accelerators. The modelling of synchrotron radiation effects has recently been reviewed, improved, and tested on the lattices of ESRF, LEP, and CLIC Final Focus System. The results were cross checked with several codes, such as AT, PLACET, Geant4, and MAD8. The implementation of space charge has been deeply restructured into a fully modular design. The linear coupling calculation has been completely reviewed and improved to ensure its robustness in the presence of strong coupling effects as is the case for some HL-LHC studies. The slicing module has been generalized to allow for generating thick slices of bending magnets, quadrupoles and solenoids. The SBEND element has been extended to take into account not only the bending angle, but also the integrated dipole strength. Patches have been added to the list of supported elements. Finally, the PTC program inside MAD-X has been extended to provide the tracking of resonance driving terms along lattices, as well as an AC dipole element.

## INTRODUCTION

MAD-X is a code to simulate beam dynamics and design beam optics, which was released in 2002. Although it has been tailored to the needs of the LHC, it still remains the most used tool for optics design inside and outside CERN. In order to extend the MAD-X capabilities to better satisfy the users' needs, in particular for the HL-LHC project and the FCC study, several new features have been implemented. An example of such a new feature is the synchrotron radiation, which has a negligible effect on the beam dynamics in the LHC, but plays a major role for the FCC-ee future collider. An element to rotate the coordinate system along the  $x$ - and  $y$ -direction has also recently been added.

Moreover, to be ready for future studies, several parts of the code and the underlying physics have been reviewed or improved. In this paper we present the clean up of space-charge code as well as of the review of the linear coupling.

An important extension to MAD-X is the Polymorphic Tracking Code (PTC) by E. Forest [1]. In this paper, some of the recently added features, including the possibility to obtain the Resonance Driving Terms (RDT) in the PTC\_TWISS table will be presented and discussed.

\* laurent.deniau@cern.ch

## SYNCHROTRON RADIATION EFFECTS

The effects of synchrotron radiation were originally implemented in MAD8 for the needs of the high-energy  $e^+e^-$  collider ring LEP at CERN [2]. This implementation was ported to MAD-X and has recently been reviewed and improved. Note that the review of the implementation of synchrotron radiation emission as a quantum phenomenon revealed the presence of a bug in MAD-X versions earlier than 5.04.01. The bug was corrected and the new implementation tested in several cases.

Synchrotron radiation emission is included in the modules TWISS, TRACK, and EMIT at four different levels:

1. No radiation, corresponding to the usual Hamiltonian dynamics.
2. Deterministic radiation, in which all particles radiate the same energy as a single particle on the closed-orbit (TWISS and TRACK). This gives a Hamiltonian system with the correct tunes and closed orbit, including the “stable phase angle” and “energy sawtooth” in electron rings.
3. Deterministic radiation with full dependence on canonical coordinates to generate radiation damping naturally (EMIT, TRACK). Since this method exposes the full dissipative structure of the non-linear phase space dynamics, it is the preferred method for dynamic aperture calculations in high-energy lepton rings [3].
4. Tracking with individual stochastic photon emissions, to provide quantum excitation and particle distributions (TRACK) and equilibrium emittances from first principles. Both the probability of photon emission, i.e. an instantaneous Poisson distribution, and the generation of the random photon-energy distribution depend on the *local* magnetic field and canonical coordinates of the particle.

The MAD-X implementation has been benchmarked against MAD8 using a conversion of the LEP-optics database to the MAD-X format. Rather than attempting to reproduce old results dating from the LEP times, this conversion allowed to run equivalent simulations with MAD8 and MAD-X so that results could be compared in detail. Quantities such as radiation damping rates, equilibrium emittances, energy sawtooth, and the Bassetti component [4] of the horizontal closed orbit  $x_c - D_x p_{1c}$ , where  $x_c$  is the position coordinate of the closed orbit, and  $p_{1c}$  is the momentum deviation of the closed orbit, are in excellent agreement between the two codes, confirming that the physics implemented in MAD8 has been preserved in MAD-X.

In addition, MAD-X was tested against other established codes capable of dealing with particle tracking in presence

of synchrotron radiation, such as PLACET [5] and AT [6], providing in all cases an excellent agreement.

Among the test cases considered, the quadrupole doublet of the final focus of CLIC was studied. In these magnets the radiation effects have a significant impact on the beam size and beam shape at the interaction point (IP), through the so-called Oide effect [7]. Figure 1 shows a direct comparison between the phase space of a multi-particle beam, composed of  $10^5$  macro-particles, at the IP of CLIC assuming a 3 TeV energy in the centre-of-mass, as simulated by the tracking code PLACET and by MAD-X, including stochastic photon emission.

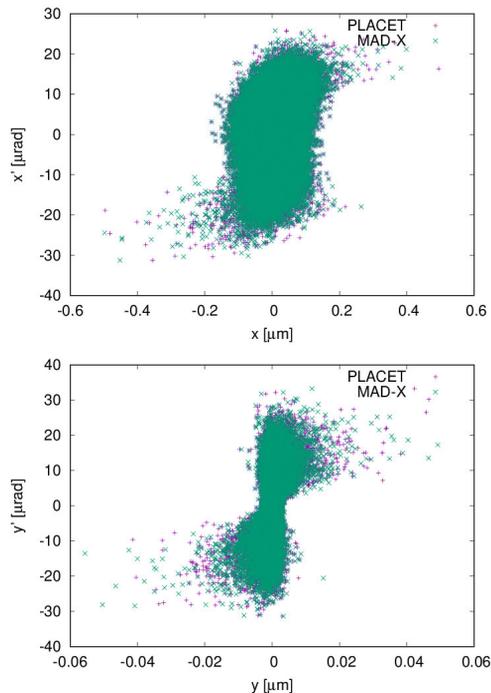


Figure 1: Comparison of phase-space distributions at the IP of CLIC, at 3 TeV centre-of-mass energy, as tracked by PLACET and MAD-X through the final-focus quadrupole doublet, with synchrotron radiation emission simulated as quantum photon emission.

## SPACE-CHARGE IMPLEMENTATION

The frozen space-charge model was introduced in MAD-X for the CERN PS injector studies since 2012 [8], and evolved over the past years. The initial implementation was interleaved within the tracking code making efficient support difficult, hence the space-charge implementation has been reviewed and cleaned up. No modifications to the actual physics modelling have been introduced for the time being. However, the code was restructured and made more modular. In the previous implementation the space-charge and beam-beam codes had been interleaved with the code of the modules TWISS and TRACK. A new FORTRAN module was created to gather together all routines related to space-charge

and beam-beam, and a new set of initialization routines have been created to minimize the coupling between the code and other modules like TWISS and TRACK.

The new code has been designed so to facilitate further separation of the space-charge routines from the beam-beam routines, and to open the possibility to create a dedicated MAD-X space-charge command in future. It will also facilitate parallel execution using OPENMP directives.

## REVIEW OF LINEAR COUPLING

This section establishes the equations found in the MAD8 physics guide [9] as implemented in MAD-X, and proposes some improvements. A review [10] has been launched after some puzzling results about (nonphysical) negative-value beta functions reported in the framework of HL-LHC studies carried out with strong linear-coupling sources.

The transverse linear coupling calculation for stable motion consists of finding a similarity  $R_M$  that transforms a map  $M$  describing the motion from  $s_1$  to  $s_2$  into its normal form  $M_{\perp}$ .

$$\vec{X}(s_2) = M\vec{X}(s_1),$$

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = R_M \begin{pmatrix} E & 0 \\ 0 & F \end{pmatrix} R_M^{-1} = R_M M_{\perp} R_M^{-1}.$$

A linear dynamics is fully described by an  $s$ -dependent quadratic Hamiltonian around a reference trajectory

$$\mathcal{H} = \frac{1}{2} \vec{X}^T H \vec{X},$$

where  $H$  is a real symmetric matrix that characterizes the equations of motion

$$x'_i = \frac{\partial \mathcal{H}}{\partial p_i} \quad \text{and} \quad p'_i = -\frac{\partial \mathcal{H}}{\partial x_i} \quad \Rightarrow \quad \vec{X}' = S H \vec{X},$$

where  $S$  is the fundamental symplectic unit matrix:

$$S = \begin{pmatrix} S_2 & 0 \\ 0 & S_2 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad S^{-1} = S^T = -S, \quad S^2 = -I.$$

The map  $M$  is the linear solution of these equations of motion and any bilinear Hamiltonian can be brought to the following form by a suitable canonical transformation:

$$\mathcal{H} = \frac{p_x^2 + p_y^2}{2(1 + \delta_p)} + \frac{1}{2}(F x^2 + G y^2) + K x y + L(y p_x - x p_y),$$

where the coefficients  $F, G, K, L$  are related to the magnet strengths  $k_0$  (dipole),  $k_1$  (quadrupole) and  $k_s$  (solenoid), and the curvature  $h$  by [10]:

$$F = h k_0 + k_1 + k_s^2, \quad G = -k_1 + k_s^2, \\ K = k_1^s, \quad L = k_s.$$

### Initialization of coupling calculation

As  $M \in \text{Sp}(4, \mathbb{R})$  then

$$\text{eig}(M) = \text{eig}(M^{-1}) = \{(\lambda_i, \lambda_i^{-1}), i = 1..2\},$$

$$\text{eig}(M + \bar{M}) = \{\Lambda_i = \lambda_i + \lambda_i^{-1}, i = 1..2\},$$

where  $\bar{M}$  denotes the symplectic conjugate of  $M$ . The characteristic polynomial for the coupled motion reads

$$\begin{aligned} \det(M + \bar{M} - \Lambda I) &= \begin{vmatrix} A + \bar{A} - \Lambda I & B + \bar{C} \\ C + \bar{B} & D + \bar{D} - \Lambda I \end{vmatrix} \\ &= \begin{vmatrix} (\text{tr } A - \Lambda)I & \overline{C + \bar{B}} \\ C + \bar{B} & (\text{tr } D - \Lambda)I \end{vmatrix} = 0, \\ &(\text{tr } A - \Lambda)(\text{tr } D - \Lambda) - |C + \bar{B}| = 0, \end{aligned}$$

which can be solved for the eigenmodes [11]

$$\begin{aligned} \Lambda^2 - (\text{tr } A + \text{tr } D)\Lambda + \text{tr } A \text{tr } D - |C + \bar{B}| &= 0, \\ \Delta &= (\text{tr } A - \text{tr } D)^2 + 4|C + \bar{B}|, \\ \Lambda_A &= \frac{1}{2}(\text{tr } A + \text{tr } D) + \frac{1}{2} \text{sign}(\text{tr } A - \text{tr } D)\sqrt{\Delta}, \\ \Lambda_D &= \Lambda_A^{-1} (\text{tr } A \text{tr } D - |C + \bar{B}|), \\ \Lambda &= \lambda + 1/\lambda \Rightarrow \lambda^2 - \Lambda\lambda + 1 = 0 \\ &\Rightarrow \lambda = \frac{1}{2} \left( \Lambda \pm \sqrt{\Lambda^2 - 4} \right), \\ \lambda \in \mathbb{C} \setminus \mathbb{R} &\Leftrightarrow \Lambda \in (-2, 2), \end{aligned}$$

and where stable linear motion implies

$$\begin{aligned} |\lambda| = 1 &\Leftrightarrow \lambda^{-1} = \lambda^* \Leftrightarrow \lambda = e^{\pm i\mu}, \mu \in [0, 2\pi), \\ \Lambda_{A,D} &= \lambda_{A,D} + 1/\lambda_{A,D} = 2 \cos \mu_{A,D}, \\ (\Lambda_A - \Lambda_D)^2 &= 4(\cos \mu_A - \cos \mu_D)^2 = \Delta. \end{aligned}$$

For any non-zero eigenvectors  $\vec{X}$  and  $\vec{Y}$ , we have [12]

$$\begin{aligned} \begin{pmatrix} (\text{tr } A - \Lambda_A)I & B + \bar{C} \\ C + \bar{B} & (\text{tr } D - \Lambda_A)I \end{pmatrix} \begin{pmatrix} X \\ R_A X \end{pmatrix} &= 0, \\ \begin{pmatrix} (\text{tr } A - \Lambda_D)I & B + \bar{C} \\ C + \bar{B} & (\text{tr } D - \Lambda_D)I \end{pmatrix} \begin{pmatrix} R_D Y \\ Y \end{pmatrix} &= 0, \end{aligned}$$

leading to the solutions (full coupling  $R_{A,D} \approx \alpha I$ )

$$R_A = \frac{C + \bar{B}}{\Lambda_A - \text{tr } D}, \quad R_D = \frac{B + \bar{C}}{\Lambda_D - \text{tr } A} = -\bar{R}_A.$$

Noting  $R = R_A = -\bar{R}_D$ , we find:

$$R = -\left( \frac{1}{2}(\text{tr } A - \text{tr } D) + \frac{1}{2} \text{sign}(\text{tr } A - \text{tr } D)\sqrt{\Delta} \right)^{-1} (C + \bar{B}).$$

The similarity  $R_M$  that block-diagonalizes  $M$  into  $M_{\perp}$  can be built from the previous eigenvectors:

$$\begin{aligned} M_{\perp} &= R_M^{-1} M R_M = g^2 \bar{R}_M M R_M \\ &= g^2 \begin{pmatrix} I & R_D \\ R_A & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & R_D \\ R_A & I \end{pmatrix} \\ &= g^2 \begin{pmatrix} I & -\bar{R} \\ R & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & \bar{R} \\ -R & I \end{pmatrix} = \begin{pmatrix} E & 0 \\ 0 & F \end{pmatrix}, \end{aligned}$$

where  $g$  has to be determined such that  $R_M^{-1} R_M = g^2 \bar{R}_M R_M = I$ . Solving last equation for  $E$  and  $F$  gives:

$$\begin{aligned} E &= g^2(A - \bar{R}C - (BR - \bar{R}DR)) \\ &= (1 + |R|)^{-1}(A - \bar{R}C)(1 + |R|) \\ &= A - \bar{R}C = A - BR, \\ F &= g^2(D + RB + (C\bar{R} + RA\bar{R})) \\ &= (1 + |R|)^{-1}(D + RB)(1 + |R|) \\ &= D + RB = D + C\bar{R}. \end{aligned}$$

The last form of these equations are mentioned in the MAD8 physics guide and used by MAD-X for initializing the coupling calculation.

The factor  $g^{-1}$  represents the strength of the coupling and  $gR$  its structure between planes  $A$  and  $D$ , i.e. the  $2 \times 2$  Rmat coupling matrix in MAD-X. To determine  $g$ , we solve:

$$g^2 \bar{R}_M R_M = g^2 \begin{pmatrix} I + \bar{R}R & 0 \\ 0 & I + R\bar{R} \end{pmatrix} = g^2 (1 + |R|) I = I.$$

That is  $g = (1 + |R|)^{-\frac{1}{2}} = |R_M|^{-\frac{1}{2}}$ , hence  $gR_M$  is symplectic and so is  $M_{\perp}$ . We can also note that  $1 + R\bar{R} > 0 \Rightarrow |C + \bar{B}| > -(\Lambda_{A,D} - \text{tr } D, A)^2$ . The Edwards–Teng parametrization of the coupling structure  $gR$  uses  $g = \cos \theta$  and  $D = -R \tan \theta$ , where  $\theta$  is the angle of their ‘‘symplectic rotation’’ of planes  $A$  and  $D$  [13].

From the characteristic polynomial using either  $\Lambda_{A,D}$ , we can express the coupling strength in terms of the eigenmodes:

$$\begin{aligned} 1 - R_A R_D &= 1 - \frac{|C + \bar{B}|}{(\Lambda_A - \text{tr } D)(\Lambda_D - \text{tr } A)} = \frac{\Lambda_D - \Lambda_A}{\Lambda_D - \text{tr } A} \\ g &= \left( \frac{\Lambda_D - \Lambda_A}{\Lambda_D - \text{tr } A} \right)^{-\frac{1}{2}} = \left( \frac{\Lambda_A - \Lambda_D}{\Lambda_A - \text{tr } D} \right)^{-\frac{1}{2}}. \end{aligned}$$

The Twiss parameters are calculated from  $E$  and  $F$  by:

$$E = \begin{pmatrix} E_{1,1} & E_{1,2} \\ E_{2,1} & E_{2,2} \end{pmatrix} = \begin{pmatrix} \cos \mu_A + \alpha_A \sin \mu_A & \beta_A \sin \mu_A \\ -\gamma_A \sin \mu_A & \cos \mu_A - \alpha_A \sin \mu_A \end{pmatrix},$$

$$\cos \mu_A = \frac{1}{2} \text{tr } E,$$

$$\sin \mu_A = \text{sign}(E_{1,2}) \sqrt{-E_{1,2} E_{2,1} - \left( \frac{E_{1,1} - E_{2,2}}{2} \right)^2}$$

$$\beta_A = \frac{E_{1,2}}{\sin \mu_A}, \quad \gamma_A = -\frac{E_{2,1}}{\sin \mu_A}, \quad \alpha_A = \frac{E_{1,1} - E_{2,2}}{2 \sin \mu_A},$$

$$M_{\perp} \in \text{Sp}(4, \mathbb{R}) \Rightarrow \det E = 1 \Rightarrow \beta_A \gamma_A - \alpha_A^2 = 1.$$

The same calculation is performed using  $F$  for the second mode.

### Propagation of coupling calculation

From the known decoupled map  $M_{1\perp}$  at point 1 (from initialization) and the transport map  $M_{12}$  from points 1 to 2,

we look after the unknown decoupled map  $M_{2\perp}$ :

$$\begin{aligned} M_2 &= M_{12}M_1M_{12}^{-1} \\ &= M_{12} \left( R_{M_1}M_{1\perp}R_{M_1}^{-1} \right) M_{12}^{-1}, \\ M_{2\perp} &= R_{M_2}^{-1}M_2R_{M_2} \\ &= \left( R_{M_2}^{-1}M_{12}R_{M_1} \right) M_{1\perp} \left( R_{M_1}^{-1}M_{12}^{-1}R_{M_2} \right) \\ &= W_{12}M_{1\perp}W_{12}^{-1}. \end{aligned}$$

Since the eigenmodes 1 and 2 in  $M_{1\perp}$  and  $M_{2\perp}$  are independent from each other,  $W_{12}$  is either block or anti-block diagonal and equality  $R_{M_2}W_{12} = M_{12}R_{M_1}$  can take the forms:

$$\begin{aligned} g_2 \begin{pmatrix} I & \bar{R}_2 \\ -R_2 & I \end{pmatrix} \begin{pmatrix} E_{12} & 0 \\ 0 & F_{12} \end{pmatrix} &= g_1 \begin{pmatrix} A_{12} & B_{12} \\ C_{12} & D_{12} \end{pmatrix} \begin{pmatrix} I & \bar{R}_1 \\ -R_1 & I \end{pmatrix}, \\ g_2 \begin{pmatrix} I & \bar{R}_2 \\ -R_2 & I \end{pmatrix} \begin{pmatrix} 0 & E_{12} \\ F_{12} & 0 \end{pmatrix} &= g_1 \begin{pmatrix} A_{12} & B_{12} \\ C_{12} & D_{12} \end{pmatrix} \begin{pmatrix} I & \bar{R}_1 \\ -R_1 & I \end{pmatrix}. \end{aligned}$$

Solving for  $E_{12}$ ,  $F_{12}$  and  $R_2$  for the block (left) and anti-block (right) diagonal cases gives:

$$\begin{array}{l|l} E_{12} = g_{12}(A_{12} - B_{12}R_1) & E_{12} = g_{12}(B_{12} + A_{12}\bar{R}_1) \\ F_{12} = g_{12}(D_{12} + C_{12}\bar{R}_1) & F_{12} = g_{12}(C_{12} - D_{12}R_1) \\ R_2 = -g_{12}(C_{12} - D_{12}R_1)E_{12}^{-1} & R_2 = -g_{12}(D_{12} + C_{12}\bar{R}_1)E_{12}^{-1} \end{array}$$

where  $g_{12} = g_1g_2^{-1}$ . The sector map  $M_{2\perp}$  at point 2 is then calculated by propagating  $M_{1\perp}$  through  $M_{12}$  using,

$$\begin{array}{l|l} E_2 = E_{12}E_1\bar{E}_{12}/|E_{12}| & E_2 = E_{12}F_1\bar{E}_{12}/|E_{12}| \\ F_2 = F_{12}F_1\bar{F}_{12}/|F_{12}| & F_2 = F_{12}E_1\bar{F}_{12}/|F_{12}| \\ R_2 = -(C_{12} - D_{12}R_1)\frac{\bar{E}_{12}}{|E_{12}|} & R_2 = -(D_{12} + C_{12}\bar{R}_1)\frac{\bar{E}_{12}}{|E_{12}|} \end{array}$$

A flip mode [14] was introduced in MAD-X to solve the problem of negative beta functions that occurred when  $|A_{12} - B_{12}R_1| < \varepsilon$  or  $|D_{12} + C_{12}\bar{R}_1| < \varepsilon$  for some small positive  $\varepsilon$ , and switch from block to anti-block diagonal solutions. So far, MAD-X uses a simplified version of these equations, and generates a warning if the number of flips is odd.

The last step consist of the propagation of the Twiss parameters through  $M_{12}$  using  $T_2 = W_{12}T_1W_{12}^{-1}$  where,

$$T^{(E,F)} = \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix},$$

and gives for  $T_2^{(E)}$  from  $E_{12}$  and  $T_1^{(E)}$ :

$$\begin{aligned} \alpha_2 &= -((E_{2,1}\beta_1 - E_{2,2}\alpha_1) \times \\ &\quad (E_{1,1}\beta_1 - E_{1,2}\alpha_1) + E_{1,2}E_{2,2})/(|E_{12}|\beta_1) \\ \beta_2 &= ((E_{1,1}\beta_1 - E_{1,2}\alpha_1)^2 + E_{1,2}^2)/(|E_{12}|\beta_1) \\ \mu_2 &= \mu_1 + \tan^{-1}(E_{1,2}, (E_{1,1}\beta_1 - E_{1,2}\alpha_1)) \\ \gamma_2 &= (1 + \alpha_2^2)/\beta_2. \end{aligned}$$

The same calculation is performed using  $F$  for the second mode.

## SLICING MODULE GENERALIZATION

MAD-X contains a slicing module called MAKETHIN. It was originally designed to provide an automatic translation of the thick lattice description to a symplectic thin-lens description, suitable for tracking codes. Previously, both thick and thin lattice description had to be generated and maintained by hand. With MAKETHIN, a single thick-lattice description was sufficient and a thin version was generated automatically when needed for thin-lens tracking. For this, MAKETHIN automatically translates all thick magnets into slices of zero-length multipoles. The number of slices can be chosen using the standard MAD-X selection, i.e. by element class, range or matching patterns in element names. The position of the slices can be chosen as equidistant (MAKETHIN, STYLE=SIMPLE) or using an optimized algorithm (selected by MAKETHIN, STYLE=TEAPOT), which minimizes the beta-beating introduced by using a small number of slices [15].

Slicing turned out to be also very useful for other purposes than thin-lens tracking. The markers generated by MAKETHIN at ends and between slices can be used to obtain the value of optical functions or perform aperture checks within elements. MAKETHIN has gradually become a rather general slicing and automatic lattice editing tool, used also for purposes like aperture or Twiss-parameter interpolation. Elements can be left thick by selecting zero slices, or also by generating thick slices for selected element types. Thick slicing was first implemented for quadrupoles, and then generalized to the slicing of bending magnets [16]. A challenge in thick slicing has been the accurate translation of fringe field effects at the edges of thick elements to the sliced sequences. For bending magnets, this is achieved by generating new thin DIPEDGE elements at the extremities of the original thick dipole, with automatic translation of rectangular dipoles to sector magnets, i.e. RBEND to SBEND.

The improvements and extension of MAKETHIN were implemented aiming at

- providing by default the best algorithms and methods,
- allowing for backward compatibility, if required using switches.

The extensions to MAKETHIN were largely driven by user requests and the need to extend the automatic slicing to new element types and attributes. The most recent extensions, implemented during this year are

- implementation of optional thick slicing of solenoids,
- translation of aperture tolerances,
- writing bending angles to multipoles if different from integrated zero-order multipole strength  $k_0 l$  (see next section).

A more general review of all element and attribute types used in MAD-X has recently been launched with the aim to simplify and unify the MAKETHIN code and to automatically transmit any new element attributes to the sliced sequence.

## ELEMENT-SPECIFIC EXTENSIONS

### *x- and y-rotation*

The  $x$ - and  $y$ -rotations are two recently-implemented features in MAD-X. Both are implemented as new elements and can be inserted at any arbitrary position in a sequence. The purpose is to rotate the reference frame and update the particle's coordinates in the new frame orientation. This is a useful way to reference other elements into a different frame. The full map used by the tracking code for a  $x$ -rotation of angle  $\theta$  is described by the equations:

$$\begin{aligned} x^f &= x^i + \frac{yp_x^i \tan \theta}{p_z^i - p_y^i \tan \theta} \\ p_x^f &= p_x^i \\ y^f &= \frac{y^i}{\cos \theta - p_y^i \tan \theta / p_z^i} \\ p_y^f &= p_y^i \cos \theta + p_z^i \sin \theta \\ t^f &= t_i - \frac{y^i(1/\beta_0 + p_t) \tan \theta}{p_z^i - p_y^i \tan \theta}, \end{aligned}$$

where  $x, y$  are the horizontal and vertical positions,  $p_x, p_y$  are the transverse momenta,  $t$  is the time difference in relation to the reference particle,  $\beta_0$  is the relativistic factor,  $p_t = (E - E_0)/P_0c$ , where  $E$  is the energy and  $P$  is the momentum, and  $p_z = \sqrt{(1 + \delta_p)^2 - p_x^2 - p_y^2}$  with  $(1 + \delta_p)^2 = 1 + 2p_t/\beta_0 + p_t^2$ . The index  $i$  refers to the initial state before the rotation and the  $f$  refers to the final state, i.e. after conversion to the new coordinate system.

It is worth noting that even though a rotation followed by a rotation in the opposite direction will return the system to its original state, the same does not hold if the procedure is done with an element that has a length in between the rotations. This can be understood by considering a particle with  $p_x = p_y = 0$  and  $p_z > 0$ . Applying a rotation around the  $y$ -axis, the particle will get a  $p_x \neq 0$ . If we then let this particle drift in the new reference frame, we will end up in a  $x \neq 0$ . If we then apply the opposite rotation the  $p_x = p_y = 0$  again, but there is still a horizontal non-zero position.

The  $x$ - and  $y$ -rotation are also implemented in TWISS for first order, where it is implemented by taking the map derivatives expanded and truncated at first order.

### *SBEND with Angle Different From $k_0l$*

The TRACK and TWISS modules have been extended to support a SBEND element with an angle  $\alpha$  different from the integrated strength  $k_0l$ . The difference between these two quantities is considered similarly to a field error, i.e. added to user-defined field errors, and treated as such in the thick map of the SBEND or transferred properly to slices, i.e. MULTIPOLE elements, by the MAKETHIN module. For the purpose of the horizontal field curvature  $h$ , this difference implies a change of curvature by  $\delta h = (\alpha/l - k_0)/(1 + \delta_p)$ .

### *Matching at Arbitrary Positions in Elements*

The MATCH command has been extended to support matching constraints at any position inside thick elements [17], and is plan for next MAD-X release. The first step was to implement the interpolation mechanism for the TWISS command. For this purpose, the SELECT command has been extended with the new flag INTERPOLATE and the new attributes AT, SLICE, and STEP to specify the (relative)positions within the selected elements by ranges and patterns. Then, the CONSTRAINT command has been extended with the new IINDEX attribute to specify the index (starting at 0) of the (relative) position in an element where the constraint must hold, as shown in the following command excerpt:

```
SELECT, FLAG=INTERPOLATE, RANGE=mq1, AT={0.5, 1};
MATCH, SEQUENCE=seq;
VARY, NAME=k1; # vary strength of quadrupole mq1
CONSTRAINT, RANGE=mq1, IINDEX=0, BETX=5;
LMDIF; # match betx at centre of mq1 varying k1
ENDMATCH;
```

## MAD-X PTC EXTENSIONS

One of the main methods of measuring non-linear beam parameters is by performing harmonic analysis of the betatron motion. The beam oscillations are provoked either by a kicker magnet or by an Alternating Current (AC) dipole. The amplitude of the lines present in the spectra are directly linked to strength of the RDTs, whose definition can be found, for instance, in [18] where the detail of the method is also presented.

Modelling RDT propagation along an accelerator structure is a very important topic in computational accelerator physics for several reasons: to compare theoretical predictions with beam-based measurements; to calculate their change upon the action of dedicated corrector magnet in view of constructing response matrices for optimized correction strategies; to test and provide optimization strategies of RDT to achieve an optimal beam performance. In this respect it is worth mentioning that while for some applications the RDTs need to be minimized for others they have to be controlled and different from zero, as is the case for the CERN PS Multi-Turn Extraction (MTE) [19, 20].

RDTs are calculated by PTC\_TWISS when the option TRACKRDTs is set to true. The PTC\_TWISS algorithm in its core propagates the  $A^{-1}$  map along the lattice, as defined by the normal form transformation  $M = A^{-1} \cdot R \cdot A$ .  $M$  is a non-linear map describing the motion along the accelerator. The  $A^{-1}$  map transforms the Cartesian coordinates  $x, p_x, y, p_y$  to phasors [21], i.e. the normal-form coordinates,  $\zeta_{x,+}, \zeta_{x,-}, \zeta_{y,+}, \zeta_{y,-}$ , where the Hamiltonian depends only on non-linear amplitudes and is independent of phase advances. Finally,  $R$  is an amplitude-dependent rotation.

$A^{-1}$  can be expressed in terms of a generating function  $F$  that is a simple polynomial

$$A^{-1} = e^{-F} \quad (1)$$

$$F = \sum_{n \geq 2} \sum_{jklm} f_{jklm} \zeta_{x,+}^j \zeta_{x,-}^k \zeta_{y,+}^l \zeta_{y,-}^m, \quad (2)$$

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

whose coefficients  $f_{ijklm}$  are the RDTs.

For each lattice element, the calculated coefficients of  $F$  are written to a dedicated table called TWISSRDT. The content of this table can be plotted or matched the same way as any other variable present in the TWISS table. Figure 2 shows an example of the comparison of the RDT  $f_{4000}$  as computed by PTC\_TWISS or from tracking simulations where the excellent agreement is clearly visible. The AC dipole element

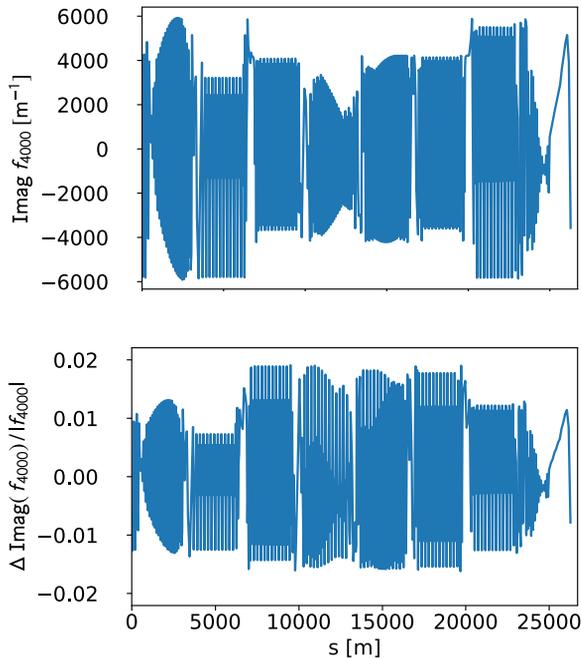


Figure 2: Imaginary part of  $f_{4000}$  from PTC\_TWISS (upper) and relative difference between RDT- and numerical-tracking simulations for the LHC (lower, courtesy of F. Carlier).

has been interfaced to the PTC\_TRACK module so that the dynamics of beams undergoing forced oscillations, as is the case during optics measurements, can now be modelled. It employs a PTC feature called *clock*, which extends the phase space with special variables oscillating proportionally to the beam's time of flight. It can be thought of as an analog hand clock representing rotating vectors (i.e. phasors) with a given frequency. Parameters of the selected magnets can be linked to a clock value at the moment of passage through this magnet, thus yielding oscillatory behaviour on a turn-by-turn basis. Currently, the implementation is limited to two distinct clock frequencies, which is enough to implement the most realistic scenario with one horizontal and one vertical AC dipole. This element also implements ramp up, plateau and ramp down times, as these features are routinely used during optics measurements to excite the beam adiabatically thus avoiding emittance blow up. AC dipoles will be implemented in PTC\_TWISS in the near future.

All the PTC options of synchrotron radiation modelling are now interfaced in the MAD-X commands and the possibility to simulate stochastic effects of the radiation were

added. Note that PTC\_TWISS also outputs damping times and equilibrium emittances.

Several other improvements, bug fixes, and performance optimization have been performed and implemented in the code, and the main ones are:

- Speed up of the sector-bend tracking with the exact Hamiltonian by an automatic detection of the maximum multipole order required in the numerical computations. Indeed, previously it was by default set to 22 to allow for magnetic errors modelling to this level. For the LHC lattice PTC\_TWISS computation became three times faster.
- A new option NORMAL of PTC\_TWISS writes all the results of the normal form analysis to a dedicated table called NONLIN. Currently, the following variables are provided to the order specified in PTC\_TWISS: the three tunes, dispersions, eigenvectors, RDTs (generating function), Hamiltonian, and one-turn map. Note that the same algorithm is available in the PTC\_NORMAL command, however, while the non-linear parameters need not to be selected for PTC\_TWISS, this is not the case for PTC\_NORMAL. It is worth stressing that the format of the output table is such that the parameters can be easily accessed in the subsequent parts of the user script, for example in matching routines.
- REGLOSS option has been implemented in PTC\_TRACK that currently produces a table of the last coordinates for each lost particle.
- 6D closed-orbit search when TOTALPATH is true was fixed and now the program correctly calculates the dependence of the beam momentum on RF frequency.
- Last, but not least, the most recent version of the PTC library from E. Forest was included in MAD-X.

## CONCLUSION

The latest round of improvements to and development of the MAD-X program fully restores some past functionality of MAD8 in terms of synchrotron radiation effects and extends the capabilities including the possibility of dealing with space-charge effects. This, together with element-specific extensions, the improvements of the linear coupling treatment, and the review of PTC commands, make MAD-X/PTC more flexible, robust, and ready to fulfill the needs of studies of planned and future accelerators.

## ACKNOWLEDGEMENTS

We would like to thank A. Franchi from ESRF at Grenoble for the help and support provided for the benchmark of radiation effects using the AT code. We are also grateful to F. Carlier from CERN for the comparison of the RDTs as computed with PTC\_TWISS against tracking data.

## REFERENCES

- [1] E. Forest, F. Schmidt, and E. McIntosh, "Introduction to the Polymorphic Tracking Code", Technical Report CERN-SL-2002-044 / KEK 2002-03, <http://cds.cern.ch/record/573082/files/sl-2002-044.pdf>.
- [2] F. Barbarin, F.C. Iselin, and J.M. Jowett, "Particle dynamics in LEP at very high-energy", in *Proc. EPAC 1994*, London, England, pp. 193-95, [http://cern.ch/accelconf/e94/pdf/epac1994\\_0193.pdf](http://cern.ch/accelconf/e94/pdf/epac1994_0193.pdf).
- [3] J.M. Jowett, "Realistic prediction of dynamic aperture and optics performance for LEP", in *Proc. PAC 1999*, New York, US, pp. 1680-82, TUP86, <http://cern.ch/accelconf/p99/papers/tup86.pdf>.
- [4] M. Bassetti, "Effects Due to the Discontinuous Replacement of Radiated Energy in an Electron Storage Ring", *Experientia* **40**, pp. 968, 1980.
- [5] D. Schulte *et al.*, "The PLACET Tracking Code", <http://cern.ch/clicsw>.
- [6] A. Terebilo, "Accelerator Toolbox for MATLAB", SLAC-PUB-8732, 2001.
- [7] K. Oide, "Synchrotron-Radiation Limit on the Focusing of Electron Beams", *Phys. Rev. Lett.* **61**, 1713, 1988.
- [8] G. Franchetti, S. Gilardoni, A. Huschauer, F. Schmidt and R. Wasef, "Space charge effects on the third order coupled resonance", *Phys. Rev. Acc. and Beam*, Vol. 20, Iss. 8, 2017.
- [9] F. C. Iselin, "The MAD program: Physical Methods Manual", CERN Report SL/92, 1994.
- [10] L. Deniau, "MAD-X Linear Coupling Review", CERN BE/ABP-HSS presentations, part 1 & part 2, 2017.
- [11] E.D. Courant, and H.S. Snyder, "Theory of the Alternating-Gradient Synchrotron", *Annals of Physics* **3**, pp. 360-408, 1958.
- [12] J. Eisenfeld, "Block diagonalization and eigenvalues", *Linear Algebra and its Application* **15**, pp. 205-215, 1976.
- [13] D.A. Edwards, and L.C. Teng, "Parametrization of linear coupled motion in periodic systems", *IEEE Trans. on Nuclear Science* **20**, pp. 885-888, 1973.
- [14] D. Sagan, and D. Rubin, "Linear analysis of coupled lattices", *Phys. Rev. ST Accel. Beams* **2**, 074001, 1999.
- [15] H. Burkhardt *et al.*, "Improved TEAPOT Method and Tracking with Thick Quadrupoles for the LHC and its Upgrade", in *Proc. IPAC 2013*, Shanghai, China, pp. 945-947, MOPWO027, <http://cern.ch/accelconf/IPAC2013/papers/mopwo027.pdf>.
- [16] H. Burkhardt, L. Deniau, and A. Latina, "Upgrade of Slicing and Tracking in MAD-X", in *Proc. IPAC 2014*, Dresden, Germany. doi:10.18429/JACoW-IPAC2014-TUPRO063
- [17] T. Gläbtle, "Report on collaboration on MAD-X extensions and GUI for tuning of transfer lines", Technical Report, 2017.
- [18] A. Franchi *et al.*, "First simultaneous measurement of sextupolar and octupolar resonance driving terms in a circular accelerator from turn-by-turn beam position monitor data", *Phys. Rev. ST Accel. Beams* **17**, 074001, 2014.
- [19] R. Capii, and M. Giovannozzi, "Novel Method for Multi-Turn Extraction: Trapping Charged Particles in Islands of Phase Space", *Phys. Rev. Lett.* **88**, 104801, 2002.
- [20] A. Huschauer *et al.*, "Transverse Beam Splitting Made Operational: Key Features of the Multi-Turn Extraction at the CERN Proton Synchrotron", *Phys. Rev. Accel. Beams* **20**, 061001, 2017.
- [21] E. Forest, "From tracking code to analysis: generalised Courant-Snyder theory for any accelerator model", *Springer*, 2016.

# SixTrack PROJECT: STATUS, RUNTIME ENVIRONMENT, AND NEW DEVELOPMENTS\*

R. De Maria<sup>†</sup>, J. Andersson, V. K. Berglyd Olsen, L. Field, M. Giovannozzi, P. D. Hermes,  
N. Høimyr, S. Kostoglou, G. Iadarola, E. McIntosh, A. Mereghetti, J. Molson, D. Pellegrini,  
T. Persson, M. Schwinzerl, CERN, Geneva, Switzerland  
E. H. Maclean, CERN, Geneva, Switzerland and University of Malta, Msida, Malta  
K. N. Sjobak, CERN, Geneva, Switzerland and University of Oslo, Oslo, Norway  
I. Zacharov, EPFL, Lausanne, Switzerland  
S. Singh, IIT Madras, India<sup>‡</sup>

## Abstract

SixTrack is a single-particle tracking code for high-energy circular accelerators routinely used at CERN for the Large Hadron Collider (LHC), its luminosity upgrade (HL-LHC), the Future Circular Collider (FCC), and the Super Proton Synchrotron (SPS) simulations. The code is based on a 6D symplectic tracking engine, which is optimised for long-term tracking simulations and delivers fully reproducible results on several platforms. It also includes multiple scattering engines for beam-matter interaction studies, as well as facilities to run integrated simulations with FLUKA and GEANT4. These features differentiate SixTrack from general-purpose, optics-design software like MAD-X. The code recently underwent a major restructuring to merge advanced features into a single branch, such as multiple ion species, interface with external codes, and high-performance input/output (XRootD, HDF5). This restructuring also removed a large number of build flags, instead enabling/disabling the functionality at run-time. In the process, the code was moved from Fortran 77 to Fortran 2018 standard, also allowing and achieving a better modularization. Physics models (beam-beam effects, RF-multipoles, current carrying wires, solenoid, and electron lenses) and methods (symplecticity check) have also been reviewed and refined to offer more accurate results. The SixDesk runtime environment allows the user to manage the large batches of simulations required for accurate predictions of the dynamic aperture. SixDesk supports CERN LSF and HTCCondor batch systems, as well as the BOINC infrastructure in the framework of the LHC@Home volunteering computing project. SixTrackLib is a new library aimed at providing a portable and flexible tracking engine for single- and multi-particle problems using the models and formalism of SixTrack. The tracking routines are implemented in a parametrized C code that is specialised to run vectorized in CPUs and GPUs, by using SIMD intrinsics, OpenCL 1.2, and CUDA technologies. This contribution presents the status of the code and an outlook on future developments of SixTrack, SixDesk, and SixTrackLib.

\* Research supported by the HL-LHC project

<sup>†</sup> riccardo.de.maria@cern.ch

<sup>‡</sup> Work supported by Google Summer of Code 2018

## INTRODUCTION

SixTrack [1, 2] is a 6D single-particle symplectic tracking code able to compute the trajectories of individual relativistic charged particles in circular accelerators for studying dynamic aperture (DA) or evaluating the performance of beam-intercepting devices like collimators [3]. It can compute linear and non-linear optics functions, time-dependent effects, and extract indicators of chaos from tracking data. SixTrack implements scattering routines and aperture calculations to compute “loss maps”, i.e., leakage from collimators as a function of longitudinal position along the ring, and collimation efficiency [4].

Different from a general-purpose code like MAD-X [5, 6], SixTrack is optimised for speed and numerical reproducibility. It can be also linked with the BOINC library to use the volunteering computing project LHC@Home [7]. SixTrack studies, such as estimation of dynamic aperture of large storage rings like the Large Hadron Collider (LHC) or the Future Circular Collider (FCC), require massive computing resources, since they consist of scans over large parameter spaces for probing non-linear beam dynamics over long periods.

The SixDesk runtime environment manages SixTrack simulations from input generation, job queue management (using HTCCondor or LSF in the CERN BATCH service and customised software in CERN Boinc server), to collecting and post-processing results.

SixTrackLib is a new library built from scratch in C with the main aim of offering a portable tracking engine for other codes and offloading SixTrack simulation to GPUs.

This paper summarises the main existing features of SixTrack, SixDesk and SixTrackLib and provide detail about the main development lines.

## MAIN FEATURES

SixTrack tracks an ensemble of particles defined by a set of coordinates through several beam-line elements, using symplectic maps [8–10], or scattering elements.

### *Coordinates*

The set of coordinates is larger than the minimum needed to describe the motion. Additional variables are used to store energy-related quantities used in the tracking maps

that are updated only on energy changes, which does not occur very frequently in synchrotrons in absence of radiation effects, to save computational time. Thick maps for dipole and quadrupoles also reuse the energy-dependent factors of the first- and second-order polynomial of the map that are recalculated at each energy change. Furthermore, different ion species, such as debris from interaction with matter, can be tracked at the same time using an extension of the usual symplectic formalism [11].

Variables used internally in tracking are not canonical, however, once they are converted to canonical form, the maps are symplectic. Different from other codes, SixTrack uses

$$\sigma = s - \beta_0 ct$$

as the longitudinal coordinate during tracking to avoid rounding errors associated to the relativistic  $\beta$  when updating time delays in drifts and

$$\left( \zeta = \frac{\beta}{\beta_0} \sigma, \quad \delta = \frac{P - P_0}{P_0} \right)$$

as conjugate canonical variables in 6D optics calculations which use explicitly symplectic maps.

### Beam-line Elements

Table 1 shows the different types of beam-line elements implemented in SixTrack. Thin multipoles are used in conjunction with the MAKETHIN and SIXTRACK commands in MAD-X to implement symplectic integrators of thick maps. Thin multipoles include the effect of the curvature, when present, up to the second order. The tracking maps have been recently reviewed and benchmarked against MAD-X and its optics module for consistency.

Table 1: Physical Elements Implemented in SixTrack

Drift expanded	Drift exact [12]
Single thin multipole	Thin multiple block
Thick dipole-quadrupole	Thin solenoid
Accelerating cavities	RF-multipoles [13]
4D Beam-beam	6D beam-beam [14]
Wire [15]	Hollow electron lens [16, 17]

### Scattering

SixTrack embeds the K2 scattering engine [18, 19], capable of simulating the basic scattering processes undergone by an ultra-relativistic proton in the multi-TeV range when passing through matter. The simulated processes range from ionisation energy loss and multiple Coulomb scattering to point-wise interactions like Coulomb, elastic, and inelastic events, including single diffractive scattering. Compound materials of interest for the low-impedance upgrade of the LHC collimators are implemented via averaged nuclear and atomic properties [20]. Other scattering models can be imported and made available in the SixTrack executable, such as that of Merlin [21] and Geant4 [22, 23].

A new scatter block is under development to offer a general framework for simulating scattering events in SixTrack. Currently, it supports beam scattering against a target specified as an area density distribution at a thin marker inserted into the lattice. Internally, the scattering module supports elastic scattering through Monte Carlo sampling of experimental data from Totem. Alternatively, scattering events can be generated on the fly by Pythia8 [24], in which case elastic and diffractive processes are supported.

### Optics Calculations

SixTrack contains matrix code for 5D optics calculation and a 6D tracking engine using Truncated Power Series Algebra library (TPSA [25]) for 6D optics calculation. The 6D tracking engine uses canonical variables and it provides a cross-check of the symplecticity of the one-turn-map. Coupled Twiss parameters (using the Mais-Ripken formalism [26]) can be extracted along the lattice. The optics parameters are optionally used in the beam-beam elements for self-consistent simulations. The 6D optics module has been recently improved by removing some unnecessary ultra-relativistic approximations which introduced small symplectic errors.

### Dynamic Effects

A general functionality for dynamically-changing simulation settings on a turn-by-turn basis has been implemented [27, 28]. This allows setting magnet strengths including multipoles, RF amplitude and phase, reference energy, and beam-beam element as a function of turn number. This can be useful for a number of different studies, e.g. magnet snap-back in the LHC [29], HL-LHC crab cavity failure scenarios [30–32], studies of beam losses during energy ramp [33], and hollow electron-lens modulation [34]. The settings can be internally computed as a function of turn number, or loaded from a file. These functions are specified using a flexible language that allows combining functions to achieve the required effect. The architecture of the functionality makes it easy to add support for new elements or new functions.

### Post Processing

Long-term tracking with SixTrack is used extensively at CERN for studying the DA, with a typical study consisting of up to  $\sim 2 \times 10^6$  individual tracking simulations over  $10^5 - 10^6$  turns (see Figure 1 for an example).

Tracking data are post processed during the study and summary files, containing the main results of the simulation for each initial condition, are returned back to the user. In particular, tracking summary files for each initial condition identify particle loss/survival, final surviving turn numbers and the inferred particle amplitudes.

Particle's invariants are calculated for each initial condition and based on the average invariant over a user-defined range of turns. An initial estimate of the invariant is obtained by assuming no coupling between the planes of motion, via the usual relation for the Courant-Snyder ellipse, e.g. for the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

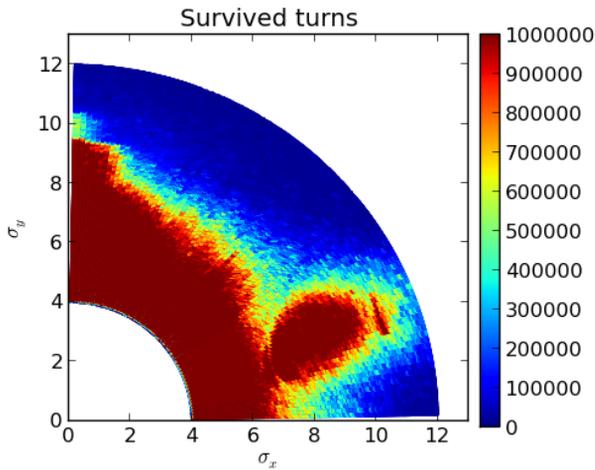


Figure 1: Survival plot of a fine phase-space scan for the LHC. The simulations was divided in task of 60 particles pairs covering the phase space in  $2\sigma$  and  $1.5^\circ$  steps.

horizontal plane. Alternatively, an estimate of the decoupled single-particle emittance for the three oscillation modes can be calculated from the eigenvectors of the motion ( $\bar{v}$ ), which may be constructed from the one-turn map, see for example [26]. Various parameters relevant to the nonlinear motion, such as smear and detuning, are also evaluated.

In addition to quantities relevant to particle survival, estimates of the long-term stability are obtained through a Lyapunov-like analysis performed by examining the phase-space separation of initially close by particle pairs. In particular, the angular separation in phase for the three oscillation modes

$$\frac{1}{\pi} \sqrt{\frac{\Delta\phi_1^2 + \Delta\phi_2^2 + \Delta\phi_3^2}{N}},$$

where  $N = (1, 2, 3)$  for  $(2D, 4D, 6D)$  motion, respectively, is considered. Linear fits to the logarithm of the separation as function of the logarithm of the turn number identify the maximum separation rate between each particle pair. This quantity is returned, along with the maximum separation in phase, to provide approximate indicators of the onset of chaotic motion in place of the far more computationally intensive Lyapunov exponents [35, 36].

Summary files for the outcome of each initial condition are collectively post-processed by the user using external tools, in order to identify minimum boundaries in the  $(\sigma_x, \sigma_y)$  space for particle survival over the tracked number of turns, as well as to study the evolution of DA as a function of the turn number.

### Frequency Analysis

A collection of routines for frequency analysis has been linked in SixTrack, namely PLATO [37] and a C++ implementation of the Numerical Analysis of Fundamental Frequency (NAFF) algorithm has been developed [38]. These algorithms allow for a more refined, compared to plain FFT,

tune determination with a much faster convergence, i.e. requiring a shorter number of turns. By comparing the tune determination at different time intervals, diffusive frequency maps can be computed [39]. With the resolution of the frequency map, resonance lines become visible, even in the case of a tune modulation from a quadrupolar ripple (the triplets in IR1 and 5) with frequency of 550 Hz and relative amplitude of  $10^{-7}$ , as shown in Figure 2.

### Input and Output

Initial conditions can be given in amplitude steps or taken from an external file. A dump module offers multiple ways to extract tracking data both in terms of type observable (physical coordinates, canonical coordinates, normalised coordinates, averages and first order distribution momenta) in a selection of turns and observation points. Data are written in ASCII and, in a few cases, a binary option is also available. Support for output of simulation data to a HDF5 [40] files and ROOT [41] is also currently being developed.

Furthermore, it is planned to develop a new way to generate the distribution that is used as initial conditions for tracking. This will provide the functionality to create matched or mismatched distributions in both physical and normalised coordinates.

### Interfaces to External Programs

Collimation studies can also be performed running SixTrack coupled [42] to Fluka [43, 44]. In this configuration, the two codes exchange particles at run time, with the aim of combining the refined tracking through the accelerator lattice, performed by SixTrack, with the detailed scattering models, implemented in Fluka, when the beam reaches intercepting devices. The use in the Fluka-SixTrack coupling of the same Fluka geometries used for subsequent energy-deposition calculations run with Fluka allows an excellent level of consistency of results.

Additionally, a more generalised interface “BDEX” for interfacing external codes is also included, enabling for example tracking of multiple bunches or coupling to cavity simulation codes. Here, the exact protocol can be implemented as a plug-in to SixTrack [45].

### Building and Testing

A CMake-based build and test system has recently been added [45]. This greatly simplified the maintenance of the dependencies between the various build options, as well as the setup for building on the large range of supported platforms.

The testing framework CTest is also provided as part of CMake. For SixTrack, this is used to verify that the executables are still providing the expected output after the code has been modified, and to track the changes to the output. Furthermore, it is used for checking that the results from versions compiled for different platforms are in agreement, which is vital for BOINC. The main benefit of using CTest is that test running is fully automatic and gives a simple

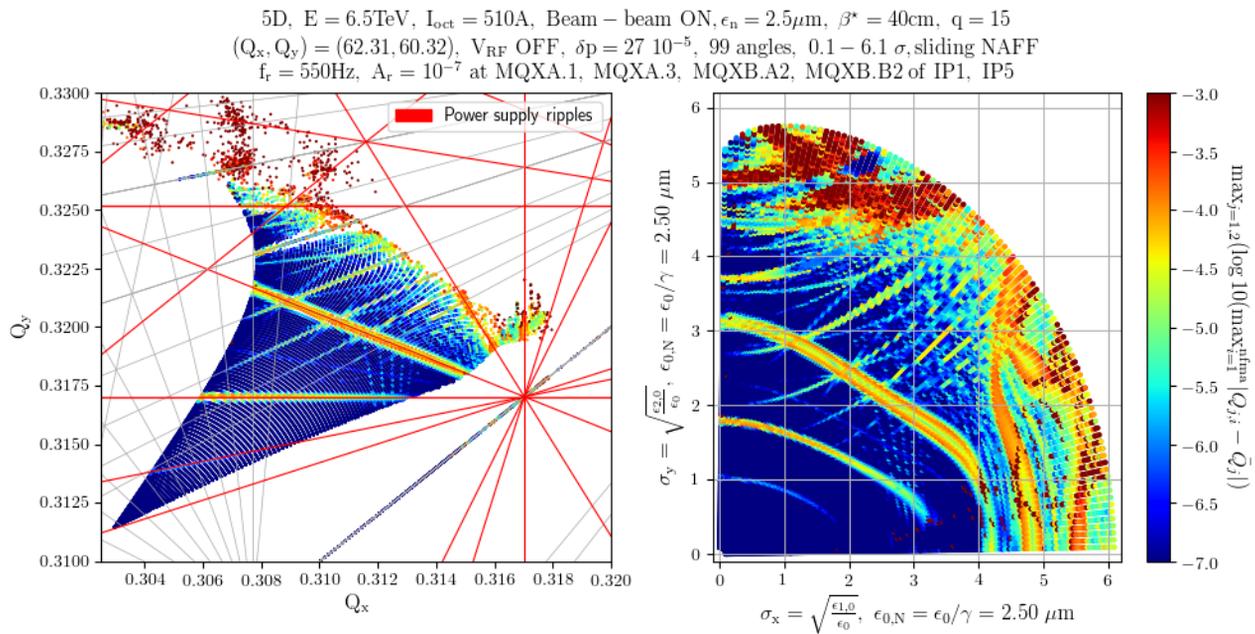


Figure 2: Frequency map using the NAFF method for LHC in the presence of a tune modulation.

pass/fail output, which is also integrated with GitHub for pull requests.

### Performance

Thanks to the recent re-factoring, the internal particles arrays are fully dynamic, therefore the number of particles that can be tracked in parallel is limited by the system memory and not by a build-time flag. A machine model like the LHC, using about 18k elements and 4.6k high-order multipole blocks, needs about 220  $\mu\text{s}$  per particle per turn on a single CPU core at 3.4 GHz. Typical studies requires of the order of  $10^9 - 10^{12}$  particle turns and even more for parameter scans. For this reason, SixTrack is often used in conjunction with high-performance computing facilities described in the following section.

## RUNTIME ENVIRONMENTS

The SixDesk environment [46] is the simulation framework used to manage and control the large amount of information necessary for, and produced by, SixTrack studies of dynamic aperture. It supports CERN batch systems [47] as well as the BOINC platform for volunteering computing available at the LHC@Home project [7]. The SixDeskDB post-processing tool collects data from SixDesk, performs post-processing analysis, and prepares reports and plots. It also offers a Python API for interactive analysis. Similarly to the SixTrack code, the SixDesk environment and SixDeskDB are continuously updated, extending the coverage of the studies and keeping the environment up to date with the latest developments in the CERN IT infrastructure.

### LHC@Home and the CERN Batch System

Volunteer computing has been used successfully at CERN since 2004 with the LHC@Home project; it has provided additional computing power for CPU-intensive applications with small data sets, as well as an outreach channel for CERN activities. LHC@Home started off with SixTrack, which had been successively ported from mainframe to supercomputer to emulator farms and PCs. In order to run on the largest number of volunteer computers, SixTrack is compiled for the most common operating systems, architectures, and CPU instruction sets.

In terms of computing power provided by the volunteers to LHC@home, the system is capable of handling  $1 \times 10^5$  tasks on average, with peaks of  $3.5 \times 10^5$  tasks simultaneously running on  $2.4 \times 10^4$  hosts observed during SixTrack intense simulation campaigns (see Figure 3). Every SixTrack task is run twice to eliminate random host errors and minimise the impact of a failing host. The LHC@Home capacity available for SixTrack can be compared to the average of  $2.5 \times 10^5$  running tasks on  $1.4 \times 10^5$  processor cores in the CERN computer centre, which is fully loaded with tasks from analysis and reconstruction of collisions recorded by LHC experiments, and has limited spare capacity for beam dynamics simulations.

The CERN batch system is presently managed by means of the HTCondor [48] package. Contrary to BOINC, most suitable for a steady stream of work units, the CERN batch system provides users with a responsive computing resource. Also, contrary to LHC@Home, no redundancy is implemented during task submission since the code run in a controlled environment, although very rarely hardware errors do appear in the results.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

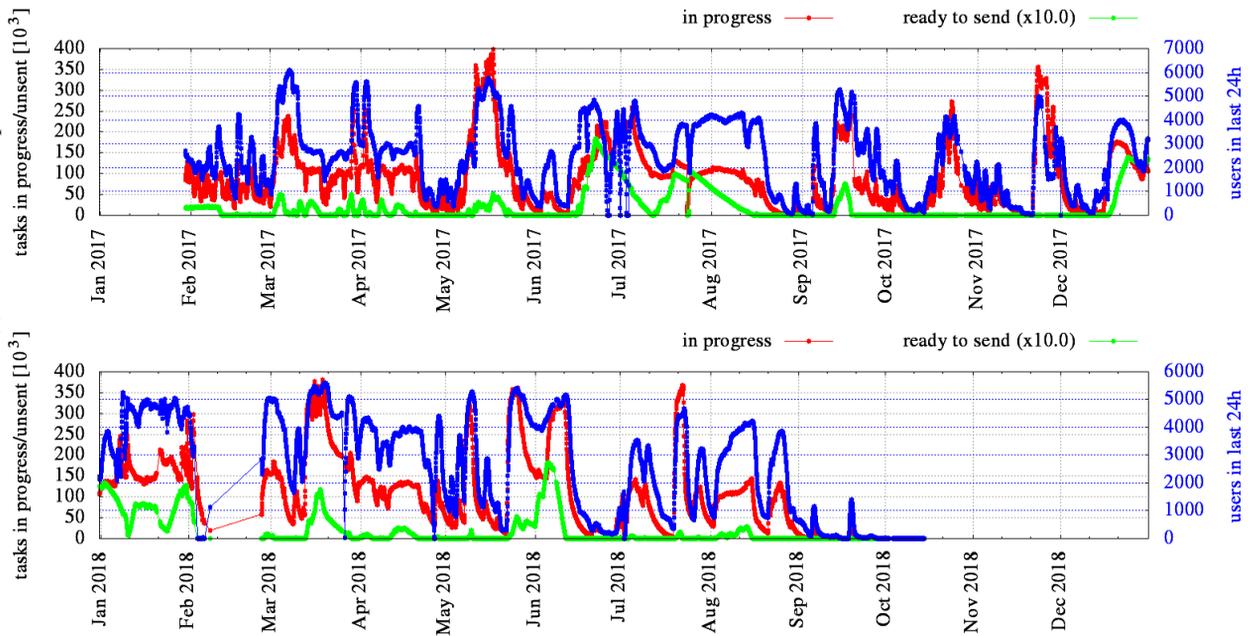


Figure 3: Summary of tasks and users during the last two years (upper frame: 2017; lower frame: 2018 to date) of the SixTrack application in the LHC@Home project. The number of users tends to increase, with the number of tasks absorbing most of the time all pending tasks.

### Developments

One of the main development lines of SixDesk is porting the collimation studies to the BOINC platform for volunteer computing. This entailed a thorough revision of the collimation part, currently on-going, to make results numerically stable and reproducible across platforms. The possibility to interrupt and restart the computation (check-point/restart capability), which is essential to run on BOINC, is being added as well.

Other lines of development include: the addition of new parameters for dedicated scans of dynamic aperture; the possibility of running chains of jobs in BOINC, for simulating extended periods of beam time in the ring; a pre-filtering stage of submission to the CERN batch system prior to submission to BOINC, to avoid short tasks, with consequent inefficient use of volunteer resources, like bandwidth and time.

### SixTrackLib

In the context of single-particle simulations, tracking requires no interaction between the calculations carried out for particle  $p_i$  and  $p_{j \neq i}$  from a set of  $N$  particles. The memory requirements for representing each  $p_i$  typically ranges from  $10^1$  to  $10^3$  Bytes. The machine description can, over a single turn, be considered constant, although different elements and sections of the ring require a different amount of local resources. Still, SixTrack presents itself as an ideal candidate for a parallel implementation: strongly CPU bound with inherent parallelism and resource requirements not inherently scaling with the number of parallel processes.

Introducing parallelism into a mature code-base like SixTrack, even from such a favourable starting point, is challeng-

ing. It entails a high levels of complexity due to competing paradigms and concepts of parallel computing. In particular, a fast-changing technological landscape in combination with a diverse, multi-vendor and long-tailed selection of hardware available via initiatives like LHC@Home [7], as well as the realities of limited development resources are the main decision-making factors. These and other constraints motivated the design and ongoing development of a new, stand-alone library providing the core functionality of SixTrack.

SixTrackLib [49] is an open-source library developed from scratch using C and C++, allowing users to off-load the particle tracking onto supported HPC resources. As of this writing, it provides a) a representation for a set of particles; b) a set of beam-elements (drifts, multipoles, cavities, 4D and 6D beam-beam elements, etc.); c) a set of maps describing the tracking of the particles over the beam-elements; d) a dedicated generic buffer for managing and transferring data to the computing nodes; e) implementations and abstractions for different computing environments (auto-vectorized CPU code, OpenCL [50], CUDA [51]); f) high-level APIs in C, C++ and (under development) Python.

The chosen design allows for the complete separation of business logic from the modelled physics, allowing the latter to be shared across all architectures. Maps and tracking functions are implemented in a sub-set of the C99 language, in terms of the provided abstractions. The physics parts are exposed to the user in a modular and header-only fashion, allowing to tap into SixTrackLib under a wide-range of use cases currently out of reach for a stand-alone application such as SixTrack.

First simulations and tests prove the feasibility of the approach and reproduce results delivered by SixTrack. Performance analysis confirms that the main limiting factors for performance and scalability are: a) the finite availability of resources such as registers and high-bandwidth/low-latency memory on computing nodes; b) the ability to compensate for any occurring latencies by having enough parallel tasks scheduled to prevent computing units from stalling or idling.

Consider for example the simulation of  $1 \leq N \leq 10^7$  particles on a lattice with 18657 beam-elements, representing the LHC without beam-beam interactions. Evaluating SixTrackLib on a CPU-based OpenCL implementation (Intel Xeon E5-2630 20x2.2 GHz hyper-threads) and a range of high-end GPUs (NVIDIA Tesla V100 PCIe 16GB GPU) as well as consumer-grade GPUs (NVIDIA GTX 1050Ti 4GB, AMD RX560 4GB) demonstrates parallel speed-ups approaching (for large  $N$ ) factors of  $10^1$  to and exceeding  $10^2$  (Fig. 4).

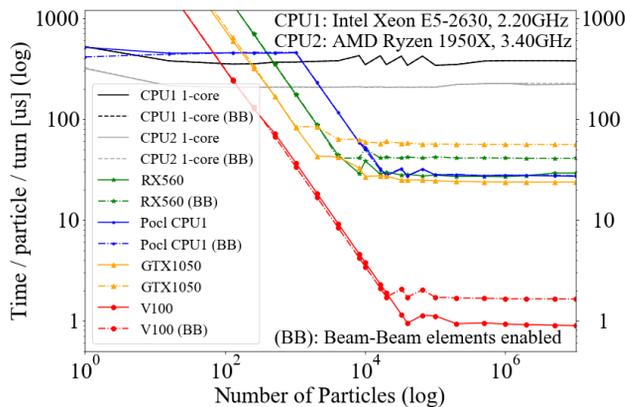


Figure 4: Benchmarking results of a LHC study using SixTrackLib on different target hardware. Increasing the complexity of the parallel code by enabling beam-beam elements (BB), but not using them, leads to decreased performance on GPUs, but not on CPU-based systems.

While enabling code-paths for handling beam-beam interactions in the parallel tracking code leads, unsurprisingly, to virtually no change on CPU-based systems, the increased complexity and pressure on resources impairs the performance on all studied GPU systems (cf. results in Fig. 4 labelled with (BB)). On lattices actually featuring complex features like beam-beam elements, this observation motivates studying ways to split the monolithic parallel code into smaller specialised blocks and to execute these blocks in sequence, thereby trading in synchronisation and dispatching overheads for a potentially better utilisation of hardware resources.

## CONCLUSIONS

The SixTrack tracking code is the main code used to simulate long-term stability, collimation cleaning, and machine failure scenarios in the LHC, SPS and FCC due its unique features of speed and integration with HPC resources. It

comes with a fully developed running environment to perform easily the massive numerical simulations that include scans on beam and ring parameters and the option of using different computing resources, from standard batch services to volunteer computing.

In spite of its maturity, SixTrack is still in an intense development phase. On the short time scale, it is planned to merge into a single code the features that were developed in the framework of the studies of the LHC collimation system. On a longer time scale, the main lines of development include tighter integration of existing features, interoperability with other codes, and deployment on new architectures such as GPU.

## ACKNOWLEDGEMENTS

We wish to thank all volunteers that supported and continue to support the SixTrack project and the related studies by donating to the LHC@Home Project their CPU power and technical advice. We also wish to thank F. Schmidt for his continuous interest in SixTrack development.

## REFERENCES

- [1] F. Schmidt *et al.*, “SixTrack Version 4.2.16 Single Particle Tracking Code Treating Transverse Motion with Synchrotron Oscillations in a Symplectic Manner”, CERN/SL/94-56, 2012.
- [2] SixTrack Project website, <http://cern.ch/sixtrack>
- [3] R. Bruce *et al.*, “Status of Sixtrack with collimation”, in *Proc. Tracking for Collimation Workshop*, CERN, Geneva, Switzerland, 2018.
- [4] G. Demolaize *et al.*, “A New Version of SixTrack with Collimation and Aperture interfaces”, in *Proc. PAC’05*, Knoxville, TN, USA, May 2005, paper FPAT081.
- [5] MAD-X Project website, <http://cern.ch/mad>
- [6] L. Deniau *et al.*, “Upgrade of MAD-X for HL-LHC project and FCC studies”, presented at ICAP’18, Key West, FL, USA, Oct 2018, paper TUPAF01, this conference.
- [7] J. Barranco *et al.*, “LHC@Home: a BOINC-based volunteer computing infrastructure for physics studies at CERN”, *Open Engineering*, vol. 7, pp. 378, 2017.
- [8] G. Ripken and F. Schmidt, “A symplectic six-dimensional thin-lens formalism for tracking”, DESY 95-63, CERN/SL/95-12(AP), 1995.
- [9] K. Heinemann, G. Ripken, and F. Schmidt, “Construction of nonlinear symplectic six-dimensional thin-lens maps by exponentiation”, DESY 95-189, 1995.
- [10] D.P. Barber *et al.*, “A non-linear canonical formalism for the coupled synchro-betatron motion of protons with arbitrary energy”, DESY 87-36, 1987.
- [11] P. Hermes *et al.*, “Symplectic Tracking of Multi-Isotopic Heavy-Ion Beams in SixTrack”, in *Proc. IPAC’16*, Busan, Korea, May 2016. doi: 10.18429/JACoW-IPAC2016-TUPMW015
- [12] M. Fjellstrom, “Particle Tracking in Circular Accelerators Using the Exact Hamiltonian in SixTrack”, CERN-THESIS-2013-248, 2013.

13th Int. Computational Accelerator Physics Conf.  
ISBN: 978-3-95450-200-4

[13] A. Latina and R. De Maria, “RF multipole implementation”, CERN-ATS-Note-2012-088 TECH, 2012.

[14] L.H.A. Leunissen *et al.*, “6D beam-beam kick including coupled motion”, *Phys. Rev. ST Accel. Beams*, vol. 3, pp. 124002, 2000.

[15] A. Patapenka, “First order transport map of wire element”, in *3rd General SixTrack Meeting*, 7 December 2016, <https://indico.cern.ch/event/590257/>.

[16] V. Previtalli *et al.*, “Numerical Simulations of a Hollow Electron Lens as a Scraping Device for the LHC”, in *Proc. IPAC’13*, Shanghai, China, May 2013, paper MOPWO044.

[17] M. Fitterer *et al.*, “Implementation of Hollow Electron Lenses in SixTrack and first simulation results for the LHC”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi: 10.18429/JACoW-IPAC2017-THPAB041

[18] T. Trenkler and J. B. Jeanneret, “K2, a software package evaluating collimation systems in circular colliders (manual)”, CERN SL/94105 (AP), 1994.

[19] C. Tambasco, “An improved scattering routine for collimation tracking studies at LHC”, CERN-THESIS-2014-014, 2014.

[20] E. Quaranta, “Investigation of collimator materials for the High Luminosity Large Hadron Collider”, CERN-THESIS-2017-101, 2017.

[21] Merlin source code repository, <https://github.com/Merlin-Collaboration/Merlin>

[22] Geant website, <https://geant4.web.cern.ch/>.

[23] K. Sjobak *et al.*, “New features of the 2017 SixTrack release”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi: 10.18429/JACoW-IPAC2017-THPAB047

[24] Pythia8 website, <http://home.thep.lu.se/Pythia/>.

[25] M. Berz, “The new method of TPSA Algebra for the description of beam dynamics to high orders”, LANL Tech. Rep. AT-6:ATN-86-16, 1986.

[26] H. Mais and G. Ripken, “Theory of coupled synchro-betatron oscillations”, DESY-M-82-05, DESY, 1982.

[27] K. Sjobak *et al.*, “General functionality for turn-dependent element properties in SixTrack”, in *Proc. IPAC’15*, Richmond, USA, May 2015. doi: 10.18429/JACoW-IPAC2015-MOPJE069

[28] K. Sjobak *et al.*, “Dynamic simulations in Sixtrack”, in *Proc. Tracking for Collimation Workshop*, CERN, Geneva, Switzerland, 2018.

[29] L. Bottura, L. Walckiers, and R. Wolf, “Field Errors Decay and “Snap-Back” in LHC Model Dipoles”, CERN-LHC-Project-Report-55, 1996.

[30] A. Santamaría García *et al.*, “Limits on failure scenarios for crab cavities in the HL-LHC”, in *Proc. IPAC’15*, Richmond, USA, May 2015. doi: 10.18429/JACoW-IPAC2015-THPF095

[31] K. Sjobak, R. Bruce, H. Burkhardt, A. MacPherson, A. Santamaría García, and R. Kwee-Hinzmann, “Time Scale of Crab Cavity Failures Relevant for High Luminosity LHC”, in *Proc. IPAC’16*, Busan, Korea, May 2016. doi: 10.18429/JACoW-IPAC2016-THPOY043

ICAP2018, Key West, FL, USA  
JACoW Publishing  
doi:10.18429/JACoW-ICAP2018-TUPAF02

- [32] A. Santamaría García *et al.*, “Machine protection from fast crab cavity failures in the High Luminosity LHC”, in *Proc. IPAC’16*, Busan, Korea, May 2016. doi: 10.18429/JACoW-IPAC2016-TUPMW025
- [33] S.J. Wretborn *et al.*, “Study of off-momentum losses at the start of the ramp in the Large Hadron Collider”, CERN-ACC-NOTE-2017-0065, 2017.
- [34] M. Fitterer *et al.*, “Implementation of hollow electron lenses in SixTrack and first simulation results for the HL-LHC”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi: 10.18429/JACoW-IPAC2017-THPAB041
- [35] F. Schmidt, F. Willeke, and F. Zimmermann, “Comparison of methods to determine long-term stability in proton storage rings”, CERN-SL-91-14-AP, 1991.
- [36] M. Böoge and F. Schmidt, “Estimates for Long-Term Stability for the LHC”, LHC Project Report 114, 1997.
- [37] M. Giovannozzi *et al.*, “PLATO: a program library for the analysis of nonlinear betatronic motion”, *Nucl. Instrum. and Methods A*, vol. 388, pp. 1, 1996.
- [38] S. Kostoglou *et al.*, “Development of Computational Tools for Noise Studies in the LHC”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi: 10.18429/JACoW-IPAC2017-THPAB044
- [39] J. Laskar, C. Froeschle, and C. Celletti, “The measure of chaos by the numerical analysis of the fundamental frequencies. Application to the standard mapping”, *Physica D*, vol. 56, pp. 253, 1992.
- [40] HDF5 website, <https://www.hdfgroup.org/HDF5/>.
- [41] ROOT website, <https://root.cern.ch/>.
- [42] V. Vlachoudis *et al.*, “Status of Fluka coupling to SixTrack”, in *Proc. Tracking for collimation Workshop*, CERN, Geneva, Switzerland, 2018.
- [43] T.T. Bohlen *et al.*, “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications”, *Nuclear Data Sheets*, vol. 120, 2014.
- [44] A. Ferrari, P.R. Sala, A. Fassò, and J. Ranft, “FLUKA: a multi-particle transport code”, CERN-2005-10, INFN/TC\_05/11, SLAC-R-773, 2005.
- [45] K. Sjobak *et al.*, “New Features of the 2017 SixTrack release”, in *Proc. IPAC’17*, Copenhagen, Denmark, May 2017. doi: 10.18429/JACoW-IPAC2017-THPAB047
- [46] E. McIntosh and R. De Maria, “The SixDesk Run Environment for SixTrack”, CERN-ATS-Note-2012-089 TECH, 2012.
- [47] CERN Batch service, <http://information-technology.web.cern.ch/services/batch>
- [48] High Throughput Condor (HTCondor), <https://research.cs.wisc.edu/htcondor>
- [49] SixTrackLib source code repository, <http://github.com/SixTrack/SixTrackLib>
- [50] J. Stone *et al.*, “OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems”, *Computing in Science & Engineering*, vol. 12, pp. 3, 2010.
- [51] J. Nickolls *et al.*, “Scalable Parallel Programming with CUDA”, *ACM Queue*, vol. 6, pp. 2, 2008.

# SIMULATIONS OF LONGITUDINAL BEAM STABILISATION IN THE CERN SPS WITH BLOND

J. Repond\*<sup>1</sup>, K. Iliakis, M. Schwarz, E. Shaposhnikova  
CERN, Geneva, Switzerland

<sup>1</sup>also at École Polytechnique Fédérale de Lausanne, Switzerland

## Abstract

The Super Proton Synchrotron (SPS) at CERN, the Large Hadron Collider (LHC) injector, will be pushed to its limits for the production of the High Luminosity LHC proton beam while beam quality and stability in the longitudinal plane are influenced by many effects. Particle simulation codes are an essential tool to study the beam instabilities. BLOND, developed at CERN, is a 2D particle-tracking simulation code, modelling the longitudinal phase space motion of single and multi-bunch beams in multi-harmonic RF systems. Computation of collective effects due to the machine impedance and space charge is done on a multi-turn basis. Various beam and cavity control loops of the RF system are implemented (phase, frequency and synchro-loops, and one-turn delay feedback) as well as RF phase noise injection used for controlled emittance blow-up. The longitudinal beam stability simulations during long SPS acceleration cycle ( $\sim 20$  s) include a variety of effects (beam loading, particle losses, controlled blow-up, double RF system operation, low-level RF control, injected bunch distribution, etc.). Simulations for the large number of bunches in the nominal LHC batch (288) use the longitudinal SPS impedance model containing broad and narrow-band resonances between 50 MHz and 4 GHz. This paper presents a study of beam stabilisation in the double harmonic RF system of the SPS system with results substantiated, where possible, by beam measurements.

## INTRODUCTION

The High-Luminosity Large Hadron Collider (HL-LHC) project [1] is the next milestone at CERN for the LHC and its experiments. The linac and the three synchrotrons in the injector chain will be upgraded to enable the production of HL-LHC proton beam with a bunch intensity  $N_b$  twice that of the current setup, as specified by the LHC Injector Upgrade (LIU) project [2].

The LIU target for the SPS, the LHC injector, is to produce four batches of 72 bunches spaced by 25 ns with an intensity of  $2.4 \times 10^{11}$  particles per bunch (ppb), each batch separated by 225–250 ns. Large particle losses, increasing with intensity, are observed at the SPS flat bottom [3] and multi-bunch longitudinal instabilities limit the ability to increase the bunch intensity [4]. The maximum bunch length allowed for the extraction to the LHC injection is fixed at 1.9 ns with an average value along the batches of 1.65 ns.

To reach the LIU target, major upgrades are necessary. The SPS RF system will have more cavities, more power

available and a better control of the beam loading through the low-level RF control loops (LLRF). Moreover, the longitudinal beam-coupling impedance of the machine will be reduced, but the baseline improvements may be insufficient to ensure beam stability at HL-LHC intensities [4]. Further impedance reduction would be useful but is limited by technical and budget considerations. Therefore, different ways of enhancing beam stability also have been investigated.

Currently, to provide a good quality beam to the LHC, a second RF system operating at 800 MHz supports the main 200 MHz RF system of the SPS. It increases the synchrotron frequency spread inside the bunch and provides more effective Landau damping of beam instabilities [5]. The longitudinal beam dynamics of the bunch train in the SPS is, in general, too complex to be treated with analytical estimations for instability growth rates in a single RF system. The double RF system and the large number of contributors to the impedance make particle tracking simulations a powerful tool in the analysis of instability mechanisms. Moreover, beam measurements in conditions close to those after LIU upgrade cannot be achieved since the present RF system is limited in power for LIU beam intensities. Predictions of future performance and longitudinal instability thresholds rely mainly on numerical simulations.

The particle tracking simulation code BLOND (Beam LONGitudinal Dynamics) [6] was used to study effects of the second RF system on beam stability and results are substantiated with beam measurements where available. In the first part of the paper we present the simulation code and the features of the SPS simulations. Then, the effects of the 800 MHz RF system on beam stability at flat top are investigated. Very promising results have been obtained in simulation at highest energy but they cannot be applied at flat bottom as explained in the third part of the publication. Finally, the goal was to find an optimum RF program for the 800 MHz RF system during the full acceleration cycle to enhance beam stability, and results are presented in the last part.

## FEATURES OF PARTICLE TRACKING SIMULATIONS IN THE SPS WITH BLOND

Developed at CERN, BLOND is a 2D particle tracking simulation code, modelling the longitudinal phase space motion of single and multi-bunch beams in multi-harmonic RF systems [6]. The particle motion is tracked through a sequence of longitudinal energy kicks and drifts. The equations of longitudinal motion are discretised in time on a turn-by-turn basis with a time step equal to the revolution

\* joel.repond@cern.ch

period  $T_{\text{rev}}$  which is  $23.1 \mu\text{s}$  in the SPS. Collective effects are taken into account by computing the induced voltage for a given impedance source, possibly on a multi-turn basis, added to the RF voltage. Various beam control loops of the LLRF system are tailor-made for each of the CERN synchrotrons; for example, the phase, frequency and synchro-loops, the one-turn delay feedback and the RF phase noise injection used for controlled emittance blow-up. The code is initially written in Python but the computationally intensive parts are optimized in C++ [7]. It has been benchmarked against measurements in different CERN accelerators [3,8,9] and also against other simulations codes like PyOrbit [10], Headtail [11] and ESME [12]. The code has been proven to be reliable and is now used to study performance of rings at CERN and even outside the laboratory.

Applied to the SPS, BLOND is an efficient tool in investigating instability mechanisms. There are many features which can be added in BLOND simulations, these include: the beam-coupling longitudinal impedance model, the large number of bunches in the beam, the bunch distribution defined by the injector, the double RF operation, the LLRF controls. During a nominal SPS cycle, four batches are injected every 3.6 s from the Proton Synchrotron (PS) with a synchronous momentum  $p_s$  of  $25.92 \text{ GeV}/c$  during the flat bottom which lasts 11.1 s. Then the beam is accelerated in 8.3 s to  $451.15 \text{ GeV}/c$  and is extracted to the LHC after half a second. Usually, only 72 bunches (or less) can be simulated because batches are weakly coupled by the SPS impedance sources [13], this keeps the computational time reasonable. At flat bottom, the bunch distribution defined by bunch rotation in the PS [14] leads to a full bucket and particle loss during direct bunch-to-bucket transfer [3]. The space-charge effect is not negligible at the injection energy [15] and is always included to the full SPS impedance model in the simulations. Other effects which impact beam stability during the cycle are the beam loading in the 200 MHz RF Travelling Wave Cavities (TWC), the particle loss, the controlled emittance blow-up applied during acceleration and the action of low-level RF controls.

### SPS Longitudinal Impedance Model

The longitudinal SPS impedance model contains broad and narrow-band resonant modes between 50 MHz and 4 GHz [2, 16, 17], see Fig. 1.

The major contributors to the impedance model are the 200 MHz TWC. Both the accelerating and High Order Mode (HOM) bands contribute significantly. The fundamental pass-band impedance is reduced by the one-turn delay feedback and the feedforward whereas the HOM band at 630 MHz is damped by means of RF couplers. The two cavities at 800 MHz used for beam stability are of travelling-wave type and are included in the impedance model. The model also contains the kicker magnets with broad-band impedance, vacuum flanges and other vacuum equipment acting mainly at high frequencies (above 1 GHz). Many smaller contributions from beam instrumentation devices, resistive wall impedance, and space-charge are also included. The

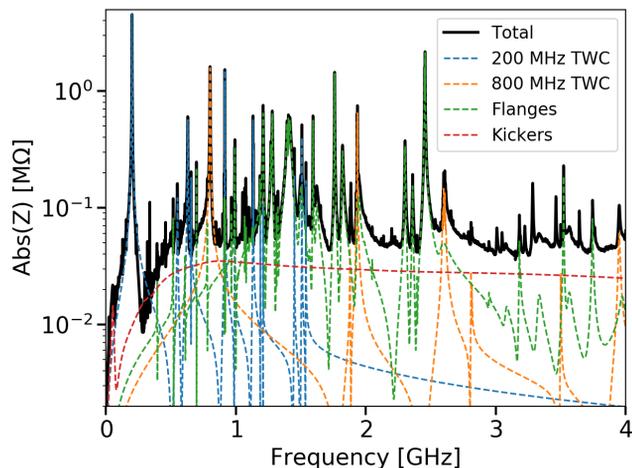


Figure 1: Longitudinal impedance model for the present configuration of the SPS. The contributions from the RF cavities, the vacuum flanges and the kicker magnets are also shown separately [2, 16, 17].

impedance of all these devices has been stimulated and/or measured over many years.

To simulate the collective effects arising from the current impedance model, careful convergence study has been performed and the best available results are presented. A large number of macroparticles (usually  $10^6$  per bunch) are needed with a sufficient number of points in the wakefield calculation in frequency domain [9].

### Double RF Operation in the SPS

The second SPS RF system operates at 800 MHz as a Landau system and is necessary to ensure stability of the LHC beam from intensity three times lower than nominal ( $1.15 \times 10^{11}$  ppb). If the amplitudes of the 200 MHz and 800 MHz voltage are respectively  $V_{200}$  and  $V_{800}$ , the total voltage provided to a particle at phase  $\phi$  is

$$V(\phi) = V_{200} [\sin \phi + r \sin(n\phi + \phi_{800})], \quad (1)$$

where  $r = V_{800}/V_{200}$  and  $n = h_{800}/h_{200} = 4$  is the ratio of the harmonic numbers. The relative phase  $\phi_{800}$  has a big impact on the synchrotron frequency distribution and can be determined to maximise the synchrotron frequency spread in the bunch center. At a given time in the cycle, the synchronous phase in a single RF system  $\phi_{s0}$  is linked to the energy gain of the synchronous particle  $\delta E_s$  by

$$\delta E_s = V_{200} \sin \phi_{s0}. \quad (2)$$

For the same energy gain  $\delta E_s$ , the synchronous phase  $\phi_s$  in double RF is related to  $\phi_{s0}$  by

$$\sin \phi_{s0} = \sin \phi_s + r \sin(n\phi_s + \phi_{800}). \quad (3)$$

The synchrotron frequency in the bunch center  $f_s(0) = \omega_s(0)/2\pi$  is modified by the second RF as follow

$$\omega_s^2(0) = \frac{\omega_{s0}^2(0)}{\cos \phi_{s0}} [\cos \phi_s + r n \cos(n\phi_s + \phi_{800})], \quad (4)$$

where  $\omega_{s0}(0)$  is the synchrotron angular frequency in single RF. The value of  $\phi_{800}$  maximising the synchrotron frequency

spread is determined by Eq. (4) and the new synchronous phase is found from Eq. (3). At flat bottom or flat top in the SPS (above transition energy),  $\phi_{s0} = \phi_s = \pi$ . Two possible values of  $\phi_{800}$ , 0 and  $\pi$ , maximise the synchrotron frequency spread. The first one ( $\phi_{800} = 0$ ) is called the Bunch Lengthening Mode (BLM) and the second ( $\phi_{800} = \pi$ ) is called the Bunch Shortening Mode (BSM). The names come from the effect these two modes have on the bunch length for  $n = 2$ .

The synchrotron frequency distribution can be written as a function of the single particle emittance (action variable). This emittance corresponds to the area enclosed by the particle trajectory in phase space and is measured usually in  $eVs$ . For  $n = 4$ , depending on the voltage ratio and the phase between both RF systems, the derivative of the synchrotron frequency distribution goes to zero and a plateau appears in the distribution, see Fig. 2. Particles in this region develop a large coherent response [18]. The Landau damping is lost and instabilities can be triggered by any perturbation.

In the nominal operation of the SPS, the BSM is used because only this configuration provides beam stability during ramp. The nominal bunch emittance is small enough to stay away from the plateau of the synchrotron frequency distribution for operational values of the voltage ratio ( $r = 0.1$ ). This is not true in BLM where the flat portion appears for emittances smaller than nominal. The situation is also different at flat bottom where bunches are longer compared to flat top. During the cycle, the relative phase of the 800 MHz RF system is approximated in BSM by  $\phi_{800} = \pi - 4\phi_{s0}$  [5].

## EFFECT OF 800 MHz RF SYSTEM ON BEAM STABILITY AT SPS FLAT TOP

At high intensity, coupled-bunch instabilities are observed during the ramp. The mitigation measures for impedances giving the lowest stability threshold have been identified [4], but other possible cure also have been investigated, see, for example, [19]. The optimisation of the 800 MHz operation is one of them.

The intensity threshold, for constant emittance, increases with the relative synchrotron frequency spread [18], which is increased by a larger ratio  $r$ . For higher voltage ratios up to 0.3, bunches with nominal emittance (0.6 eVs after controlled emittance blow-up) are not affected by the flat portion of the synchrotron frequency distribution, see Fig. 2, and beam stability is improved when the voltage ratio increases. This effect has been seen in simulations and then tested in measurements. A batch of 12 bunches was used in measurements to be able to accelerate high intensity bunches since for 72 bunches the beam loading is too high for the RF power available. In this experiment, the nominal emittance of 0.35 eVs was used without controlled emittance blow-up during ramp. Figure 3 shows the average bunch length at flat top with error bars representing the maximum and minimum bunch length measured along the batch as a function of the bunch intensity in the case  $r = 0.1$  (a) and  $r = 0.25$  (b), kept during whole cycle. In the first case, large oscillations

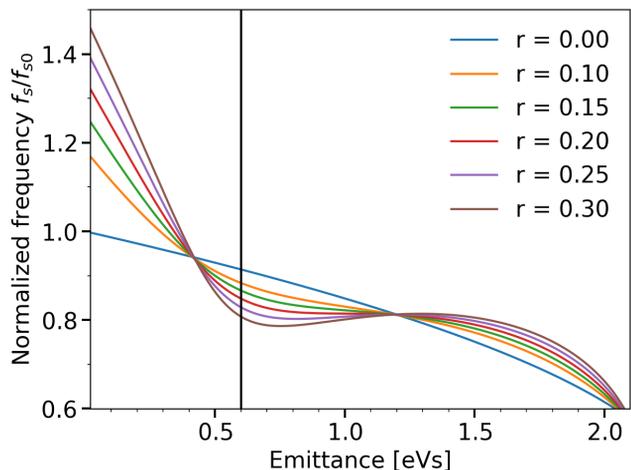


Figure 2: Synchrotron frequency distribution for the double RF system ( $n = 4$ ) in BSM at SPS flat top as a function of the single particle emittance. For different values of the voltage ratio  $r$ , the frequencies are normalized by the synchrotron frequency in the bunch center computed for  $r = 0$ . The 200 MHz voltage is  $V_{200} = 7$  MV. The vertical black line indicates the nominal bunch emittance at flat top (0.6 eVs).

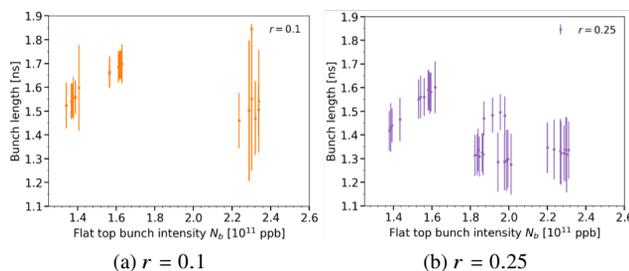


Figure 3: Average bunch length at flat top measured for batches of 12 bunches with nominal emittance (0.35 eVs). The error bars represent the maximum and minimum bunch length measured along the batch. The cases  $r = 0.1$  (a) and  $r = 0.25$  (b), kept during whole cycle, are presented. The 200 MHz voltage at flat top is 7 MV and feedback and feedforward were activated during whole cycle.

are observed for higher bunch intensity ( $N_b > 2.2 \times 10^{11}$ ) which are suppressed when  $r = 0.25$ . The bunch length is computed from the Full Width at Half Maximum (FWHM) of the bunch profile, rescaled to  $4\sigma$  assuming a Gaussian bunch. Since the measured stability threshold is reproduced at flat top in simulations of 12 bunches (see [20]), simulations are then used to study the effect of the 800 MHz RF system on beam stability with a nominal batch containing 72 bunches. Operation also confirms the increase of beam stability with a larger voltage ratio  $r$  only on flat top.

### Simulated Intensity Threshold for 72 Bunches

The instability threshold is at a minimum value at flat top. The simulations were done at a constant momentum of 451.15 GeV/c. Bunches which do not become unstable during acceleration can be considered as matched to the RF bucket with intensity effects at flat top. Then, 72 bunches

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

spaced by 25 ns were generated with a bunch distributions matched to the RF bucket including intensity effects and parameters in agreement with measurements. The bunch distribution has the form of the binomial function

$$F(J) = F_0 \left(1 - \frac{J}{J_0}\right)^\mu, \quad J \in [0, J_0], \quad (5)$$

where  $J = 2\pi\epsilon$  is the action with  $\epsilon$  being the bunch emittance. The emittance is computed in simulation by integrating the unperturbed potential well over the  $4\sigma$  Gaussian bunch length obtained from the FWHM of the bunch profile;  $J_0/2\pi$  is the initial bunch emittance and  $\mu = 1.5$  was chosen, in agreement with measurements. Bunch emittance and intensity values were scanned to obtain the stability map.

The transient beam loading in the main 200 MHz RF system saturates quickly after injection and in the stationary regime, the impedance of the 200 MHz harmonic is reduced by the one-turn-delay feedback and feedforward by 20 dB, in agreement with measurements [3]. The present SPS impedance model has been used. The maximum voltage at 200 MHz was 7 MV for all intensities. Due to beam loading, the available voltage is intensity dependent and 7 MV can be obtained only at intensities comparable to nominal. However, the goal was to observe the effect of the fourth harmonic RF system; the power limitation will be raised in future. The simulated time at flat top was two seconds (compared to the 500 ms in the SPS operation) to observe slowly growing instabilities. However, in relevant intensity range, up to  $2.5 \times 10^{11}$  ppb, the multi-bunch instabilities are violent and appear before half a second. The ratio  $r$  was varied between 0.1 and 0.3. Larger values were not considered due to hardware limitations. A maximum ratio of 0.1 will be achievable for HL-LHC intensity after RF upgrades since the 200 MHz voltage at flat top will be increased to 10 MV. Figure 4 shows a good agreement of simulations with the reference measurements for the nominal case ( $r = 0.1$ ) [4] and demonstrates that the SPS is already pushed to its limits for a batch of 72 bunches. Increasing the voltage ratio on flat top up to  $r = 0.3$  improves the stability threshold. For the largest value, the intensity limit is increased by 150%. Simulations for situations after LIU upgrade also show that an increase of the voltage ratio can improve the stability threshold even beyond the scope of the HL-LHC project. However, other limitations will start to play a role—beam loading in the 200 MHz RF system for example.

## EFFECT OF 800 MHz RF SYSTEM ON BEAM STABILITY AT FLAT BOTTOM

The significant improvement of beam stability with a larger voltage ratio  $r$  cannot be obtained at flat bottom. In this case, some particles within the nominal injected emittance have a region in synchrotron frequency where the derivative goes to zero. With  $V_{200} = 4.5$  MV, a flat portion in synchrotron frequency distribution appears for the voltage ratio above  $r = 0.15$ , see Fig. 5. Measurements also show larger bunch length oscillations at flat bottom when the voltage ratio  $r$  is increasing, see Fig. 6. For a ratio  $r = 0.25$ , the

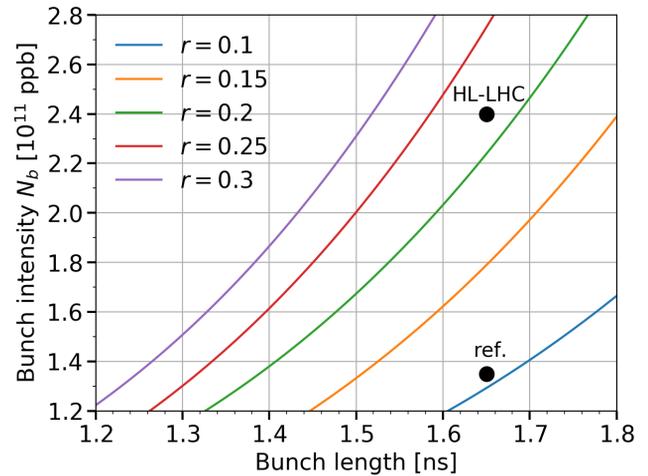


Figure 4: Simulated stability threshold at flat top as a function of the bunch length for 72 bunches spaced by 25 ns. The longitudinal impedance model (Fig. 1) is used.  $V_{200} = 7$  MV and  $V_{800} = r V_{200}$ . A reference measurement for four batches with 72 bunches spaced by 25 ns is included [4], as well as the LIU intensity target. The maximum amplitude of the bunch length oscillations during cycle (normalised by the average) was used as a criterion to separate stable from unstable beams.

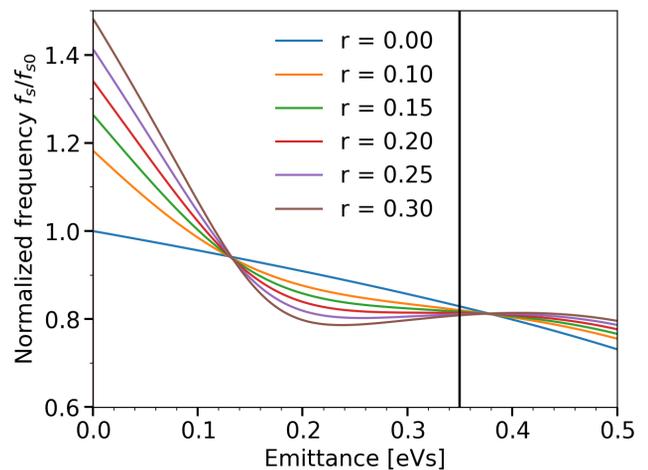


Figure 5: Normalised synchrotron frequency distribution for the double RF system in BSM at SPS flat bottom, as a function of the emittance for different value of the voltage ratio  $r$ . The 200 MHz voltage is  $V_{200} = 4.5$  MV. The vertical black line indicates the nominal bunch emittance of 0.35 eVs.

bunch length oscillations increase compared to the lower value  $r = 0.1$ . Larger losses also appear at the start of acceleration.

When the feedback and feedforward were deactivated during whole cycle, a beam instability has been observed at flat bottom for different bunch train at lower intensity, comparable to nominal. The instability is likely caused by the impedance of the 200 MHz main harmonic which will be further reduced after planned RF upgrades. However, if another impedance source contributes to the instabilities, the 800 MHz RF system will lack efficiency to mitigate it, since

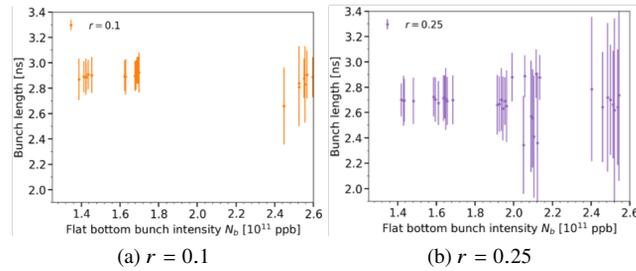


Figure 6: Average bunch length at flat bottom measured for batches of 12 bunches with nominal emittance (0.35 eVs). The error bars represent the maximum and minimum bunch length measured along the batch. The cases  $r = 0.1$  (a) and  $r = 0.25$  (b), kept during whole cycle, are presented. The 200 MHz voltage at flat bottom was 4.5 MV, feedback and feedforward were activated during whole cycle.

the bucket is full. With the one-turn delay feedback and feedforward activated at flat bottom, it has been observed that a voltage ratio of 0.1 provides better stability for trains of 48 bunches with intensities above nominal. To remove the plateau in the synchrotron frequency distribution, it is also possible to shift the relative phase  $\phi_{800}$  away from the BSM. Improvements of the stability threshold with a phase shift have been shown in simulations [21]. However, the longitudinal acceptance, already full, is reduced.

The intensity thresholds measured for batches of 12 bunches without feedback and feedforward are presented in Fig. 7. First, simulations have been carried out with bunches matched to the RF bucket (with intensity effects). The maximum amplitude of the bunch length oscillations during cycle (normalised by the average) was used as a criterion to separate stable beam from the unstable one and similarly in measurements. As one can see in this case, the stability limit is far above the measured one. The simulations done with realistic bunches gave much better results. This indicates that the bunch distribution defined by the injector has a large effect on the instability occurring during the 10 s flat bottom cycle. The realistic distribution is defined by the bunch rotation in the PS and has an S-shape [3]. Particles completely fill the RF bucket after filamentation and resulting bunch profile has larger components interacting with the high frequency part of the machine longitudinal impedance.

The bunch distribution after rotation was generated by simulations in the PS without intensity effects. Bunches were matched at PS flat top with the distribution from Eq. (5) using a large number of macroparticles ( $3.6 \times 10^7$ ). The nominal RF program for bunch rotation in the PS is used. Then, each of the 12 bunches in the SPS are generated as a subset of one million macroparticles randomly selected. Simulations are compared with measurements done in single RF in Fig. 8. The measured stability threshold is reproduced if the rotated bunch distribution is used in simulation. The double RF operation with a voltage ratio  $r = 0.1$  does not improve the beam stability.

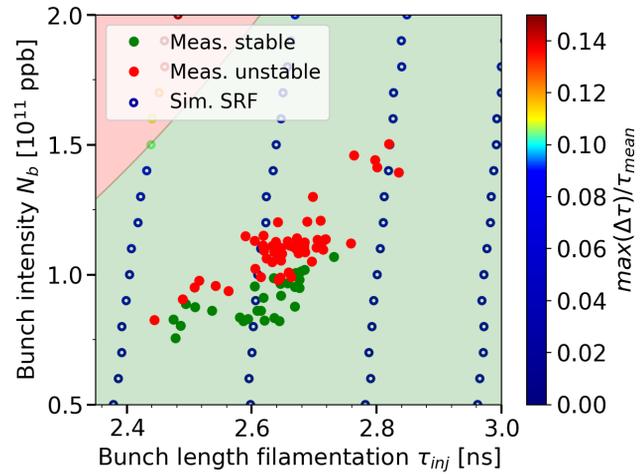


Figure 7: Stability threshold at flat bottom as a function of the bunch length after filamentation for 12 bunches matched to the RF bucket with intensity effects. The 200 MHz voltage is 4.5 MV and  $r = 0$  (SRF), feedback and feedforward are deactivated. The longitudinal impedance model (Fig. 1) is used. Beam measurements in the same configuration are included. For simulations, colours of circles correspond to the maximum amplitude of the bunch length oscillations during cycle normalised by the average.

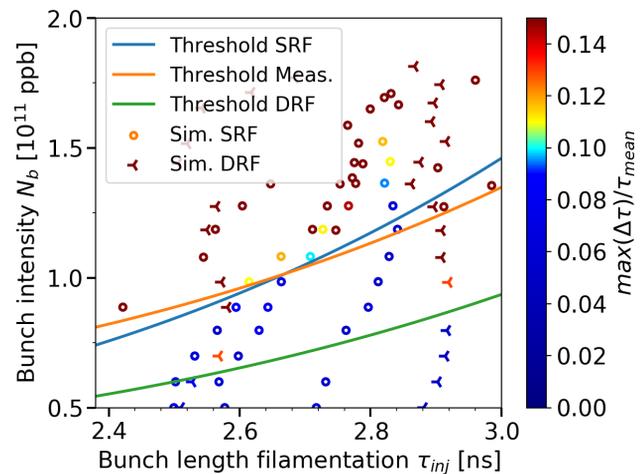


Figure 8: Intensity limit of 12 bunches as a function of the bunch length after filamentation at flat bottom. The bunch distribution is generated by simulations of the rotation in the PS. The single RF case (SRF) is compared with the intensity threshold measured under the same conditions. The results for double RF operation (DRF,  $r = 0.1$ ) are also shown. The feedback and feedforward were deactivated. For simulations, colours of circles correspond to the maximum amplitude of the bunch length oscillations during cycle normalised by the average.

Different values of  $\mu$  were used for the bunch generation in the PS. Figure 8 presents the results with  $\mu = 1$  from Eq. (5) but similar stability limits are obtained for larger value of  $\mu$  up to 2. Larger values of  $\mu$  have been also studied and the intensity threshold decreases significantly. Assuming that the flat bottom instability can be cured by the one-turn delay

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

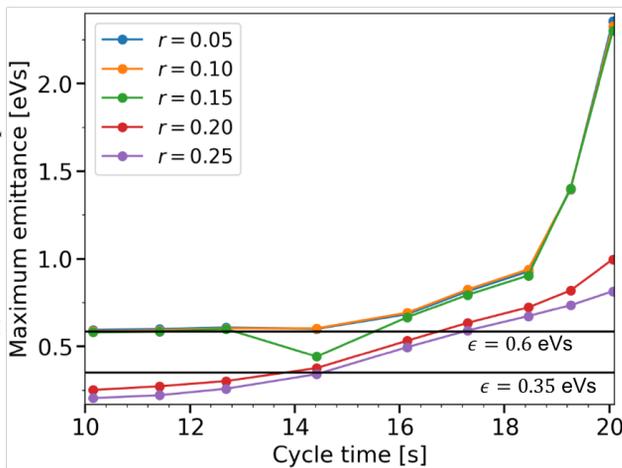


Figure 9: Maximum bunch emittance  $\epsilon_{max}$  during cycle for different voltage ratios  $r$ , where  $\epsilon_{max}$  is the critical emittance defined by Eq. (6) if it exists or the bucket area otherwise. The two horizontal lines show the nominal emittances at flat bottom (0.35 eVs) and at flat top (0.6 eVs).

feedback and feedforward systems, to improve the beam stability during all the cycle the voltage ratio should be kept at low value ( $r \leq 0.1$ ) at flat bottom and increased during acceleration to reach the largest value at flat top.

## VOLTAGE OPTIMISATION DURING CYCLE

To determine an optimal voltage ratio during cycle, we define the critical emittance

$$\epsilon_c = \min\{0 < \epsilon' \leq A \text{ such that } \frac{\partial}{\partial \epsilon} f_s(\epsilon = \epsilon') = 0\}, \quad (6)$$

where  $A$  is the bucket acceptance. If  $\epsilon_c$  exists, the synchrotron frequency distribution has a plateau and  $\epsilon_c$  is the maximum allowed bunch emittance  $\epsilon_{max}$ . If  $\epsilon_c$  does not exist, the maximum emittance is the acceptance. The synchrotron frequency and its derivative are computed numerically during cycle without intensity effects. The evolution of  $\epsilon_{max}$  is shown for different voltage ratios in Fig. 9.

A voltage ratio  $r \geq 0.15$  at flat bottom creates a plateau in the synchrotron frequency distribution for the nominal emittance, so the voltage ratio was fixed to  $r = 0.1$ . During acceleration, after 16 s (before blow-up),  $r$  can be increased to 0.15 and after 18 s the ratio can be increased to 0.25. The resulting voltage program is plotted in Fig. 10. These settings have been tested in real conditions with up to four batches of 12 bunches and improvement of beam stability was demonstrated for two different SPS optics (Q20 and Q22).

However, one should also keep in mind that intensity effects modify the synchrotron frequency distribution. In simulations for high intensity ( $2.3 \times 10^{11}$  ppb) the synchrotron frequency distribution is affected by the induced voltage differently for each bunch. The frequency at zero amplitude is reduced by 4% for the first bunch and by 11% for the twelfth bunch. As a next step in the optimisation of the voltage

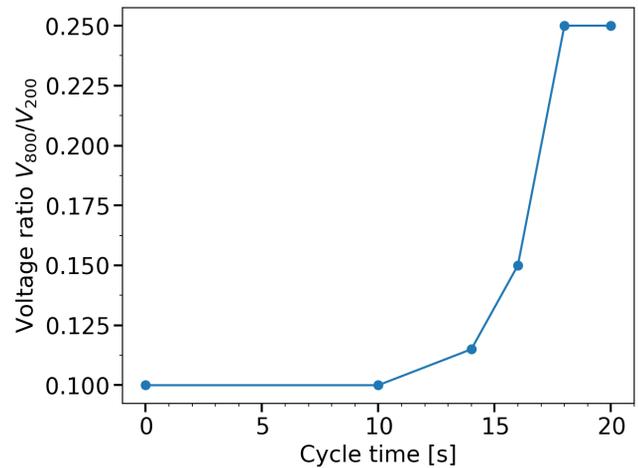


Figure 10: Optimised voltage ratio  $r$  between the two SPS RF systems during acceleration cycle for the LHC proton beam.

program, the collective effects could be taken into account in the design of the voltage program.

## CONCLUSION

The fourth harmonic RF system is one of the main causes of beam instabilities in the SPS. Simulations have shown the possibility to significantly (by 150%) improve the stability threshold at flat top by increasing the voltage ratio between main and fourth harmonic RF systems. In the present operation, the voltage ratio of two RF systems is fixed at 0.1 during whole cycle. At flat bottom, larger bunch length oscillations are observed when the intensity increases up to  $2.5 \times 10^{11}$  ppb with voltage ratios larger than  $r = 0.1$ . They are caused by the plateau in the synchrotron frequency distribution. If the voltage ratio  $r = 0.1$  is used at flat bottom and increased during ramp to reach  $r = 0.25$  at flat top, the stability threshold is improved. The stability enhancement with these settings has been demonstrated in operation with four batches of 12 bunches for an injected intensity of  $N_b = 2.3 \times 10^{11}$  ppb.

## ACKNOWLEDGEMENTS

We would like to thank T. Bohl, A. Lasheen, A. Farricker, G. Papotti, H. Timko and our colleagues from the SPS OP for fruitful input, discussions and help.

## REFERENCES

- [1] G. Apollinari *et al.*, *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, ser. CERN Yellow Reports: Monographs. Geneva: CERN, 2017.
- [2] J. Coupard *et al.*, *LHC Injectors Upgrade: Technical Design Report, Vol I: Protons*. CERN, Geneva: CERN-ACC-2014-0337, 2014.
- [3] M. Schwarz, H. Bartosik, E. Shaposhnikova, A. Lasheen, J. Repond, and H. Timko, "Studies of Capture and Flat-Bottom Losses in the SPS," in *Proc. HB2018*, Daejeon, Korea, 17–22 Jun 2018, pp. 180–185. doi: 10.18429/JACoW-HB2018-TUP2WA03

- [4] E. Shaposhnikova, T. Argyropoulos, T. Bohl, P. Cruikshank, B. Goddard, T. Kaltenbacher, A. Lasheen, J. Perez Espinos, J. Repond, B. Salvant, and C. Vollinger, "Removing Known SPS Intensity Limitations for High Luminosity LHC Goals," in *Proc. IPAC'16*, Busan, Korea, 8–13 May 2016, pp. 989–991. doi:10.18429/JACoW-IPAC2016-MOP0Y058
- [5] T. Bohl, T. Linnecar, E. Shaposhnikova, and J. Tückmantel, "Study of Different Operating Modes of the 4th RF Harmonic Landau Damping System in the CERN SPS," in *Proc., EPAC'98*, Stockholm, Sweden, 22–26 Jun 1998, pp. 978–980. <http://jacow.org/e98/PAPERS/THP09A.PDF>
- [6] "Beam Longitudinal Dynamics code (BLonD)," <http://blond.web.cern.ch>, 2018.
- [7] K. Iliakis *et al.*, "BLonD++: Performance Analysis and Optimizations for Enabling Complex, Accurate and Fast Beam Dynamics Studies," in *Proc. SAMOS'18*. Samos, Greece: to be published, 15–19 Jul 2018.
- [8] H. Timko, J. Esteban Müller, A. Lasheen, and D. Quartullo, "Benchmarking the Beam Longitudinal Dynamics Code BLonD," in *Proc. IPAC'16*, Busan, Korea, 8–13 May 2016, pp. 3094–3097. doi:10.18429/JACoW-IPAC2016-WEPOY045
- [9] D. Quartullo, M. Migliorati, and J. Repond, "Comparison of Different Methods to Calculate Induced Voltage in Longitudinal Beam Dynamics Codes," in *Proc. IPAC'17*, Copenhagen, Denmark, 14–19 May 2017, pp. 4465–4468. doi:10.18429/JACoW-IPAC2017-THPVA022
- [10] A. Shishlo *et al.*, "The particle accelerator simulation code pyorbit," *Procedia Computer Science*, vol. 51, pp. 1272–1281, 2015, international Conference On Computational Science, ICCS 2015. <http://www.sciencedirect.com/science/article/pii/S1877050915011205>
- [11] G. Rumolo and F. Zimmermann, "Practical user guide for HEADTAIL," SL-Note-2002-036-AP, Geneva, Nov 2002. <http://cds.cern.ch/record/702717>
- [12] J. A. MacLachlan, "ESME: Longitudinal Phase Space Particle Tracking. Program Documentation." FERMI LAB-TM-1274, Fermilab, Batavia, 1984.
- [13] E. Shaposhnikova, T. Argyropoulos, T. Bohl, A. Lasheen, J. Repond, and H. Timko, "Identification and reduction of the CERN SPS impedance," in *Proc. HB2016*, Malmö, Sweden, 3–8 Jul 2016, pp. 260–265. doi:10.18429/JACoW-HB2016-MOPR028
- [14] A. Lasheen, H. Damerou, J. Repond, M. Schwarz, and E. Shaposhnikova, "Improvement of the Longitudinal Beam Transfer from PS to SPS at CERN; Improvement of the Longitudinal Beam Transfer from PS to SPS at CERN," in *Proc. IPAC'18*, Vancouver, BC Canada, 2018, pp. 3060–3063. doi:10.18429/JACoW-IPAC2018-THPAF042
- [15] A. Lasheen, "Longitudinal Space Charge in the SPS." CERN-ACC-NOTE-2016-0074, Dec 2016. <https://cds.cern.ch/record/2238995>
- [16] G. Dôme, "The SPS acceleration system travelling wave drift-tube structure for the CERN SPS," in *Proc. 9th Linear Accelerator Conference*, Chalk River, Canada, 14–17 Sep 1976, pp. 138–147. <http://cds.cern.ch/record/319440?ln=en>
- [17] J. E. Campelo *et al.*, "An Extended SPS Longitudinal Impedance Model," in *Proc IPAC'15*, Richmond, VA, USA, 3–8 May 2015, pp. 360–362. doi:10.18429/JACoW-IPAC2015-MOPJE035
- [18] E. Shaposhnikova, "Bunched beam transfer matrices in single and double rf systems," CERN-SL-94-19-RF, Geneva, Aug 1994. <https://cds.cern.ch/record/267121>
- [19] J. Repond and E. Shaposhnikova, "Effect of HOM Frequency Shift on Bunch Train Stability," in *proc. of the ICFA Mini-Workshop Workshop Impedances and Beam Instabilities in Particle Accelerators*, Benevento, Italy, 18–22 Sep 2017, pp. 93–97.
- [20] J. Repond, "LIU-SPS Beam Dynamics Working Group meeting, Longitudinal stability of 12 bunches," [https://paf-spsu.web.cern.ch/paf-spsu/meetings/2018/m\\_28\\_august/SPSU\\_stability\\_flatbottom\\_v2.pptx](https://paf-spsu.web.cern.ch/paf-spsu/meetings/2018/m_28_august/SPSU_stability_flatbottom_v2.pptx), Jul 2018.
- [21] T. Argyropoulos, E. Shaposhnikova, and A. Burov, "Loss of Landau Damping for Inductive Impedance in a Double RF System," in *Proc. IPAC'13*, Shanghai, China, 12–17 May 2013, pp. 1796–1798. <http://JACoW.org/IPAC2013/papers/tupwa040.pdf>

# A FULL FIELD-MAP MODELING OF CORNELL-BNL CBETA 4-PASS ENERGY RECOVERY LINAC\*

F. Méot<sup>†</sup>, N. Tsoupas, S. Brooks, D. Trbojevic, BNL C-AD, Upton, NY, USA  
J. Crittenden, Cornell University (CLASSE), Ithaca, NY, USA

## Abstract

The Cornell-BNL Electron Test Accelerator (CBETA), a four-pass, 150 MeV energy recovery linac (ERL), is now in construction at Cornell. Commissioning will commence in March 2019. A particularity of CBETA is that a single channel loop recirculates the four energies (42, 78, 114 and 150 MeV). The return loop arcs are based on fixed-field alternating gradient (FFAG) optics. The loop is comprised of 107 quadrupole-doublet cells, built using Halbach permanent magnet technology. Spreader and combiner sections (4 independent beam lines each) connect the 36 MeV linac to the FFAG arcs. We introduce here to a start-to-end simulation of the 4-pass ERL, entirely, and exclusively, based on the use of magnetic field maps to model the magnets.

## INTRODUCTION

The Cornell-BNL Electron Test Accelerator (CBETA), a four-pass, 150 MeV energy recovery linac (ERL), is now in construction at Cornell. A particularity of CBETA is in its single channel loop recirculating four energies, 42, 78, 114 and 150 MeV, four-pass up, four-pass down. The return loop arcs (FA-TA and TB-FB sections, Fig. 1) are based on fixed-field alternating gradient (FFAG) optics. The loop is comprised of 107 quadrupole-doublet cells, built using Halbach permanent magnet technology. Spreader (SX) and combiner (RX) sections (4 independent beam lines each) connect the 36 MeV linac to the FFAG arcs. This paper introduces to a start-to-end simulation of the 4-pass ERL, entirely, and exclusively, based on the use of magnetic field maps to model the magnets, now under development in view of the commissioning of CBETA which will commence in March 2019.

The OPERA field maps of the return loop Halbach magnets are produced at BNL. The OPERA field maps of most of the spreader and combiner line conventional electro-magnets are produced at Cornell.

## Why Use Field Maps?

There is a variety of reasons for that:

- All necessary material is available or will soon be: the return loop Halbach magnet field maps have been produced during the design [1], the spreader and combiner section conventional magnet field maps (dipoles and quadrupoles) are under production. Thus, as it yields highest simulation accuracy, why not just do it? And,

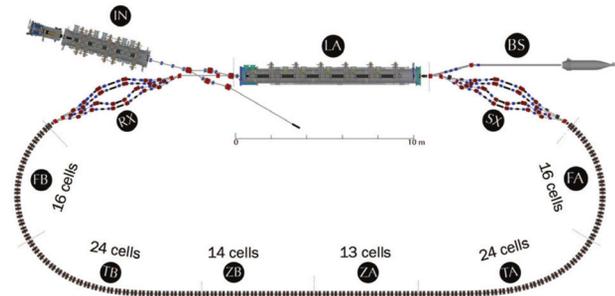


Figure 1: CBETA 150 MeV ERL [2]. The linac is 36 MeV, four different energies circulate concurrently in the single-channel return loop: 42, 78, 114 and 150 MeV (hence, 4 spreader (SX) and recombiner lines (RX), at linac downstream and upstream ends, respectively).

in passing, forget about questionable mapping approximations.

- FFAG experience dictates to do so: as early as in the 1950s, Frank Cole wrote on the virtues of the use of field maps and Runge-Kutta ray-tracing in designing and operating the MURA scaling FFAG rings [3]:

*"[...] digital computation to explore nonlinear problems in spiral-sector orbits. This was not done by mapping in the usual sense of the term, but by step-by step integration of the equations of motion, using the fourth-order Runge-Kutta method. It was a marvelous productive year for the [MURA] group."*

Kyushu University and KURRI 150 MeV scaling FFAG proton rings (amongst others in Japan) were designed, and are operated, using 3D OPERA field maps of the cell dipoles [4]; the RACCAM spiral FFAG dipole constructed and measured in 2009 was designed and optimized, successfully, based on field map simulations [5–7]; the optics of the EMMA linear FFAG ring accelerator at Daresbury (CBETA arc cell is similar to EMMA's) was studied using OPERA field maps of its QF-QD cell magnets [8].

- Using field maps yields closest-to-real-life modeling of the Halbach doublets return loop, over the all 8 passes (4 accelerated, 4 decelerated).

Now, the method must be validated. This will be discussed here and includes showing the feasibility of

- using separate field maps, especially of the QF and BD focusing quadrupole and combined function defocusing dipole in the return loop,
- including field overlapping between neighboring magnets, all along the return loop,
- and accounting for iron yoke corrector magnets superimposed on the Halbach magnets.

\* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy

<sup>†</sup> fmeot@bnl.gov

The interest of using separate field maps is in the flexibility in the modeling, allowing in particular,

- independent fine-tuning of QF, QD and BD Halbach magnet strengths,
- an independent power-supply knob for each corrector field map,
- the possibility of independent field and positioning errors and compensation,
- easier connection between CBETA sectors (FA, TA, ZA, etc., Fig. 1).

## OVERVIEW

The rest of this technical note consists in a series of figures with self-explanatory captions (Figs. 2–14), together with some comments and sample input data lists to the ray-tracing code used in this modeling of CBETA [9, 10]. This Section gives an overview of the methods and present outcomes. This is a work in progress, thus this note will conclude on partial completion, the plan being to have a complete simulation in due time, in particular a 1-pass up, 1-pass down loop ready for the start of the commissioning.

Note that the code used is under development at Radiasoft [11], which includes its installation in the SIREPO environment [12]. Figure 2 shows preliminary aspects of the latter, more is to come in near future.

### OPERA Simulation of CBETA Arc Cell

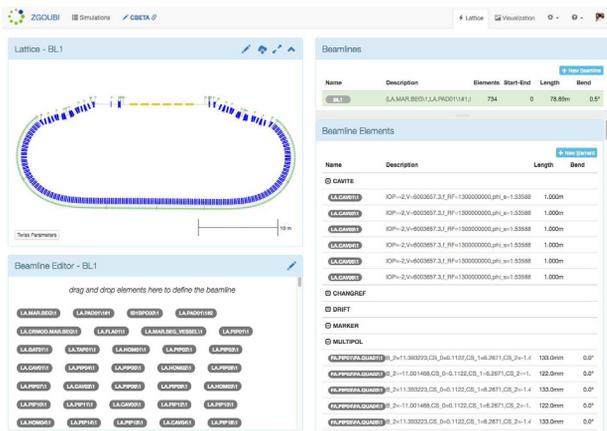


Figure 2: Layout of the CBETA 42 MeV pass, in SIREPO environment [12].

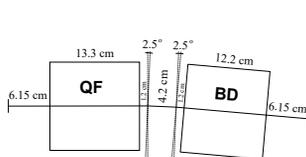
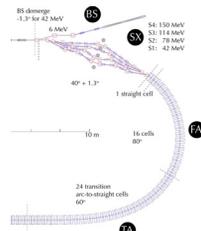


Figure 3: CBETA FFAQ QF-BD cell in the FA and FB sections of the arcs (Figs. 1, 4). Figure 4: SX and FA-TA arc.



Optical sequence of the arc cell (Fig. 3) in Zgoubi, case of a single full-cell field map:

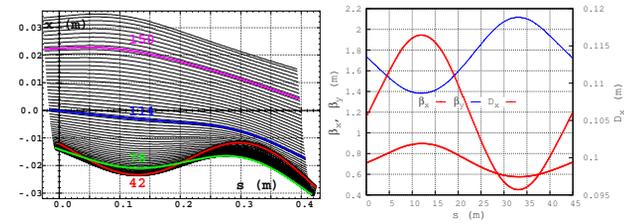


Figure 5: An energy scan of the orbits across the arc cell, including the 4 design energies (left) and the 42 MeV optical functions (right), derived from the OPERA field maps modeling of the QF and BD Halbach magnets.

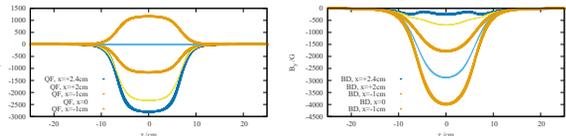
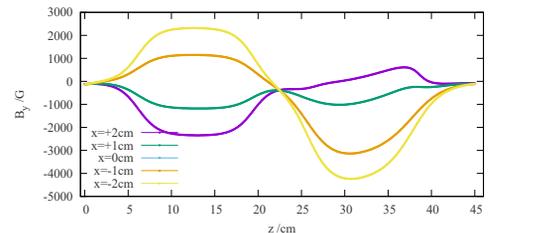
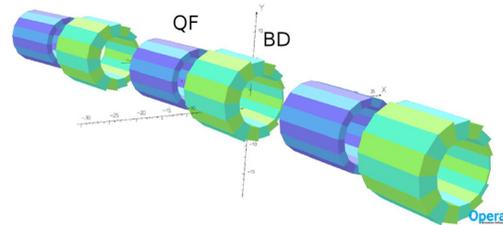


Figure 6: Top: the OPERA field map of a full cell is computed from the middle QF-BD doublet of a series of three, to ensure periodicity of the field. Middle: the resulting mid-plane field across the cell, samples taken at various distances x from the cylinder axis. Bottom: mid-plane field across the magnets, at various distances x from the cylinder axis, case of separate computation of the two field maps.

```
'TOSCA' QF+BD
0 0
-9.69871600E-04 1.000 1.000 1.000
HEADER_8 ZroBXY
451 83 27 15.1 1.
3cellFieldMap.table
1 -508.5 44.49 2.2E4 ! MOTION BOUNDARY
2
.2
2 0.000 0.000 0.000
'CHANGREF'
XS -0.678391 YS -1.8870962 ZR -5.0
```

Optical sequence of the arc cell (Fig. 3) in Zgoubi, case of separate QF, BD maps:

```
'DRIFT'
6.15
'DRIFT'
-18.35 !=(50cm - 13.3cm)/2 (50cm is field map extent)
'TOSCA' QF
0 0
-9.76E-04 1. 1. 1.
HEADER_8 ZroBXY
501 83 1 15.1 1.
```

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

```

QF-3D-fieldMap.table
0 0 0
2
.2
2 0.00E+00 0.00E+00 0.00E+00
'DRIFT'
-18.35 ! =(50cm - 13.3cm)/2 (50cm is field map extent)
'DRIFT' ED1
1.2
'CHANGREF' CORNER
ZR -2.50
'DRIFT' BPM
4.2
'CHANGREF' CORNER
ZR -2.50
'DRIFT' ED1
1.2
'DRIFT'
-18.9 ! =(50cm - 12.2cm)/2 (50cm is field map extent)
'TOSCA' BD
0 0
-9.76E-04 1.00E+00 1.00E+00 1.00E+00
HEADER_8 ZroBXY
501 83 1 15.1 1.0
BD-3D-fieldMap.table
0 0 0
2
.2
2 0. -.019 0. ! Y-offset -0.019cm = inward
'DRIFT'
-18.9 ! =(50cm - 12.2cm)/2 (50cm is field map extent)
'DRIFT' HD2
6.15
    
```

### Beam Optics Validations

**1: First order parameters of the arc cell** They are displayed in Figs. 7 and 8. Table 1 details the path length at the four design energies, depending on the field map modeling method. Differences do not exceed a few ppm.

Table 1: Path Length, Detailed Values

Path length across cell (cm)				
E (MeV)	42	78	114	150
Single 3D map	44.4846	44.3298	44.3898	44.5806
Two 2D or 3D maps	44.4845	44.3291	44.3884	44.5797

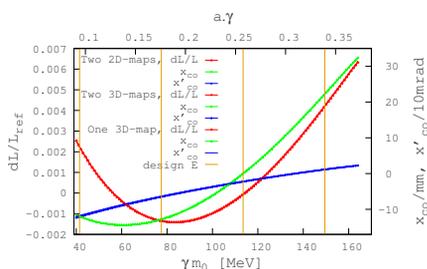


Figure 7: Separate 2D or 3D field maps of QF and BD, or 3-D full-cell single map, yield the same closed orbit coordinates (at the center of the long drift, here), and the same trajectory lengthening, all superimposed on this graph.

**2: Dynamical admittance** The dynamical admittance at a given energy, here, is taken as the maximum stable invariant that makes it through a 400 cell channel, for that energy: beyond that invariant, particles get kicked away under the effect of field or kinematical non-linearities. Results are displayed in Fig. 9.

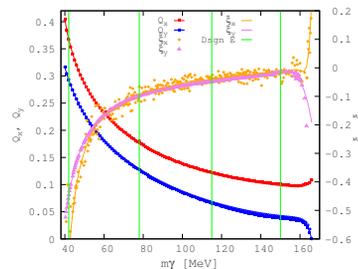


Figure 8: Separate 2D or 3D field maps of QF and BD, or 3-D full-cell single map, yield same paraxial tunes and chromaticities.

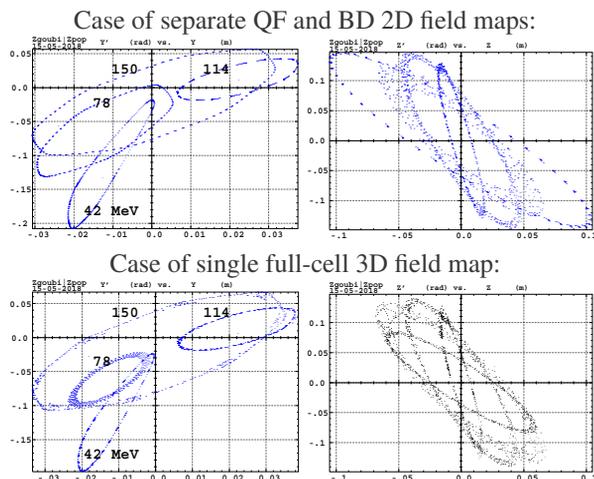


Figure 9: Left column: horizontal motion; right column: vertical motion. Observation plan is at the middle of the long drift. Non-linearities at the origin of the limited amplitude are from the field and from kinematic terms in the motion. The maximum invariant values are ~meter normalized, they are comparable in the two cases, two separate field maps of a single full-cell map - and far beyond  $\mu\text{m}$  CBETA beam emittance.

**3: Dynamical admittance, energy scan** A similar exercise to the previous one, repeated for a series of energies ranging from 39 to 170 MeV (Fig. 10).

### Closer to CBETA FFAG Cell

We want the cell model even fancier, Fig. 11. Namely, including the H and V orbit correction dipoles (iron yoke electromagnets), on top of respectively the F and D Halbach FFAG magnets. This requires two independent field maps. In the case of a full-cell single field map for instance, as was done for the EMMA FFAG ring [8], each one of the two additional field maps comprises the corrector pair, however,

- one corrector pair field map has the F-corrector on and the D-corrector off,
- the other one has the F-corrector off and the D-corrector on. This allows independent knobs for these correctors.

The complete return loop is at present operational, fully field-map.

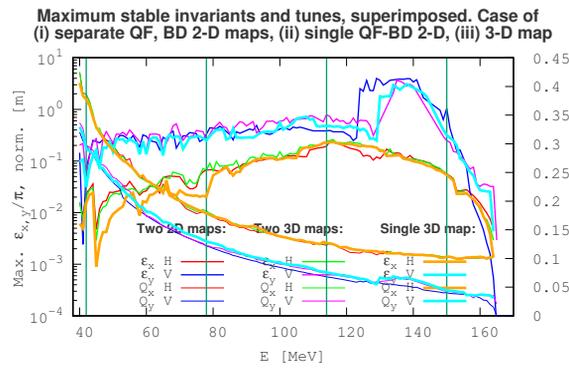


Figure 10: “H”: horizontal motion (initial V invariant is taken very small). “V”: vertical motion (initial H invariant is taken very small). The DA curves in this graph are the surfaces of the phase space curves as shown in Fig. 9, repeated for different energies. The H and V tunes are for these maximum invariants, computed using a DFT. It can be seen that the 42 MeV beam is placed away from the Walkinshaw resonance (the dip in the vertical acceptance, to the left of the 42 MeV vertical bar), and from the  $Q_x = 1/3$  resonance (the dip in the horizontal acceptance, to its right). The superimposition shows that the three different field map models yield comparable results.

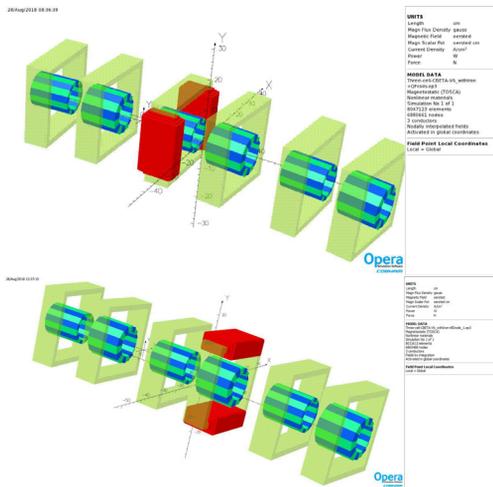


Figure 11: OPERA simulation of the full-cell H and V orbit correction dipoles (iron yoke electromagnets), on top of respectively the F and D Halbach FFAG magnet.

Code sequence for an arc cell, case of single full-cell field maps:

```
'TOSCA' QF+BD map + corrector maps
0 0
-9.69871600E-04 1. 1. 1.
HEADER_8 ZroBXY
451 83 27 15.3 1. 0.001 0.001 ! 3 independent knobs
3D-Cell-fieldMap.table ! FFAG qf-BD doublet
FConDCoff-3D-fieldMap.table ! F corrector
FCoffDCon-3D-fieldMap.table ! D corrector
1 482.028 42.172 -20328 ! integration boundary
2
.2 ! integration step size
2 0.0 0.0 0.0 ! magnet positioning
'CHANGREF' ! magnet positioning:
XS -0.6586 YS -3.2061 ZR -5.0 YS 1.2047
```

Code sequence for an arc cell, case of separate QF, BD and corrector field maps:

```
'DRIFT'
5.6 -18.35
'TOSCA' QF
0 0
-9.69871600E-04 1.00E+00 1.00E+00 1.00E+00
HEADER_8 ZroBXY
501 83 1 15.2 1. 0.
QF-2D-fieldMap.table
FCorr-2D-fieldMap.table
0 0 0 0
2
.1
2 0.00E+00 0.00E+00 0.00E+00
'DRIFT'
-18.35
'DRIFT'
1.2
'CHANGREF' CORNER
ZR -2.50
'DRIFT'
4.2
'CHANGREF' CORNER
ZR -2.50
'DRIFT'
1.2
'DRIFT'
-18.9
'TOSCA'
0 0
-9.69871600E-04 1.00E+00 1.00E+00 1.00E+00
HEADER_8 ZroBXY
501 83 1 15.1 1.0
501 83 1 15.2 1. 0.
BD-2D-fieldMap.table
DCorr-2D-fieldMap.table
0 0 0 0
2
.1
2 0.00E+00 3.60319403E-04 0.00E+00
'DRIFT'
-18.9 + 6.7
```

SX and RX line models in Zgoubi are under construction, replacing the analytical field model of the dipole and quadrupoles by their field map, step by step. Any such change of an optical element causes a slight change in the optical functions, Fig. 12, necessitating a retuning of the orbit and optical functions (textsli.e., re-matching between SX (or RX) and the FFAG arc).

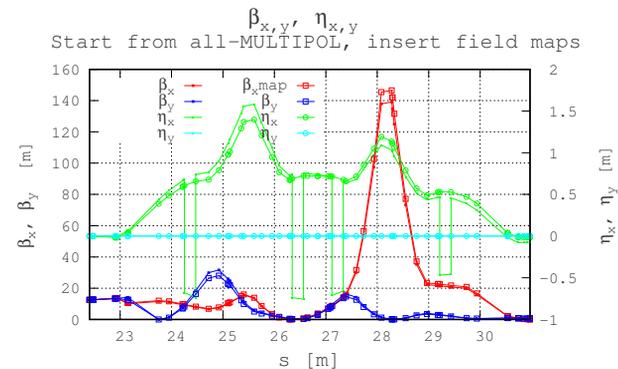


Figure 12: A change in the modeling of an element along the line (SX here, 42 MeV spreader line), from an analytical field model to its OPERA field map, perturbs the optics and necessitates a re-matching of the orbit and optical functions.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

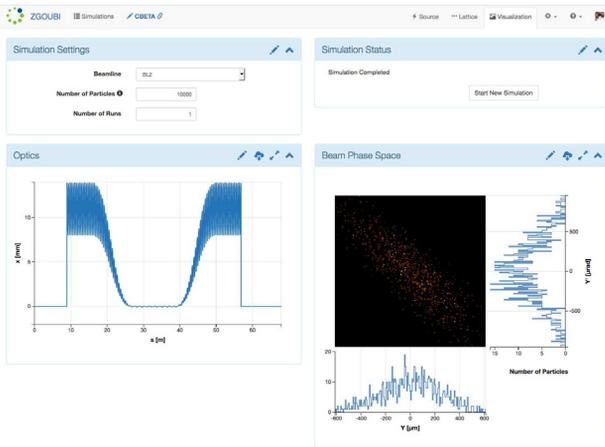


Figure 13: Working on the first turn (42 MeV) in SIREPO environment [12]. The orbit around the loop is shown, here, together with the projection of a bunch in the horizontal phase-space after a turn from linac exit to linac entrance.

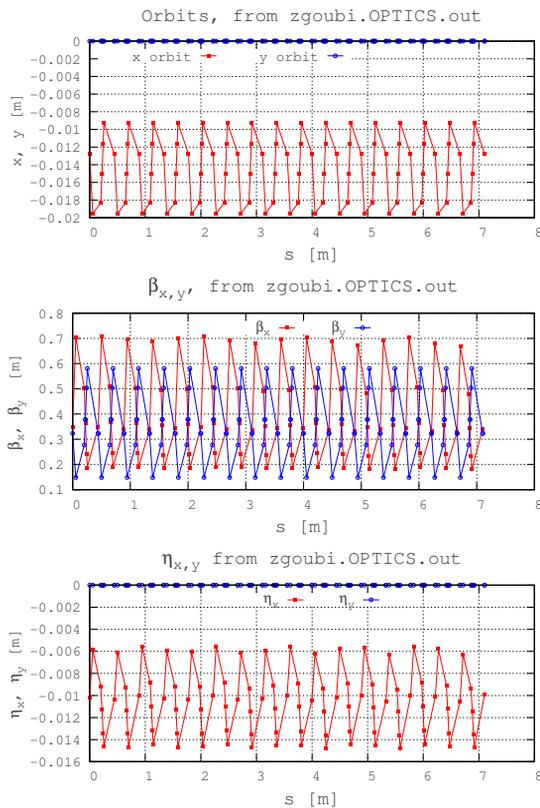


Figure 14: 42 MeV orbit and optical functions in TA arc, observed at a few points along the cell (lines are to guide the eye).

## REFERENCES

- [1] F. Méot *et al.*, “Beam dynamics validation of the Halbach Technology FFAg Cell for Cornell-BNL Energy Recovery Linac”, *NIM A*, vol. 896, pp. 60-67, 2018.
- [2] J. Barley *et al.*, *CBETA Design Report*, Jan 2017.
- [3] F. T. Cole, “O Camelot ! A Memoir of the MURA Years,” April 1, 1994, <http://jacow.org/c01/cyc2001/extra/Cole.pdf>
- [4] See the FFAg workshop series: <https://www.bnl.gov/ffag17/pastWorkshops.php>
- [5] F. Méot, “RACCAM: A status, including magnet prototyping and magnetic measurements”, FFAg 2009 Workshop, Fermilab, Batavia, IL, USA. <https://indico.fnal.gov/event/2672/session/2/contribution/8/material/slides/1.pdf>
- [6] S. Antoine *et al.*, “Principle design of a protontherapy, rapid-cycling, variable energy spiral FFAg,” *NIM A*, vol. 602, pp. 293–305, 2009.
- [7] T. Planche *et al.*, “Design of a prototype gap shaping spiral dipole for a variable energy protontherapy FFAg,” *NIM A*, vol. 604, pp. 435-442, 2009.
- [8] Y. Giboudot and F. Méot, “Optical matching of EMMA cell parameters using field map sets,” in *Proc. PAC’09*, Vancouver, BC, Canada, paper TH5PFP040.
- [9] [https://zgoubi.sourceforge.io/ZGOUBI\\_DOCS/Zgoubi.pdf](https://zgoubi.sourceforge.io/ZGOUBI_DOCS/Zgoubi.pdf)
- [10] <https://www.osti.gov/scitech/biblio/1062013-zgoubi-users-guide>
- [11] D. Abell, “Zgoubi: Recent Developments and Future Plans”, presented at ICAP2018, Key West, FL, USA, paper TUPAF04, this conference.
- [12] P. Moeller, Radiansoft, priv. comm., Oct. 2018. “An interface to accelerator codes” <https://beta.SIREPO.com/#/accel>

# MULTI PASS ENERGY RECOVERY LINAC DESIGN WITH A SINGLE FIXED FIELD MAGNET RETURN LINE\*

D. Trbojevic<sup>†</sup>, J. Scott Berg, Stephen Brooks, Francois Meot, Nicholas Tsoupas  
BNL, Upton, NY, USA

William Lou, Cornell University (CLASSE), Ithaca, NY, USA

## Abstract

We present a new approach of the Energy Recovery Linac Design for the future projects: PERLE (Powerful Energy Recovery Linac for Experiments), LHeC/FCCeH and eRHIC. The concept uses superconducting linacs and a single fixed field beam line with multiple energy passes of electron beams. This represents an update to the existing CBETA (Cornell University Brookhaven National Laboratory ERL Test Accelerator) where the superconducting linac uses a single fixed field magnet beam line with four energy passes during acceleration and four passes during the energy recovery. To match the single fixed field beam line to the linac the CBETA uses the spreaders and combiners on both sides of the linac, while the new concept eliminates them. The arc cells from the single fixed field beam line are connected to the linac with adiabatic transition arcs where cells increase in length. The orbits of different energies merge into a single orbit through the interleaved linac within the straight sections as in the CBETA project. The betatron functions from the arcs are matched to the linac. The time of flight of different electron energies is corrected for the central orbits by additional correction magnet controlled induced beam oscillations.

## INTRODUCTION

The Energy Recovery Linacs (ERLs) and Recirculating Linacs (RAL) are considered to be a part of the future Electron Ion Colliders in several world programs: LHeC (CERN) [1], FCC eh [2], eRHIC(BNL) [3], ELIC (Jefferson Lab) [4] and EIC@HIAF (China) [5]. A proposal presented in this report describes a solution of ERL where the electron beam is brought back to the linac with a single large energy acceptance beam line using a concept of linear fixed field alternating (FFA) gradient. The concept of the FFA beam transport with large momentum acceptance is not a novelty. There are three experimentally confirmed proof-of-principles: EMMA-Electron Model for Many Applications [6], ATF (Accelerator Test Facility) [7] experiment with 12 FFA cells, and Fractional Arc Test of the Cornell University-Brookhaven National Laboratory ERL Test Accelerator-CBETA [8]. A comparison of these three examples are compiled and shown by Stephen Brooks [9] in Fig. 1 and in Table 1.

We present a new concept of the ERLs where the large momentum acceptance linear FFA magnet beam lines bring

\* Work supported by Work performed under Contract Number DE-AC02-98CH10886

<sup>†</sup> dejan@bnl.gov

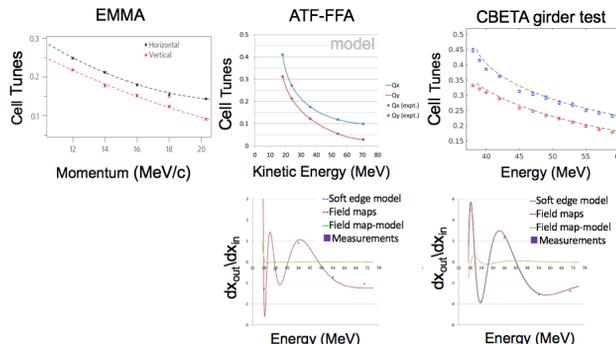


Figure 1: Comparisons on tune dependence on Energy in the three FFA examples.

the electron beam back to the superconducting linac without spreaders and combiners. In multiple passes through the ERLs the acceleration of electrons generated by the linac is too fast to consider any return beam line by using fast cycling magnets. With the large momentum acceptance linear FFA gradient magnets this can be achieved. The first test of the concept will be achieved in 2019 by the ERL CBETA at Cornell University. There are many advantages and simplifications: 1) the replacement of multiple returning beam lines with a single beam line reduces the cost and complications; 2) established technology of the Halbach type permanent magnets used at CBETA project will confirm the reduction of cost and simplification of the beam lines; the magnet aperture remains to be very small while transporting multiple energy beams. In the present study the superconducting linac is made of 5-6 cavity cells, each separated with small permanent FFA triplets. The same type of triplet cells con-

Table 1: EMMA-ATF-CBETA Comparisons

Parameter	EMMA	ATF-FFA	CBETA girder	CBETA Future
Energy (MeV)	10–20	18–71	37.5–59	41–150
Mom. ratio	1.953	3.837	1.553	3.572
$\rho$ of curv. (m)	2.637	2.014	5.0879	5.0879
Avrg. dip. (T)	0.026	0.118	0.0983	0.0983
Total angle (°)	360.0	40.0	20.0	280.0
Oper. mode	ring	Tr. line	Tr. Line	ERL
Acceleration?	YES	none	none	linac
Lattice	Doublet	~FODO	Doublet	Doublet
Cell Length (m)	0.395	0.234	0.444	0.444
Cell per turn	42	54	72	107.5
Length (m)	16.57	1.406	1.776	47.73

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

tinues towards the FFA arc but with an adiabatic reduction of the cell lengths. The betatron and dispersion functions for each energy are adiabatically matched to the corresponding functions of the arc FFA cells.

Recently the National Academy of Sciences released a study of “An Assessment of U.S.-Based Electron Ion-Collider Science.” ... “The principal goals of the study were to evaluate the significance of the science that would be enabled by the construction of an EIC, its benefits to U.S. leadership in nuclear physics, and the benefits to other fields of science of a U.S.-based EIC.” ... “Several presentations to the committee specifically addressed the challenges and **necessary innovations** in accelerator science needed for constructing an EIC capable of addressing the most important science questions” ... “To reach the performance goals of the proposed EIC conceptual designs, a number of accelerator advances are required. Several of these advances are common to all EIC designs and include the following: advanced magnet designs, strong hadron beam cooling, **high current multi turn ERL technology**, crab cavity operation with hadron beams, the generation of polarized 3He beams, and development and benchmarking of simulation tools. The successful implementation of an EIC requires the successful validation of these key concepts through high-fidelity simulations and demonstration experiments. The following subsections review these enabling technologies, the present state of the art, and required research and development to meet EIC facility specifications and realize EIC science: **Energy Recovery Linacs.**” ... “The ERLs required for electron cooling are at scales much larger than supported by present-day experience, so a number of accelerator physics and technology challenges still need to be overcome with focused R&D and great attention to detailed simulations. The challenges center around the following three major areas: 1. Achieving high electron source brightness 2. Maintaining high beam brightness through the accelerator transport-beam dynamics of an unprecedented number of spatially superposed bunches in the SRF linacs; very precise phase and amplitude control 3. Dealing with unprecedented beam currents in SRF linacs (halo mitigation, beam breakup instabilities, higher order mode dissipation). Many of these R&D issues are being investigated vigorously in dedicated test facilities under construction and commissioning in laboratories around the world. Specifically, the 4-pass Fixed Field Alternating Gradient R&D loop for eRHIC, see Box 4.2 **CBETA and Fixed Field Alternating Gradient Optics** for Electron Acceleration, it could illuminate key issues including multi-turn beam-breakup instability thresholds for proof of possible cavity designs, halo and mitigation, beam-ion effects, and operational challenges such as instrumentation and stability of multi-turn beams”. The suggestions obtained in the National Academy of Sciences report emphasize the importance of the CBETA project and further developments towards improvements of the concept.

The presented study is divided into couple of sections: first section describes the arc cell design to allow transport of multiple energy electrons; the second section describes a

design of the long straight section to be used for the placement of the linac superconducting cavities. Acceleration of electrons through the linac requires use of the “normalized to betatron functions” as the momentum changes along the length of the linac. To make the matching conditions on both sides of the linac the triplet quadrupoles strength along the linac needs to be adjusted accordingly. The adiabatic transition from the arc FFA cells towards the linac is explained in the next section. Overall solution with orbits, betatron functions, dispersion function is shown in the fourth section. The time of flight adjustment is described in the fifth section while the summary and conclusions are shown in the sixth section.

## LINEAR FIXED FIELD ALTERNATING (FFA) GRADIENT ARCS

The principle of the linear FFA gradient acceleration has been previously described in details [10] and [11]. The main idea of the principle is to keep the value of dispersion function  $\Delta x = D_x \cdot \delta p/p$  as low as possible or to have a control of the 'dispersion action'  $\mathcal{H} = (D_x/\sqrt{\beta})^2 + (D'_x\sqrt{\beta} + \alpha D_x/\sqrt{\beta})^2$ . The accelerator physics program 'Bmad' [12] is used throughout this study. The main arc FFA cell design is shown in Fig. 2. Figure 3 shows few arc cells with betatron functions and orbits.

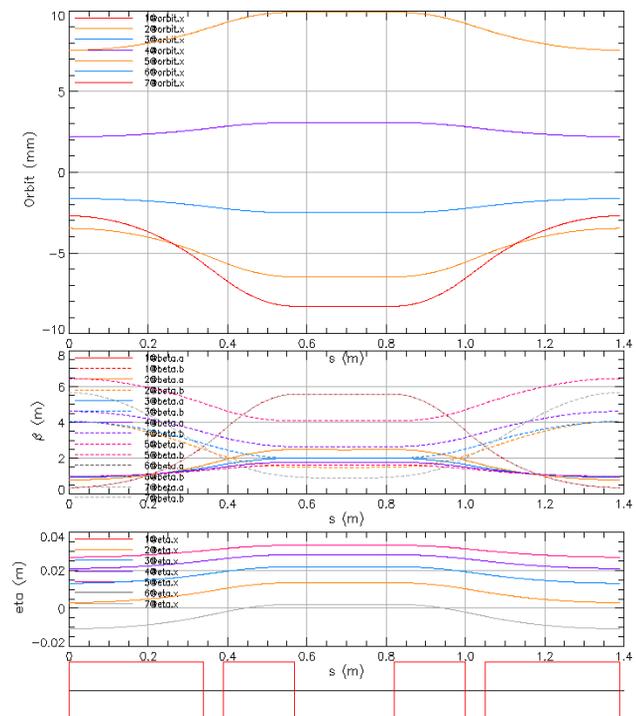


Figure 2: Lattice functions and orbits in the linear FFA arc cell.

## STRAIGHT SECTION

The linear FFA racetrack design has two straight sections on opposites sides connected to the two arcs with the adi-

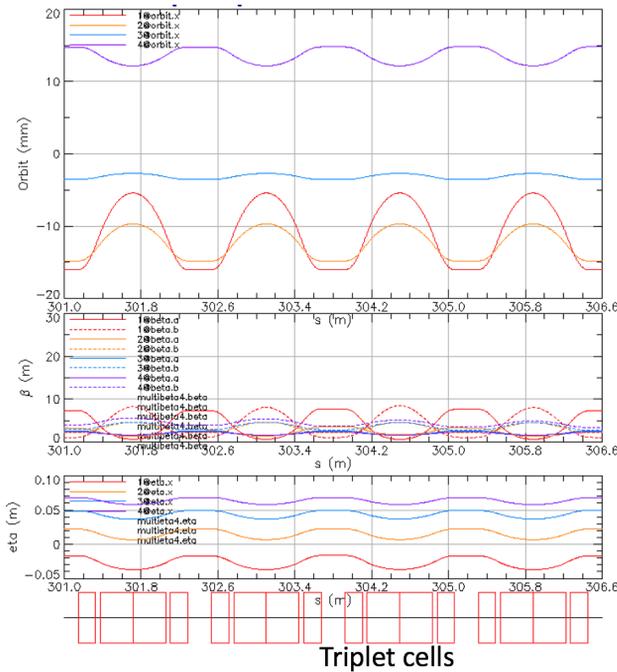


Figure 3: Lattice functions and orbits in few basic linear FFA arc cells.

adiabatic transitions. A problem of the RF straight section is solved in two steps: First, the straight section is designed with 42 cells made of the triplet quadrupoles and drifts without RF cavities matched to the FFA gradient arcs with adiabatic transitions. The second step was done by retuning the 42 triplets with the 1.7 m superconducting 5 cell cavities placed with 3.2-meter long cells, making the total length of each straight section equal to 134.2 m. The accelerator physics code Bmad has two ways to present the acceleration through the linac: `lcavity` and `rfcavity` where the `lcavity` represents the accelerating cavity without constant reference energy, while the `rfcavity` is the storage ring cavity with constant reference energy without acceleration. The transverse trajectory in Bmad through an `lcavity` is modeled using equations developed by Rosenzweig and Serafini [13] modified to give the correct phase-space area at non ultra-relativistic energies. First step in the the ERL design is to create the straight section with long enough drifts between the triplet quadrupoles to allow placement of the superconducting RF 5 cell modules. The lattice functions of the straight section cell are shown in Fig. 4. The adiabatic transition from the arc FFA cell to the straight section is shown in details in the next section. The second step in the straight section design is to introduce the 1.7 m long cavity modules of in the available 2.1 m long drift space. The 42 triplet quadrupoles are the variables with constraints related to the 6 different energies: the first energy is the lowest energy from the injector getting into the straight section. All other betatron function of initial and end values of the linac are already obtained in the results from the step one described above. The triplet quadrupole matching for the

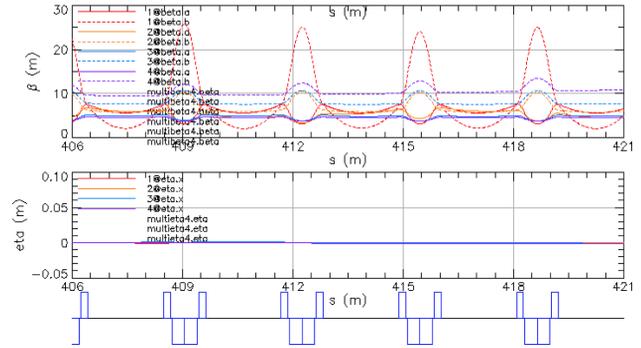


Figure 4: Lattice functions and orbits in the straight section cells.

six energies, or in the Bmad program corresponding six “universes,” is possible as there are 12 constraints on both sides of the linac of the  $\beta_x, \beta_y, \alpha_x,$  and  $\alpha_y$  with  $42 \times 3$  variables of the triplet gradients.

### ADIABATIC TRANSITION

A transition from the linear FFA arc cells to the straight section requires that electron orbits of all energies merge into a single orbit without any orbit offsets. In addition the lattice and dispersion functions are matched on both sides. This is accomplished with adiabatic reduction of dipole field down to zero. The combined function magnet properties remain the same but the bending magnetic field is being adiabatically reduced to zero. A problem of merging all arc orbits into a single straight section orbit with the lattice functions was solved previously in the former electron ion collider eRHIC design. The adiabatic function was a polynomial of the third order. The CBETA project follows S. Berg [14] optimized adiabatic dependence as shown in Eq. (1):

$$f_T(x) = \frac{1}{2} + \left(x - \frac{1}{2}\right) = \sum_{k=0}^{2} a_k \binom{2k}{k} x^k (1-x)^k. \quad (1)$$

The adiabatic transition in this design is different as the length of the cells is adiabatically increased from the 1.3889 m FFA arc value to the 3.2 m. Details of the adiabatic transition are shown in Fig. 5. The superconducting cavity placement in the straight section is shown in Fig. 6.

### RACETRACK ERL WITH LINEAR FFA

The complete layout of the ERL racetrack with the betatron functions in shown in Fig. 7.

The orbit merging in the racetrack from the arc cells through the adiabatic transition cells is shown in Fig. 8.

### TIME OF FLIGHT CORRECTION

The most important parameter of the ERL is the electron bunch arrival to the linac as the acceleration is at the top of the sinusoidal RF function. The energy recovery requires a change of the RF phase so the bunches arrive at the bottom of the sinusoidal RF function. This phase flip occurs after

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

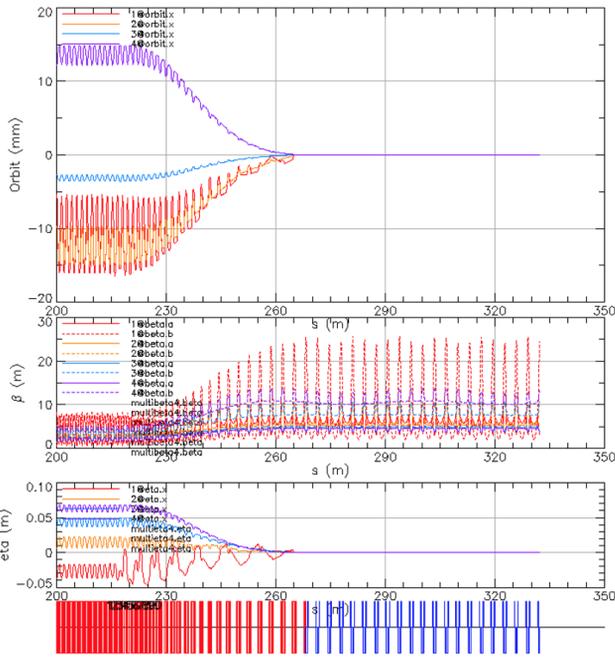


Figure 5: Lattice functions and orbits in the adiabatic transition and straight section cells.

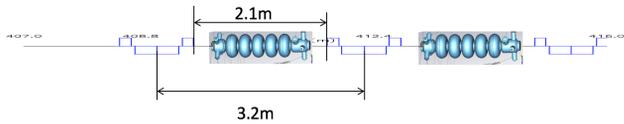


Figure 6: The superconductivity RF modules placement in the straight section.

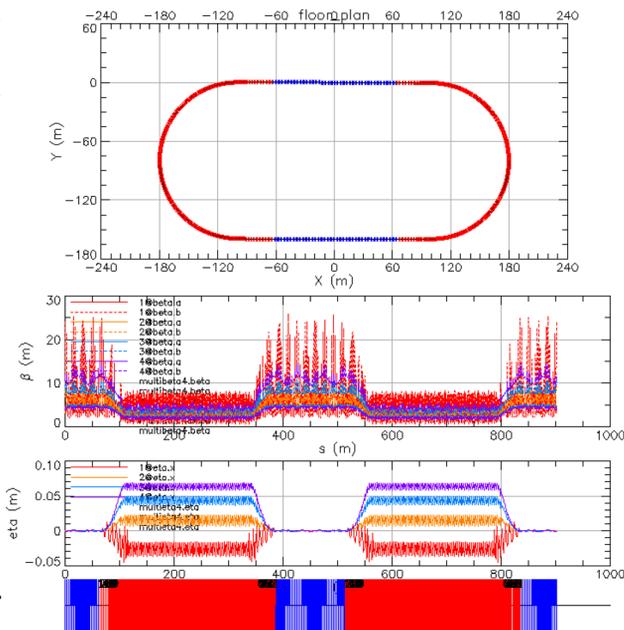


Figure 7: Lattice functions and orbits in the basic straight section cell.

the collisions. The electrons are extracted from the adiabatic transition part brought to the collisions with a separate beam

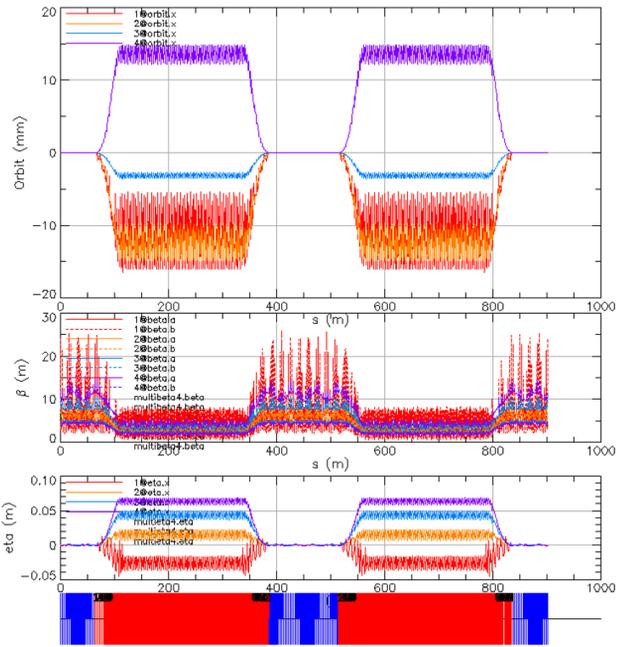


Figure 8: Orbits and dispersion merging from the arc cells to the straight section for different energies in the racetrack.

line to ions and brought back by the separate beam line with an 0.5 phase difference at the highest energy. The FFA gradient arcs have a parabolic function with respect to the time of flight. The lowest energy has the same value of the time of flight as the highest energy while the medium energies correspond to the minimum of the parabolic function. The orbits in the FFA arc cells oscillate around the middle “central” energy as shown in Fig. 3. The main idea of the time of flight correction is that the central orbit can become longer if additional oscillations are introduced to the mostly circular orbit. This was first tested by using two correction dipoles with opposite kicks. This produced the same time of flight after the first attempt. More sophisticated method of time of flight correction was developed by Stephen Brooks and soon will be shown in different publication. The main idea is to use dependence of FFA arc cell of the tunes with respect to energy and choose correctors at positions where the betatron phases are equal to zero and make additional oscillation to the orbits. Results of the path length corrections are shown in Fig. 9.

## CONCLUSION

We have shown an example of ERL with the multi-turn single beam line returning the beam to the linac. The arc FFA gradient cells are made of triplet combined function magnets with a 1.3889 m long cells. The arc FFA gradient cells are matched to the straight sections with adiabatic transition cells where the length of the cells is increasing gradually up to 3.2 m, according to the to the function shown in Eq. (1). The matching was done in two steps: the first step is to match the straight section cells made of triplet magnets separated by the drifts. The second step in matching is to introduce in

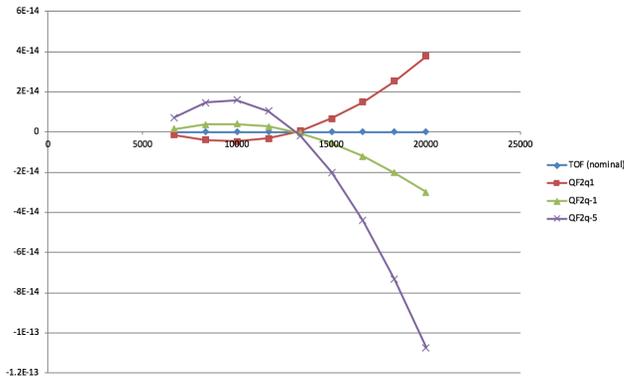


Figure 9: Stephen Brooks method of time of flight correction.

one of the straight sections 1.7 m long superconducting RF modules and then redo the matching of the betatron functions. The beginning and end of the straight section corresponds to the two values of energy where after each pass energy steps are equal to 1.25 GeV. The previously found initial conditions correspond to specific energy. There are many advantages of this proposal especially for the future Electron Ion Colliders: the previous splitter and combiner beam lines used to match the linac without any quadrupoles where necessary to the time of flight correction and betatron function matching. They are eliminated by introduction of the triplet magnets between the RF superconducting modules.

## REFERENCES

[1] LHeC Study Group, J. L. Abelleira Fernandez *et al.* "A Large Hadron Electron Collider at CERN", *Journal of Physics G: Nuclear and Particle Physics*, vol. 39, no. 7, July 2012, pp. 075001. doi:10.1088/0954-3899/39/7/075001

[2] FCC eh: <http://lhec.web.cern.ch>

[3] eRHIC Design Study, "An Electron Ion Collider at BNL", arXiv:1409.1633, 2014.

[4] ELIC: <https://casa.jlab.org/overview.shtml>

[5] EIC@HIAF, <http://snst-hu.lzu.edu.cn/wiki/index.php/Eic>

[6] S. Machida *et al.*, "Acceleration in the linear non-scaling fixed-field alternating-gradient accelerator EMMA", *Nature Physics*, Jan 2012. doi:10.1038/nphys2179

[7] R. C. Johnson (ed.), "Small Accelerator, Big Promise", [https://www.eetimes.com/document.asp?doc\\_id=1332173](https://www.eetimes.com/document.asp?doc_id=1332173)

[8] CBETA, [https://www.classe.cornell.edu/CBETA\\_PM/](https://www.classe.cornell.edu/CBETA_PM/)

[9] Stephen Brooks, "CBETA Magnet Production", presented at FFA'18 in Kyoto, Japan, Sep 2018. <https://indico.rcnp.osaka-u.ac.jp/event/1143/contributions/1213/>

[10] D. Trbojevic, E. D. Courant, and A. Garren, HEMC'99 Workshop, Montauk, New York, October 1999.

[11] D. Trbojevic, E. D. Courant, and M. Blaskiewicz, *Phys. Rev. Spec. Topics Accelerators and Beams*, vol. 8, pp. 050101, 2005.

[12] D. Sagan, *Bmad manual*, <https://www.classe.cornell.edu/bmad/manual.html>

[13] J. Rosenzweig and L. Serafini, "Transverse particle motion in radio-frequency linear accelerators", *Phys Rev E*, vol. 49, p. 1599, 1994. doi:10.1103/PhysRevE.49.1599

[14] J. S. Berg, <https://indico.bnl.gov/event/2764/contributions/7146/attachments/6425/7812/JSBerg-170909.pdf>

# EXPERIENCE WITH CBETA ONLINE MODELING TOOLS

C. Gulliford, D. Sagan, A. Bartnik, J. Dobbins, Cornell University, Ithaca, NY, USA  
J. S. Berg, BNL, Upton, Long Island, New York, USA,  
Antonett Nunez-delPrado, Department of Physics, University of Central Florida

## Abstract

The Cornell-Brookhaven CBETA machine is a four pass Energy Recovery Linac (ERL) with a Non-scaling Fixed-Field Alternating gradient (NS-FFA) arc. For online modeling of single particle dynamics in CBETA, a customized version of the Tao program, which is based upon the Bmad toolkit, has been developed. This online program, called CBETA-V, is interfaced to CBETA's EPICS control system. This paper describes the online modeling system and initial experience during machine running.

## INTRODUCTION

The Cornell-Brookhaven Energy recovery linac Test Accelerator (CBETA) [1], currently under construction at Cornell University, is a 4-pass, 150 MeV Energy Recovery Linac (ERL), utilizing a Non-Scaling Fixed Field Alternating-gradient (NS-FFA) permanent magnet return loop. CBETA is a joint collaboration of Brookhaven National Laboratory (BNL) and the Cornell Laboratory for Accelerator based Sciences and Education (CLASSE).

The CBETA project builds on the significant advancements in high-brightness photoelectron sources and Superconducting RF (SRF) technology developed at Cornell [2–5], as well as the FFA magnet and lattice design expertise from BNL. One aim of CBETA is to establish the operation of a multi-turn SRF based ERL utilizing a compact FFA return loop with large energy acceptance (a factor of roughly 3.6 in energy), and thus demonstrate the feasibility of one possible cost-reduction technology under consideration for the eRHIC Electron-Ion Collider (EIC) currently being designed at BNL. The CBETA project involves the study and measurement of many critical phenomena relevant to proposed EIC machine designs: the Beam-Breakup (BBU) instability, halo-development and collimation, growth in energy spread from Coherent Synchrotron Radiation (CSR), and CSR mi-

cro bunching. The CBETA project should provide valuable insight for both the EIC and ERL communities [1].

As part of the CBETA commissioning sequence, a combined test of the elements of all of the critical subsystems required for the CBETA project was done in the spring of 2018. This “Fractional Arc Test” (FAT) involved the injector, the Main Linac Cryomodule (MLC), the low energy splitter line, and a first prototype production permanent magnet girder featuring four cells of the FFA return loop (see Fig. 1). Besides hardware, the FAT commissioning involved testing and benchmarking of CBETA-V, the CBETA online model. This paper describes the online modeling system and initial experience during machine running.

## ONLINE SIMULATION ENVIRONMENT

The online single particle dynamics simulation model CBETA-V is based upon Bmad [6] and Tao [7]. Bmad is a modular, object-oriented subroutine library for simulating charged particle beams in high-energy accelerators and storage rings. Tao is a general purpose design and simulation program based upon Bmad and includes several optimization routines allowing the user to correct orbits, fit measured data, etc.

There were a number of advantages to basing CBETA-V on Bmad and Tao. For one, the majority of the CBETA lattice design was done using Bmad and Tao. This, and the fact that any Bmad based program is able to read Bmad lattice files, meant that offline and online simulations could be seamlessly interfaced. Additionally, the modular nature of both Bmad and Tao meant that CBETA-V development essentially consisted of creating a custom version of Tao which had the ability to communicate with the CBETA online EPICS database [8]. This was a relatively simple procedure requiring development of only about 1500 lines of code, and resulted in an online program which had all the capabilities of Tao. The ease of which CBETA-V was implemented is due in no small part to the fact that Bmad was originally developed for use with online modeling. Chan-

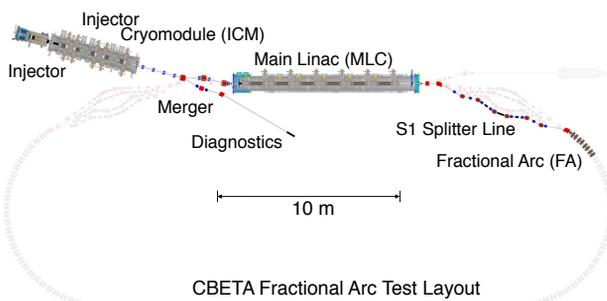


Figure 1: Schematic of the CBETA machine highlighting the components installed for Fractional Arc Test.

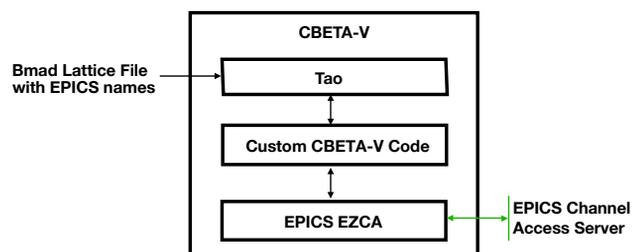


Figure 2: Schematic of the CBETA-V application showing the linking of the Tao with the EPICS EZCA library.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

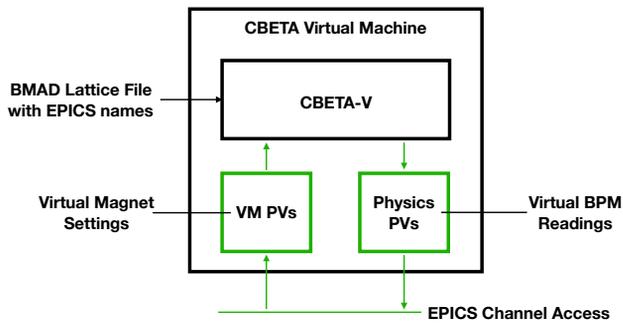


Figure 3: Schematic of the CBETA Virtual Machine application showing the communication between an outside EPICS user and CBETA-V via the CBETA Virtual Machine “virtual” control system.

nel Access between CBETA-V and the EPICS database is achieved via the EZCA [8] C/Fortran interface library as illustrated in Fig. 2. Additionally, CBETA-V (and Tao) can be interfaced to Python using the Python foreign function library ctypes or the Python pexpect module. The Bmad lattice files for CBETA have EPICS Process Variable (PV) information attached to corresponding physical elements which allows CBETA-V to translate between EPICS PV values and the CBETA machine state. The calibration constants needed to convert machine readback quantities (such as magnet currents) into field strengths (such as quadrupole focusing strengths) are incorporated into the EPICS database so that CBETA-V works independently of any calibration constants.

In order to simulate space charge effects in the low energy part of CBETA, the General Particle Tracer (GPT) program [9] is used for simulations from the cathode through the first pass of the main linac. MATLAB was interfaced to GPT to facilitate communication between GPT and EPICS. Features of this MATLAB/GPT program include the ability to save and load optics settings and simulation results to and from a file, the ability to load injector settings from the machine and independently adjust them in simulation, as well as the ability to visualize all relevant simulation data [3].

Building on CBETA-V, a second application, called the CBETA Virtual Machine (CBETA-VM), was designed. This software creates its own copy of the CBETA EPICS control system, allowing users to command the virtual optical elements in the simulation via standard EPICS commands. Figure 3 shows a schematic of this application. By changing any of the element strengths in the model, the software computes all relevant single particle tracking parameters (that is, centroid orbit, dispersion, transfer matrix, etc), and publishes the results to its own EPICS records, thus making the virtual machine data available to the user via EPICS in exactly the same manner as real machine data. This allows for any automated measurement procedure to command and take data from either/both the real or virtual machine. This provides the ability to easily produce simulated predictions of measurement results, as well as the ability to realistically

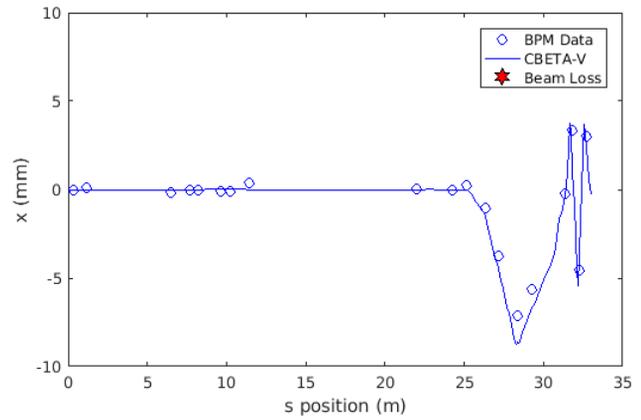


Figure 4: Real time comparison of an orbit bump. Circles indicate data read from the CBETA BPMs, while the blue line indicates the simulated values.

debug automated measurement procedures. Many of the experimental procedures used in CBETA were developed and tested offline in this manner. The software also provides a “sync” mode where the Virtual Machine continuously monitors the status of the real CBETA settings, and updates the simulated machine data upon detecting a change in the settings of the real machine, thus providing a useful online diagnostic.

## REAL TIME COMPARISON WITH MEASUREMENTS

By serving simulated physics data from CBETA-V via EPICS records, the CBETA-VM enables any client application which can read EPICS PVs capable of accessing simulation data by reading the corresponding virtual EPICS records. For machine data that can be read out continuously, such as BPM data, this makes visual comparison of the measured and simulated data straightforward. Figure 4 shows an example snap shot of an orbit bump generated in the S1 splitter line. With CBETA-VM in sync mode, changing the S1 corrector responsible for the bump produced a simulated curve and measured data points that closely moved together.

In general, data (such as orbit data) from the real machine is limited to a finite number of locations throughout the machine. This makes the real time inclusion of simulated smooth data curves in the same displays as the real data useful as it “completes” the real data by filling in the values in between diagnostic points. In addition to this, once the accuracy of the CBETA-VM was established, it provided accurate simulated information for which there is no corresponding read out that is continuous in time. Quantities useful in operation include the beam energy, dispersion, BPM time of arrival through the FAT lattice,  $R_{56}$  transfer matrix element.

## OFFLINE MEASUREMENT ANALYSIS

Some important quantities are not directly measured but must be calculated from direct measurements. For example, to measure the dispersion, the voltage of the last MLC cavity

Content from this work may be used under the terms of the CC BY 3.0 licence © 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

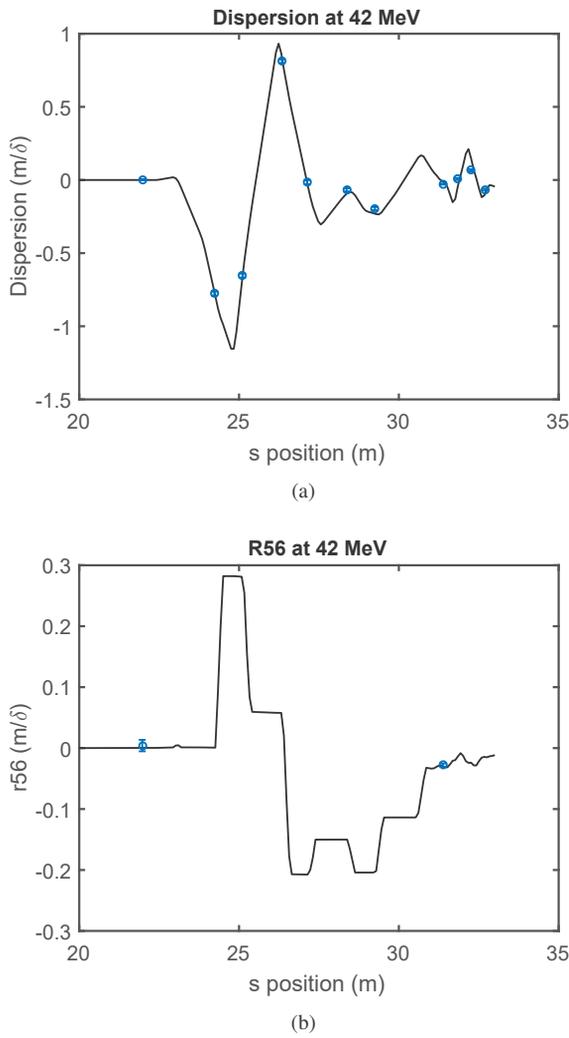


Figure 5: Comparison of the measured (blue circles) and simulated (black line) data. Measurements were made at the nominal 42 MeV energy. (a) Dispersion measurement, and (b) The time of arrival derivative with respect to energy,  $R_{56}$ . The two data points correspond to the first BPM after the MLC and the first FA BPM in the fractional arc.

was scanned and the resulting orbit position changes measured downstream. The dispersion was computed by fitting curves to the position changes on each BPM and extracting the linear variation around the nominal voltage set-point. Since such quantities as the dispersion require some amount of time to take data, the analysis essentially has to be done “offline”.

Figure 5(a) shows measured dispersion data along with the simulated prediction at the nominal S1 energy of 42 MeV. In addition to recording the BPM positions downstream of the linac during the dispersion measurement, the procedure also saved the BPM time of arrival  $\phi$  (in units of the RF phase) on the first BPM after the MLC as well as the first FA BPM in the fractional arc. This allows for the determination of the  $R_{56}$  matrix element through the splitter line. Figure 5(b) displays the resulting measured and simulated  $R_{56}$ , defined as  $R_{56} = (c/\omega)d\phi/d\delta$  where  $\delta$  denotes a relative

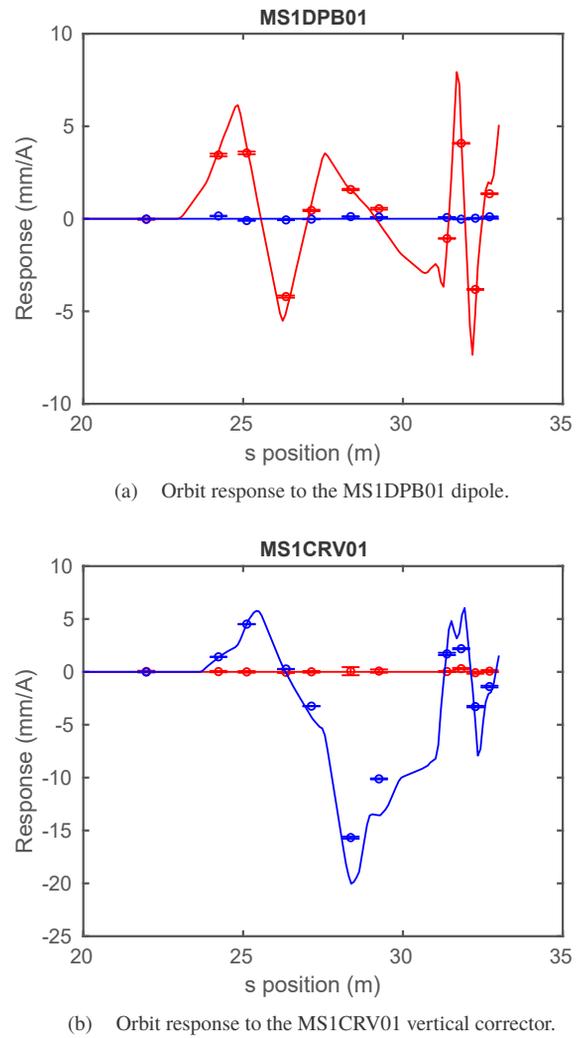


Figure 6: Measured orbit response (points) compared to simulation (lines). Horizontal response is shown in red, and vertical in blue. (a) Response to the variation of the first S1 splitter dipole. (b) Response to the variation of the first S1 splitter vertical corrector.

energy change. The excellent agreement between simulation and measurement seen in both the dispersion and  $R_{56}$  data required adjusting the simulated S1 quadrupole settings by 1%. Similar agreement were found for the comparison of the measured dispersion and  $R_{56}$  at beam energies ranging from 38.5 to 59 MeV.

In addition to the above results, the orbit response matrix was also measured during FAT commissioning. To do so, every corrector and dipole magnet in the machine was scanned and the resulting orbit position changes recorded. The slope of the fit of position versus magnet strength at the nominal magnet settings gives the response matrix. Figures 6(a) and 6(b) show the BPM response to the variation of two S1 magnets downstream of the MLC. The measured data and simulated responses agree well, especially for the horizontal orbit.

## TESTING ORBIT CORRECTION PROCEDURES OFFLINE

Orbit correction studies for the CBETA machine up until FAT commissioning were carried out offline from the EPICS control system/real machine. Unfortunately, the tight schedule of the Fractional Arc Test did not allow for significant tests of orbit correction software during the experiment. Consequently, the CBETA-VM was used to perform virtual orbit correction experiments originally planned for during commissioning [10]. Orbit corrections were done using Singular Value Decomposition (SVD) since that technique had proved robust in the past.

One test involved the viability of orbit correction in the injector section following the ICM and before the MLC. The test proceeded as follows: first the orbit response matrix  $R$  is computed in the lattice model. Random errors to the quadrupole calibrations in the beamline were assigned by scaling the corresponding quadrupole currents using a 25% RMS normal distribution. In addition, random quadrupole offsets with an RMS spread of 1 mm in both the horizontal and vertical planes were introduced. Finally, the correctors in the sections of the machine between the ICM and MLC were randomly set in order to produce initial “uncorrected” orbits. Fig 7(a) shows 100 examples of these uncorrected orbits. In this plot (as well Fig. 7(b)), the red dots indicate BPM readings.

These orbits were then corrected using SVD, with a singular value tolerance of 0.4345 used for finding the pseudo-inverse of  $R$ . Because  $R$  was computed for the nominal machine state (on-axis orbit, no quadrupole scaling errors or misalignments), the correction algorithm was iterated 10 times. Figure 7(b) shows the resulting orbits which have been reduced in scale by a factor of roughly 100.

To further quantify the residual orbit, the RMS of the virtual BPM readings,

$$\sigma_{Res} = \sqrt{\frac{1}{2N} \sum_i (x_i^2 + y_i^2)}, \quad (1)$$

after each iteration of the SVD procedure was calculated and shown in Fig. 8. Here  $i$  runs over all the BPMs in the FAT layout. From this we conclude that the SVD algorithm produces sub-micron residual orbit error within a few (roughly 3-4) iterations.

In addition to this test, the steering of the beam onto the periodic orbit in the FA section of the FAT layout was also solved using SVD. When the beam is steered onto the periodic orbit, the FA BPMs should read the same value horizontally (vertically the orbit should be zero). Before correction, denote the horizontal positions on the FA BPMs by  $\mathbf{x}$ , and denote the desired periodic orbit by  $\mathbf{C} = C \cdot (1, 1, \dots, 1)^T$  where the constant  $C$  is the as yet unknown periodic offset. The matrix problem for finding the change in corrector currents  $\Delta \mathbf{I}$  that will give periodic orbit readings

can be written as:

$$\begin{pmatrix} R_{11} & R_{12} & \dots & R_{1N} & 1 \\ R_{21} & R_{22} & & R_{2N} & 1 \\ R_{31} & & & R_{3N} & 1 \\ \vdots & & & \vdots & \vdots \\ R_{M1} & & & R_{MN} & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta I_1 \\ \Delta I_2 \\ \vdots \\ \Delta I_N \\ -C \end{pmatrix} = - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_M \end{pmatrix} \quad (2)$$

where the  $R_{ij}$  are the elements of the corrector to BPM response matrix  $R$ . Inverting this equation using SVD allows for the determination of both the corrector currents  $\mathbf{I}$  and periodic BPM reading  $C$  simultaneously.

To test this, the response matrix from the last two dipole magnets in the S1 splitter line to the four BPMs in the FA section was computed, and the matrix in Eq. (2) formed. Using the fact that the periodic orbit solution in the FA section is a function of energy, ten non-periodic orbits were constructed by scanning the beam energy from 39 to 59 MeV (corresponding to the energy ranged demonstrated in the FAT). The SVD steering algorithm was then applied and iterated ten times at each energy. Figure 9(a) shows the ten different non-periodic initial orbits.

The results of the SVD operation are shown in Fig. 9(b) and show that the periodic orbit has been found. To get a sense of the level at which the orbit is periodic on the FA BPMs, the error in the periodicity of the orbit is defined as

$$\sigma_{Res} = \sqrt{\frac{1}{2N} \sum_{i,j} (x_i - x_j)^2}. \quad (3)$$

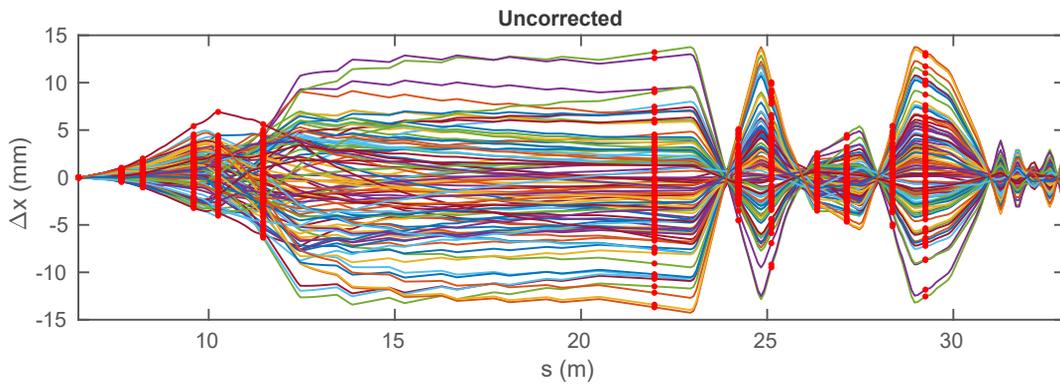
where  $i$  and  $j$  run over the four BPMs in the FFA fractional arc. Clearly this quantity vanishes when the orbit is periodic. Figure 10 shows the residuals found as a function of SVD iteration for the ten example energies shown in Figs. 9(a) and 9(b). Note that for some of the energies in this example, the graph of the RMS orbit residual only extends to about three to five iterations. In these examples, applying SVD algorithm resulted in a perfectly periodic orbit and thus the RMS orbit residual is exactly zero (which is not shown on a log scale plot). As can be seen, the SVD technique works well in both examples and provides an important proof of principle for the online procedure.

## CONCLUSION

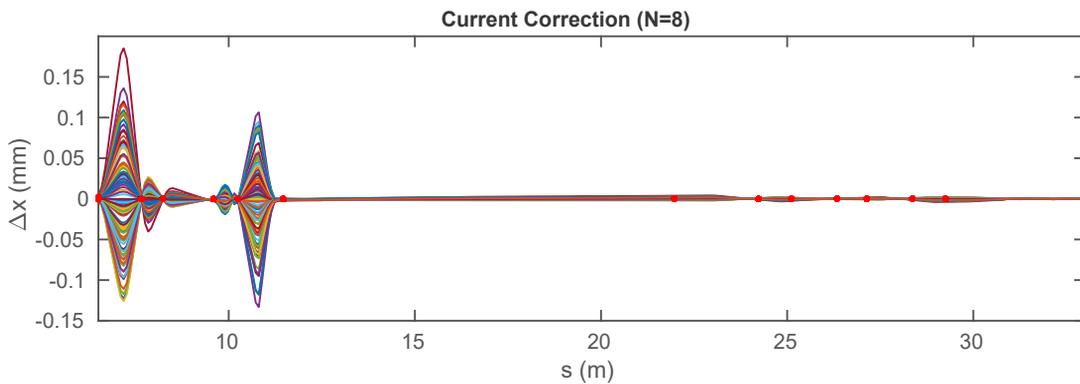
The flexibility built into the Bmad toolkit and the Tao program allowed for the timely construction of both online and offline models for CBETA and results from the FAT experiment verified the rationality of the design approach as well as the usefulness of both the CBETA-V online model and virtual machine. The flexibility of the design allows modifications to be added as needed relatively simply.

With this in mind, work on including a single pass CBETA lattice into CBETA-V has begun which will allow for additional offline testing of more complicated orbit correction and steering algorithms using the CBETA-VM prior to the next beam commissioning period. In addition to this work, significant effort is under way to fully take advantage of the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.



(a) Uncorrected orbits.



(b) Corrected orbits.

Figure 7: (a) Uncorrected orbits and (b) SVD corrected orbits. The red dots indicate BPM position/readings.

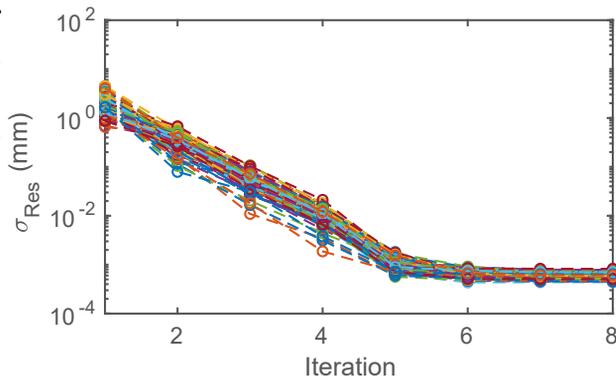


Figure 8: Correction error of the 100 simulated orbits as a function of SVD iteration.

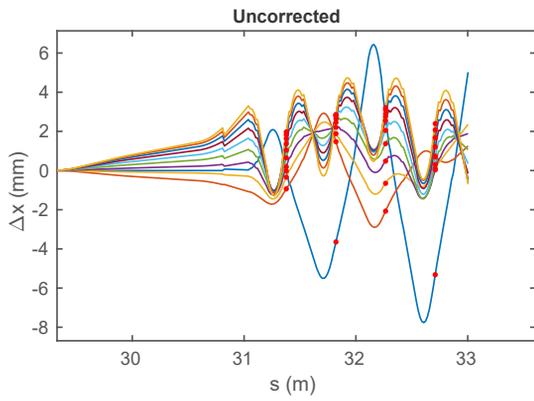
flexibility of the Python language in restructuring the virtual machine wrapper. The purpose of this future work is to generalize and modularize the virtual machine layer so that users can add new physics tasks to the software at runtime, as well as allowing users to plug in different accelerator physics codes. Tests are underway of the latest version, with the hopes of this being ready for the next CBETA commissioning period as well.

### ACKNOWLEDGEMENT

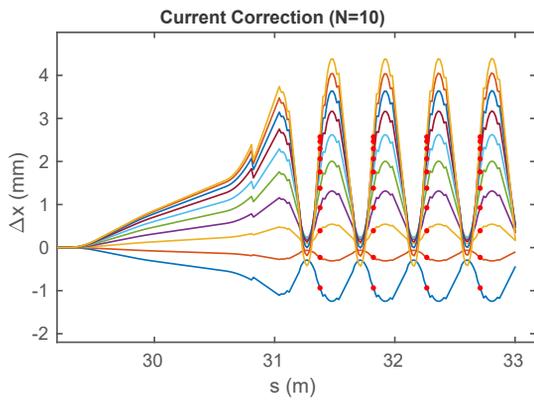
This work was funded by NYSERDA, the New York State Energy Research and Development Agency.

### REFERENCES

- [1] G. H. Hoffstaetter, *et al.*, *CBETA Design Report*, Cornell-BNL ERL Test Accelerator, ArXiv e-prints, June 2017.
- [2] B. Dunham, *et al.*, “Record High-average Current from a High-brightness Photoinjector”, *Applied Physics Letters*, vol. 102, no. 3, pp. 034105, 2013.
- [3] C. Gulliford, *et al.*, “Demonstration of Low Emittance in the Cornell Energy Recovery Linac Injector Prototype”, *Phys. Rev. ST Accel. Beams*, vol. 16, pp. 073401, July 2013.
- [4] C. Gulliford, *et al.*, “Demonstration of cathode emittance dominated high bunch chargebeams in a DC gun-based photoinjector”, *Applied Physics Letters*, vol. 106, no. 9, pp. 094101, 2015.
- [5] R. Eichhorn, *et al.*, “The Cornell Main Linac Cryomodule: A Full Scale, High Q Accelerator Module for CW Application”, *Physics Procedia*, vol. 67, pp. 785–790, 2015.
- [6] D. Sagan, “Bmad: A relativistic charged particle simulation”, *Nuc. Instrum. Methods Phys. Res. A*, vol. 558, pp 356-59, 2006.
- [7] D. Sagan and J. C. Smith, “The Tao Accelerator Simulation Program”, in *Proc. PAC’05*, pp. 4159–4161, May 2005.
- [8] [epics.anl.gov/extensions/ezca/index.php](http://epics.anl.gov/extensions/ezca/index.php)
- [9] [www.pulsar.nl/gpt/](http://www.pulsar.nl/gpt/)
- [10] A. Nunez-delPrado and C. Gulliford, *CBETA-V: Advanced Online Simulation Tools for the CBETA Machine*, Cornell Internal Report, 2018.



(a) Horizontal orbit before SVD.



(b) Horizontal orbit after SVD.

Figure 9: Steering onto the periodic orbit using SVD: (a) shows the initial non-periodic orbit for ten different beam energies, (b) shows the results of steering onto the period orbit using the last two dipoles in the splitter S1 section. The red circles indicate the BPM position and readout values.

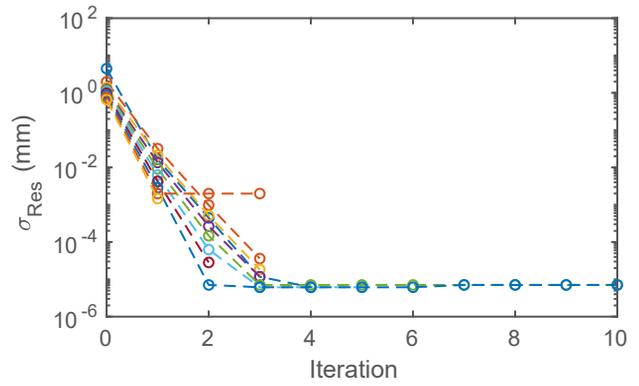


Figure 10: Correction error of the 10 simulated orbits with energies ranging from 39 to 52 MeV, as a function of the number of iterations of orbit correction.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

# LONGITUDINAL BEAM DYNAMICS WITH A HIGHER-HARMONIC CAVITY FOR BUNCH LENGTHENING\*

G. Bassi<sup>†</sup>, J. Tagger, Brookhaven National Laboratory, Upton, 11973 NY, USA

## Abstract

We discuss the longitudinal beam dynamics in storage rings in the presence of a higher-harmonic cavity (HHC) system for bunch lengthening. We first review the general conditions for HHC operations, either in active or passive mode, assuming the stability of the system. For uniform filling patterns, a distinction is made between operations with a normal-conducting HHC, where optimal conditions for bunch lengthening can be satisfied, and operations with super-conducting HHC, where optimal conditions can be met only approximately. The option to operate the NSLS-II storage ring with a passive, super-conducting third harmonic cavity (3HC) system is discussed next. The stability and performance of the system in the presence of a gap in the uniform filling, which corresponds to the present mode of operation of the NSLS-II storage ring, is investigated with self-consistent Vlasov-Fokker-Planck simulations performed with the code SPACE [1].

## INTRODUCTION

Higher-harmonic cavities (HHCs) play a crucial role for stable operations of present and future low-emittance storage rings. The primary benefic effect provided by the HHC is bunch lengthening without energy spread increase, with consequent beam lifetime improvement and reduction of the effect of intrabeam scattering on the transverse emittance [2]. Besides bunch lengthening, the highly nonlinear potential well distortion produced by the HHC introduces a strong dependence of the synchrotron tune on the amplitude of synchrotron oscillations. The induced anharmonic motion with enhanced synchrotron tune spread provides a powerful mechanism, known as Landau damping, for the suppression of collective instabilities. Moreover, the increase in bunch length and synchrotron tune spread can enhance the stabilizing effect of positive chromaticity on the transverse oscillations and help to stabilize higher-order head-tail modes [2]. The option considered for the NSLS-II storage ring is to operate with a passive superconducting 3HC [3, 4], a choice motivated by the successful development and operation of the superconducting 3HC system at the ELETTRA [5] and SLS [6] storage rings, a system that has been developed in the framework of the SUPER-3HC project [7]. The SUPER-3HC project represented the first superconducting application of a HHC system in storage rings, taking advantage of the very high quality factor of the superconducting cavity and the associated narrow bandwidth, allowing for the tuning of the 3HC very near to the

third harmonic of the beam, without exciting longitudinal instabilities [5]. The success of the 3HC operation at the ELETTRA storage ring is substantiated by a beam lifetime improvement by more than a factor of three with respect to the nominal value, an improvement that has led to a change in the refilling frequency of the storage ring, allowing a refilling every 48 hr instead of every 24 hr, with benefit for the reliability and stability of user's operations and relevant benefit even for the machine thermal stability [5]. The success with the operation of a 3HC at the SLS storage ring is substantiated by a bunch lengthening up to a factor of three and a beam lifetime increase greater than a factor of two, achieved with stable conditions at the design current of 400 mA [6]. The success experienced at the ELETTRA and SLS storage rings has clearly shown that the very high quality factor of the superconducting HHC renders the performance of the HHC system less sensitive to high-order modes (HOMs) driven longitudinal coupled bunch instabilities, which is a major issue with normal conducting HHCs, where powerful longitudinal feedback systems are often needed for stable operations. Performance limiting factors, however, such as transients effects induced by non uniform filling patterns and the beam phase instability [8], can be detrimental for stable HHC operations, and need to be carefully investigated with detailed design studies. Accurate numerical simulations represent an essential part of the aforementioned design studies, with their goal to determine feasible conditions of operation and their range of applicability. To this end, the stability and performance of the passive superconducting 3HC system for the NSLS-II storage ring is studied numerically with the parallel, particle tracking code SPACE [1], which allows to follow self-consistently the dynamics of  $h$  bunches, where  $h$  is the number of RF buckets, in arbitrary multi-bunch configurations. The specific goal of the numerical simulations is to determine stable HHC cavity settings and to study the performance limitation due to a gap in the uniform filling, which represents the nominal NSLS-II mode of operation.

## OPERATIONS WITH HIGHER-HARMONIC CAVITIES

In the discussion of the theoretical conditions for optimal bunch lengthening, we assume a stable, beam loading compensated HHC system characterized by an equilibrium multi-bunch configuration. Radiation damping and quantum fluctuations are excluded from the analysis. The overall stability of the HHC system, including radiation damping and quantum fluctuations, together with the inclusion of a model for beam loading compensation, will be addressed in the next section with time-dependent Vlasov-Fokker-Planck simulations.

\* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy

<sup>†</sup> gbassi@bnl.gov

## Active Higher-Harmonic Cavity

We assume that the voltage  $V(\tau)$  seen by a particle in the beam with arrival time  $\tau$  is

$$V(\tau) = V_{rf}[\sin(\omega_{rf}\tau + \phi_s) - r \sin(m\omega_{rf}\tau + \phi_m)] - \frac{U_s}{e} \\ =: V_c(\tau) - \frac{U_s}{e}, \quad (1)$$

where  $V_{rf}$  is the amplitude of the voltage of the main rf cavity,  $\omega_{rf} = h\omega_0$ , where  $h$  is harmonic number and  $\omega_0$  the angular revolution frequency,  $m$  is the order of the HHC and  $r$  the ratio of HHC to main cavity amplitude voltage,  $U_s$  energy loss per turn,  $e$  the electron charge,  $\phi_s$  and  $\phi_m$  the phases of the synchronous particle in the main and HHC respectively. Here  $V_c(\tau)$  is the total rf voltage produced by the main rf cavity and by HHC.

The longitudinal dynamics in the double RF system described by Eq. (1) has been comprehensively discussed, together with optimal conditions for bunch lengthening, by Hofmann and S. Myers in 1980 [9]. See also [10]. Here we summarize the main results.

Table 1: NSLSII Parameters

Parameter	Symbol	Value	Unit
Energy reference particle	$E_0$	3	GeV
Average current	$I_0$	500	mA
Gap in the uniform filling	$g$	260	
Harmonic number	$h$	1320	
Circumference	$C$	792	m
Bunch duration	$\sigma_\tau$	14.5	ps
Energy spread	$\sigma_p$	0.00087	
Energy loss per turn	$U_s$	674	keV
Momentum compaction	$\alpha$	0.00037	
Revolution frequency	$f_0$	378.5	kHz

Table 2: RF Parameters Main (2 Cavities) and 3HC (1 Cavity)

Per Cavity Parameters	Symbol	Value	Unit
main frequency	$\omega_{rf}$	$2\pi \times 499.68$	MHz
HHC frequency	$3\omega_{rf}$	$2\pi \times 1499.04$	MHz
main voltage	$V_{rf}$	1.7	MV
main shunt impedance	$R_M$	2.97	MΩ
main quality factor	$Q_M$	66817	
HHC shunt impedance	$R_H$	22880	MΩ
HHC quality factor	$Q_H$	$2.6 \times 10^8$	

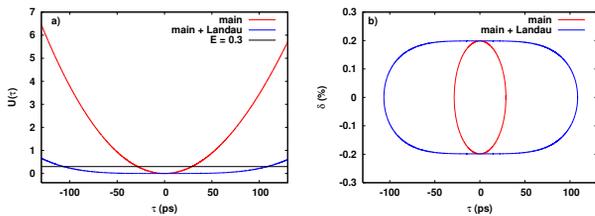


Figure 1: a) Potential energy  $U(\tau)$  without HHC (red line) and with HHC (blue line) with parameters of the NSLS-II storage ring (see Table 1) satisfying Eq. (5-7). b) Phase space portrait corresponding to a) for  $E = 0.3$ .

To compensate for the energy loss  $U_s$ , we require that the voltage seen by the synchronous particle is zero, i.e.  $V(0) = 0$ . In addition, we require  $V'(0) = V''(0) = 0$ , where  $' = d/d\tau$ . Thus

$$\sin \phi_s = r \sin \phi_m + \frac{U_s}{eV_{rf}}, \quad (2)$$

$$\cos \phi_s = rm \cos \phi_m, \quad (3)$$

$$\sin \phi_s = rm^2 \sin \phi_m, \quad (4)$$

which, solved for  $\phi_s$ ,  $\phi_m$  and  $r$  give

$$\sin \phi_s = \frac{m^2}{m^2 - 1} \sin \phi_{s0}, \quad \sin \phi_{s0} = \frac{U_s}{eV_{rf}}, \quad (5)$$

$$\tan \phi_m = \frac{m \sin \phi_{s0}}{\sqrt{(m^2 - 1)^2 - m^4 \sin^2 \phi_{s0}}}, \quad (6)$$

$$r = \frac{1}{m} \sqrt{1 - \frac{m^2}{m^2 - 1} \sin^2 \phi_{s0}}. \quad (7)$$

where we introduced  $\phi_{s0}$ , the synchronous phase in absence of the HHC. With the voltage given by Eq. (1), from the Hamiltonian

$$H(\tau, \delta) = \frac{\eta}{2} \delta^2 + U(\tau),$$

$$U(\tau) = \frac{eV_{rf}}{E_0 T_0 \omega_{rf}} \left[ \cos(\omega_{rf}\tau + \phi_s) - \cos \phi_s + \frac{r}{m} \cos \phi_m \right. \\ \left. - \frac{r}{m} \cos(m\omega_{rf}\tau + \phi_m) + \omega_{rf}\tau \sin \phi_{s0} \right], \quad (8)$$

follow the longitudinal equations of motion

$$\dot{\tau} = \frac{\partial H}{\partial \delta} = \eta \delta,$$

$$\dot{\delta} = -\frac{\partial H}{\partial \tau} = \frac{eV_{rf}}{E_0 T_0} \left[ \sin(\omega_{rf}\tau + \phi_s) - r \sin(m\omega_{rf}\tau + \phi_m) \right. \\ \left. - \sin \phi_{s0} \right], \quad (9)$$

where  $' = d/dt$ ,  $\eta = \alpha - \gamma_0^{-2}$  is the slippage factor, where  $\gamma_0$  is the Lorentz factor,  $\delta = (E - E_0)/E_0$  is the relative energy deviation with respect to the synchronous particle with energy  $E_0$ , and the arbitrary constant in the definition of  $U(\tau)$  has been chosen in order to satisfy  $U(0) = 0$ . Since  $U$  does not depend explicitly on time,  $H$  is a constant of motion and setting  $E = H$  we have  $\delta(\tau) = \pm \sqrt{2(E - U(\tau))/\eta}$ ,  $E = \text{const.}$  In Fig. 1a we show the potential energy  $U(\tau)$  with only the main rf cavity (red line) and with a third-harmonic cavity (blue line) with parameters satisfying Eqs. (5-7) for the NSLS-II storage ring (see Table 1). In Fig. 1b we show the corresponding phase space portrait for  $E = 0.3$ . The optimal conditions satisfied by the voltage  $V(\tau)$  induce a bunch lengthening without an increase of the energy spread. According to Table 1,  $r = 0.329 \approx 1/3$ , thus the peak voltage induced by the harmonic cavity system is roughly one third the peak voltage  $V_{rf}$  of the main cavity system. In the case of no energy loss ( $U_s = 0$ ) the conditions (2)-(4) simplify to  $\phi_s = \phi_m = 0$  and  $r = 1/m$ . In Fig.2a we plot the

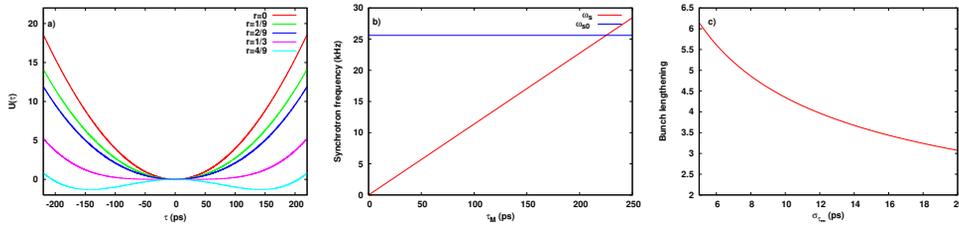


Figure 2: a) Potential energy  $U(\tau)$  in the active case ( $m = 3$ ) and for no energy loss ( $U_s = 0$ ). b) Synchrotron frequency  $\omega_s$  for the quartic potential (red line) and synchrotron frequency  $\omega_{s0}$  for the quadratic potential (blue line). c) Bunch lengthening  $u$  vs. bunch length  $\sigma_{\tau_m}$  (w/o HHC).

potential energy  $U(\tau)$  for  $m = 3$  and different values of  $r$ . For  $r = 4/9$  the potential energy has two stable fixed points close to  $\pm 150$  ps.

**Small Oscillations** For small oscillations ( $\tau \ll 1$ ) the potential energy  $U(\tau)$  without HHC can be approximated by a quadratic function of  $\tau$

$$U(\tau) = -\frac{eV_{rf}\omega_{rf}\cos\phi_{s0}}{2E_0T_0}\tau^2 =: \frac{\omega_{s0}^2}{2\eta}\tau^2, \quad (10)$$

while with the addition of the harmonic cavity the potential given by Eq.(8) can be approximated by a quartic

$$U(\tau) = -\frac{eV_{rf}(m^2 - 1)\omega_{rf}^3\cos\phi_s}{24E_0T_0}\tau^4. \quad (11)$$

For a potential energy satisfying  $U(-\tau) = U(\tau)$  and  $U(\tau) > 0$  for  $\tau > 0$ , the trajectory is confined in the region  $[-\tau_M, \tau_M] \times [-\delta_M, \delta_M]$  where  $\tau_M$  and  $\delta_M$  satisfy  $U(\tau_M) = E$  and  $\eta\delta_M^2/2 = E$  respectively, thus the amplitude of the trajectory is  $d = 2\tau_M$ . It can be shown that the synchrotron frequency  $\omega_s$  for the quartic potential (11) reads

$$\omega_s(\tau_M) = \frac{\pi}{2}\sqrt{\frac{m^2 - 1}{6}}\sqrt{\frac{\cos\phi_s}{\cos\phi_{s0}}}\frac{\omega_{rf}\omega_{s0}}{K(1/\sqrt{2})}\tau_M, \quad (12)$$

where  $K$  is the complete integral of the first kind. The dependence of the synchrotron frequency on  $\tau_M$  provides Landau damping for beam stability. In Fig. 2b we plot  $\omega_s$  as a function of  $\tau_M$ . It can also be shown that the bunch lengthening factor  $u$  for an equilibrium distribution  $\rho_e$  in the quartic potential (11) reads

$$u := \frac{\sigma_{\tau L}}{\sigma_{\tau m}} = \sqrt{\frac{\Gamma(3/4)}{\Gamma(1/4)}}\left(\frac{24\cos\phi_{s0}}{(m^2 - 1)\omega_{rf}^2\cos\phi_s}\right)^{1/4}\frac{1}{\sqrt{\sigma_{\tau m}}}, \quad (13)$$

where  $\Gamma$  is the Gamma function,  $\sigma_{\tau m} = \eta\sigma_\delta/\omega_{s0}$  is the equilibrium bunch length with only the main cavity, and  $\sigma_{\tau L}$  is the equilibrium bunch length with the harmonic cavity. The bunch lengthening factor  $u$  as a function of  $\sigma_{\tau m}$  is plotted in Fig. 2c. With NSLS-II parameters ( $\sigma_{\tau m} = \sigma_\tau$ ) the bunch lengthening factor reads  $u = 3.7$ .

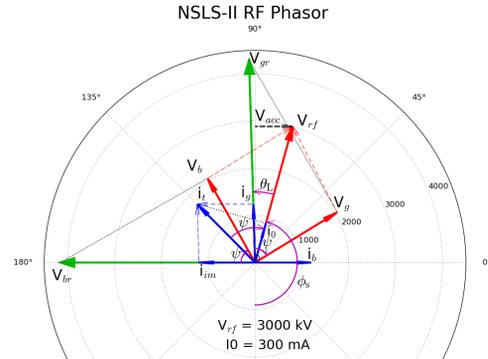


Figure 3: RF phasor of the NSLS-II storage ring during operations with a stored beam current  $I_0 = 300$ mA,  $V_{rf} = 3$  MV,  $\phi_s = 164.5^\circ$ ,  $\theta_L = -17^\circ$ ,  $V_b = 2062$  kV,  $V_g = 2132$  kV and detuning angle  $\psi = 60^\circ$ .

### Passive Higher-Harmonic Cavity

For passive HHC operations, the total RF voltage is given by the sum of the voltage produced by the powered main cavity and the beam voltage induced by the beam in both cavities. We assume in this section that the main cavity is beam loading compensated.

**Operations with Normal-Conducting Cavities.** In the case of stationary bunches uniformly distributed around the ring, and for a narrow-band resonator wake with frequency  $\omega_r$ , shunt impedance  $R_s$  and quality factor  $Q$ , the voltage acting on the beam reads

$$V_C(\tau) = V_{rf}\sin(\phi_s + \omega_{rf}\tau) - i_m R_s \cos\psi \cos(\psi + m\omega_{rf}\tau), \quad (14)$$

where  $i_{im} = 2I_0\tilde{\rho}(\omega)$  and the detuning angle  $\psi$  satisfies

$$\tan\psi = 2Q\delta, \delta = \frac{1}{2}\left(\frac{\omega_r}{m\omega_{rf}} - \frac{m\omega_{rf}}{\omega_r}\right) \approx \frac{\omega_r - m\omega_{rf}}{m\omega_{rf}}. \quad (15)$$

Here  $\tilde{\rho}(\omega)$  is the Fourier transform of the bunch density  $\rho(\tau)$  satisfying  $\rho(-\tau) = \rho(\tau)$ . For Gaussian bunches  $i_{im} = 2I_0e^{-\frac{1}{2}(m\omega_{rf}\sigma_\tau)^2}$ . Imposing the same the conditions (2–4) for the active HHC, by comparing Eq. (1) and Eq. (14) it follows ( $0 < \phi_s < \pi \implies 0 < \psi < \pi/2 \implies \cos\psi > 0$ )

$$\tan\psi = -\cot\phi_m = -m\cot\phi_s, \quad (16)$$

$$R_s = \frac{rV_{rf}}{i_{im}\cos\psi} = \frac{V_{rf}\sin\phi_s}{i_{im}m^2\cos^2\psi}, \quad (17)$$

where we used  $\sin \phi_m = \cos \psi$  ( $\tan \psi = -\cot \phi_m \implies \psi = \phi_m - \pi/2 \implies \sin \phi_m = \cos \psi$ ). Therefore the conditions for passive HHC operations corresponding to the active case (5–7) are

$$\sin \phi_s = \frac{m^2}{m^2 - 1} \sin \phi_{s0}, \quad (18)$$

$$\tan \psi = -\frac{\sqrt{(m^2 - 1)^2 - m^4 \sin^2 \phi_{s0}}}{m \sin \phi_{s0}}, \quad (19)$$

$$R_s = \frac{V_{rf}(m^2 - 1)(1 - \sin \phi_{s0})}{i_{im} m^2 \sin \phi_{s0}}. \quad (20)$$

An important difference to active case is that  $R_s$  is uniquely determined and a function of the beam current  $I_0$ . Notice, however, that these conditions do not impose any constraint on the value of  $Q$ , therefore do not determine uniquely the detuning frequency  $\Delta\omega = \omega_r - m\omega_{rf}$ . The optimal parameters for passive HHC operations of the NSLS-II storage ring according to Tables 1 and 2 are therefore  $\sin \phi_s = 0.4592$ ,  $\tan \psi = 5.8 \implies \psi = 80.22^\circ$  and  $R_s = 9.02 \text{ M}\Omega$ .

**Operations with Super-Conducting Cavities.** According to Table 2, the shunt impedance of the HHC is  $R_H = 22880 \text{ M}\Omega$ , much bigger than the optimal value  $R_s = 9.02 \text{ M}\Omega$ , so the optimal conditions for passive operations can not be met. Good conditions, however, can be found by comparing Eq. (1) and Eq. (14) at  $\tau = 0$ , which gives  $R_s = rV_{rf}/(i_{im} \cos \psi)$ , and by noticing from Eq. (7) that  $r \approx 1/m$ , since to good approximation  $\sin \phi_{s0}^2 \ll 1$ . We therefore impose on the detuning angle  $\psi$  the condition  $\cos \psi = V_{rf}/(mi_{im}R_H)$ , which implies that the detuning frequency  $\Delta\omega_H$  approximately satisfies

$$\Delta\omega_H = \frac{m^2\omega_{rf}i_bR_HV_{rf}}{2Q_HV_{rf}}, \quad (21)$$

where we used Eq. (15) and the fact that  $\sin \psi \ll 1$ . With parameters listed in Table 2, it follows that  $\Delta\omega_H = 2\pi \times 58.24 \text{ kHz}$ .

## NUMERICAL SIMULATIONS

With the inclusion of a model for beam loading compensation, time dependent simulations of the Vlasov-Fokker-Planck equation allow for the study of the overall stability of the HHC system. Moreover, numerical simulations allow for the study of transient effects induced by arbitrary multi-bunch configurations, such as a gap in the uniform filling pattern for ion clearing, which corresponds to the nominal configuration of the NSLS-II storage ring. The numerical simulations of the Vlasov-Fokker-Planck equation discussed in this paper are done with the parallel code SPACE, a particle tracking code that allows for the simultaneous study of short- and long-range wakefield effects in storage rings. The general strategy adopted by SPACE to study multibunch effects is to distribute each bunch to one processor, each with  $\mathcal{N}$  simulations particles representing the bunch population, thus performing the short- and long-range wakefield

calculation in serial and parallel respectively. For more details on the code SPACE see [1]. For steady state beam loading compensation, the algorithm implemented in SPACE is based on the standard phasor diagram, shown in Fig. 3 with parameters of one of the operational settings of the NSLS-II storage ring. The numerical simulations discussed in this paper have been done on the supercomputers Cori and Edison at NERSC [11]. The equations of motion for bunch  $n$  ( $n = 0, \dots, h - 1$ ), shown here without radiation damping and quantum fluctuations, for the general NSLS-II operations with two main cavities and one HC read

$$\begin{aligned} \dot{\tau} &= \eta\delta, \\ \dot{\delta} &= \frac{e}{E_0T_0} \left[ \sum_{i=1}^2 V_{gr,i} \cos \psi_i \sin(\omega_{rf}\tau + \phi_s - \theta_{L,i} + \psi_i) \right. \\ &\quad \left. - V_n(\tau, s) - \frac{U_s}{e} \right], \end{aligned} \quad (22)$$

where  $V_{gr,i}$ ,  $\psi_i$  and  $\theta_{L,i}$  ( $i = 1, 2$ ) correspond to the generator voltage, detuning angle and load angle of the two main cavities respectively, and  $V_n(\tau, s)$  is the total beam loading voltage acting on bunch  $n$ . The numerical simulations discussed in this paper assume the two main cavities with same beam loading parameters, which correspond to the standard mode of operation of the NSLS-II storage ring. By projecting the current phasors shown in Fig. 3 along and perpendicular to the RF voltage phasor,  $V_{gr}$  and  $\psi$  satisfy

$$\tan \psi = \left( 1 + \frac{i_{im,M}}{i_0} \sin \phi_s \right) \tan \theta_L + \frac{i_{im,M}}{i_0} \cos \phi_s, \quad (23)$$

$$V_{gr} = \frac{V_{rf}}{\cos \theta_L} \left( 1 + \frac{i_{im,M}}{i_0} \sin \phi_s \right). \quad (24)$$

where  $i_{im,M} = 2I_0\tilde{\rho}(\omega_{rf})$  is the image current in the main cavity, and  $i_0 = V_{rf}/R_M$ . In the analysis of the performance of the NSLS-II HHC system, we study first the case with a uniform filling pattern, and compare the results with the nominal case, which corresponds to a gap of 260 bunches, (80% fractional filling), and with the case with a gap of 130 bunches (90% fractional filling). In the discussion that follows we omit the subscript  $H$  to label the detuning frequency of the HHC.

### Uniform Fillings

In Fig. 4 we show numerical simulations for values of the HHC detuning frequency  $\Delta f = 45 \text{ kHz}$ ,  $55 \text{ kHz}$  and  $65 \text{ kHz}$ , above and below the value  $\Delta f = 58.24 \text{ kHz}$  calculated in Sect. II for good bunch lengthening conditions. The longitudinal density of the bunches after 100,000 turns is shown in Fig. 4a for  $\Delta f = 45 \text{ kHz}$ , in Fig. 4b for  $\Delta f = 55 \text{ kHz}$  and Fig. 4c for  $\Delta f = 65 \text{ kHz}$ . The bunch lengthening is uniform across the bunch train for  $\Delta f = 55 \text{ kHz}$  and  $\Delta f = 65 \text{ kHz}$ , with values  $\sigma_\tau = 50 \text{ ps}$  and  $\sigma_\tau = 36 \text{ ps}$  respectively, as shown in Fig. 4e and Fig. 4f, corresponding to the bunch lengthening factors  $u = 3.45$  and  $u = 2.48$ . For  $\Delta f = 45 \text{ kHz}$ , the longitudinal density of the bunches shows a double peaked structure and a non-uniform bunch

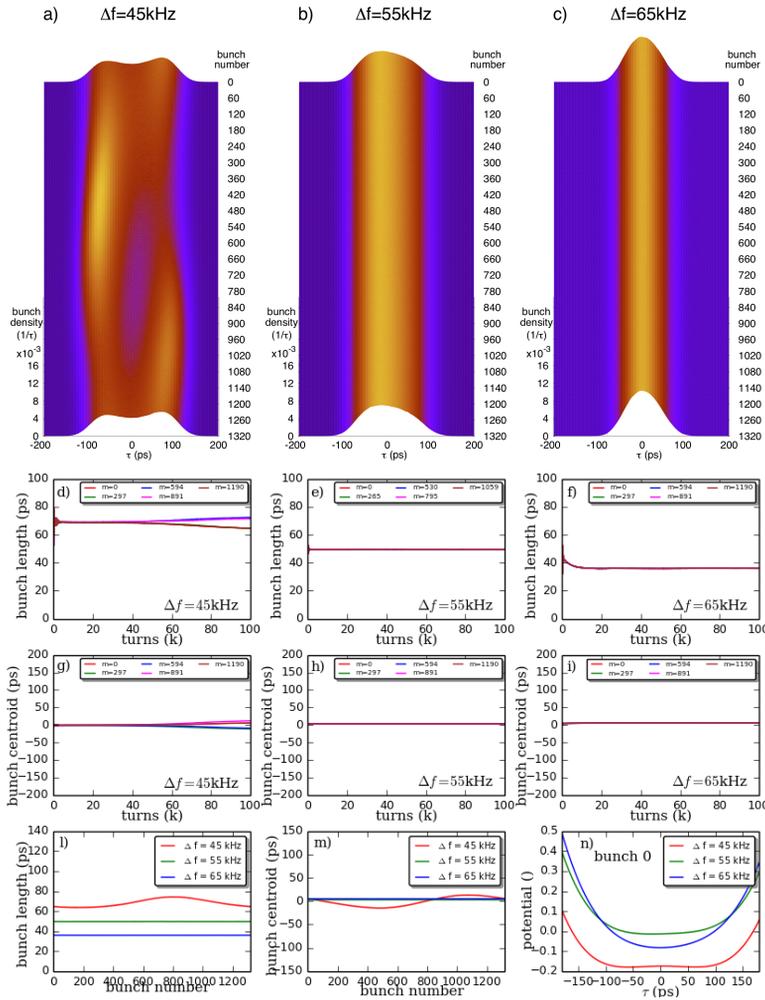


Figure 4: Numerical simulations in the case of uniform filling with detuning frequency  $\Delta f = 45$  kHz, 55 kHz and 65 kHz.

lengthening, as shown in Fig. 4a, Fig. 4g and Fig. 4l. Long term simulations up to 500,000 turns, as plotted in Fig. 4a and Fig. 4g, show that for  $\Delta f = 45$  kHz the HHC system is weakly unstable, signing the transition to an “overstretching” regime, with average bunch length across the train of  $\approx 70$  ps. The potential well of bunch  $n = 0$ , showing two local minima, is shown by the red trace in Fig. 4n.

### Gap in the Uniform Filling: $g = 130$ and $g = 260$

The case of a gap in the uniform filling corresponds to a train of  $M = h - g$  bunches, where  $h$  is the harmonic number and  $g$  is the gap. The case with nominal gap,  $g = 260$ , corresponding to a 80% fractional filling, is compared with the case  $g = 130$ , corresponding to a 90% fractional filling. The main effect introduced by a gap in the uniform filling is a monotonic variation of the bunch centroid across the train, and a reduced, non uniform bunch lengthening. Fig. 5 and Fig. 6 show numerical simulations up to 100,000 turns with gaps  $g = 130$  and  $g = 260$  respectively, for the same HHC detuning frequencies of the uniform case. The monotonic variation of the bunch centroid across the train is evident from the longitudinal density of the bunches

shown in Fig. 5a-c, from the time evolution of the bunch centroids shown in Fig. 5g-i, and from Fig. 5m, where it can be noticed that the range of variation of the bunch centroids increases with the decrease of the HHC detuning frequency. Fig. 5d-f and Fig. 5l show the non uniform bunch lengthening across the train, with a similar average value  $\langle \sigma_\tau \rangle \approx 35$  ps for the different detuning frequencies. We notice that for  $\Delta f = 45$  kHz the bunches in the center of the train have longer bunch length than the bunches in the periphery of the train. The case of nominal gap,  $g = 260$ , is discussed in Fig. 6. A detuning frequency threshold is observed in this case. For detuning frequencies above the threshold, as shown for  $\Delta f = 65$  kHz in Fig. 6c, Fig. 6f and Fig. 6l a stable equilibrium is reached after 100,000 turns, with average bunch length across the train  $\langle \sigma_\tau \rangle \approx 27$  ps, while for detuning frequencies below threshold, as shown in Fig. 6a, Fig. 6d for  $\Delta f = 45$  kHz and Fig. 6b, Fig. 6e and Fig. 6h for  $\Delta f = 55$  kHz, an unstable regime with saturation is observed, with both the bunch lengths and bunch centroids exhibiting a well defined mode of oscillation. The numerical simulations discussed so far have been done with load angle  $\theta_L = 0$ . In the attempt to improve stability, the two unstable

Content from this work may be used under the terms of the CC BY 3.0 licence under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

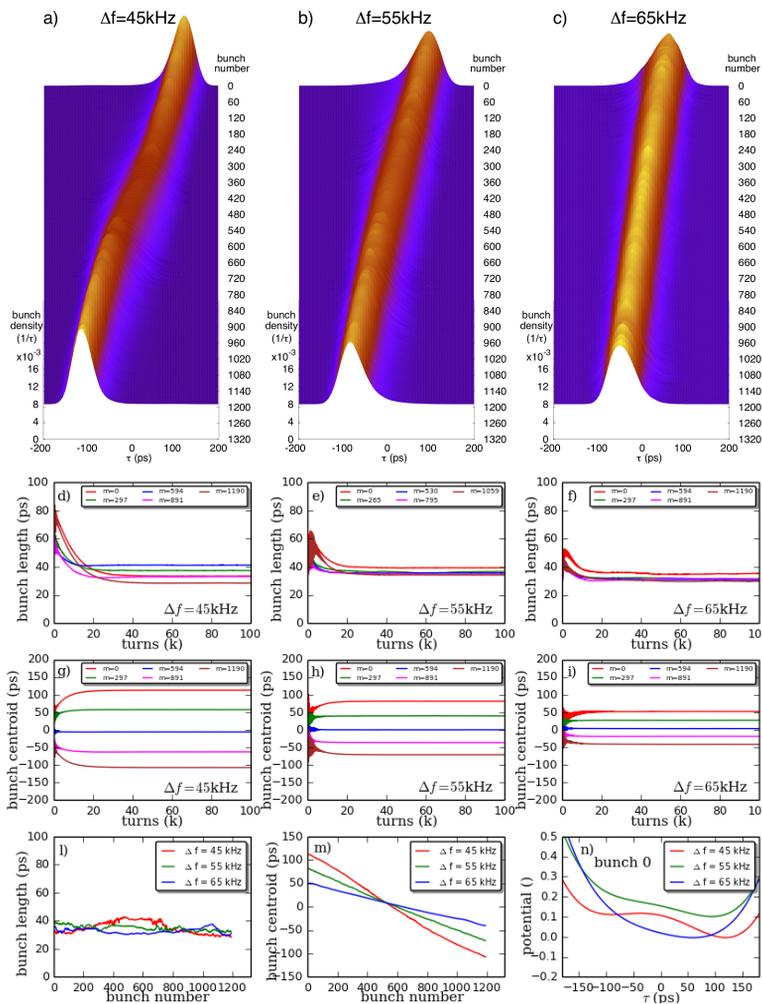


Figure 5: Numerical simulations up to 100,000 turns with a gap  $g = 130$ . HHC detuning frequencies as in Fig. 4.

configurations at the nominal gap  $g = 260$  for  $\Delta f = 45$  kHz and 55 kHz have been simulated with  $\theta_L < 0$ . In both cases the numerical results show that the introduction of a negative load angle is partially effective in stabilizing the HHC system, with the “stabilizing” load angle in the range  $[-20^\circ, 0^\circ]$  for  $\Delta f = 45$  kHz, and  $[-40^\circ, -20^\circ]$  for  $\Delta f = 55$  kHz.

## CONCLUSION

The numerical result clearly show a reduction in both performance and stability of the HHC system with the increase of the gap in the uniform filling, with the case of a gap  $g = 130$ , corresponding to a 90% fractional filling, stable at all the detuning frequencies considered. On the other hand, the nominal case with  $g = 260$ , corresponding to a 80% fractional filling, has shown to be unstable for some values of the detuning frequencies. Moreover, the case with  $g = 130$  has shown a superior performance in terms of bunch lengthening with respect to the nominal case. The performance of stable HHC settings for the nominal case  $g = 260$  and the case  $g = 130$ , as a function of detuning frequency  $\Delta f$  and load angle  $\theta_L$ , is shown in Fig. 7, both in terms of average bunch lengthening and uniformity of the bunch centroid and bunch

length across the train. The average bunch lengthening factor for the 80% and 90% fractional filling is approximately 2 and 2.5 respectively, to be compared with the bunch lengthening factor of the uniform filling case, which, according to Fig. 4g, is approximately 3.5 for  $\Delta = 55$  kHz. The performance reduction in the average bunch lengthening due to a gap in the uniform filling is therefore 45% for  $g = 260$  and 30% for  $g = 130$ . Machine studies are planned at the NSLS-II storage ring to revisit the need of the nominal 80% fractional filling pattern for ion clearing, with the goal to increase the fractional filling towards a more uniform filling pattern. Arbitrary, more general multibunch configurations, such as filling patterns with multiple, smaller gaps than the nominal, are also under consideration. To this end, an analytical calculation to determine the beam loading voltage induced by arbitrary, stationary bunches has been done and implemented in a numerical code for fast parametric scans and guidance in Vlasov-Fokker-Planck simulations of the HHC system [12].

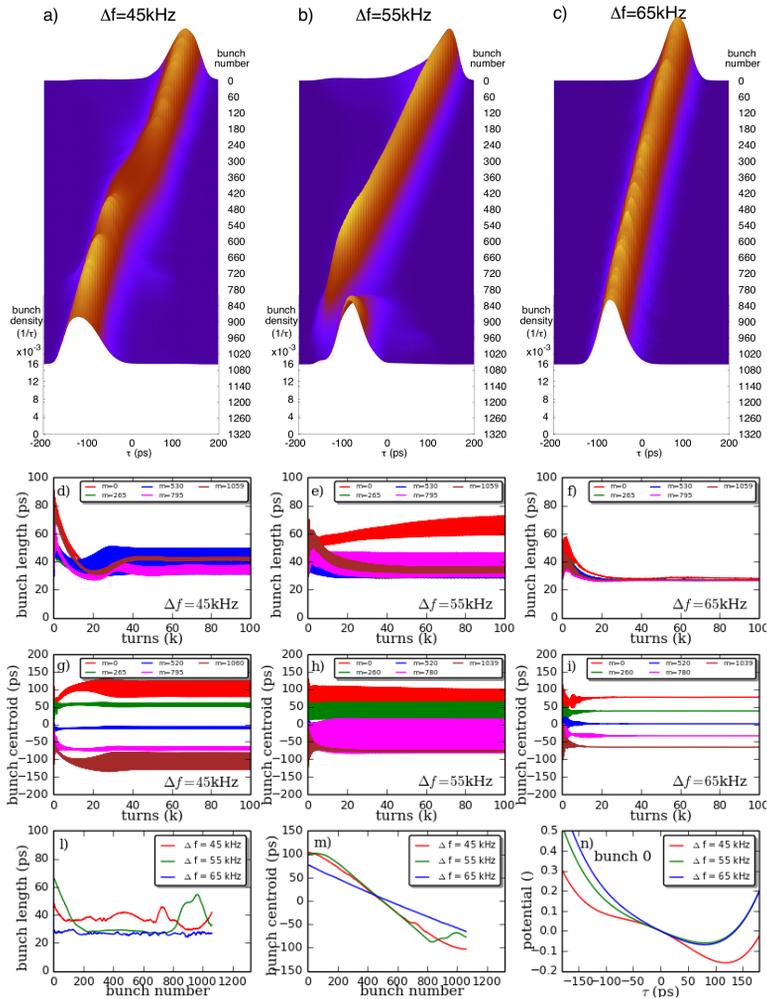


Figure 6: Numerical simulations up to 100,000 turns with nominal gap  $g = 260$ . HHC detuning frequencies as in Fig. 4.

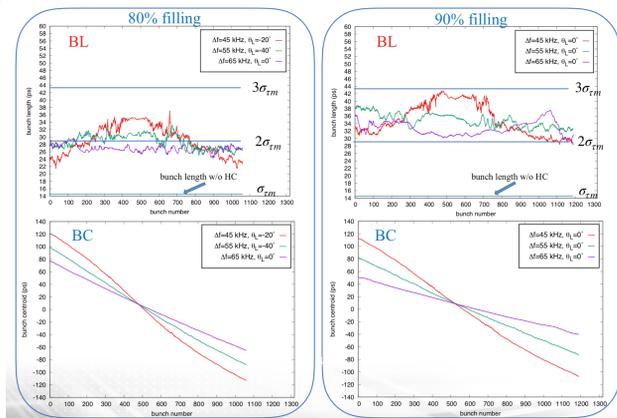


Figure 7: Performance of the HHC stable settings as a function of the detuning frequency  $\Delta$  and load angle  $\theta_L$  for  $g = 260$  (80% fractional filling) and  $g = 130$  (90% fractional filling). The bunch length and bunch centroid are labeled with BL and BC respectively.

## REFERENCES

[1] G. Bassi, A. Blednykh and V. Smaluk, *Phys. Rev. Accel. Beams* **19**, 024401, 2016.

[2] NSLS-II Conceptual Design Report, 2006.  
 [3] N. Towne, NSLS-II Tech Note 19, 2006; N. Towne, NSLS-II Tech Note 37, 2007; N. Towne, NSLS-II Tech Note 27, 2007; N. Towne, NSLS-II Tech Note 18, 2006.  
 [4] G. Bassi and J. Tagger, NSLS-II Tech Note 286, 2018.  
 [5] G. Penco and M. Svandrlík, *Phys. Rev. ST Accel. Beams* **9**, 044401, 2006.  
 [6] M. Predozzi *et al.*, in *Proc. 11th Workshop on RF Superconductivity*, Lübeck/Travemünde, Germany, paper MOP25, 2003.  
 [7] G. Penco, Ph.D. thesis, Sincrotrone Trieste – University of Milan, 2004.  
 [8] R. A. Bosh, K. J. Kleman, and J. J. Bisognano, “Robinson Instabilities with higher-harmonic cavity”, *Phys. Rev. ST Accel. Beams* **4**, 074401, 2001; R. A. Bosh, *Phys. Rev. ST Accel. Beams* **8**, 084401, 2005.  
 [9] A. Hofmann and Myers, in *Proc. 11th Int. Conf. on High Energy Accelerators*, Geneva, 1980.  
 [10] K.Y. Ng, “Physics of Intensity Dependent Beam Instabilities”, Fermilab-FN-0713.  
 [11] NERSC, <http://www.nersc.org>  
 [12] G. Bassi and J. Tagger, in preparation.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

# CALCULATION OF THE AGS OPTICS BASED ON 3D FIELDS DERIVED FROM EXPERIMENTALLY MEASURED FIELDS ON MEDIAN PLANE\*

N. Tsoupas<sup>†</sup>, J. S. Berg, S. Brooks, F. Méot, V. Ptitsyn, D. Trbojevic  
 Brookhaven National Laboratory, Upton, NY, USA

## Abstract

Closed orbit calculations of the Alternating Gradient Synchrotron (AGS) were performed and the beam parameters at the Fast Beam Extraction (FEB) point of the AGS [1] were calculated using a modified RAYTRACE computer code [2] to generate 3D fields from measured field maps on the median plane of the AGS combined function magnets. The algorithm which generates 3D fields from field maps on a plane is described in reference [3] which discusses the details of the mathematical foundation of this approach. In this paper we discuss results from studies reported in Refs. [1,4] that are based on the 3D fields generated from measured field components on a rectangular grid of a plane. A brief overview of the algorithm used will be given, and one of the two methods of calculating the required field derivatives on the plane will be discussed.

## INTRODUCTION

The AGS is one of the pre-acceleration stages of the RHIC complex. Figure 1 is an aerial picture of the site with the green trace indicating the tunnel of the AGS. The 240 com-

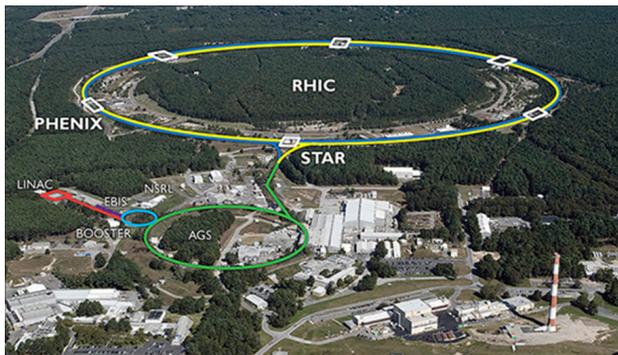


Figure 1: Area view of the RHIC complex. The green trace indicates the location of the AGS tunnel.

combined function main magnets of the AGS are placed inside the AGS tunnel whose schematic diagram is showing in Fig. 2. The AGS main magnets are separated in 12 superperiods of 20 magnets per superperiod spanning an arc of  $30^\circ$  as shown in Fig. 2.

Figure 3 is a schematic diagram of the main magnets layout (small rectangles) in each superperiod. The “+” and “-” signs on each magnets indicate the focusing and defocusing quadrupole property of each combined function magnet for positive ions circulating in the synchrotron.

\* Work supported by the US Department of Energy  
<sup>†</sup> tsoupas@bnl.gov

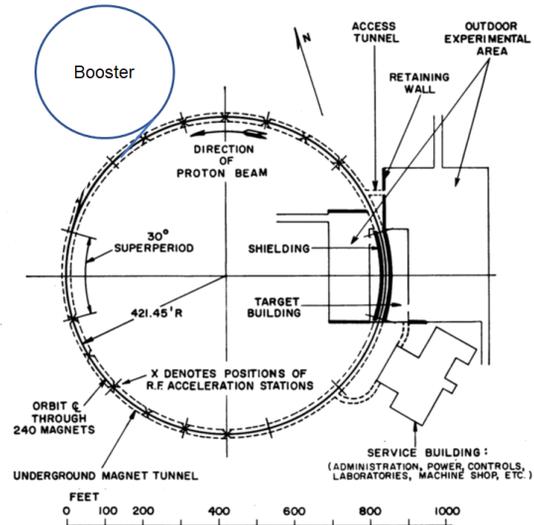


Figure 2: Schematic diagram of the AGS tunnel. The 240 main AGS magnets are separated in 12 superperiods with 20 magnets per superperiod which spans an arc of  $30^\circ$

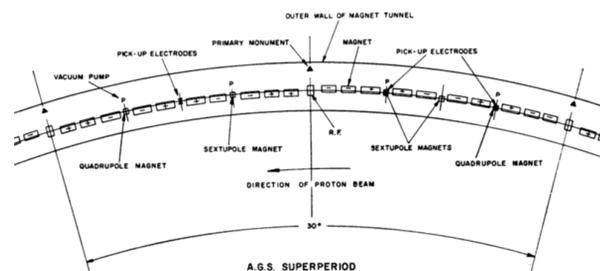


Figure 3: Drawing of the AGS superperiod which consist of 20 combined function main magnets. The “+” and “-” signs on each magnets indicate the focusing and defocusing quadrupole property of each combined function magnet for positive ions circulating in the synchrotron. There is one tune quadrupoles in each of the straight sections SS03,SS17 and one chromaticity sextupole in each of the straight sections SS07,SS13 of each superperiod.

There is one tune quadrupoles in each of the straight sections SS03,SS17 and one chromaticity sextupole in each of the straight sections SS07, SS13 of each superperiod. Three pickup electrodes are also located in the straight sections SS03, SS05, and SS13 respectively. Additional information on the AGS appears in Refs. [1,5].

## BEAM OPTICS OF AGS

The beam optics of the AGS for a charged particle beam which is not subject to space charge forces depends on the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

magnetic field of the magnetic elements of the AGS ring. These fields can be either calculated using various methods, or can be measured. In this study we make use of the measured magnetic fields of the AGS combined function magnets. These fields were measured on a grid located on the median plane of the combined function magnets as it is described in one of the following sections. In the next section we introduce the reader to the AGS main magnets.

### The AGS Combined Function Magnets

The 240 AGS main magnets are made of three type of magnets named A, B, and C. Magnets type A and B have the same cross section and lengths 90" and 75" respectively, and magnet C has a length of 90". The table in Fig. 4 lists the lengths of the three type of magnets and shows pictures of their cross sections. By rotating the magnets A, B, C, by 180° about the vertical axis the focusing property of the magnets changes from focusing to defocusing.

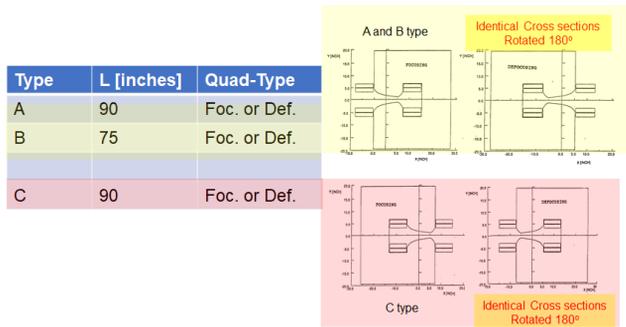


Figure 4: The three types A, B, and C of the AGS combined function magnets. Magnets type A and B have the same cross section but different lengths. By rotating the magnets by 180° about the vertical axis the focusing property of the magnets changes from focusing to defocusing.

Figure 5 is a picture of a C-type AGS combined function magnet.

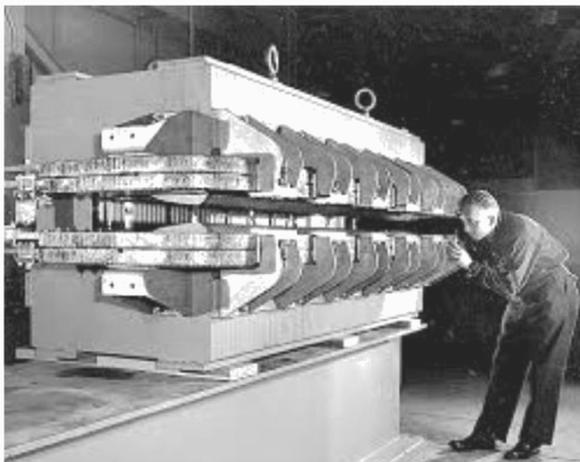


Figure 5: A picture of a C-type AGS combined function magnet.

### The Field Maps of the AGS Main Magnets

The field map of a single type A and a single type C magnet were measured on the median plane of each magnet. Figure 6 is a top view of the magnets showing the areas (ABCD) and (EFGH) over which the field maps of magnet type A were performed (picture on the left) and the area (ABCD) over which the field maps of magnet type C was performed (picture on the right). For magnet type A (left

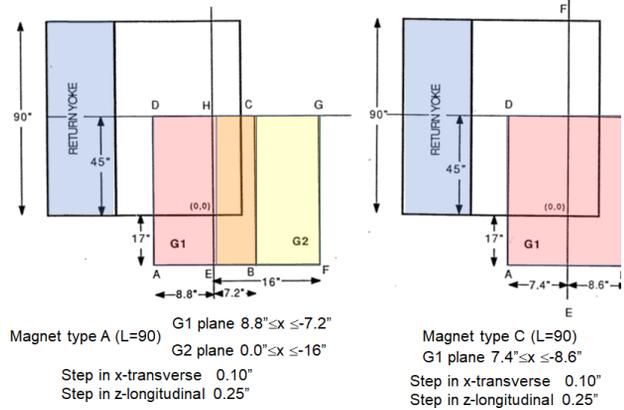


Figure 6: Top view of magnets type A (left) and C (right). The dimensions of the areas G1 (ABCD) and G2 (EFGH) over which the field maps were performed on each magnet appear in the figure.

picture in Fig. 6) two field maps were measured, the G1 over the area (ABCD) and G2 over the area (EFGH). The rectangular grid size of each field map was 0.1" in the transverse x-direction and 0.25" in the longitudinal z-direction. For magnet type C (right picture in Fig. 6) one field map G1 over the area (ABCD) was measured with a rectangular grid size of 0.1" in the transverse x-direction and 0.25" in the longitudinal z-direction. The dimensions of the areas over which the field maps were measured appear on Fig. 6. Although these field maps cover only half of the magnet's median plane, the symmetry of the magnet was used to complete the field map over the entire area of the magnet's median plane.

### Brief Description of the Algorithm

This section provides a brief description of the algorithm which is used in the RAYTRACE code [2] to calculate the field components of the AGS main magnets from the measured field maps on the median plane of the magnet.

Figure 7 shows the grid points (intersection points of the red lines) on a plane where the magnetic field components (yellow arrows) are measured. The algorithm calculates the magnetic field components (blue arrows) at any given point in space at a distance y from the plane. The algorithm is based on the Taylor series expansion of the magnetic field components at a point located at a distance y from the plane in terms of the y coordinate. This expansion is shown in Eq. (1).

$$B_i(x, y, z) = \sum_{j=0}^4 \frac{1}{j!} \frac{\partial^j B_i(x, y, z)}{\partial y^j} \Big|_{y=0} y^j = \sum_{j=0}^4 a_{ij}(x, z) y^j \quad (1)$$

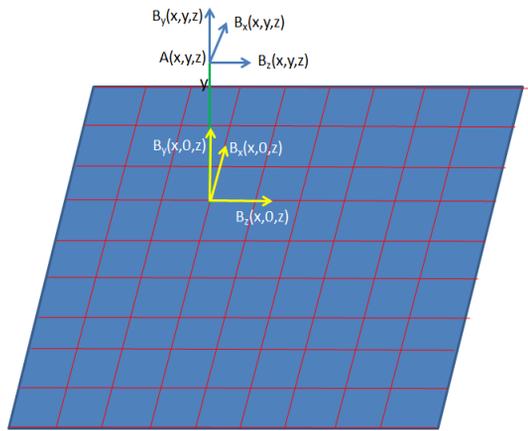


Figure 7: A schematic diagram of a grid on a plane. The magnetic field components (yellow arrows) at any grid point on the plane are measured. The algorithm calculates the magnetic field components (blue arrows) at any given point in space at a distance  $y$  from the plane.

Since the values of the field components on the plane ( $y=0$ ) are known the task is to express the coefficients of expansion  $a_{ij}(x, z)$  in terms of these field components on the plane and also in terms of their partial derivatives of these field components with respect to the  $x$  and  $z$  spatial coordinates. This is done by constraining the magnetic fields of Eq. (1) to satisfy the Maxwell's equations:

$$\vec{\nabla} \cdot \vec{B}(x, y, z) = 0 \quad \text{and} \quad \vec{\nabla} \times \vec{B}(x, y, z) = 0 \quad (2)$$

In **APPENDIX I** the coefficients  $a_{ij}$  are expressed in term of the known measured field components on the plane and their spatial derivatives with respect to  $x$  and  $z$ .

Reference [6] provides a detailed derivation of the coefficients  $a_{ij}$  in terms of the measured fields on the plane and their partial spatial derivatives.

## CLOSED ORBIT CALCULATIONS AND BEAM EXTRACTION

This section describes the procedure that was used to extract the beam bunches from the AGS and report the results of the beam optics just before and during the fast beam extraction process from the AGS. All the calculations are based on the numerical integration of the equation of motion of particles moving in the magnetic field which was derived from the experimentally measured field maps using the algorithm mentioned earlier. The RAYTRACE code was employed to calculate the fields and integrate the equation of motion of the particles in the AGS [1]. At the extraction energy and prior to the beam extraction two local beam bumps are generated in the AGS. One of the local beam bumps brings the beam inside the “G10” extraction kicker and the second local beam bump brings the beam close to the “H10” extraction septum. The green trace in Fig. 8 is the closed orbit in the AGS prior to fast beam extraction. The local “G10” and “H10” beam bumps are shown as part of the green trace. The

red trace corresponds to trajectory of the extracted beam. Details on the formation of the local extraction beam bumps “G10” and “H10” appear in Ref. [1].

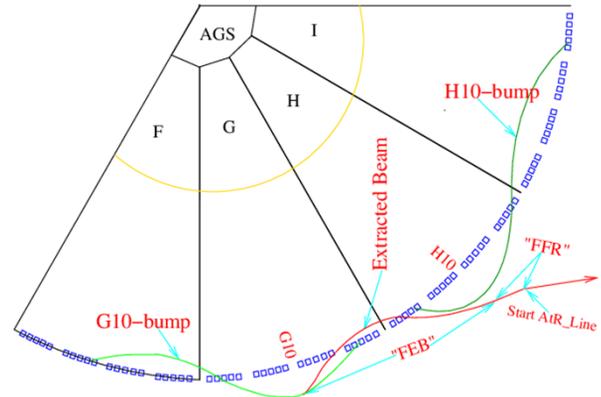


Figure 8: A schematic diagram of a section of the AGS. The green trace is the closed orbit in the AGS prior to fast beam extraction. The local “G10” and “H10” beam bumps are shown as part of the green trace. The red trace corresponds to trajectory of the extracted beam.

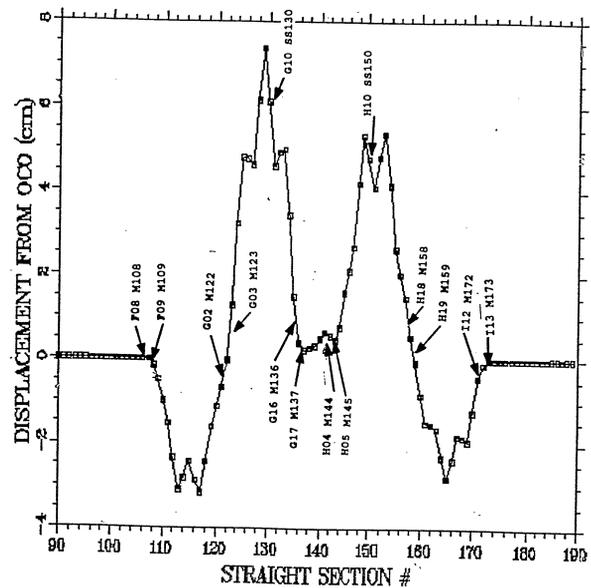


Figure 9: Displacement of the closed orbit in the straight sections starting from (SS90) and ending at (SS190). Both the local beam bumps G10 and H10 are on. The label at the beginning of each small arrow is the name of the combined function magnet.

Figure 9 shows the beam displacement at the middle of the straight sections of the AGS when the closed orbit extraction bumps “G10” and “H10” are excited. The label at the beginning of the small arrow corresponds to the name of the combined function magnet.

Figure 10 shows the horizontal  $\beta_x$  (left) and  $\beta_y$  (right) functions of the closed orbit at the middle of the straight sections with the extraction bumps “G10” and “H10” on.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

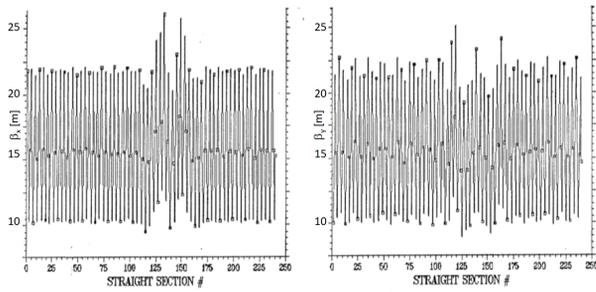


Figure 10: The  $\beta_x$  (left) and  $\beta_y$  (right) functions at the middle of the straight sections of the AGS with the extraction bumps “G10” and “H10” on. The line is to guide the eye.

Figure 11 shows the horizontal  $\eta_x$  function of the closed orbit at the middle of the straight sections with the extraction bumps “G10” and “H10” on.

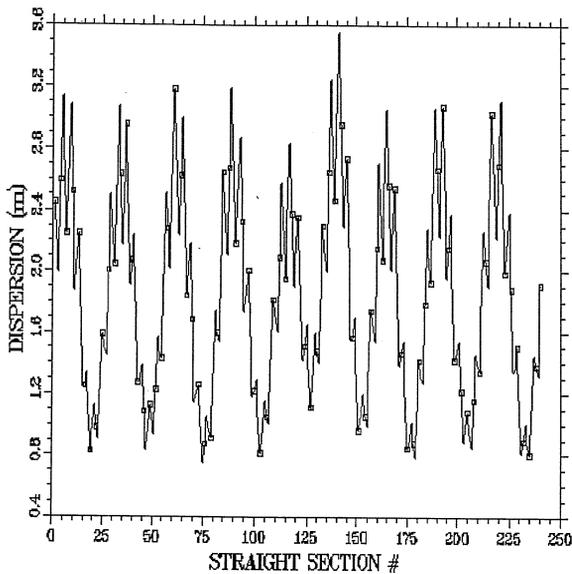


Figure 11: The horizontal  $\eta_x$  function at the middle of the straight sections of the AGS with the extraction bumps “G10” and “H10” on. The line is to guide the eye.

## THE FAST BEAM EXTRACTION

Upon the formation of the closed orbit in AGS with the local “G10” and “H10” beam bumps on as described in the previous section, the bunched beam can be extracted from the AGS by energizing the G10 fast kicker which generates a magnetic field pulse of duration  $\sim 200$  nsec and provides a kick to the beam of 1.5 mrad. The location of the G10 kicker is shown in Fig. 12 and the red trace initiating from the G10 kicker represents the extracted beam which is displaced at the location of the H10 septum by  $\sim 6$  cm to enter in the main field region of the H10 septum which deflects the beam by  $\sim 20$  mrad to extract it from the AGS. The table in Fig. 12 shows the R-matrix elements between the location of the G10 kicker and the start of the AGS-to-RHIC (AtR) transfer line shown in Fig. 12. The AtR line is the beam transfer line

from the AGS to the Relativistic Heavy Ion Collider (RHIC). The knowledge of this R-matrix and the knowledge of the beta and eta functions from the closed orbit calculations at the location of the G10 kicker provide the values of the beta and eta functions at the beginning of the AtR line as shown by the matrix equations at the bottom of Fig. 12. Again the R-matrix from the location of the G10 kicker to the beginning of the AtR line was calculated by raytracing many rays in the magnetic field which was computed from the measured field maps using the algorithm mentioned earlier.

R Matrix G10-kicker to Start-AtR					
-1.8261	-17.443	0.0	0.0	0.0	-2.5889
0.17912	1.1634	0.0	0.0	0.0	0.2517
0.0	0.0	1.125	-15.016	0.0	0.0
0.0	0.0	0.186	-1.5855	0.0	0.0
-0.0041	1.3785	0.0	0.0	1	0.0
0.0	0.0	0.0	0.0	0.0	1.0

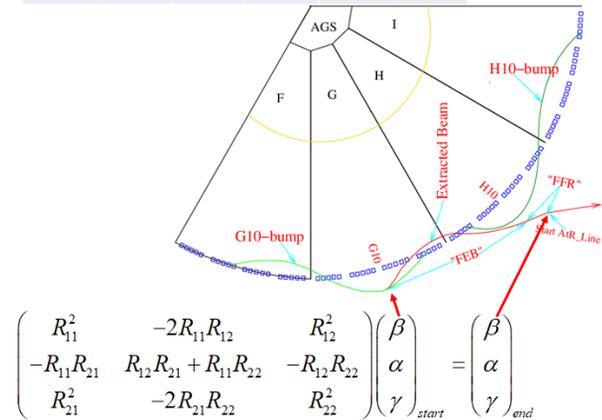


Figure 12: The red trace initiating from the location of the G10 kicker represents the trajectory of the reference orbit of the extracted beam. The table on top of the figure is the R-matrix between the location of the G10 kicker and the beginning of the AtR line.

## CONCLUSION

From measured field maps at the median plane of the AGS combined function main magnets, the magnetic fields in circulating beam region were calculated by the use of an algorithm discussed in this paper. From these fields the beam parameters ( $\beta_{x,y}$  and  $\eta_{x,y}$ ) of the closed orbit of the circulating beam in AGS beam were calculated. Subsequently the extraction kicker “G10” of the AGS was energized and the beam was extracted from the AGS. Using the magnetic fields along the extraction channel the R-matrix elements from the location of the G10 kicker to the beginning of the AtR line were calculated and therefore the beam parameters at the beginning of the AtR line were calculated.

## APPENDIX I

**Expression of the coefficients  $a_{ij}$**  Applying Maxwell’s Equations (2) on the magnetic field which is given by Equation (1) the following expressions of the coefficients  $a_{ij}$  are derived.

### The 0<sup>th</sup> order $a_{i0}$ coefficients

$$a_{x0} = B_x(x, 0, z), a_{y0} = B_y(x, 0, z), a_{z0} = B_z(x, 0, z) \quad (3)$$

### The 1<sup>st</sup> order $a_{i1}$ coefficients

$$a_{x1} = \frac{\partial B_y(x, 0, z)}{\partial x} \quad (4)$$

$$a_{y1}(x, z) = - \left( \frac{\partial B_x(x, 0, z)}{\partial x} + \frac{\partial B_z(x, 0, z)}{\partial z} \right) \quad (5)$$

$$a_{z1} = \frac{\partial B_y(x, 0, z)}{\partial z} \quad (6)$$

### The 2<sup>nd</sup> order $a_{i2}$ coefficients

$$a_{x2} = -\frac{1}{2} \left( \frac{\partial^2 B_x(x, 0, z)}{\partial x^2} + \frac{\partial^2 B_z(x, 0, z)}{\partial x \partial z} \right) \quad (7)$$

$$a_{y2} = -\frac{1}{2} \left( \frac{\partial^2 B_y(x, 0, z)}{\partial x^2} + \frac{\partial^2 B_y(x, 0, z)}{\partial z^2} \right) \quad (8)$$

$$a_{z2} = -\frac{1}{2} \left( \frac{\partial^2 B_x(x, 0, z)}{\partial z \partial x} + \frac{\partial^2 B_z(x, 0, z)}{\partial z^2} \right) \quad (9)$$

### The 3<sup>rd</sup> order $a_{i3}$ coefficients

$$a_{x3} = -\frac{1}{6} \left( \frac{\partial^3 B_y(x, 0, z)}{\partial x^3} + \frac{\partial^3 B_y(x, 0, z)}{\partial x \partial z^2} \right) \quad (10)$$

$$a_{y3} = \frac{1}{6} \left( \frac{\partial^3 B_x(x, 0, z)}{\partial x^3} + \frac{\partial^3 B_z(x, 0, z)}{\partial x^2 \partial z} + \frac{\partial^3 B_x(x, 0, z)}{\partial z^2 \partial x} + \frac{\partial^3 B_z(x, 0, z)}{\partial z^3} \right)$$

$$a_{z3} = -\frac{1}{6} \left( \frac{\partial^3 B_y(x, 0, z)}{\partial z \partial x^2} + \frac{\partial^3 B_y(x, 0, z)}{\partial z^3} \right) \quad (11)$$

### The 4<sup>th</sup> order $a_{i4}$ coefficients

$$a_{x4} = \frac{1}{24} \left( \frac{\partial^4 B_x(x, 0, z)}{\partial x^4} + \frac{\partial^4 B_z(x, 0, z)}{\partial x^3 \partial z} + \frac{\partial^4 B_x(x, 0, z)}{\partial x^2 \partial z^2} + \frac{\partial^4 B_z(x, 0, z)}{\partial x \partial z^3} \right) \quad (12)$$

$$a_{y4} = -\frac{1}{24} \left( \frac{\partial^4 B_y(x, 0, z)}{\partial x^4} + 2 \frac{\partial^4 B_y(x, 0, z)}{\partial x^2 \partial z^2} + \frac{\partial^4 B_y(x, 0, z)}{\partial z^4} \right) \quad (13)$$

$$a_{z4} = \frac{1}{24} \left( \frac{\partial^4 B_x(x, 0, z)}{\partial z \partial x^3} + \frac{\partial^4 B_z(x, 0, z)}{\partial x^2 \partial z^2} + \frac{\partial^4 B_x(x, 0, z)}{\partial z^3 \partial x} + \frac{\partial^4 B_z(x, 0, z)}{\partial z^4} \right) \quad (14)$$

### Calculation of the Partial Field Derivatives

The partial derivatives of the field components appearing in the expression of the coefficients  $a_{ij}$  can be calculated either numerically or using the “fit a function method” which is the method used in the studies discussed in Refs. [1, 4]. Both methods are discussed in details in Ref. [6].

**Fit a Function method** Figure 13 shows the grid points of the global  $(x, z)$  coordinate system where the field components  $[B_x(x, y = 0, z), B_y(x, y = 0, z), B_z(x, y = 0, z)]$  are measured, and also shows two of the many “small-grid-areas” (ABCD), (EFGH) in which the global grid is partitioned.

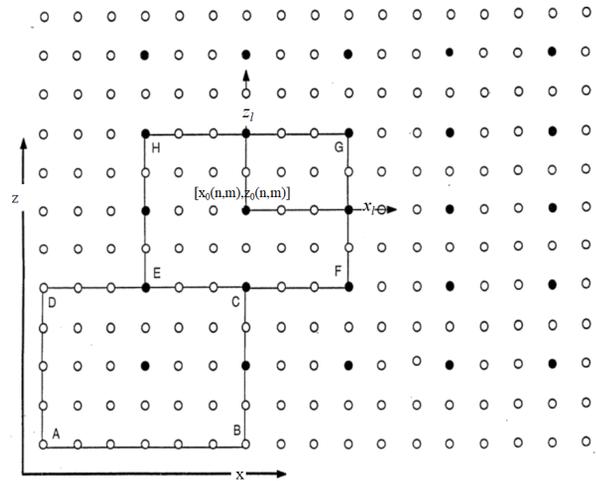


Figure 13: The large 2D grid in the  $(x, z)$  coordinate system is partitioned in many “small-grid-areas” two of these small areas (ABCD), (EFGH) are shown. The small grid areas may overlap, and each area can be characterized by the  $(n, m)$  indices and its local  $(x_l, y_l)$  coordinate system. The local coordinates  $(x_l, y_l)$  of each small grid are related to the global  $(x, z)$  coordinates through the equations  $x = x_0(n, m) + x_l$  and  $z = z_0(n, m) + z_l$ .

The small grid areas may partially overlap with each other, and each area can be characterized by the  $(n, m)$  indices and its associated  $(x_l, y_l)$  local coordinate system. With this method a polynomial function shown in Eq. (15), is fitted to the experimentally measured values of the magnetic field components. The local coordinates  $(x_l, y_l)$  of each small grid are related to the global coordinates  $(x, z)$  through the equations  $x = x_0(n, m) + x_l$  and  $z = z_0(n, m) + z_l$ , where  $x_0(n, m)$  and  $z_0(n, m)$  are the global coordinates of the centers  $(x_l, y_l) = (0, 0)$  of the “small-grid-area” which is characterized by the indices  $(n, m)$ .

$$B_i(n, m, x_l, z_l) = \sum_{j=0}^4 \sum_{k=0}^4 c_{i,n,m,j,k} (x_l)^j (z_l)^k \quad (15)$$

In Eq. (15) the index  $i$  corresponds to the field component with  $i = (1, 2, 3) \Leftrightarrow (x, y, z)$  and the indices  $n, m$  correspond to the particular “small-grid-area” of the global grid. The  $x_l, z_l$  variables are the local coordinates of this “small-grid-area”

which are related to the global coordinates  $(x, z)$  through equations mentioned earlier. The coefficients  $c_{i,n,m,j,k}$  are calculated using the method of Singular Value Decomposition (SVD) [7] which is applied to solve  $N$  equations with  $M$  unknowns ( $N \geq M$ ). This method of fitting a function to the experimentally measured field components at the grid points of a rectangular grid on the median plane of a magnet has been used in the RAYTRACE code to calculate the beam optics of the AGS synchrotron [1] using the median plane field maps of the AGS combined function magnets.

## REFERENCES

- [1] N. Tsoupas *et al.*, “Closed orbit calculations at AGS and Extraction Beam Parameters at H13,” AD/RHIC/RD-75, Oct. 1994.
- [2] S. B. Kowalski and H. A. Enge, “The Ion-Optical Program Raytrace,” *NIM A*, vol. 258, pp. 407, 1987.
- [3] K. Makino, M. Berz, and C. Johnstone, “High-Order Out-Of-Plane Expansion for 3D Fields,” *Int. Journal of Modern Physics A*, vol. 26, pp. 1807-1821, 2011.
- [4] N. Tsoupas *et al.*, “Effects of Dipole Magnet Inhomogeneity on the Beam Ellipsoid,” *NIM A*, vol. 258, pp. 421-425, 1987.
- [5] “Alternating Gradient Synchrotron Project Construction Completion Report,” CAD report BNL 1966
- [6] N. Tsoupas *et al.*, “Algorithm to calculate off-plane magnetic field from an on-plane field map,” C-A/AP/605, May 2018.
- [7] W. H. Press *et al.*, “Singular Value Decomposition”, in *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, MA, USA.

# ANALYTICAL CALCULATIONS FOR THOMSON BACKSCATTERING BASED LIGHT SOURCES

P. Volz\*, A. Meseck, Helmholtz-Zentrums Berlin für Materialien und Energie, Berlin, Germany

## Abstract

There is a rising interest in Thomson-backscattering based light sources, as scattering intense laser radiation on MeV electrons produces high energy photons that would require GeV or even TeV electron beams when using conventional undulators or dipoles. Particularly, medium energy high brightness beams delivered by LINACs or Energy Recovery LINACs, such as bERLinPro being built at Helmholtz-Zentrum Berlin, seem suitable for these sources. In order to study the merit of Thomson-backscattering-based light sources, we are developing an analytical code to simulate the characteristics of the Thomson scattered radiation. The code calculates the distribution of scattered radiation depending on the incident angle and polarization of the laser radiation. Also the impact of the incident laser polarization and the full 6D bunch profile, including microbunching, are incorporated. The Status of the code and first results will be presented.

## INTRODUCTION

Shortly after the invention of the LASER the idea of Thomson-backscattering light sources emerged [1]. Only in recent years however did lasers become powerful enough to make these sources viable due to the small Thomson scattering cross section. In the case of Thomson backscattering a relativistic electron beam interacts with a counter-propagating laser field. The backscattered photons travel in the direction of the electron beam in a small cone with an opening angle proportional to  $1/\gamma$ . The scattered laser photons experience a Doppler shift according to the energy of the electrons they are scattered on. This allows Thomson backscattering sources to produce very high energy photons, from relatively low energy electron beams, that would otherwise require GeV electron energies. Thomson scattering is the low energy limit of Compton scattering. The Thomson limit is accurate if the photon energy in the particle's rest frame is significantly lower than its rest mass.

Nowadays the demand for beam time at hard X-ray synchrotron facilities heavily outweighs supply. Such facilities however are very cost prohibitive to build and operate. Thomson-backscattering light sources provide an alternative to conventional sources at a cost that would be manageable for smaller laboratories and universities. Furthermore, in recent years there have been advances in the development of high brightness electron beam sources in both classical linac, and energy recovery linac (ERL) configuration. These sources provide electron beams with very low energy spread and emittance which results in less quality degradation of the backscattered light. This has opened up new possibilities

for high performance ERL based Thomson backscattering light sources. The design and development of such a source requires a fast code that takes into account the relevant properties of the electron beam and laser pulse to calculate the critical properties of the backscattered radiation field. The development of such a code is the goal of the presented work.

First we will present a short description of our code, followed by some test cases to validate its results. Then we will present our first results in simulating the radiation of microbunched beams. Finishing off with an outlook of what improvements are planned.

## CODE DESCRIPTION

The goal of our code is to calculate the spatial and spectral radiation distribution for different Thomson scattering events. It has to include the laser polarization, the incident angle between the laser and the electron beam, and the coherent effects resulting from the full 6D bunch profile of the beam. As of now emittance cannot be fully incorporated by the code because evaluation of transverse momentum distribution is not yet implemented. The laser is treated as a flat top pulse with no rise time or fringe effects. The 6D particle configuration can be imported from ascii files in ASTRA [2] output format. The underlying calculations are based on an Evaluation of Liénard-Wiechert potentials. The derivation of the formulas is shown in detail in [3] and [4]. The code does not use numerical integration. The essential calculations are based on complex Bessel functions.

Our code can calculate the spatial intensity distribution of the radiation produced by electrons interacting with a circularly or linearly polarized laser. The incident angle between the laser and the electron beam as well as the detector size and position can be chosen freely. The detector is modeled as grid of pixels. The number of pixels together with the size of the detector dictates the resolution. The backscattered radiation generated by individual particles are added up at each pixel. For multiple particles either the intensities or the amplitude of the radiation generated by the individual particles can be added. As will be explained in the following sections, the addition of amplitudes is necessary for the correct calculation of coherent effects.

A simulation run with 200 k particles and a detector resolution of  $80 \times 80$  pixels takes around 2000 s on a current workstation CPU (single core load). The computation time scales linearly with the number of particles.

## CODE VALIDATION

To validate our code we simulated some simple scenarios for which we have a clear expectation of results.

\* paul.volz@helmholtz-berlin.de

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## Thomson Scattering on Single Electrons

The simplest case is the scattering of laser radiation on a single electron. For this case there are a number of simulations and analytical studies available, e.g. [3,4]. In particular, the spatial radiation distribution of scattering on a single electron for both circular and linear polarization of the incident laser has been derived in [3] and [4]. In [3], figure 2 shows the intensity distribution as a function of photon energy and polar angle to the detector. Our code can produce those plots as shown in Fig. 1. Our code also reproduces

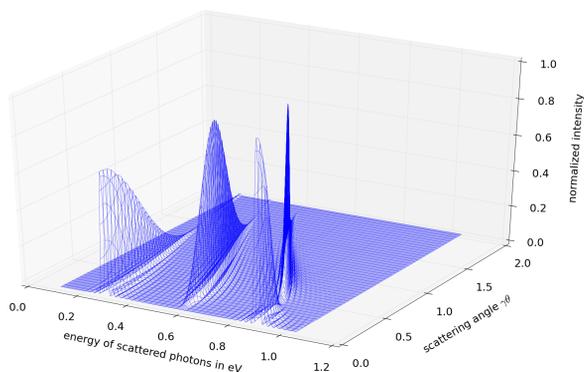


Figure 1: Normalized intensity as a function of scattered photon energy and scattering angle  $\gamma\theta$  of the radiation scattered by a relativistic electron ( $\gamma = 69.5$ ) from a counterpropagating linearly polarized laser pulse.

the results for both polarizations and for different incident angles shown in figures 3, 4, 6, 7, and 9 in [4]. Figure 2 shows the results for a linearly polarized laser pulse and Fig. 3 shows the circularly polarized case.

### Coherent Addition

When looking at the radiation produced by more than one electron we need to make the distinction between coherent and incoherent addition. In the incoherent case the total intensity is just a superposition of intensities generated by scattering on single electrons. This case is examined for example by P. Tomassini et al. [5]. If we want to fully investigate the effects of the electron bunch structure we also have to take into account coherent addition. In this case we have to consider the complex amplitudes of the radiation produced by each electron. This allows us to take into account the phase of the propagating radiation and to calculate interference at the detector accordingly. In our code the phase is calculated by analyzing the path length to every detector pixel for the radiation field generated by each electron together with its frequency. The Thomson backscattering process can be compared to an electron traveling through an undulator. This way we can form some expectations based on what we know about coherent addition of electron spectra in undulators. There are however some key differences between an undulator field and the one produced by a laser pulse. The field of an undulator is constant in time. Two electrons passing through the undulator will witness the same field at a given point in space. The laser field on the

other hand oscillates in both time and space as the pulse propagates. This introduces another phase factor that needs to be taken into account. As a proof of concept we simulated the case of two electrons at different longitudinal distances from each other.

For two electrons at the same position we expect a four-fold increase in intensity when adding coherently, compared to a two-fold increase for incoherent addition. Figure 4 shows the comparison between the intensities of radiation produced by a single electron and by two identical electrons treated coherently. As expected we see a four fold increase.

For two electrons half a wavelength apart the naive expectation considering an undulator field would be a phase difference of  $180^\circ$  and therefore zero intensity at the center of the detector. Away from the detector center the path length and therefore the phase difference changes so we would expect an interference pattern. Figure 5 shows the spectrum for the described case. In the center of the detector there is an intensity peak instead of a minimum. This is due to the additional phase factor of the laser field. If we omit this phase factor we get the expected result from an undulator case as shown in Fig. 6. The same test can be done for constructive interference with a phase difference of  $360^\circ$ . This means the two electrons are a full wave length apart. Figure 7 shows the coherent addition of scattered radiation including the laser phase factor and Fig. 8 shows the case where the laser phase has been omitted. This shows that the code produces the expected results for coherent addition of electron spectra and we can proceed to investigate microbunching effects.

## RESULTS

As a first test we simulate a circularly polarized laser pulse scattering head on with an electron bunch comprised of 200 k particles. The electron bunch and laser parameters are listed in Table 1. We compare three different electron beams, two

Table 1: Parameters Used in Simulations

Electron Bunch	
energy	35.5 MeV
$\gamma$	69.5
Laser Pulse	
$\lambda$	2.665 cm
number of periods in pulse	7
pulse duration	622 ps
$a_0$	2
intensity	7.79 GW/cm <sup>2</sup>

microbunched beams with different energy spreads and a Gaussian beam with similar energy spread. To accentuate the effects of microbunching all three beams had no transverse momentum or spread. The electric field amplitudes of the radiation generated by the single particles were added to correctly incorporate coherent effects. This causes a lot of noise because most of the radiation gets canceled out by

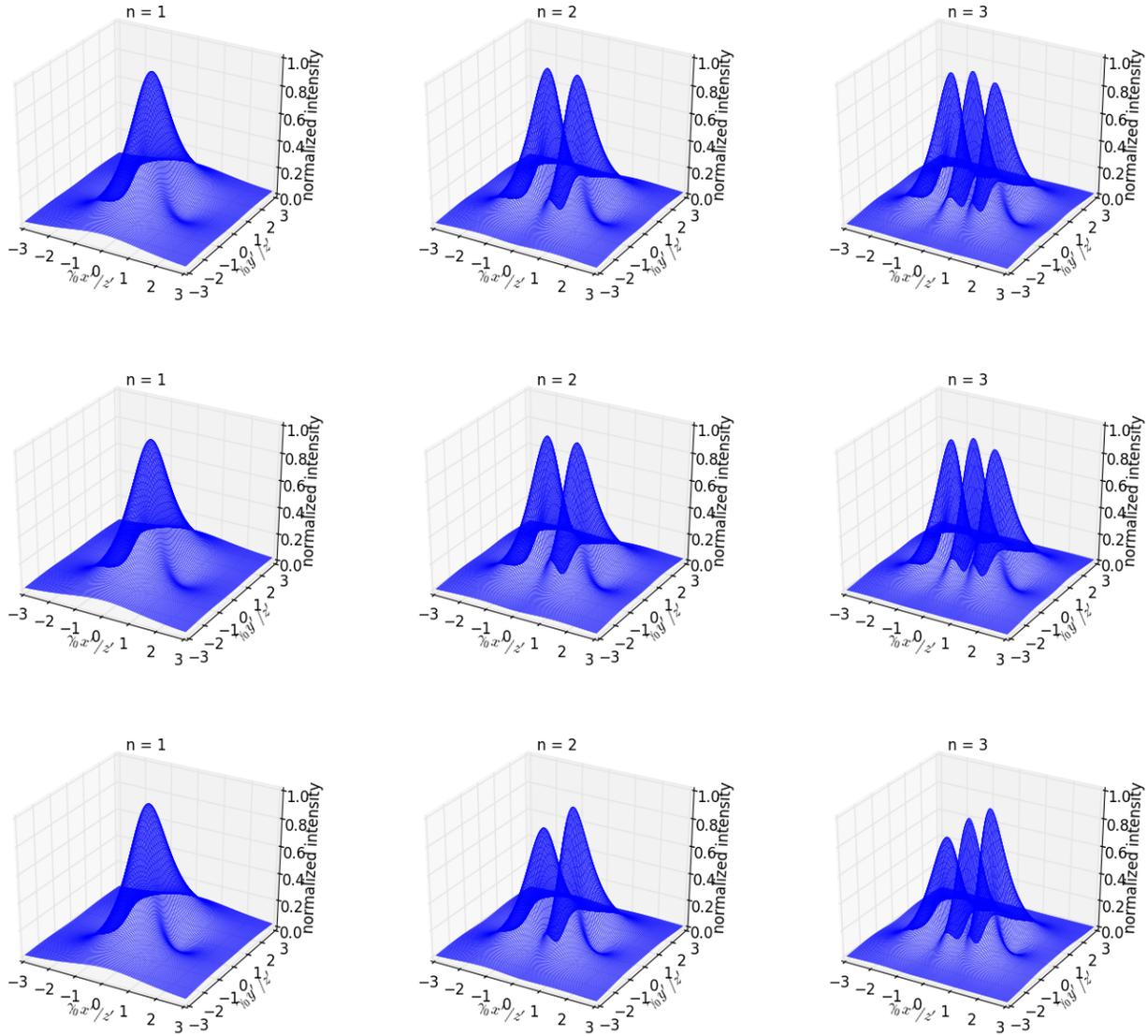


Figure 2: Normalized Intensity of radiation scattered by a relativistic electron ( $\gamma = 10$ ) from a high intensity ( $a_0 = 2$ ) linearly polarized laser pulse, viewed in plane of the detector. The detector is located at  $z'$  and centered on the electron beam axis. Distances in  $x', y'$  are measured in units  $\gamma_0(x'/z'), \gamma_0(y'/z') \propto \gamma_0\theta$ . Head-on scattering on top, transverse scattering with the electron moving perpendicular to the laser's plane of polarization in the middle, transverse scattering with the electron moving in the laser's plane of polarization on the bottom. The first three harmonics are shown.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

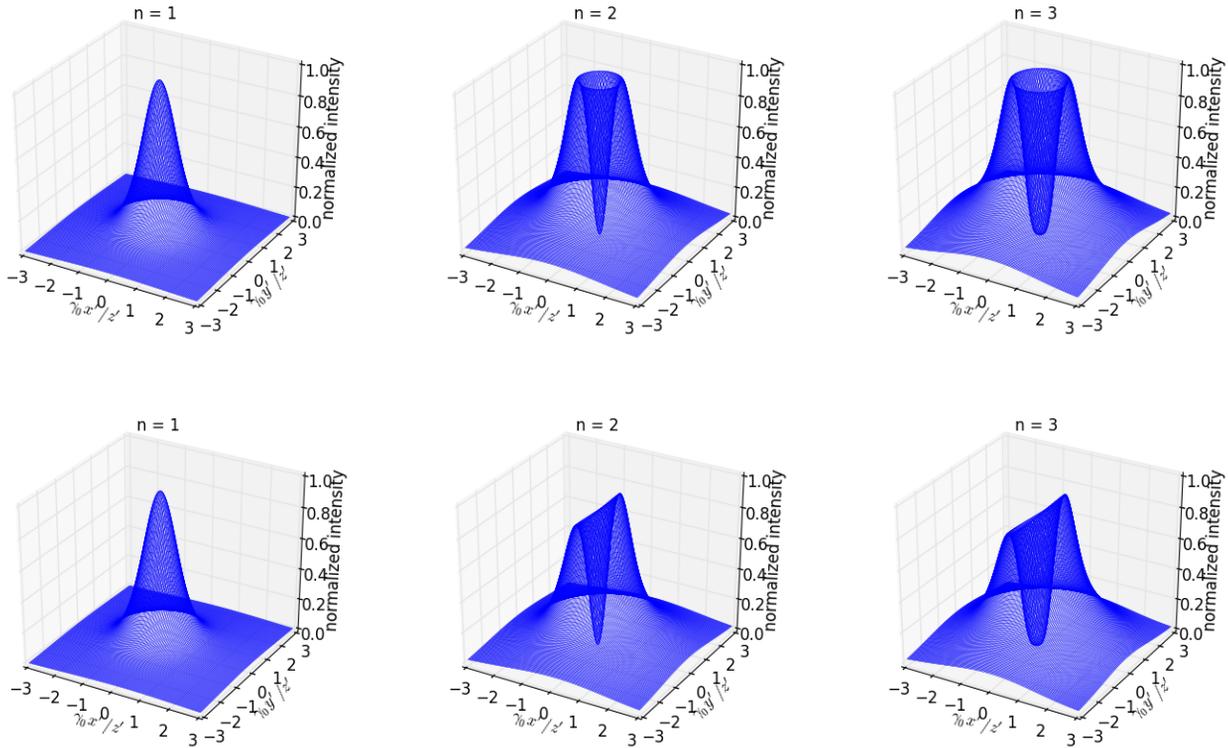


Figure 3: Normalized Intensity of radiation scattered by a relativistic electron ( $\gamma = 10$ ) from a high intensity ( $a_0 = 2$ ) circularly polarized laser pulse, viewed in plane of the detector. The detector is located at  $z'$  and centered on the electron beam axis. Distances in  $x', y'$  are measured in units  $\gamma_0(x'/z'), \gamma_0(y'/z') \propto \gamma_0\theta$ . Head-on scattering on top, transverse scattering on the bottom. The first three harmonics are shown.

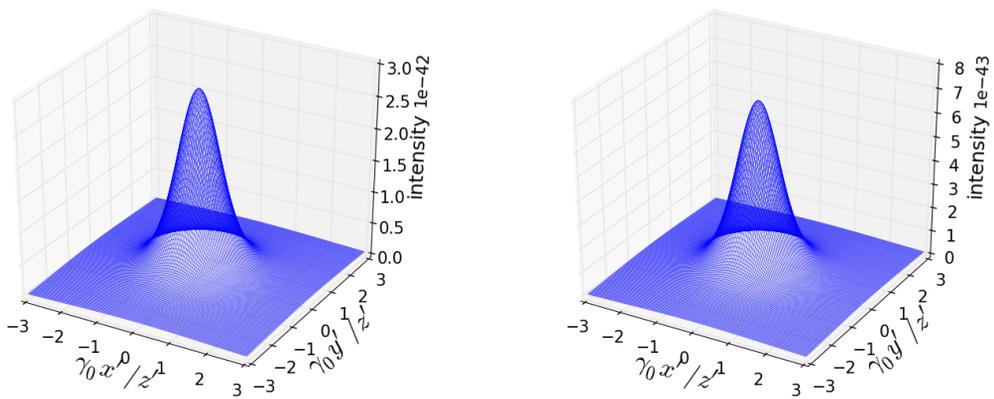


Figure 4: Intensity of radiation in arbitrary units produced by a single electron on the right and by two identical electrons treated coherently on the left. The coherent addition of radiation increases the intensity four-fold.

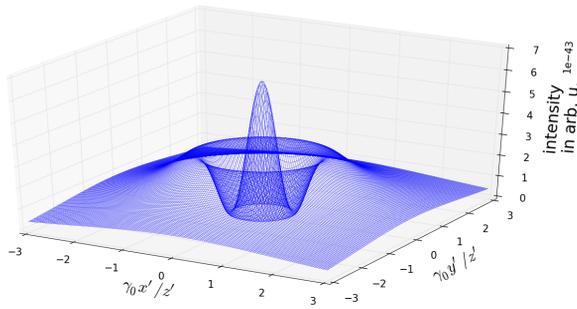


Figure 5: Two electrons with a longitudinal distance of half a wavelength. Laser wavelength of 500 nm and  $\gamma = 10$ .

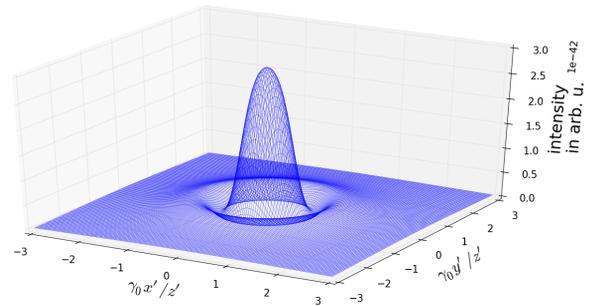


Figure 8: Two electrons with a longitudinal distance of a full wavelength. Laser wavelength of 500 nm and  $\gamma = 10$ . The laser phase factor has been omitted.

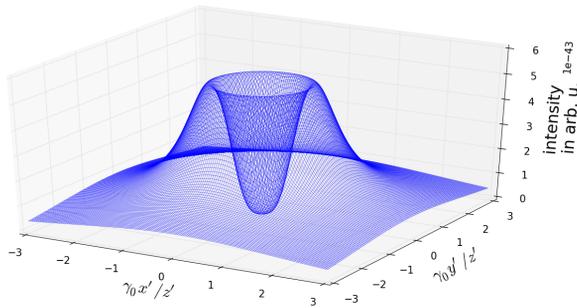


Figure 6: Two electrons with a longitudinal distance of half a wavelength. Laser wavelength of 500 nm and  $\gamma = 10$ . The laser phase factor has been omitted.

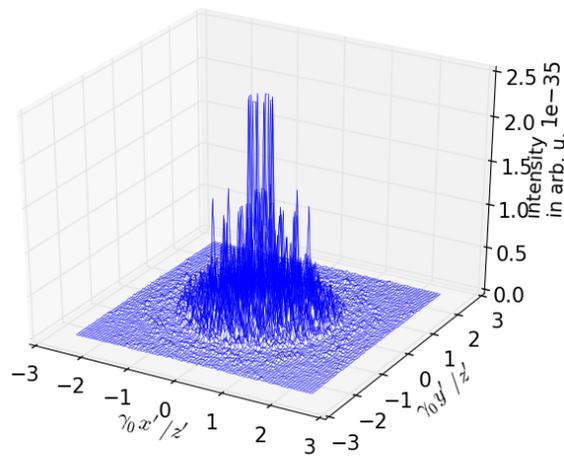


Figure 9: Coherent addition of radiation generated by electrons from a bunch with a Gaussian longitudinal distribution and energy spread.

phase difference to radiation generated by other particles in the bunch thus fewer particles effectively contribute to the spectrum. This noise is amplified by the pseudo-random noise reduction algorithms that most particle tracking codes use to generate their electron bunches. In this case the bunch was generated with ASTRA, which uses Hammersley sets. Figure 9 shows the spacial intensity distribution of the radiation generated by a Gaussian electron bunch when the field amplitudes are summed up. As a comparison Fig. 10 shows the incoherent intensity generated by the same bunch.

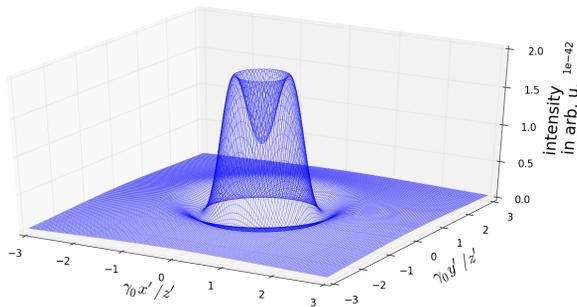


Figure 7: Two electrons with a longitudinal distance of a full wavelength. Laser wavelength of 500 nm and  $\gamma = 10$ .

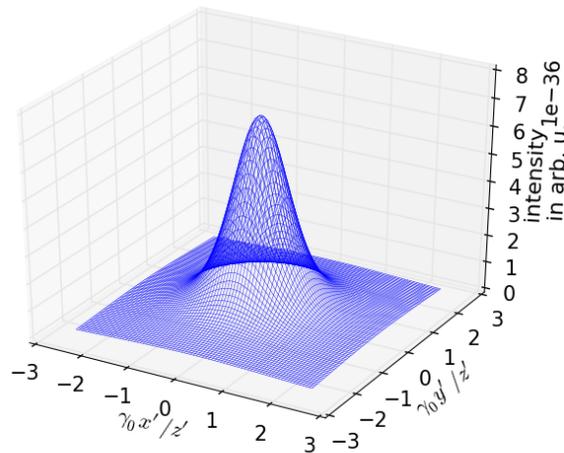


Figure 10: Incoherent addition of radiation generated by electrons from a bunch with a Gaussian longitudinal distribution and energy spread.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## Microbunching

Microbunching refers to a density modulation within an electron bunch. This allows for a coherent emission of photons at distinct frequencies and therefore greatly increased intensity. To test the feasibility of microbunching in Thomson backscattering sources, we simulated two microbunched beams with different energy spreads. The energy spread and longitudinal particle distribution of one case is shown in Fig. 11. The corresponding spatial intensity distribution of the coherent radiation is shown in Fig. 12. Figure 13 shows the case with higher energy spread. A higher energy spread seems to be detrimental to the intensity. However, compared to Fig. 9 the differences aren't immediately obvious.

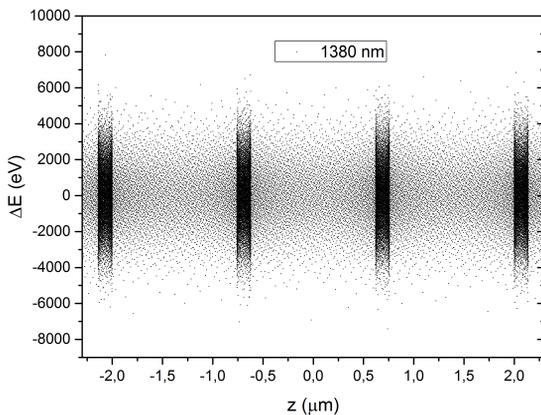


Figure 11: Energy spread and longitudinal electron distribution of the simulated bunch.

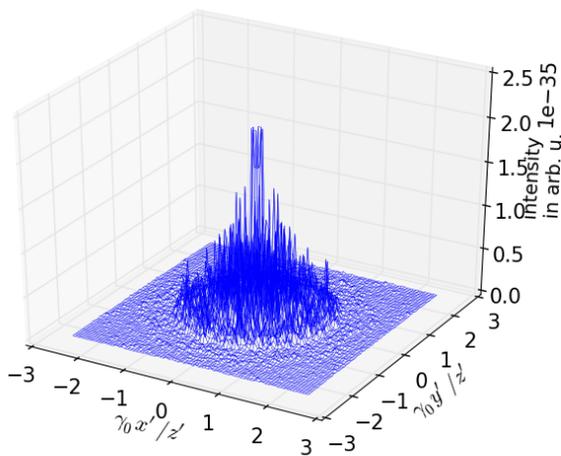


Figure 12: Coherent addition of radiation generated by electrons from a microbunched beam and a relatively small energy spread.

## CONCLUSION AND OUTLOOK

A fast analytical code has been developed to simulate coherent radiation interactions in Thomson backscattering

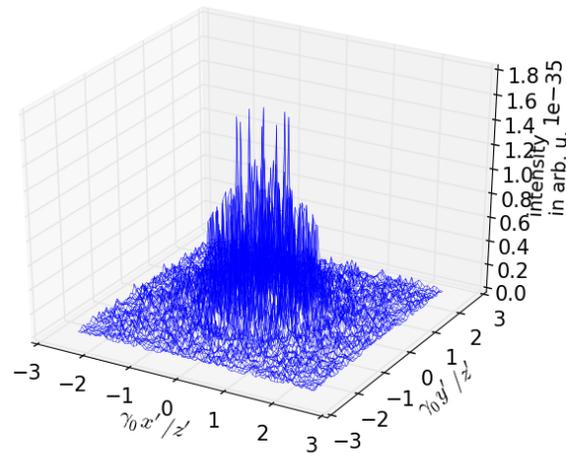


Figure 13: Coherent addition of radiation generated by electrons from a microbunched beam and a relatively large energy spread.

events. From simple tests we can conclude that our code is able to correctly calculate the coherent interaction of radiation produced by different particles as well as the incoherent superposition of Intensity. The simulated bunches have very strong microbunching of 60 %. It is possible that the resonance is not exactly hit by the laser or that most of the radiation cancels out due to the aforementioned pseudo-random methods of bunch generation. The next step is to fully implement emittance of electron bunches. The noise level for coherent treatment of many particles needs to be improved as well before first design studies utilizing our code can start.

## ACKNOWLEDGEMENTS

We would like to thank Michael Scheer for the fruitful discussions about coherent addition of synchrotron radiation.

## REFERENCES

- [1] C. Bemporad *et al.*, “High-Energy Photons from Compton Scattering of Light on 6.0-GeV Electrons”, *Phys. Rev. B*, vol. 138, issue 6, pp. B1546–B1549, 1965.
- [2] Klaus Floettmann, *A Space Charge Tracking Algorithm*, <http://www.desy.de/~mpyf1o/>
- [3] E. Esarey, S. K. Ride, *et al.*, “Nonlinear Thomson scattering of intense laser pulses from beams and plasmas”, *Phys. Rev. E*, vol. 48, issue 4, pp. 3003–3021, 1993.
- [4] S. K. Ride, E. Esarey, *et al.*, “Thomson scattering of intense lasers from electron beams at arbitrary interaction angles”, *Phys. Rev. E*, vol. 52, issue 5, pp. 5425–5442, 1995.
- [5] P. Tomassini *et al.*, “Thomson backscattering X-rays from ultra-relativistic electron bunches and temporally shaped laser pulses”, *Appl. Phys. B*, vol. 80, issue 4–5, pp. 419–436, 2005.

# A HOLISTIC APPROACH TO SIMULATING BEAM LOSSES IN THE LARGE HADRON COLLIDER USING BDSIM

S. D. Walker\*, A. Abramov, S. T. Boogert,  
H. Garcia-Morales, S. M. Gibson, H. Pikhartova, W. Shields, L. J. Nevay  
Royal Holloway, University of London, Egham, TW20 0EX, United Kingdom

## Abstract

To fully understand the beam losses, subsequent radiation, energy deposition, backgrounds and activation in particle accelerators, a holistic approach combining a 3-D model, physics processes and accelerator tracking is required. Beam Delivery Simulation (BDSIM) is a program developed to simulate the passage of particles, both primary and secondary, in particle accelerators and calculate the energy deposited by these particles via material interactions using the Geant4 physics library. A Geant4 accelerator model is built from an optical description of a lattice by procedurally placing a set of predefined accelerator components. These generic components can be refined to an arbitrary degree of detail with the use of user-defined geometries, detectors, field maps, and more. A detailed model of the Large Hadron Collider has been created in BDSIM, validated with existing tracking codes and applied to study beam loss patterns.

## INTRODUCTION

The Large Hadron Collider (LHC) at CERN is at the forefront of the accelerator energy frontier, with a design energy of 7 TeV and with a stored energy of 386 MJ per beam [1]. This extremely large stored energy presents a challenge to protect the experiments, and machine elements both from irradiation and prevent any superconducting magnets from quenching, where as little as  $1 \text{ mJ cm}^{-3}$  is sufficient to cause a quench [2]. Beam losses are inevitable in any machine and it is due to the aforementioned factors that a dedicated collimation system has been designed and built. It is primarily located in two insertion regions (IRs)—IR3 for momentum cleaning, and IR7, for betatron cleaning. Common to both is the concept of a collimation hierarchy, which consists of a sequence of collimators with increasing apertures, such that large amplitude particles will first hit the primary (smallest aperture) collimator, followed by the secondary collimators (wider aperture), and finally the absorbers (larger still). Added to this are tertiary collimators (TCTs) on either side of the experimental IRs, which protect the final focus magnets and reduce beam-induced backgrounds. This design has proven exceedingly successful in protecting the machine.

Detecting beam losses reliably in critical regions requires the presence of 3600 beam loss monitors (BLMs) placed around the ring [2]. These are used to detect abnormal beam conditions, and if one detects losses above a given threshold, a beam dump is triggered. In order to characterise the pattern of losses around the ring and study the effectiveness of the

collimation system, special runs are performed where a low-intensity beam is blown up to produce losses and the signal from the BLMs is recorded. This is referred to as a loss map of the machine.

To ensure the collimation system works effectively both in normal-functioning as well as in adverse scenarios, such as an asynchronous beam dump, effective simulation tools are necessary. The tool of choice used at CERN for collimation studies is SixTrack, and is used to generate loss maps [3]. SixTrack is a fully symplectic 6-D thin lens tracking code which was originally used for dynamic aperture studies, but was extended for use in aiding the design and implementation of LHC collimation system [4]. SixTrack is often paired with another standard CERN code, FLUKA [5], for irradiation and beam background studies to study specific areas of interest. Together these have demonstrated themselves to be extremely effective in aiding the design and optimisation of the LHC collimation system.

SixTrack's approach to primary impacts on collimators is to call Monte Carlo scattering procedures, whereby the primary is either lost in an inelastic collision, or undergoes an elastic process and is reintroduced to the tracker. Elsewhere, if a primary particle exceeds the aperture at a given point, it is treated as lost immediately at that location. Secondary particles that would generally stem from these impacts are not treated.

Beam Delivery Simulation (BDSIM) is a novel code which seeks to track the passage of the primary particle as well as any resulting secondary particles [6]. As a result it is it will be more capable of capturing the details in LHC loss maps which are otherwise missing in existing tools, and present a more holistic method for simulating beam losses in particle accelerators.

In this paper, preliminary results comparing LHC loss maps from BDSIM, SixTrack, and BLM data from a recent run are presented.

## BDSIM

BDSIM is a C++ particle tracking code based on a collection of high energy physics libraries, including Geant4 [7], CLHEP [8], and ROOT [9]. It automatically builds a Geant4 3-D accelerator model from a set of generic components which enables the seamless tracking of both primary and secondary particles throughout an accelerator or detector. In using Geant4 it has access to all of the standard particle physics processes, but is supplemented with accelerator tracking routines. Standard Geant4 numerical integrators are replaced with transfer matrices for elements such as drifts,

\* stuart.walker@rhul.ac.uk

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

dipoles (including fringe fields) and quadrupoles. This approach only is possible because of the use of a curvilinear transform, provided by BDSIM, from the Geant4 Cartesian coordinate system, to the curvilinear system. Higher thick order multipoles are handled with the use of symplectic Euler integrators.

BDSIM models are described with the use of a MAD-X-style ASCII input and typically converted from an existing optical description, such as MAD-X TFS [10] using the Python package *pybdsim*, allowing one to build a Geant4 model of a given accelerator within minutes. Whilst models are built from a set of generic components, the user may choose to provide more detailed geometries, field maps, and more to further improve their simulation's accuracy. BDSIM supports all MAD-X apertures, which are shown in Figure 1, including the LHC aperture shown on the top left, that includes the copper beam screen and cooling tubes.

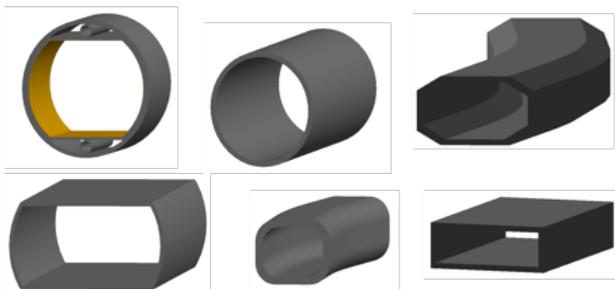


Figure 1: A selection of MAD-X apertures as rendered in BDSIM. Note that the top left LHC aperture has been supplemented with additional detail.

Geant4 mandates that there must be no overlaps between solids, otherwise the particle tracking is prone to becoming stuck in a loop, and fail. Therefore it is absolutely necessary that overlaps are not present in any BDSIM model. BDSIM guarantees that the default-provided generic geometries respect this requirement. Not only do overlaps result in problematic behaviour, but also coplanar faces do as well. This means that not only must there be no overlaps, but also there must be a gap between solids. In BDSIM this manifests itself most noticeably as the introduction of a small 1 nm gap between all accelerator components. The introduction of this small gap has minimal impact for linear accelerators, but ultimately manifests itself as an emittance growth over many turns in a circular collider model. This is not a physical process, but instead simply a limitation of Geant4 as applied to accelerators.

In order to mitigate the detrimental effects of geometry safety separations, multi-turn tracking is synchronised with a 14<sup>th</sup> order one turn map (OTM) from PTC. For any one turn the tracking is first done in the OTM and the results are cached. The same full-turn tracking is then done in BDSIM, which includes a comprehensive set of physics interactions, accurate aperture intersection calculation and tracking in external magnetic fields outside the beam pipe. At the end of the turn, if and only if a beam particle is primary and has not

undergone an interaction in this turn, its coordinates are set to the coordinates from the OTM. All secondary particles and primary particles that have interacted keep the coordinates from BDSIM and the tracking continues. This approach ensures stable tracking of particles that survive inside the aperture for many turns, while leveraging all the benefits of the 3D model for the particles that survive. Figure 2 shows a reference particle tracked in BDSIM with and without the use of the one turn map. The tracking accuracy is clearly drastically improved when using the one turn map.

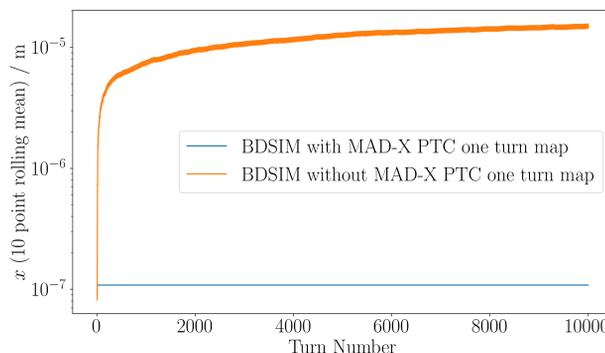


Figure 2: BDSIM long term tracking comparison for a reference particle. The transverse position  $x$  is recorded at the end of each turn for 10,000 turns with and without the use of a one turn map to correct the particle's trajectory.

## MODELLING

BDSIM and SixTrack models were both prepared from the same MAD-X TFS optical description of a recent LHC configuration. The model parameters are the 2018, 6.5 TeV,  $\beta^* = 30$  cm, “end-of-squeeze” optics and a summary of the collimator openings is shown in Table 1. The only difference is that SixTrack, a thin lens tracker, uses a thin description, whereas BDSIM was prepared from the thick version to ensure the desired geometry is built.

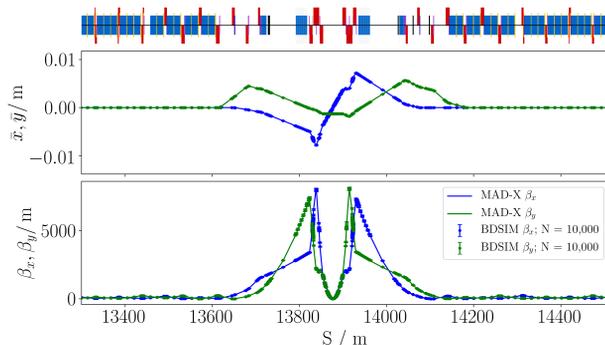


Figure 3: A comparison of the beam centroids (top) and the horizontal and vertical Twiss  $\beta$ -functions (bottom) between MAD-X and BDSIM.

BDSIM's tracking with respect to this model is validated by tracking  $10^4$  protons over a single turn, where the optical

Table 1: The Collimator Openings, Corresponding to the 2018 6.5 TeV,  $\beta^* = 30$  cm, “End-of-Squeeze” Optics Used in the Comparisons Described in this Paper

Collimator	Opening / $\sigma$
TCP IR7	5.0
TCSG IR7	6.5
TCLA IR7	10.0
TCP IR3	15.0
TCSG IR3	18.0
TCLA IR3	20.0
TCSP IR6	7.4
TCDQ IR6	7.4
TCT IR2	37.0
TCT IR8	15.0
TCT IR1/5	8.5

functions are extracted directly from the beam distribution using moments of the beam sigma matrix of up to 4<sup>th</sup> order and comparing with MAD-X. Excellent agreement between BDSIM and MAD-X is shown in Figure 3. However, to correct BDSIM’s longer term tracking, as described in more detail in the previous section, a 14<sup>th</sup> order Taylor map from MAD-X PTC is used to reset primaries onto their correct trajectory at the end of each turn.

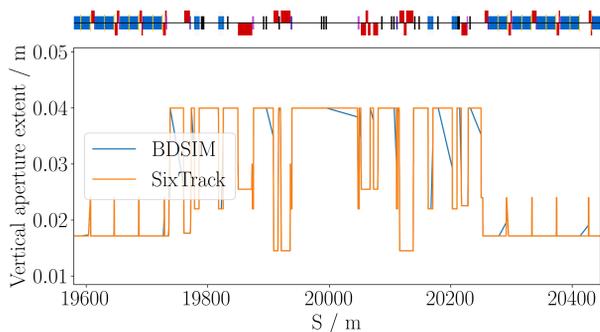


Figure 4: A comparison of the vertical aperture extents between BDSIM and SixTrack. The difference between the two shown is an artefact of the way apertures are described in BDSIM. A BDSIM does not have an aperture, but instead an *opening*, whereas the aperture is defined throughout for SixTrack, even where apertures are located.

An accurate aperture model is mandatory for a faithful simulation of losses in the LHC. SixTrack uses an interpolated aperture description with a resolution of 10 cm by default. To further enhance the geometrical description, this same aperture description was loaded from SixTrack into BDSIM. A comparison of the aperture model in IR5 is shown in Figure 4, where excellent agreement is shown. The geometric description was supplemented with further details to improve the model accuracy: the *lhcdetailed* aperture type, shown in Figure 1, was used throughout; the detailed

LHC magnet geometries; and the correct handedness for the magnet geometries featuring two beampipes.

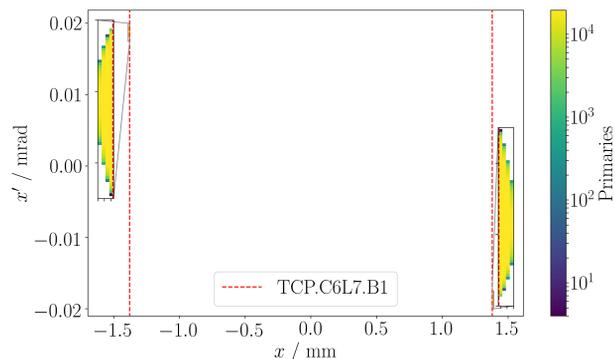


Figure 5: Primary distribution with collimator jaws marked. The primary distribution is shown zoomed inset as the region in phase space is extremely small.

In this paper only results from the simulation of hits on the the horizontal primary collimator (TCP.C6L7.B1) in IR7 have been simulated. To reduce the simulation time, only primaries with an initial position overlapping with the collimator jaws are generated. Again, these same primaries are loaded into BDSIM from SixTrack. The initial primary distribution is shown in Figure 5, with the jaws of the collimator marked in red. This primary distribution results in an impact parameter distribution shown in Figure 6. The impact parameter is simply the shortest transverse distance from the impact on the collimator jaw to its edge. The  $\leq 10 \mu\text{m}$  impact parameter is relevant here as this is typical of beam halo particles impacting on the collimator jaws.

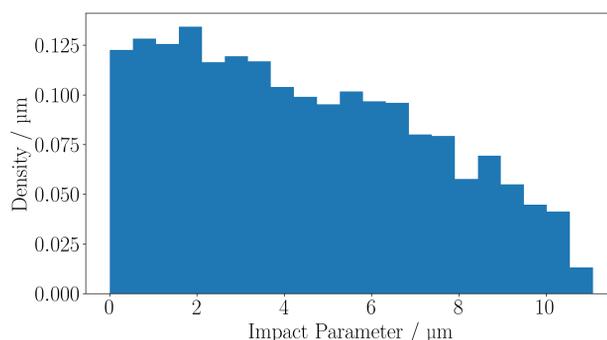


Figure 6: Probability density for the impact parameters on the horizontal collimator in IR7.

## RESULTS

$6.4 \times 10^6$  protons in SixTrack, and  $3 \times 10^6$  protons in BDSIM were simulated. When comparing measured losses to predicted losses it is important to make the distinction of what the results compared are. In the case of standard SixTrack, the total energy of the particle is considered to have been deposited at the aperture impact location. In contrast, the BLMs record losses that escape the magnet and have

thus propagated from the aperture impact location through the beampipe, coils, yoke and cryostat fixtures. In BDSIM the whole shower development is captured and energy deposition is recorded along the full path of any nominal or secondary particles in the corresponding materials. As a result, a direct one-to-one comparison is inappropriate. Instead BDSIM energy deposition, and SixTrack losses, both normalised with respect to their corresponding peaks in IR7 are shown, as well as the BLM dose, again normalised with respect to the peak in IR7.

Figure 7 shows a comparison between the three different loss maps for the whole machine. The maps are colour coded to highlight in particular losses in cold regions, as these are the regions in which it is most necessary to avoid losses. The two largest peaks in the collimation insertions, the larger in IR7 and the smaller in IR3, are correctly reproduced in BDSIM and good agreement is shown between all three loss maps. Low-level noise is present throughout the BLM loss map, and as this is completely missing in the SixTrack loss map, one might imagine it is a result of secondary particles depositing energy in these regions. However, whilst BDSIM does record some energy deposition in these regions, it is mostly missing in BDSIM as well. This suggests that this energy deposition does not stem from primary losses in IR7. Instead it could be originating in losses from the beam, which neither SixTrack nor BDSIM account for. Finally a longer tail in the region following IR7 in the BDSIM loss map compared to the SixTrack loss map is clearly visible. This is likely explained as secondary particles which originate from primary losses in IR7 travelling further downstream and depositing energy.

In Figure 8, one can see a zoomed section of IR7 and the dispersion suppressor immediately following it. Inspecting this region is useful for a number of regions, firstly to ensure the collimator hierarchy is obeyed, where the number of losses in the collimators decreases further downstream. Secondly, the cold section immediately following IR7 is the dispersion suppressor, which is generally the cold region in the LHC exposed to the greatest number of losses, and therefore the section closest to quenching.

A number of features are notable in Figure 8. The collimation hierarchy is recreated as expected, however in BDSIM the peak energy deposition actually occurs in the warm quadrupole immediately following the primary collimator. The BLM data in this region shows that the dose in this quadrupole is indeed very large, but still marginally smaller than detected in the collimator BLMs. This could simply be due to the fact that BDSIM integrates all energy deposition within a given S bin, whereas the actual BLM volume is much smaller than this. It may also be necessary to further improve the geometry in this location to resolve these features more accurately: it was mentioned above that the quadrupole geometry used was the geometry of superconducting quadrupoles. However one can plainly see that in this case, the quadrupoles are warm, and therefore normal conducting, not superconducting.

The BLMs placed between the collimators show that there are sizeable losses in these regions. These features, which are almost entirely missing in SixTrack, appear to be faithfully recreated in BDSIM, with similar losses relative to the maximum.

In the dispersion suppressor more detail is present in the loss map generated with BDSIM. The losses are much more smeared out, which is most likely explained by the fact that SixTrack will immediately kill all particles lost in this region, as they are aperture losses, whereas BDSIM will model any primary elastic or inelastic scattering and track any resulting secondary particles. However the presence of two populations in this region, clear in SixTrack, is still discernible in BDSIM. Finally, the relative sizes of the losses in this region appear to be more comparable to the BLM data in BDSIM than SixTrack, where in this region the BLM losses are on the order of  $10^{-4}$ , in BDSIM also  $10^{-4}$ , and in SixTrack between  $10^{-5}$  and  $10^{-6}$ .

## CONCLUSION

A Geant4 accelerator model for the LHC has been built using BDSIM and used to generate loss maps. These BDSIM loss maps have been contrasted and compared with ones from the standard CERN LHC collimation code SixTrack, as well as BLM data from a recent qualification run.

Whilst the results presented here are preliminary in advance of further, more detailed studies, it is nevertheless apparent that many of the features present in the BLM data which are missing in SixTrack seem to be recreated with BDSIM. This includes the warm losses present between the collimators, explained simply as energy deposited by secondaries, as well as the more smeared out losses in the dispersion suppressor. Some features remain unexplained, such as the low-level noise-like present in the BLM data.

Further refinement of the model geometry can be explored to improve the accuracy of the model. For example, the differences in the geometries between warm and cold magnets is not accounted for in this model, instead only cold magnet geometries are used. Moreover BLM elements can be placed along the ring model in a one-to-one correspondence with those in the actual LHC. This will enable a direct comparison between detector dosages calculated with BDSIM and real LHC BLM dosages.

On the BDSIM side, developments to further improve its use for LHC collimation studies are planned. Firstly, additional data pertaining to losses in collimators will be stored. Secondly, BDSIM can be engineered to treat collimator impacts and aperture losses in the same way as SixTrack—enabling a direct one-to-one comparison between the two codes. Finally, development of a dedicated tracker is in progress. This will allow for a dramatic decrease in simulation time, allowing for greater statistics, as well as tracking which is both more accurate and symplectic.

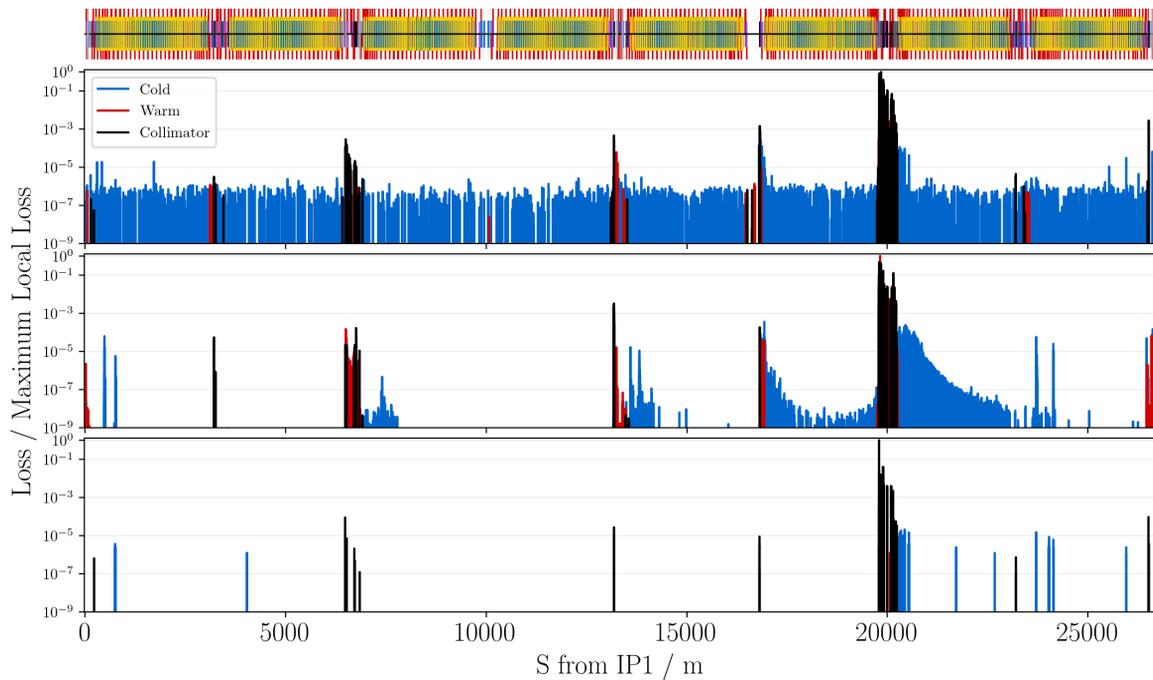


Figure 7: Loss map comparison for the whole LHC between beam loss monitor signal (top), BDSIM energy deposition (middle), and SixTrack losses (bottom). Each is normalised to the corresponding peak in IR7. Cold, warm, and collimator losses are encoded to highlight the various features, and a diagram of the machine is displayed along the top.

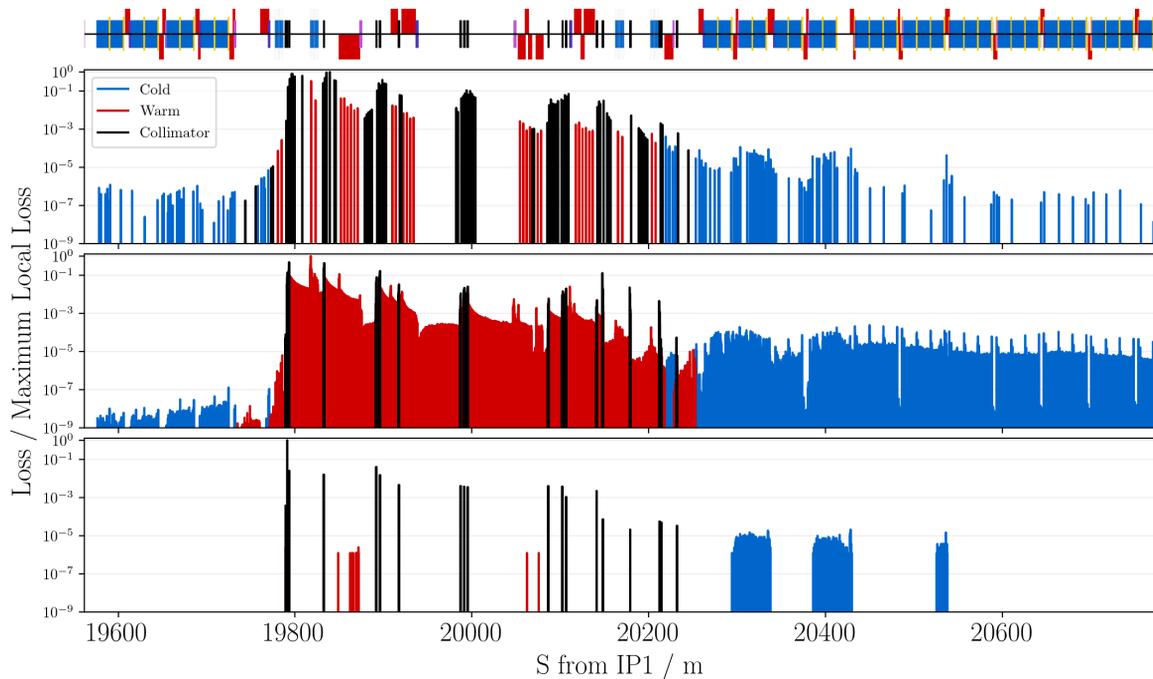


Figure 8: Loss map comparison, zoomed to IR7 where the betatron collimation is located, between beam loss monitor signal (top), BDSIM energy deposition (middle), and SixTrack losses (bottom). Each is normalised to the corresponding peak in IR7. Cold, warm, and collimator losses are encoded to highlight the various features, and a diagram of the machine is displayed along the top.

### ACKNOWLEDGEMENTS

A great deal of gratitude towards three people: to N. Fuster-Martinez for help in acquiring and processing LHC BLM data, R. Bruce for various LHC collimation details, and J. Molson for indispensable help with using SixTrack.

### REFERENCES

- [1] L. Evans and P. Bryant, "LHC machine," *Journal of instrumentation*, vol. 3, no. 08, pp. S08001, 2008.
- [2] E. B. Holzer, *et al.*, "Beam loss monitoring for LHC machine protection," *Phys. Procedia*, vol. 37, pp. 2055-2062, 2012.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

- [3] F. Schmidt, "SIXTRACK: version 1, single particle tracking code treating transverse motion with synchrotron oscillations in a symplectic manner," CERN, Geneva, Switzerland, No. CERN-SPS-88-51-AMS-rev. CM-P00049314, 1990.
- [4] G. Robert-Demolaize, R. Assmann, S. Redaelli, and F. Schmidt, "A new version of SixTrack with collimation and aperture interface," in *Proc. PAC'05*, Knoxville, TN, USA, May 2005, pp. 4084-4086, paper FPAT081.
- [5] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, "FLUKA: a multi-particle transport code (program version 2005)," No. INFN-TC-05-11, 2005.
- [6] L. Nevay, *et al.*, "BDSIM: an accelerator tracking code with particle-matter interactions," arXiv preprint arXiv:1808.10745, 2018.
- [7] S. Agostinelli, *et al.*, "GEANT4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250-303, 2003.
- [8] L. Lönnblad, "CLHEP: a project for designing a C++ class library for high energy physics," *Comput. Phys. Commun.*, vol. 84, no. CERN-TH-7175-94, pp. 307-316, 1994.
- [9] R. Brun and F. Rademakers. "ROOT—an object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1-2, pp. 81-86, 1997.
- [10] H. Grote and F. Schmidt. "MAD-X—an upgrade from MAD8," in *Proc. PAC'03*, Portland, OR, USA, May 2003, pp. 3497-3499, paper FPAG014.

# ANALYSIS OF THE BEAM LOSS MECHANISM DURING THE ENERGY RAMP-UP AT THE SAGA-LS

Y. Iwasaki<sup>†</sup>, SAGA Light Source, [841-0005] Tosu, Japan

## Abstract

The accelerator of the SAGA Light Source consists of the 255 MeV injector linac and the 1.4 GeV storage ring. The accumulated electron beam current of the storage ring is about 300 mA. The energy of the electrons is raised up to 1.4 GeV in 4 minutes in the storage ring. At the moment of the beam acceleration, the electron beam is lost. The amount of beam loss is normally about 5 mA to 30 mA. All electrons are sometimes lost. We investigated the relationship between the beam loss and the betatron tune shifts. The tune shifts during the beam acceleration were analyzed from the measured data of the output currents of the magnets power supplies by using beam tracking code of TRACY2. It was found that the anomalous output of the power supply of bending magnets was one of the causes of the beam loss.

## INTRODUCTION

The accelerator of the SAGA Light Source (SAGA-LS) consists of the 255 MeV injector linac and the 1.4 GeV storage ring [1, 2]. The accumulated electron beam current of the storage ring is about 300 mA. The energy of the electrons is raised up to 1.4 GeV in 4 minutes in the storage ring. At the moment of the beam acceleration (the beam energy is lower than 400 MeV), the electron beam is lost (see Figure 1). The amount of beam loss is normally about 5 mA to 30 mA. All electrons are sometimes lost. To understand the beam loss mechanism, which depends on the beam current, we developed high-speed logging

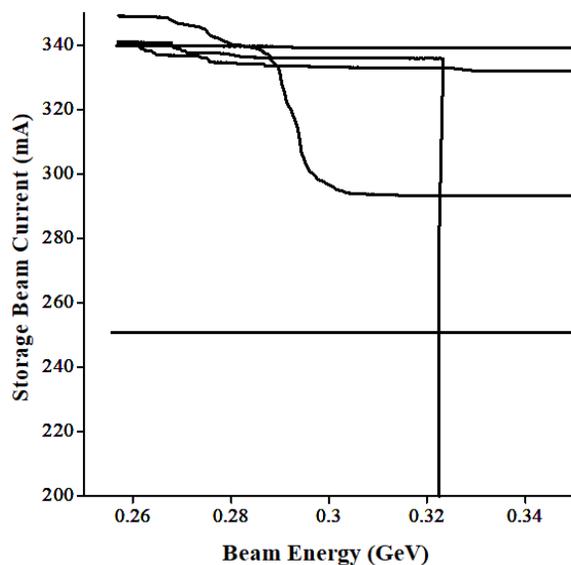


Figure 1: Beam loss at the energy ramp-up.

<sup>†</sup> iwasaki@saga-ls.jp

system of 100 kHz for monitoring the beam current and the magnets power supplies using National Instruments PXI. We investigated the relationship between the beam loss and the betatron tune shifts. The tune shifts during the beam acceleration were analyzed from the measured data of the output currents of the magnets power supplies by using beam tracking code of TRACY2 [3]. To estimate the K-value of the quadrupole magnets, we used orbit response matrix analysis method [4]. By adopting the new high-speed logging system, the time structure of the beam loss process was clearly observed. In this paper, we describe the data acquisition and the data processing system, and the result of the analysis.

## METHODS

Commonly we use N.I. Fieldpoints, PLCs and PCs system for controlling and monitoring the SAGA-LS accelerator [5]. The beam loss occurs at the moment of the energy ramp-up and it observed like a step function by using slow (1 Hz) monitoring system. Therefore, we developed high-speed logging system of 100 kHz for monitoring the beam current and the magnets power supplies using National Instruments PXI. Figure 2 shows the data acquisition and analysis system for investigation of the beam loss. Since the signals of output currents of the power supplies are highly noisy, the low-pass filter (100 Hz) was performed to the measured data. The data sets were thinning out to 1/100 to calculate tunes and Twiss parameters step by step. Orbit response matrix analysis method [4] was adopted to estimate effective K-value of quadrupole magnets from the measured output currents of the power supplies of the quadrupole magnets.

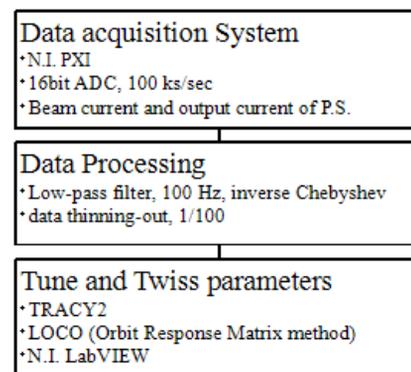


Figure 2: Data acquisition and analyses System.

The TRACY2 was used to calculate the tunes and Twiss parameters. The tracking code of TRACY2 was called from the National Instruments LabVIEW. The Data processing and the optics calculation were performed under the environment of LabVIEW for Microsoft Windows.

## RESULTS

Figure 3 shows the output current of the bending magnets at the case of the all electrons were lost. As can be seen in Figure 3, measured data of the output currents of the power supplies contained high frequency components. After the data processing of low-pass filter, the output current of the power supply of the bending magnets is obtained. Figure 4 shows the anomalous output of the power supply of bending magnets. Since the ramp-up pattern is monotonically increases and fixed, power down of the power supply during the ramp-up couldn't be expected. Figure 5 shows the large betatron tune shifts at the beam loss. The beam energy was estimated by using the magnetic field measurement data and the monitoring value of the power supply of the bending magnets. Figure 6 shows the beam current at the period.

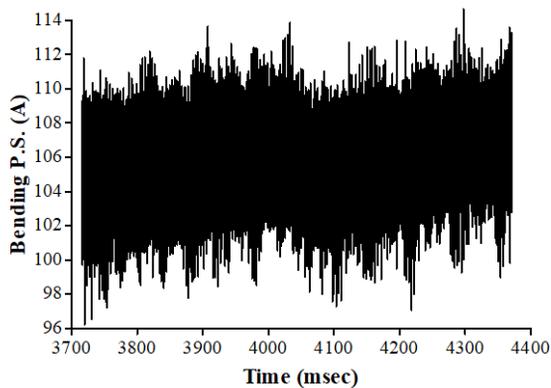


Figure 3: Output current of the power supply of the bending magnets measured by using PXI system at the all beam loss.

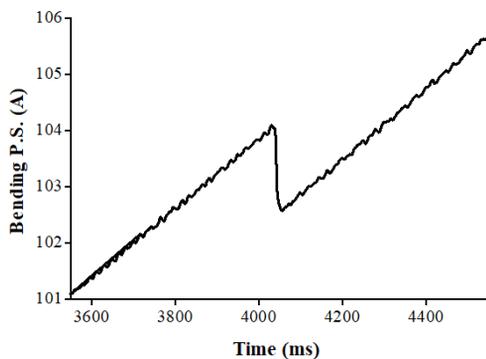


Figure 4: Output current of the power supply of the bending magnets after low-pass filtered.

We confirmed that the beam loss due to the anomalous power down during ramp-up has occurred several times. On the other hand, there were the cases that couldn't be explained by the power supply faults.

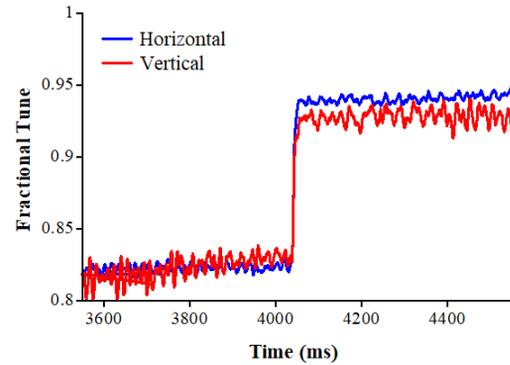


Figure 5: Calculated tune shifts at the all beam loss.

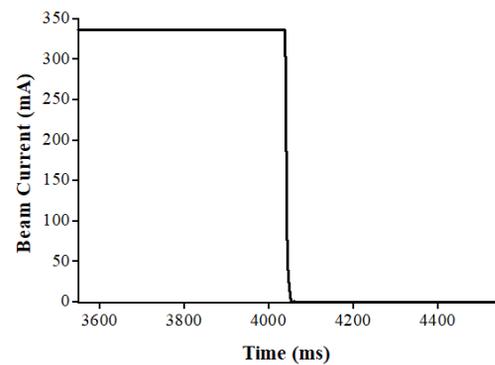


Figure 6: Beam current at the all beam loss.

## CONCLUSION

The first data acquisition system of the PXI and the data processing of low-pass filter were useful to find the anomalous output of the power supply of the bending magnets. The anomalous output of the power supply of bending magnets was one of the causes of the beam loss. The longitudinal motion will be taken into account for the further investigation of the beam loss mechanism.

## REFERENCES

- [1] T. Tomimasu *et al.*, "The SAGA Synchrotron Light Source in 2003", in *Proc. PAC'03*, Portland, pp. 902-904, 2003.
- [2] Y. Iwasaki *et al.*, "Lattice Design of SAGA Synchrotron Light Source", in *Proc. PAC'03*, Portland, pp. 3270-3272, 2003.
- [3] W. Decking, D. Robin, "Dynamic Aperture Measurements at the Advanced Light Source", in *Proc. PAC'99*, New York, pp. 1581-1583, 1999.
- [4] J. Safranek, "Experimental determination of storage ring optics using orbit response measurements", *Nucl. Instr. and Meth. A* 388, pp. 27-36, 1997.
- [5] H. Ohgaki *et al.*, "PC-LABVIEW Based Control System in SAGA-LS", in *Proc. PAC'05*, Knoxville, May, 2005, pp. 3976-3978.

# pyaopt OPTIMIZATION SUITE AND ITS APPLICATIONS TO AN SRF CAVITY DESIGN FOR UEMS\*

Ao Liu<sup>†</sup>, Roman Kostin, Euclid Techlabs, Bolingbrook, IL, USA  
 C. Jing, Pavel Avrakhov, Euclid Beamlabs, Bolingbrook, IL, USA

## Abstract

Designing and commissioning particle accelerators need great optimization efforts. This is particularly true for large accelerators with complex components that provide stable beam such as light sources and colliders, where nonlinearities of the beam play an important role. Currently, many design optimizations are provided by individual user-created automated problem-finding and solution-proposing algorithms, which requires an extensive amount of computing resources. Heuristic algorithms such as Genetic Algorithms (GA) and Simulated Annealing (SA) are commonly implemented. They are either created for individual tasks, or are implemented directly in simulation codes, such as OPAL or IMPACT3D. An optimization suite that is independent of the accelerator codes is needed for general application studies. Meanwhile, researchers now have access to the HPC resources, which can be utilized for parallelization of codes. We propose a Python-based optimization suite for general applications. In this paper, we introduce the pyaopt suite by giving some details of its applications, including a design of an SRF photogun for UEMs.

## INTRODUCTION

Recently, there has been multiple new applications of heuristic algorithms in the particle accelerator community. The fields include secondary particle collection [1], DA optimizations [2, 3], and space charge calculations [4]. In most of these cases, algorithms were customized for specific physics problems, or built in a specific simulation program. In fact, the number of programs that include the Genetic Algorithm (GA) as the multi-objective optimizer is rapidly increasing [5, 6]. However, for many accelerator physicists and engineers, these algorithms are still inaccessible to some extent: there is no easy way to use them in a “plug and play” fashion.

The design of Python advanced optimization pyaopt suite aims at delivering a package that has an API for users to conveniently describe the optimization problem, select the optimization algorithm and start the job. It not only includes widely-accepted algorithms such as the GA, Simulated Annealing (SA) and the Particle Swarm Algorithm (PSA), but also gradient-based (deterministic) algorithms, such as the Gauss-Newton method, etc. The goal of the Python-based package is to let users run optimization jobs in any environments, including a personal computer, a small-scale cluster, or a HPC supercomputer. Users may select the mode such

\* SRF cavity design work supported by DOE under contract DE-SC0018621

<sup>†</sup> a.liu@euclidtechlabs.com

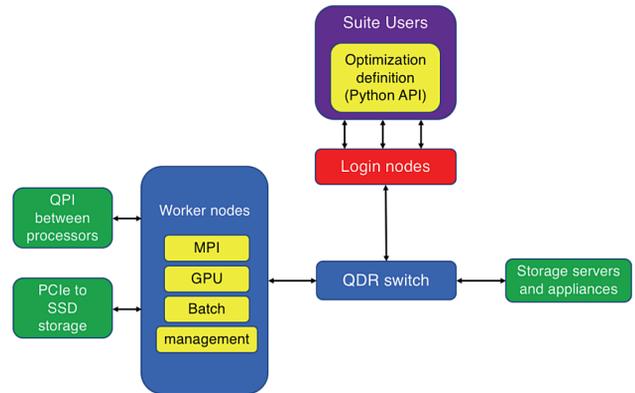


Figure 1: The flowchart of running optimization jobs on an HPC machine.

that the pyaopt job manager can handle the job submission, monitoring, and logging. The idea of running the jobs on an HPC is illustrated in Figure 1.

pyaopt includes a few customized metaheuristic algorithms and some deterministic algorithms. We introduce a selection of the metaheuristics below:

- pyaopt-GA, which is based on the NSGA-II [7] and can do both single-objective GA (SOGA) and multi-objective GA (MOGA). The customization is on the crowding distance (CD) of individuals, which represents the similarities of them, and the “rescue method”: judgment day, which is used only when the whole population ceases improving prematurely. The algorithm is enhanced by MPI [8], such that calculations of dominance and CD are distributed on different ranks.
- pyaopt-SA, which is based on the standard annealing formula  $P = e^{(f(x)-f(x')/T)}$  for  $f(x) < f(x')$ , where  $f(x)$  is the fitness value for solution  $x$ , and  $T$  is the current temperature. The customization is on the adaptive cooling schedule,  $\Delta T$  per iteration, and on the cooling range assignment for MPI implementation. Although it is already a common practice to normalize the fitness value to an expected one and implemented by many SA users, using an adaptive cooling schedule further helps to prevent the system to converge prematurely. As for the search range where multiple workers are present, users can choose how many slices each variable’s range needs to be divided, based on the number of workers available. Then pyaopt allocates each combination of range slices evenly to the workers.
- pyaopt-ANN, which is based on multiple artificial neural network (ANN) algorithms. The parallelization is

done through both the forward and backward propagations of data, in simply a batch fashion.

pyaopt can be installed on multi-platforms, thanks to the installation capabilities brought by the `setuptools` [9] package. Our idea is to make this process painless, such that modules targeted for different computing accelerator architecture that are available on the machine can be automatically detected by the setup code, or specified by the users (when the installation frontend cannot see the heterogeneity from the worker's point of view).

The variable ranges are specified in a JSON [10] file (we are also investigating the HDF5 [11] format to contain the metadata of all the input, output and log files). The API provides the proposed combinations of variables (each combination of variable values is hereafter referred as "individual"), and users specify the functions to be called by `pyaopt` for evaluation of individuals.

## EXAMPLES OF APPLICATIONS

In this section we use some applications of `pyaopt` to demonstrate its capability to be used on different areas of accelerator physics.

### *nuSTORM Magnetic Horn*

The neutrinos from STORed Muons (*nuSTORM*) uses a magnetic horn to capture the secondary pions generated from bombarding a long target rod with high energy protons. Because of the finite length of the target rod, the original point-to-parallel principle of a simple horn design with double parabolic surface is no longer optimal. The horn has to be re-designed for each target that has different materials, lengths and diameters, even when the primary proton beam parameters are fixed. In this example, we used a 46 cm Inconel target, with a radius of 3 mm. We used the MOGA for this purpose, where one of the two objectives is to maximize the number of pions within the transverse phase space, and other is to maximize that in a the momentum acceptance described by a derived formula. Figure 2 shows the variation of fitness values for the dominant elite candidate in each generation. This treatment of converting a single-objective, time-consuming multiparticle tracking-based optimization was efficient in dramatically saving the optimization time and increase the acceptable pions at the end of the pion beamline by 13%.

### *nuSTORM Muon Storage Ring*

The physics objective is to store as many muons in the 2 mm-rad full transverse admittance and  $3.8 \pm 10\%$  GeV/c. Since this is an extra large beam, multiple nonlinearity terms of the beam optics become critical as stop bands for beam circulation in the ring. Sextupoles are introduced in the lattice, in both standalone sextupole magnets and also as combined-function. Therefore, instead of isolating the nonlinearity terms one by one and evaluate the importance of each, we chose to directly rely on multi-particle tracking result as the single optimization objective. Since 90% of the

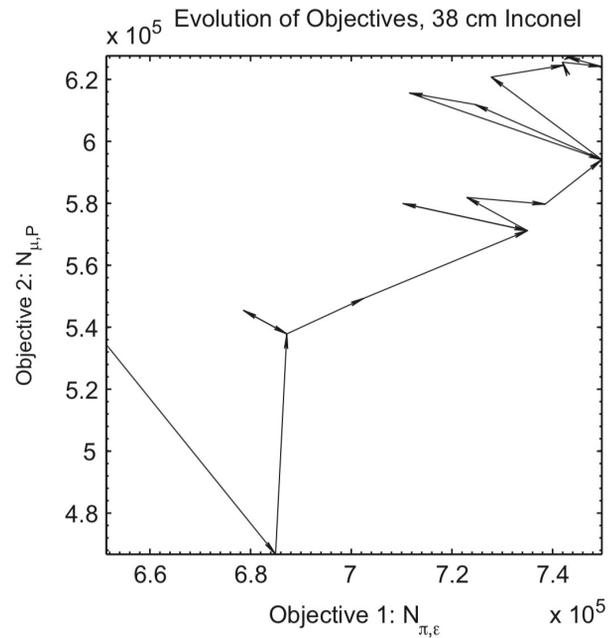


Figure 2: The variation of fitness values of the dominant elite candidate in each generation.

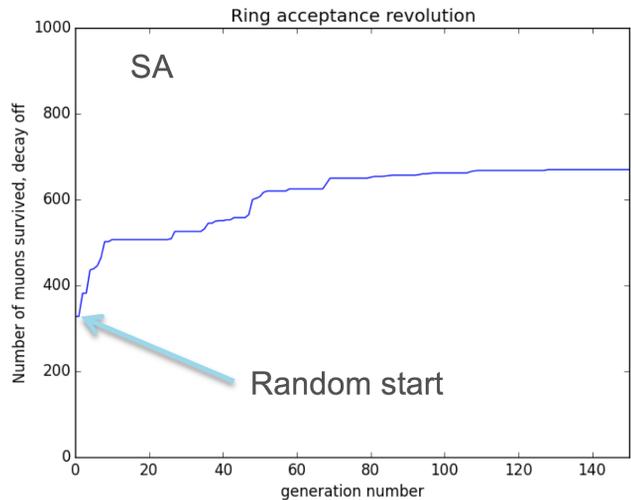


Figure 3: The variation of fitness value of the best solution in each generation.

muons will decay to neutrinos in approximately 100 turns, the percentage of survived muons after 100 turns was used as the evaluation function. We then chose SA for this task. The variation of fitness value per generation is shown in Figure 3. The momentum acceptance is compared for before and after the sextupole correction with the optimized setting in Figure 4.

The suite was also used on the optimization for a Step IV lattice of the international Muon Ionization Cooling Experiment (MICE) [12], 6D ionization cooling channel, etc. Because the suite was designed for general optimization problems, we foresee more areas of applications. In the next

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

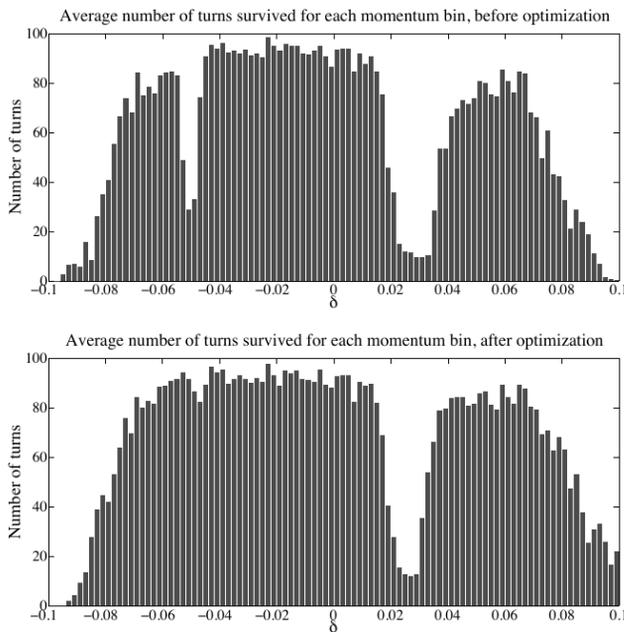


Figure 4: The improvement of momentum acceptance after the sextupole correction with the optimized setting.

section we discuss about its application on the design of an SRF photogun for UEMs.

## SRF PHOTOGUN FOR UEMS

Benefiting from the rapid progress on RF photocathode gun technologies in the past two decades, the development of MeV-range ultrafast electron diffraction/microscopy (UED and UEM) has been identified as an enabling instrumentation, which may lead to breakthroughs in fundamental science and applied technologies [13–15]. In a UED/UEM, stable femtosecond (fs) electron bunches that are synchronized with fs laser pulses is required. Currently, there are room temperature RF photocathode electron guns for generating MeV electrons for UED/UEM. However, the shot-to-shot stability for those machines is still low to fully satisfy requirements from the UED/UEM community. Here we propose a 1.3 cell, 1.3 GHz SRF cavity as the UEM electron gun. The innovations of this structure include but are not limited by:

- It uses a Euclid-designed, ILC type SRF cavity cell with a novel detachable coupler, which was inherited from our previously completed DoE SBIR project (DE-SC0002479). The advantage of using this cell is that the manufacturing and operation time for the whole SRF cavity is dramatically reduced.
- It uses the backwall of the first 0.3 cell as the photocathode, where the quantum efficiency (QE) of the high RRR Niobium (Nb) is up to  $10^{-5}$ .
- By using the novel technologies of conduction cooling and coating of  $Nb_3Sn$ , which are what Euclid and Fermilab are collaborating on now, the peak axial electric

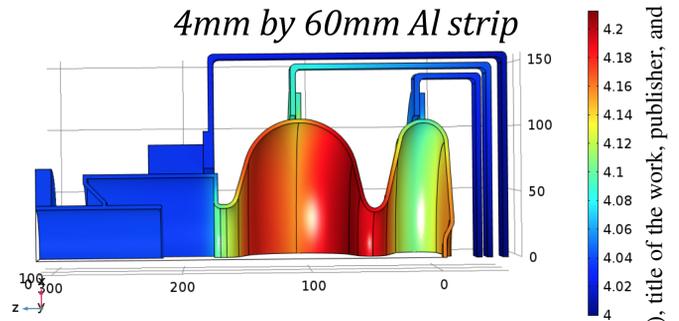


Figure 5: Simulation of the conduction cooling scheme in COMSOL. Figure courtesy of R. Kostin, Euclid Techlabs.

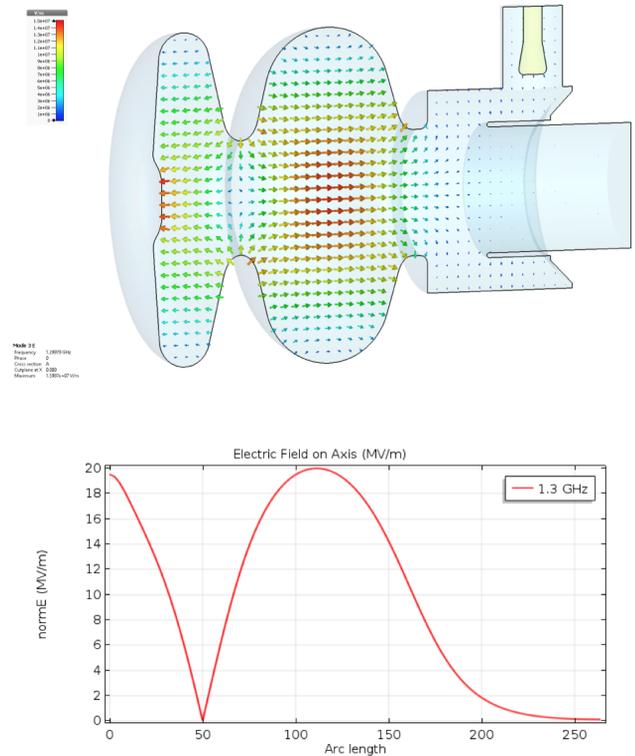


Figure 6: CST simulation of the 1.3 cell SRF cavity and the corresponding axial field, normalized to a maximum value of 20 MV/m.

field ( $E_z$ ) can reach 26 MV/m. Moreover, the conduction cooling allows one to use a cryocooler, without liquid helium, to cool down an SRF structure. See Figure 5 for a COMSOL simulation of the conduction cooling scheme. See also [16] for the published news on the scheme by Fermilab.

Figure 6 shows the designed cavity in CST, and the corresponding field scaled to a conservative estimate of peak axial  $E_z$  of 20 MV/m.

The back wall geometry was preliminarily optimized to provide transverse RF focusing when the beam is generated at the cathode. It has a unique “step” design where the flat face is used for photocathode, and the curved geometry provide the transverse field needed. pyaopt is able to parallelize

the RF simulation in Superfish [17] on a Linux cluster or the OSX platform via using a WineHQ container. The resultant field is then used in Astra [18] for multi-particle tracking. In Table 1, beam parameters suitable for UED/UEM applications, simulated in Astra with space charge effect with the gun design shown in Figure 6, are listed and compared. Further more thorough optimizations will be done in the future studies.

Table 1: Beam parameters for UED/UEM Applications

Parameter name [unit]	Value	
Application	UED	UEM
Beam Energy [MeV]	1.655	1.655
Charge per pulse [fC]	5	500
Laser pulse length [fs]	6.4	6.4
Laser spot size [ $\mu\text{m}$ ]	36	180
Bunch length [fs]	167	741
Beam emittance [nm]	6.6	39
Relative energy spread [1]	$1.3 \times 10^{-5}$	$6.4 \times 10^{-5}$

## FUTURE WORK

The Python wrapper, pyCUDA, will be implemented in pyaopt in the future to utilize users' NVIDIA GPU accelerators, or GPU on NERSC [19]. Furthermore, in order to implement algorithms that are more robust against noises, such that the suite can be deployed on experimental jobs, we will add the RCDS [20] algorithm to pyaopt soon in the future. More test cases, such as a collaboration with lattice design work at BNL [3] will also be considered.

## CONCLUSIONS

We are actively developing a Python-based optimization suite, pyaopt to let users conveniently describe and run optimization problems on personal computers, small-scale clusters or HPC supercomputers. pyaopt includes a selection of deterministic and metaheuristic algorithms and allow users to run them in parallel mode. We showed two test cases for the GA and SA of pyaopt on nuSTORM related studies. The algorithm was also tested by cases of optimizations for MICE and 6D ionization cooling channel designs. In all cases the suite works efficiently in reducing the computing time and finding optimal solutions. By using WineHQ, we are able to combine RF and tracking simulations by Superfish and Astra on OSX or Linux platforms. A preliminary optimization on the SRF photogun design shows promising beam qualities for it to be applied to UED/UEM. The SRF photogun has multiple advantages over the room-temperature photoguns, including its superstability, CW mode operation enabled by the conduction cooling, etc.

## ACKNOWLEDGMENTS

The SRF photogun design work is supported by DoE SBIR grant DE-SC0018621, and the current optimization application study is supported by DoE SBIR grant DE-SC0018191. The authors gratefully thank Dr. A. Bross,

D. Neuffer from Fermilab, and Dr. P. Snopok from IIT for their inputs on the muon accelerator-related optimizations.

## REFERENCES

- [1] A. Liu *et al.*, "Design and Simulation of the nuSTORM Pion Beamline", *Nuclear Inst. and Methods in Physics Research A*, 2015. doi:10.1016/j.nima.2015.08.020
- [2] A. Liu *et al.*, "A FODO racetrack ring for nuSTORM: design and optimization", *JINST*, 2017. doi:10.1088/1748-0221/12/07/P07018
- [3] Y. Li *et al.*, "Genetic Algorithm Enhanced by Machine Learning for Dynamic Aperture Optimization", presented at ICAP'18, Key West, FL, USA, Oct 2018, paper SAPAF01, this conference.
- [4] J. Qiang, "Advances in Simulation of High Brightness/High Intensity Beams", presented at ICAP'18, Key West, FL, USA, Oct 2018, paper SAPLG01, this conference.
- [5] OPAL <http://gitlab.psi.ch/OPAL/src/wikis/home>
- [6] ELEGANT, <https://www.aps.anl.gov/Accelerator-Operations-Physics/Software#elegant>
- [7] K. Deb *et al.*, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182 - 197, Apr 2002. doi: 10.1109/4235.996017
- [8] Open MPI, <https://www.open-mpi.org/>
- [9] setuptools, <https://setuptools.readthedocs.io/en/latest/>
- [10] JSON, <https://www.json.org/>
- [11] HDF5, <https://www.hdfgroup.org/HDF5/>
- [12] A. Liu, on behalf of the MICE collaboration, "MICE Step IV Optics without the M1 Coil in SSD", in *Proc. IPAC'16*, Busan, Korea, May 2016. doi:10.18429/JACoW-IPAC2016-TUPMY006
- [13] T. Chase *et al.*, "Ultrafast electron diffraction from non-equilibrium phonons in femtosecond laser heated Au films", *Applied Physics Letters*, vol. 108, no. 4, pp. 041909, 2016. doi:10.1063/1.4940981
- [14] Ming-Fu Lin *et al.*, "Ultrafast non-radiative dynamics of atomically thin MoSe<sub>2</sub>", *Nature Communications*, vol. 8, pp. 1745, 2017. doi:10.1038/s41467-017-01844-2
- [15] Xiaoxi Wu *et al.*, "Light-induced picosecond rotational disordering of the inorganic sublattice in hybrid perovskites", *Sci. Adv.*, vol. 3, no. 7, pp. e1602388, 2017. doi:10.1126/sciadv.1602388
- [16] Fermilab news, <https://science.energy.gov/hep/highlights/2018/hep-2018-10-b/>
- [17] Superfish, LANL, <https://laacg.lanl.gov/laacg/>
- [18] Astra, DESY, <http://www.desy.de/~mpyflo/>
- [19] NERSC computing center, <http://www.nersc.gov>
- [20] X. Huang *et al.*, "An algorithm for online optimization of accelerators", *NIM A*, vol. 726, pp. 77-83, Oct 2013. doi: 10.1016/j.nima.2013.05.046

# MEAN-FIELD DENSITY EVOLUTION OF BUNCHED PARTICLES WITH NON-ZERO INITIAL VELOCITY

B. S. Zerbe\*, P. M. Duxbury†

Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA

## Abstract

Reed presented a 1D mean-field model of initially cold pancake-beam expansion demonstrating that the evolution of the entire spatial distribution can be solved for all time where the 1D assumption holds. This model is relevant to ultra-fast electron microscopy as it describes the evolution of the distribution within the photoelectron gun, and this model is similar to Anderson's sheet beam density time dependence except that Reed's theory applies to freely expanding beams instead of beams within a focussing channel. Our recent work generalized Reed's analysis to cylindrical and spherical geometries demonstrating the presence of a shock that is seen in the Coulomb explosion literature under these geometries and further discussed the absence of a shock in the 1D model. This work is relevant as it offers a mechanistic explanation of the ring-like density shock that arises in non-equilibrium pancake-beams within the photoelectron gun; moreover, this shock is coincident with a region of high-temperature electrons providing an explanation for why experimentally aperturing the electron bunch results in a greater than 10-fold improvement in beam emittance, possibly even resulting in bunch emittance below the intrinsic emittance of the cathode. However, this theory has been developed for cold-bunches, i.e. bunches of electrons with 0 initial momentum. Here, we briefly review this new theory and extend the cylindrical- and spherical- symmetric distribution to ensembles that have non-zero initial momentum distributions that are symmetric but otherwise unrestricted demonstrating how initial velocity distributions couple to the shocks seen in the less general formulation. Further, we derive and demonstrate how the laminar condition may be propagated through beam foci.

## INTRODUCTION

Freely expanding ensembles of charged particles are fundamental to accelerator physics. Although continuous beams near the particle source are relatively diffuse, bunched beams can reach densities where space-charge effects dominate the expansion. In such a regime, the expansion dynamics are similar to the dynamics of Coulomb explosion that are well studied in the laser ablation field, where it is well established that shocks that form at the periphery of the distribution [1–6]. Our group recently found that in an ultrafast electron microscope experimentally aperturing a high density bunch of electrons after they exit the photocathode gun can result in a significant improvement to the brightness. Simulation results suggest that this effect is due to a den-

sity shock, akin to the shock seen in the Coulomb explosion literature, of high-temperature electrons that form at the longitudinal median of the bunch and migrate out to the transverse edge [7].

It has been known for decades that charge redistribution near the particle source is the origin of a major portion of the emittance seen in standard beams [8]. More than 30 years ago, Anderson presented 1D and cylindrical mean-field fluid models of beam dynamics for ensembles of particles with arbitrary initial distributions relevant while the beam remains laminar [9]. These models describe the transverse density and emittance evolution in the presence of a focussing force, and specifically they provide insight into emittance oscillation that is important for emittance compensation [10, 11]. While it is reasonable to make an analogy between that mechanism and the freely expanding charge redistribution we see during Coulomb explosion, Anderson's models are inappropriate for a freely expanding bunch as they assume the focussing force is non-zero and radially inward. Therefore, other models are needed to describe the freely expanding case.

Within the ultrafast electron microscopy (UEM) literature, numerous works postulated 1D models for non-relativistic longitudinal free expansion [12–14], and Reed eventually settled upon the same mean-field fluid approach used by Anderson but without any external fields [15]. Again this model was to describe the longitudinal density evolution of initially dense “pancake” bunches — named so as they have much shorter longitudinal widths than transverse radius — that can be assumed to be planar symmetric instead of Anderson's description of a cylindrical symmetric beam's transverse density evolution. Reed's mean-field model accurately describes the longitudinal expansion while planar symmetry can be assumed [15]. However, Reed was concerned that no Coulomb explosion-like shock was seen in the model even when non-uniform initial conditions were assumed, in stark disagreement to what had been previously found within the Coulomb explosion literature. We recently demonstrated that such a shock cannot occur in the non-relativistic 1D model without careful tuning of the initial velocity distribution [16]. In contrast, we showed that these shocks spontaneously occur in higher dimensions for non-uniform distributions [16], so that the theoretical results found in the UEM community are consistent with the shocks found in the Coulomb explosion literature.

To demonstrate these results, we generalized Reed's model to higher dimension by deriving closed form analytic expressions that describe arbitrary density evolution under cylindrical and spherical symmetries. We discovered that the shocks arise due to relative bunching of particles

\* zerbe@pa.msu.edu

† duxbury@pa.msu.edu

that can be described by dimensionally-dependent evolution functions,  $f_d$  with  $d = 2, 3$  for the cylindrical and spherical symmetric cases, respectively, that are multiplied by the factor  $D_{0d} = \frac{d}{2} \left( \frac{\rho_0}{\bar{\rho}_0} - 1 \right)$  where  $\rho_0 = \rho_0(r_0)$  and  $\bar{\rho}_0 = \bar{\rho}_0(r_0)$  represent the initial probability-like density and initial average probability-like density at  $r_0$ , respectively.  $D_{0d}$  is determined entirely from the initial conditions of the distribution, and captures the difference in behavior from that of a uniform distribution, as for the uniform distribution  $D_{0d} = 0$ . In that case, the density evolution is free from the changes caused by the function  $f_d$  and follow a simple power law. We note that most analysis in accelerator physics is based on ensembles with spatially uniform distributions, where the complications introduced by the function  $f_d$  are seldom treated analytically, though in many experimental contexts non-uniform distributions are endemic.

As density peaks may arise in the planar symmetric model by carefully tuning the initial velocity distribution, we postulate that the peaks under cylindrical and spherical symmetry should be able to be likewise controlled by the initial velocity distribution. However, our previous model assumes cold initial conditions. Here we present an extension of our previous model that includes arbitrary initial velocities that can be written as a single-value function of the radius of the appropriate symmetry. We demonstrate that this model reproduces particle-in-cell (PIC), implemented in warp [17], simulations. We also show that this model breaks down when an inward velocity that is linear in the radius is assumed for the Gaussian distribution; however, the model correctly predicts the focus and subsequent expansion when a more complicated, non-linear initial velocity profile is assumed.

## DERIVATION

In this section, we present a derivation of the density evolution equations with arbitrary initial velocity,  $v_0 = v_0(r)$  in the  $\hat{r}$  direction, under cylindrical and spherical symmetries. This analysis follows from our earlier work [16] with the following differences: 1) we assume non-zero radial velocity and 2) we adopt slightly modified notation that we have recently developed for a relativistic extension of our initial analysis (un-published).

Consider an ensemble of particles with cylindrical symmetry. Define the time-dependent probability-like density (fraction of entire distribution per unit area),  $\rho_2(r, t)$ , and denote the initial probability-like density as  $\rho_{02} = \rho_{02}(r_0) = \rho_2(r_0, t = 0)$ . With the initial conditions, we have,

$$P_{02} = \int_0^{r_0} 2\pi r \rho_{02}(r) dr, \quad E_0(r_0) = E_{02} = \frac{\Lambda_{\text{tot}} P_{02}}{2\pi \epsilon_0 r_0},$$

where  $\Lambda_{\text{tot}}$  is the total charge per unit length along the cylindrical charge distribution and  $P_{02}$  is the cumulative probability. Notice that the quantity  $\Lambda_0 P_{02}$  represents the charge per unit length inside radius  $r_0$ , so further define the average

probabilistic-like density as

$$\bar{\rho}_{02} = \frac{P_{02}}{\pi r_0^2}. \quad (1)$$

Consider an ensemble of particles with spherical symmetry. Define the time-dependent probability-like density (fraction of entire distribution per unit volume),  $\rho_3(r, t)$ , and denote the initial probability-like density as  $\rho_{03} = \rho_{03}(r_0) = \rho_3(r_0, t = 0)$ . With the initial conditions, we have,

$$P_{03} = \int_0^{r_0} 4\pi r^2 \rho_{03}(r) dr, \quad E_0(r_0) = E_{03} = \frac{Q_{\text{tot}} P_{03}}{4\pi \epsilon_0 r_0^2},$$

where  $Q_{\text{tot}}$  is the total charge in the system and  $P_{03}$  is the cumulative probability. Again notice that  $P_{03}$  represents the fraction of the particles that lie inside radius  $r_0$  and  $Q_{\text{tot}} P_{03}$  gives the charge inside radius  $r_0$ , so further define the average probability-like density as

$$\bar{\rho}_{03} = \frac{P_{03}}{\frac{4}{3}\pi r_0^3}. \quad (2)$$

Assuming the distribution undergoes laminar flow, the electric field experienced by a particle at radial position  $r(r_0, t)$  under cylindrical and spherical symmetries, respectively, is

$$E_2(r) = E_{02} \frac{r_0}{r}, \quad E_3(r) = E_{03} \left( \frac{r_0}{r} \right)^2.$$

Under the laminar assumption,  $E_{02}$  and  $E_{03}$  are constants, and the change in potential energy is found by integrating the force  $qE$ , and we find,

$$\Delta U_2 = \mathcal{E}_{r2} \ln \left( \frac{r_0}{r} \right), \quad \Delta U_3 = \mathcal{E}_{r3} \left( \frac{r_0}{r} - 1 \right) \quad (3)$$

for the cylindrical and spherical cases respectively. Here  $\mathcal{E}_{r2} = \frac{q\Lambda_{\text{tot}} P_{02}}{2\pi \epsilon_0}$  for the cylindrical case and where  $\mathcal{E}_{r3} = \frac{qQ_{\text{tot}} P_{03}}{4\pi \epsilon_0 r_0}$  for the spherical case. Notice that by this convention,  $\Delta U_d < 0$ . Further introduce a fictitious velocity,  $v_{\text{rd}}$  for  $d = 2, 3$ , such that  $v_{\text{rd}} = +\sqrt{\frac{2\mathcal{E}_{\text{rd}}}{m}}$ . Using conservation of energy in the non-relativistic regime with initial energy  $\mathcal{E}_0 = \frac{1}{2}mv_0^2$ , we can solve for the velocity,

$$\frac{v_2}{v_{r2}} = \pm \sqrt{\frac{\mathcal{E}_0}{\mathcal{E}_{r2}} + \ln \left( \frac{r}{r_0} \right)}, \quad \frac{v_3}{v_{r3}} = \pm \sqrt{\frac{\mathcal{E}_0}{\mathcal{E}_{r3}} + 1 - \frac{r_0}{r}},$$

where the  $\pm$  is determined by whether the particle is traveling away or toward the origin and the subscript again indicates the appropriate symmetry. In other words, the velocity equations become double valued for  $r < r_0$  when  $v_0 < 0$  as both the negative and positive square roots occur; specifically, there is a radius,  $r_{\text{td}} < r_0$  with  $d = 1, 2$ , at which the Lagrangian particle reaches 0 velocity and turns-around, and the velocities between this  $r_{\text{td}}$  and  $r_0$  are symmetric —

differing only by their sign. By setting  $v = 0$ ,  $r_{td}$  can be derived

$$r_{t2} = r_0 e^{-\frac{v_0^2}{v_{r2}^2}}, \quad r_{t3} = \frac{r_0}{1 + \frac{v_0^2}{v_{r3}^2}}. \quad (4)$$

With this notation, the velocities can be rewritten as

$$\frac{v_2}{v_{t2}} = \pm \sqrt{\ln\left(\frac{r}{r_{t2}}\right)}, \quad \frac{v_3}{v_{t3}} = \pm \sqrt{1 - \frac{r_{t3}}{r}}, \quad (5)$$

where  $v_{t2} = \sqrt{\frac{q\Lambda_{tot}P_{02}}{\pi m \epsilon_0}} = v_{r2}$  and  $v_{t3} = \sqrt{\frac{qQ_{tot}P_{03}}{2\pi m \epsilon_0 r_{t3}}} = v_{r3} \sqrt{\frac{r_0}{r_{t3}}}$ . We use these turn-around radii to define the average probability-like densities

$$\bar{\rho}_{t2} = \frac{P_{02}}{\pi r_{t2}^2}, \quad \bar{\rho}_{t3} = \frac{P_{03}}{\frac{4}{3}\pi r_{t3}^3}, \quad (6)$$

and the associated plasma frequencies

$$\omega_{t2} = \sqrt{\frac{q\Lambda_{tot}\bar{\rho}_{t2}}{\epsilon_0 m}} = \frac{v_{t2}}{r_{t2}}, \quad \omega_{t3} = \sqrt{\frac{2}{3}} \sqrt{\frac{qQ_{tot}\bar{\rho}_{t3}}{\epsilon_0 m}} = \frac{v_{t3}}{r_{t3}}, \quad (7)$$

thus effectively mapping this problem to the cold freely-expanding case. The main difference, now, is that  $r_{td}$  is a function of both  $r_0$  and  $v_0$ , and  $\omega_{t3}$  is now a function of both  $r_0$  and  $v_0$  instead of solely  $r_0$ . Furthermore,  $r_{td}$  does not necessarily occur at the same time for all Lagrangian particles, so it is not precisely cold expansion-like but is mathematically similar. This will complicate the the derivation of  $r'$  where  $' \equiv \frac{d}{dr_0}$ , but it will much simplify the derivation and interpretation of the time-position relation.

To derive the time-position relation for a specific Lagrangian particle, we consider the normal time-position relation with  $r_0$  replaced by  $r_{td}$ . If  $v_0 > 0$ , then the time-position relation is the same as the cold expansion relations less the time it would take the particle to travel from  $r_{td}$  to  $r_0$ , call this  $t_d$  for  $d = 1, 2$ . If  $v_0 < 0$ , then the particle needs to travel from  $r_0$  to  $r_{td}$  before undergoing cold free expansion. As this process is symmetric to the expansion from  $r_{td}$  to  $r_0$ , the alteration is again  $t_{td}$ . Denote  $t_{fid}$  as the portion of the time-position relation defined by the cold free-expansion from  $r_{td}$ . Thus,  $t = \pm t_{fid} - t_{td}$  where the  $\pm$  sign is determined by whether the Lagrangian particle is moving away or toward the origin, respectively,  $t_d$  has the same sign as  $v_0$ , and  $d = 1, 2$  for the cylindrical and spherical symmetric case, respectively. The parameter  $t_{fid}$  can be determined from our previous work:

$$t_{fid} = \frac{2}{\omega_{t2}} e^{y_2^2} F(y_2), \quad (8)$$

$$t_{fid} = \frac{1}{\omega_{t3}} \left( \tanh^{-1} y_3 + \frac{y_3}{1 - y_3^2} \right). \quad (9)$$

where  $y_2 = \sqrt{\ln\left(\frac{r}{r_{t2}}\right)}$ ,  $y_3 = \sqrt{1 - \frac{r_{t3}}{r}}$ , and  $F(\cdot)$  represents the Dawson function. From these equations, we can also obtain  $t_{td}$

$$t_{t2} = \frac{2}{\omega_{t2}} e^{\frac{v_0^2}{v_{r2}^2}} F\left(\frac{v_0}{v_{r2}}\right), \quad (10)$$

$$t_{t3} = \frac{1}{\omega_{t3}} \left( \tanh^{-1} \left( \frac{v_0}{\sqrt{v_0^2 + v_{r3}^2}} \right) + \frac{v_0 \sqrt{v_0^2 + v_{r3}^2}}{v_{r3}^2} \right). \quad (11)$$

Implicit differentiation of  $t$  allows us to determine  $r' = \frac{dr}{dr_0}$  which is used in the density evolution expression

$$\rho_d(r, t) = \frac{\rho_{0d}}{\left(\frac{r}{r_0}\right)^{d-1} r'}. \quad (12)$$

To obtain an expression for  $r'$ , we need to take the derivative of the time with respect to  $r_0$  while holding  $t$  constant, and then we solve for  $r'$ . We present the results of this process written in terms of time, the ratio  $\frac{r}{r_{td}}$ , and the initial conditions

$$r' = \begin{cases} -y_d r_{td} \omega_{td} t'_{td} + y_d r_{td} \omega'_{td} t_{fid} + \frac{r}{r_{td}} r'_{td}, & t < -t_{td}, \\ y_d r_{td} \omega_{td} t'_{td} + y_d r_{td} \omega'_{td} t_{fid} + \frac{r}{r_{td}} r'_{td}, & t \geq -t_{td}, \end{cases} \quad (13)$$

for  $d = 2, 3$  for the cylindrical and spherical symmetric case, respectively. Notice that all of the derivatives on the right hand side can be written in terms of  $r_0$ ,  $v_0$ ,  $v'_0$ , and  $\rho_{0d}$ ; namely

$$r'_{td} = \frac{r_{td}^{d-1}}{r_0^{d-1}} \left( 1 - 2 \frac{v_0}{v_{rd}} \frac{r_0 v'_0}{v_{rd}} + d \frac{v_0^2}{v_{rd}^2} \frac{\rho_{0d}}{\bar{\rho}_{0d}} \right), \quad (14)$$

$$\omega'_{d3} = \frac{d}{2} \frac{\omega_{td}}{r_0} \left( \frac{\rho_{0d}}{\bar{\rho}_{0d}} - \frac{r_0}{r_{td}} r'_{td} \right), \quad (15)$$

$$t'_{t2} = -\frac{t_{t2}}{\omega_{t2}} \omega'_{t2} + \frac{2}{\omega_{t2}} \frac{v_0}{v_{r2}} e^{\frac{v_0^2}{v_{r2}^2}} \frac{1}{r_0} \left( \frac{r_0 v'_0}{v_0} - \frac{\rho_{02}}{\bar{\rho}_{02}} \right), \quad (16)$$

$$t'_{t3} = -\frac{t_{t3}}{\omega_{t3}} \omega'_{t3} + \frac{1}{\omega_{t3}} \frac{v_0}{v_{r3}} \sqrt{1 + \frac{v_0^2}{v_{r3}^2}} \frac{1}{r_0} \left( 1 + 2 \frac{r_0 v'_0}{v_0} - 3 \frac{\rho_{03}}{\bar{\rho}_{03}} \right), \quad (17)$$

so Eq. (13) leads to an analytic form for  $\rho_d(r, t)$  through Eq. (12). Note that the condition on the time corresponds to the same  $\pm$  condition seen with the velocity and the time-position relation.

## COMPARISON TO CYLINDRICALLY-SYMMETRIC SIMULATIONS

We first demonstrate the use of these equation with initially uniform distributions under cylindrical symmetry. Within the initial distribution, we introduce a velocity term that is linear in the initial position, specifically it has the

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

form  $v_0(r_0) = C \frac{r_0}{R}$  where  $C$  is a simulation dependent constant and  $R$  is the initial radius of the uniform distribution. This form for the velocity was chosen as it models the linear kick received by a distribution as it passes through a typical focussing lens. For our demonstration, we chose  $\Sigma_{\text{tot}} = 2 \times 10^7 \frac{e}{m}$  and  $R = 1$  mm, and this corresponds to a  $v_{r2} \approx 10^5 \frac{m}{s} \frac{r_0}{R}$ . We used the electrostatic Poisson solver in PIC from warp [17] to simulate the evolution of the bunch. Figure 1 shows the results for two values chosen for  $C$ , one positive and one negative, both having magnitude equal to the constant associated with  $v_{r2}$ . Notice the excellent agreement between theory and simulation in all cases; this is in part due to both the initial velocity and  $v_{r2}$  having the same functional form,  $\frac{r_0}{R}$ , thus the distribution remains laminar.

We next demonstrate the use of these equations with initially Gaussian distributions under cylindrical symmetry. Again we introduce the same initial velocity relation,  $v_0(r_0) = C \frac{r_0}{R}$ , choosing  $\Sigma_{\text{tot}} = 4 \times 10^7 \frac{e}{m}$  and  $\sigma_r = 1$  m, which

corresponds to a  $v_{r2} \approx 1.4 \times 10^5 \frac{m}{s} \sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$ . Unlike the uniform case, the functional form for the initial velocity and the velocity scale differ. As we are interested in the emergence of the shock, we focus our analysis on whether a shock emerges, and if so, the period of time during which the shock emerges. Figure 2 shows the evolution for the three positive values of  $C$ . We see that for  $C = 10^4 \frac{m}{s}$  the shock emerges around 22 ns instead of the 20 ns emergence seen in the cold case [16]. For  $C = 5 \times 10^4 \frac{m}{s}$ , the shock is less noticeable and emerges in the vicinity of 50 ns. We do not see the emergence of the shock when  $C = 10^5 \frac{m}{s}$  even at times  $> 100$  ns.

Figure 3 shows the evolution for the three negative values of  $C$ . We see that for  $C = -10^4 \frac{m}{s}$  the shock emerges around 18 ns instead of the 20 ns emergence seen in the cold case, and for  $C = -5 \times 10^4 \frac{m}{s}$  the shock seems to emerge in the vicinity of 11 ns although the variation of the simulated density from the theoretical expectation is much larger for this simulation than for the previously investigated simulations. Interestingly, the model predicts qualitatively different behavior than what is seen in simulation for  $C = -10^5 \frac{m}{s}$ . Specifically, the mean-field fluid model predicts that the distribution begins to expand much earlier than what is seen in simulation. We believe this is due to many Lagrangian particles violating the laminar assumption leading to the incorrect assignment of force to a large proportion of the Lagrangian particles. Specifically, as

$v_0 = C \frac{r_0}{R}$  and  $v_{r2} \approx 1.4 \times 10^5 \frac{m}{s} \sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$  results in a the Lagrangian particles mixing as they got to  $r_{t2}$ .

To address this concern, we again simulate the cylindrically-symmetric distribution but with  $v_0 = C \sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$  for  $C < 0$ . This velocity profile has the main advantage of  $\frac{v_0}{v_{r2}} = \frac{C}{10^5 \frac{m}{s}}$ , which is independent of  $r_0$ . This results in  $r_{t2} = \alpha r_0$  where  $\alpha = e^{-\frac{c^2}{10^{10} \frac{m^2}{s^2}}}$  where  $\alpha$  is indepen-

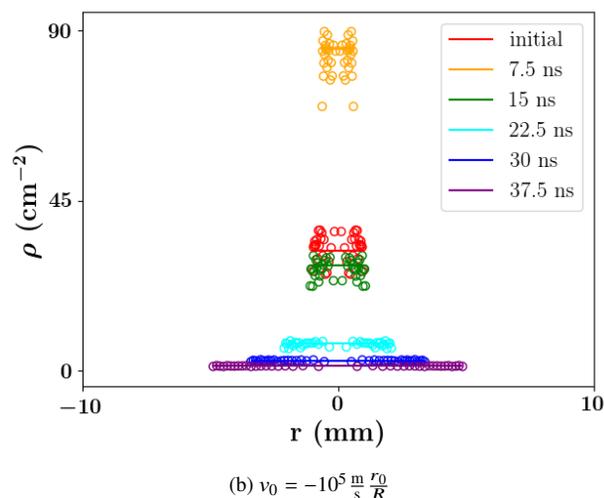
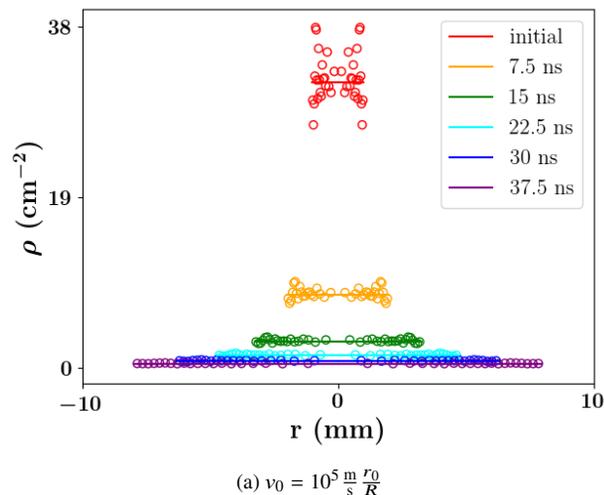


Figure 1: The evolution of uniformly distributed electrons with density of  $2 \times 10^7 \frac{e}{m}$  in a  $R = 1$  mm radius and where  $r_0$  represent the radial position of the particle. The scale of the initial velocity was chosen to be approximately the same size as the scale of  $v_{r2}$ . Solid lines are from the theory presented in this paper and circles are from a single PIC simulation for each figure. Notice that the mean-field fluid model captures the evolution of the bunch in both cases. Specifically, notice that the model correctly captures the contraction and re-expansion of the uniform distribution in the negative case.

dent of  $r_0$  and  $t_{t2} = \frac{2\alpha}{\omega_{02}} e^{-\frac{c^2}{10^{10} \frac{m^2}{s^2}}} F\left(\frac{C}{10^5 \frac{m}{s}}\right)$ ; that is, the turn around points are simply scaled from the initial Gaussian, although they still occur at different times as  $\omega_{02}$  is still dependent on  $r_0$ . As can be seen in Fig. 4, this distribution does appear to remain laminar through the focus as the theory is now in agreement with simulation when  $v_0 > v_{r2}$ ; however, this comes at a cost of an early-emergence of the shock that can be seen at 10 ns in Fig. 4b.

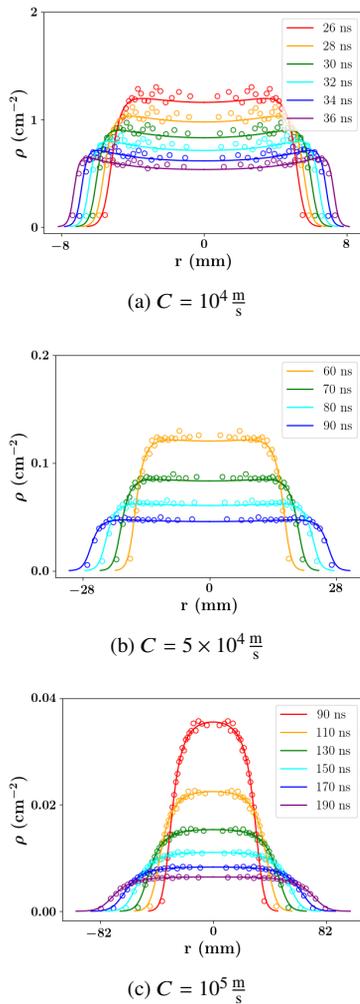


Figure 2: The evolution of Gaussian distributed electrons with density of  $4 \times 10^7 \frac{e}{m}$  with  $\sigma_r = 1$  mm and with initial velocity of the particle given by  $v_0 = C \frac{r_0}{R}$  where  $r_0$  represent the radial position of the particle and  $C > 0$ . Solid lines are from the theory presented in this paper and circles are from a single PIC simulation for each figure. Notice that the mean-field fluid model captures the evolution of the bunch in all cases. Also notice that large value of  $C$  appears to transform the evolution of the bunch into a uniform-like structure and that the bunch apparently loses the emergence of a shock.

## COMPARISON TO SPHERICALLY-SYMMETRIC SIMULATIONS

We now demonstrate that the analysis for systems with spherical symmetry is also accurate for a wide range of initial conditions. As for the cylindrical case, we introduce a velocity term that is linear in the initial position, specifically it has the form  $v_0(r_0) = C \frac{r_0}{R}$  where  $C$  is a simulation dependent constant and  $R$  is the initial radius of the uniform distribution. For our demonstration, we chose  $Q_{\text{tot}} = 2 \times 10^4 e$  and  $R = 1$  mm, and this again corresponds

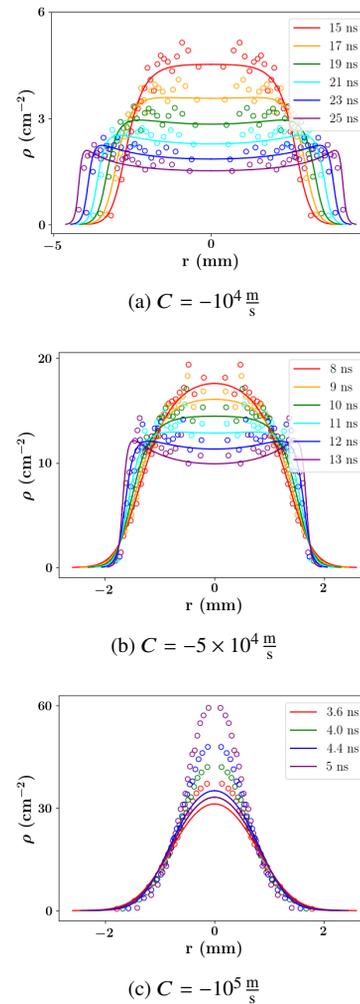


Figure 3: The evolution of Gaussian distributed electrons with density of  $4 \times 10^7 \frac{e}{m}$  with  $\sigma_r = 1$  mm and with initial velocity of the particle given by  $v_0 = C \frac{r_0}{R}$  where  $r_0$  represent the radial position of the particle and  $C < 0$ . Solid lines are from the theory presented in this paper and circles are from a single PIC simulation for each figure. While the model may be an acceptable approximation for small negative values of  $C$ , the mean-field fluid model gets progressively worse as  $C$  becomes more negative and provides qualitatively incorrect predictions when  $C = -10^5$ . The reasons for this are discussed in the text.

to a  $v_{r3} \approx 10^5 \frac{\text{m}}{\text{s}} \frac{r_0}{R}$ . We used the electrostatic Poisson solver in PIC from warp [17] to simulate the evolution of the bunch. Figure 5 shows the results for the same two values chosen for  $C$ ,  $\pm 10^5 \frac{\text{m}}{\text{s}}$ , again approximately equal to the constant associated with  $v_{r3}$ . Notice the excellent agreement between theory and simulation in all cases thus validating the use of the spherically-symmetric formulation.

## CONCLUSIONS

Here, we presented a mean-field fluid model for the evolution of cylindrically and spherically symmetric charged

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

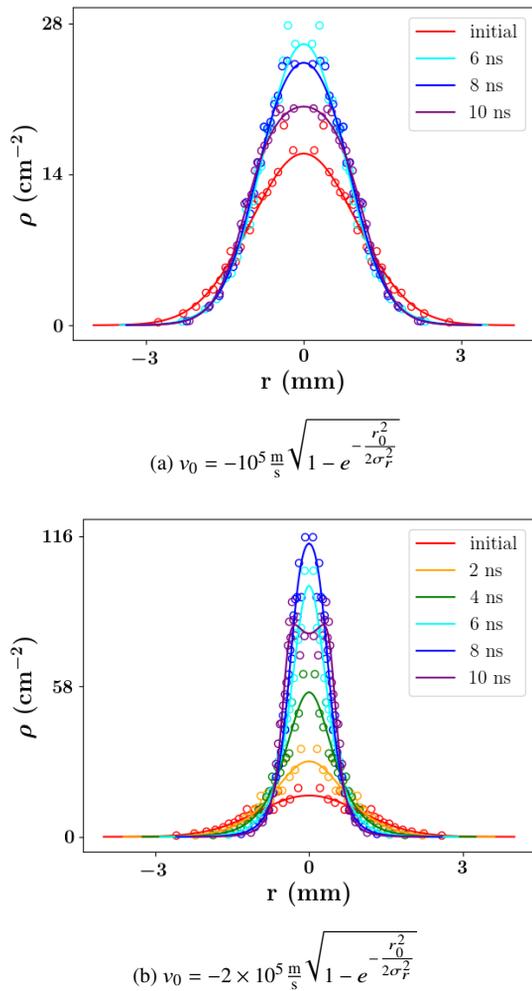


Figure 4: The evolution of Gaussian distributed electrons with density of  $4 \times 10^7 \frac{e}{\text{m}}$  in a  $R = 1 \text{ mm}$  radius and where  $r_0$  represent the radial position of the particle. The functional form of the initial velocity was chosen to be similar to  $v_{r2}$ . Solid lines are from the theory presented in this paper and circles are from a single PIC simulation for each figure. Notice that the mean-field fluid model captures the evolution of the bunch in both cases despite the model failing for linear initial velocity of the same scale as seen in Fig. 3. Notice that for (b), a shock emerges between 8 and 10 ns.

bunches with arbitrary initial distribution and initial velocity that can be written as a function of the radial coordinate. We demonstrated that this model predicts the density evolution of the initially uniform bunch when the initial velocity distribution is linear under both spherical and cylindrical geometries. In the cylindrical geometry, we showed that the shock that arises in the cold Gaussian distribution can be suppressed by introducing a initial radially-outward velocity distribution whose linear proportionality constant is of the order or greater than  $\sqrt{\frac{q\Lambda_{\text{tot}}}{\pi m \epsilon_0}}$ . However, when an analogous negative linear velocity distribution is introduced, the model disagrees with simulations as the initial velocity results in the violation of the laminar assumption. Nonetheless, by

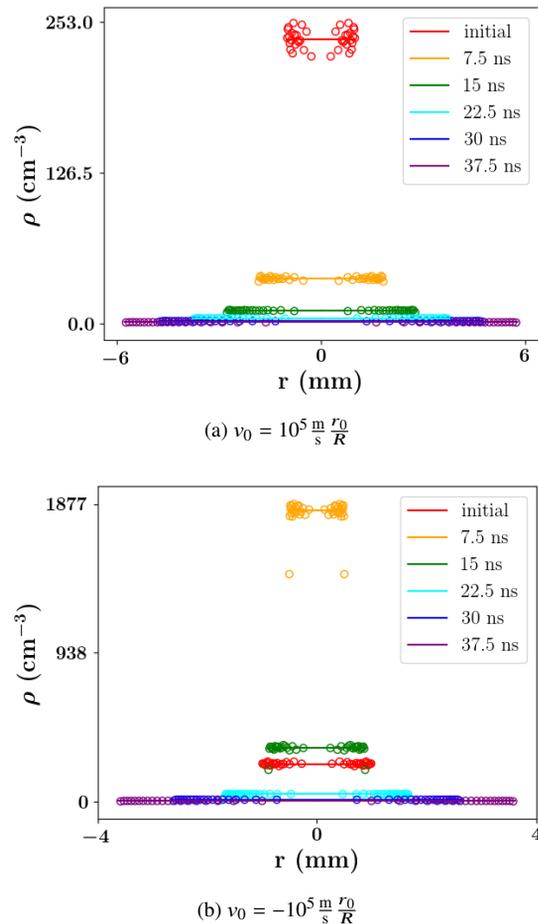


Figure 5: The evolution of uniformly distributed electrons with density of  $Q_{\text{tot}} = 2 \times 10^4 e$  with  $R = 1 \text{ mm}$  and where  $r_0$  represent the initial radial position of the particle. Solid lines are from the theory presented in this paper and circles are from a single PIC simulation for each figure. Like the cylindrically symmetric case with this initial velocity distribution, notice that the theory is in agreement with the simulations capturing the density of the bunch both as it contracts as well as expands.

adjusting the functional form of the initially velocity distribution from  $\frac{r_0}{R}$  to  $\sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$ , we demonstrated that the model can predict the evolution of the initially Gaussian distribution through the focus including the emergence of a shock. This suggests that the laminar assumption for the Gaussian distribution is not violated by this functional form of the velocity distribution, at least for the duration of time we simulated.

The velocity scales derived in this paper,  $v_{r2}$  and  $v_{r3}$ , present the means to qualitatively understand when the laminar assumption can be made. For the uniform distribution,  $v_{rd} \propto \frac{r_0}{R}$ . Thus linear momentum kicks should result in the evolution of the distribution remaining laminar even when the kick is inward; however, if the inward kick has a functional where the slope of the function is beyond linear, say

$v_0 \propto \frac{r_0^2}{R^2}$ , outer Lagrangian particle trajectories will cross the trajectories of inner Lagrangian particles and the laminar assumption will be violated. Analogously for the Gaussian

distribution,  $v_{rd} \propto \sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$ , so it is the slope of this function that matters; that is, the linear kick, where  $v_{rd} \propto \frac{r_0}{R}$ ,

has slope beyond  $\sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$  resulting in violation of the

laminar assumption. On the other hand, using  $\sqrt{1 - e^{-\frac{r_0^2}{2\sigma_r^2}}}$  as the functional form of the velocity distribution does retain as seen in this work; thus such focussing follows the laminar assumption to the point where the laminar assumption is violated by the shock dynamics as we have discussed in our previous work [16].

In other words, the model we have presented here provides an accurate description of the density evolution of a beam as it expands and focusses as long as the beam dynamics exhibits laminar flow. The model also lends important insight into what parameters drive the beam into non-laminar conditions; specifically shortly after the emergence of a shock and when the focussing kick has a functional form beyond what is needed for the specific distribution. Of course, our theoretical initial distributions are still technically cold as  $v_0$  is exactly specified by  $r_0$ ; however, the initial spatial distribution in the simulations was sampled, and this process does make the beam warm. Despite the beam being warm, though, the model correctly predicts the focussing behavior of both the initially uniform and Gaussian distributions provided that the laminar criteria are met. Presumably there if the temperature is high enough, the model will fail, and an exploration of this condition is one of our current goals; as is extending the analysis presented here to the relativistic regime. The question of whether it is better to remain within this laminar regime or to allow mixing is also worth investigating as we now understand many of the conditions to prevent Lagrangian particle mixing.

## ACKNOWLEDGEMENTS

This work was supported by NSF Grant numbers 1803719 and RC108666. Useful discussions with Steve Lund and Chong-Yu Ruan are gratefully acknowledged.

## REFERENCES

- [1] M. Grech, R. Nuter, A. Mikaberidze, P. Di Cintio, L. Gremillet, E. Lefebvre, U. Saalman, J. M. Rost, and S. Skupin, "Coulomb explosion of uniformly charged spheroids," *Physical Review E*, vol. 84, no. 5, pp. 056404–056414, 2011.
- [2] A. E. Kaplan, B. Y. Dubetsky, and P. L. Shkolnikov, "Shock shells in coulomb explosions of nanoclusters," *Physical review letters*, vol. 91, no. 14, p. 143401, 2003.
- [3] V. Kovalev and V. Y. Bychenkov, "Kinetic description of the coulomb explosion of a spherically symmetric cluster," *Journal of Experimental and Theoretical Physics*, vol. 101, no. 2, pp. 212–223, 2005.

- [4] I. Last, I. Schek, and J. Jortner, "Energetics and dynamics of coulomb explosion of highly charged clusters," *The Journal of chemical physics*, vol. 107, no. 17, pp. 6685–6692, 1997.
- [5] D. Murphy, R. Speirs, D. Sheludko, C. Putkunz, A. McCulloch, B. Sparkes, and R. Scholten, "Detailed observation of space-charge dynamics using ultracold ion bunches," *Nature communications*, vol. 5, 2014.
- [6] V. P. Degtyareva, M. A. Monastyrsky, M. Y. Schelev, and V. A. Tarasov, "Computer studies on dynamics of ultrashort electron packets in streak tubes and diffractometers," *Optical Engineering*, vol. 37, no. 8, pp. 2227–2232, 1998.
- [7] J. Williams, F. Zhou, T. Sun, K. Chang, K. Makino, M. Berz, P. Duxbury, and C.-Y. Ruan, "Active control of bright electron beams with rf optics for femtosecond microscopy," *Structural Dynamics*, vol. 4, 2017.
- [8] T. P. Wangler, "Emittance growth from space-charge forces," Los Alamos National Lab., NM (United States), Tech. Rep., 1991.
- [9] O. Anderson, "Internal dynamics and emittance growth in space charge dominated beams," *Part. Accel.*, vol. 21, pp. 197–226, 1987.
- [10] L. Serafini and J. B. Rosenzweig, "Envelope analysis of intense relativistic quasilaminar beams in rf photoinjectors: ma theory of emittance compensation," *Physical Review E*, vol. 55, no. 6, p. 7565, 1997.
- [11] J. Rosenzweig, A. Cook, R. England, M. Dunning, S. Anderson, and M. Ferrario, "Emittance compensation with dynamically optimized photoelectron beam profiles," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 557, no. 1, pp. 87–93, 2006.
- [12] O. J. Luiten, S. B. vanderGeer, M. J. deLoos, F. B. Kiewiet, and M. J. vanderWiel, "How to realize uniform three-dimensional ellipsoidal electron bunches," *Physical review letters*, vol. 93, no. 9, p. 094802, 2004.
- [13] B. J. Siwick, J. R. Dwyer, R. E. Jordan, and R. J. Dwayne Miller, "Ultrafast electron optics: Propagation dynamics of femtosecond electron packets," *Journal of Applied Physics*, vol. 92, no. 3, pp. 1643–1648, 2002.
- [14] B.-L. Qian and H. E. Elsayed-Ali, "Electron pulse broadening due to space charge effects in a photoelectron gun for electron diffraction and streak camera systems," *Journal of Applied Physics*, vol. 91, no. 1, pp. 462–468, 2002.
- [15] B. W. Reed, "Femtosecond electron pulse propagation for ultrafast electron diffraction," *Journal of Applied Physics*, vol. 100, no. 3, pp. 034916–034931, 2006.
- [16] B. Zerbe, X. Xiang, C.-Y. Ruan, S. Lund, and P. Duxbury, "Dynamical bunching and density peaks in expanding coulomb clouds," *Physical Review Accelerators and Beams*, vol. 21, no. 6, p. 064201, 2018.
- [17] A. Friedman, R. H. Cohen, D. P. Grote, S. M. Lund, W. M. Sharp, J.-L. Vay, I. Haber, and R. A. Kishek, "Computational methods in the warp code framework for kinetic simulations of particle beams and plasmas," *IEEE Transactions on Plasma Science*, vol. 42, no. 5, pp. 1321–1334, 2014.

# FEL SIMULATION USING THE LIE METHOD\*

Kilean Hwang<sup>†</sup>, Ji Qiang, Lawrence Berkeley National Laboratory, Berkeley, USA

## Abstract

Advances in numerical methods for free-electron-laser (FEL) simulation under wiggler period averaging (WPA) are presented. First, WPA is generalized using perturbation Lie map method. The conventional WPA is identified as the leading order contribution. Next, a widely used shot-noise modeling method is improved along with a particle migration scheme across the numerical mesh. The artificial shot noise arising from particle migration is suppressed. The improved model also allows using arbitrary mesh size, slippage resolution, and integration step size. These advances will improve modeling of longitudinal beam profile evolution for fast FEL simulation.

## INTRODUCTION

FEL design optimization often involve multiple times of numerical simulations with different system parameters. Such a study requires highly efficient simulation code. The WPA is the one of the best options. Indeed, most of the start-to-end design codes choose to incorporate the WPA FEL simulation code [1–5]. In this proceeding, we review the advances in numerical methods within the WPA framework presented in our previous work [6].

First, we generalize the WPA using the perturbation Lie map method. The conventional WPA is identified as the leading order contribution. The next order corrections we includes are coupling between betatron and wiggling motion, transverse field gradient, and longitudinal field envelope variation.

Second, we present an improved shot-noise model within WPA framework. Unless particle migration across the numerical mesh is artificially suppressed, as in many old WPA codes [7], there can be large artificial shot noise due to the nature of the shot-noise modeling method [8]. We solve this problem by re-interpreting and combining the two widely used shot-noise modeling methods of Refs. [8] and [9]. The improved modeling can further benefit smoothness of numerical mesh through arbitrary weight and shape functions. Here, the weight function refers to the integral kernel used for particle deposition on numerical mesh points. The shape function refers to the shape of the field representation at numerical mesh points used for field interpolation from mesh points to particles' coordinates. This, in turn, allows arbitrary mesh size, integration step size and slippage resolution. Especially, the arbitrary slippage resolution comes with many other benefits. It can naturally simulate correct slippage through non-resonant transport line other than wiggler and allows applying the operator split-composition method [10] on field solver for better accuracy. Last but the

least, the particle loading method can naturally accept the particle data from upstream simulation enabling start-to-end simulation seamless.

## GENERALIZATION OF WIGGLER PERIOD AVERAGING

In general, when a Hamiltonian can be decomposed into integrable part and a small parametric potential, one can build a perturbation map in order of the small parameters. In an undulator, if a map is constructed over a undulator period, the lowest order of the wiggling motion average out leaving the small coupling effects between the fast wiggling and slow betatron motion.

### Magnus Series

We split the Hamiltonian into  $H = S + F(z) + V(z)$  where  $S$  is the wiggler period averaged Hamiltonian representing slow motion,  $V$  is the radiation field potential,  $F$  is the rest representing the fast wiggling motion. Accordingly, we factor the Lie map as the following [11]:

$$\begin{aligned} \mathcal{H}(\lambda_u) &= \mathcal{S}(\lambda_u) \mathcal{F}(\lambda_u) \mathcal{V}(\lambda_u), \\ \mathcal{S}(\lambda_u) &= e^{\mathcal{G}_S(\lambda_u)}, \\ \mathcal{F}(\lambda_u) &= e^{\mathcal{G}_F(\lambda_u)}, \\ \mathcal{V}(\lambda_u) &= e^{\mathcal{G}_V(\lambda_u)}, \end{aligned} \quad (1)$$

where  $\lambda_u$  is the wiggler period. The generators can be written in terms of the Magnus's series

$$\begin{aligned} \mathcal{G}_F(z) &= -\int_0^z dz_1 : F_1^{\text{int}} : + \frac{1}{2!} \int_0^z dz_1 \int_0^{z_1} dz_2 : [2, 1] : \\ &\quad + \frac{1}{3!} \int_0^z dz_1 \int_0^{z_1} dz_2 \int_0^{z_2} dz_3 : [3, [2, 1]] + [[3, 2], 1] : \\ \mathcal{G}_V(z) &= -\int_0^z dz : V^{\text{int}}(z) : \end{aligned} \quad (3)$$

where the pair of colons is Dragt's notation [11] of the Poisson bracket, i.e.,  $: A : B = [A, B]$ , and the numbers in the Poisson bracket is an abbreviation of  $i \equiv F_i^{\text{int}}$  with  $F_i^{\text{int}} \equiv \mathcal{S}(z_i) F(z_i)$ . Since the radiation field strength is much weaker than external field strength only the 1st sequence is taken for  $\mathcal{G}_V$  [6].

### Hamiltonian

Starting from the following Hamiltonian,

$$H(\mathbf{x}, \mathbf{p}, ct, -\gamma; z) = -\sqrt{\gamma^2 - 1 - (p_x - a_x)^2 - (p_y - a_y)^2},$$

where  $ct$  is the time multiplied by the speed of light and it's canonical momentum pair is negative of the normalized

\* Work supported by the Director of the Office of Science of the US Department of Energy under Contract no. DEAC02-05CH11231

<sup>†</sup> kilean@lbl.gov

energy  $-\gamma$ . The normalized vector potential of the planar wiggler is assumed to be

$$\begin{aligned} a_x &= K \cosh(k_x x) \cosh(k_y y) \cos(k_u z) + a_r, \\ a_y &= K \frac{k_x}{k_y} \sinh(k_x x) \sinh(k_y y) \cos(k_u z), \end{aligned} \quad (5)$$

where  $k_u$  is the wiggler wave number,  $k_x$  and  $k_y$  are the natural focusing strength of the wiggler,  $K$  is the normalized wiggler strength. We use the following ansatz for the vector potential  $a_r$

$$a_r = \Re \sum_{h \geq 1} K_h(\mathbf{x}, t; z) e^{ihk_r(z-ct)}, \quad (6)$$

where,  $K_h$  is the  $h$ -th harmonic field envelope, and  $k_r$  is the fundamental wave number. The generating function,

$$G_2(ct, \eta) = [k_r(z-ct) + k_u z] \eta \quad (7)$$

transforms the Hamiltonian into

$$H = (k_u + k_r) \eta - \sqrt{k_r^2 \eta^2 - 1 - (p_x - a_x)^2 - (p_y - a_y)^2},$$

whose new conjugate variables are

$$\theta \equiv k_r(z-ct) + k_u z, \quad \eta \equiv \gamma/k_r. \quad (8)$$

Then, the split parts of the Hamiltonian are

$$\begin{aligned} S &\equiv \frac{k_u}{k_s} \gamma + \frac{1}{2\gamma} \left[ 1 + p_x^2 + p_y^2 + \frac{K^2}{2} (1 + k_x^2 x^2 + k_y^2 y^2) \right] \\ &+ \frac{K^2}{4\gamma} \left[ \frac{1}{3} (k_x^4 x^4 + k_y^4 y^4) + k_x^2 k_u^2 x^2 y^2 \right] \\ &+ \frac{1}{(2\gamma)^3} \left( 1 + K^2 + \frac{3}{8} K^4 \right) + O\left(\frac{q_\perp^6}{\gamma}, \frac{q_\perp^2}{\gamma^3}, \frac{1}{\gamma^5}\right), \end{aligned} \quad (9)$$

$$\begin{aligned} F &\equiv \frac{K_{\text{eff}}^2}{4\gamma} \cos(2k_u z) + \frac{K_{\text{eff}}}{\gamma} p_x \cos(k_u z) \\ &+ O\left(\frac{q_\perp^3}{\gamma}, \frac{1}{\gamma^3}\right), \end{aligned} \quad (10)$$

$$\begin{aligned} V &\equiv -\Re \sum_h \left[ \frac{K_{\text{eff}}}{\gamma} \cos(k_u z) + \frac{p_x}{\gamma} \right] K_h e^{ih(\theta - k_u z)} \\ &+ O\left(\frac{K_h q_\perp^2}{\gamma}, \frac{K_h}{\gamma^3}, \frac{K_h^2}{\gamma}\right), \end{aligned} \quad (11)$$

where  $q_\perp \in \{k_x x, p_x, k_y y, p_y\}$  and

$$K_{\text{eff}} = K \left( 1 + k_x^2 \frac{x^2}{2} + k_y^2 \frac{y^2}{2} \right) \quad (12)$$

is the effective wiggler strength.

### Lie Map Generators

The generator of the slow map is trivial:

$$\mathcal{G}_S(\lambda_u) = -\lambda_u : S : \quad (13)$$

Evaluating Eq. (3), the the fast map generator becomes

$$\mathcal{G}_F(\lambda_u) = -\lambda_u \frac{K^4 k_x^2}{16 k_u^2 \gamma^3}. \quad (14)$$

This corresponds to the coupling between slow betatron oscillation and fast wiggling oscillation. It is negligible in most cases as it scales as  $\gamma^{-2}$  compared to the wiggling motion in Eq. (10). The smallness is due to the large frequency ratio between the betatron and wiggling oscillation. When strong quadrupole is present on top of the wiggler field, such coupling can be more relevant.

Before we evaluate Eq. (4), let us write the propagated field potential  $V^{\text{int}} \equiv \mathcal{S} \mathcal{F} V$  by

$$V^{\text{int}} = - \left[ \frac{K_{\text{eff}}}{\gamma} \cos(k_u z) + \frac{p_x}{\gamma} \right] K_h^{\text{int}} e^{ih(\theta^{\text{int}} - k_u z)}, \quad (15)$$

where the real value operator and the summation over  $h$  is assumed for simplicity, and  $K_h^{\text{int}} \equiv \mathcal{S} \mathcal{F} K_h$  and  $\theta^{\text{int}} \equiv \mathcal{S} \mathcal{F} \theta$  are the propagated field envelope and FEL phase respectively. We neglected the propagation on terms in the square bracket of Eq. (15) assuming propagation on FEL phase and field envelope is more relevant.

In order to evaluate the integration in Eq. (4), we need to model the field envelope variation over the integration range. As the WPA presumes small field envelope variation over a wiggler period, we model as the following:

$$K_h^{\text{int}}(z) = \mathbb{K}_h + \frac{K_{\text{eff}}}{k_u \gamma} \sin(k_u z) \frac{\partial}{\partial x} \mathbb{K}_h + z \partial_z \mathbb{K}_h, \quad (16)$$

where  $\mathbb{K}_h = \frac{1}{z} \int_0^z K_h dz$  is an averaged field envelope,  $\partial_z \mathbb{K}_h$  represent the first order longitudinal variation and the transverse gradient term is from the propagation by  $\mathcal{F}(z)$ . On the other hand, the propagated FEL phase is

$$\theta^{\text{int}} = \theta + \dot{\theta} z - \xi \sin(2k_u z) - \zeta \sin(k_u z), \quad (17)$$

where

$$\begin{aligned} \dot{\theta} &\equiv k_u - \frac{k_r}{2\gamma^2} \left[ 1 + p_x^2 + p_y^2 + \frac{K_{\text{eff}}^2}{2} \right], \\ \xi &\equiv \frac{k_r K_{\text{eff}}^2}{8 k_u \gamma^2}, \\ \zeta &\equiv \frac{k_r K}{k_u \gamma^2} p_x. \end{aligned} \quad (18)$$

Note that  $\dot{\theta} = 0$  exactly on resonance. The inclusion of it encompasses small off-resonant effects. Finally, the generator of the field potential map can be written by

$$\begin{aligned} \mathcal{G}_V &= \lambda_u \Re \sum_h \frac{e^{ih\theta}}{\gamma} \\ &\times \left[ K_{\text{eff}} \int_C^h + p_x \int_1^h + K \int_{z_C}^h \partial_z + \frac{K_{\text{eff}}^2}{k_u \gamma} \int_{SC}^h \partial_x \right] \mathbb{K}_h, \end{aligned} \quad (19)$$

where the integration parameters are [6]:

$$\int_C^h = \frac{1}{2} \left( J_{-\frac{h-1}{2}}^{h\xi_R} + J_{-\frac{h+1}{2}}^{h\xi_R} \right) \left( 1 + \frac{ih\dot{\theta}\lambda_u}{2} \right) - \frac{1}{2} \frac{h\dot{\theta}}{k_u} \left( \sum_{l \neq -\frac{h-1}{2}} \frac{J_l^{h\xi_R}}{(2l+h-1)} + \sum_{l \neq -\frac{h+1}{2}} \frac{J_l^{h\xi_R}}{(2l+h+1)} \right) - \frac{1}{2} \Delta\xi \left( \frac{h-1}{2\xi_R} J_{-\frac{h-1}{2}}^{h\xi_R} + h J_{-\frac{h-3}{2}}^{h\xi_R} + \frac{h+1}{2\xi_R} J_{-\frac{h+1}{2}}^{h\xi_R} + h J_{-\frac{h-1}{2}}^{h\xi_R} \right) + \frac{h\xi}{2} \frac{1}{2} \left( J_{-\frac{h+2}{2}}^{h\xi_R} - J_{-\frac{h-2}{2}}^{h\xi_R} \right), \quad (20)$$

$$\int_1^h = J_{-\frac{h}{2}}^{h\xi_R} \left( 1 + \frac{ih\dot{\theta}\lambda_u}{2} \right) - \frac{h\dot{\theta}}{k_u} \sum_{l \neq -\frac{h}{2}} \frac{J_l^{h\xi_R}}{(2l+h)} - \Delta\xi \left( \frac{h}{2\xi_R} J_{-\frac{h}{2}}^{h\xi_R} + h J_{-\frac{h}{2}+1}^{h\xi_R} \right) + \frac{h\xi}{2} \left( J_{-\frac{h+1}{2}}^{h\xi_R} - J_{-\frac{h-1}{2}}^{h\xi_R} \right), \quad (21)$$

$$\int_{zC}^h = \frac{\lambda_u}{4} \left( J_{-\frac{h-1}{2}}^{h\xi_R} + J_{-\frac{h+1}{2}}^{h\xi_R} \right) + \frac{i\lambda_u}{4\pi} \left( \sum_{l \neq -\frac{h-1}{2}} \frac{J_l^{h\xi_R}}{(2l+h-1)} + \sum_{l \neq -\frac{h+1}{2}} \frac{J_l^{h\xi_R}}{(2l+h+1)} \right), \quad (22)$$

$$\int_{SC}^h = \frac{1}{4i} \left( J_{-\frac{h-2}{2}}^{h\xi_R} - J_{-\frac{h+2}{2}}^{h\xi_R} \right). \quad (23)$$

Here, the integration parameters are calculated to the first order of  $\dot{\theta}$ ,  $\zeta$ ,  $\Delta\xi \equiv \xi - \xi_R$  with  $\xi_R \equiv \frac{k_r K^2}{8k_u \gamma_R^2}$  and  $\gamma_R$  being the resonant energy. The first kind Bessel function of order  $i$  and argument  $h\xi_R$  is abbreviated as  $J_i^{h\xi_R}$  and only integer  $i$  is allowed. Note that the integration parameter  $\int_{SC}$  coupled with the transverse gradient in Eq. (19) vanishes for odd harmonics but can be as large as the leading order terms for even harmonics.

### Effective Hamiltonian

The factorized map Eq. (2) is not yet practically useful for numerical implementation since each factorized map is not solvable and the step size is fixed by one wiggler period. A trick is to concatenate the map using Baker-Campbell-Hausdorff (BCH) formula and define an effective Hamiltonian [11],

$$H_{\text{eff}} = -\frac{1}{\lambda_u} (\mathcal{G}_S + \mathcal{G}_F + \mathcal{G}_V) - \frac{1}{2\lambda_u} (: \mathcal{G}_S : \mathcal{G}_F + : \mathcal{G}_S : \mathcal{G}_V + : \mathcal{G}_F : \mathcal{G}_V) + \dots \quad (24)$$

Note that the effective Hamiltonian does not have explicit dependency on  $z$ . Therefore, we can apply numerical methods like Runge-Kutta (RK) with arbitrary step size to solve the effective Hamiltonian. For example, a popular FEL simulation code GENESIS uses the 4th order RK method [7]. Figure 1 shows an order of magnitude improvement in accuracy of particle pusher compared to the conventional WPA. We used the GENESIS pusher to represent typical WPA. Since, we used converging small enough step sizes, the error from the exact solution originate from the differences of the effective equations of motion in use. However, such an accuracy improvement on particle pusher can easily become obscured by other source of errors such as paraxial approximation of field solver, and numerical discretization.

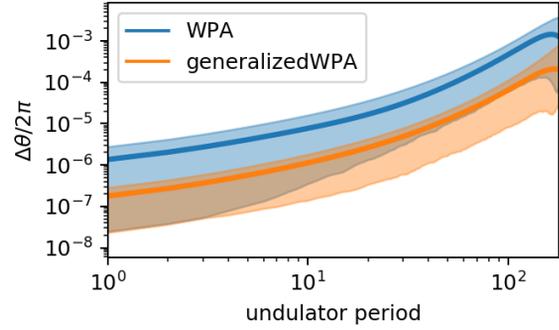


Figure 1: Comparison of particle pushers derived from Lie map and conventional WPA. The error is defined by  $\Delta\theta \equiv |\theta - \theta_{\text{ref}}|$  where  $\theta_{\text{ref}}$  is from a converging small enough step size integration of exact Hamiltonian and the two  $\theta$  are from the 4th order RK tracking of the conventional WPA and the generalized WPA. Exponentially growing gaussian field envelope from 0.4 MW to 1 GW in power,  $\sigma_r = 56 \mu\text{m}$  and  $\lambda_r = 27 \text{nm}$  is assumed. Electron beam parameters used are  $\sigma_{x,y} = 56 \mu\text{m}$ ,  $\epsilon_{x,y} = 0.6 \mu\text{m}$ ,  $\gamma = 1000$  and  $\Delta\gamma/\gamma = 2 \times 10^{-4}$ . Wiggler parameters used are  $K = 1.5$ ,  $\lambda_u = 2.5 \text{cm}$ , and  $k_x = k_y = k_u/\sqrt{2}$ . The thick line represents the average and the shadowed area corresponds to the range of error of simulated particles' population.

## NUMERICAL SHOT NOISE MODELING

### Review of 1D Model

Here, we review two widely used 1D shot noise modeling methods in Ref. [8] and [9]. Figure 2 illustrates these two methods schematically. In both methods, the first step is to populate particles uniformly along the temporal coordinate with equal charge weight to remove artificial temporal shot-

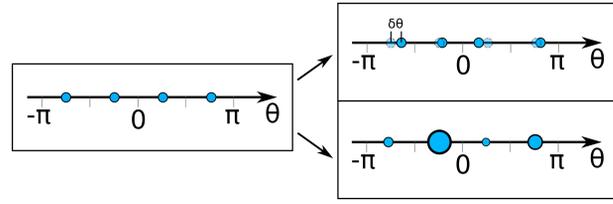


Figure 2: Schematic description of the 1D shot-noise modeling methods. Unperturbed uniform particle loading on left. Temporal coordinate perturbation on the upper right. Charge weight perturbation on the lower right.

noise. The bunching factor at this step is

$$b_h^0 = \frac{1}{N_e} \sum_{j=1}^M m_j e^{ih\theta_j} = 0. \quad (25)$$

Here, index 0 denotes vanishing bunching factor,  $N_e$  is the number of electrons,  $M$  is the number of the simulated particles and  $m_j = N_e/M$  is the electron number weight of the  $j$ -th particle. It vanishes as all the weights are equal, and the temporal coordinates  $\theta_j = \theta_0 + j\Delta\theta$  are uniformly distributed with a equal distance  $\Delta\theta = 2\pi/M$ .

The second step is to add proper perturbation on the temporal coordinates or charge weights so that, at least, the root-mean-square (RMS) bunching factor is that of physical shot-noise  $\langle b_h b_h^* \rangle = 1/N_e$ .

The perturbation on temporal coordinate of  $j$ -th particle is given by [8]

$$\delta\theta_j \equiv \sum_{h'=1}^{M/2} \xi_{h'} e^{-ih'\theta_j}, \quad (26)$$

where the number of particles is assumed to be twice of the maximum harmonic number to be modeled and  $\xi_h$  is a random variable. Then, the bunching factor becomes

$$b_h = \frac{1}{N_e} \sum_{j=1}^M m_j e^{ih(\theta_j + \delta\theta_j)} \simeq ih\xi_h. \quad (27)$$

Therefore, we require  $\langle \xi_h \xi_h^* \rangle = 1/(h^2 N_e)$  so that

$$\langle b_h b_h^* \rangle \simeq h^2 \langle \xi_h \xi_h^* \rangle \equiv \frac{1}{N_e}. \quad (28)$$

On the other hand, when the charge weight perturbation [9] is applied, the RMS bunching factor becomes

$$\begin{aligned} \langle b_h b_h^* \rangle &= \frac{1}{N_e^2} \sum_j^M \sum_k^M \langle \tilde{m}_j \tilde{m}_k \rangle e^{ih(\theta_j - \theta_k)} \quad (29) \\ &= \frac{1}{N_e^2} \sum_j^M \left( \langle \tilde{m}_j^2 \rangle - \langle \tilde{m}_j \rangle^2 \right) = \frac{1}{N_e}, \quad (30) \end{aligned}$$

where  $\tilde{m}_j \equiv m_j + \delta m_j$  is the perturbed number weight with

$$\langle \tilde{m}_j \rangle = \langle \tilde{m}_j^2 \rangle - \langle \tilde{m}_j \rangle^2 = \frac{N_e}{M}. \quad (31)$$

Note that this is a property of the Poisson distribution [9].

## Review of 6D Extension

One way to extend the 1D shot-noise method is the 6D volume division method. This is a natural extension of the 1D charge perturbation based on Poisson principle. Density function is represented by particles sitting at the center of each 6D volume division. As each volume does not overlap, all the particles are statistically independent. However, it requires a lot of particles as division over 6 dimension can be huge.

Another way is to mirror same 5D coordinates  $x, y, p_x, p_y, \gamma$  among a set of particles called ‘beamlet’ [8] which models 1D shot-noise. Note that the member particles of a beamlet are not statistically independent as they share the same 5D phase-space coordinates. However, there can be large numerical shot-noise upon particles migration across the numerical mesh. This is because, the migration break the cancellation of bunching factor in Eq. (25).

Figure 3 illustrates these two 6D extension methods schematically.

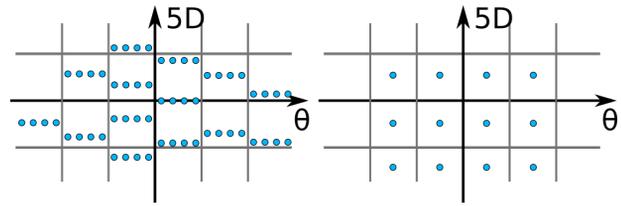


Figure 3: Schematic description of the 6D extension methods of the 1D shot-noise modeling methods. The 5D mirroring method on the left and the 6D volume division method on the right.

## Particle Loading and Migration

As the 6D volume division requires a lot of particles, we adopt the 5D mirroring strategy. Our idea is to interpret one beamlet as one statistically independent entity whose phase-space coordinate is given by the average over the member particles in it. This is based on the observation that the member particles are not statistically independent and the motion of the beamlets describe the macroscopic ( $\geq \lambda_r$ ) dynamics while the motion of the individual member particles of the beamlet describe microscopic ( $\leq \lambda_r$ ) dynamics.

This interpretation allows us to load particles naturally. First, the beamlet is loaded from a random generator or from external upstream tracking code. Then, each beamlet is divided into  $M = 2h_{\max}$  particles whose temporal coordinates and charge follows the 1D shot-noise model while the average coordinate is that of the beamlet. Here  $h_{\max}$  is the maximum harmonic number to be modeled. This procedure is described in Fig. 4.

Since the beamlet is an independent entity, we migrate all the particles composing a beamlet when the beamlet migrate across the numerical mesh. This method grants us much smoother numerical discretization as the weight and shape functions are evaluated at the beamlet position regardless

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

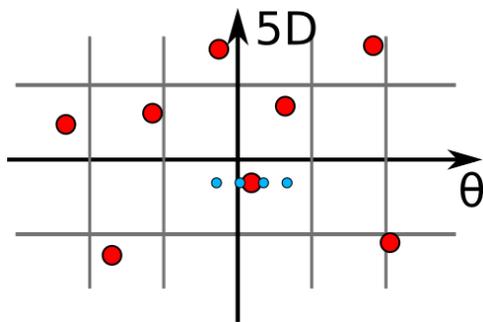


Figure 4: Schematic illustration of beamlets (red) and member particles (blue) composing a beamlet.

of individual member particle's relative coordinates. Figure 5 presents a benchmark between the beamlet migration and GENESIS v. 1. 3 in Self Amplified Spontaneous Emission (SASE) simulation showing a good agreement between the two codes. However, there can be significant deviation from correct model when the artificial shot-noise due to individual particle migration dominate initial emission as shown in Fig. 6.

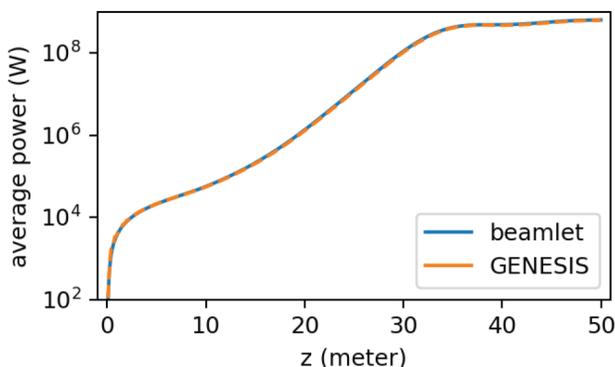


Figure 5: Bench mark between GENESIS v. 1. 3 and beamlet migration of SASE simulation using Next Generation Light Source (NGLS) parameters [12]. Blue is the beamlet migration. Dashed orange is GENESIS v. 1. 3

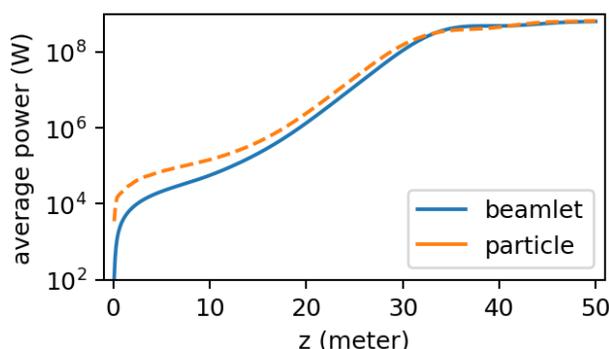


Figure 6: Illustration of artificial shot-noise due to particle migration. Same parameters used as in Fig. 5. Blue is from the beamlet migration. Dashed orange is from the individual particle migration.

## SLIPPAGE RESOLUTION

The beamlet migration enables us to use arbitrary slippage resolution through moving window. The typical implementation of slippage is to copy the field data from the previous temporal mesh point to the next temporal mesh point. This procedure as well as the moving window can be best understood by a pseudo code in Fig. 7. Note that the slippage

```

    Data Copy from previous slice
    for i in [0,1,2,...,nt]:
        Fld.data[:, :, nt-i] = Fld.data[:, :, nt-i-1]

    Moving wondow : change of domain
    Fld.domain.theta[:]
    = Fld.domain.theta[:] + dtheta
    
```

Figure 7: Pseudo code illustrating slippage implementation of the copying data and the moving window. The first two indices of Fld.data are for the transverse mesh points while the last index is for the temporal mesh point. Here, nt is the number of temporal mesh points. The domain range represented by Fld.domain is used by deposition and interpolation algorithm. Therefore, change of the domain range by dtheta effectively slip the field by dtheta.

resolution of the moving window is arbitrary while the slippage resolution of the copying data is one temporal mesh size. Furthermore, moving window with beamlet migration can naturally and correctly model the slippage through arbitrary length of non-resonant transport like drift. Figure 8 shows kinks on power gain curve due to rough slippage resolution for the copying data. The power difference is especially notable in the drift-quadrupole line.

## SPLIT AND COMPOSITION METHOD IN FIELD SOLVER

The field solver can be split into two operation - diffusion and slippage. Let  $\mathcal{F}_\perp$  and  $\mathcal{F}_\parallel$  represent the diffusion and slippage operator respectively. Then, the following composition method of step size  $\Delta z$ ,

$$\mathcal{F}_\parallel \left( \frac{\Delta z}{2} \right) \mathcal{F}_\perp (\Delta z) \mathcal{F}_\parallel \left( \frac{\Delta z}{2} \right) \quad (32)$$

is a second order method provided that  $\mathcal{F}_\perp$  and  $\mathcal{F}_\parallel$  are one-step method of order higher than two [10]. The arbitrary slippage resolution though moving window allow us to split  $\mathcal{F}_\parallel$  and build the 2nd order composition method. Figure 9 shows comparison of the power gain curve error using the 1st order the 2nd order method.

## CONCLUSION

Several advances in numerical methods for FEL simulation under the WPA are presented. We generalized the WPA using the perturbation Lie map method. The perturbative correctional terms to WPA includes coupling between fast

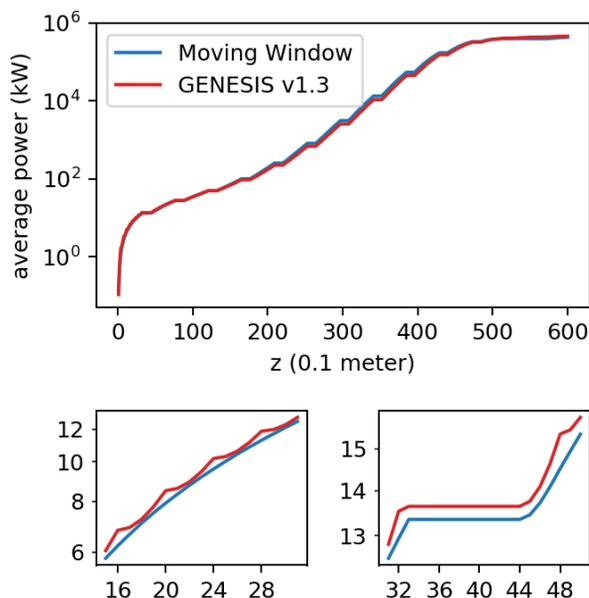


Figure 8: Effect of slippage resolution on power gain curve. Similar parameters are used as Fig. 6. except that drift and quadrupole elements are added here. The integration step size is 5 wiggler period ( $=0.1$  m) and the temporal mesh size is 20 wavelength. The moving window is compared with GENESIS v.1.3 which uses the copying data for field slip.

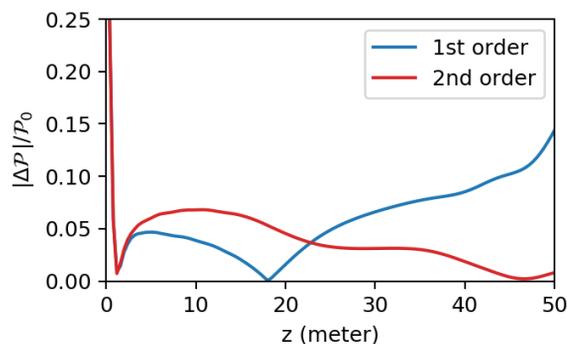


Figure 9: Comparison of the convergence between the two integration orderings.  $P_0$  is the reference power curve,  $\Delta P$  is the difference between the power curve simulated with a large step size ( $\Delta z = 20\lambda_u$ ) and the reference curve. Same parameters are used as Fig. 6. The large deviation at the initial stage is due to shot-noise.

wiggling and slow betatron motion, transverse field gradient, and longitudinal field variation. In addition, we improved the shot-noise modeling method. This allows us to combine advantages of two widely used shot-noise modeling methods in Refs. [8] and [9]. Such an improvement also leads us to develop a particle migration scheme suppressing artificial shot-noise upon migration. This enables much smoother numerical discretization and the temporal beam profile evolution modeling. All these methods are implemented in

the parallel beam dynamics simulation framework IMPACT code suite [13].

## ACKNOWLEDGEMENTS

We appreciate valuable discussion with Gregory Penn. This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and used computer resources at the National Energy Research Scientific Computing Center.

## REFERENCE

- [1] S. Reiche, C. Pellegrini, J. Rosenzweig, P. Emma, and P. Krejčík, “Start-to-end simulation for the LCLS X-ray FEL”, *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 483, pp. 70–74, 2002.
- [2] M. Dohlus, K. Flottmann, O. S. Kozlov, T. Limberg, Ph. Piot, E. L. Saldin, E. A. Schneidmiller, and M. V. Yurkov, “Start to end simulations of SASE FEL at the TESLA Test Facility. Phase 1”, *Nucl. Instrum. Meth., Sect. A*, vol. 530, pp. 217–233, 2004.
- [3] Y. Ding *et al.*, “Start-to-End Simulations of the LCLS Accelerator and FEL Performance at Very Low Charge”, *Proc. PAC’09*, Vancouver, Canada, 2009, paper WE5RFP040
- [4] J. Qiang *et al.*, “Start-to-End Simulation of the LCLS-II Beam Delivery System with Real Number of Electrons”, *Proc. FEL2015*, Daejeon, Korea, 2015, paper WEP070
- [5] J. Qiang, Y. Ding, P. Emma, Z. Huang, D. Ratner, T. O. Raubenheimer, M. Venturini, and F. Zhou, “Start-to-end simulation of the shot-noise driven microbunching instability experiment at the Linac Coherent Light Source”, *Phys. Rev. Accel. Beams*, vol. 20, pp. 054402, 2017.
- [6] K. Hwang, J. Qiang, “Advances in numerical methods under wiggler period averaging for free electron laser simulation”, *Phys. Rev. Accel. Beams* (submitted in 2018).
- [7] S. Reiche, “GENESIS 1.3: A fully 3D time-dependent FEL simulation code”, *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 429, p. 243, 1999.
- [8] William M. Fawley, *Phys. Rev. ST Accel. Beams*, vol. 5, pp. 070701, 2002.
- [9] B. W. J. McNeil, M. W. Poole, and G. R. M. Robb, *Phys. Rev. ST Accel. Beams*, vol. 6, pp. 070701, 2003.
- [10] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration*, Springer, 2006.
- [11] Alex J. Dragt, “Lie Methods for Nonlinear Dynamics”. <http://www.physics.umd.edu/dsat/dsatliemethods.html>
- [12] J. Qiang, J. Corlett, C. E. Mitchell, C. F. Papadopoulos, G. Penn, M. Placidi, M. Reinsch, R. D. Ryne, F. Sannibale, C. Sun, M. Venturini, P. Emma, and S. Reiche, “Start-to-end simulation of x-ray radiation of a next generation light source using the real number of electrons”, *Phys. Rev. ST Accel. Beams*, vol. 17, pp. 030701, 2014.
- [13] J. Qiang, R. Ryne, S. Habib, and V. Decyk, *J. Comp. Phys.*, vol. 163, pp. 434, 2000.

# START-TO-END SIMULATIONS OF THz SASE FEL PROOF-OF-PRINCIPLE EXPERIMENT AT PITZ

M. Krasilnikov\*, P. Boonpornprasert, F. Stephan, DESY, Zeuthen, Germany  
E.A. Schneidmiller, M.V. Yurkov, DESY, Hamburg, Germany  
H.-D. Nuhn, SLAC, Menlo Park, California, USA

## Abstract

The Photo Injector Test facility at DESY in Zeuthen (PITZ) develops high brightness electron sources for modern linac-based Free Electron Lasers (FELs). The PITZ accelerator has been proposed as a prototype for a tunable, high power THz source for pump and probe experiments at the European XFEL. A Self-Amplified Spontaneous Emission (SASE) FEL is considered to generate the THz pulses. High radiation power can be achieved by utilizing high charge (4 nC) shaped electron bunches from the PITZ photo injector. THz pulse energy of up to several mJ is expected from preliminary simulations for 100  $\mu\text{m}$  radiation wavelength. For the proof-of-principle experiments a re-usage of LCLS-I undulators at the end of the PITZ beamline is under studies. One of the challenges for this setup is transport and matching of the space charge dominated electron beam through the narrow vacuum chamber. Start-to-end simulations for the entire experimental setup - from the photocathode to the SASE THz generation in the undulator section - have been performed by combination of several codes: ASTRA, SCO and GENESIS 1.3. The space charge effect and its impact onto the output THz radiation have been studied. The results of these simulations will be presented and discussed.

## INTRODUCTION

The Photo Injector Test facility at DESY in Zeuthen (PITZ) has been suggested as a prototype for developments on an accelerator based high power tunable THz source for pump and probe experiments at the European XFEL [1]. The SASE FEL is considered as main option to generate THz pulses at PITZ using a high bunch charge (4 nC) operation mode of the photo injector. In order to prepare a proof-of-principle experiment start-to-end beam dynamics simulations have been performed. They include generation of electron bunches in the RF photogun, further acceleration by the booster cavity, further transport ( $\sim 25$  m) of the space charge dominated electron beam and its matching into the undulator section. A measured field profile of a typical LCLS-I undulator has been used to reconstruct a 3D magnetic field map used for the beam transport through the undulator section. Best obtained matching solution was used to simulate THz SASE FEL with a centre radiation wavelength of  $\sim 100$   $\mu\text{m}$ .

\* mikhail.kraskilnikov@desy.de

## RF GUN AND BOOSTER

The PITZ RF photogun with a peak cathode field of 60 MV/m operated at the launch phase of maximum mean momentum gain (MMMG) is used to generate 4 nC electron bunches by applying photocathode laser pulses with a flattop temporal profile (21.5 ps FWHM) and with a radially homogeneous transverse distribution. Preliminary emittance optimization yielded the optimum photocathode laser spot size which is larger than the whole cathode size. In order to be closer to the practical case a 5 mm diameter of the photocathode laser spot (coinciding with the size of the photocathode) has been used for further optimizations. Beam mean longitudinal momentum of  $\sim 16.7$  MeV/c required for generating THz radiation with  $\sim 100$   $\mu\text{m}$  radiation wavelength is achieved using a booster cavity. For each booster gradient (peak electric field) the booster phase (w.r.t. MMMG) was tuned in order to yield the required final mean momentum of electron beam. Corresponding curve is shown in Fig. 1a. Besides the mean beam momentum booster cavity gradient and phase were tuned in order to yield a small correlated energy spread ( $\langle zE \rangle \rightarrow 0$ ) of the electron beam close to the undulator ( $z=29$  m). This two-fold optimization resulted in a peak booster field of 12.85 MV/m and a phase of  $-26$  deg. w.r.t. MMMG (Fig. 1a). The main gun solenoid was tuned to control electron beam size and emittance. Beam dynamics simulations were performed using the ASTRA code [2] with 200000 macroparticles. Simulated rms normalized emittance at the location of the first emittance measurement station (EMSY1 at  $z=5.277$  m) is shown in Fig. 1b for the booster MMMG phase and for the optimized correlated energy spread (working point, WP). The chosen absolute value of the gun solenoid peak field of the main solenoid (212.85 mT) is by  $\sim 5\%$  lower than that delivering the minimum emittance for the booster in on-crest operation. Corresponding curves of the beam projected transverse emittance and correlated energy spread along the beam line are shown in Fig. 1c. The WP setup corresponds to a more flat emittance along the beamline, whereas a significant increase of the emittance after the minimum is clearly seen for the case of the MMMG booster phase.

Transverse and longitudinal phase spaces of the optimized electron beam (WP) at the EMSY1 location are shown in Fig. 2. This beam setup was used as a starting point for studies on the space charge dominated beam transport towards the undulator section.

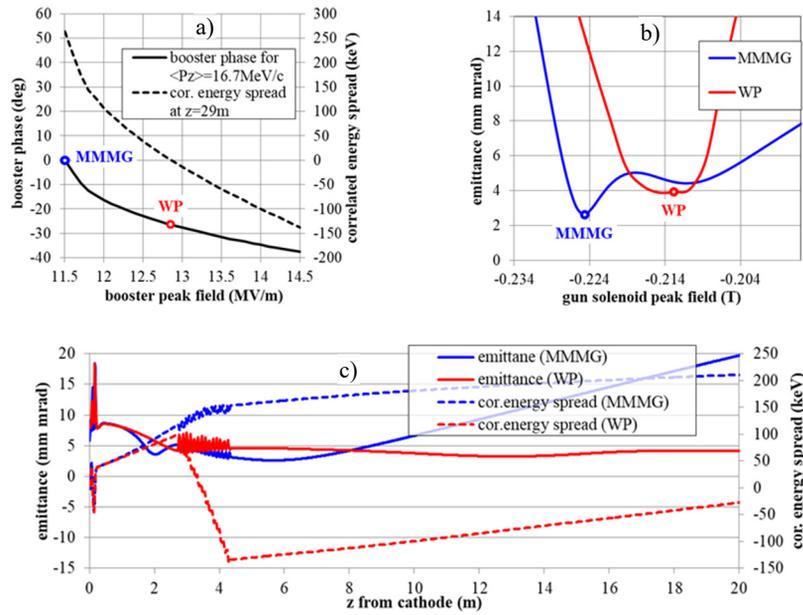


Figure 1: Optimization of RF gun and booster. a) Booster phase w.r.t. MMMG delivering the final beam mean momentum of 16.7 MeV/c. The correlated energy spread at  $z=29 \text{ m}$  is shown at the right axis. b) Beam transverse normalized rms emittance at EMSY1 as a function of gun solenoid peak field. c) Beam transverse emittance and correlated energy spread along the beam line simulated for points marked in a) and b).

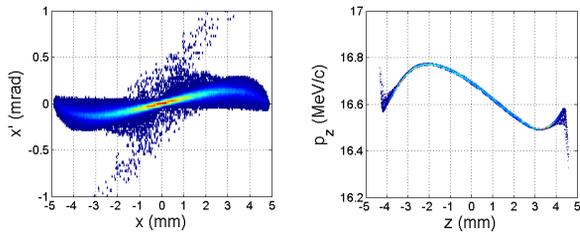


Figure 2: Transverse (left) and longitudinal (right) phase spaces of electron beam at  $z=5.277 \text{ m}$ .

## BEAM TRANSPORT TO UNDULATOR

Currently for a proof-of-principle SASE THz experiment the installation of LCLS-I undulators is foreseen in the PITZ tunnel annex. The concrete wall between main and annex tunnels is 1.5 m thick and starts at  $\sim 24 \text{ m}$  w.r.t. the photocathode plane. This space is considered for electron beam drift only without any focusing elements inside and assuming a standard beam pipe ( $\varnothing 35 \text{ mm}$ ). The beam transport through the undulator vacuum chamber (a race-track profile with  $5 \times 11 \text{ mm}$  cross section and  $3.4 \text{ m}$  length [3]) is an even much harder task. In order to test a feasibility of such a transport a fast space charge tracking code Space Charge Optimizer (SCO) [4] was used. The results of ASTRA tracking till EMSY1 position ( $5.277 \text{ m}$ ) were used as an input for the SCO by applying a corresponding interface. Three triplets of quadrupole magnets were involved into optimization, first two of them were chosen from the magnets available in the present PITZ beamline, and the last one was assumed to be installed at the end of the main tunnel. The solution obtained from the optimized SCO tracking was plugged into the ASTRA

input lattice. Results of SCO and ASTRA simulations are shown in Fig. 3 demonstrating a rather good agreement despite the SCO includes only linear space charge forces.

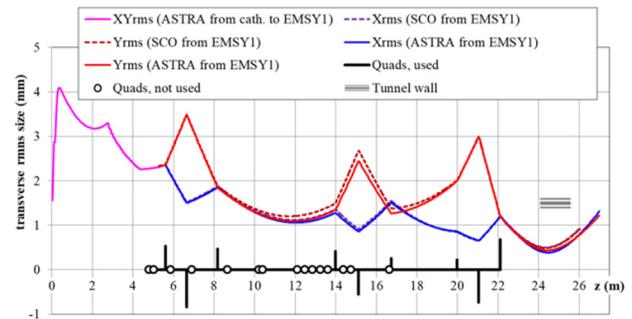


Figure 3: Electron beam rms size along the accelerator simulated using SCO and ASTRA. Gradients of applied quadrupoles are shown schematically on bottom. Circles denote available (but not used) quadrupoles.

## MODELING OF UNDULATOR FIELD

The LCLS-I undulator module is a 3.4-m-long permanent magnet planar hybrid structure with 113 periods of  $\lambda_U = 30 \text{ mm}$  and a magnetic gap of 6.8 mm [3]. The on-axis field measurements data for SN07 undulator [5] was used to reconstruct a 3D field map required for particle tracking. The undulator peak on-axis field is 1.28 T which corresponds to the undulator parameter of  $K=3.585$ . Applying Fourier transformation to the measured field profile  $B_y(x=0, y=0, z)$  centered around  $z=0$  ( $|z| \leq L/2$ ) the vertical component of the magnetic field reads:

$$B_y(0,0,z) = \sum_{n=0}^{\infty} \{a_n \cos(k_n z) + b_n \sin(k_n z)\}, \quad (1)$$

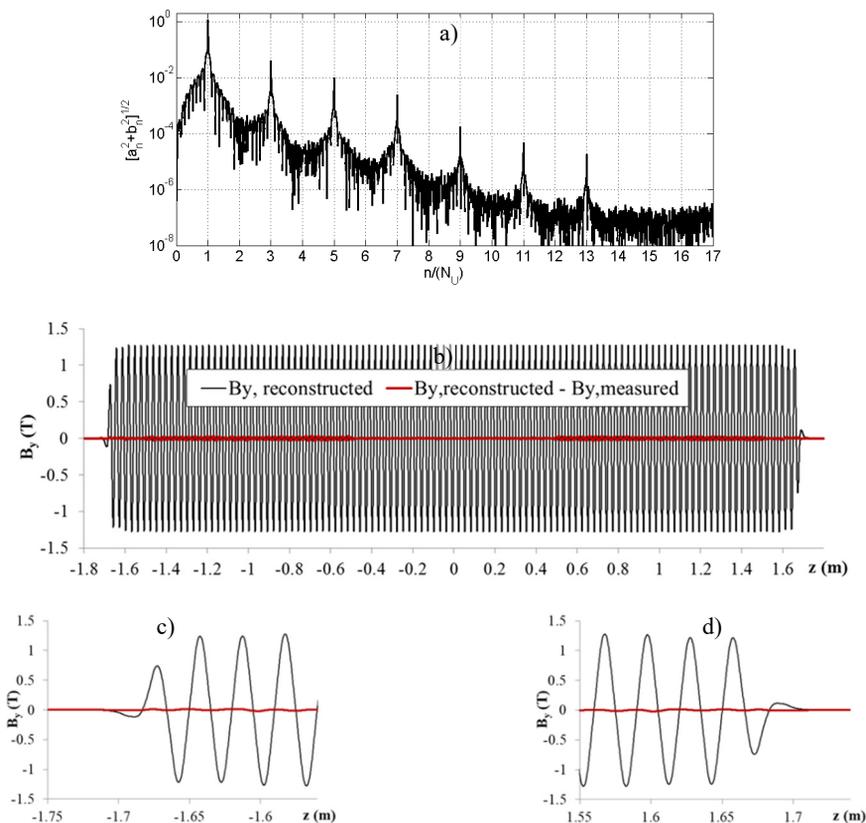


Figure 4: a) Fourier spectrum of the measured undulator field. b) Reconstructed field profile compared with its discrepancy to the measured data. Zoom of the left (c) and right (d) edges of the field profile.

where  $L = N_U \lambda_U$  is the undulator effective length and  $k_n = 2\pi n/L$  is the wavenumber of the  $n$ -th Fourier harmonic. Specific shape of the field profile, including spectral content of the regular periods and the end poles length and strength slope require rather large number of harmonics  $N_h$  to be taken into account to reproduce the measured field profile with sufficient accuracy. Proper centring and fine treatment of the measured field profile yield vanishing first and second field integrals. This corresponds to conditions  $a_n \approx 0$  and  $\sum_{n=1}^{\infty} [(-1)^n b_n/n] \approx 0$  for the first and second integral, correspondingly. These optimization resulted in  $L = 120\lambda_U$  (despite nominally the undulator contains only 113 periods) and  $N_h = 17$  is a harmonic number of the fundamental wavelength  $\lambda_U$ . This corresponds to the first 2040 ( $N_U N_h$ ) terms of the series (1) taken into account. Fourier spectrum and the field profile reconstructed from it are shown in Fig. 4.

Using a 2D approximation (i.e. no field variation in the horizontal direction) the field satisfying Maxwell's equations can be written as follows:

$$B_y = \sum_{n=1}^{N_h \cdot N_U} \{[a_n \cos(k_n z) + b_n \sin(k_n z)] \cdot \cosh(k_n y)\}, \quad (2)$$

$$B_z = \sum_{n=1}^{N_h \cdot N_U} \{[b_n \sin(k_n z) - a_n \cos(k_n z)] \cdot \sinh(k_n y)\}.$$

Equation (2) has been applied to create 3D field maps of the undulator for their subsequent use in ASTRA and CST Particle Studio [6]. Field imported into the CST is shown in Fig. 5 where the right edge of the structure is depicted. A static magnetic field option of 3D cavity name list was used in ASTRA [2] to simulate particle dynamics in the undulator.

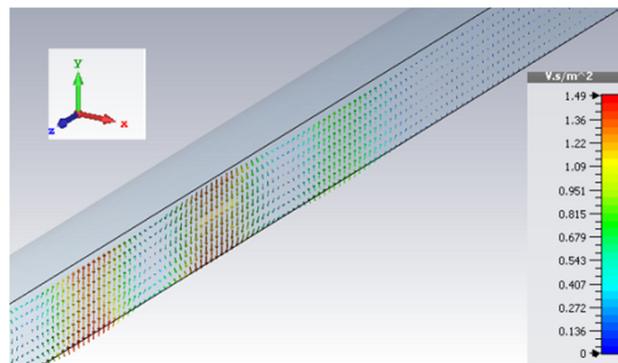


Figure 5: CST model of the undulator field.

The results of the on-axis reference particle tracking by the two above mentioned simulation tools coincide within good accuracy (Fig. 6).

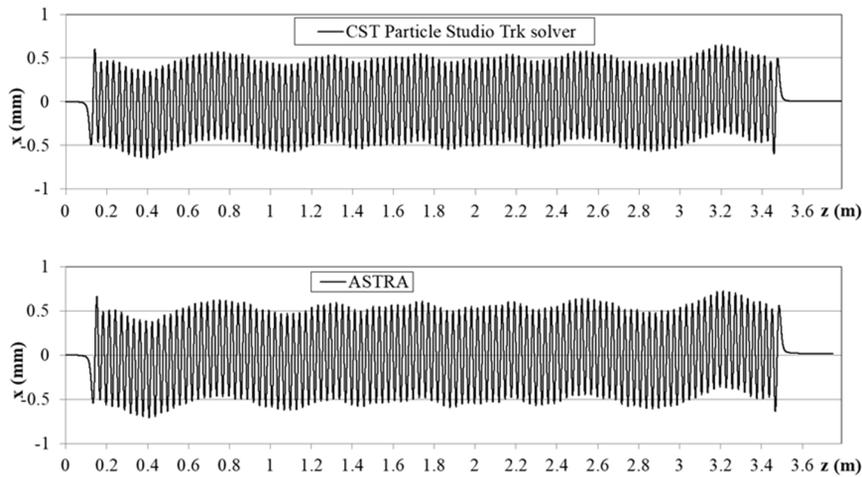


Figure 6: Horizontal trajectory of the on-axis reference particle simulated by: CST Particle Studio Trk solver (upper plot) and ASTRA code (bottom plot).

Off-axis reference particle tracking revealed rather strong vertical focusing by the undulator field (2). The results of these tracking were used for the 4 nC beam matching into the undulator. As a first approximation an ideal model electron distribution with a flat-top temporal (7 mm FWHM) and Gaussian transverse phase spaces (projected emittance of 4 mm mrad) was used. After tuning of the input beam Twiss parameters by running ASTRA an optimum setup has been found (Fig. 7). Results of this matching have shown that for the optimum beam transport through the undulator the electron beam should be rather X-Y asymmetric.

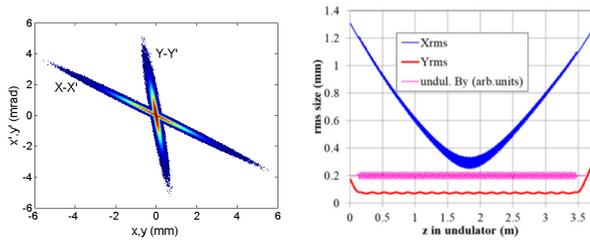


Figure 7: Matching of the model beam into the undulator: left - input transverse phase spaces; right - rms sizes of electron beam in the undulator simulated by ASTRA.

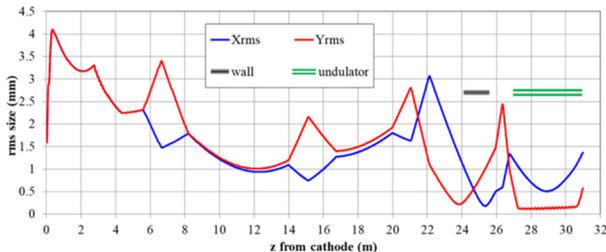


Figure 8: Beam transport in the PITZ linac, including wall and undulator section (ASTRA simulations).

Obtained input parameters were used for the optimization with a beam tracked from the cathode. Finally, undulator section has been included as well. In order to prepare asymmetric matching into the undulator keeping reasonable size of the electron beam inside the wall between the

main tunnel and the annex several quadrupoles were re-tuned. Results of these optimizations are shown in Fig. 8. Transverse and longitudinal phase spaces of electron beam as well as its slice parameters at the undulator entrance are shown in Fig. 9.

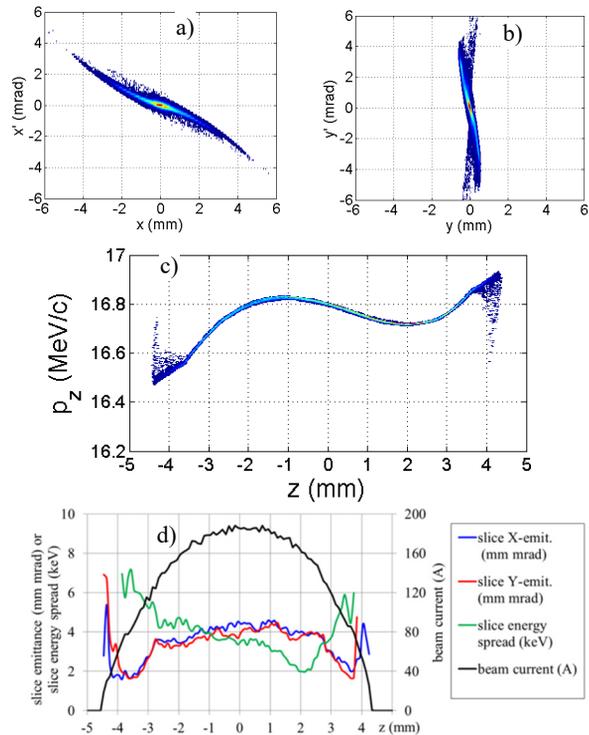


Figure 9: Phase space of the electron beam at the undulator entrance: a) horizontal, b) vertical, c) longitudinal. d) Slice parameters of electron bunch: beam current, slice emittance and slice energy spread.

### THz SASE FEL SIMULATIONS

Electron beam parameters (Fig. 9) were used to simulate THz SASE FEL with GENESIS 1.3 code [7]. Only fundamental wavelength  $\lambda_U$  of the undulator field was included. Number of undulator periods was set to 113; no

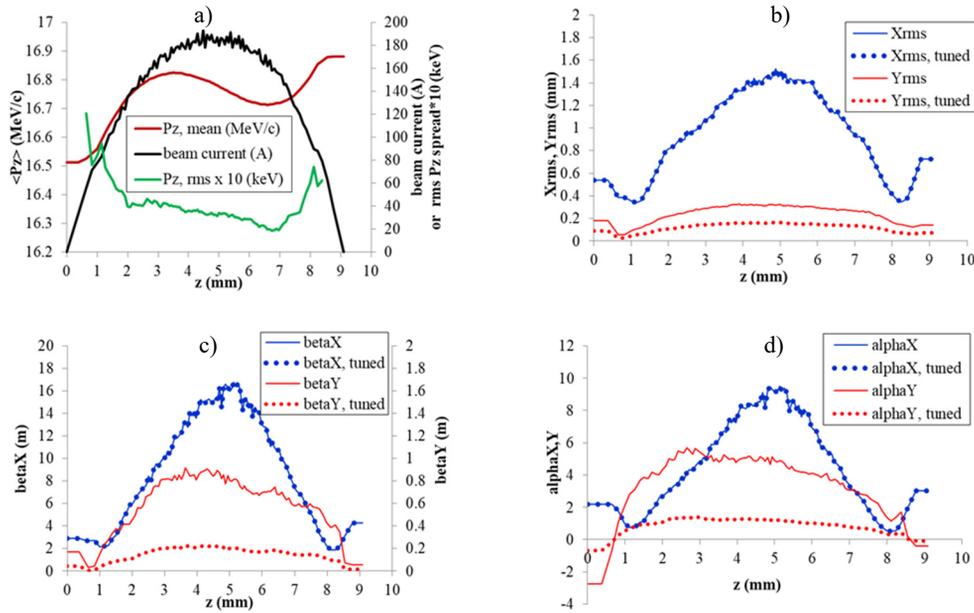


Figure 10: Input electron beam parameters used for GENESIS simulations: a) beam current, mean momentum and slice rms momentum spread; b) transverse rms sizes along the bunch; c) beta functions; d) alpha functions along the bunch.

end cell features were included. Due to these and some other systematic limitations of the simulation tool (e.g., simplified space charge model) an additional tuning w.r.t. input electron beam Twiss parameters was performed to maximize the output THz pulse energy. Mainly vertical Twiss parameters of the input electron beam ( $\beta_y, \alpha_y$ ) were varied by simultaneous scaling of corresponding slice parameters. The emittance (slice and projected) remained unchanged. Main beam parameters along the electron bunch are shown in Fig. 10 for the electron beam directly plugged from the start-to-end simulations (nominal beam) and for the beam with ( $\beta_y, \alpha_y$ ) scaled by a factor of 0.25 (tuned beam).

Results of GENESIS 1.3 simulations using both input beams are shown in Fig. 11. Grey curves refer to single shot realizations, the black curve corresponds to the average over a hundred realizations (simulation seeds). Only one LCLS-I undulator was used for these simulations. The output THz average pulse energy of 440  $\mu$ J for the nominal beam was increased to 600  $\mu$ J by the above mentioned tuning of the vertical Twiss parameters of the input electron beam. Average THz pulse at the undulator exit (middle plots in Fig. 11) has peak power of 32 and 38 MW for nominal and tuned beam correspondingly, the

rms duration for both cases is  $\sim 6$  ps. The average spectrum (black curves in the bottom plots of Fig. 11) has a centre at  $\sim 107 \mu$ m and a width of  $\sim 5 \mu$ m (FWHM). Main parameters of THz pulses obtained from the statistical analysis of hundred realizations are summarized in the Table 1. This includes also an arrival rms time jitter of  $\sim 1$ -2 ps calculated from the centre positions of simulated THz pulses.

Previous simulations of the THz SASE FEL for the PITZ setup with APPLE-II type undulator [8] yielded rather high level of the radiation pulse energy (up to  $\sim 3$  mJ at 100  $\mu$ m wavelength). Current start-to-end simulations of the proof-of-principle experiment resulted in a reduction of this level by a factor of 5. This can be related to the planar type of the LCLS-I undulator in contrary to the helical undulator APPLE-II used in [8]. The planar undulator assumes asymmetric beam matching and therefore less efficient interaction of electrons with radiated field, whereas for the helical undulator the focusing in both transverse planes makes this interaction more homogeneous and efficient. Another source of the reduced THz power output is space charge effect, which should be taken into account by matching and transport of 4 nC and 16 MeV electron bunches in the narrow undulator gap.

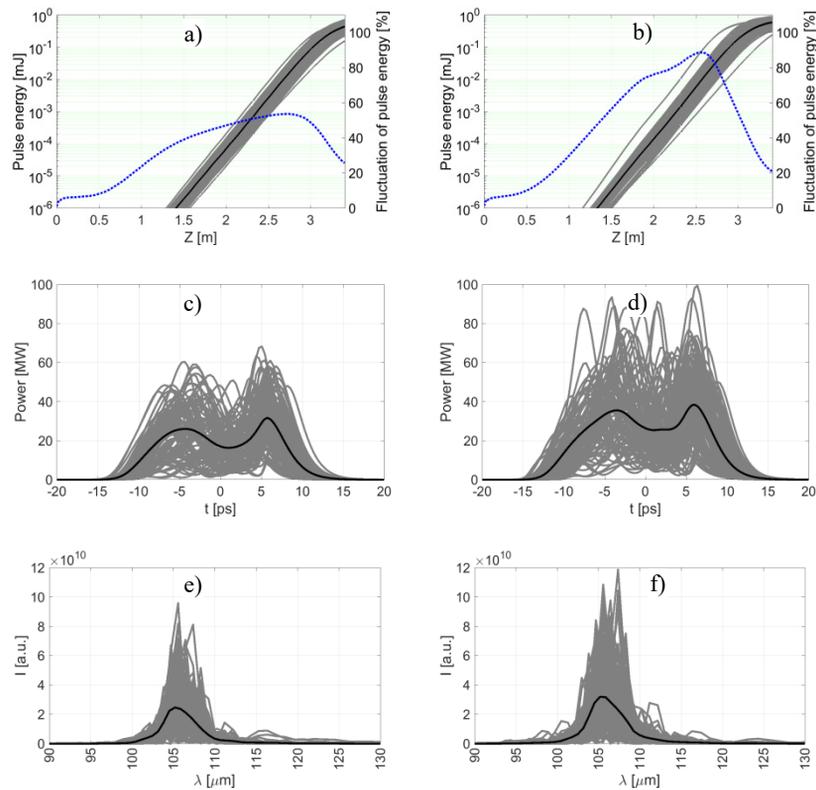


Figure 11: Results of THz SASE FEL simulations with GENESIS code. Pulse energy along the undulator for the nominal (a) and tuned (b) beams. The blue dotted line shows the fluctuation of the pulse energy along the undulator axis. Radiation pulse profile at the undulator exit for the nominal (c) and tuned (d) beams. Corresponding spectra at the undulator exit for the nominal (e) and tuned (f) beams.

Table 1: Simulated THz radiation properties

Parameter	Nominal beam	Tuned beam
Pulse energy (mJ)	0.44±0.11	0.60±0.13
Peak power (MW)	43.0±10.2	58.5±14.3
Pulse rms duration (ps)	5.6±0.7	5.7±0.7
Arrival rms time jitter (ps)	1.7	1.4
Centre wavelength (μm)	106.5	106.8
Spectrum width FWHM (μm)	4.5	4.8

## CONCLUSIONS AND OUTLOOK

Start-to-end beam dynamics simulations have been performed for the proof-of-principle experiment on THz SASE FEL generation at PITZ by using a LCLS-I undulator. Space charge dominated electron beam transport through the PITZ accelerator was optimized combining and iterating ASTRA and SCO codes. A model to generate 3D magnetic field map of the undulator has been developed and implemented. Tracking of the reference particle in the undulator field using CST Trk solver and ASTRA yields similar results. A strategy for the matching of

the 4 nC electron bunch into the planar undulator was proposed. Asymmetric beam matching obtained with Gaussian beam was applied to the electron beam tracked from the photocathode and refined by tracking in the undulator field with included space charge effect. Obtained electron beam 6D phase space at the undulator entrance was used as an input for the THz SASE FEL simulations by means of GENESIS 1.3. Additional tuning by scaling of electron beam vertical phase space resulted in the simulated THz pulse energy increase from 440 μJ to 600 μJ at the centre wavelength of ~100 μm. Several effects are still not considered: possible impact of the narrow vacuum chamber of the undulator (wakefield of electron bunch and waveguide effect of the FEL process). Also the space charge model used in ASTRA and GENESIS simulations inside undulator has a limited applicability. Impact of possible undulator imperfections onto beam transport and FEL radiation has still to be estimated as well.

## ACKNOWLEDGEMENTS

The authors would like to thank Vladimir Balandin and Nina Golubeva for helpful discussions on simulations of the beam trajectory in undulator.

## REFERENCES

- [1] E. Schneidmiller *et al.*, “Tunable IR/THz source for pump probe experiment at European XFEL,” in *Proc. FEL'12*, Nara, Japan, 2012, pp. 503-506.
- [2] ASTRA particle tracking code.  
<http://www.desy.de/~mpyf1o/>.
- [3] LCLS Conceptual Design Report, SLAC-R-593, UC-414.
- [4] A. V. Bondarenko, A. N. Matveenko, “Implementation of 2D-emittance compensation scheme in the BERLinPro injector”, in *Proc. FEL'11*, Shanghai, China, 2011, pp. 564–567.
- [5] LCLS Undulator Segment Measurement Results, L143-112000-007.
- [6] Computer Simulation Technology, Darmstadt, Germany: CST Studio. <http://www.cst.com>.
- [7] S. Reiche, "GENESIS 1.3 User Manual" 2004.  
<http://genesis.web.psi.ch/index.html>
- [8] P. Boonpornprasert, M. Krasilnikov, and F. Stephan, “Calculations for a THz SASE FEL based on the measured electron beam parameters at PITZ”, in *Proc. FEL'17*, Santa Fe, NM, USA, Aug. 2017, pp. 419-421.

# COMPUTATION OF EIGENMODES IN THE BESSY VSR CAVITY CHAIN BY MEANS OF CONCATENATION STRATEGIES\*

Thomas Flisgen<sup>†</sup>, Adolfo Vélez

Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB), 12489 Berlin, Germany

Johann Heller, Shahnam Gorgi Zadeh, and Ursula van Rienen

Institute of General Electrical Engineering, University of Rostock, 18059 Rostock, Germany

## Abstract

The computation of eigenmodes in chains of superconducting cavities with asymmetric couplers is a demanding problem. This problem typically requires the use of high-performance computers in combination with dedicated software packages. Alternatively, the eigenmodes of chains of superconducting cavities can be determined by the so-called State-Space Concatenation (SSC) approach that has been developed at the University of Rostock. SSC is based on the decomposition of the full chain into individual segments. Subsequently, the RF properties of every segment are described by reduced-order models. These reduced-order models are concatenated to a reduced-order model of the entire chain by means of algebraic side constraints arising from continuity conditions of the fields across the decomposition planes. The constructed reduced-order model describes the RF properties of the complete structure so that the field distributions, the coupling impedances and the external quality factors of the eigenmodes of the full cavity chain are available. In contrast to direct methods, SSC allows for the computation of the eigenmodes of cavity chains using desktop computers. The current contribution revises the scheme using the BESSY VSR cavity chain as an example. In addition, a comparison between a direct computation of a specific localized mode is described.

## INTRODUCTION

The computation of eigenmodes of superconducting RF resonators used for the acceleration of charged particles is a standard task in computational accelerator physics. Complementary to the characterization of the accelerating mode, higher-order modes are of special interest as they can interact with the beam as well and may lead to additional cryogenic load or beam instabilities. Often eigenmode computations are restricted to single cavities with couplers to reduce computational efforts, despite the fact that the cavities are arranged in chains and are connected via the beam pipes. These chains are accommodated in cryomodules providing the cryogenic infrastructure to cool the resonators so that their surfaces become superconducting.

The consideration of single cavities with couplers is a reasonable approximation for characterizing the accelerating mode as the field distribution of this mode is on purpose

confined in the cavity. However, the restriction to a single cavity often becomes invalid for higher-order modes, in particular if mode resonant frequencies are larger than the fundamental cutoff frequency of the beam pipe connecting the adjacent cavities. The field distributions of higher-order modes in cavity chains are much more complex than in single cavities as the fields can be distributed along the entire cavity chain or along parts of it.

Direct approaches to determine the eigenmode spectrum of cavity chains require high-performance computers [1–4]. Alternatively, the State-Space Concatenation (SSC) [5–9] approach allows for computing the eigenmodes in complex chains of cavities with asymmetric couplers using desktop computers. The scheme is based on decomposing the complex cavity chain into segments. The electromagnetic fields of the segments are described using state-space equations (coupled systems of ordinary differential equations) obtained from analytical calculations or from numerical discretization techniques such as the Finite-Integration Technique [10, 11]. Typically, the aforementioned state-space equations have many degrees of freedom to account for the distributed nature of the underlying partial differential equations. Fortunately, the number of degrees of freedom for each state-space model can be significantly reduced using model-order reduction (MOR) approaches [12–14]. Subsequently, all reduced-order models are concatenated by means of algebraic side constraints, which ensure that the tangential electric and magnetic fields are continuous across the surfaces of the decomposition planes. This concatenation delivers a very compact description of the complex structure in terms of its electromagnetic properties and allows for the determination of its eigenmodes by computing the eigenvalues and eigenvectors of comparably small matrices.

In this paper, the SSC scheme is revised using the BESSY VSR chain of superconducting cavities as an application example. The presented work has been conducted in the framework of a collaboration between the University of Rostock and the Helmholtz-Zentrum Berlin. Predominantly, this article is based on the internal report [15], which comprises all details of the computations. The field patterns and properties of the computed eigenmodes are listed in a compendium attached to [15]. All computations for the internal report have been conducted by J. Heller. In addition to the results provided by the internal report, this article presents a comparison between results from SSC and a direct computation using CST Studio Suite® (CST) [16]. Note that

\* The research leading to these results was supported by the German Bundesministerium für Bildung und Forschung, Land Berlin and grants of Helmholtz Association.

<sup>†</sup> thomas.flisgen@helmholtz-berlin.de

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

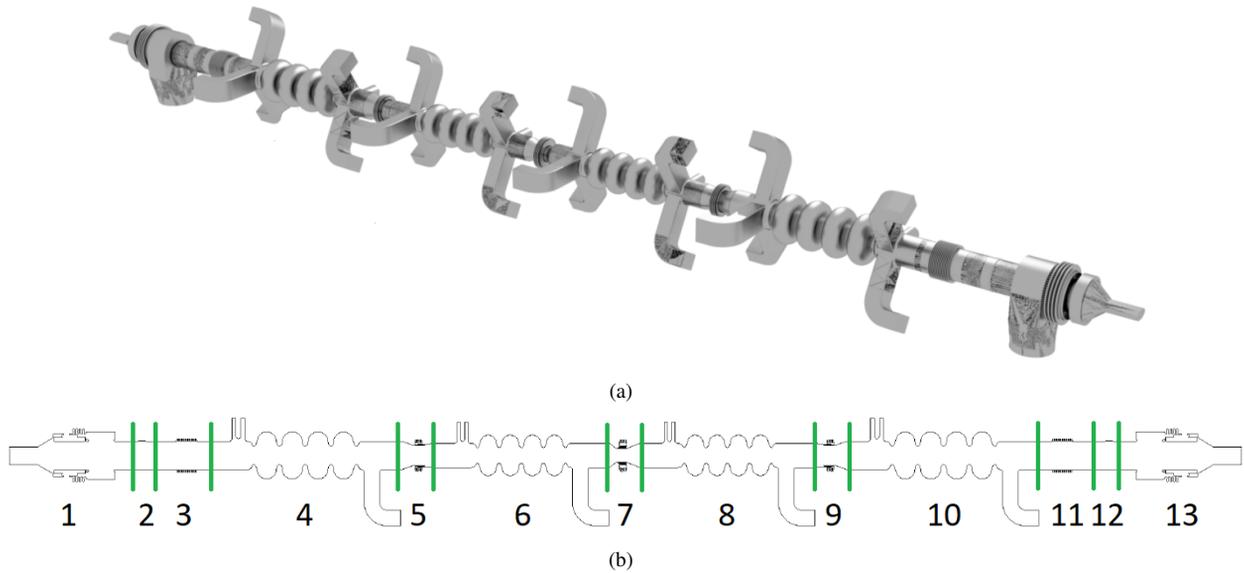


Figure 1: (a) Geometry of the BESSY VSR cavity string. Its main components are two cavities whose  $TM_{01}-\pi$ -modes resonate at 1.5 GHz and two cavities whose  $TM_{01}-\pi$ -modes resonate at 1.75 GHz. All cavities are constructed by means of ellipses. The 1.5 GHz cavities are located at both ends of the chain whereas the 1.75 GHz cavities are in between. Each cavity is equipped with one fundamental power coupler and five waveguides for the absorption of higher-order modes. At both ends of the complete chain endgroups with tapers, bellows, pump domes and dielectric absorbers are located. The figure is adapted from [15]. (b) Decomposition of the chain of cavities into individual segments. The green lines denote the decomposition planes. The arising segments are: 1. endgroup, 2. valve, 3. bellow, 4. cavity (1.5 GHz), 5. shielded bellow, 6. cavity (1.75 GHz), 7. collimating shielded bellow, 8. cavity (1.75 GHz), 9. shielded bellow, 10. cavity (1.5 GHz), 11. bellow, 12. valve, and 13. endgroup.

further comparisons between direct computations and SSC computations are available in [5, 6, 8, 9].

### EIGENMODES OF THE BESSY VSR CHAIN USING SSC

The third generation light source BESSY II is operated by the Helmholtz-Zentrum at the Wilhelm-Conrad-Röntgen-Campus in Berlin Adlershof in Germany. BESSY II is a user facility providing photon pulses ranging from the Terahertz to the hard X-ray regime. Its main component is a ring with a circumference of 240 m which can store currents up to 300 mA with the energy of 1.7 GeV. Currently, the upgrade of BESSY II to BESSY VSR (Variable pulse-length Storage Ring) is in preparation. BESSY VSR allows for simultaneously storing long and short pulses in the machine [17–19].

The upgrade requires the insertion of a cryomodule accommodating two superconducting 1.5 GHz and two superconducting 1.75 GHz four-cell resonators into the existing BESSY II ring. Figure 1 depicts the layout of the string of superconducting cavities. The first and the last cavities in the string are constructed such that their  $TM_{01}-\pi$ -modes resonate with 1.5 GHz, whereas the two cavities in the center of the chain are designed so that their  $TM_{01}-\pi$ -modes resonate with 1.75 GHz. The four cavities in the BESSY VSR cavity chain are connected by means of bellows. The

chain is equipped with pump domes, dielectric absorbers, bellows and tapers at both ends. Please refer to Figure 2 in [20] for a detailed view of the endgroups.

The superposition of the accelerating fields of all four cavities depicted in Figure 1 results in a beating pattern. In fact, the derivative with respect to the longitudinal direction of the voltages cancels for every second bunch whereas for the remaining bunches the derivative of the voltages constructively is added up. Together with the optics of the machine, the voltage beating pattern leads to long and short pulses simultaneously stored in the ring.

#### *Eigenmodes and their Properties*

The electric fields  $\mathbf{E}_n(\mathbf{r})$  of the eigenmodes in the cavity string fulfill Helmholtz equation

$$\Delta \mathbf{E}_n(\mathbf{r}) + \varepsilon \mu \omega_n^2 \mathbf{E}_n(\mathbf{r}) = \mathbf{0} \quad (1)$$

with the boundary conditions

$$\mathbf{n} \times \mathbf{E}_n(\mathbf{r}) = \mathbf{0} \text{ on } \partial\Omega_{\text{PEC}} \quad (2)$$

and

$$\mathbf{n} \cdot \mathbf{E}_n(\mathbf{r}) = 0 \text{ on } \partial\Omega_{\text{Port}}. \quad (3)$$

Here,  $\varepsilon$  and  $\mu$  denote the permittivity and the permeability of free space, respectively. The resonant angular frequency of the  $n$ th eigenmode is denoted by  $\omega_n$ . Perfect electric

conducting boundary conditions are assumed on the boundary  $\partial\Omega_{\text{PEC}}$  of the superconducting cavities, whereas perfect magnetic conducting boundary conditions are enforced on the boundary  $\Omega_{\text{port}}$  of the external waveguide ports. External waveguide ports are assigned at the beam pipe at both ends of the chain, at the ends of the fundamental power couplers and at the ends of the higher-order mode couplers.

In addition to the frequencies  $f_n = \omega_n/2\pi$  and the field distributions  $\mathbf{E}_n(\mathbf{r})$  of the eigenmodes, their coupling impedances

$$\left(\frac{r}{Q}\right)_n = \frac{1}{\omega_n W_n} \left| \int_{z_{\min}}^{z_{\max}} E_{n,z}(x_0, y_0, z) e^{j\omega_n z/c} dz \right|^2 \quad (4)$$

are of special interest. Here,  $W_n$  denotes the energy stored in the  $n$ th eigenmode,  $z_{\min}$  and  $z_{\max}$  the beginning and the end of the chain in longitudinal direction,  $E_{n,z}(x_0, y_0, z)$  the on-axis longitudinal component of the electric field of the  $n$ th eigenmode and  $c = 1/\sqrt{\epsilon\mu}$  the speed of light in vacuum.

Another important quantity related to resonant modes is the external quality factor defined by

$$Q_{\text{ext},n} = \frac{\omega_n W_n}{P_{\text{loss},n}}, \quad (5)$$

where  $P_{\text{loss},n}$  is the propagation of energy through the open waveguide ports of the structure. Note that the external quality factors  $Q_{\text{ext},n}$  do not directly result from Eqs. (1–3) because perfect magnetic boundary conditions are enforced on the port planes  $\Omega_{\text{port}}$ . Consequently, the Poynting vector normal to the port boundary equals zero. Nonetheless,  $Q_{\text{ext},n}$  can be approximated from the lossless eigenmodes using a later described perturbation approach.

It is worth to note that the coupling impedance  $(r/Q)_n$  specifies the coupling of the eigenmode with the bunch of charged particles, whereas the external quality factor  $Q_{\text{ext},n}$  quantifies the coupling of the mode to the waveguide ports.

### Decomposition of the Chain into Segments

In a first step, the BESSY VSR cavity chain is decomposed into individual segments. The decomposition planes are depicted in Figure 1(b) by green lines. In total, 13 segments are obtained. Following Figure 1(b) from the left to the right, the segments are endgroup, valve, bellow, cavity (1.5 GHz), shielded bellow, cavity (1.75 GHz), collimating shielded bellow, cavity (1.75 GHz), shielded bellow, cavity (1.5 GHz), bellow, valve, and endgroup. Note that the decomposition planes are chosen at regions of constant cross section to keep the number of 2D port modes required for the field expansion in the cut planes small. To each 2D port mode a modal voltage corresponding to the electric field distribution of the port mode and a modal current corresponding to the magnetic field distribution of the port mode are assigned. It is distinguished between internal ports located at the cutplanes arising from the decomposition and external ports located at the ends of the beam pipe or at ends of power and higher-order mode couplers. Depending on the cross section of the cutplanes 8, 16, or 20 2D port

modes are considered for the field expansion in these planes. Consequently, 8, 16, or 20 modal voltages and currents are required per cutplane per segment. Please refer to column five in Table 3.1 in [15] for more details.

### Electromagnetic Properties of the Segments

To describe the electromagnetic properties of the  $R = 13$  segments, each substructure is discretized by means of a hexahedral mesh using the commercial software CST [16]. Subsequently, the discrete wave equation with excitation is exported to MATLAB [21], so that the second-order state-space systems

$$\frac{d^2}{dt^2} \mathbf{x}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{B}_r \frac{d}{dt} \mathbf{i}_r(t) \quad (6)$$

are available for each segment. Here,  $1 \leq r \leq R$  is the segment index,  $\mathbf{x}_r(t) \in \mathbb{R}^{N_d}$  is a time-dependent state vector,  $\mathbf{A}_r \in \mathbb{R}^{N_d \times N_d}$  the system matrix,  $\mathbf{B}_r \in \mathbb{R}^{N_d \times N_i}$  the input matrix, and  $\mathbf{i}_r(t) \in \mathbb{R}^{N_i}$  the time-dependent excitation term. The excitation term comprises modal port currents of the waveguide ports of the  $r$ th segment. In close analogy, the vector

$$\mathbf{v}_r(t) = \mathbf{B}_r^T \mathbf{x}_r(t) \quad (7)$$

lists the time-dependent modal voltages of the  $r$ th segment. The total number of 2D port modes of each segment is given by  $N_i$ .

Following the Finite-Integration Technique (FIT) [10, 11], the state matrix can be chosen to be

$$\mathbf{A}_r = -\mathbf{M}_{\epsilon,r}^{-1/2} \mathbf{C}_r^T \mathbf{M}_{\mu,r}^{-1} \mathbf{C}_r \mathbf{M}_{\epsilon,r}^{-1/2}, \quad (8)$$

where the discrete representation of the curl operator is denoted by  $\mathbf{C}_r$ . The matrices  $\mathbf{M}_{\epsilon,r}$  and  $\mathbf{M}_{\mu,r}$  are diagonal and comprise properties of the grid and averaged material parameters. The input matrix (or output matrix transposed) is given by

$$\mathbf{B}_r = \mathbf{M}_{\epsilon,r}^{-1/2} \mathbf{R}_r. \quad (9)$$

The columns of the matrix  $\mathbf{R}_r$  comprise the electric field distributions of the 2D port modes in a lexicographic order. Note that the sampled 3D electric field distribution in the  $r$ th segment is given in a lexicographic order by

$$\mathbf{e}_r(t) = \mathbf{D}_{s,r}^{-1} \mathbf{M}_{\epsilon,r}^{-1/2} \mathbf{x}_r(t), \quad (10)$$

where  $\mathbf{D}_{s,r}$  is a diagonal matrix holding the lengths of the edges of the primary grid.

### Model-Order Reduction

The number of degrees of freedom of each state-space model can be significantly reduced by means of model-order reduction approaches. The reduction is conducted by expressing a reduced state vector  $\mathbf{x}_{\text{rd},r}(t) \in \mathbb{R}^{N_{\text{dr}}}$  using a semi-orthogonal reduction matrix  $\mathbf{W}_r \in \mathbb{R}^{N_d \times N_{\text{dr}}}$ :

$$\mathbf{x}_r(t) = \mathbf{W}_r \mathbf{x}_{\text{rd},r}(t). \quad (11)$$

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

As a matter of fact, there are various ways to construct the reduction matrix. In the framework of this study, a proper orthogonal decomposition is used which requires a finite set of frequency-domain field distributions of the state-space system Eq. (6). It is a key property that the reduction matrix has much more rows than columns, i.e.  $N_d \gg N_{dr}$ . Replacing Eq. (11) in Eq. (6) and multiplying the obtained equation with  $\mathbf{W}_r^T$  from the left hand side gives

$$\frac{d^2}{dt^2} \mathbf{x}_{rd,r}(t) = \underbrace{\mathbf{W}_r^T \mathbf{A}_r \mathbf{W}_r}_{\mathbf{A}_{rd,r}} \mathbf{x}_{rd,r}(t) + \underbrace{\mathbf{W}_r^T \mathbf{B}_r}_{\mathbf{B}_{rd,r}} \frac{d}{dt} \mathbf{i}_r(t) \quad (12)$$

on account of the semi-orthogonality of the reduction matrix ( $\mathbf{W}_r^T \mathbf{W}_r = \mathbf{I}$ ).

Replacing Eq. (11) in Eq. (7) delivers the reduced-order output equation

$$\mathbf{v}_r(t) = \underbrace{\mathbf{B}_r^T \mathbf{W}_r}_{\mathbf{B}_{rd,r}^T} \mathbf{x}_{rd,r}(t). \quad (13)$$

For instance, the model-order reduction allows for reducing the number of degrees of freedom from  $N_d \approx 5.5 \times 10^6$  to  $N_{dr} \approx 9.2 \times 10^2$  for the 1.5 GHz cavities and from  $N_d \approx 5.5 \times 10^4$  to  $N_{dr} \approx 2.5 \times 10^2$  for the valves. Please refer to column three of Table 3.1 in [15] for the number of degrees of freedom for the remaining segments.

The computation of the reduction matrices was performed on an Intel(R) Xeon(R) CPU E5-2687W @ 3.4 GHz with 256 GB of RAM using Windows Server 2012. The total computation time to determine all reduction matrices based on an accuracy criterion was approximately 6 d and 2 h. Column nine of Table 3.1 in [15] presents the computing times for the construction of the reduction matrix for each of the  $R = 13$  segments.

### Concatenation of State-Space Models

To concatenate the reduced-order state-space models of the individual segments to a reduced-order state-space model of the full structure, all state-space models are collated in terms of a block system. For instance, this reads

$$\mathbf{x}_b(t) = \left( \mathbf{x}_{rd,1}^T(t) \mathbf{x}_{rd,2}^T(t) \dots \mathbf{x}_{rd,r}^T(t) \dots \mathbf{x}_{rd,R}^T(t) \right)^T \quad (14)$$

for the state vector of the block system. In a next step, this block system is modified to ensure that Kirchhoff's laws are fulfilled for modal voltages and modal currents of ports to be coupled. Following [6, Appendix C.2], Kirchhoff's laws result from continuity constraints of tangential electric and magnetic fields. The incorporation of Kirchhoff's laws into the block state-space system delivers a state-space system of the full structure. This system is again reduced to obtain

$$\frac{d^2}{dt^2} \mathbf{x}_{cr}(t) = \mathbf{A}_{cr} \mathbf{x}_{cr}(t) + \mathbf{B}_{cr} \frac{d}{dt} \mathbf{i}_{ext}(t) \quad (15)$$

with the output equation

$$\mathbf{v}_{ext}(t) = \mathbf{B}_{cr}^T \mathbf{x}_{cr}(t). \quad (16)$$

This system is a very compact description of the electromagnetic properties of the BESSY VSR cavity chain. Note that the relationship between the reduced state-vector of the concatenated system and the state-vector of the block system is given by

$$\mathbf{x}_b(t) = \mathbf{W}_c \mathbf{x}_{cr}(t), \quad (17)$$

with the semi-orthogonal reduction matrix  $\mathbf{W}_c \in \mathbb{R}^{6,921 \times 3,573}$  of the concatenated system.

### Determination of Eigenmodes

As perfect magnetic conducting boundary conditions are assumed on the surfaces of the external waveguide ports

$$\mathbf{i}_{ext}(t) = \mathbf{0} \quad (18)$$

is enforced for the eigenmode computations. Note that modal currents correspond to the tangential magnetic fields of the respective 2D port modes on the port surfaces. Using the constraint Eq. (18) and transforming the reduced-order state-equation Eq. (15) of the complete structure into frequency domain leads to the eigenvalue problem

$$\underbrace{\mathbf{A}_{cr}}_{\mathbf{v}_{cr,n}} \underbrace{\mathbf{x}_{cr}}_{\lambda_n} = \underbrace{-\omega^2}_{\mathbf{v}_{cr,n}} \underbrace{\mathbf{x}_{cr}}_{\mathbf{v}_{cr,n}}. \quad (19)$$

The eigenvectors  $\mathbf{v}_{cr,n}$  as well as the eigenvalues are real-valued because of the symmetry of the matrix  $\mathbf{A}_{cr}$ . A comparison of constants in Eq. (19) gives the following relation between resonant frequencies of the eigenmodes of the entire chain and the eigenvalues of the system matrix:

$$f_n = \frac{\omega_n}{2\pi} = \frac{1}{2\pi} \sqrt{-\lambda_n} \in \mathbb{R}. \quad (20)$$

The frequencies are real-valued, because  $\lambda_n$  is smaller than or equal to zero on account of the negative semi-definiteness of  $\mathbf{A}_{cr}$ .

The field distributions of the  $n$ th eigenmode are determined based on the eigenvectors  $\mathbf{v}_{cr,n}$ . In a first step, the state-vector of the block system used for the concatenation is computed:

$$\mathbf{v}_{b,n} = \mathbf{W}_c \mathbf{v}_{cr,n}. \quad (21)$$

Then, the vector  $\mathbf{v}_{b,n}$  is partitioned following its definition in Eq. (14). Subsequently, the reduced-order state vectors  $\mathbf{v}_{rd,r,n}$  of each segment are employed to reconstruct the electric field distribution of the  $n$ th mode in the  $r$ th segment:

$$\mathbf{e}_{r,n} = \mathbf{D}_{s,r}^{-1} \mathbf{M}_{\varepsilon,r}^{-1/2} \mathbf{W}_r \mathbf{v}_{rd,r,n}. \quad (22)$$

The sampled electric field distributions in  $\mathbf{e}_{r,n}$  are stored in a lexicographic order and are transferred to a 3D field so that they can be exported to ParaView [22]. A processing script is used to automatically create field plots of all modes with resonant frequencies in the interval from 500 MHz to 3.6 GHz. The field plots show the absolute value of the electric fields in two different cutplanes. Based on the electric fields and the resonant frequencies, the coupling impedances  $(r/Q)_n$  are also determined.

## Determination of Quality Factors

The external quality factors of lossy modes in the BESSY VSR cavity chain are determined using a perturbation approach. In a first step, the state-space model Eqs. (15) and (16) of the lossless closed cavity chain is transferred to a model with first-order derivatives with twice as many states as the model with second-order derivatives:

$$\frac{d}{dt} \tilde{\mathbf{x}}_{\text{cr}}(t) = \tilde{\mathbf{A}}_{\text{cr}} \tilde{\mathbf{x}}_{\text{cr}}(t) + \tilde{\mathbf{B}}_{\text{cr}} \mathbf{i}_{\text{ext}}(t) \quad (23)$$

with the corresponding output equation

$$\mathbf{v}_{\text{ext}}(t) = \tilde{\mathbf{B}}_{\text{cr}}^T \tilde{\mathbf{x}}_{\text{cr}}(t). \quad (24)$$

The matrices with the tilde directly result from the quantities without the tilde. To account for open boundary conditions, the external modal excitation currents are chosen based on the modal voltages in frequency domain:

$$\mathbf{i}_{\text{ext}}(j\omega) = -\underline{\mathbf{D}}_z^{-1}(j\omega) \mathbf{v}_{\text{ext}}(j\omega), \quad (25)$$

whereas  $\underline{\mathbf{D}}_z(j\omega)$  is the diagonal matrix holding the frequency-dependent wave impedances of the 2D ports modes at the external waveguide ports. Combining Eq. (25) with the state-space system Eq. (23) and Eq. (24) and transferring the resulting statement into frequency domain gives the nonlinear eigenvalue problem

$$\left[ \tilde{\mathbf{A}}_{\text{cr}} - \tilde{\mathbf{B}}_{\text{cr}}^T \underbrace{\underline{\mathbf{D}}_z^{-1}(j\omega_n)}_{\lambda_n} \tilde{\mathbf{B}}_{\text{cr}} \right] \tilde{\mathbf{x}}_{\text{cr},n} = \underbrace{j\omega_n}_{\lambda_n} \tilde{\mathbf{x}}_{\text{cr},n}. \quad (26)$$

The reader is referred to [23, 24] for a complete description of the nonlinear eigenvalue problem arising from external quality factor computations. The frequencies of the lossy modes and their external quality factors are determined by

$$f_{\text{ext},n} = \frac{\Im\{\lambda_n\}}{2\pi}, \quad Q_{\text{ext},n} = -\frac{\Im\{\lambda_n\}}{2\Re\{\lambda_n\}}. \quad (27)$$

Note that the external quality factors are not directly linked to the lossless eigenmodes as the introduction of losses leads to a coupling of all eigenmodes. Consequently, the resonant frequencies and field distributions for lossy modes are different from lossless modes. Therefore, it is in general difficult to directly connect external quality factors to eigenmodes obeying Eq. (1) with the boundary conditions Eqs. (2) and (3).

## NUMERICAL RESULTS

The central result of the described computations is a modal compendium listing the resonant frequencies, the coupling impedances, and the field distributions of 1,576 eigenmodes which were found in the interval 500 MHz to 3.6 GHz. The modal compendium is part of [15] and comprises a large variety of modes with complex field patterns such as cavity modes, multi-cavity modes, bellow modes or combinations of these. Note that the complete and rigorous

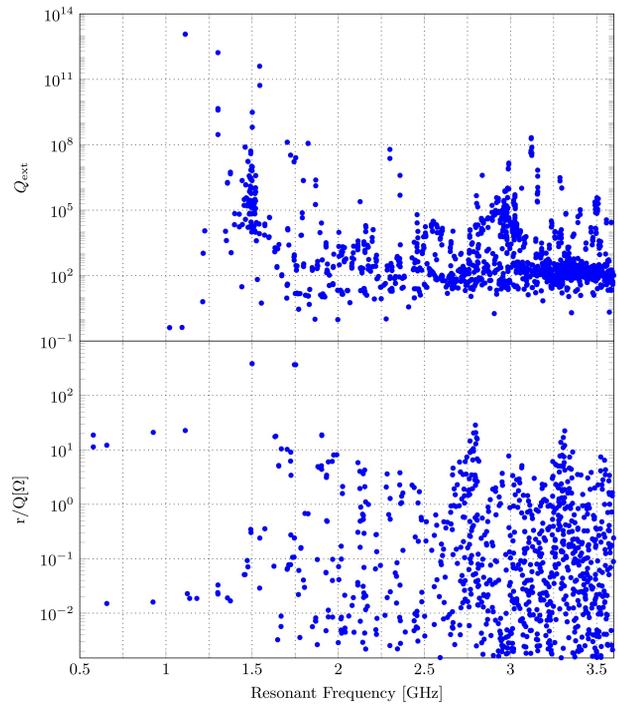


Figure 2: Semilog plots of the external quality factors (top) and the coupling impedances (bottom) of modes in the BESSY VSR cavity chain.

discussion of the results is not feasible on account of the large amount of generated data.

The diagram in the upper part of Figure 2 shows the external quality factors of modes in the BESSY VSR cavity chain. It is worth to mention that the modes with excessively large external quality factors ( $10^{11} \dots 10^{14}$ ) in the interval 578 MHz to 1.54 GHz are modes either being localized in the collimators or close to the dielectric absorbers. Both parts are assumed to be lossless in the computation, so that the quality factors of confined modes in these segments are estimated to be large. In practice however, these segments are not lossless so their quality factors are significantly reduced as factors of these modes are governed by dielectric and surface losses and not by external losses.

The diagram in the lower part of Figure 2 depicts the coupling impedances. The  $\text{TM}_{01-\pi}$  modes of the 1.5 GHz resonators and the 1.75 GHz resonators with their  $(r/Q)_n$  in the order of  $10^2 \Omega$  are readily identifiable. On purpose, these modes have the largest coupling impedances. In addition to these modes, various modes with comparably large coupling impedances in the order of  $1 \Omega$  to  $10 \Omega$  exist. These modes are of potential danger for the operation of BESSY VSR, in particular, if their resonant frequencies are close to the  $\Delta f = 250$  MHz harmonics of the periodic BESSY VSR current. Please refer to Figure 3 in [20] for the spectrum of the BESSY VSR current.

For a review of the set of modes relevant for the operation of BESSY VSR, the reader is referred to [15]. The follow-

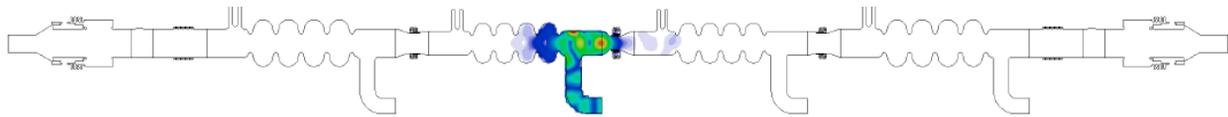


Figure 3: Absolute value of the electric field distribution of mode  $n = 888$ . The colorbar is presented in Figure 5(a). The field distribution is predominantly located in the waveguide connecting the cavity with the collimating shielded bellow and the higher-order mode waveguide absorber. The frequency of the mode is  $f_{888}^{SSC} = 2.9899$  GHz and its coupling impedance is  $(r/Q)_{888}^{SSC} \approx 7.73 \Omega$ .

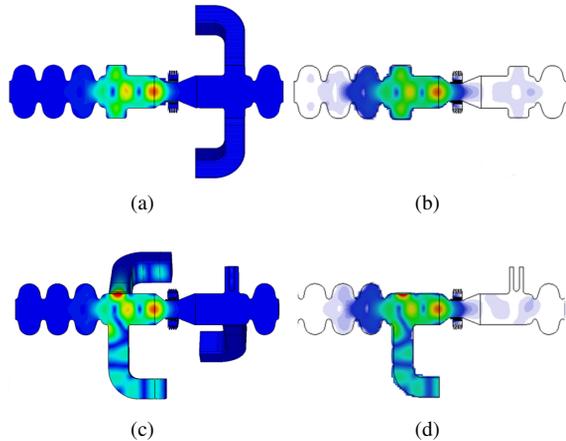


Figure 4: Electric field distribution of mode  $n = 888$  on orthogonal cutplanes: (a) and (c) depict the result directly delivered by CST whereas (b) and (d) show the result delivered by SSC. The respective colorbars are presented in Figure 5. Note that the waveguides for the absorption of higher-order modes are elongated by  $\lambda/4$  in the direct run to emulate perfect magnetic conducting boundary conditions with perfect electric boundary conditions.

ing subsection is focused on mode  $n = 888$  to conduct a plausibility analysis of SSC.

### Comparison of SSC with Direct Computation

Figure 3 depicts the absolute value of the electric field distribution of mode  $n = 888$  in the BESSY VSR cavity chain. The mode has a comparably large coupling impedance  $(r/Q)_{888}^{SSC} \approx 7.7298 \Omega$  and its frequency  $f_{888}^{SSC} = 2.9899$  GHz is close to the  $\Delta f = 250$  MHz harmonics of the beam current. Thus, this mode is of interest although it is expected to have a small external quality factor because of its strong coupling to the higher-order mode absorbers.

Figure 3 shows that the field energy of the mode is confined across two segments, namely at one 1.75 GHz resonator and at the collimating shielded bellow. Therefore, this eigenmode is suitable to compare the SSC result with a direct eigenmode computation using CST [16]. The direct eigenmode computation is performed for a geometry comprising three cells of the 1.75 GHz resonator with higher-order mode couplers, the collimating shielded bellow and one cell of the next 1.75 GHz resonator with fundamental power coupler and input coupler. Please refer Figure 4(a) and Figure 4(c) for cutplanes of the structure considered by the direct eigen-

mode computation. Note that the higher-order waveguide absorbers are slightly elongated by  $\lambda/4 = c/4/\sqrt{f_{888}^2 - f_{co}^2}$ , because CST does not allow to specify perfect magnetic boundary conditions within the computational domain. Here,  $\lambda$  is the wavelength in the higher-order mode waveguide coupler and  $f_{co}$  the cutoff frequency of the fundamental waveguide mode. A tetrahedral mesh with 218,444 tetrahedrons is used to generate a discrete representation of Eq. (1) for the substructure depicted in 4(a) and Figure 4(c). The tetrahedral mesh is chosen because it is more suitable to discretize the smooth curvatures of the geometry. Moreover, in general, the properties of the eigenmodes should neither depend on the discretization (tetrahedral or hexahedral) nor on the numerical approach (direct or SSC). The eigenmode solver of CST [16] is requested to search for 10 modes with resonant frequencies larger than 2.98 GHz. The computational time required to determine these 10 eigenmodes is approximately 13 min on an Intel(R) Xeon(R) CPU E5-2687W @ 3.4 GHz with 256 GB of RAM using Windows Server 2012.

The resonant frequency arising from the direct computation is  $f_{888}^{CST} = 2.9863$  GHz, so that the relative difference in the frequency between direct computation and SSC amounts to less than  $1.2 \times 10^{-3}$ . The coupling impedance delivered by the direct computation amounts to  $(r/Q)_{888}^{CST} = 8.8095 \Omega$ , so that a relative difference of  $\approx 1.2 \times 10^{-1}$  results.

## SUMMARY AND CONCLUSION

The SSC scheme allows for the determination of eigenmodes of long chains of cavities using workstations. The key of the method is the combination of non-overlapping domain decomposition with model-order reduction. The scheme allows for the creation of eigenmode compendia, which systematically comprise field distributions, resonant frequencies, coupling impedances, and quality factors.

In addition to the various comparisons of SSC results with direct computations for test structures (refer for instance to [5, 6, 8, 9]), this contribution presents a comparison using another structure. The confinement of the studied mode in two segments of the chain allows for a comparison with a direct computation using CST.

As a central result, the resonant frequency agrees very well for this mode and the coupling impedance agrees reasonably well. The differences are attributed to the different discretization techniques (tetrahedral vs. hexahedral mesh) and the different numerical approaches (direct vs. SSC).

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## ACKNOWLEDGEMENTS

The authors would like to thank Hans-Walter Glock, Felix Glöckner, and Andranik Tsakanian from Helmholtz-Zentrum Berlin for various fruitful discussions related to the BESSY VSR cavity chain.

## APPENDIX

Figure 5(a) presents the colorbar for the absolute values of the electric field strength delivered by SSC whereas Figure 5(b) depicts the corresponding colorbar for the direct computation using CST.

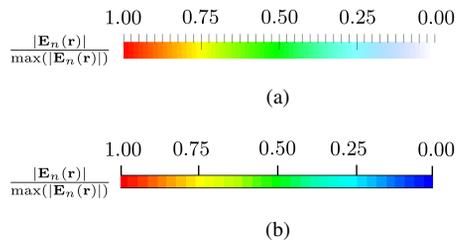


Figure 5: (a) Colorbar for SSC field plots. (b) Colorbar for CST field plots.

## REFERENCES

- [1] R. V. Beeumen *et al.*, “Computing resonant modes of accelerator cavities by solving nonlinear eigenvalue problems via rational approximation,” *Journal of Computational Physics*, vol. 374, pp. 1031–1043, 2018, ISSN: 0021-9991.
- [2] I. R. R. Shinton, R. M. Jones, Z. Li, and P. Zhang, “Simulations of Higher Order Modes in the ACC39 Module of FLASH,” in *Proceedings of International Particle Accelerator Conference 2012*, New Orleans, Louisiana, USA, May 2012, pp. 1900–1902.
- [3] K. Ko *et al.*, “Advances in Parallel Electromagnetic Codes for Accelerator Science and Development,” in *Proceedings of the Linear Accelerator Conference 2010*, Tsukuba, Japan, 2010, pp. 1028–1032.
- [4] F. Yaman, W. Ackermann, and T. Weiland, “Comparison of Eigenvalue Solvers for Large Sparse Matrix Pencils,” in *Proceedings of the 11th International Computational Accelerator Physics Conference 2012*, Rostock-Warnemünde, Germany, 2012, pp. 287–289.
- [5] T. Flisgen, H.-W. Glock, and U. van Rienen, “Compact Time-Domain Models of Complex RF Structures based on the real Eigenmodes of Segments,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 6, pp. 2282–2294, 2013. doi: 10.1109/TMTT.2013.2260765.
- [6] T. Flisgen, “Compact state-space models for complex superconducting radio-frequency structures based on model order reduction and concatenation methods,” available as CERN-ACC-2015-0145, PhD thesis, University of Rostock, 2015.
- [7] T. Flisgen *et al.*, “Eigenmode Compendium of the Third Harmonic Module of the European XFEL Computed by SSC,” *Physical Review Accelerators and Beams*, vol. 20, no. 4, Apr. 2017. doi: 10.1103/PhysRevAccelBeams.20.042002.
- [8] T. Flisgen, J. Heller, and U. van Rienen, “Computation of Eigenmodes in Long and Complex Accelerating Structures by Means of Concatenation Strategies,” in *Proceedings of the 5th International Particle Accelerator Conference 2014 (IPAC 2014)*, Dresden, Germany, 2014, pp. 947–949.
- [9] J. Heller, T. Flisgen, and U. van Rienen, “Computational Benefits Using an Advanced Concatenation Scheme Based on Reduced Order Models for RF Structures,” *Physics Procedia*, vol. 79, pp. 38–45, 2015. doi: 10.1016/j.phpro.2015.11.060.
- [10] 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.
- [11] U. Rienen, *Numerical Methods in Computational Electrodynamics: Linear Systems in Practical Applications*. Berlin, Germany: Springer, 2001, ISBN: 3540676295.
- [12] T. Wittig, *Zur Reduzierung der Modellordnung in elektromagnetischen Feldsimulationen*. Cuvillier Verlag, 2004, ISBN: 9783865372086.
- [13] T. Wittig, R. Schuhmann, and T. Weiland, “Model order reduction for large systems in computational electromagnetics,” *Linear Algebra and its Appl.*, vol. 415, no. 2–3, pp. 499–530, Jun. 2006.
- [14] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005, ISBN: 9780898718713.
- [15] J. Heller and U. van Rienen, “Report on SSC calculations for BESSY VSR,” University of Rostock, Internal Report, 2017.
- [16] CST AG, *CST STUDIO SUITE 2018*. Bad Nauheimer Str. 19, 64289 Darmstadt, Germany.
- [17] G. Wüstefeld, A. Jankowiak, J. Knobloch, and M. Ries, “Simultaneous Long and Short Electron Bunches in the BESSY II Storage Ring,” in *2nd International Particle Accelerator Conference (IPAC’11)*, San Sebastián, Spain, 2011.
- [18] A. Jankowiak *et al.*, “Technical Design Study BESSY VSR,” Helmholtz-Zentrum Berlin, Internal Report, 2015.
- [19] A. Vélez *et al.*, “The SRF Module Developments for BESSY VSR,” in *8th International Particle Accelerator Conference (IPAC’17)*, Copenhagen, Denmark, 2017.
- [20] T. Flisgen, H.-W. Glock, and A. Tsakanian, “Estimation of Dielectric Losses in the BESSY VSR Warm Beam Pipe Absorbers,” in *9th International Particle Accelerator Conference (IPAC’18)*, Vancouver, BC, Canada, 2018.
- [21] The MathWorks Inc., *Matlab version r2015b*.
- [22] A. Henderson and J. Ahrens, *The Paraview guide: a parallel visualization application*. New York: Kitware, Inc., 2004, ISBN: 1930934149.
- [23] J. Heller, “Numerical simulation of electromagnetic fields in complex multi-cavity superconducting radio frequency resonators,” PhD thesis, Universität Rostock, 2018.
- [24] H. W. Pommerenke, J. Heller, and U. Rienen, “Efficient Computation of Lossy Higher Order Modes in Complex SRF Cavities Using Reduced Order Models and Nonlinear Eigenvalue Problem Algorithms,” in *Proceedings of the 13th International Computational Accelerator Physics Conference 2018*, Key West, Florida, USA, 2018, this conference.

# FIRST STEPS TOWARDS A NEW FINITE ELEMENT SOLVER FOR MOEVE PIC TRACKING\*

Ursula van Rienen<sup>1†</sup>, Dawei Zheng, Johann Heller, Christian Bahls

Institute of General Electrical Engineering, University of Rostock, 18059 Rostock, Germany

<sup>1</sup>also at Department Life, Light & Matter, University of Rostock, 18051 Rostock, Germany

## Abstract

A relevant task in designing high-brilliance light sources based on high-current linear accelerators (e.g. Energy Recovery Linacs (ERLs)) consists in systematic investigations of ion dynamics in the vacuum chamber of such machines. This is of high importance since the parasitic ions generated by the electron beam turned out to be a current-limiting factor for many synchrotron radiation sources. In particular, the planned high current operation at ERL facilities requires a precise analysis and an accurate development of appropriate measures for the suppression of ion-induced beam instabilities. The longitudinal transport of ions through the whole accelerator plays a key role for the establishment of the ion concentration in the machine. Using the Particle-in-Cell (PIC) method, we started redesigning our code MOEVE PIC Tracking in order to allow for the fast estimation of the effects of ions on the beam dynamics. For that, we exchanged the previously used Finite Difference (FD) method for the solution of Poisson's equation within the PIC solver by a solver based on the Finite Element Method (FEM). Employing higher order FEM, we expect to gain improved convergence rates and thus lower computational times. We chose the Open Source Framework FEniCS for our new implementation.

## INTRODUCTION

MOEVE was developed as a Particle-in-Cell (PIC) solver at the University Rostock. It is an abbreviation and stands for Multigrid for non-equidistant grids to solve Poisson's equation. The software was originally developed in C by G. Pöplau et al. [1] and employs the Finite-Difference Technique (FD) to numerically discretize Poisson's equation. The discretized system of linear equations is solved iteratively by a geometric multigrid method.

MOEVE has been used successfully to simulate the interaction of electron beams with ionized residual gas [2, 3], several investigations for the clearing of ions with clearing electrodes and/or clearing gaps [4] and the simulation of transverse wake functions [5]. MOEVE has also been implemented in the tracking code GPT [6] and ASTRA [7].

## Ion Clearing

Any residual gas in the vacuum chamber of an accelerator can be ionized rapidly by the electron beam. The resulting

ion distribution is denoted as ion cloud. For many synchrotron radiation sources, these parasitic ions generated by the electron beam are a current-limiting factor. They often lead to beam instabilities, beam loss and they prevent a continuous filling of electron bunches into the ring shape machine.

In the existing synchrotron accelerators, mainly two strategies are used to ensure a minimum stability in standard operational regimes: clearing gaps and special electrodes for removing and neutralizing the ions. In certain high-current operating conditions ion effects are important, as they lead to beam instabilities. In particular, the planned high-current operation at ERL facilities requires a precise analysis and an accurate development of appropriate measures for the suppression of ion-induced beam instabilities [8].

The longitudinal transport of ions through the whole accelerator plays a key role for the establishment of the ion concentration in the machine. This aspect of the dynamics has implications on both the beam dynamics and the ion clearing efficiency but it has not been deeply studied up to now. In particular, the extent to which the accelerating resonators contribute to the transport is largely unclear. Thus, we are targeting a fast, systematic investigation of ion dynamics in the vacuum chamber of the machines involving the impact on the beam and its application to reduce the effects related to ionized residual gas in high-current electron machines. This study follows our previous investigations on ion trapping in high-current storage rings and linear accelerators [2–4, 9, 10].

## Prior Limitations of MOEVE Due to Finite Differences

Any PIC software consists of five different main modules. These are

1. Charge weighting
2. Discretization of Poisson's equation
3. Solver for Poisson's equation
4. Field interpolation at particles position
5. Update scheme of the particle distribution

MOEVE's current limitations, caused by the underlying FD discretization, affect the discretization and solution of Poisson's equation. A comparably large number of degrees of freedom (DOFs) are required for the accurate solution, especially because the tensor product grid in the FD method

\* Work supported by the German Federal Ministry for Research and Education BMBF under contract 015K16HRA.

† ursula.van-rienen@uni-rostock.de

does not approximate the geometry well. The current FD implementation in MOEVE does not allow for the discretization of arbitrary domains without strong reductions in accuracy [11], as it would be required for the tasks at hand.

The PIC method scales in its complexity with the number of macro-particles and the required mesh cells of the discretization. While the number of macro-particles can not be further reduced (since this would lead to an insufficient accuracy) one can reduce the number of mesh cells by using a different discretization technique (e.g. the Finite Element Method (FEM)).

Using appropriate ansatz functions in FEM, e.g. *Crouzeix-Raviart* elements as in [12], one could even improve this convergence by at least one order<sup>1</sup> compared to the FD discretization, thus allowing a reduction in the number of mesh cells.

## REPLACING THE FD SOLVER WITH FEM FROM FENICS

When using a PIC method, we compute the accelerating field of a charge density  $\rho(x)$  from the solution of Poisson's equation on the domain  $\Omega$ :

$$-\Delta u(x) = \frac{\rho(x)}{\varepsilon} \quad \forall x \in \Omega. \quad (1)$$

Usually, in our applications  $\varepsilon$  will be the vacuum permittivity  $\varepsilon_0$ .

For a unique solution  $u(x)$  we have to impose boundary conditions on  $\partial\Omega$ :

$$u(x) = g_D(x) \quad \forall x \in \partial\Omega_D, \quad (2)$$

$$\frac{\partial u(x)}{\partial n(x)} = g_N(x) \quad \forall x \in \partial\Omega_N. \quad (3)$$

The accelerating field  $\vec{E}$  can then be computed as the negative gradient  $\nabla u(x)$  of the solution  $u(x)$ .

For ease and speed-up of development and to attain a certain flexibility in the selection of function-spaces for ansatz and test functions, we use FEniCS [13] as implemented in the C++/Python library dolfin [14] for the automated solution of the system of equations arising from the FEM formulation of Eqs. (1), (2), and (3).

FEniCS allows to directly write down the weak formulation of Poisson's equation:

$$\int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx = \int_{\Omega} \frac{\rho(x)}{\varepsilon_0} v(x) dx \quad \forall v \in V \quad (4)$$

as a pair of a bilinear form  $a(u, v)$ :

$$a(u, v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx \quad (5)$$

and a linear form  $L(v)$ :

$$L(v) = \int_{\Omega} \frac{\rho(x)}{\varepsilon_0} v(x) dx. \quad (6)$$

<sup>1</sup> This would lead to a quadratic (or better) convergence in the force with FEM, compared to a linear convergence using FD.

To be able to do this one has to import the Python module dolfin:

```
from dolfin import *
```

and to specify the discrete function spaces (depending on the mesh used):

```
V = FunctionSpace(mesh, "CG", degree)
u = TrialFunction(V)
v = TestFunction(V)
```

One can directly write down the bilinear form as:

```
a = dot(grad(u), grad(v))*dx(mesh)
```

We can now prepare and assemble the system matrix A for the solver included in FEniCS:

```
template = PETScMatrix()
A = assemble(a, tensor=template)
```

The linear form  $L$  in the weak formulation of Eq. (1) as given in Eq. (4) depends on the charge density arising from the charges in the domain.

Starting from a constant  $\rho = 0$  one can assemble the right hand side linear form  $L$  and the corresponding vector rhs:

```
L = Constant(0.0)*v*dx(mesh)
rhs = assemble(L)
```

using the charge weighting implemented by the method `PointSource` from dolfin to add macro-particles to the right hand side:

```
macro_particles = []
for i in range(Number_of_Particles):
    macro_particles.append((Particle[i],
                           charge[i]/eps0))
delta = PointSource(V, macro_particles)
delta.apply(rhs)
```

After defining and applying the boundary condition  $g_D$  on  $\partial\Omega_D$ <sup>2</sup>

```
bc = DirichletBC(V, g_D, "on_boundary")
bc.apply(A)
bc.apply(rhs)
```

and setting up one of the solvers provided through dolfin (here the conjugate gradient method):

```
solver = PETScKrylovSolver("cg", "default")
solver.parameters["relative_tolerance"] =
    residual
solver.set_operator(A)
```

one can solve for the unknown potential  $u(x)$ :

<sup>2</sup> For ease of exposition, we choose to only show the implementation using a Dirichlet boundary condition  $g_D$  on  $\partial\Omega$ .

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

```
u_x = Function(V)
solver.solve(u_x.vector(), rhs)
```

The electric field  $\vec{E}$  can then be computed from the gradient  $\nabla u(x)$  of the solution.

```
e_temp = -grad(u_x)
```

To be applicable as an interpolated field it has to be projected onto an appropriate function space. Possibilities include

- the matching Raviart-Thomas finite element space

```
if V_field == "RT":
    Efield = project(e_temp,
                    FunctionSpace(mesh, \
                    "RT", degree))
```

- the Brezzi-Douglas-Marini finite element space

```
elif V_field == "BDM":
    Efield = project(e_temp,
                    FunctionSpace(mesh, \
                    "BDM", degree-1))
```

- a corresponding Discontinuous-Galerkin vector function space

```
elif V_field == "DG":
    Efield = project(e_temp,
                    VectorFunctionSpace(mesh, \
                    "DG", degree-1))
```

- and the continuous Lagrange(Courant) vector function space

```
else:
    Efield = project(e_temp,
                    VectorFunctionSpace(mesh, \
                    "CG", degree-1))
```

The computed field can next be used to accelerate the particles using the well-known Boris pusher [15].

## FIRST RESULTS

In this section, we show first results for a simple model problem as obtained from our FEniCS implementation for MOEVE and compare with ASTRA [7]. The model problem regards tracking of an electron bunch of Gaussian distribution in all directions for a short drift space without external electromagnetic field and without ion cloud.

We have tracked an electron bunch with the new FEM solver for a drift distance of 3.0 m without any external electromagnetic field or ion cloud. The initial bunch is generated by ASTRA. The bunch profile is listed in Table 1. The rms bunch size and the emittance growth are plotted in Figures 1

and 2, respectively. The results are compared with ASTRA for the transverse directions. The emittance was computed according [16].

Table 1: Bunch Profile for Tracking

Parameters of the electron bunch	
Number of macro particles	5,000
Beam energy	15 MeV
Beam energy spread	1.49 keV
Beam charge	-0.4 nC
Transverse emittance	$1.0 \pi$ mm·mrad
Bunch length	0.88 mm
rms bunch radius	0.362 mm

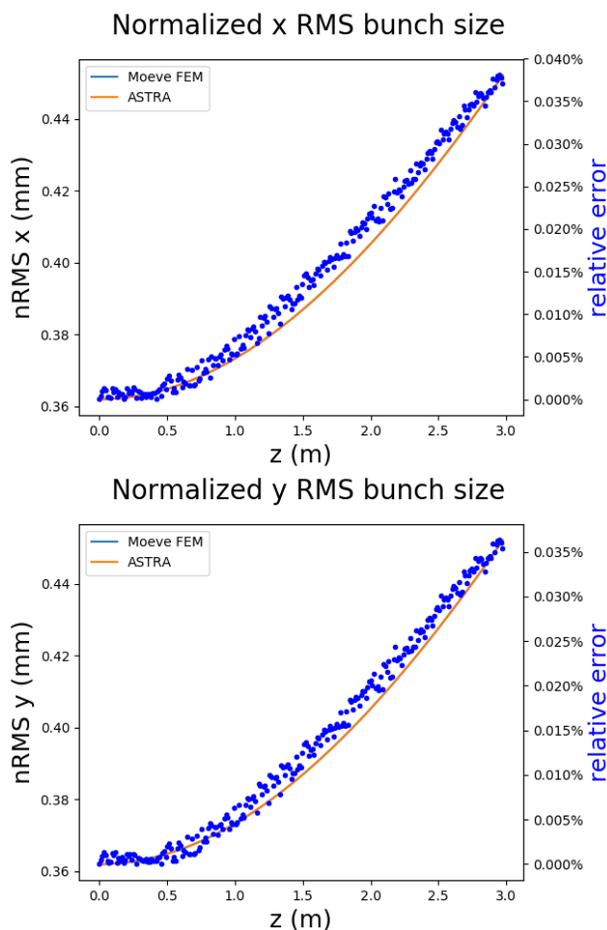


Figure 1: Bunch size growth in transverse directions for a drift distance of 3.0 m without external electromagnetic fields as computed by MOEVE based on FEM and ASTRA, respectively. Both curves agree very well. The relative error between both results is shown as well in the same plot, each.

It can be seen that the results of MOEVE with the FEM-based FEniCS implementation and ASTRA agree very well both for the transverse bunch size growth and the emittance. Regarding the relative error in the transverse bunch size growth, it is observable that the relative error grows follow-

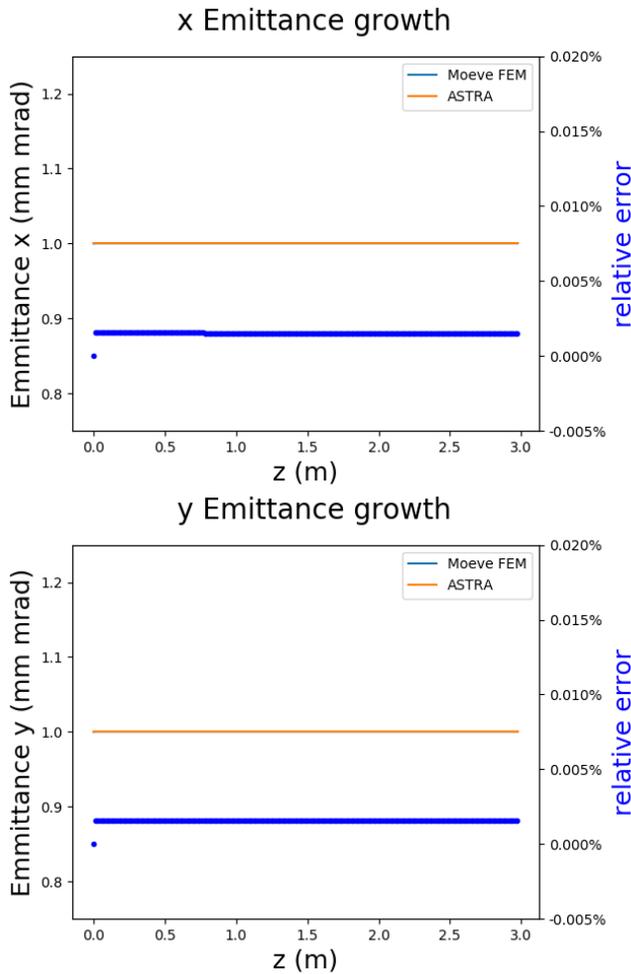


Figure 2: Emittance growth in transverse directions for a drift distance of 3.0 m without external electromagnetic fields as computed by MOEVE based on FEM and ASTRA, respectively. Both curves agree very well. The relative error between both results is shown as well in the same plot, each.

ing a very similar functional behaviour as the transverse bunch size growth itself. In that, the relative error after a drift of 3.0 m without external electric field reaches about 0.0371% and 0.0353% for  $x$  and  $y$ , respectively. On the other hand, the emittance stays constant over the drift. So does the relative error between the results of MOEVE with the FEM-based FEniCS implementation and ASTRA. Its value is about 0.0295% and 0.0317% for  $x$  and  $y$ , respectively. It should be noted that the new implementation in FEniCS needs only  $8.0 \times 10^4$  DoFs for a mesh of 0.3 m in  $z$ -direction then moving along the drift distance of a 3.0 m long vacuum chamber while ASTRA uses a grid of  $64^3 \approx 2.6 \times 10^5$  DoFs, which covers the bunch area only. We used an Intel Xeon workstation with 3.7 GHz CPU for both simulations. The CPU time of our current FEniCS implementation is not yet compatible to ASTRA since the procedures are not optimized yet — e.g. adaptive time stepping and other efficiency measures (see below) have not been implemented by now.

## CONCLUSION AND FUTURE PERSPECTIVE

In a pilot study, we developed an FEM-based model to track electron bunches. A first study on a simple model problem, tracking through a drift space without external electric field, showed very good agreement with results obtained by ASTRA.

Next, we aim to improve this FEM-based numerical model to study ion cloud dynamics using realistic geometries for the accelerator components. To achieve a high computational performance, we will employ MPI parallelization throughout the code. Also, the charge weighting with `PointSource` as well as the speed of interpolating the accelerating field at the particles' position can still be improved.

Additionally, we want to use the flexibility of the FEM approach to implement adaptive hp-refinements, i.e. with respect to element size ( $h$ ) and to the polynomial degree ( $p$ ) of the FEM approach. Elements that have macro-particles allocated to them (e.g. that are close to the bunch or covering it) should use a low-order approximation and be small [17], while further away from high charge densities one can use larger elements with a high order of approximation.

For validation, we will employ measurement results obtained at the Electron Stretcher Accelerator (ELSA) in Bonn [18]. Then, we will study the ion cloud dynamics to be expected in the proposed ERL bERLinPro at Helmholtz-Zentrum Berlin (HZB) [10]. In general, this numerical model can serve for ion cloud studies to estimate and reduce the effects related to ionized residual gas in high-current electron machines.

## REFERENCES

- [1] G. Pöplau. MOEVE: Multigrid Poisson Solver for Non-Equidistant Tensor Product Meshes. Germany, (2003).
- [2] G. Pöplau, A. Meseck, and U. van Rienen. Simulation of the interaction of an electron beam with ionized residual gas, in *Proc. IPAC 2011*, 2011, pp. 2250-2252.
- [3] G. Pöplau, U. van Rienen, A. Meseck. Numerical studies of the behavior of ionized residual gas in an energy recovering linac, *Phys. Rev. ST Accel. Beams* 18, 044401 (2015).
- [4] G. Pöplau, et al. Simulations for ion clearing in an ERL, in *Proc. ICAP 2012*, 2012.
- [5] A. Markovic, G. Pöplau, and U. van Rienen. Computation of the 2D transverse wake function of an electron-cloud for different parameters, in *Proc. IPAC 2012*, 2012.
- [6] S. van der Geer, M. de Loos. The general particle tracer code. Design implementation and application, 2001.
- [7] K. Flöttmann. ASTRA: A space charge tracking algorithm. Manual, Version 3 (2011): 2014.
- [8] G. H. Hoffstaetter, M. Liepe. Ion clearing in an ERL, *Nuclear Instruments and Methods in Physics Research Section A* 557, 2006, pp. 205-212.
- [9] E. Brentegani, W. Hillert, A. Meseck, D. Sauerland, U. van Rienen. Simulation of the Distribution of Parasitic Ions in the Potential of an Electron Beam, in *Proc. ICAP 2015*, Shanghai, China (2015).

- Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.
- [10] B. Kuske, N. Paulick, A. Jankowiak, J. Knobloch. Conceptual Design Report (CDR). HZB, Berlin, Germany (2011).
- [11] A. Markovic. Simulation of the interaction of positively charged Beams and electron clouds. PhD Thesis. Rostock University, 2013.
- [12] C. Bahls. Space charged calculations using refinements on structured and unstructured grids. PhD Thesis. Rostock University, 2015.
- [13] A. Logg, K.-A. Mardal, G. N. Wells et al. Automated Solution of Differential Equations by the Finite Element Method. Springer, 2012
- [14] A. Logg, G. N. Wells and J. Hake. DOLFIN: a C++/Python Finite Element Library. Automated Solution of Differential Equations by the Finite Element Method. Volume 84 of Lecture Notes in Computational Science and Engineering, Edited by A. Logg, K.-A. Mardal and G. N. Wells, Springer, chapter 10, 2012
- [15] J.B. Boris. Relativistic plasma simulation-optimization of a hybrid code", in *Proceedings of the 4th Conference on Numerical Simulation of Plasmas*, November 1970. Naval Res. Lab., Washington, D.C., pp. 3-67.
- [16] K. Floettmann. Some basic features of the beam emittance, *Phys. Rev. ST Accel. Beams* 6, 034202, 2005.
- [17] C. Bahls, G. Poeplau, U. van Rienen. Using Nudg++ to Solve Poisson's Equation on Unstructured Grids, in *Proc. Scientific Computing in Electrical Engineering SCEE 2008*. Springer, 2010.
- [18] D. Sauerland, W. Hillert, A. Meseck. Estimation of the ion density in accelerators using the beam transfer function technique, in *Proc. IPAC 2015*, Richmond, VA, USA (2015).

# STATISTICAL ANALYSIS OF THE EIGENMODE SPECTRUM IN THE SRF CAVITIES WITH MECHANICAL IMPERFECTIONS\*

A. Lunin<sup>†</sup>, T. Khabiboulline, N. Solyak, A. Sukhanov, V. Yakovlev, Fermilab, 60510 Batavia, USA

## Abstract

The superconducting radio frequency (SRF) technology is progressing rapidly over the last decades toward high accelerating gradients and low surface resistance making feasible the particle accelerators operation with high beam currents and long duty factors. However, the coherent RF losses due to high order modes (HOMs) excitation becomes a limiting factor for these regimes. In spite of the operating mode, which is tuned separately, the parameters of HOMs vary from one cavity to another due to finite mechanical tolerances during cavities fabrication. It is vital to know in advance the spread of HOM parameters in order to predict unexpected cryogenic losses, overheating of beam line components and to keep stable beam dynamics. In this paper we present the method of generating the unique cavity geometry with imperfections while preserving operating mode frequency and field flatness. Based on the eigenmode spectrum calculation of a series of randomly generated cavities, we can accumulate the data for the evaluation of the HOM statistics. Finally, we describe the procedure for the estimation of the probability of the resonant HOM losses in the SRF resonators. The study of these effects leads to specifications of SRF cavity and cryomodule and can significantly impact the efficiency and reliability of the machine operation.

## INTRODUCTION

Over recent decades the progress in SRF technology has made it feasible for a number of applications of the particle accelerators to operate in the continuous wave (CW) regime with a high beam current. There is an active demand on such machines based on multiple projects in the industry, high energy physics and material science, such as developing subcritical fission reactors based on an accelerator driven system (ADS), next generations of neutrino facilities and neutron spallation sources (PIP-II, ESS), radioactive ion beam facilities (RIBs) and free electron lasers (FELs) [1-6]. Variety of experimental programs often require a complex beam pattern and an ultra-short bunch length. Figure 1 shows typical examples of the dense beam frequencies spectrum in the PIP-II proton linac and the broadband power spectrum of wake fields generated by a series of 25  $\mu\text{m}$  rms bunches in the LCLS-II cryomodule. Evidently a combination of a large average beam current, a high bunch repetition rate and a broadband generated wake fields might result in significant cavity rf losses. The most danger comes out of the trapped HOMs in a case of

their coherent excitation by the beam. The later causes excessive cryogenic loads, overheating of beam line components and beam emittance dilution.

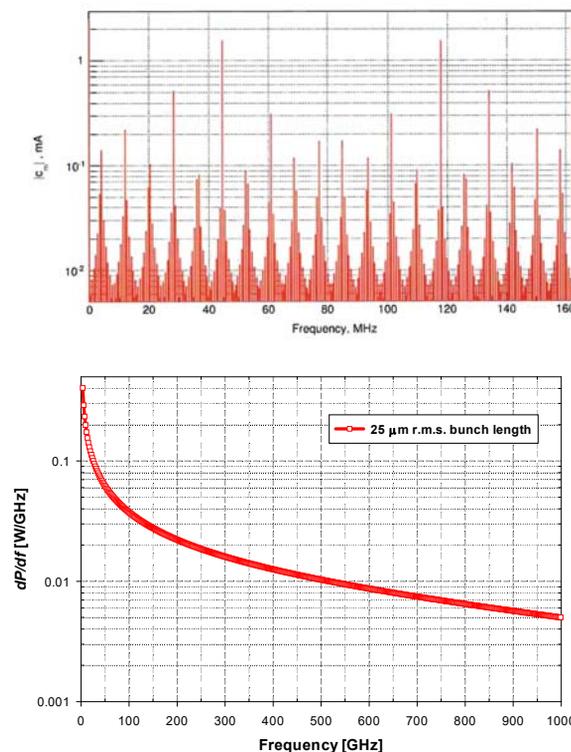


Figure 1: Beam frequency spectrum in the PIP-II linac (up) and power spectrum of wake fields generated by a series of ultra-short bunches in the LCLS-II cryomodule (down)

Due to a nature of SRF cavities they are very good resonance systems with multiple low loss eigenmodes with high intrinsic quality factors. For the coherent excitation one of the beam harmonics must coincide or be close to HOM frequencies. At the same time the HOM spectra in actual cavities will have significant frequency spreads comparing to the cavity with ideal geometry due to mechanical errors. Because of a randomness of mechanical errors, the resonant HOM excitation by the beam is inherently the probabilistic issue. The idea is illustrated on Figure 2, where the left sketch shows overlapping of the beam spectrum line and the HOM frequency spread with a high probability of coherent HOM excitation. The problem is complicated if we consider the propagating HOMs with frequencies above the beam pipe cut off. In this case the probability of mode trapping depends on the frequencies of neighbour cavities and, thus, taking into account the stochastic behaviour of cavity HOMs spectrum is essential for a proper analysis of the HOMs excitation.

\* Work supported by Fermi Research Alliance, LLC, under Contract DE-AC02-07CH11359 with the U.S. DOE

<sup>†</sup> lunin@fnal.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

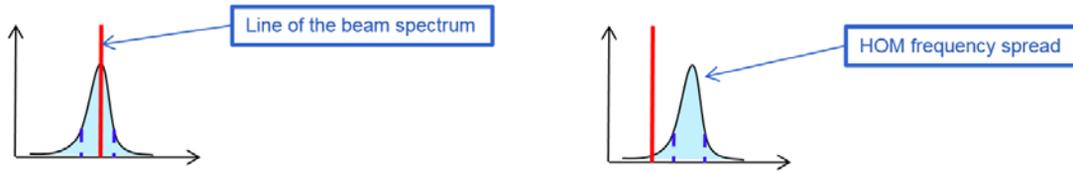


Figure 2: Stochastic HOM excitation by the pulsed beam. High excitation probability (left) corresponds to the case of the beam harmonic frequency within the HOM frequency spread and low probability (right) represents the beam spectrum line which is far off the HOM frequency deviation bandwidth.

Recent studies of the HOMs excitation in SRF cavities performed by different research groups was based either on the pre-deterministic approach, when the given spreads of HOM parameters are used for an evaluation of the worst-case scenario, or on the experimental data of frequencies and quality factors measured for most dangerous HOM passbands [7-9]. Both methods give only approximate results since they use rough or limited estimations for HOMs frequencies, tend to overstate quality factors and don't consider possible deviations of HOMs shunt impedances. The accurate evaluation of HOMs coherent effects is important to the design stage of SRF particle accelerators because it determines mechanical, thermal and electromagnetic requirements for expensive accelerator components and might set certain limits on machine operational scenarios. In this paper we propose the method of modelling the eigenmode spectrum in the multi-cell SRF cavities with finite mechanical tolerances defined by the fabrication technology. The method is based on the generation of unique cavity cell geometries with random dimension errors and the instant tuning of individual cells frequencies to preserve cavity operating mode field flatness. By doing the eigenmode spectrum calculation of a series of randomly generated cavities we can accumulate the data for the evaluation the HOM statistics. We used the proposed technique for HOMs analysis in high energy (HE) 650 MHz cavities of the PIP-II linac and 3.9 GHz cavities of the LCLS-II project [10,11]. Finally, we compared our results with available HOMs experimental data and present the procedure for the estimations of cumulative probabilities of resonant HOM losses and the beam emittance dilution in the string of SRF cavities.

## EIGENFREQUENCY ANALYSIS OF THE SRF CAVITIES WITH MECHANICAL IMPERFECTIONS

Conventional thin-walled niobium SRF cavities consist of multiple shell components welded together. Mechanical forming of such components and further electron-beam welding introduce significant uncertainty for the final cavity geometry. Typical maximum deviations of cavity profiles in respect to the ideal shape are about  $\pm 200 \mu\text{m}$  and  $\pm 100 \mu\text{m}$  for 1.3 GHz and 3.9 GHz cavities respectively [12,13]. Therefore, cavities get tuned for adjusting operating frequencies and preserving the field flatness in multicell cavities. Since each cavity has a unique geometry, the HOM spectrums vary from cavity to cavity and, then, the

beam to cavity interaction has a probabilistic nature. The ideal geometry of elliptical cavity cells can be characterized by eight parameters and another few parameters are required to describe mechanical imperfections like the cell to cell non-concentricity and the cell transverse deformation. Figure 3 shows ideal geometries of the end and regular cells for a typical multicell elliptical cavity.

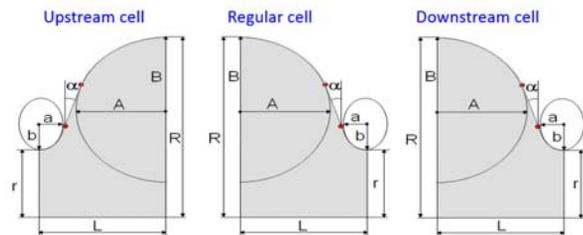


Figure 3: Nominal geometries of the multicell elliptical cavity.

For collecting accurate HOM statistics we perform the eigenmode analysis of HOM spectrum in the cavity with mechanical errors. The procedure is based on adding random components to the geometrical parameters of each cavity cell:

$$P_n^i = P_n^{nom} + |\Delta_{tol}| [2R \text{rnd}(1) - 1],$$

where  $P_n^i$  is a random dimension of the individual cell,  $P_n^{nom}$  is the nominal value of the  $n^{\text{th}}$  geometrical parameter of the half-cell,  $i$  is the cell number,  $\Delta_{tol}$  is the mechanical tolerance of a cavity fabrication and  $\text{rnd}(1)$  is the uniform random function in the range of 0 to 1. For a preliminary tuning of the cavity operating frequency we need to calculate a frequency-dependent sensitivity for each of cavity geometrical parameters. Next frequencies of the half-cells are tuned by adjusting the cell lengths similarly as it happens with real cavities to preserve the field flatness. The half-cell tuning frequency balance is described then:

$$\Delta L^i \frac{\partial f}{\partial L^i} = - \sum_{n=1}^N \left[ \Delta P_n^i \frac{\partial f}{\partial P_n^i} \right],$$

where  $\Delta L^i$  is the length compensation of the  $i^{\text{th}}$  half-cell,  $\Delta P_n^i = P_n^i - P_n^{nom}$  is the random error of the  $i^{\text{th}}$  half-cell dimension,  $\partial f / \partial L^i$  and  $\partial f / \partial P_n^i$  are frequency-dependent sensitivities of the half-cell length and the  $n^{\text{th}}$  geometrical parameter. We assume here that small deviations of each parameter won't influence the sensitivity of other parameters, all mechanical tolerances are the same and uncorrelated with each other.

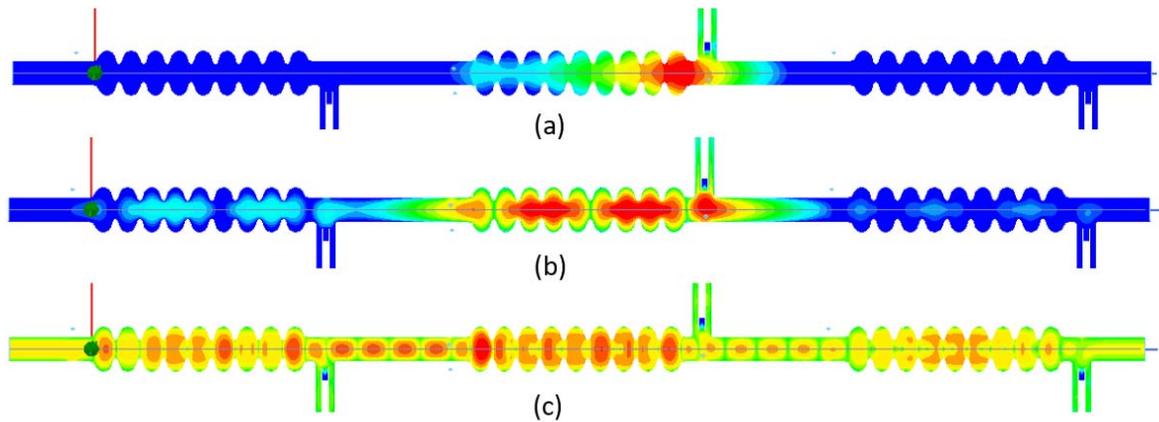


Figure 4: Trapped modes in the infinite chain of SRF cavities with random HOM spectrum: (a) type I high-Q HOM is localized within the cavity volume, (b) type II medium-Q HOM occupies the cavity and adjacent beam pipes and (c) type III low-Q HOM distributed along the cavities chain. All plots show the complex magnitude of electric field, where a blue colour corresponds to the zero amplitude.

Conventional accelerating cryomodule consists of a series of superconducting cavities connected by copper plated normal conducting bellows. Because of the unique HOM spectrum of each cavity common modes in the chain of identical ideal cavities will split into modes of individual resonators in the real cryomodule. Following this approach, we can categorize the trapped HOMs in the cryomodule into three types depending on their quality factors. The first kind of trapped HOMs is weakly coupled with the beam pipe and most of the HOM stored energy is localized within the superconducting cavity volume resulting in a high-quality factor. The second kind of trapped HOMs has a good coupling with the beam pipe but the signal is reflected by the neighbouring cavities and a part of the stored energy fills in the volume of adjacent beam pipes causing an additional signal dampening through far upstream and downstream coupler ports and by normal conducting bellows. The third type of HOM represents the case where the HOM signal can propagate through the neighbouring cavities. Then the stored energy is distributed along the cryomodule beam line and damped by many coupler ports and by ohmic losses in interconnecting cavity bellows. Such modes have low quality factors and we don't count them as a dangerous resonant mode. Figure 4 illustrates typical electric field distribution for what is described above, three types of trapped HOMs in the chain of SRF cavities with random HOM spectrums. Evidently the single cavity model is adequate for the HOM type I analysis, while the HOM types II and III require at least the chain of three cavities for accurate simulations. Further increasing the number of cavities might improve results for the low-Q HOMs but at the same time results in unnecessary complication of the eigenmode analysis and a longer simulation time. Thus, we conclude that the chain of three cavities is an optimum choice for the calculation of resonant HOMs spectrum cryomodule and longer chains give a little or no impact to the overall result.

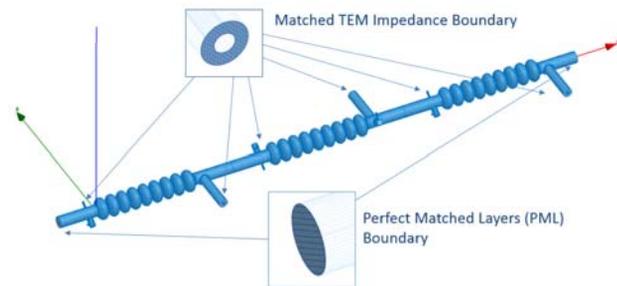


Figure 5: ANSYS HFSS model for the eigenmode analysis of the chain of three SRF cavities.

Finally, we create the 3D model of three random cavities with auxiliary coupler ports. For example, Figure 5 shows the ANSYS HFSS model of the chain of three elliptical cavities with matched boundary conditions [14]. All coaxial TEM ports are terminated with the constant free space impedance boundary, while the free radiation to the round beam pipe is ensured by the perfectly matched layers boundary. Since the number of trapped HOMs in the cavity chains are quite large, it is preferable to perform the HOM spectra simulation in series of few tens of eigenmodes starting from the operating passband for achieving better convergence and accuracy. The upper frequency limit is defined by the increasing number of propagating modes in the beam pipe at higher frequencies where the HOM spectrum becomes almost continuous. For mode sorting it is necessary to set up and calculate secondary values during the HOMs analysis: local stored energy in each cavity with adjacent beam pipes, longitudinal and transverse shunt impedances and partial quality factors for coupler ports. By sorting the HOMs compendium we can exclude the end cavities spectra and recognize monopole, dipole and quadrupole HOMs passbands. As the last step we calculated the HOMs statistics, mean and rms values for frequencies, shunt impedances and quality factors. Having in hand

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

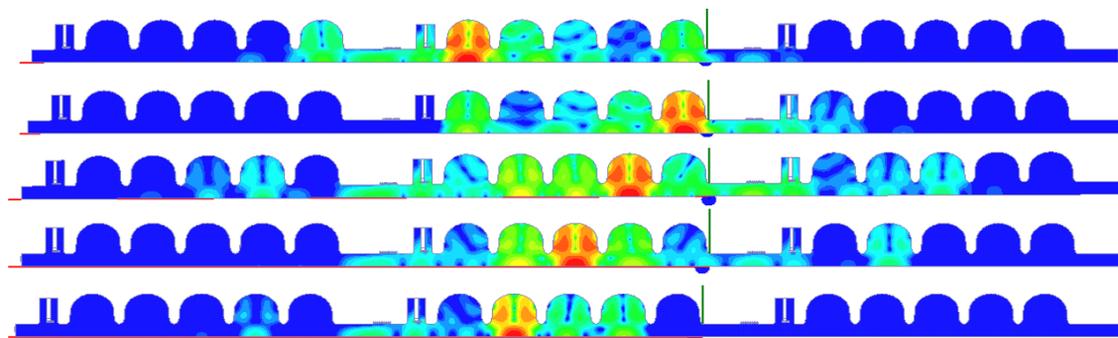


Figure 6: Stochastic HOM excitation by the pulsed beam. High excitation probability (left) corresponds to the case of the beam harmonic frequency within the HOM frequency spread and low probability (right) represents the beam spectrum line which is far off the HOM frequency deviation bandwidth.

HOMs statistics it is possible to perform a Monte Carlo analysis for a coherent HOMs excitation in the chain of SRF cavities by the beam with any given average current and bunch patterns thus, derive the probabilities for excessive cryogenic losses and beam emittance dilution.

### APPLICATION OF THE STOCHASTIC ANALYSIS OF HOM SPECTRUM

Originally, we developed a stochastic approach for the HOMs analysis in the high energy (HE) 650 MHz five cells elliptical structure of the Project-X linac [15]. During the stage of rf design it was found that the 5<sup>th</sup> monopole band in a cavity with ideal geometry has extremely narrow pass-band of few tens of kilohertz while the expecting HOMs frequencies deviation due to mechanical errors is at least few megahertz. In this case, when coupling between cells is weak and varies from cell to cell, the usual pass-band structure of  $N$  modes of  $m\pi/N$ -kind, where  $N$  is the number of cells and  $m$  runs from 1 to  $N$ , may change [16]. The field of a cavity mode may be concentrated in a single cell, or two adjacent cells. The calculated distribution of electric fields for the 5<sup>th</sup> monopole band is presented in Figure 6 for the chain of three random HE 650 MHz cavities. Later we verified field distributions of the 5<sup>th</sup> monopole band by bead pull measurements on the prototype niobium cavities. Typical measured field distributions are summarized in Figure 7. Evidently the experimental results are in perfect agreement with both theoretical and numerical predictions.

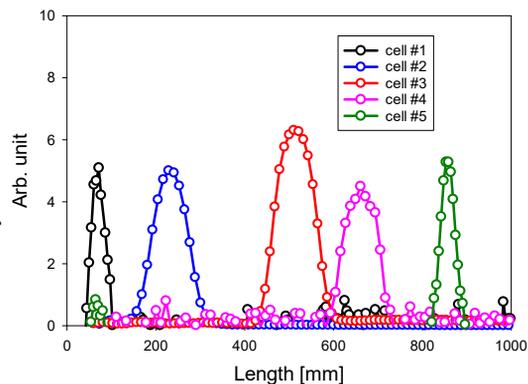


Figure 7: Bead pull measurement of on-axis field profiles of the 5<sup>th</sup> monopole band in the HE 650 MHz structure

For the second time we calculated HOM statistics for the third harmonic cavity of LCLS-II project [11]. Figure 8 shows frequency standard deviations for modes in the first monopole band. There is a good agreement between calculations and measured 2 K data for the similar 3.9 GHz cavity developed for the XFEL project [17]. The average calculated field flatness of the operating mode is above 80% for both studies, which is close to the conventional specification of minimal 90% field flatness for multicell cavities. Therefore, we conclude that proposed method of stochastic analysis of HOM parameters in the SRF cavities with mechanical imperfections provide a reliable data for the further statistical analysis of the coherent HOM excitation.

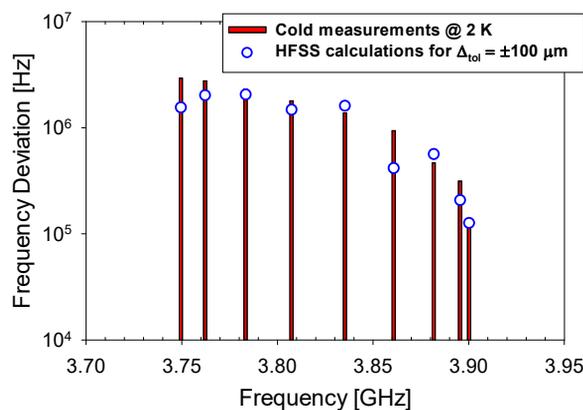


Figure 8: Frequency standard deviations for the 1<sup>st</sup> monopole band of the 3.9 GHz cavity.

By using the predicted deviations of monopole and dipole HOMs frequencies, shunt impedances and quality factors we can generate the chains of cavities with random HOMs spectra and, thus, estimate probabilities of the rf losses or the beam emittance dilution in real superconducting linacs. As an example, the resulting cumulative probability of rf losses are presented in Figure 9 for the HE 650 MHz PIP-II cryomodule (left) and the LCLS-II third harmonic section (right) respectively.

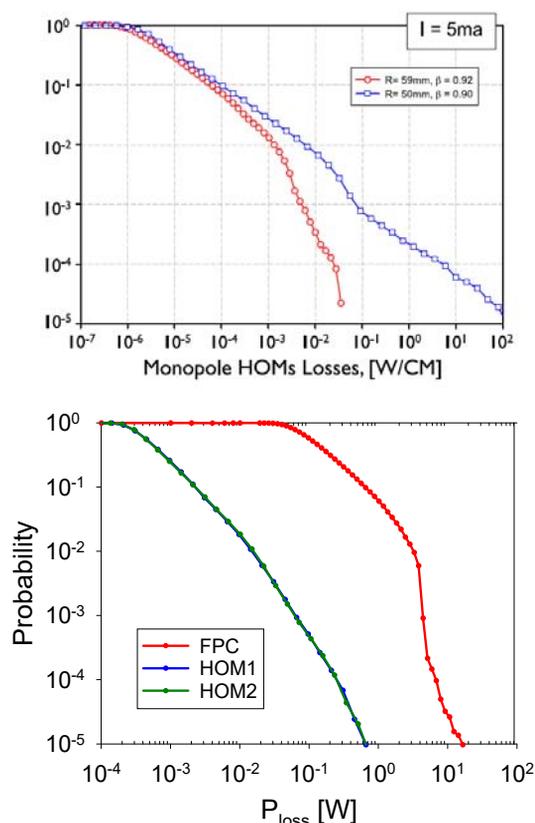


Figure 9: Cumulative probabilities of monopole HOMs RF losses in the HE 650 MHz PIP-II (up) and 3.9 GHz LCLS-II (down) cryomodules.

The accurate quantitative estimation of the negative effects caused by the resonant HOM excitations is important for a comparison of different cavity designs and various regimes of the beam operation. Based on HOMs statistics we chose an optimal design for the HE 650 MHz cavity for PIP-II project and modified the end group in the third harmonic cavity for the LCLS-II linac [18,19]. In summary, we conclude that the proposed method of statistical HOM spectra evaluation in the SRF cavities with mechanical imperfections is straightforward and reliable instrument for the risk analysis of the coherent HOM excitation in superconducting particle accelerators.

### CONCLUSION

The statistical analysis of the eigenmode spectrum in SRF cavities is reliable and accurate tool for quantitative evaluation of the coherent HOM excitation by the beam with arbitrary time structure. The outcome of HOM analysis resulted in critical decisions for the design of superconducting accelerating cavities. Simplification of the cavity production and operation is a significant part of the overall cost reduction for both machines. The proposed technique can be easily adapted and used for other superconducting particle accelerators operating at high average beam current and high duty factor regimes.

### REFERENCES

- [1] H. A. Abderrahim *et al.*, “MYRRHA, a Multipurpose Hybrid Research Reactor for High-End Applications”, *Nucl. Phys. News*, vol. 20, issue 1, pp. 24, 2010.
- [2] V. Lebedev *et al.*, “The PIP-II Conceptual Design Report”, Fermilab, Batavia, FERMILAB-TM-2649-AD-APC, 2017.
- [3] S. Peggs (Eds.), “ESS Technical Design Report”, Lund, ESS-doc-274, 2013.
- [4] FRIB, <http://www.frib.msu.edu>
- [5] M. Altarelli *et al.*, (Eds.), “The European X-Ray Free-Electron Laser Technical Design Report”, DESY, Hamburg, Preprint DESY 2006-097, 2006.
- [6] T. Raubenheimer *et al.*, “SLAC LCLS II Final Design Report”, SLAC, Menlo Park, LCLSII-1.1-DR-0251, 2015.
- [7] M. Schuh *et al.*, “Influence of higher order modes on the beam stability in the high power superconducting proton linac”, *Phys. Rev. ST Accel. Beams*, vol. 14, pp. 051001, 2011.
- [8] N. Valles *et al.*, “HOM studies of the Cornell ERL Main Linac Cavity in the Horizontal Test Cryomodule”, in *Proc. SRF 2013*, Paris, France, Sep. 2013, doi:10.18429/JACoW-SRF2013-THP071
- [9] K. Bane, *et al.*, “A Study of Resonant Excitation of Longitudinal HOMs in the Cryomodules of LCLS-II”, in *Proc. SRF 2015*, Whistler, BC, Canada, Sep. 2015. doi:10.18429/JACoW-SRF2015-THPB007
- [10] A. Sukhanov *et al.*, “High order modes in Project-X Linac”, *Nucl. Instr. Methods Phys. Res., Sect. A* 734, pp. 9, 2014.
- [11] A. Lunin, *et al.*, “Resonant excitation of high order modes in the 3.9 GHz cavity of the Linac Coherent Light Source”, *Phys. Rev. ST Accel. Beams* 21, 022001 (2018).
- [12] A. Sulimov *et al.*, “Estimation of Small Geometry Deviation for TESLA-Shape Cavities due to Inner Surface Polishing”, in *Proc. SRF 2013*, Paris, France, Sep. 2013. doi:10.18429/JACoW-SRF2013-TUP053
- [13] N. Solyak *et al.*, “Development of the Third Harmonic SC Cavity at Fermilab”, in *Proc. PAC’03*, Portland, OR, USA, May 2003. doi:10.18429/JACoW-PAC2003-TPAB014
- [14] ANSYS, Inc., <http://www.ansys.com/Products/Electronics/ANSYS-HFSS>
- [15] A. Lunin *et al.*, “Alternative  $\beta=0.92$  cavity for Project X”, Fermilab, USA, Project X Document 1179-v1, 2013.
- [16] H. Padamsee, J. Knobloch, T. Hays, *RF Superconductivity for Accelerators*, Wiley, New York, 1998.
- [17] T. Khabibouline *et al.*, “Higher Order Modes of a 3rd Harmonic Cavity with an Increased End-cup Iris”, Fermilab, Technical Report FERMILAB-TM-2210, 2003.
- [18] A. Lunin *et al.*, “Alternative Cavity for the HE Part of the Project X Linac”, in *Proc. IPAC’12*, New Orleans, USA, May 2012. doi:10.18429/JACoW-IPAC2012-WEPPR029
- [19] A. Lunin *et al.*, “Redesign of the End Group in the 3.9 GHz LCLS II Cavity”, in *Proc. LINAC’16*, East Lansing, USA, Sep. 2016. doi:10.18429/JACoW-LINAC2016-MOPLR007

# EFFICIENT COMPUTATION OF LOSSY HIGHER ORDER MODES IN COMPLEX SRF CAVITIES USING REDUCED ORDER MODELS AND NONLINEAR EIGENVALUE PROBLEM ALGORITHMS\*

H. W. Pommerenke<sup>†</sup>, J. D. Heller, U. van Rienen<sup>1</sup>

Institute of General Electrical Engineering, University of Rostock, Germany

<sup>1</sup>also at Department Life, Light & Matter, University of Rostock, Germany

## Abstract

Superconducting radio frequency (SRF) cavities meet the demanding performance requirements of modern accelerators and high-brilliance light sources. For the operation and design of such resonators, a very precise knowledge of their electromagnetic resonances is required. The non-trivial cavity shape demands a numerical solution of Maxwell's equations to compute the resonant eigenfrequencies, eigenmodes, and their losses. For large and complex structures this is hardly possible on conventional hardware due to the high number of degrees of freedom required to obtain an accurate solution. In previous work it has been shown that the considered problems can be solved on workstation computers without extensive simplification of the structure itself by a combination of State-Space Concatenation (SSC) and Newton iteration to solve the arising nonlinear eigenvalue problem (NLEVP).

First, SSC is applied to the complex, closed and thus lossless RF structure. SSC employs a combination of model order reduction and domain decomposition, greatly reducing the computational effort by effectively limiting the considered frequency domain. Next, a perturbation approach based on SSC is used to describe the resonances of the same geometry subject to external losses. This results in a NLEVP which can be solved efficiently by Newton's method. In this paper, we expand the NLEVP solution algorithm by a contour integral technique, which increases the completeness of the solution set.

## INTRODUCTION

Superconducting radio frequency (SRF) cavities are essential components of modern particle accelerators, as they provide the radio frequency (RF) electromagnetic fields used to accelerate charged particles to high energies. The design of RF cavities requires a precise knowledge of their resonant frequencies  $f$ , field distributions, and power losses  $P$ . This usually requires solving an eigenvalue problem, where the eigenvalues and eigenvectors correspond to the frequencies and field distributions, respectively. In this context, the eigenvectors are also denoted as eigenmodes of the cavity.

A dimensionless measure for power losses in general is the quality factor

$$Q = 2\pi fW/P, \quad (1)$$

\* This work has been supported by the German Federal Ministry of Education and Research (grant no. 05K13HR1).

<sup>†</sup> Currently also at CERN, Geneva, Switzerland.

Email: hermann.pommerenke@uni-rostock.de

which is the ratio between the energy loss per oscillation and the total energy  $W$  stored in the electromagnetic field. Generally, there are dielectric, magnetic, surface and external losses. The latter occur when energy is propagating out of the cavity through its openings, e.g. a coupler or the beam pipe. For an SRF cavity, the external losses are several orders of magnitude larger than other loss mechanisms, since the structure is both superconducting and evacuated [1]. Therefore, the external quality factor  $Q_{\text{ext}}$  is often equivalent to the total  $Q$  (and in the following denoted as such). External losses are of significant importance for eigenmodes, whose resonant frequencies are larger than that of the operating mode used for acceleration. They are denoted as higher order modes (HOM). These usually unwanted modes are excited by the current of the passing beam and may influence the beam in an unwanted manner, e.g. by deviation from its optimum trajectory or emittance growth [2, 3]. One usually designs cavities such that HOM energy is dissipated quickly and the mode is practically completely damped before the next particle bunch arrives. The structures must thus feature low  $Q$  factors regarding the HOMs. Besides available openings like the beam pipe or the power coupler, HOM couplers are utilized. Nevertheless, there exist additional HOMs whose interaction with the couplers is almost non-existent and which therefore have very high  $Q$  factors. Identification and computation of these trapped modes is particularly important in SRF cavity design [1, 4, 5].

Even for comparatively simple structures, an analytical solution of Maxwell's equations [6] is not available. Numerical methods such as the Finite Element Method (FEM) [7] or Finite Integration Technique (FIT) [8, 9] are therefore employed. If one solely considers closed lossless cavities, this leads to a linear eigenvalue problem (LEVP), whose solution can be acquired by a variety of methods. However, the precise computation of external losses is accomplished by applying suitable boundary conditions to the cavity's openings leading to a complex-valued, nonlinear eigenvalue problem (NLEVP), whose solution requires significantly more effort.

The above-mentioned numerical methods show disadvantageous scaling behavior regarding size and complexity of the structure. Especially large and complex structures, e.g. a sequence of cavities and couplers like in Fig. 1, require many degrees of freedom (DOF) for an accurate solution. In a direct approach, these problems can only be solved on powerful computational infrastructure which is costly and rarely available. Another possibility is to only con-

sider a part of the structure by restricting the computational domain. While this approach is suitable if the fields are confined in one cavity, a significant portion of HOM energy is stored in electromagnetic fields that may fill the entire structure [1] (Fig. 1(b)). Such fields cannot be computed if the domain is limited to a single cavity.

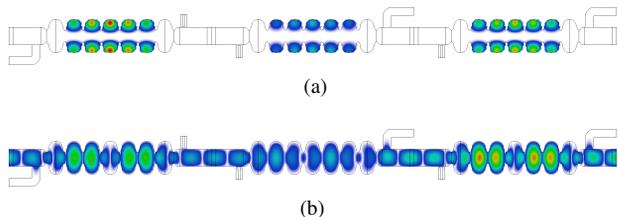


Figure 1: Electric field of two HOMs of the bERLinPro linac [10]. (a) Mode resonating at 3.68 GHz that is confined within the cavities. (b) Mode resonating at 1.86 GHz with a field distribution filling the entire structure. The plots are taken from [11].

In [11] it was shown by Heller that employing the State-Space Concatenation Scheme (SSC) [12, 13] together with suitable boundary conditions leads to a NLEVP which can be solved efficiently by Newton’s method [14]. In this paper, we extend upon these results by combining them with a contour integral algorithm proposed by Beyn [15] to increase the completeness of the solution set. The proposed technique has been used to investigate current examples from accelerator physics.

## STATE-SPACE CONCATENATION

SSC has been suggested by Flisgen [12, 13] and is a combination of domain decomposition and model order reduction (MOR) techniques to solve Maxwell’s equations for large complex SRF structures. We will introduce it briefly here; for an exhaustive explanation we refer to [13, 16–18].

First, the investigated structure is decomposed into non-overlapping segments. The substructures are treated separately, with the cutting planes considered as waveguide ports. For each segment, a State-Space model (SSM) is assembled. An incomplete eigendecomposition is performed, computing a finite number<sup>1</sup> of 3D eigenmodes around the frequency interval of interest. Due to poor convergence of the incomplete decomposition, a very high number of modes would be necessary, since even the structure’s behavior within a certain frequency range is influenced by modes outside said range. Instead, the orthogonal basis is expanded by so-called snapshots computed from the frequency response of the segment to certain excitations. This is done for a total of  $N_{3D}$  modes. This technique is denoted as Corrected Modal Expansion (CME) [17]. The 3D modes represent the internal states of the system comprised in the state vector  $\mathbf{x}$ . For each waveguide port, 2D port modes are computed, with  $N_{2D}$  denoting the total number of port modes of the segment.

<sup>1</sup> The amount of 3D eigenmodes is by orders of magnitude smaller than the DOFs.

Their amplitudes are comprised in the vector of modal currents  $\mathbf{i}$  and modal voltages  $\mathbf{v}$ . The coupling between 3D eigenmodes and 2D port modes is obtained as the inner product of the respective fields. In frequency domain, the assembled first-order SSM<sup>2</sup> reads as

$$\begin{aligned} s\mathbf{x}(s) &= \mathbf{A}\mathbf{x}(s) + \mathbf{B}\mathbf{i}(s) \\ \mathbf{v}(s) &= \mathbf{C}\mathbf{x}(s), \end{aligned} \quad (2)$$

with  $\mathbf{A} \in \mathbb{R}^{2N_{3D} \times 2N_{3D}}$ ,  $\mathbf{B} \in \mathbb{R}^{2N_{3D} \times N_{2D}}$  and  $\mathbf{C} = \mathbf{B}^T$ .

The individual SSMs of the segments are concatenated by assembling  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  from (2) into block matrices. Redundant modal currents and voltages at connected ports are eliminated using Kirchhoff’s laws, so that only those of external ports remain. The MOR by CME is again applied to the concatenated SSM, further reducing the DOFs. The resulting reduced SSM of the complex structure is again of the same form as (2). The external modal currents and voltages can be used to assign excitations or boundary conditions to the structure.

SSC is not the only Maxwell solution approach involving a domain decomposition and concatenation approach. Notable methods are the mode-matching techniques e.g. [4, 5, 19–22], the Coupled S-Parameter Calculation [18, 23–25], the Generalized Scattering Matrix approach [26–28] or the description of RF structures by means of circuit theory, e.g. [17, 29]. However, to the best of the authors’ knowledge, only SSC allows simultaneous and direct access to time and frequency domain calculations, MOR and 3D field distributions [11–13].

## EXTERNAL LOSSES IN SRF CAVITIES

For the computation of external losses of an SRF cavity, the openings are modeled as having infinitely long waveguides attached to them, i.e. being reflection-free. Therefore impedance matching is required: the termination impedance (i.e. the quotient of modal voltage and current) of each port mode must be the wave impedance of said mode [30].

Each eigenmode of the lossy structure is described as a weighted sum of the eigenmodes of the corresponding lossless structure. In this perturbation ansatz usually one of the lossless eigenmodes dominates. For the description of certain perturbations, very high frequencies are necessary; in that case a large number of lossless eigenmodes must be considered in the preceding MOR. The resulting NLEVP read as [11, 16, 30]

$$\mathbf{T}(\lambda)\mathbf{x} = \left( \mathbf{A} - \mathbf{B}\mathbf{G}(\lambda)\mathbf{B}^T - \lambda\mathbf{I} \right) \mathbf{x} = \mathbf{0}, \quad (3)$$

where  $\lambda$  and  $\mathbf{x}$  denote the eigenvalue and eigenvector, respectively. The individual frequencies and external  $Q$  factors are obtained by

$$f = \frac{\Im\{\lambda\}}{2\pi} \quad Q = -\frac{\Im\{\lambda\}}{2\Re\{\lambda\}}, \quad (4)$$

<sup>2</sup> The SSM may also be assembled as a symmetric second order system with half the degrees of freedom. Both can also be represented in time domain.

while the field distributions can be reconstructed from the respective eigenvectors.

The diagonal matrix  $\mathbf{G}(\lambda)$  contains the reciprocal wave impedances (i.e. wave admittances) of the port modes which depend nonlinearly on the eigenvalues and are given for transversal electric (TE) and magnetic (TM) port modes by [31]

$$\begin{aligned} Z_{\text{wave}}^{\text{TE}}(\lambda) &= Z_0 \frac{\lambda}{\sqrt{\lambda^2 + \omega_{\text{co}}^2}} \\ Z_{\text{wave}}^{\text{TM}}(\lambda) &= Z_0 \frac{\sqrt{\lambda^2 + \omega_{\text{co}}^2}}{\lambda}, \end{aligned} \quad (5)$$

where  $Z_0 \approx 377 \Omega$  is the impedance of free space and  $\omega_{\text{co}}$  denotes the cutoff frequency of the respective port mode. For transversal electromagnetic (TEM) waves this is the impedance of the connected transmission line, usually  $Z_{\text{wave}}^{\text{TEM}} = 50 \Omega$ .

The introduction of the wave impedances (5) causes the operator  $\mathbf{T}(\lambda)$  to be both nonlinear and meromorphic: isolated poles occur at  $\lambda = 0$  and at the cutoff frequencies  $\omega_{\text{co}}^{\text{TM}}$  of the TM port modes.

## SOLVING THE NLEVP

The solution of NLEVPs  $\mathbf{T}(\lambda)\mathbf{x} = \mathbf{0}$  such as (3) is significantly more demanding than that of well-known LEVPs, for which a wide range of solution methods exist [32–34]. The NLEVP is subject of active ongoing research. An overview over NLEVP solution approaches is given by [33, 35–39], notably the Newton iteration [14], contour integral methods [15, 40], or methods based on generalized QR decompositions [41] or Rayleigh functionals [42, 43].

In this work, we utilized Newton’s method as well as a contour integral algorithm suggested by Beyn [15]. Both are introduced in the following.

### Newton’s Method

The well-known Newton’s method [14] is used to successively approximate zeros of functions or operators. Solving the NLEVP (3) can be interpreted as finding zeros of  $\mathbf{T}(\lambda)\mathbf{x}$  in an  $(N + 1)$ -dimensional search space, where all components of the eigenvector  $\mathbf{x} \in \mathbb{C}^N$  and the eigenvalue  $\lambda$  itself must be determined. Hence, a suitable formulation of the Newton iteration can be found [33, 36, 37].

Without loss of generality, the eigenvector is normalized according to  $\mathbf{v}^H \mathbf{x} = 1$  and (3) is reformulated to

$$\mathbf{P} \begin{pmatrix} \mathbf{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{T}(\lambda)\mathbf{x} \\ \mathbf{v}^H \mathbf{x} - 1 \end{pmatrix} = \mathbf{0}. \quad (6)$$

Applying the iteration rule yields

$$\begin{pmatrix} \mathbf{x}_{\nu+1} \\ \lambda_{\nu+1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{\nu} \\ \lambda_{\nu} \end{pmatrix} - \left( \mathbf{P}' \begin{pmatrix} \mathbf{x}_{\nu} \\ \lambda_{\nu} \end{pmatrix} \right)^{-1} \mathbf{P} \begin{pmatrix} \mathbf{x}_{\nu} \\ \lambda_{\nu} \end{pmatrix}, \quad (7)$$

where  $\nu$  denotes the iteration index and  $\mathbf{P}'$  is called Fréchet derivative [44] of  $\mathbf{P}$ . The initial pair  $(\lambda_0, \mathbf{x}_0)$  must be known

a priori. The initial pairs are usually obtained by solving a linearized system or (additionally) sampling the domain of interest using a grid or Monte Carlo methods.

As (7) is difficult to implement, one introduces the search direction  $\mathbf{u}$  as an auxiliary quantity from which the approximated eigensolution is computed:

$$\begin{aligned} \mathbf{u}_{\nu+1} &= \mathbf{T}^{-1}(\lambda_{\nu}) \frac{\partial \mathbf{T}}{\partial \lambda}(\lambda_{\nu}) \mathbf{x}_{\nu} \\ \mathbf{x}_{\nu+1} &= c_{\nu+1} \mathbf{u}_{\nu+1} \\ \lambda_{\nu+1} &= \lambda_{\nu} - \frac{\mathbf{v}^H \mathbf{x}_{\nu}}{\mathbf{v}^H \mathbf{u}_{\nu+1}}, \end{aligned} \quad (8)$$

where  $c_{\nu}$  is a normalization factor that influences the convergence speed<sup>3</sup>. In general, the iteration (8) stops when an error measure falls below a certain limit<sup>4</sup>.

When employing Newton’s method, convergence against previously computed eigenpairs must be avoided, which is called deflation. Most techniques of the LEVP solvers cannot be used [33], since the eigenvectors of the NLEVP are generally linearly dependent. A possibility is to choose  $\mathbf{v}$  in each new solution attempt orthogonal to all previously computed eigenvectors. More sophisticated deflation techniques can be developed based on minimal invariant pairs, requiring an alternative formulation of the algorithm [39].

### Beyn’s Algorithm

The integral algorithm introduced by Beyn [15] allows for a complete solution of a NLEVP within a finite enclosed sub domain of the complex plane. We explain it briefly and refer to [15, 45, 46] for a detailed derivation.

The algorithm is based on Keldyš’s theorem [45] stating that the inverse of  $\mathbf{T}(z)$  may be expanded into a Laurent series, whose principal part can be expressed in terms of the right and left eigenvectors  $\mathbf{x}, \mathbf{y}$  belonging to  $\lambda$ :

$$\mathbf{T}^{-1}(z) = \sum_{j=1}^k \frac{1}{z - \lambda_j} \mathbf{x}_j \mathbf{y}_j + (\text{holomorphic part}). \quad (9)$$

Choosing a rectangular matrix  $\Psi$  at random, the integrals

$$\mathbf{L}_p = \frac{1}{2\pi j} \oint_{\Gamma} z^p \mathbf{T}^{-1}(z) \Psi dz, \quad p = 0, 1 \quad (10)$$

are computed using numerical integration along a closed contour  $\Gamma$ . The computational cost is dominated by solving the linear system  $\mathbf{T}^{-1}(z)\Psi$  at every quadrature sampling point. The convergence behavior depends on the condition of the operator  $\mathbf{T}(z)$ , the number of sampling points, and chosen quadrature rule [15, 47, 48].

From (9) it follows by Cauchy’s integral formula [34] that  $\mathbf{L}_0 = \mathbf{X}\mathbf{Y}\Psi$  and  $\mathbf{L}_1 = \mathbf{X}\lambda\mathbf{Y}\Psi$ . The matrices  $\mathbf{X}, \mathbf{Y}$  and

<sup>3</sup> In practice one chooses e.g.  $c_{\nu} = 1/||\mathbf{u}_{\nu}||$ ; however it has been observed that  $|c_{\nu}| \ll 1$  leads to instability of the algorithm in rare cases. In [33] the condition  $|c_{\nu}| < 1$  is mentioned to guarantee convergence.

<sup>4</sup> To account for potentially occurring non-convergence, the algorithm should additionally be stopped when a maximum number of iterations is reached.

$A$  contain the left and right eigenvectors and eigenvalues, respectively, that are enclosed by  $\Gamma$ , but are independent of the specific contour shape. They are extracted using singular value decomposition.

In this work, we combine Beyn's algorithm with a subsequent Newton iteration to improve convergence as suggested in [36, 38], since the algorithm on its own shows disadvantageous convergence behavior when employed to the SRF cavity NLEVP. The solutions found by Beyn's algorithm are used as starting pairs for Newton's method. This has initially been tested in [48] and is demonstrated below.

The alternative Contour Integral Slicing Method proposed by [40] should also be acknowledged. It is available within the SLEPc library [49] combined with a Newton iteration in a similar fashion.

## APPLICATION EXAMPLES

In the current setup implemented in [11], the SRF structure is discretized by FIT in CST Microwave Studio (CST MWS) [50]. The assembled matrices are transferred to and further processed in MATLAB [51] using a collection of scripts written in Python [52] and Visual Basic for Applications. The arising linear systems and decompositions are computed using linear algebra software packages like LAPACK [53] and ARPACK [54], which are made available by MATLAB and the Python packages NumPy [55] and SciPy [56]. The field plots are generated using Paraview [57].

All mentioned computations have been performed on an Intel Xeon E5-2687W CPU with 3.4 GHz clock rate, 256 GB RAM and Windows Server 2012.

### Academic Example

The hypothetical minimalistic resonator depicted in Fig. 2 serves as a proof of principle. The cavity has two waveguide ports symbolizing a beam pipe and a HOM coupler. The structure is discretized in CST MWS using FIT with 233,000 mesh cells and Perfect Boundary Approximation [58]; after the MOR the arising NLEVP is of order  $N = 178$ .

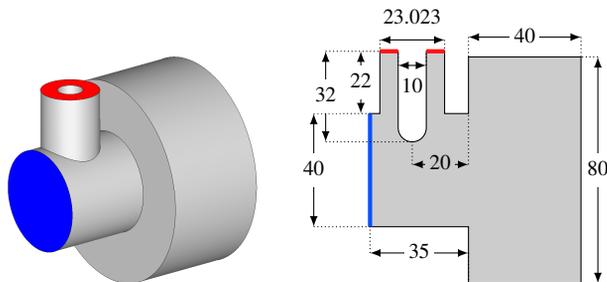


Figure 2: CAD model and sketch of the minimalistic resonator used as an academic example [11, 48]. Beam pipe and HOM coupler are highlighted in blue and red respectively. Geometric dimensions in mm.

Figure 3 shows the convergence behavior of Beyn's algorithm applied to this example with different numbers of

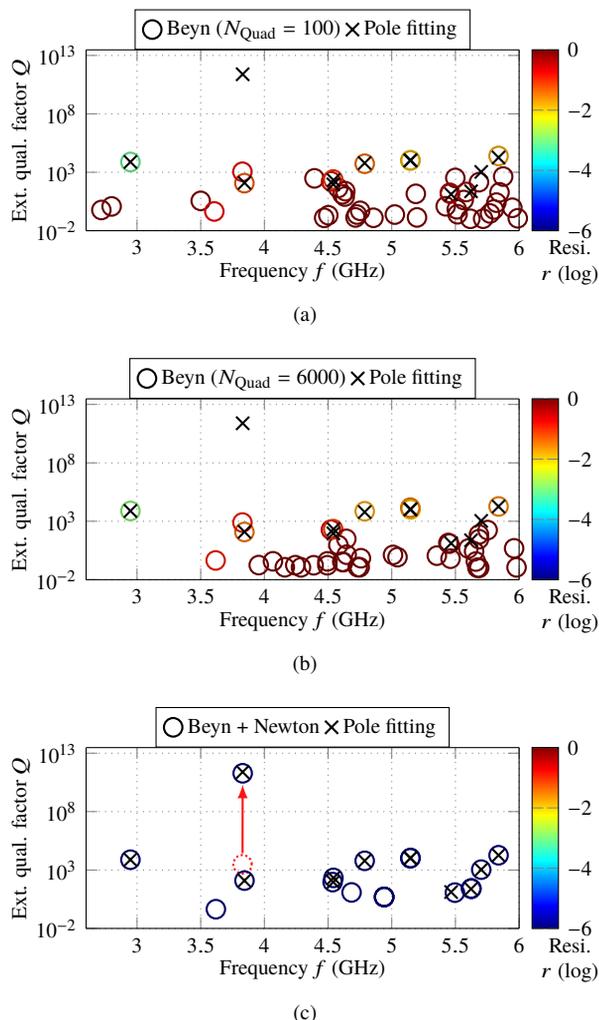


Figure 3: Eigensolutions of the minimalistic resonator with corresponding residuals  $r$  computed using Beyn's algorithm with different numbers of quadrature points: (a)  $N_{\text{Quad}} = 100$ , (b)  $N_{\text{Quad}} = 6000$ . The residual does not decrease for larger numbers of points, but can be further reduced by consecutive Newton iteration (c).

quadrature points  $N_{\text{Quad}}$ . The relative residual

$$r_n = \frac{\|T(\lambda_n)\mathbf{x}_n\|}{\|\mathbf{x}_n\|} \quad (11)$$

is used as an error measure. Even for very large  $N_{\text{Quad}}$ , the residual cannot be reduced below a certain limit. However, the convergence can drastically be improved by employing Newton's method. For comparison, reference solutions are computed from the scattering parameters of the structure using pole fitting [11, 59, 60].

The time consumption of different setups is depicted in Fig. 4. The Newton iteration on its own computes solutions individually and fast, but without guaranteeing completeness. Employing Beyn's algorithm with comparably low numbers of quadrature points and using its solution as initial values for the Newton iteration yields the optimum solution strategy for this structure.

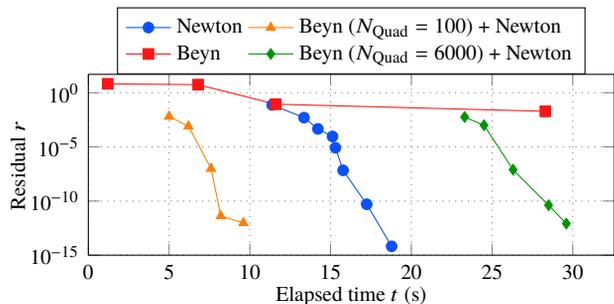


Figure 4: Time consumption of the different algorithms when solving the NLEVP of the minimalistic resonator.

### FLASH Third Harmonic Cavity

The SRF structure depicted in Fig. 5 is part of the Third Harmonic Module [61, 62] of the Free Electron Laser in Hamburg (FLASH). It consists of a nine-cell 3.9 GHz cavity, an input coupler and two HOM couplers. For the MOR by SSC, it is decomposed into three segments. Due to the three couplers in addition to the beam pipe, five external waveguide ports are defined, and the NLEVP is of order  $N = 780$ .

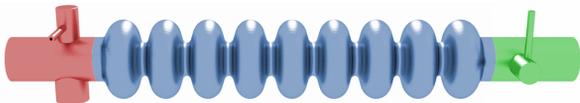


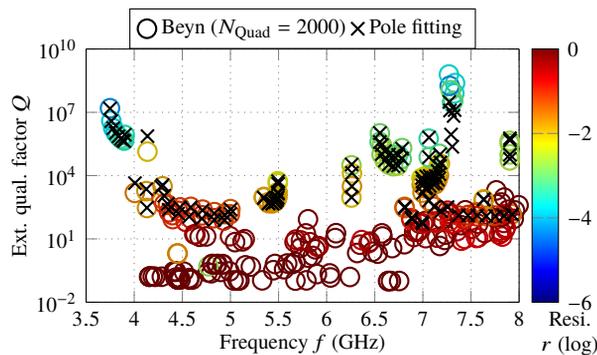
Figure 5: CAD model of the FLASH cavity. The colors indicate the chosen domain decomposition for SSC.

Figure 6(a) shows the results obtained by Beyn's algorithm with  $N_{\text{Quad}} = 2000$  quadrature points and the corresponding residuals. While for most higher- $Q$  modes an acceptable residual is achieved, the lower- $Q$  modes do not converge even for by orders of magnitude larger  $N_{\text{Quad}}$  ( $10^{-1} \leq r \leq 10^{-4}$ ). We can again improve the eigensolutions by individual Newton iteration. To achieve a residual below  $10^{-5}$  down to  $10^{-10}$ , 4 min are required by Beyn's algorithm, whereas the Newton iteration consumes an additional 6 min. The corresponding solutions are shown in Fig. 6(b). Especially in the more relevant cases of larger  $Q$  factors, i.e. potentially dangerous parasitic eigenmodes, a very good agreement is observed with the pole fitting solutions [11, 59, 60].

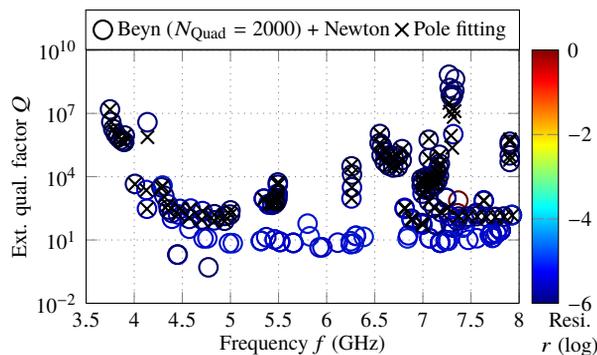
Next we compare the eigensolutions of the FLASH cavity found by Newton's method alone and when combined with Beyn's algorithm. Figure 7 depicts the computed frequencies and external  $Q$  factors and the electric field of a few selected modes is shown in Fig. 8. Newton's method solves the NLEVP with residuals  $10^{-5} \leq r \leq 10^{-10}$  in about 7 min. The combination of both algorithms finds a few additional solutions mainly in the low- $Q$  range, but is slightly slower (10 min). The difference in computational speed by a few minutes when solving the NLEVP is insignificant compared to the time consumption of the preceding MOR: the generation of the reduced SSM of the concatenated structure<sup>5</sup>

<sup>5</sup> The initial FIT mesh has roughly  $3 \cdot 10^6$  DOF. 200 to 300 3D eigenmodes are computed in each segment. For the external waveguide ports 25 port modes are considered in total.

takes roughly 7 h. However, the MOR by SSC approach is much faster than alternative methods.



(a)



(b)

Figure 6: Computed frequencies and  $Q$  factors of the FLASH cavity with corresponding residuals  $r$ . (a) Solutions computed by Beyn's algorithm with  $N_{\text{Quad}} = 2000$  quadrature points. Even for much larger  $N_{\text{Quad}}$ , the residuals do not improve significantly. (b) The convergence can be improved for individual eigenmodes by employing Newton iteration.

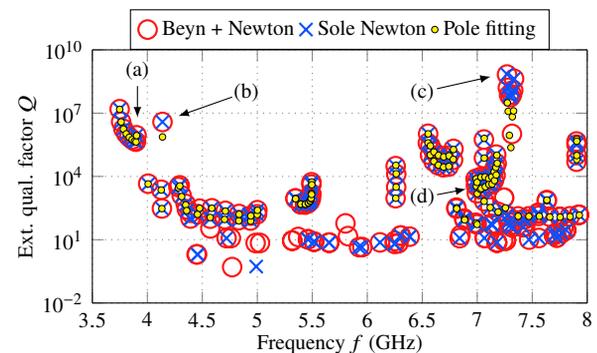


Figure 7: Eigensolutions of the FLASH cavity computed from the NLEVP by sole Newton iteration and by Beyn's algorithm with consecutive Newton iteration, with pole fitting solutions as reference. The markings (a)-(d) refer to the field distributions depicted in Fig. 8.

## CONCLUSION

The method developed in [11–13] allows the computation of resonant frequencies, external losses and field distribu-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

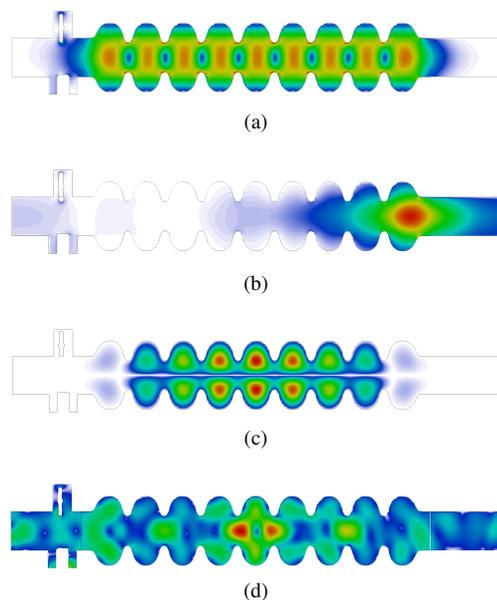


Figure 8: Electric fields of selected eigenmodes of the FLASH cavity: (a) Accelerating  $\pi$ -mode at 3.90 GHz. (b) Potentially dangerous trapped mode at 4.14 GHz with high  $Q$  and non-zero  $R/Q$ . (c) Trapped mode at 7.27 GHz and the highest  $Q$  within the analyzed spectrum. (d) Mode at 7.10 GHz with strong coupling to the cavity ports.

tion of complex SRF structures without extensive geometric simplification on workstation computers. The MOR and domain decomposition by SSC significantly reduce the computational effort. Using a perturbation approach and nonlinear boundary conditions, a NLEVP describing the external losses is obtained. The Newton iteration [14] can solve the arising NLEVP efficiently. This paper extends the approach by solving the NLEVP using Beyn's contour integral algorithm [15]. The convergence of this algorithm is observed to be limited for the type for NLEVP arising from SRF cavities. However, using it to compute initial eigenpairs for a consecutive Newton iteration shows promising results. More eigenmodes can be found in a comparable amount of time than by solely employing Newton's method. The approach has been demonstrated on an academic and a real-life SRF resonator.

## REFERENCES

- [1] H. Padamsee, J. Knobloch, and T. Hays, *RF Superconductivity: Science, Technology and Applications*. Wiley-WCH, 2008, ISBN: 9783527408429. doi: 10.1002/9783527627172.
- [2] S. U. de Silva and J. R. Delayen, "Wakefield analysis of superconducting rf-dipole cavities," in *Proceedings of the 28th Linear Accelerator Conference (LINAC'16), East Lansing, MI, USA, 2017*, pp. 206–208. doi: 10.18429/JACoW-LINAC2016-MOPLR031.
- [3] T. Weiland and R. Wanzenberg, "Wake fields and impedances," in *Frontiers of Particle Beams: Intensity Limitations*, Springer, 1992, pp. 39–79. doi: 10.1007/3-540-55250-2\_26.

- [4] U. van Rienen, "Higher order mode analysis of tapered discharged waveguides using the mode matching technique," *Particle Accelerators*, vol. 41, pp. 173–201, 1993.
- [5] T. Weiland, U. van Rienen, P. Hülsmann, W. F. O. Müller, and H. Klein, "Investigations of trapped higher order modes using a 36-cell test structure," *Physical Review Special Topics-Accelerators and Beams*, vol. 2, no. 4, p. 042001, 1999. doi: 10.1103/PhysRevSTAB.2.042001.
- [6] J. C. Maxwell, *A Treatise on Electricity and Magnetism*. Clarendon Press, 1873, ISBN: 9780511709333. doi: 10.1017/cbo9780511709333.
- [7] A. Bondeson, T. Rylander, and P. Ingelström, *Computational Electromagnetics*, ser. Texts in Applied Mathematics. Springer New York, 2005, ISBN: 9780387261584.
- [8] T. Weiland, "A discretization model for the solution of Maxwell's equations for six-component fields," *AE Ü*, vol. 31, no. 3, pp. 116–120, 1977.
- [9] U. van Rienen, *Numerical Methods in Computational Electrodynamics: Linear Systems in Practical Applications*, ser. Lecture Notes in Computational Science and Engineering. Springer Berlin Heidelberg, 2012, ISBN: 9783642568022.
- [10] T. Galek, J. Heller, T. Flisgen, and U. van Rienen, *Report on ssc calculations for bERLinPro module containing three srf cavities*, Internal Document, 2016.
- [11] J. D. Heller, "Numerical simulation of electromagnetic fields in complex multi-cavity superconducting radio frequency resonators," PhD thesis, Universität Rostock, 2018.
- [12] T. Flisgen, H.-W. Glock, and U. van Rienen, "Compact time-domain models of complex rf structures based on the real eigenmodes of segments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 6, pp. 2282–2294, 2013. doi: 10.1109/TMTT.2013.2260765.
- [13] T. Flisgen, "Compact state-space models for complex superconducting radio-frequency structures based on model order reduction and concatenation methods," available as CERN-ACC-2015-0145, PhD thesis, University of Rostock, 2015.
- [14] P. M. Anselone and L. B. Rall, "The solution of characteristic value-vector problems by Newton's method," *Numerische Mathematik*, vol. 11, no. 1, pp. 38–45, 1968. doi: 10.1007/BF02165469.
- [15] W.-J. Beyn, "An integral method for solving nonlinear eigenvalue problems," *Linear Algebra and Its Applications*, vol. 436, no. 10, pp. 3839–3863, 2012. doi: 10.1016/j.laa.2011.03.030.
- [16] T. Flisgen, A. Vélez, J. D. Heller, S. Zadeh, and U. van Rienen, "Computation of eigenmodes for the bessy vsr cavity chain by means of concatenation strategies," in *Proceedings of the 13th International Computational Accelerator Physics Conference (ICAP'18), Key West, FL, USA, To appear, 2018*.
- [17] T. Wittig, *Zur Reduzierung der Modellordnung in elektromagnetischen Feldsimulationen*. Cuvillier Verlag, 2004, ISBN: 9783865372086.
- [18] H.-W. Glock, K. Rothemund, and U. van Rienen, "CSC-A procedure for coupled S-parameter calculations," *IEEE Transactions on Magnetics*, vol. 38, no. 2, pp. 1173–1176, 2002. doi: 10.1109/20.996300.
- [19] E. Kühn, "A mode-matching method for solving field problems in waveguide and resonator circuits," *Archiv fuer Elektronik und Übertragungstechnik*, vol. 27, pp. 511–518, 1973.
- [20] R. Mittra, Y. Hou, and V. Jamnejad, "Analysis of open dielectric waveguides using mode-matching technique and vari-

- ational methods,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 28, no. 1, pp. 36–43, 1980. doi: 10.1109/TMTT.1980.1130003.
- [21] W. Wessel, T. Sieverding, and F. Arndt, “Mode-matching analysis of general waveguide multiport junctions,” in *Microwave Symposium Digest, 1999 IEEE MTT-S International*, IEEE, vol. 3, 1999, pp. 1273–1276. doi: 10.1109/MWSYM.1999.779619.
- [22] S. J. Cooke, “Reduced-order simulation of large accelerator structures,” *Physics of Plasmas*, vol. 15, no. 5, p. 056706, 2008. doi: 10.1063/1.2870086.
- [23] K. Rothemund, D. Hecht, U. van Rienen, *et al.*, “Calculation of rf properties of the third harmonic cavity,” in *Proceedings of the 22nd Linear Accelerator Conference (LINAC’04)*, Lübeck, Germany, vol. 14243, 2004, p. 4.
- [24] U. van Rienen, H.-W. Glock, D. Hecht, and K. Rothemund, “Computation of higher order modes in TESLA structures,” *International Journal of Applied Electromagnetics and Mechanics*, vol. 14, no. 1-4, pp. 237–242, 2001.
- [25] K. Rothemund, H.-W. Glock, and U. van Rienen, “Eigenmode calculation of complex rf-structures using s-parameters,” *IEEE Transactions on Magnetics*, vol. 36, no. 4, pp. 1501–1503, 2000. doi: 10.1109/20.877722.
- [26] I. R. R. Shinton and R. M. Jones, “Large scale linac simulations using a globalised scattering matrix approach,” in *Proceedings of the 11th European Particle Accelerator Conference (EPAC’08)*, Genoa, Italy, 2008.
- [27] I. R. R. Shinton, R. M. Jones, Z. Li, and P. Zhang, “Simulations of higher order modes in the ACC39 module of FLASH,” in *3rd International Particle Accelerator Conference (IPAC’12)*, New Orleans, LA, USA, 2012.
- [28] R. M. Jones, L. Shi, and N. Baboi, “Simulation of electromagnetic scattering through the E-XFEL third harmonic cavity module,” in *Proceedings of the 7th International Particle Accelerator Conference (IPAC’16)*, Busan, Korea, 2016. doi: 10.18429/JACoW-IPAC2016-WEPOY007.
- [29] M. Liepe, “Superconducting multicell cavities for linear colliders,” PhD thesis, Universität Hamburg, 2001.
- [30] T. Flisgen, J. Heller, and U. van Rienen, “Time-domain absorbing boundary terminations for waveguide ports based on state-space models,” *IEEE Transactions on Magnetics*, vol. 50, no. 2, pp. 145–148, 2014. doi: 10.1109/TMAG.2013.2283065.
- [31] T.-H. Loh and C. Mias, “Implementation of an exact modal absorbing boundary termination condition for the application of the finite-element time-domain technique to discontinuity problems in closed homogeneous waveguides,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 882–888, 2004. doi: 10.1109/TMTT.2004.823559.
- [32] S. Börm and C. Mehl, *Numerical Methods for Eigenvalue Problems*, ser. De Gruyter Graduate Lectures. Walter de Gruyter, 2012, ISBN: 9783110250336. doi: 10.1515/9783110250374.
- [33] A. Ruhe, “Algorithms for the nonlinear eigenvalue problem,” *SIAM Journal on Numerical Analysis*, vol. 10, no. 4, pp. 674–689, 1973. doi: 10.1137/0710059.
- [34] I. N. Bronshtein, K. A. Semendyayev, G. Musiol, and H. Mühlig, *Handbook of Mathematics*. Springer Berlin Heidelberg, 2015, ISBN: 9783662462218.
- [35] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, *Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide*. SIAM, 2000, ISBN: 9780898714715. doi: 10.1137/1.9780898719581.
- [36] H. Voss, “Nonlinear eigenvalue problems,” in *Handbook of Linear Algebra*, vol. 164, Chapman and Hall/CRC Boca Raton, FL, 2013.
- [37] V. Mehrmann and H. Voss, “Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods,” *GAMM-Mitteilungen*, vol. 27, no. 2, pp. 121–152, 2004. doi: 10.1002/gamm.201490007.
- [38] S. Güttel and F. Tisseur, “The nonlinear eigenvalue problem,” *Acta Numerica*, vol. 26, pp. 1–94, 2017. doi: 10.1017/S0962492917000034.
- [39] C. Effenberger, “Robust solution methods for nonlinear eigenvalue problems,” PhD thesis, École polytechnique fédérale de Lausanne, 2013.
- [40] T. Sakurai and H. Sugiura, “A projection method for generalized eigenvalue problems using numerical integration,” *Journal of Computational and Applied Mathematics*, vol. 159, no. 1, pp. 119–128, 2003. doi: 10.1016/S0377-0427(03)00565-X.
- [41] V. N. Kublanovskaja, “On an application of Newton’s method to the determination of eigenvalues of  $\lambda$ -matrices,” *Doklady Akademii Nauk SSSR*, vol. 188, pp. 1004–1005, 1969.
- [42] R. J. Duffin, “A minimax theory for overdamped networks,” *Journal of Rational Mechanics and Analysis*, vol. 4, pp. 221–233, 1955. doi: 10.1512/iumj.1955.4.54007.
- [43] K. P. Hadelér, “Mehrparametrische und nichtlineare Eigenwertaufgaben,” *Archive for Rational Mechanics and Analysis*, vol. 27, no. 4, pp. 306–328, 1967. doi: 10.1007/bf00281717.
- [44] L. B. Rall, *Computational Solution of Nonlinear Operator Equations*. Wiley New York, 1969, ISBN: 9783110920291. doi: 10.1515/9783110920291.
- [45] M. V. Keldyš, “On the characteristic values and characteristic functions of certain classes of non-selfadjoint equations,” *Doklady Akademii Nauk SSSR*, vol. 77, pp. 11–14, 1951.
- [46] R. Mennicken and M. Möller, *Non-Self-Adjoint Boundary Eigenvalue Problems*, 1st ed., ser. North Holland Mathematics Studies. North Holland, 2003, vol. 192, ISBN: 9780444514479.
- [47] L. N. Trefethen and J. A. C. Weideman, “The exponentially convergent trapezoidal rule,” *SIAM Review*, vol. 56, no. 3, pp. 385–458, 2014. doi: 10.1137/130932132.
- [48] H. W. Pommerenke, “Untersuchung von konturintegralmethoden zur lokal vollständigen lösung nichtlinearer eigenwertprobleme, die auf den maxwell-gleichungen basieren,” Master’s thesis, Universität Rostock, 2017.
- [49] V. Hernandez, J. E. Roman, and V. Vidal, “SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems,” *ACM Trans. Math. Software*, vol. 31, no. 3, pp. 351–362, 2005. doi: 10.1145/1089014.1089019.
- [50] Computer Simulation Technology AG, *CST Studio Suite 2017*.
- [51] T. M. Inc., *Matlab version r2015b*.
- [52] G. van Rossum and F. L. Drake Jr., *Python Reference Manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [53] E. Anderson *et al.*, *LAPACK Users’ Guide*, 3rd ed. Society for Industrial and Applied Mathematics, 1999, ISBN: 0898714478. doi: 10.1137/1.9780898719604.
- [54] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK users’ guide: Solution of large scale eigenvalue problems with implicitly restarted arnoldi methods*, Computer Based Learning Unit, University of Leeds, 1997. doi: 10.1137/1.

- 9780898719628.
- [55] S. van der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy array: A structure for efficient numerical computation,” *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011. doi: 10.1109/MCSE.2011.37.
- [56] E. Jones, T. Oliphant, P. Peterson, *et al.*, *SciPy: Open source scientific tools for Python*, 2001. <http://www.scipy.org>
- [57] A. Henderson, J. Ahrens, C. Law, *et al.*, *The ParaView Guide*. Kitware Clifton Park, NY, 2004, ISBN: 9781930934306.
- [58] B. Krietenstein, P. Thoma, R. Schuhmann, and T. Weiland, “The perfect boundary approximation technique facing the big challenge of high precision computation,” in *Proceedings of the 19th Linear Accelerator Conference (LINAC’98)*, Chicago, IL, USA, 1998.
- [59] T. Galek, T. Flisgen, U. van Rienen, B. Riemann, and A. Neumann, “Traveling poles elimination scheme and calculations of external quality factors of HOMs in SC cavities,” in *Proceedings of the 11th International Computational Accelerator Physics Conference (ICAP’12)*, Rostock-Warnemünde, Germany, 2012, pp. 152–154.
- [60] B. Gustavsen, “Improving the pole relocating properties of vector fitting,” *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1587–1592, 2006. doi: 10.1109/TPWRD.2005.860281.
- [61] J. Sekutowicz, R. Wanzenberg, W. F. O. Müller, and T. Weiland, “A design of a 3rd harmonic cavity for the TTF 2 photoinjector,” *DESY-TESLA-FEL*, vol. 5, p. 2002, 2002.
- [62] N. Solyak *et al.*, “Development of the third harmonic sc cavity at fermilab,” in *Proceedings of the Particle Accelerator Conference, Portland, OR, USA*, IEEE, vol. 2, 2003, pp. 1213–1215. doi: 10.1109/PAC.2003.1289656.

# NONLINEAR OPTICS AT UMER: LESSONS LEARNED IN SIMULATION

K. Ruisard\*, Oak Ridge National Laboratory, Oak Ridge, Tennessee

B. Beaudoin, I. Haber, D. Matthew, and T. Koeth,

Institute for Research in Electronics and Applied Physics, College Park, Maryland

## Abstract

Design of accelerator lattices with nonlinear optics to suppress transverse resonances is a novel approach and may be crucial for enabling low-loss high-intensity beam transport. Large amplitude-dependent tune spreads, driven by nonlinear field inserts, damp resonant response to driving terms. This presentation will focus on simulations of the UMER lattice operated as a quasi-integrable system (one invariant of transverse motion) with a single strong octupole insert. We will discuss the evolution of simulation models, including the observation of losses associated with the original operating point near a fourth-order resonance. Other operating points farther from this resonance are considered and shown to be more promising.

## INTRODUCTION

Nonlinear integrable optics (NLIO) is a novel implementation of focusing optics for accelerator rings. Proposed by Danilov and Nagaitsev [1], this technique is expected to mitigate resonant beam losses in circular machines. This is of particular interest at the “intensity frontier,” where even low-level losses can threaten machine components and personnel safety.

Nonlinear terms in the transverse focusing potential have long been known to counteract resonant interactions in rings. In the presence of nonlinear forces, the coupling of regular driving terms to particle orbits is reduced and collective motions such as envelope modes decohere. The most well-known example is octupole-induced Landau damping, in which an octupole-induced tune shift in the particle distribution can damp transverse collective instability [2]. Simulation studies of NLIO systems shows fast decoherence of envelope modes, which are a known mechanism for halo formation [3].

In general, introducing nonlinearities reduces dynamic aperture due to chaotic orbits near resonance overlap, which has previously restricted the use of nonlinear correctors to weak perturbations of the linear Hamiltonian. The breakthrough of NLIO is the identification of a family of highly nonlinear, physically-realizable magnetic potentials in which transverse particle orbits conserve coupled, quadratic invariants of motion that are distinct from the Courant-Snyder invariants.

This paper describes progress towards an experimental demonstration of quasi-integrable optics (QIO) at the University of Maryland Electron Ring (UMER). This variation on the NLIO theory utilizes an octupole potential (rather than the fully-integrable fields discussed in reference [1]) that

allows one invariant of transverse motion: the Hamiltonian in normalized coordinates<sup>1</sup>:

$$H_N = \frac{1}{2} \left( p_{x,N}^2 + p_{y,N}^2 + x_N^2 + y_N^2 \right) + \frac{\kappa}{4} \left( x_N^4 + y_N^4 - 6y_N^2 x_N^2 \right). \quad (1)$$

Although motion is not fully integrable (only one invariant for 2D motion), the invariant corresponds with particle amplitude resulting in chaotic but bounded motion [3].

This proceedings discusses simulation results for the QIO as designed for UMER. We probe dynamics within the octupole insert “as designed” and show clear improvement for one insert configuration over another. We also compare transport properties across a range of tune operating points while seeking to maximize octupole-induced tune spread and preserve stable particle orbits.

## NONLINEAR OPTICS PROGRAM AT UMER

UMER is a scaled, 10 keV ( $\beta = 0.195$ ) electron ring designed for the study of high-intensity beam dynamics relevant to higher-energy ion rings. Different space charge densities are selected by aperturing the beam near the source, in the range  $\nu/\nu_0 = 0.85 \rightarrow 0.14$  for nominal UMER tune 6.7 (incoherent shifts  $\Delta\nu = 0.3 \rightarrow 5.7$ ) [4, 5].

A proof-of-principle QIO experiment has been designed for UMER. The experiment layout, shown in Figure 1, include a single octupole insert element. This effort uses existing UMER quadrupole optics to meet requirements for linear lattice focusing, which are outlined below. The RMS envelope solution for the linear optics as designed is shown in Figure 2. Details of implementing this solution in the UMER ring are discussed in references [6, 7]. A custom-designed octupole insertion, consisting of seven independently-wired octupole PCBs, has been fabricated and is capable of meeting requirement 2 to within RMS error of 2%. Design of the octupole element is covered in reference [8].

For initial tests of the QIO concept, we desire beams with lower space-charge tune shift than the typical UMER range, as the NLIO/QIO theory is based on single-particle dynamics. An ultra-low-current, high emittance beam was characterized for use in initial experiments. A beam with current 10 to 100  $\mu\text{A}$  is generated by operating the UMER triode electron gun in voltage amplification mode. This beam has low tune shift due to its large emittance; quadrupole scan emittance measurement at 40  $\mu\text{A}$  output current returns

<sup>1</sup>  $x_N \equiv \frac{x}{\sqrt{\beta_x(s)}}, p_{x,N} \equiv p_x \sqrt{\beta_x(s)} + \frac{\alpha_x x}{\sqrt{\beta_x(s)}}$

\* ruisardkj@ornl.gov

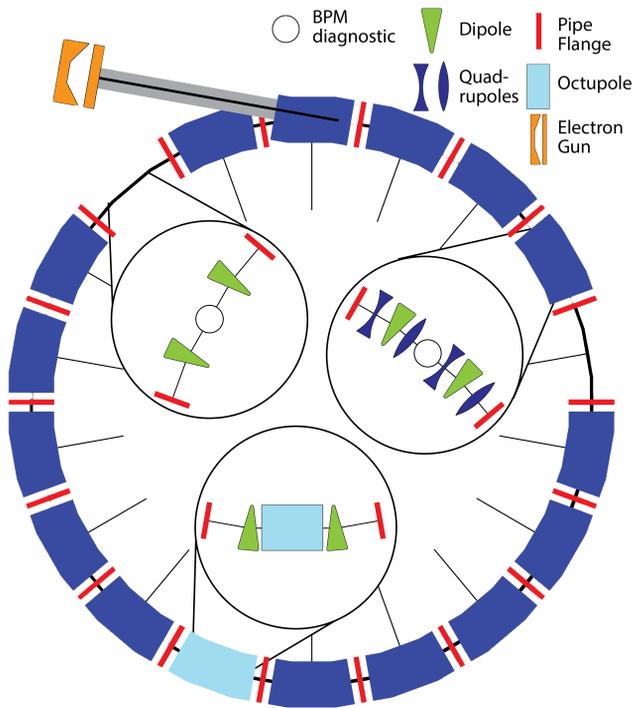


Figure 1: Diagram of UMER QIO lattice, with insets showing detail for standard 20° section (Q-D-Q-Q-D-Q) and octupole section (O-D-OCTU-D-O). Spokes indicate diagnostic locations and red bands indicate flange locations.

$\epsilon_x = 300$  mm-mrad,  $\epsilon_y = 100$  mm-mrad (unnorm.,  $4 \times \text{RMS}$ ). For the simulations discussed here, a 100  $\mu\text{m}$ , 60  $\mu\text{A}$  semi-Gaussian beam is used as the test case.

### Choice of Operating Point

The building blocks of a NLIO/QIO lattice are nonlinear insertion elements embedded within a linear-focusing (quadrupole) lattice, which is required for transverse confinement. The conditions for integrability are:

1. Beam envelope is round through the nonlinear insertion ( $\beta_x = \beta_y$ ); this is done by forming a round waist  $\beta_*$  in the nonlinear insert.
2. Nonlinear potential is scaled to cancel  $s$ -dependence of  $H_N$  (for QIO,  $G_3(s) \propto \beta^{-3}(s)$ ); this is required for normalized orbits to move through a constant potential.
3. Linear-focusing transport between insertions must have phase advance  $n\pi$ ; this is required for particle motion to be quasi-continuous in the nonlinear potential.

Choice of lattice operating tune is constrained mainly by the minimum achievable waist size,  $\beta_*$ , given the existing quadrupole optics.  $\beta_* = 0.3$  was identified as a “safe” waist size, with  $2 \times$  edge distance to the pipe wall at the largest transverse beam extent.

Original plans for nonlinear UMER included a “segmented” 64-cm octupole channel over a UMER 20° section,

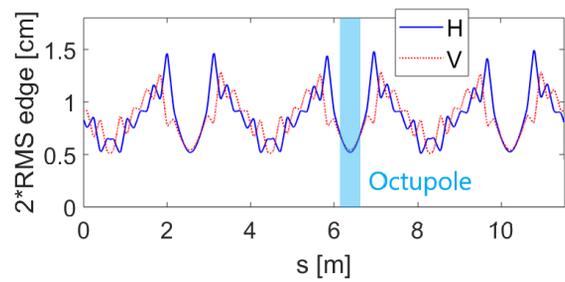


Figure 2: Periodic KV envelope solution for 100 mm-mrad, 60  $\mu\text{A}$  beam at  $\nu_x = \nu_y = 3.26$ . Injection is at  $s = 0$  plane.

which encompasses two 10° bends. This lattice would have fractional tune 0.263 to meet the quasi-integrable condition; this fractional tune also corresponds with the maximum theoretical tune spread. Restricting the channel length to 25 cm (the longest contiguous element that can be placed in UMER) reduces the maximum achievable tune spread from 0.26 to 0.13.

As seen in Figure 3, the largest accessible tune operating point that can be achieved for a large-emittance beam and 25-cm insert with existing optics is  $\sim 0.35$ . Higher tunes require either a smaller waist size  $\beta_*$ , which will lead to scraping, or use of multiple insert regions, which is not considered at this time but may be possible as an extension of this work.

### Simulation Models

In order to isolate dynamics within the nonlinear potential, PIC simulations with the WARP code [9] are performed on a “simple model” of the QIO system. This simple model consists only of the octupole element and an ideal, thin lens, symmetric focusing kick as a proxy for the linear focusing sections. Effects accruing over the length of the lattice (such as the space charge force) are excluded in this model. We also examine PIC simulation over full QIO experiment configuration included linear optics modeled as hard-edged quadrupoles. In the ring model used here, dipoles are excluded for simplicity. The UMER dipoles introduce a significant linear focusing component due to fringe fields and a sextupole term in the PCB dipole circuits. For the insert, we alternately use an ideal octupole potential or a gridded field element representative of the octupole channel as de-

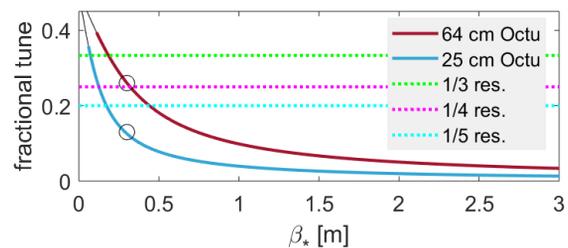
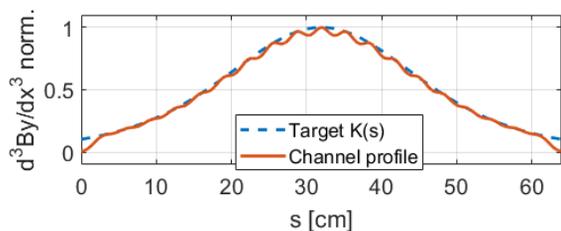
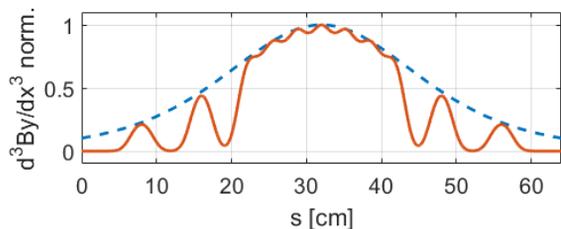


Figure 3: Accessible tune operating points as a function of  $\beta_*$  waist size in a single-channel UMER QIO lattice. Circles indicate nominal operating points at  $\beta_* = 0.3$  m.

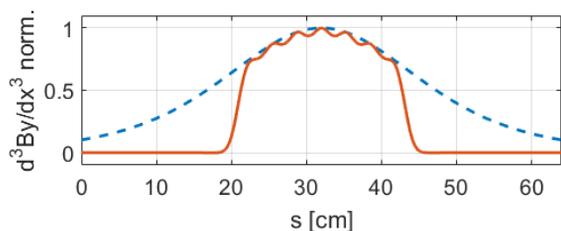
Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.



(a) Octupole channel assuming full population of 20° arc.



(b) Octupole channel with clearance for mounts, including octupole circuits co-housed with dipoles.



(c) Octupole channel restricted to straight section between dipoles.

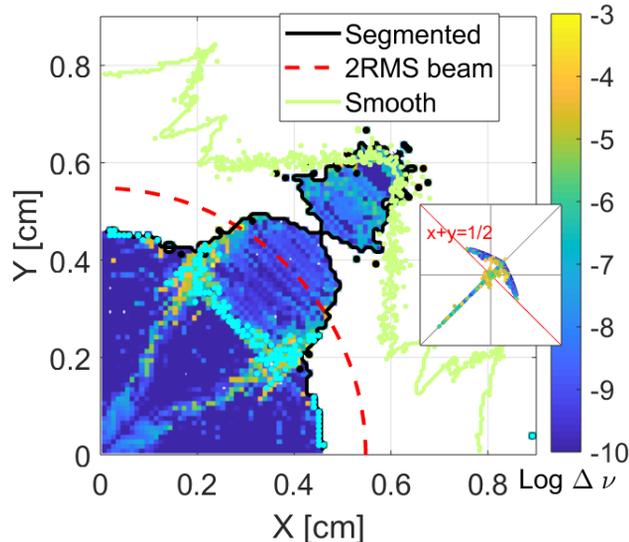
Figure 4: Octupole channel profile configurations. While 4(a) is ideal, only 4(b) and 4(b) are realizable in the experiment.

signed, which is generated by the Biot-Savart solution for the PCB circuit. 40,000 macro-particles are used in the space-charge model, while the frequency map is sampled using a zero-charge, gridded “witness distribution.”

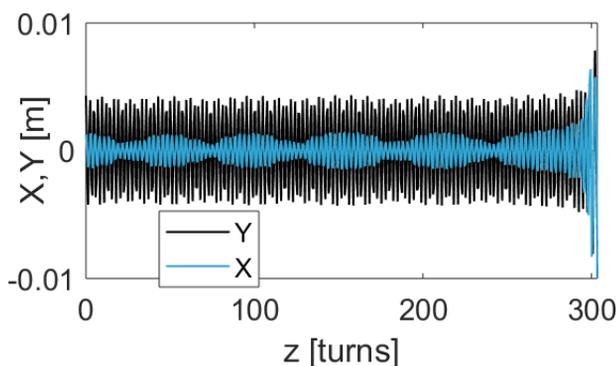
## CHOOSING OCTUPOLE INSERT CONFIGURATION

Adding an octupole insertion to UMER can be done by modifying a single, modular 20° (64 cm) arc. As mentioned, initial designs assume a fully-occupied but segmented arc, in which octupoles are placed at every available location with gaps for mechanical clearance. The desired octupole profile  $d^3B_y(s)/dx^3 \propto \beta^{-3}(s)$  (Requirement 2) is painted in along the “octupole section” using short (4.65 cm) PCB magnets. Three configurations are shown in Fig. 4. The scheme shown in Fig. 4(a) is purely hypothetical, as the presence of two bending dipoles in the arc restrict octupole PCB placement.

The three cases shown in Fig. 4 were compared using the dipole-free simple model with thin-lens focusing and gridded fields from Biot-Savart solution of the octupole circuits (the added complication of introducing bends is ignored in this analysis). The resulting frequency map for



(a) Configuration space plot of aperture and resonant structure.



(b) X and Y trajectory of particle at edge of stable boundary. Particle initial condition is indicated on Fig. 5(a).

Figure 5: Frequency map of simple QIO lattice at peak field 50 T/m<sup>3</sup> ( $\kappa = 3984$ ) and fractional tune 0.26 for 1024 passes. Light green contour/footprint indicates the smooth, 64 cm channel (Fig. 4(a)).

1024 passes with peak field 50 T/m<sup>3</sup> ( $\kappa = 3984$ ) is shown in Fig. 5. Very poor performance is seen in the “segmented channel” (Fig. 4(b)) as compared to the other configurations shown, even before considering the effect of bends.

Particles appear to be lost along the  $\nu_x + \nu_y = 1/2$  coupling resonance. Figure 5(b) shows a typical orbit of a particle near the stable boundary, where the small-amplitude plane quickly grows after some length of time. As a result of this poor performance seen here, plans for a 64-cm “segmented” octupole insertion are discarded, and the design shifts to a 25-cm, 10° insert.

## BEAM STABILITY IN QIO UMER RING

Simulation of the proposed experiment over the full ring is done both at the original operating point based on the assumption of a 64 cm octupole insert (fractional tune 0.263) and at a new operating point adjusted for a shorter insert

(fractional tune 0.126). As stated above, dipole effects are ignored in these test cases and all elements are hard-edged. For the proposed experiments, a peak octupole gradient of  $50 \text{ T/m}^3$  places the fixed point at  $3.2\sigma$  (for a beam with  $100 \text{ mm-mrad } 4\times\text{RMS}$  emittance). This requires only  $0.97 \text{ A}$  in the central octupole circuit, well within the safety limit of the UMER octupoles.

### PIC Simulation at Nominal 3.26 Operating Point

The first full ring simulations were run at the original operating point with fractional tune 0.26. Figure 6 shows results from a  $60 \mu\text{A}$  beam at the initial design tune of  $\nu_{0,x} = \nu_{0,y} = 3.26$ . Large loss of stable aperture is seen when compared to the 64-cm simple model. This tune is also very near the fourth order  $\nu = 0.25$  resonance, which is strongly driven by the octupole term. Although not a feature seen in simple model calculations, the beam distribution shown in Fig. 6(c) depicts clear fourth-order resonant structure. However, the boundary of stable orbits corresponds with the  $\nu_y + \nu_x = 1/2$  coupling resonance rather than  $\nu = 0.25$ . The reduced stability may be due to overlap between these two conditions.

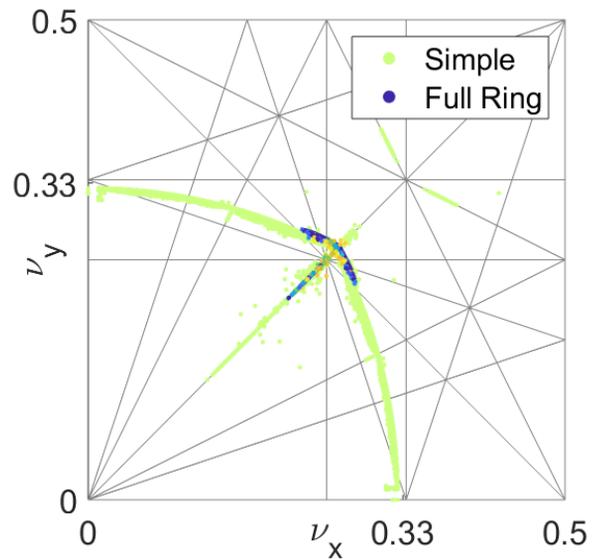
In the case shown here, the octupole insert length is limited to 25-cm to correspond with the channel as designed. This places the operating point quite far from the quasi-integrable condition  $\psi = n\pi$ . This is reflected in the large fluctuations in the quantity  $H_N$  (Eq. (1)), which on average over all stable orbits exhibits 8.02% variation.

### PIC Simulation at $\nu = 0.13$

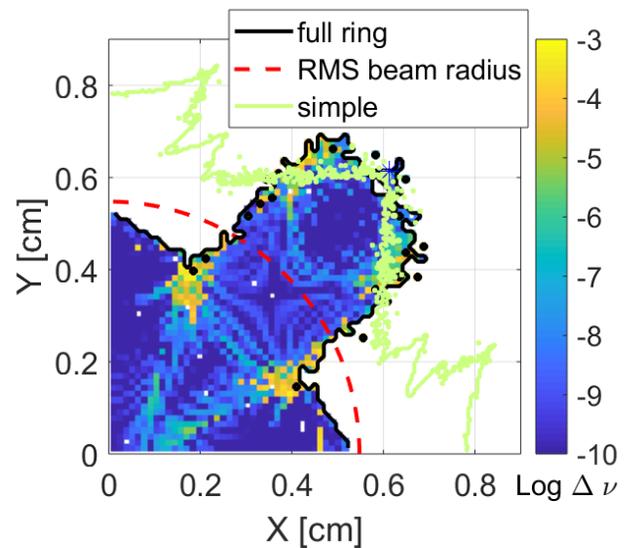
In comparison, adjusting the linear optics to shift the tune towards the quasi-integrable condition at  $\nu_x = \nu_y = 3.13$  yields much better transport. The envelope solution is similar to that shown in Fig. 2 so the adjusted case is not reproduced here. As shown in Fig. 7, the dynamics agree well with simple model predictions. The enhanced stability is reflected in better conservation of  $H_N$ . A less than 1% variation is observed for all particles in the zero-charge limit. An average 4% variation is seen when the  $60 \mu\text{A}$  current is included, but this is mainly due to contribution from low-amplitude, highly-depressed particles (see discussion in conclusion section). The main drawback is that operating at lower tune puts a limit on the maximum stable tune spread, which is reduced from  $\sim 0.26$  to  $\sim 0.13$ .

### A Comment on Space Charge

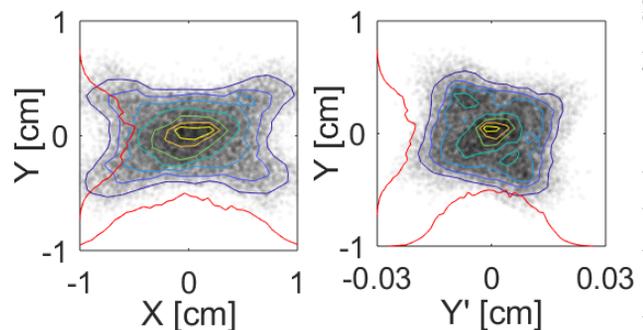
Simulations were run with the low-charge, high-emittance  $60 \mu\text{A}$  beam, which has a predicted incoherent tune spread of 0.005. Compared to the “zero-charge” case, the effect of space charge on the stable aperture and induced tune spread is small. However, variation of the quantity  $H_N$  increased in the presence of space charge. Figure 8 shows the dependence of  $H_N$  variation on initial particle radius  $r$  (comparable to particle amplitude). While there is no trend in the “zero-charge” case, when the  $60 \mu\text{A}$  current is considered variation increases, with the largest variations occurring at small amplitudes (the most shielded particles).



(a) Tune footprint with up to fourth order resonance lines indicated.



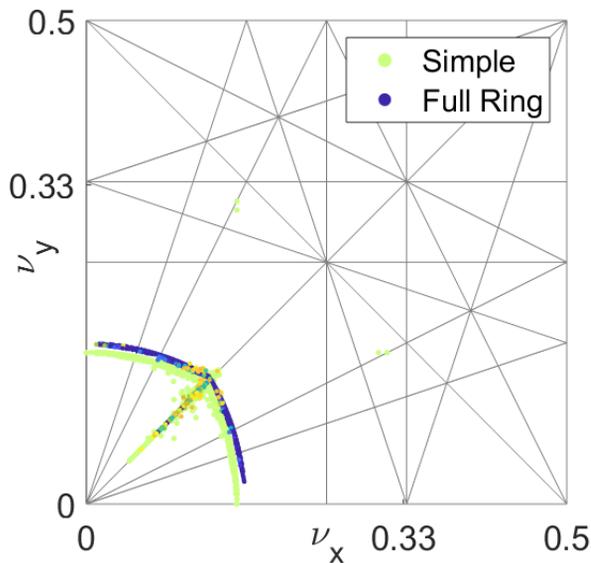
(b) Configuration space plot of aperture and resonant structure.



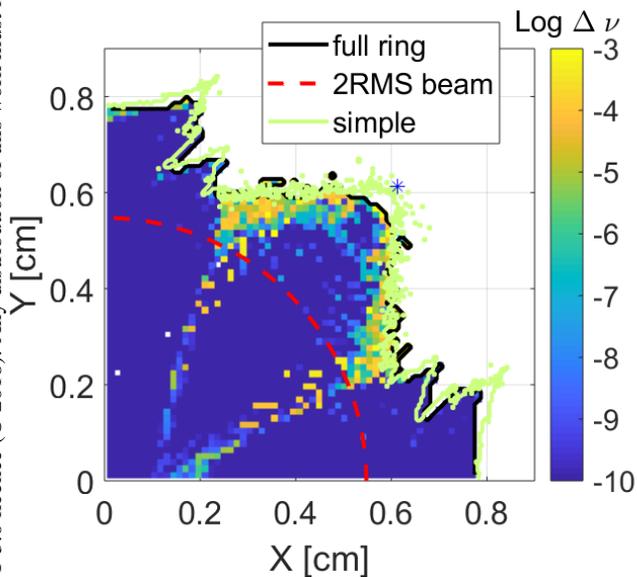
(c) Snapshot of beam distribution at turn 128 showing clear fourth-order resonant structure.

Figure 6: Frequency map of QIO lattice at peak field  $50 \text{ T/m}^3$  ( $\kappa = 2390$ ) and fractional tune 0.26 at turn 384. Best-case simple model is compared with hard-edged model of full ring.

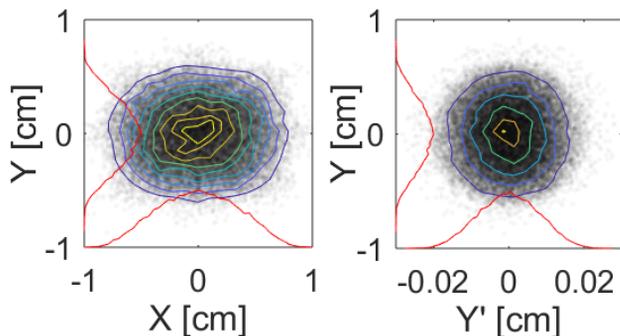
Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.



(a) Tune footprint with up to fourth order resonance lines indicated.



(b) Configuration space plot of aperture and resonant structure.



(c) Snapshot of beam distribution at turn 264.

Figure 7: Frequency map of QIO lattice at peak field  $50 \text{ T/m}^3$  ( $\kappa = 3984$ ) and fractional tune 0.13 at turn 896.

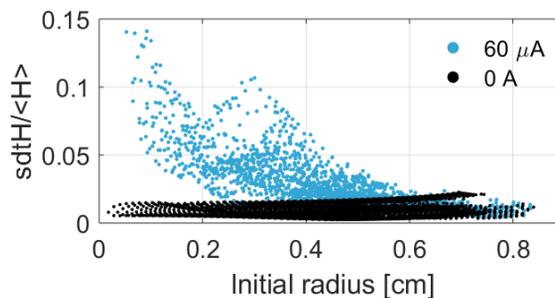
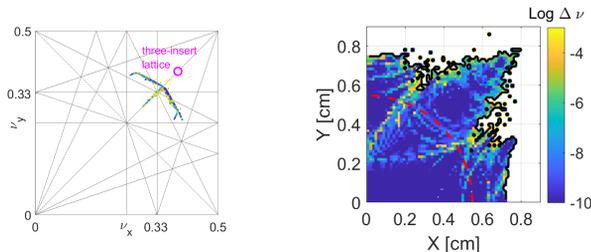


Figure 8: Dependence of  $H_N$  variation on initial orbit radius  $r$  after 264 turns in the  $\nu = 3.13$  QIO lattice.



(a) Tune footprint with up to fourth order resonance lines indicated. (b) Configuration space plot of aperture and resonant structure.

Figure 9: Frequency map of QIO lattice at peak field  $50 \text{ T/m}^3$  ( $\kappa = 3984$ ) and fractional tune 0.35 at turn 384. Asymmetry is due to tune error  $\nu_x - \nu_y = 0.01$ .

However, these orbits appear to remain bounded. For the work shown here, linear optics were optimized for quasi-integrable transport of an equivalent KV beam (in other words, optimized for the maximally depressed particle orbit).

### Alternative Operating Points for Increased Tune Spread

Two strategies for increasing the octupole-induced tune spread are considered. In the first, the fractional tune is increased by adjusting the focusing of the linear optics. Second, fractional tune is increased by reducing waist size  $\beta_*$ . Results shown in this section are calculated in the zero-charge limit.

Figure 9 shows transport results for a lattice at tune  $\nu = 3.35$ , which is achieved by adjusting linear optics without changing envelope properties in the octupole section. Dynamics appear very similar to the case as  $\nu = 3.13$ , with a slight asymmetry due to unequal tunes in this solution ( $\nu_x - \nu_y = 0.01$ ). The variation of  $H_N$  is larger in this case, 1.3%, but this is expected as the QI condition is not met. However, this design is near the tune  $\nu = 3.39$  where installing three insertions (one at each waist) satisfies the QI condition. Multi-insert configurations may be considered to enhance the tune spread near this operating point in future work.

Tune for the single-channel lattice can also be increased by decreasing waist size  $\beta_*$ . This can be achieved for the 25-

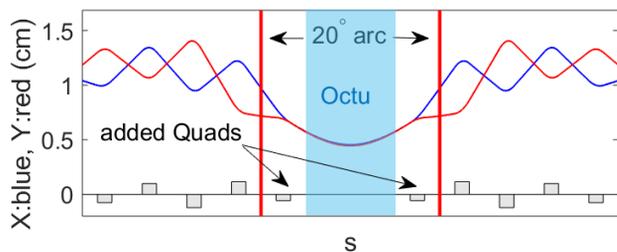
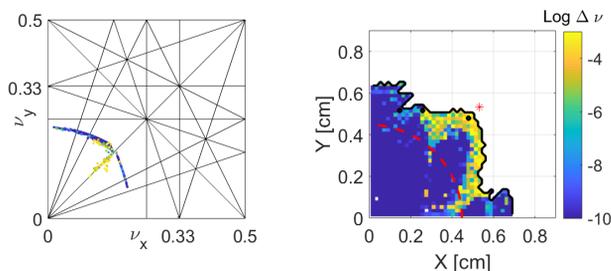


Figure 10: Alternative lattice solution included additional quads placed in same 20° section as octupole insert.



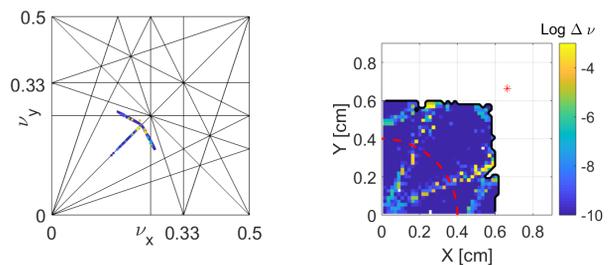
(a) Tune footprint with up to fourth order resonance lines indicated. (b) Configuration space plot of aperture and resonant structure.

Figure 11: Frequency map of QIO lattice at peak field 150 T/m<sup>3</sup> ( $\kappa = 3514$ ) and fractional tune 0.18 at turn 512.

cm insertion by placing quadrupoles nearer the insertion as shown in Fig. 10 (In the previous cases, these locations were unoccupied drifts). Figure 11 shows the case for  $\beta_* = 0.2$  m,  $\nu = 3.18$  and Fig. 12 shows  $\beta_* = 0.16$  m,  $\nu = 3.23$ . In both cases the peak octupole field must be increased to 150 T/m<sup>3</sup> to maintain a large tune spread in the beam distribution. This requires 2.92 A in the octupole PCB, which is near the safety limit for UMER magnets. 6.4 A is required to achieve  $\kappa$  equal to the cases shown above, which is possible with water cooling to protect the magnet circuit.

## SUMMARY AND CONCLUSION

In this investigation, multiple operating points are explored for use in the UMER QIO experiment. Results for



(a) Tune footprint with up to fourth order resonance lines indicated. (b) Configuration space plot of aperture and resonant structure.

Figure 12: Frequency map of QIO lattice at peak field 150 T/m<sup>3</sup> ( $\kappa = 1813$ ) and fractional tune 0.22 at turn 512.

all cases are summarized in Table 1. Shifting of operating points is motivated in part by a study of various octupole channel configurations. The original design included a “segmented” 64-cm octupole insert, in which a long, straight insert is flanked by single octupole magnets as space allows. This was found to be very detrimental to lattice stability, and plans were shifted to focus on a single, 25-cm straight insert.

In full ring simulations, the original operating point of  $\nu = 3.26$  is found to have poor transport, due both to proximity to the driven fourth-orders resonance and violation of the QI condition. A more promising result is found at  $\nu = 3.13$  for the single-insert configuration. However, the maximum possible tune spread is halved as a result. To remedy this, we explore the possibility of increasing the operating tune (and therefore the tune spread) by adjusting optics outside and inside the insertion region. Increasing tune advance through the octupole by decreasing waist size  $\beta_*$  is limited by safety limits on octupole circuit heat load. However, adjusting for a higher tune with three insertions is promising, and may warrant further investigation.

Table 1: Predicted performance of two tune operating points for PIC simulation of QIO ring with hard-edged elements and peak octupole gradient 50 T/m<sup>3</sup> ( $\kappa = 3948$ ). Conservation of invariant  $\langle H_N \rangle / H_N$  is taken as average over all stable particle orbits within the 2×RMS beam radius.

$\nu$	$\beta_*$ [m]	max/RMS		$\langle \frac{H_N}{H_N} \rangle$ [%]
		$\sigma_\nu$	eff. $r_{max}$	
0.26	0.30	0.017/0.006	0.48 cm	8.02
0.13	0.30	0.064/0.019	0.67 cm	0.92
0.35	0.30	0.062/0.019	0.65 cm	1.32
0.22 *†	0.16	0.056/0.019	0.58 cm	3.64
0.18 *‡	0.20	0.063/0.021	0.54 cm	14.47

\* Peak field 150 T/m<sup>3</sup>

†  $\kappa = 1813$

‡  $\kappa = 3514$

## ACKNOWLEDGMENTS

Many thanks to members of the UMER group, whose comments and support have helped this project along. These include Rami Kishek, Levon Dovlatyan, Moiz Siddiqi, Dave Sutter and Eric Montgomery. Special thanks especially to Santiago Bernal for his work on low-current UMER beams [5] and Heidi B. Komkov for her contribution to characterizing the octupole magnets and constructing the octupole insert [8, 10]). Thanks also to the FNAL IOTA team for leading the way in this topic [11]. Also thanks to the SNS Accelerator Physics group, for supporting the presentation of this research at ICAP’18.

Funding for this project was provided through DOE-HEP Award DESC0010301, NSF Award PHY1414681 and the NSF GRFP program. This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy.

## REFERENCES

- [1] V. Danilov and S. Nagaitsev, "Nonlinear accelerator lattices with one and two analytic invariants," *PRSTAB*, vol. 13, no. 8, p. 084002, Aug. 2010.
- [2] A. W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators*. John Wiley & Sons, 1993.
- [3] S. D. Webb, D. L. Bruhwiler, A. Valishev, S. N. Nagaitsev, and V. V. Danilov, "Chromatic and Dispersive Effects in Nonlinear Integrable Optics," in *Proceedings of HB2014*, East-Lansing, MI, 2014, p. 10.
- [4] R. A. Kishek *et al.*, "The University of Maryland Electron Ring Program," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 733, pp. 233–237, 2014.
- [5] S. Bernal *et al.*, "Ultra-low current beams in UMER to model space-charge effects in high-energy proton and ion machines," in *AIP Conf. Proc.*, vol. 1812, National Harbor, MD: AIP, 2017, p. 110006.
- [6] K. J. Ruisard *et al.*, "Tuning low-current beams for nonlinear quasi-integrable optics experiments at the University of Maryland Electron Ring," in *Proceedings of IPAC2018*, 2018, pp. 3585–3588.
- [7] K. Ruisard, H. Baumgartner, B. Beaudoin, I. Haber, M. Teperman, and T. Koeth, "Experimental plans for single-channel strong octupole fields at the University of Maryland Electron Ring," in *Proceedings of NAPAC2016*, paper TUPOB12, Chicago, IL, 2016, pp. 507–510.
- [8] H. Baumgartner *et al.*, "Quantification of octupole magnets at the University of Maryland Electron Ring," in *Proceedings of NAPAC2016*, Chicago, IL, 2016, pp. 503–506.
- [9] A. Friedman *et al.*, "Computational methods in the warp code framework for kinetic simulations of particle beams and plasmas," *IEEE Transactions on Plasma Science*, vol. 42, no. 5, pp. 1321–1334, 2014.
- [10] H. Baumgartner, K. Ruisard, D. Matthew, T. Koeth, I. Haber, and B. Beaudoin, "Initial tests of nonlinear quasi-integrable optics at the university of maryland electron ring," in *Proceedings of IPAC18*, Vancouver, Canada, 2018.
- [11] S. Antipov *et al.*, "IOTA (Integrable Optics Test Accelerator): facility and experimental beam physics program," *J Instr.*, vol. 12, no. 03, Mar. 2017.

# ESS ACCELERATOR LATTICE DESIGN STUDIES AND AUTOMATIC SYNOPTIC DEPLOYMENT

Y. Levinsen, M. Eshraqi, T. Grandsaert, A. Jansson, H. Kocevar, Ø. Midttun, N. Milas, R. Miyamoto, C. Plostinar, A. Ponton, R. de Prisco, T. Shea, ESS, Lund, Sweden  
H. D. Thomsen, Aarhus University, Aarhus, Denmark

## Abstract

The European Spallation Source is currently under construction in the south of Sweden. A highly brilliant neutron source with a 5 MW proton driver will provide state of the art experimental facilities for neutron science. A peak proton beam power in the accelerator of 125 MW means that excellent control over the beam losses becomes essential. The beam physics design of the ESS accelerator is in a TraceWin format, for which we have developed revision control setup, automated regression analysis and deployment of synoptic viewer and tabulated spreadsheets. This allows for an integrated representation of the data that are always kept synchronized and available to other engineering disciplines. The design of the accelerator lattice has gone through several major and minor iterations which are all carefully analysed. In this contribution we present the status of the latest studies, which includes the first complete end-to-end study beginning from the ion source.

## INTRODUCTION

The construction of the European Spallation Source (ESS) is currently ongoing at full force [1], with the first part of the accelerator under commissioning now in the second half of 2018. The ESS is designed to provide the neutron instruments with the world brightest neutron source, coming from the spallation process of a 5 MW proton beam hitting a rotating tungsten target [2]. ESS is built outside of Lund, Sweden, and is a European Infrastructure Research Consortium (ERIC) [3], with 12 founding member countries. A large fraction of the contributions from the member states to the ESS project is done in form of in-kind contracts, and there are currently 38 in-kind partners involved in the ESS project. The ESS user programme is planned to start in 2023.

The ESS accelerator layout is shown in Fig. 1. A microwave discharge ion source is producing approximately 3 ms of stable proton beam pulse of 75 keV at around 70 mA, which is accelerated through an RFQ and DTL, together with two transport sections that make up the normal-conducting front end. After that there are three families of superconducting cavities that bring the beam energy up from around 90 MeV to the final 2 GeV beam energy that is painted onto the rotating tungsten target.

To maintain control over the changes in the beam physics design lattice, and to try to keep the beam physics simulations as close to reality as possible, we have developed a deployment procedure for changing the beam physics lattice files. This procedure involves the use of modern revision

control systems, continuous integration, and scripting languages for automated deployment on an interactive web page. Tools which will be familiar to any programmer, but might be a less obvious use case for physicists.

In the second part of this paper, we will go through our recent progress with the large scale integrated error studies of the entire machine, starting from the ion source and up to the target. These studies are essential to confirm that the design can deliver a performance according to specifications, while keeping the losses low enough to not cause problems in the machine.

## AUTOMATED LATTICE DEPLOYMENT SETUP

A challenge most accelerator projects face is how to translate beam physics design to accurate locations for all machine elements. Further, during the transition from a pure design phase to an installation and commissioning phase, the physics design might still change, which one wants to make sure to propagate to the appropriate databases when it involves changing physical locations and/or dimensions, or changes such as polarity switches which involves cable routing changes. In the end, most projects end up with some discrepancies between the files used for beam physics studies, and the actual machine installed. This complicates the work for beam physicists, who then need to evaluate which differences may have a relevant impact to beam physics studies, and need to go hunting for errors whenever there is a discrepancy between the machine behaviour and what is expected from simulations.

For the ESS, we have predominantly been using TraceWin [4, 5] to simulate the machine, so the beam physics files are stored in TraceWin format. In the beginning these were manually updated and kept in a synchronised folder, with one person being the main responsible for collecting the files for the different sections of the machine and combining to an integrated lattice description. An improvement on this procedure was to store all lattice files in a revision control system (git), so that all changes were authored and could be tracked properly. We extended this with a slightly stricter change control process for what we define as the baseline branch, in order to make sure that all involved parties are aware of and agree to changes to the official machine description.

This quickly ended up being the most accurate and up to date description of element locations, which typically means engineers and other non-physicists started being interested in the data. These users do not know the structure of the TraceWin format, and further, the TraceWin files are not

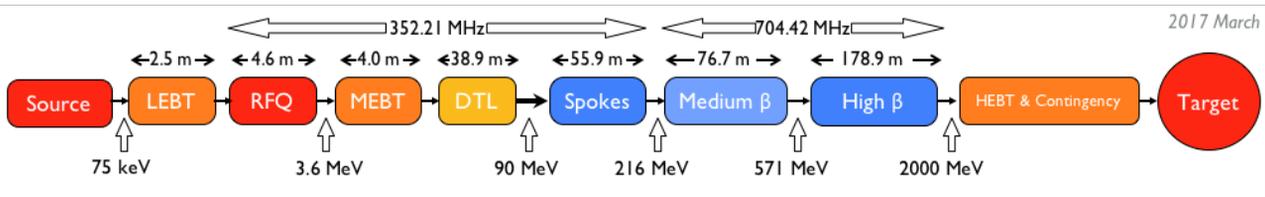


Figure 1: The overall layout of the ESS linac.

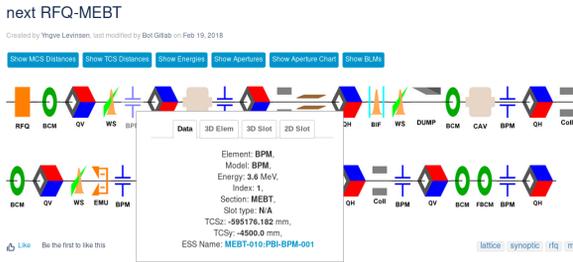


Figure 2: A screenshot demonstrating the synoptic viewer.

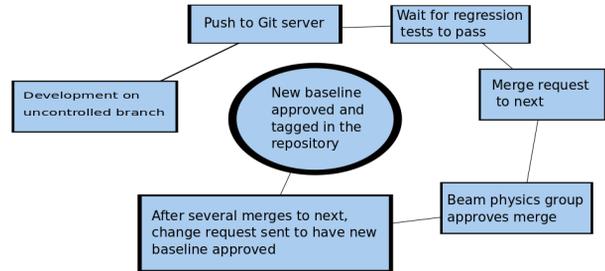


Figure 3: Our work flow for new lattice changes.

really suitable for storing database-like information about the elements. Information such as absolute location of elements are only indirectly available in these files, so a translation is necessary to make them useful to other collaborators.

An initiative was therefore started to tabulate the relevant data from the TraceWin files. This tabulation was automated through a Python script. A few extra descriptors (markers) were needed to fully automate the process, but this was kept to a minimum so that the original files would stay “physicist readable”.

This automated process was expanded on, as it now became easy to present the machine in an interactive synoptic viewer. This was realised with a second layer of Python scripts that use the tabulated data as input, and translates them to HTML files that are published on a Confluence Wiki page [6]. A screenshot from this viewer is shown in Fig. 2. All beamline and diagnostic elements are shown in order, and when one clicks on any element one can find useful information about the location with respect to target and source, beam energy at this location, drawings of the element etc. There are several requests to add more information, but we try to only accommodate those in cases when the process can remain automated.

Our current setup allows anyone editing or simulating beam physics files to do so in a way where they can be sure they work on the common and up to date lattice descriptions, and changes they make can be propagated back to the official branches through merge requests that are transparent to everyone involved. The full work flow is shown in Fig. 3. We currently have two controlled versions, one “next” branch which contains a largely stable compilation of the latest changes from the physicists, and one “baseline” branch that contains the latest officially approved lattice. Every time the baseline is updated, a new tag is created so that they can accurately be referred to in publications etc. All other branches in the repository are uncontrolled, which means

anyone can add what they want to these branches. When someone has changes that they believe should go in the official branch, they send a merge request from their branch to the next branch. This request is reviewed and approved by the beam physics team. When the next branch contain a significant amount of changes since baseline, the beam physics section prepares a change request which is reviewed and approved by the technical board. The technical board meets 3-4 times per year, which sets a limit on how many times the baseline can be updated.

## ERROR STUDIES

The entire ESS lattice has been studied in larger integrated end-to-end studies a few times already, see e.g. [7, 8]. These studies have so far started at the end of the RFQ, adding errors on the input beam that should mimic the real errors from the source, LEBT and RFQ. For the ESS lattice it is important to have excellent knowledge of the expected loss levels in the machine. The machine will deliver 5 MW proton beam power, while the non localised losses are required to stay below 1 W/m at energies beyond neutron production threshold. That means that the relative amount of losses that can be accepted are on the order of  $10^{-4}$  per metre at the front end, and down to  $10^{-7}$  for the 2 GeV beam.

In addition to losses, we also say that the emittance growth should not be more than approximately 10% in each section, which multiplies up to a maximum of around 100% from the MEBT to target. Since a large amount of macro particles is required to get good statistics in the beam halo (i.e. good loss patterns), one can make use of the emittance growth as a faster but indirect indication that the real beam going through the same machine may cause losses.

We first apply static errors to the machine according to our requirements. In particular, the RF phase and amplitude

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

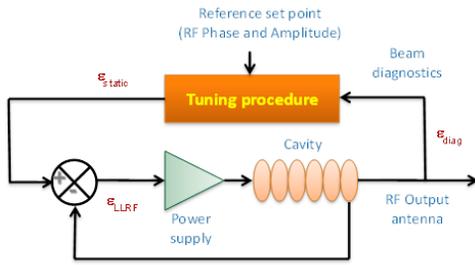


Figure 4: The schematic from the TraceWin manual explaining the concept of the tune cavity procedure [4].

tolerances are expected to be limiting factors for the machine performance. We then run our simulated correction using the beam parameters the diagnostic devices will give us at the location of the diagnostic devices. This matching is an iterative process and requires many simulations of the beam through the machine, which in some cases can take quite some time, even when using envelope calculation. Once the machine is corrected, we then apply the dynamic error tolerances, and finally we do a macro-particle tracking simulation to get accurate prediction of the beam behaviour through a linac that includes realistic errors. It should be noted that in the current configuration, the beam profile is not used for the correction procedure except for in the A2T area.

Such a simulation is performed many times over, and we then get statistical confidence that we will be able to keep at least  $N\%$  of the machines below the required loss level, and that the emittance stays within reason. We typically aim for up to 99% confidence, which requires around 1000 simulated machines in total to achieve certainty.

For the front-end, there has been studies on how much the output beam from the RFQ varies [9, 10]. We have looked at solenoid scans together with RFQ to see how the transmission through the front-end varies [11]. Lately we have added simulation of the ion source as well using the IBSimu plasma code [12], to get a better agreement between the simulated beam and the actual beam we get from the source that is now under commissioning [13, 14].

## Setup

The RFQ errors come from machining of the individual vanes, brazing of vanes to form a segment, and the alignment of each section during assembly of the complete RFQ. These inaccuracies cause errors in the quadrupolar fields of the vanes, as well as introducing dipolar terms. An extensive set of simulation tools have been developed in Python, to evaluate both the defined tolerances, and when available, include the measured vane profiles, brazing errors and alignment errors in the simulation [9, 15]. One can also do a combination, where measured data are used where available, and simulated errors based on defined tolerances are used for the rest.

We have added a new RF tuning procedure in the simulation, which has recently become available in TraceWin (TUNE\_CAVITY). The tune cavity procedure is explained in the diagram in Fig. 4. This procedure tries to directly translate the errors we define to how they affect the tuning procedure in the cavity, that involves both the diagnostics that measure the response, as well as errors in the LLRF that provide the tuning feedback to the power supply. This should more closely resemble how a real RF tuning is performed. We should add that we are working on some improvements in the tuning configuration of the DTL that did not make it into the lattice in time for this publication. Hence we might see slightly higher losses than expected in the DTL and downstream of the DTL.

## Results

In the error study presented here, we are for the first time including the LEBT and RFQ in the error study directly. For each machine, we take a 1 M sample out of a 10 M IBSimu simulation of the source, which we track through the LEBT and a RFQ vane profile that has been given random errors according to our tolerances. This is used as input for the usual MEBT-A2T error study, where we now no longer add any further input beam errors. At a later stage, we will look into adding errors in the LEBT. We expect that we should be able to reliably correct for the static errors of the LEBT by optimizing the mismatch factor in envelope mode, which is described by [16]

$$M = \left[ 1 + \frac{\Delta + \sqrt{\Delta(\Delta + 4)}}{2} \right]^{1/2} - 1, \quad (1)$$

where

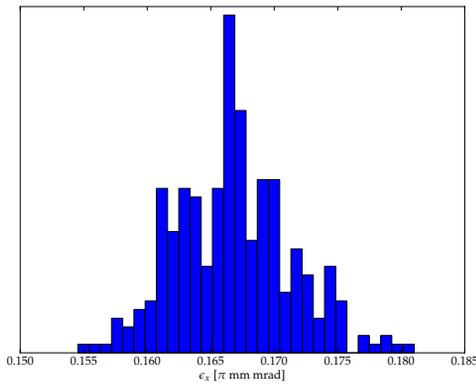
$$\Delta = (\Delta\alpha)^2 - \Delta\beta\Delta\gamma, \quad (2)$$

comparing the difference in Twiss parameters between the matched beam and the measured/simulated beam. Optimizing for the mismatch can be compared to optimizing for transmission through the RFQ, but the latter will be much too time-consuming for an error study of this scale.

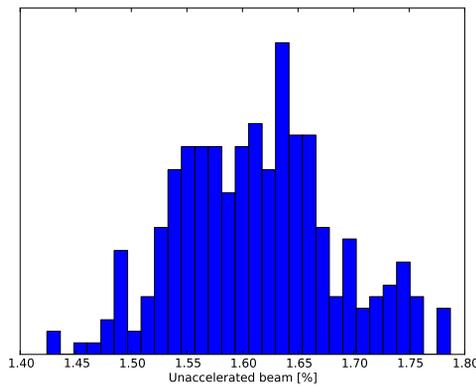
If we look at the beams coming out of the RFQ from the different machines, we see about 5-7% spread in the emittances (of the accelerated part of the beam), and around 1.5% of unaccelerated beam, as shown in Fig. 5. We only show the horizontal emittance, but both planes show a similar distribution.

The tracking from a few hundred machines gives the loss pattern shown in Fig. 6. The losses in the RFQ starts around 1 m, which is where the acceleration starts. Around 3-4% of the beam is lost in the RFQ, and another 1.6% is unaccelerated and largely contributes to that first peak we see in the MEBT. The DTL has a tight aperture compared to the downstream linac, so it effectively functions as some sort of collimator for the superconducting section. The spoke section hardly sees any beam losses in our simulations. This is in agreement with earlier simulations, the DTL effectively functions as a sort of collimator for the spoke section, since

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.



(a) Horizontal Emittance



(b) Unaccelerated fraction

Figure 5: The spread of the beam distributions coming out of the different RFQ's.

the aperture in the spoke section is much larger. The losses in the elliptical cavities (magenta) are essentially all originating from the frequency jump. While we do believe the loss pattern looks largely reasonable, we do believe the absolute numbers can be brought down with a further refinement of the procedure. For this reason we have left the vertical scale as arbitrary units for now.

The spike in the HEBT region is in the area after the neutron shield wall, just before the target. In this area losses on the order of 1 kW are expected. The losses in the simulation is a bit higher than what we would like to see, but not alarming, as we expect further refinement of the tuning configuration should bring the losses down. It is further not unreasonable to expect that correcting for the beam profile at profile measurement locations in the linac might improve the conditions at this location, where we defocus the beam significantly in both planes. Transversal errors generally becomes more relevant when the  $\beta$ -function increases.

## SUMMARY

We have developed a set of practices to maintain control over the changes done in the beam physics lattice files

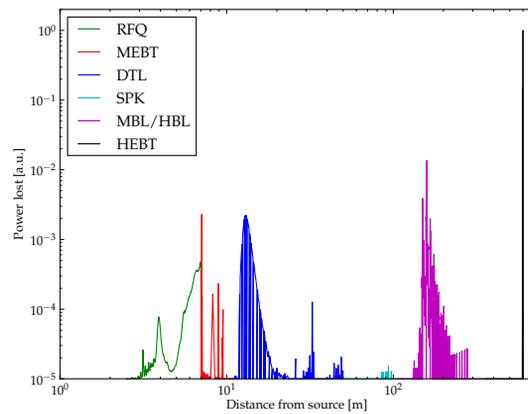


Figure 6: The average losses in W/m from the simulated machines, with a different colour for the different sections of the machine.

for ESS. Using this basis, we have then developed a set of tools to automate retrieval of useful engineering data from these files, and added tests that automatically checks that all representations of the data are consistent. This has made the lattice files useful to several others outside of the beam physics group.

In our latest error study we have done some significant changes, introducing RFQ errors as well as making use of a new RF tuning procedure. The results are already showing reasonable agreement with the old data. Further refinement of the simulated tuning procedure should most likely reduce the losses in the high energy part of the machine compared to the results presented here. We believe this is a good step towards a more realistic simulation of the machine correction, which provides a deeper understanding of the defined error tolerances.

## REFERENCES

- [1] M. Lindroos *et al.*, “ESS Progressing into Construction,” in *IPAC'16, Busan, South Korea*, Jun. 2016. DOI: 10.18429/JACoW-IPAC2016-FRYAA02.
- [2] R. Garoby *et al.*, “The European Spallation Source Design,” *Physica Scripta*, vol. 93, no. 1, p. 014001, 2018, ISSN: 1402-4896. DOI: 10.1088/1402-4896/aa9bff. <http://stacks.iop.org/1402-4896/93/i=1/a=014001>
- [3] <https://ec.europa.eu/research/infrastructures/index.cfm>
- [4] D. Uriot, *TraceWin Simulation Software*. <http://irfu.cea.fr/Sacm/en/logiciels/index3.php>
- [5] D. Uriot and N. Pichoff, “Status of tracewin code,” in *IPAC'15, Richmond, VA, USA, 2015*. DOI: 10.18429/JACoW-IPAC2015-MOPWA008.
- [6] <https://confluence.esss.lu.se/display/ALS>
- [7] M. Eshraqi, R. de Prisco, R. Miyamoto, and Y. I. Levinsen, “Preliminary study of the possible failure modes of the components of the ESS linac,” European Spallation Source ERIC, Tech. Rep. ESS/AD/0057, 2015.

- [8] Y. I. Levinsen, "ESS 2015 baseline lattice error study," Tech. Rep. ESS-0049433, Feb. 2016.
- [9] A. Ponton, "Beam physics analysis of the ESS RFQ non-conformities," in *Proceedings, 9th International Particle Accelerator Conference (IPAC 2018): Vancouver, BC Canada, 2018*, TUPAF067. doi: 10.18429/JACoW-IPAC2018-TUPAF067.
- [10] E. Sargsyan, "ESS RFQ error study," European Spallation Source ERIC, Tech. Rep. ESS-0011071, May 6, 2014.
- [11] Y. I. Levinsen, M. Eshraqi, L. Celona, and L. Neri, "In-depth analysis and optimization of the european spallation source front end lattice," in *IPAC'16, Busan, South Korea*, Jun. 2016. doi: 10.18429/JACoW-IPAC2016-TUPMR020.
- [12] T. Kalvas, O. Tarvainen, T. Ropponen, O. Steczkiewicz, J. Ärje, and H. Clark, "IBSIMU: A three-dimensional simulation software for charged particle optics," *Review of Scientific Instruments*, 10.1063/1.3258608, vol. 81, no. 2, 2010. doi: 10.1063/1.3258608.
- [13] O. Midttun, Y. I. Levinsen, R. Miyamoto, and C. Plostinar, "Benchmarking of the ESS LEPT in TraceWin and IBSimu," in *IPAC'17, Copenhagen, Denmark, 2017*. doi: 10.18429/JACoW-IPAC2017-THPVA013.
- [14] R. Miyamoto *et al.*, "Preparation towards the ESS linac ion source and LEPT beam commissioning on ESS site," in *IPAC'18, Vancouver, BC, Canada, May 2018*. doi: 10.18429/JACoW-IPAC2018-TUPAF064.
- [15] A. Ponton, Y. I. Levinsen, E. Sargsyan, A. C. France, O. Piquet, and B. Pottin, "Voltage error studies in the ESS RFQ," in *IPAC'16, Busan, South Korea*, Jun. 2016. doi: 10.18429/JACoW-IPAC2016-THPMB039.
- [16] T. P. Wangler, *RF Linear Accelerators*, 2nd. 2008.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

# S-BASED MACRO-PARTICLE SPECTRAL ALGORITHM FOR AN ELECTRON GUN

Paul M. Jung, Thomas Planche, TRIUMF, Vancouver, Canada

## Abstract

We derive a Hamiltonian description of a continuous particle distribution and its electrostatic potential from the Low Lagrangian. The self consistent space charge potential is discretized according to the spectral Galerkin approximation. The particle distribution is discretized using macro-particles. We choose a set of initial and boundary conditions to model the TRIUMF 300keV thermionic DC electron gun. The field modes and macro-particle coordinates are integrated self-consistently. The current status of the implementation is discussed.

## INTRODUCTION

The section of beamline we are trying to model includes the electron gun (Fig. 1) and one solenoid, a total length of 57 cm up to the first view screen. The electrons are emitted from a hot cathode. An RF grid is placed a fraction of a millimetre downstream from the cathode. It is used to modulate the emission of electrons at 650 MHz. Electrons are accelerated to 300 keV using a DC field. The distance between the cathode and the ground electrode is 12 cm. The emitting surface of the cathode has a radius of 4 mm. The nominal bunch charge is 15 pC with a bunch length of 130 ps, see [1]. The solenoid enables us to adjust the phase advance between the cathode and the view screen. At a particular phase advance, we can use the electron beam to create an image of the RF grid on the view screen see Fig. 2. Scanning the phase advance enables us to measure the transverse phase space distribution using tomography [2]. Our objective is to reproduce these measurements using an algorithm derived from the least action principle like in [3–5]. The description of thermionic emission and effects from the grid are outside the scope of this model.

Following classical field theory conventions, let an over dot ‘ $\dot{\cdot}$ ’ represent an explicit derivative with respect to time, and similarly a prime ‘ $\prime$ ’ denotes a partial derivative with respect to  $z$ . We write the vectors that lie in the transverse  $xy$  plane with a lower ‘ $\perp$ ’. For example:  $\mathbf{x}_\perp$  is the vector  $(x, y, 0)$ .

## CONTINUOUS MODEL

We start from the Low Lagrangian [6] which is a sum of two integrals:

$$L = \int d^3\mathbf{x}_0 d^3\dot{\mathbf{x}}_0 \mathcal{L}_p(\mathbf{x}(\mathbf{x}_0, \dot{\mathbf{x}}_0, t), \dot{\mathbf{x}}(\mathbf{x}_0, \dot{\mathbf{x}}_0, t); \mathbf{x}_0, \dot{\mathbf{x}}_0, t) + \int d^3\bar{\mathbf{x}} \mathcal{L}_f(\phi, \mathbf{A}; \bar{\mathbf{x}}, t), \quad (1)$$

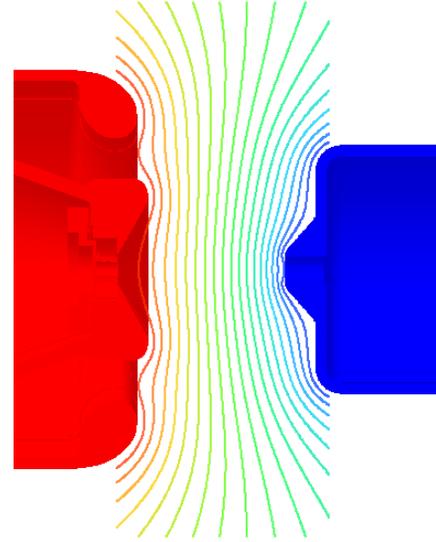


Figure 1: OPERA Model of the 300 keV TRIUMF electron gun with equipotential lines of the electric potential.

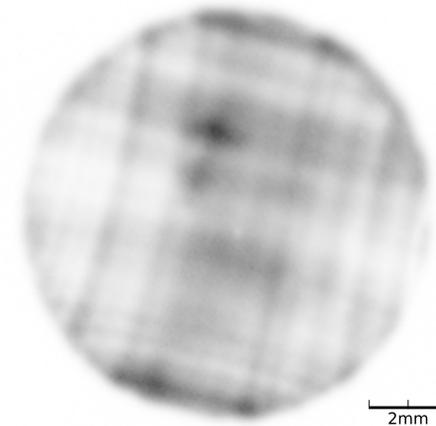


Figure 2: The view screen image, after the first solenoid.

where the Lagrangian densities are:

$$\begin{aligned} \mathcal{L}_p(\mathbf{x}, \dot{\mathbf{x}}; \mathbf{x}_0, \dot{\mathbf{x}}_0, t) &= \\ f(\mathbf{x}_0, \dot{\mathbf{x}}_0) \left( -mc^2 \sqrt{1 - |\dot{\mathbf{x}}|^2/c^2} - q\phi(\mathbf{x}, t) + q\dot{\mathbf{x}} \cdot \mathbf{A}(\mathbf{x}, t) \right), & (2) \\ \mathcal{L}_f(\phi, \mathbf{A}; \mathbf{x}, t) &= \\ \frac{\epsilon_0}{2} \left( |\nabla\phi(\mathbf{x}, t) + \dot{\mathbf{A}}(\mathbf{x}, t)|^2 - c^2 |\nabla \times \mathbf{A}(\mathbf{x}, t)|^2 \right). & (3) \end{aligned}$$

and  $\bar{\mathbf{x}}$  is a dummy variable of integration.

To describe the self field we make the assumption that in the centre of mass frame the self field is completely described by the scalar potential  $\phi(\mathbf{x}, t)$ , and the vector potential is zero. We assume that the beam is travelling in the positive  $z$ -direction, with unit vector  $\hat{z}$ . Now, by applying an active

Lorentz transformation we find that this field corresponds to:

$$\phi(\mathbf{x}, t) = \gamma_0 \phi(\mathbf{x}, t), \quad (4)$$

$$\mathbf{A}(\mathbf{x}, t) = \frac{\beta_0}{c} \gamma_0 \phi(\mathbf{x}, t) \hat{z} = \frac{\beta_0}{c} \phi(\mathbf{x}, t) \hat{z}, \quad (5)$$

in the laboratory frame. The fields before and after transformation are functions of the coordinates in the laboratory frame.  $c\beta_0$  is the centre of mass velocity and  $\gamma_0$  is the corresponding Lorentz factor. We also assume:

$$\dot{\mathbf{x}} \cdot \hat{z} \approx c\beta_0. \quad (6)$$

Now, we can write the new Lagrangian densities as:

$$\mathcal{L}_p(\mathbf{x}, \dot{\mathbf{x}}; \mathbf{x}_0, \dot{\mathbf{x}}_0, t) = -fmc^2 \sqrt{1 - |\dot{\mathbf{x}}|^2/c^2} - fq\gamma_0^{-2} \phi(\mathbf{x}, t), \quad (7)$$

$$\mathcal{L}_f(\phi; \bar{\mathbf{x}}, t) =$$

$$\frac{\epsilon_0}{2} \left( \gamma_0^{-2} |\nabla_{\perp} \phi(\bar{\mathbf{x}}, t)|^2 + \left| \partial_z \phi(\bar{\mathbf{x}}, t) + c^{-1} \partial_t (\beta_0 \phi(\bar{\mathbf{x}}, t)) \right|^2 \right), \quad (8)$$

so we have described the self field using only a scalar potential.

To check the reasonableness of the field Lagrangian we look at the equation of motion for the scalar field which is:

$$\left( \partial_z + \frac{\beta_0}{c} \partial_t \right)^2 \phi + \frac{\beta_0'}{c} \dot{\phi} + (1 - \beta_0^2) \nabla_{\perp}^2 \phi = 0. \quad (9)$$

In the ultra-relativistic limit  $\beta_0 = 1$  we find that the transverse dynamics of  $\phi$  become frozen and we have the following equation of motion:

$$\left( \partial_z + \frac{1}{c} \partial_t \right)^2 \phi = 0. \quad (10)$$

The solutions to this equation are wave-fronts travelling in the  $z$  direction with speed  $c$ . In the stationary limit  $\beta_0 = 0$  we find the Laplace equation:

$$\phi'' + \nabla_{\perp}^2 \phi = 0. \quad (11)$$

Thus far we have described a Lagrangian system with two main assumptions, the self-field in the beam frame is electrostatic and that  $\Delta\beta/\beta_0 \ll 1$ .

### Change of Independent Variable

Changing the independent variable in the Lagrangian is done using coordinate transformation [7, 8]. We find that the new Lagrangian density is:

$$\mathcal{L}_p(\mathbf{x}_{\perp}, t, \mathbf{x}'_{\perp}, t'; z) = -fmc \sqrt{(ct')^2 - |\mathbf{x}'_{\perp}|^2} - fq\gamma_0^{-2} t' \phi(\mathbf{x}_{\perp}, t, z), \quad (12)$$

and the field Lagrangian density is unchanged.

### Hamiltonian

Since we have taken  $z$  as the independent variable, we can take  $\gamma_0$  and  $\beta_0$  to be solely functions of  $z$ . The momentum density canonically conjugated to particle position is:

$$P_{\mathbf{x}_{\perp}}(x_0, y_0, t_0, x'_0, y'_0, t'_0, z) = \frac{\partial \mathcal{L}_p}{\partial \dot{\mathbf{x}}'_{\perp}} = \frac{fmc \mathbf{x}'_{\perp}}{\sqrt{(ct')^2 - |\mathbf{x}'_{\perp}|^2}}, \quad (13)$$

$$\begin{aligned} -E(x_0, y_0, t_0, x'_0, y'_0, t'_0, z) &= \frac{\partial \mathcal{L}_p}{\partial t'} \\ &= \frac{-fmc^3 t'}{\sqrt{(ct')^2 - |\mathbf{x}'_{\perp}|^2} - 1} - fq\gamma_0^{-2} \phi. \end{aligned} \quad (14)$$

As for the scalar potential, we have that:

$$\pi_{\phi}(\mathbf{x}_{\perp}, t, z) = \frac{\partial \mathcal{L}_f}{\partial \phi'} = \epsilon_0 \left( \partial_z + \frac{\beta_0}{c} \partial_t \right) \phi. \quad (15)$$

So, we find the Hamiltonian to be:

$$\begin{aligned} H &= \int d x_0 d y_0 d t_0 d x'_0 d y'_0 d t'_0 \mathcal{H}_p(\mathbf{x}_{\perp}, t, \mathbf{P}_{\perp}, E; x_0, y_0, t_0, x'_0, y'_0, t'_0) \\ &+ \int d^2 \bar{\mathbf{x}}_{\perp} d \bar{r} \mathcal{H}_f(\phi, \pi_{\phi}; \bar{\mathbf{x}}_{\perp}, \bar{r}, z), \end{aligned} \quad (16)$$

where the Hamiltonian densities are given by the Legendre transform:

$$\begin{aligned} \mathcal{H}_p &= P_{\mathbf{x}_{\perp}} \cdot \dot{\mathbf{x}}'_{\perp} - Et' - \mathcal{L}_p \\ &= -\sqrt{\frac{1}{c^2} \left( E - fq\gamma_0^{-2} \phi(\mathbf{x}_{\perp}, t, z) \right)^2 - |\mathbf{P}_{\perp}|^2 - (mfc)^2}, \end{aligned} \quad (17)$$

$$\mathcal{H}_f = \pi_{\phi} \phi' - \mathcal{L}_f = \frac{\pi_{\phi}^2}{2\epsilon_0} - \frac{\beta_0}{c} \pi_{\phi} \dot{\phi} - \frac{\epsilon_0}{2\gamma_0^2} (\nabla_{\perp} \phi)^2. \quad (18)$$

Lastly, we can examine the equations of motion:

$$\begin{aligned} \dot{\mathbf{x}}'_{\perp} &= \frac{\mathbf{P}_{\perp}}{P_z}, & \dot{\mathbf{P}}_{\perp} &= fq\gamma_0^{-2} t' \nabla_{\perp} \phi(\mathbf{x}_{\perp}, t, z), \\ t' &= \frac{E - fq\gamma_0^{-2} \phi(\mathbf{x}_{\perp}, t, z)}{c^2 P_z}, & E' &= fq\gamma_0^{-2} t' \partial_t \phi(\mathbf{x}_{\perp}, t, z), \end{aligned} \quad (19)$$

where

$$P_z = \sqrt{\frac{1}{c^2} \left( E - fq\gamma_0^{-2} \phi(\mathbf{x}_{\perp}, t, z) \right)^2 - |\mathbf{P}_{\perp}|^2 - (mfc)^2}, \quad (20)$$

as well as the equations of motion for the scalar potential canonical pair:

$$\phi' = \frac{\pi_{\phi}}{\epsilon_0} - \frac{\beta_0}{c} \dot{\phi}, \quad \pi'_{\phi} = \frac{\epsilon_0}{\gamma_0^2} \nabla_{\perp}^2 \phi + \frac{\beta_0}{c} \dot{\phi}. \quad (21)$$

These equations of motion are useful to get an intuitive picture of the model. To obtain the discrete Hamiltonian we can now discretize our Lagrangian system and follow the same steps.

## DISCRETEIZATION

Our choice of discretization scheme is:

$$f(\mathbf{x}_0, \dot{\mathbf{x}}_0) = \sum_j w^j \delta^{(3)}(\mathbf{x}_0^j - \mathbf{x}_0) \delta^{(3)}(\dot{\mathbf{x}}_0^j - \dot{\mathbf{x}}_0), \quad (22)$$

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

$$\sum_{nm\ell} \Phi_{nm\ell}(z) \cos\left(\frac{n\pi x}{L_x}\right) \cos\left(\frac{m\pi y}{L_y}\right) \cos\left(\frac{\ell\pi \Delta t}{L_t}\right), \quad (23)$$

where the particle distribution is a sum of Dirac delta functions which gives us point-like model particles. The basis functions of the scalar potential are chosen such that each of

them satisfies the boundary conditions, the Galerkin approximation. The field is contained in the box of size  $L_x \times L_y \times L_t$  and is zero at the boundaries. The self field is periodic in time with the period being the RF period, and zero at the boundaries. The field mode labels  $n, m, \ell$  are positive odd integers. This selects the modes that satisfy the boundary conditions and are even functions about each axis.

Substituting these into the Lagrangian, simplifying and solving for the Hamiltonian gives:

$$H = \sum_j H^j(\mathbf{x}_\perp^j, \mathbf{P}_\perp^j, t^j, E^j; z) + \sum_{nm\ell} H_{nm\ell}(\Phi_{nm\ell}, \Pi_{nm\ell}; z), \quad (24)$$

where the model particle Hamiltonian for particle  $j$  is:

$$H^j = -\sqrt{\frac{1}{c^2} \left( P_t^j - q w^j \gamma_0^{-2} \phi(\mathbf{x}_\perp^j, t^j, z) \right)^2 - |\mathbf{P}_\perp^j|^2 - (m w^j c)^2} - (\Delta E^j + E_0) t_0' + \Delta t^j E_0', \quad (25)$$

and for field mode  $n, m, \ell$  it is:

$$H_{nm\ell} = \frac{1}{2V} \Pi_{nm\ell}^2 - \frac{V}{2\gamma_0^2} \left( \left( \frac{n\pi}{L_x} \right)^2 + \left( \frac{m\pi}{L_y} \right)^2 + \left( \frac{\beta_0 \gamma_0 \ell \pi}{c L_t} \right)^2 \right) \Phi_{nm\ell}^2, \quad (26)$$

where  $V$  is a volume factor given by  $V = \frac{1}{8} \epsilon_0 L_x L_y L_t$ . Note that the  $\pi_\phi \phi$  term became decoupled in this Hamiltonian because of the orthogonality of the basis functions and its derivatives. The discretized Hamiltonian yields an equation of motion for each of the discrete degrees of freedom.

The equations of motion for the macro-particles are:

$$\mathbf{x}_\perp^{j'} = \frac{\mathbf{P}_\perp^j}{P_z^j}, \quad \Delta t^{j'} = \frac{P_t^j - q w^j \gamma_0^{-2} \phi(\mathbf{x}_\perp^j, t^j, z)}{c^2 P_z^j} + t_0', \quad (27)$$

$$\begin{aligned} \mathbf{P}_\perp^{j'} &= w^j q \gamma_0^{-2} (t_0' - \Delta t^{j'}) \nabla_\perp \phi(\mathbf{x}_\perp^j, t^j, z), \\ \Delta E^{j'} &= w^j q \gamma_0^{-2} (t_0' - \Delta t^{j'}) \phi(\mathbf{x}_\perp^j, t^j, z) - E_0', \end{aligned} \quad (28)$$

where the longitudinal particle momentum is calculated by:

$$P_z^j = -\sqrt{\frac{1}{c^2} \left( P_t^j - q w^j \gamma_0^{-2} \phi(\mathbf{x}_\perp^j, t^j, z) \right)^2 - |\mathbf{P}_\perp^j|^2 - (m w^j c)^2}. \quad (29)$$

Also, the equations of motion for the field modes are given by:

$$\begin{aligned} \Phi'_{nm\ell} &= \frac{1}{V} \Pi_{nm\ell}, \\ \Pi'_{nm\ell} &= \frac{V}{\gamma_0^2} \left( \left( \frac{n\pi}{L_x} \right)^2 + \left( \frac{m\pi}{L_y} \right)^2 + \left( \frac{\beta_0 \gamma_0 \ell \pi}{c L_t} \right)^2 \right) \Phi_{nm\ell} \\ &+ \sum_j \frac{q w^j}{c^2 \gamma_0^2} (\Delta t^{j'} - t_0') \cos \left( \frac{n\pi x^j}{L_x} \right) \cos \left( \frac{m\pi y^j}{L_y} \right) \cos \left( \frac{\ell \pi \Delta t^j}{L_t} \right), \end{aligned} \quad (30)$$

## IMPLEMENTATION AND FUTURE WORK

The current implementation is written in Python as vectorized Numpy code. The system of differential equations These equations of motion are a consistent set of coupled first order ordinary differential equations.

is solved using the `scipy.integrate` module. One integration method is chosen and the equations are integrated simultaneously with appropriate tolerances for adaptive step size methods. All of the integration methods provided by the module were tested and integrating the fields was unconditionally unstable.

Future work is to understand and address the problems with integration.

## REFERENCES

- [1] F. Ames *et al.*, “The TRIUMF ARIEL RF Modulated Thermionic Electron Source”, in *Proc. LINAC’16*, East Lansing, MI, USA, pp. 458-461, 2017. doi:10.18429/JACoW-LINAC2016-TUPRC020
- [2] M. Marchetto *et al.*, “Commissioning and Operation of the ARIEL Electron LINAC at TRIUMF”, in *Proc. IPAC’15*, Richmond, VA, USA, pp. 2444-2449, 2015. doi:10.18429/JACoW-IPAC2015-WEYC3
- [3] H. Qin and X. Guan, “Variational symplectic integrator for long-time simulations of the guiding-center motion of charged particles in general magnetic fields”, *Phys. Rev. Lett.*, vol. 100, pp. 035006. doi:10.1103/PhysRevLett.100.035006
- [4] H. Qin *et al.*, “Canonical symplectic particle-in-cell method for long-term large-scale simulations of the Vlasov-Maxwell equations”, *Nuclear Fusion*, vol. 56, no. 1, pp. 014001, 2016. doi:10.1088/0029-5515/56/1/014001
- [5] S. D. Webb, “A spectral canonical electrostatic algorithm”, *Plasma Physics and Controlled Fusion*, vol. 58, no. 3, pp. 034007, 2016. doi:10.1088/0741-3335/58/3/034007
- [6] F. E. Low, “A Lagrangian formulation of the Boltzmann-Vlasov equation for plasmas”, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 248, 1958. doi:10.1098/rspa.1958.0244
- [7] R. Broucke, “On changes of independent variable in dynamical systems and applications to regularization”, *Icarus*, vol. 7, no. 1-3, pp. 221-232, 1967. doi:10.1016/0019-1035(67)90066-8
- [8] T. Planche and P. M. Jung, “Symplectic And Self-consistent Algorithms For Particle Accelerator Simulation”, in *Proc. ICAP’18*, Key West, Florida, USA, Oct 2018. doi:10.18429/JACoW-ICAP2018-SUPAF04

# CONSTRAINED MULTI-OBJECTIVE SHAPE OPTIMIZATION OF SUPERCONDUCTING RF CAVITIES TO COUNTERACT DANGEROUS HIGHER ORDER MODES

M. Kranjčević\*, P. Arbenz, Computer Science Department, ETH Zurich, 8092 Zürich, Switzerland  
A. Adelman, Paul Scherrer Institut (PSI), 5232 Villigen, Switzerland  
S. Gorgi Zadeh, U. van Rienen, University of Rostock, 18059 Rostock, Germany

## Abstract

High current storage rings, such as the Z operating mode of the FCC-ee (FCC-ee-Z), require superconducting radio frequency (RF) cavities that are optimized with respect to both the fundamental mode and the dangerous higher order modes (HOMs). In this paper, in order to optimize the shape of the RF cavity, a constrained multi-objective optimization problem is solved using a massively parallel implementation of an evolutionary algorithm. Additionally, a frequency-fixing scheme is employed to deal with the constraint on the frequency of the fundamental mode. Finally, the computed Pareto front approximation and an RF cavity shape with desired properties are shown.

## INTRODUCTION

Superconducting RF cavities are mainly optimized with respect to the properties of the fundamental mode [1]. However, in high current machines, such as the FCC-ee-Z [2], monopole and dipole modes are major sources of beam instability. The first monopole HOM band can be untrapped by enlarging the beam pipe radius, but the first dipole band remains trapped in the cavity and requires a special damping mechanism. In order to ease the HOM damping of the first dipole band, in this paper, a multi-objective shape optimization of a single-cell cavity that takes into account both the fundamental mode and the first dipole band is performed. The optimization algorithm is described on the concrete problem of optimizing the shape of the axisymmetric cavity for the FCC-ee-Z, but the same method can be used with other objectives and parameterizations.

## MULTI-OBJECTIVE OPTIMIZATION

In this paper four objective functions have to be optimized simultaneously. First, the distance between the frequency of the first dipole mode,  $f_1$ , which is typically the TE<sub>111</sub> mode, and the frequency of the fundamental mode,  $f_0$ , has to be maximized. Second, the distance between  $f_1$  and the frequency of the second dipole mode,  $f_2$ , which is typically the TM<sub>110</sub> mode, has to be minimized. These two objectives simplify the design of the HOM couplers for damping the first dipole band. Third, the sum of the transverse shunt impedances of the dipole modes has to be minimized. The following definition of the transverse shunt impedance for

the dipole modes is used [3]

$$\frac{R}{Q_{\perp}} = \frac{(V_{||}(r=r_0) - V_{||}(r=0))^2}{k^2 r_0^2 \omega U},$$

where  $k$  is the wave number,  $r_0$  the offset from the axis,  $\omega$  the angular frequency, and  $U$  the stored energy. Fourth,  $G_0 \cdot R/Q_0$  ( $G_0$  is the geometry factor) of the fundamental mode has to be maximized, since it is inversely related to the dissipated power on the surface of the cavity [4]. In addition to these four objectives,  $f_0$  has to be fixed to the operating frequency of 400.79 MHz.

Seven variables ( $R_{eq}$ ,  $R_i$ ,  $L$ ,  $A$ ,  $B$ ,  $a$  and  $b$ ) uniquely describe the shape of an elliptical cavity as shown in Fig. 1. The wall slope angle  $\alpha$  can be computed from these design variables and, in order to avoid re-entrant shape cavities, it must be at least 90°.

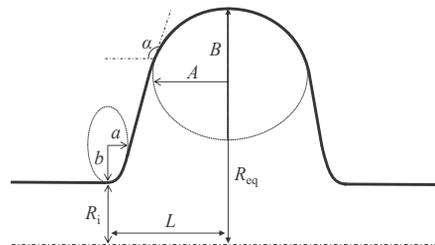


Figure 1: Parameterization of a single-cell elliptical cavity.

This can be formulated as the constrained multi-objective optimization problem

$$\begin{aligned} \min_{R_i, L, A, B, a, b} \quad & \underbrace{(f_0 - f_1)}_{F_1}, \underbrace{|f_1 - f_2|}_{F_2}, \underbrace{\frac{R}{Q_{\perp 1}} + \frac{R}{Q_{\perp 2}}}_{F_3}, \underbrace{-G_0 \cdot \frac{R}{Q_0}}_{F_4}, \\ \text{subject to} \quad & f_0 = 400.79 \text{ MHz and } \alpha \geq 90^\circ. \end{aligned} \quad (1)$$

Since in a single-cell cavity there is no restriction on the length  $2L$  of the cell (because the particle only passes through one cell), and since this length highly affects  $f_1$ ,  $L$  is also taken to be a design variable. On the other hand, the variable  $R_{eq}$  has the highest influence on the value of  $f_0$ , so it is not considered to be a design variable in the optimization, but rather used to tune  $f_0$  to the desired value.

As part of the field leaks into the beam pipe, the cell and the beam pipe are simulated together and  $f_0$  is tuned to 400.79 MHz taking the beam pipe effect into account. In this paper, the beam pipe length is set to the value of the wave length  $\lambda = 748$  mm.

\* marija.kranjcevic@inf.ethz.ch

## FORWARD SOLVER

When solving time-harmonic Maxwell's equations in a vacuuated axisymmetric RF cavity with perfectly electrically conducting (PEC) boundary conditions (BC) using the finite element method, one gets a generalized eigenvalue problem (GEVP) for each  $m \in \mathbb{N}_0$  [5, 6]. The parameter  $m$  is the azimuthal mode number, and it is 0 for monopole modes, 1 for dipole modes, etc. Since the cross section of the single-cell elliptical cavity, as shown in Fig. 1, is symmetric, it is sufficient to solve time-harmonic Maxwell's equations in only half of it, once with PEC and once with perfectly magnetically conducting (PMC) BC on the symmetry plane.

For a specific value of  $R_{eq}, R_i, L, A, B, a$  and  $b$ , half of the cross section of the corresponding cavity is created and meshed. In order to compute the properties of the fundamental mode, the smallest eigenpair of the GEVP corresponding to  $m = 0$  and PEC BC is found (TM<sub>010</sub>). Similarly, to compute the properties of the dipole modes, the smallest eigenpair of the GEVPs corresponding to  $m = 1$  and PEC (TM<sub>110</sub>) and PMC (TE<sub>111</sub>) BC on the cross section symmetry plane is found.

## OPTIMIZATION ALGORITHM

### Evolutionary Algorithm (EA)

A design point  $\mathbf{d}_1 = (R_{i,1}, L_1, A_1, B_1, a_1, b_1)$  dominates  $\mathbf{d}_2$  if it is not worse in any of the objectives, and it is strictly better in at least one objective. A massively parallel implementation of an EA [7], combined with the axisymmetric Maxwell eigensolver [8], is used to find points that are not dominated by any other point, called Pareto optimal points. The basic steps of an EA are given in Algorithm 1. The individuals (i.e., the RF cavities) comprising the first generation are chosen randomly, i.e., their design variable values are chosen randomly from a given interval (line 1). The values used in this paper are given in Table 1. These individuals are then evaluated, i.e., their objective function values are computed (line 2). After that, a predetermined number of cycles is performed, each resulting in a new generation (lines 3-7). In every cycle, crossover and mutation operators are used to create new individuals (lines 4-5) which are subsequently evaluated (line 6). The new generation is chosen to comprise approximately  $N$  fittest individuals (line 7).

#### Algorithm 1 Evolutionary algorithm

- 1: random population of individuals,  $I_i, i = 1, \dots, N$
- 2: evaluate the population
- 3: **for** a predetermined number of generations **do**
- 4:     **for** pairs of individuals  $I_i, I_{i+1}$  **do**
- 5:         crossover( $I_i, I_{i+1}$ ), mutate( $I_i$ ), mutate( $I_{i+1}$ )
- 6:     evaluate new individuals
- 7:     choose  $N$  fittest individuals for the next generation

Table 1: Design Variable Bounds, in mm

Variable	$R_i$	$L$	$A$	$B$	$a$	$b$
Lower bound	145	120	40	40	10	10
Upper bound	160	190	140	140	70	70

### Constraint Handling

For each design point  $\mathbf{d} = (R_i, L, A, B, a, b)$  it is first necessary to determine the value of  $R_{eq}$  for which the frequency of the fundamental mode is  $f_0 = 400.79$  MHz. This can be done with a zero-finding method. In this paper, TOMS 748 [9] is used, with the initial guess (in mm)  $R_{eq} \in [325, 375]$ . Furthermore, each time  $f_0$  is computed for this fixed  $\mathbf{d}$  and a specific  $R_{eq}$ , it is enough to solve the eigenproblem corresponding to  $m = 0$  and PEC BC, and, as long as  $f_0$  is far away from the desired value, a coarse mesh can be used to speed up this process. If such a value of  $R_{eq}$  is not found, the individual is declared invalid and discarded from the population. Similarly, the individual is discarded if the value of the wall slope angle  $\alpha$  is below  $90^\circ$ .

## RESULTS

The optimizations were run on the Euler cluster<sup>1</sup> (Euler I and II) of ETH Zurich. The coarse eigensolves use a mesh with around 10'000 triangles, and the fine ones around 300'000 triangles. Solving just one coarse eigenproblem (meshing, computing 3 smallest eigenpairs and the objective function values) takes around 2 s. Solving a fine one takes around 90 s (24 s for meshing, 64 s for computing the eigenpairs, and 2 s for computing the objective function values). On average, 4 fine eigensolves are necessary to find the value of  $R_{eq}$  and the properties of the fundamental mode. After that, two more fine eigensolves (using an existing mesh) are needed to compute the properties of the dipole modes. Computing 50 generations of the EA with  $N = 100$  (where almost 30% of the evaluated individuals get discarded from the optimization) using 96 processes takes around 13h.

The 50-th generation of an optimization with  $N = 100$  is illustrated in Figs. 2 and 3. Each square represents an individual in the generation, i.e., an RF cavity shape. For all of these individuals  $f_0 = 400.79$  MHz and  $\alpha \geq 90$ , i.e., both constraints from (1) are satisfied. In Fig. 2, the  $x$  and  $y$  coordinates represent the values of  $F_1$  and  $F_3$ , respectively, and the color shows the value of  $F_2$ . The functions  $F_1$  and  $F_3$  are inversely correlated, and  $F_1$  and  $F_2$  do not seem to be conflicting, i.e., for lower values of  $F_1$ , the values of  $F_2$  are usually also lower. In Fig. 3, the  $x$  and  $y$  coordinates represent the values of  $F_2$  and  $F_3$ , respectively, and the value of  $F_4$  is indicated by the color. The functions  $F_2$  and  $F_3$  are also inversely correlated.

The aperture radius  $R_i$  and the cell length  $2L$  have a high impact on the objective functions. The value of  $f_1$  is more sensitive to the changes in  $R_i$  and  $L$  than  $f_2$ . If  $f_1$  increases

<sup>1</sup> <https://scicomp.ethz.ch/wiki/Euler>

and gets closer to  $f_2$ , both  $F_1$  and  $F_2$  improve (note that  $f_0$  is fixed). A decrease in  $R_i$  typically increases  $f_1$ ,  $f_2$  and  $G_0 \cdot R/Q_0$ , which improves  $F_1$ ,  $F_2$  and  $F_4$ , but also increases  $R/Q_{\perp 1}$  and  $R/Q_{\perp 2}$ , so  $F_3$  deteriorates. A decrease in  $L$ , on the other hand, decreases  $G_0 \cdot R/Q_0$  and  $R/Q_{\perp 1}$ , and increases  $f_1$  and  $R/Q_{\perp 2}$  (which is typically much larger than  $R/Q_{\perp 1}$ ). Therefore, there is an inverse correlation between the sum of the impedances of the first dipole band ( $F_3$ ) and their frequency difference ( $F_2$  and  $F_1$ ). The  $x$  axis in Fig. 3 is a reshuffled version of the  $x$  axis in Fig. 2, based on  $f_2$ . Since both  $L$  and  $R_i$  have a high influence on  $f_1$  and only  $L$  highly impacts  $f_2$ , these two variables do not always move in the same direction and a wiggly behavior is observed in the Pareto front approximation (orange line) in Fig. 3.

The values of design variables and objective functions for a chosen individual are given in Table 2. Note that the design variables are allowed to go outside of the specified bounds if it helps to improve the objective functions. Some additional information is given in Table 3, and the shape of the RF cavity is shown in Fig. 4.

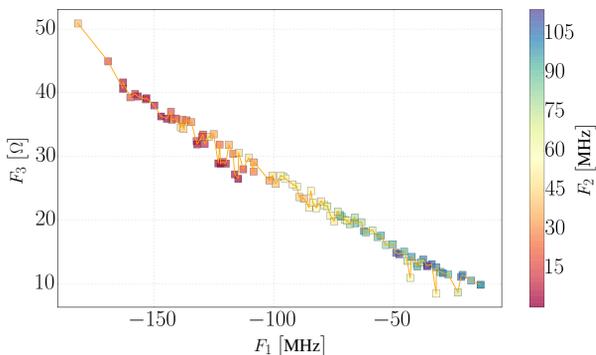


Figure 2: The relationship between  $F_1$ ,  $F_2$  and  $F_3$  for the individuals in the 50-th generation of the EA.

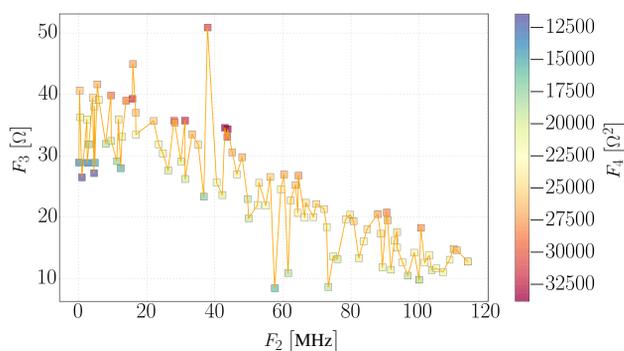


Figure 3: The relationship between  $F_2$ ,  $F_3$  and  $F_4$  for the individuals in the 50-th generation of the EA.

## CONCLUSIONS

In this paper an optimization algorithm for solving constrained multi-objective shape optimization problems for RF cavities was proposed and applied to the problem of optimizing the shape of the superconducting RF cavity for the FCC-ee-Z with respect to both the fundamental mode



Figure 4: The electric field of the fundamental mode in half of the chosen RF cavity.

Table 2: Description of the Chosen RF Cavity

Variable	$R_i$ [mm]	$L$ [mm]	$A$ [mm]	$B$ [mm]
Value	141.614	146.270	103.54	127.521
Variable	$a$ [mm]	$b$ [mm]	$R_{eq}$ [mm]	$\alpha$ [°]
Value	41.921	45.812	339.166	91.697
Objective	$F_1$ [MHz]	$F_2$ [MHz]	$F_3$ [Ω]	$F_4$ [Ω <sup>2</sup> ]
Value	-147.03	0.40	36.3	-2.13e3

Table 3: Additional Information on the Chosen RF Cavity

<b>TM<sub>010</sub></b>	$f_0 = 400.79$ MHz	$\frac{R}{Q_0} = 94.9$ Ω
	$\frac{E_{pk}}{E_{acc}} = 1.92$	$\frac{B_{pk}}{E_{acc}} = 4.16$ $\frac{\text{mT}}{\text{MV/m}}$
<b>TE<sub>111</sub></b>	$f_1 = 547.82$ MHz	$\frac{R}{Q_{\perp 1}} = 5.10$ Ω
<b>TM<sub>110</sub></b>	$f_2 = 548.22$ MHz	$\frac{R}{Q_{\perp 2}} = 31.2$ Ω

and the first dipole band. The proposed algorithm and its implementation could be used to optimize the shape of other axisymmetric RF structures, taking into account also the HOMs corresponding to arbitrary azimuthal mode numbers.

## ACKNOWLEDGEMENTS

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the project RI 814/29-1. The computations were executed on the Euler compute cluster of ETH Zurich at the expense of a Paul Scherrer Institut (PSI) grant.

## REFERENCES

- [1] V. Shemelin, S. Gorgi Zadeh, J. Heller, and U. van Rienen, “Systematical study on superconducting radio frequency elliptic cavity shapes applicable to future high energy accelerators and energy recovery linacs”, *Phys. Rev. Accel. Beams*, vol. 19, 2016. doi:10.1103/PhysRevAccelBeams.19.102002
- [2] S. Gorgi Zadeh, R. Calag, F. Gerigk, and U. van Rienen, “FCC-ee Hybrid RF Scheme”, in *Proc. IPAC’18*, Vancouver, BC, Canada, Apr-May 2018. doi:10.18429/JACoW-IPAC2018-MOPMF036
- [3] B. P. Xiao *et al.*, “Higher Order Mode Filter Design for Double Quarter Wave Crab Cavity for the LHC High Luminosity Upgrade”, in *Proc. IPAC’15*, Richmond, VA, USA, May 2015. doi:10.18429/JACoW-IPAC2015-WEPWI059

- Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.
- [4] J. Sekutowicz *et al.*, “Cavities for JLAB’s 12 GeV Upgrade”, in *Proc. PAC’03*, Portland, OR, USA, May 2003, paper TPAB085. <http://jacow.org/p03/PAPERS/TPAB085.PDF> and doi: 10.1109/PAC.2003.1289717
  - [5] P. Arbenz and O. Chinellato, “On solving complex-symmetric eigenvalue problems arising in the design of axisymmetric VCSEL devices”, *Appl. Numer. Math.*, vol. 58, no. 4, pp. 381-394, 2008. doi: 10.1016/j.apnum.2007.01.019
  - [6] O. Chinellato, “The complex-symmetric Jacobi–Davidson algorithm and its application to the computation of some resonance frequencies of anisotropic lossy axisymmetric cavities”, ETH Zurich, Diss. ETH No. 16243, 2005.
  - [7] Y. Ineichen *et al.*, “A fast and scalable low dimensional solver for charged particle dynamics in large particle accelerators”, *Comput. Sci. Res. Dev.*, vol. 28, no. 2, pp. 185-192, 2013. doi:10.1007/s00450-012-0216-2
  - [8] M. Kranjčević, A. Adelman, P. Arbenz, A. Citterio, and L. Stingelin, “Multi-objective shape optimization of radio frequency cavities using an evolutionary algorithm”, ArXiv e-prints, 2018. arXiv:1810.02990
  - [9] G. E. Alefeld, F. A. Potra, and Y. Shi, “Algorithm 748: Enclosing Zeros of Continuous Functions”, *ACM Trans. Math. Softw.*, vol. 21, no. 3, pp. 327-344, 1995. doi:10.1145/210089.210111

# BEAMLINE MAP COMPUTATION FOR PARAXIAL OPTICS

B. Nash\*, N. Goldring, J. P. Edelen, S. Webb  
Radiasoft LLC, Boulder, CO, USA

## Abstract

Modeling of radiation transport is an important topic tightly coupled to many charged particle dynamics simulations for synchrotron light sources and FEL facilities. The radiation results from the electron dynamics and then passes through beamlines, either directly to an experiment or may be recirculated back to interact with the electron beam in the case of an FEL oscillator. The Wigner function representation of these wavefronts have been described in the literature, and is the closest relation to the phase space description of charged particle dynamics. We describe this formalism and the computation of phase space maps using the code SRW, applying this to the case of a 4 crystal FELO 1:1 imaging beamline, resulting in a substantial speed-up in computation time.

## INTRODUCTION

Optical beamlines for radiation transport are crucial components of many scientific facilities. They may be used to transport radiation from the electron beam source in a synchrotron light source, or in recirculation optics in a free electron laser oscillator (FELO) to improve longitudinal coherence of the radiation. Numerous codes exist to model the radiation transport through the beamline elements typically either using a ray tracing, geometrical optics approach (e.g. SHADOW [1]), or a physical optics wavefront propagation approach (e.g. SRW [2]). The wavefront propagation captures more of the optical physics, but can be highly intensive computationally.

For FELO and synchrotron light source modeling, one requires many passes of radiation through a beamline, while varying either the initial conditions, or some beamline parameters. A more compact representation of a beamline is desired for such calculations.

In this paper, we consider a map based approach to beamline modeling that, once computed, allows a large range of initial conditions to be rapidly transported through the beamline. The formalism we use for representing the wavefront is based on the Wigner function, pioneered in x-ray optics by K. J. Kim [3]. Although the general non-linear map applied to the Wigner function may be quite complex, in the case of linear transport (so-called ABCD matrix, in the optics literature), the transformation is quite straight forward.

We provide a proof of principle for this method, applied to an FELO recirculation optics beamline, with radiation transport of increasing complexity. The beamline is shown in Fig. 1. The radiation transport starts at the end of the undulator and diffracts off the crystals in Bragg geometry. A single ideal lens is used for focussing the radiation back

to the beginning of the undulator for the next pass. We first consider Gaussian wavefronts, in which case, only transport of second moments is necessary. We set up the beamline in SRW and propagate the initial wavefront. We also compute the transfer matrix for the beamline to transport the Wigner function. For the Gaussian case, only the moments need be transported, but we also apply the method to the numerical Gaussian to check our calculation of the Wigner function. Finally, we consider a non-Gaussian case of an  $m_x = 2$  Hermite Gaussian mode, to show the generality of the method. In each case, we compare the SRW simulation to the linear transport of the Wigner function, to confirm that the method is sound. One important difference between our Wigner function transport and the wavefront propagation, is the absorption effect in the crystals. However, since ideal crystals don't affect the wavefront distribution, only the intensity, the two effects can be treated separately.

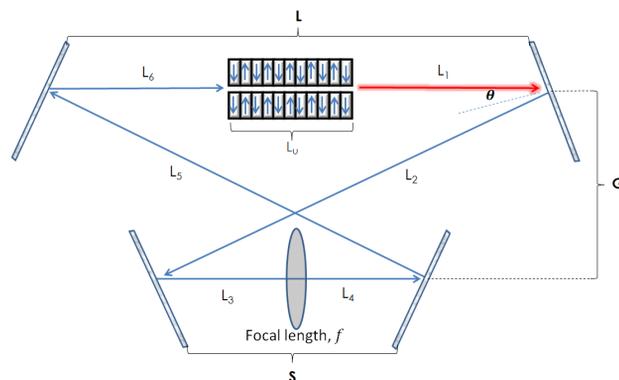


Figure 1: Four crystal FELO beamline schematic as described in reference [4].

## LINEAR PARAXIAL OPTICS PROPAGATION

We start by briefly reviewing the evolution equations for a wavefront with wavelength  $\lambda$  propagating through empty space. By this means we will set our notation, and clarify the issue of separability, which we will be assuming. We consider one component of an electric field travelling in the  $z$  direction, which we write as

$$E(x, y, z; t) = \bar{E}(x, y, z)e^{i(kz - \omega t)}, \quad (1)$$

where  $k = \frac{2\pi}{\lambda}$  and  $\omega = ck$ . The paraxial Helmholtz equation for the evolution of the electric field in free space is given by

$$\nabla_{\perp}^2 E + 2ik \frac{\partial E}{\partial z} = 0, \quad (2)$$

where

$$\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad (3)$$

\* bnash@radiasoft.net

and we have assumed

$$\frac{\partial^2 \bar{E}}{\partial z^2} \ll 2ik \frac{\partial \bar{E}}{\partial z} \quad (4)$$

(i.e. a slow change in transverse envelope relative to the wavelength). Equation (2) may be solved to yield Fresnel propagation of the electric field. In this paper, we will assume electric fields which satisfy the separability condition

$$E(x, y; z) = E_0 E_x(x; z) E_y(y; z), \quad (5)$$

where  $E_0$  is a constant with units of electric field.

For the case of separable electric fields, Fresnel propagation may be written in the following form.

$$E(x, y; z + l) = E_0 e^{i(kl - \frac{\pi}{2})} E_x(x; z + l) E_y(y; z + l), \quad (6)$$

where

$$E_x(x; z + l) = \frac{1}{\sqrt{\lambda l}} \int_{-\infty}^{\infty} E_x(x'; z) e^{\frac{ik}{2l}(x-x')^2} dx' \quad (7)$$

and likewise for  $E_y$ .

The angular representation  $E(\theta)$  for either component is given by

$$E(\theta) = \frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} E(x) e^{-\frac{2\pi i}{\lambda} \theta x} dx. \quad (8)$$

In this representation, the propagation is simply given by

$$E(\theta; z + l) = E(\theta; z) e^{\frac{2\pi i l}{\lambda} \left(1 - \frac{\theta^2}{2}\right)}. \quad (9)$$

We normalize the separate field components such that<sup>1</sup>

$$\int_{-\infty}^{\infty} E^*(x) E(x) dx = 1, \quad (10)$$

$$\int_{-\infty}^{\infty} E^*(\theta) E(\theta) d\theta = 1. \quad (11)$$

The second moments of the field distribution in coordinate and angular representations may now be calculated as

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 E^*(x) E(x) dx, \quad (12)$$

$$\langle \theta^2 \rangle = \int_{-\infty}^{\infty} \theta^2 E^*(\theta) E(\theta) d\theta. \quad (13)$$

We now introduce the Wigner function defined from the electric field,  $E(x)$ , as follows

$$W(x, \theta) = \frac{1}{\lambda} \int_{-\infty}^{\infty} E^*\left(x - \frac{\phi}{2}\right) E\left(x + \frac{\phi}{2}\right) e^{-\frac{2\pi i}{\lambda} \phi \theta} d\phi, \quad (14)$$

<sup>1</sup> Note that we have normalized the electric field in the same way as wave functions are normalized in quantum mechanics. In fact, much of the formalism of quantum mechanics may now be directly applied with the proviso that Planck's constant,  $\hbar$ , be replaced by the reduced wavelength  $\lambda = \lambda/2\pi$ .

where  $W(x, \theta)$  will be normalized as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x, \theta) dx d\theta = 1. \quad (15)$$

The Wigner function can be thought of as a probability distribution in phase space except for the fact that it may become negative. The second moments are given simply as

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 W(x, \theta) dx d\theta, \quad (16)$$

$$\langle \theta^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta^2 W(x, \theta) dx d\theta, \quad (17)$$

$$\langle x\theta \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\theta W(x, \theta) dx d\theta. \quad (18)$$

For propagation of the Wigner function, we briefly give a general formulation before reducing to the simplified linear case. To do so, we assume the existence of a Hamiltonian,  $H(x, \theta; z)$ . We assume that we stay within the paraxial approximation, and ignore the effect of absorption. The evolution equation for the Wigner function is then given as follows [5]

$$\frac{\partial W(x, \theta; z)}{\partial z} = [W, H]_*, \quad (19)$$

where the Moyal bracket is defined for arbitrary phase space functions  $f$  and  $g$  as

$$[f, g]_* = \frac{1}{i\lambda} (f * g - g * f) \quad (20)$$

and the Moyal star is given by

$$* = e^{\frac{i\lambda}{2} (\overleftarrow{\partial}_x \overrightarrow{\partial}_\theta - \overleftarrow{\partial}_\theta \overrightarrow{\partial}_x)} \quad (21)$$

with the arrows representing action of the derivative, either to the left or right, depending on arrow orientation.

Fortunately, in the case of a quadratic Hamiltonian, evolution of the Wigner function is much more simple and intuitive. Only first order in  $\lambda$  is needed, and the Moyal bracket reduces to the Poisson bracket giving classical evolution (again using the quantum/classical mechanics analogy). One finds that the motion in phase space is a linear transformation. These considerations allow us to formulate our approach. In particular, consider a paraxial beamline where the geometric optics will be defined by a transfer matrix  $M$  acting on the phase space vector  $\vec{z}$ :

$$\vec{z}_f = M \vec{z}_i, \quad \vec{z} = \begin{pmatrix} x \\ \theta \end{pmatrix}. \quad (22)$$

The Wigner function will evolve along this beamline according to

$$W_f(\vec{z}) = W_i(M \vec{z}). \quad (23)$$

The electric field may be reconstructed from the Wigner function as follows [6]

$$E^*(x) E(0) = \frac{1}{\lambda} \int_{-\infty}^{\infty} W\left(\frac{x}{2}, \theta\right) e^{\frac{2\pi i}{\lambda} x \theta} d\theta. \quad (24)$$

## Hermite Gaussian Modes

The Hermite Gaussian modes will satisfy the separability condition (5). These modes are defined by

$$E_m(x) = E_{0,m} H_m \left( \frac{\sqrt{2}x}{w(z)} \right) e^{-i \frac{kx^2}{2q(z)}}, \quad (25)$$

where  $E_{0,m}$  is the normalization factor resulting in the normalization condition (10) [7].  $q(z)$  is the complex beam parameter defined in terms of the beam radius of curvature,  $R(z)$ , and beam spot size,  $w(z)$ :

$$\frac{1}{q(z)} \equiv \frac{1}{R(z)} - i \frac{\lambda}{\pi w^2(z)}. \quad (26)$$

In empty space the beam spot size evolves as

$$w(z) = w_0 \sqrt{1 - \left( \frac{z}{z_R} \right)^2} \quad (27)$$

with  $z_R$  the Raleigh length, given by

$$z_R = \frac{\pi w_0^2}{\lambda}, \quad (28)$$

and  $w_0$  is the beam spot size at waist. The radius of curvature  $R(z)$  evolves as

$$R(z) = z \left[ 1 + \left( \frac{z_R}{z} \right)^2 \right]. \quad (29)$$

For the case  $m = 0$ , we have a Gaussian field given explicitly by

$$E(x) = \frac{1}{(2\pi\sigma_x^2)^{1/4}} e^{-\frac{x^2}{4\sigma_x^2}}, \quad (30)$$

where  $\sigma_x$  is the RMS size of the distribution. One can relate  $\sigma_x$  to the beam spot size at waist via the equation  $w_0 = 2\sigma_x$ . This Gaussian distribution will satisfy the diffraction limit which implies the following relation between the RMS size and divergence

$$\sigma_x \sigma_\theta = \frac{\lambda}{4\pi}. \quad (31)$$

For the Gaussian case, the Wigner function propagation may be reduced to a transformation of the second moments. In particular, under a transfer matrix  $M$ , the second moments will evolve as follows:

$$\Sigma_f = M \Sigma_i M^T, \quad (32)$$

where the second moment matrix is given by

$$\Sigma = \begin{pmatrix} \langle x^2 \rangle & \langle x\theta \rangle \\ \langle x\theta \rangle & \langle \theta^2 \rangle \end{pmatrix}. \quad (33)$$

## MOMENT CALCULATIONS FOR FELO BEAMLINE AND COMPARISON WITH SRW

We have set up the four crystal beamline as described in reference [4]. Note however that we've used even simpler optics, with just a single ideal lens, located at the midpoint of the beamline. See Fig. 1 for the schematic. The parameters used for this example calculation are as follows. The undulator length,  $L_u$  is 10 meters. The total length,  $L$ , is 100 meters. The crystal diffraction angle,  $\theta$ , is  $\pi/8$ . The length of the lower leg of the beamline,  $S$ , is 3 meters. The other parameters are then determined by geometric relations and are as follows.  $L_1$  and  $L_6$  are 45.0 meters.  $L_2$  and  $L_5$  are 72.8 meters.  $L_3$  and  $L_4$  are 1.5 meters. The distance between the two legs of the beamline,  $G$ , is 51.5 meters.

The reflecting crystals were chosen to be diamond with a d-spacing of 0.892 Å. The crystal thickness was 10 millimeters. The real and imaginary parts of the 0-th Fourier component of crystal polarizability were  $-0.217 \times 10^{-4}$  and  $0.280 \times 10^{-7}$  respectively. The real and imaginary parts of the next Fourier component of crystal polarizability was  $-0.544 \times 10^{-5}$  and  $0.259 \times 10^{-7}$  respectively.

In order to compare to the analytical approach, we need to compute the transfer matrix  $M$  for the beamline. The effect of the crystal is only to decrease the electric field amplitude, and not change the wavefront distribution. Thus, disregarding this absorption effect, we may ignore the crystals in our analytical calculation. The transfer matrix may be calculated using the matrices for a drift,  $M_d$ , and for an ideal lens,  $M_f$ . These are given by

$$M_d = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix}, \quad (34)$$

$$M_f = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}, \quad (35)$$

where  $l$  is the drift length and  $f$  is the focal length for the ideal lens. The total transfer matrix  $M_T$  is then given by multiplying the component matrices for the beamline. The transfer matrix for the simple single lens beamline is given by

$$M_T = M_d M_f M_d \quad (36)$$

$$= \begin{pmatrix} 1 - \frac{l}{f} & 2l - \frac{l^2}{f} \\ -\frac{1}{f} & 1 - \frac{l}{f} \end{pmatrix}. \quad (37)$$

In the case that  $f = l/2$  the transfer matrix simplifies to

$$M_T = \begin{pmatrix} -1 & 0 \\ -\frac{2}{l} & -1 \end{pmatrix}. \quad (38)$$

To look at the effect of focusing errors, we have allowed the focal length to vary from the nominal value by defining  $f = \frac{l}{2} (1 + \frac{df}{f})$ . For varying values of  $\frac{df}{f}$ , we have set up a computation in SRW and also compared it to this moment

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

calculation. The results are shown in Fig. 2. The SRW calculation is a wavefront calculation. In order to compute the divergence, we convert the wavefront to angular coordinates, and then compute the RMS of the distribution. The initial RMS beam size for the Gaussian chosen was 10 micrometers. Applying Eq. (31), we find an initial divergence at the waist of 0.987 microradians.

We now give more details for the SRW simulation. The initial Gaussian wavefront was represented on a grid of size  $42 \times 42$ . Each beamline element requires specification of propagation parameters. These were chosen for drift, crystal, and lens elements to ensure accuracy and avoid excessive computational overhead. The final grid size needed to capture the full Gaussian ended up ranging from  $42 \times 42$  to  $52 \times 52$  as  $\frac{df}{f}$  was varied.

SRW was run on the Radiosoft Jupyter server [8]. The time required for the full SRW calculation was 12.8 seconds and the time required for the matrix computation was 0.21 seconds. Note that the time for the SRW calculation includes the transformation of the wavefront from coordinate to angular representation (see Eq. (8)) in order to compute RMS divergences. The matrix transformation amounts to a speedup of a factor of 608 over the SRW calculation.

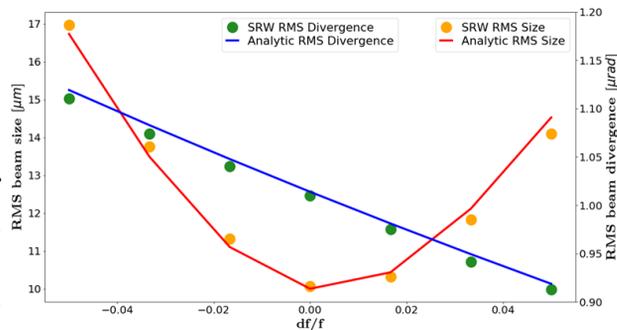


Figure 2: Comparison between RMS size and divergence as a function of focal length for the four crystal FEL beamline.

## NUMERICAL WIGNER FUNCTION EVOLUTION AND SRW COMPARISON

In order to consider non-Gaussian wavefronts, we need to move beyond calculation of second moments. This requires numerical calculation of the Wigner function and inversion to reconstruct the electric field after propagation using Eq. (24). To demonstrate this procedure and explore the efficiency of the algorithms, we have implemented the Wigner function calculation and inversion formulae numerically. We first consider Gaussian distributions and then give results for an  $m_x = 2$  Hermite Gaussian mode. The beams are assumed to start at a waist and are then propagated through the four crystal FEL recirculation beamline. Because of the separability condition for Hermite Gaussian modes, we can propagate the 2D wavefronts in SRW and compare the results between the map method and SRW for each individual component.

## Gaussian Wigner Function

We now consider the results for Gaussian wavefronts. We start with a Gaussian wavefront at a waist with an RMS beam size of  $10 \mu\text{m}$ . We consider x-rays of energy  $10 \text{ keV}$  which corresponds to a wavelength of  $1.23 \text{ \AA}$ . For the diffraction limited Gaussians satisfying Eq. (31), the divergence is found to be  $0.987 \mu\text{rad}$ . We construct such an initial Gaussian wavefront for input to SRW for propagation. We propagate the Gaussian through the FEL beamline we have described. The initial grid size was  $2100 \times 2100$  and the grid size for the propagated wavefront was  $3150 \times 3150$ . The initial and final wavefront intensity distributions are shown in Fig. 3. Because of the one to one focusing optics, the initial and final intensity distributions are equal although the wavefronts themselves will differ. Due to the absorption from the crystals in the beamline, the final intensity computed with SRW is reduced from the initial intensity. We find a reflection coefficient of  $R = 0.96$  where we define  $R$  as

$$R = \frac{\Phi_f}{\Phi_i} = \frac{\iint I_f(x, y) dx dy}{\iint I_i(x, y) dx dy}, \quad (39)$$

i.e.  $\Phi_i$  and  $\Phi_f$  are the initial and final total fluxes of the distribution.

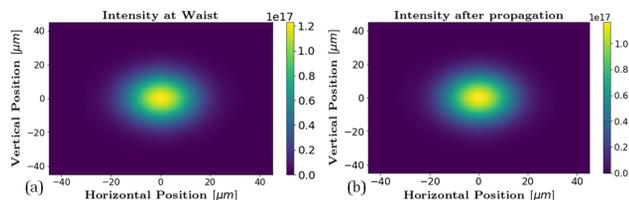


Figure 3: Gaussian intensity distribution at a) waist and b) after SRW propagation through FEL beamline.

In order to apply the map method, we have taken a horizontal slice of the initial electric field generated by SRW. We normalize this according to Eq. (10). We now compute the numerical Wigner distribution which will also be Gaussian. Next we apply the transfer matrix for the FEL recirculation optics, Eq. (38). The transformed Wigner function is then redeposited onto the initial grid. Initial transformed and redeposited Wigner functions are shown in Fig. 4. Note that for our angular variable, we have used  $q = \theta/\lambda$ .

Given the propagated Wigner function, we reconstruct the electric field using Eq. (24) and compare this to the results found from SRW. We see good agreement which is presented in Fig. 5.

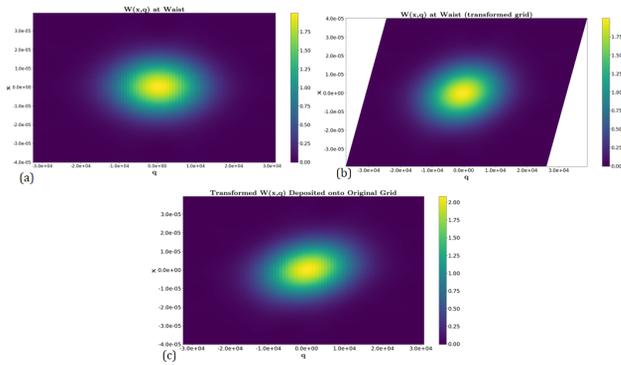


Figure 4: Numerically computed Wigner distributions a) at waist, b) after application of FELo beamline transfer matrix, and c) after redeposition of b) on initial grid.

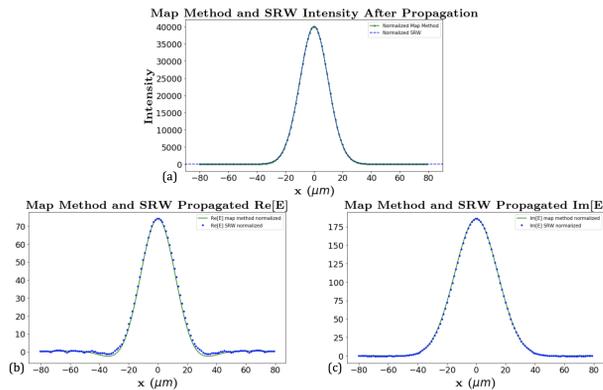


Figure 5: Comparison of SRW and map method propagation results for a) intensity, b)  $\text{Re}[E]$ , and c)  $\text{Im}[E]$ .

### $m_x = 2$ Hermite Mode Wigner Function

We now present the results for an  $m_x = 2$  Hermite Gaussian mode. As in the Gaussian case, we start at a waist and subsequently propagate the mode, using SRW, through the four crystal FELo beamline, using the same grid sizes as in the Gaussian case. We again consider x-rays of energy 10 keV, corresponding to a wavelength of 1.23 Å. Next, the exact same method as used in the Gaussian case was applied for the  $m_x = 2$  mode. The Wigner function was computed and transformed with the transfer matrix and the distribution was redeposited onto the initial grid. This is shown in Fig. 7. Note in particular the large region in which the Wigner function takes on negative values outside the central peak. Finally, the electric field was again reconstructed using Eq. (24). Again, due to the one to one focusing beamline, the intensity distributions shown in Fig. 6 are equal whereas the electric field distributions shown in Fig. 8 are not and one can see the inversion from the transfer matrix in the field components. The final comparison between intensity and real and imaginary parts of the electric field for the two different methods are shown in Fig. 9. We see that the map method has agreed well with the SRW calculation.

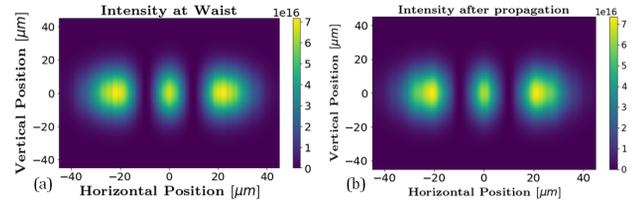


Figure 6:  $m_x = 2$  Hermite mode intensity distribution at a) waist and b) after SRW propagation through FELo beamline.

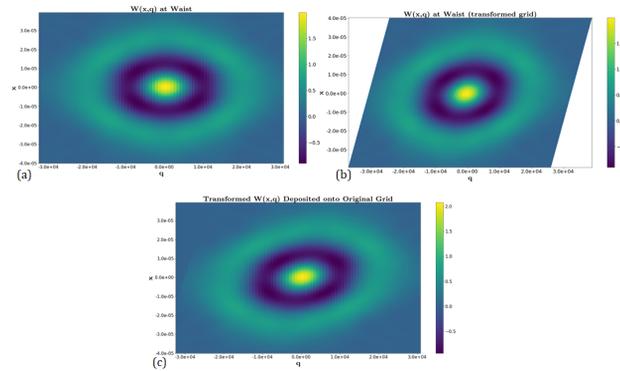


Figure 7: Numerically computed Wigner distributions a) at waist, b) after application of FELo beamline transfer matrix, and c) after redeposition of b) on initial grid.

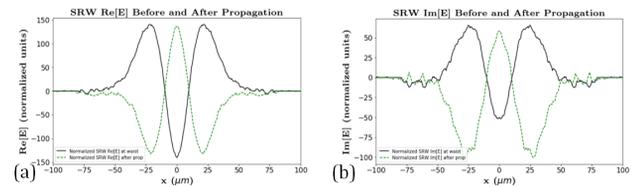


Figure 8: Horizontal electric field at waist and after SRW propagation through FELo beamline. a)  $\text{Re}[E]$  b)  $\text{Im}[E]$ .

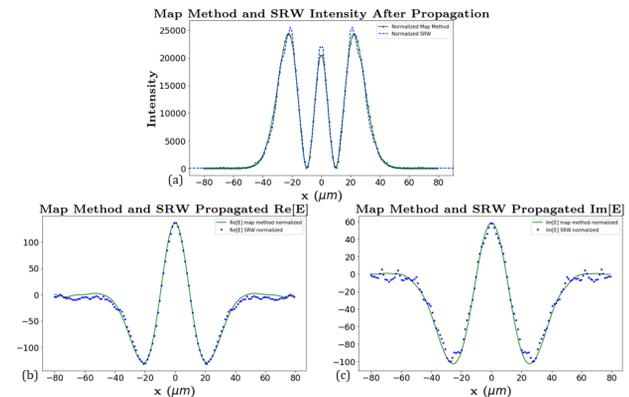


Figure 9: Comparison of SRW and map method propagation results for a) intensity, b)  $\text{Re}[E]$ , and c)  $\text{Im}[E]$ .

### Timing Comparison Between SRW and Map Method Calculations

As with the moment calculations, the aforementioned calculations were all performed on the Radiosoft Jupyter server. The initial wavefront is computed in the same

way for both methods and took several seconds for the grid size used. The propagation through the FELo beamline in SRW took 43 seconds and was approximately the same for both Gaussian and non-Gaussian wavefronts. As with the moment calculations, SRW propagation parameters were tuned to achieve adequate results. We tried to ensure good resolution on the resulting electric fields while minimizing the overall grid sizes.

The calculation of the Wigner function and propagation for the method took 38 seconds and was approximately the same for both Gaussian and non-Gaussian wavefronts. As can be seen, the algorithms as implemented are comparable in speed. A large amount of the time for the map method came from the deposition of the transformed Wigner function. A more efficient algorithm for that process is conceivable. In addition we note that for a longer and more complex beamline, the SRW calculation would increase in time whereas the map method would not, once the transfer matrix has been computed.

## CONCLUSION

We have demonstrated the use of a map based method for radiation transport through a beamline. We applied this method to a simple 4 crystal recirculation beamline which may be used for a free electron laser oscillator. For Gaussian wavefronts, only moment transport is required. We showed that such moment transport using a transport matrix for the beamline agrees very well with a more complete wavefront computation using the SRW code. The moment transport is substantially faster, by a factor of 608 for this particular case. In order to apply the method to non-Gaussian wavefronts, we computed the numerical Wigner function from the wavefront and demonstrated the linear transformation in phase space of this Wigner function. We did this both for a Gaussian and for an  $m_x = 2$  Hermite Gaussian mode. We then reconstructed the electric field and showed good agreement with SRW simulation. We considered cases in which the field profile is separable so that we could work with 2-D phase space instead of 4-D phase space. In this case, the map method was found to be comparable in speed with the SRW calculation. However, as discussed, we expect that we may be able to speed up the map calculation, and in addition it would be substantially faster in the case of a beamline containing a larger number of optical elements. Regardless of the complexity, transport of the wavefront through the beamline is all contained within a single transfer matrix. Finally, we point out that partial coherence may be included in the Wigner function and the transport of this Wigner function is no more complicated than the case of a fully coherent beam.

Next steps include development of the corresponding routines for 4-D phase space, computation of the transfer matrix for arbitrary beamlines using a ray tracing code such as SHADOW, inclusion of partial coherence in the Wigner function, treatment of polarization, and inclusion of elements such as apertures or elements where the linear transport is not applicable and require special direct transformation of the Wigner function. In this latter case, we envisage using matrices for partial maps between such elements, and still expect a substantial speed up over the full wavefront propagation.

For the FELo simulations, we expect to be able to combine these simulations with a rapid FEL model to allow us to model the build-up of coherence within a manageable computation.

## ACKNOWLEDGEMENTS

We acknowledge David Bruhwiler and Gabriel Marcus for valuable discussions on the subjects of free electron lasers and radiation propagation. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0018571.

## REFERENCES

- [1] M. Sanchez del Rio, N. Canestrari, F. Jiang, and F. Cerina, "SHADOW3: a new version of the synchrotron X-ray optics modelling package", *Journal of Synchrotron Radiation*, vol. 18, no. 5, pp. 708–716, 2011. doi:10.1107/S0909049511026306
- [2] O. Chubar and P. Elleaume, "Accurate and Efficient Computation of Synchrotron Radiation in the Near Field Region", in *Proc. EPAC'98*, Stockholm, Sweden, paper THP01G, pp. 1177-1179, Jun 1998.
- [3] K.-J. Kim, "A new formulation of synchrotron radiation optics using the Wigner distribution", LBL-20630, Berkeley, CA, USA, Nov 1985.
- [4] K.-J. Kim and Y. V. Shvyd'ko, "Tunable optical cavity for an x-ray free-electron-laser oscillator", *Phys. Rev. Special Topics - Accelerators and Beams*, vol. 12, pp. 030703, 2009. doi:10.1103/PhysRevSTAB.12.030703
- [5] T. Curtwright *et al.*, "Features of time-independent Wigner functions", 1997. arXiv:hep-th/9711183
- [6] I. V. Bazarov, "Synchrotron radiation representation in phase space", 2011. arXiv:1112.4047
- [7] A. Siegman, *Lasers*, University Science Books, Palo Alto, CA, 1986.
- [8] <https://jupyter.radiasoft.org>

# COMPUTATIONAL BEAM DYNAMICS REQUIREMENTS FOR FRIB\*

P. N. Ostroumov<sup>†</sup>, Y. Hao, T. Maruta, A. Plastun, T. Yoshimoto, T. Zhang, Q. Zhao  
Facility for Rare Isotope Beams, Michigan State University, East Lansing, MI, USA

## Abstract

The Facility for Rare Isotope Beams (FRIB) being built at Michigan State University moved to the commissioned stage in the summer of 2017. There were extensive beam dynamics simulations in the FRIB driver linac during the design stage. Recently, we have used TRACK and IMPACT simulation codes to study dynamics of ion beam contaminants extracted from the ECR together with main ion beam. The contaminant ion species can produce significant uncontrolled losses after the stripping. These studies resulted in development of beam collimation system at relatively low energy of 17 MeV/u and room temperature bunchers instead of originally planned SC cavities. Commissioning of the Front End and the first 3 cryomodules enabled detailed beam dynamics studies experimentally which were accompanied with the simulations using above-mentioned beam dynamics codes and envelope code FLAME with optimizers. There are significant challenges in understanding of beam dynamics in the FRIB linac. The most computational challenges are in the following areas: (1) Simulation of the ion beam formation and extraction from the ECR; (2) Development of the virtual accelerator model available on-line both for optimization and multi-particle simulations. The virtual model should include realistic accelerator parameters including device misalignments; (3) Large scale simulations to support high-power ramp up of the linac with minimized beam losses; (4) Extension of the existing codes for large scale simulations to support tuning of fragment separators for selected isotopes.

## INTRODUCTION

The Facility for Rare Isotope Beams (FRIB) currently being built at Michigan State University (MSU) is the next generation facility for rare isotope science. The FRIB includes a high-power driver accelerator, a target, and fragment separators. The FRIB driver linac will provide stable nuclei accelerated to 200 MeV/u for the heaviest uranium ions and higher energies for lighter ions with 400 kW power on the target [1]. FRIB features a continuous wave (CW) linac with a room-temperature 0.5 MeV/u front-end followed by a superconducting radiofrequency (SRF) linac consisting of 4 types of niobium cavities. The first SRF section includes quarter-wave resonators (QWR) with  $\beta_{OPT}=0.041$  and  $\beta_{OPT}=0.085$  which accelerate ion beams from 0.5 MeV/u to  $\sim 20$  MeV/u at the charge stripper. The optimal beta is defined as relative velocity,  $\beta_{OPT}$ , where the maximum transit time factor T is achieved. The ion beams are further accelerated with the half-wave resonators

\* Work supported by the U.S. Department of Energy Office of Science under Cooperative Agreement DE-SC0000661 and the National Science Foundation under Cooperative Agreement PHY-1102511, the State of Michigan and Michigan State University.

<sup>†</sup> ostroumov@frib.msu.edu

(HWR) of  $\beta_{OPT}=0.29$  and  $\beta_{OPT}=0.53$ . Total 316 SRF cavities are used for acceleration to the design energy of 200 MeV/u for heaviest uranium ions. 400 kW accelerated ion beams will be delivered to the target which is followed by a large acceptance high resolution fragment separator. While many isotopes will be studied in the in-flight experiments, FRIB will use upgraded National Superconducting Cyclotron Laboratory (NSCL) facilities to prepare and re-accelerate stopped isotopes up to 12 MeV/u. Currently, the re-accelerator (ReA3), consisting of a radiofrequency quadrupole (RFQ) and a superconducting radio-frequency (SRF) linac provides 3 MeV/u rare isotope beams for experiments.

The layout of the FRIB is shown in Fig. 1.

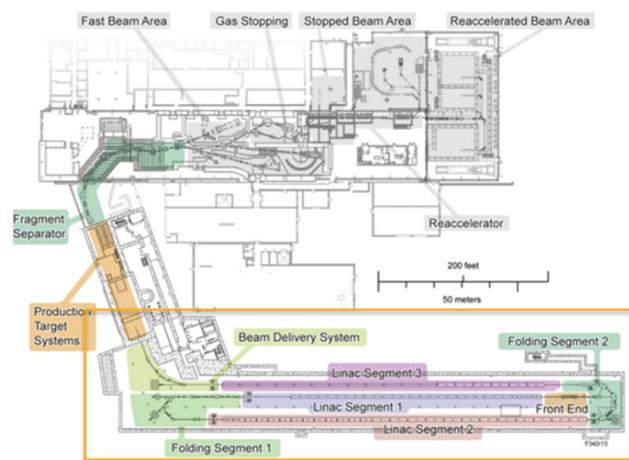


Figure 1: Layout of the FRIB driver accelerator, target, fragment separator, re-accelerator and existing infrastructure. The driver linac consist of three straight segments, LS1, LS2, LS3 and two folding segments FS1 and FS2.

## DRIVER LINAC

Due to CW mode of the FRIB driver linac, the final beam power of 400 kW can be achieved with a low beam current which is below 1 emA for all ion species. The space charge effects are mostly negligible over the entire linac except in the ion source and low energy beam transport (LEBT). FRIB linac will be equipped with the state-of-the art high intensity superconducting ECR ion source capable to produce required intensity of heaviest ions in a single charge state. However, to operate the SC ECR with a large margin the linac was designed to accelerate two charge states of heaviest ions (e.g.  $U^{33+}$  and  $U^{34+}$ ) up to the stripper [2]. To meet power requirement, a multiple charge state acceleration for the most ions heavier than argon is foreseen after the stripping at  $\sim 17$  MeV/u [3].

In the FRIB design stage we have evaluated beam dynamics of the most critical beam of uranium with high statistics simulations in realistic conditions with all types of

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

errors and misalignments using the IMPACT [4] and TRACK [5] codes on high performance computers.

### End-to-end Particle Tracking

The end-to-end simulation started with “realistic” distribution restored from emittance measurements of uranium beam extracted from the VENUS ion source [6]. The “realistic” distribution of two-charge-state uranium is then tracked through the Front End, LS1 and then five charge states (from  $U^{76+}$  through  $U^{80+}$ ) were selected after the lithium stripper followed by another two acceleration segments [3]. The final beam phase space distributions at the fragmentation target are shown in Fig. 2. Beam-on-target requirements are met even for the most challenging multi-charge state uranium beam (e.g. >96 % of particles are within 1 mm diameter of beam spot size, all particles are within angular spread of  $\pm 5$  mrad).

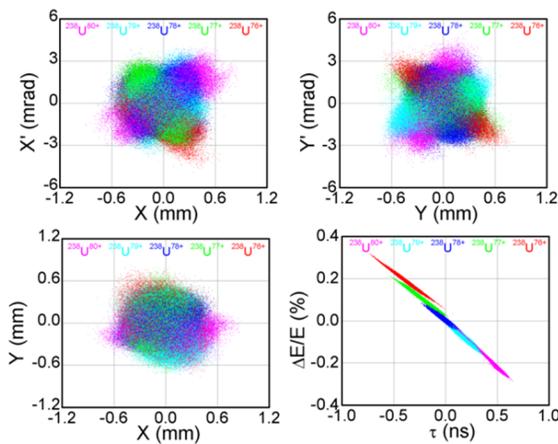


Figure 2: Transverse phase space plots (top), physical beam size (bottom-left), and longitudinal phase space (bottom-right) distributions on the target for 5-charge-state uranium without machine errors. Different colours represent 5 charge states of uranium.

Beam simulation studies with machine errors were performed to evaluate the linac performance under more realistic conditions [3]. When the element displacements are introduced, especially the misalignment of superconducting solenoids within  $\pm 1$  mm, correctors must be set properly for the beam steering using BPMs’ readings otherwise beam cannot be threaded through the linac. A total of 200 random seeds combining the errors were used in the multi-charge-state uranium beam simulations. In each seed run, one million particles were tracked from the exit of RFQ through the three linac segments to the fragmentation target. Figure 3 illustrates the maximum beam envelope (blue) at each longitudinal location of the 200 seeds together with the beam envelope without errors (green) and linac radial apertures (red). Beam evaluation results with machine errors show that the beam envelopes are well within apertures. Beam envelope growth is mainly due to misalignment (correctors were on) of the accelerator components. RF errors cause significant longitudinal emittance growth but it is not coupled into the transverse motion. No uncontrolled beam losses are observed with the nominal

errors. Although errors impact the beam distribution on target, beam-on-target requirements can be easily satisfied by final focusing quadrupoles and corrector magnets.

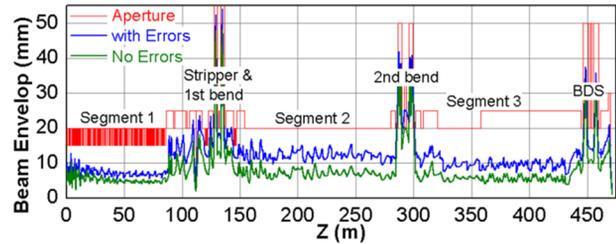


Figure 3: Beam envelopes along linac: beam element radial aperture in red, beam envelope without errors in green and with machine errors in blue.

### Recent Beam Dynamics Studies

The average charge state of the ion beam after stripper strongly depends on the ions’ atomic numbers. Therefore, after the stripper, contaminant ions will have different charge-to-mass ratios than the main beam. For example, for uranium,  $q/A=78/238=0.328$  while for fully-stripped nitrogen  $q/A=0.5$ . The intensity of the contaminants can be as high as  $\sim 1\%$  of the main beam power at the stripper which is  $\sim 40$  kW for the FRIB 400 kW design power. The contaminant beam power impinging onto the charge stripper can be up to several hundred Watts and the loss of these contaminants must be controlled.

To avoid uncontrolled losses, we have designed a set of collimators installed along the FS1 that can intercept contaminant ions at relatively low energy of  $\sim 17$  to 20 MeV/u depending on the ion species and localize losses in the designated areas with appropriate shielding as described in our recent publication [7]. The set of 10 collimators and charge selection slits slightly reduce the acceptance of the FS1 to avoid any beam losses in the LS2 and LS3. Figure 4 shows horizontal and vertical phase planes at the entrance of the LS2. As can be seen, the design beam emittance without any errors and other imperfections is well inside the FS1 acceptance (dark grey area). The latter is smaller than LS2 acceptance (pale grey area). Figure 4 shows the emittance and acceptance calculated for  $^{238}U^{78+}$ . Each uranium charge state has slightly different orientation of the beam phase space portraits and acceptances.

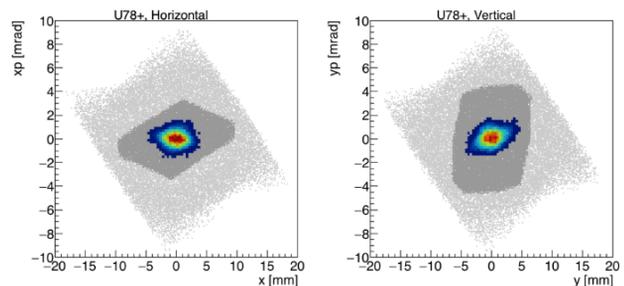


Figure 4: Transverse acceptance of the LS2 (pale grey), acceptance of the FS1 formed with all collimators (dark grey) and the design beam emittance at the entrance of the LS2.

## HIGH-LEVEL PHYSICS CONTROLS SOFTWARE

The high-level physics controls software is under active development for FRIB accelerator system and it is mainly Python-based software ecosystem known as **Physics High-level Applications and Toolkit for Accelerator System, PHANTASY**. It features the systematic solution to perform high-level physics controls in an efficient way and includes the following main components:

- The whole accelerator is represented as a hierarchical data structure. All the optics devices are modelled with unified software application programming interfaces (APIs), such that the user can talk to the devices in an object-oriented programming (OOP) way, rather than to the distributed power supply controls variables. PHANTASY provides Python classes to make this procedure standard and easy to do, the user can abstract FRIB accelerator (which is defined by the parameter named 'machine') with different segments ('segment'), e.g. LEBT, MEBT, etc. All the description of the accelerator is maintained by another package, which is updated once the machine configuration is changed.
- Interactive scripting environment for high-level physics controls. Once the accelerator is abstracted to OOP level, the users can implement the tuning algorithms to achieve various goals.
- Virtual accelerators solution. This is truly the same accelerator as the real FRIB accelerator from the view of EPICS controls; all the devices are named the same as FRIB accelerator. Powered by the so-called model engine, i.e. code to simulate the accelerator behaviour, the virtual accelerator supports testing of tuning algorithms.
- Interface to the different model engines. For instance, FLAME [8], IMPACT, TRACK etc, are developed or under development.
- Interface to different web services. For instance, 'channelfinder', which is a controls variables directory service, 'unicorn', which is home developed REST web service for unit conversion between physics and engineering fields.
- GUI applications. Finally, tuning algorithms are developed into a GUI application with PyQt5 [9], then all the users, including operators, can reach these automatically deployed apps from any workstation in the control room.

### Example of Two-Charge State Beam Tuning

Due to the different synchronous phases for each charge state, the bunch centers in the phase space oscillate with respect to each other and result to effective emittance growth, as shown in Fig. 5. The two-charge-state ion beam should be tuned to overlap phase space images at the stripper location both in the transverse and longitudinal phase planes to minimize the emittance growth due to scattering and energy straggling as shown in Fig. 5.

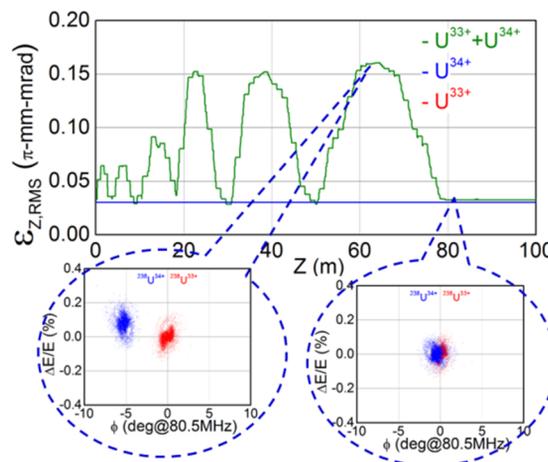


Figure 5: Longitudinal emittance of two-charge-state uranium beam along Segment 1 together with sampled particles (33+ red, 34+ blue) in longitudinal phase space.

### Beam Central Trajectory Correction

Central trajectory correction could be achieved either by global parameter optimization or applying Orbit-Response-Matrix (ORM). The latter is being widely used on light sources like synchrotron and free-electron laser facilities. At FRIB, the ORM based central trajectory correction application is originally developed against virtual accelerator. The response matrix could be measured by altering the selected correctors one by one, and meanwhile, keep the BPM or wire-scanner readings for the beam central positions. The polynomial fitting can give the correspondent term of each corrector, in both horizontal and vertical direction. Then, with the trajectory to correct, the corrector settings could be calculated based on the inverse matrix of the measured ORM, usually, a singular-value decomposition (SVD) algorithm is applied to robustly figure out the inverse matrix. All these operations are being done in a user-friendly way, from the seamlessly integrated high-level physics applications.

### Commissioning of Front End and First Three Cryomodules

A set of on-line physics applications have been developed for the setting of LEBT; optimal tuning of the Multi-Harmonic Buncher (MHB); beam central trajectory correction in LEBT, MEBT and cryomodules; quadrupole or solenoid scan for profile measurements and evaluation of rms emittance; longitudinal emittance rms evaluation by rotating beam image in the longitudinal phase space and measuring bunch length. Using these applications we were able to accelerate and characterize 33  $\mu$ A argon beam up to 2.3 MeV/u through the first 3 cryomodules without notable beam losses as shown in Fig. 6. The signals from 15 BPMs along the MEBT, 3 cryomodules and diagnostics station are shown in Fig. 7. The results of quadrupole scan and beam rms size from the downstream profile monitor are shown in Fig. 8. Similar data measured with a silicon detector for the evaluation of the longitudinal emittance are shown in Fig. 9.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

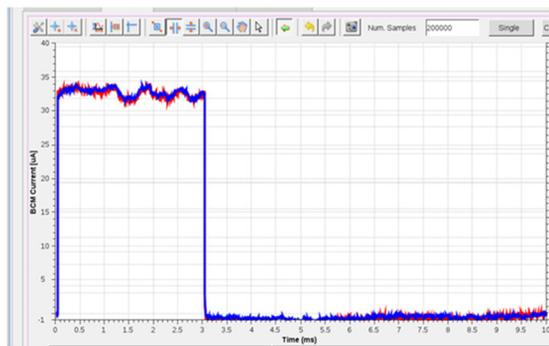


Figure 6: Beam peak current upstream (red) and downstream (blue) of the first three cryomodules. Pulse length is 3 ms at 100 Hz repetition rate.

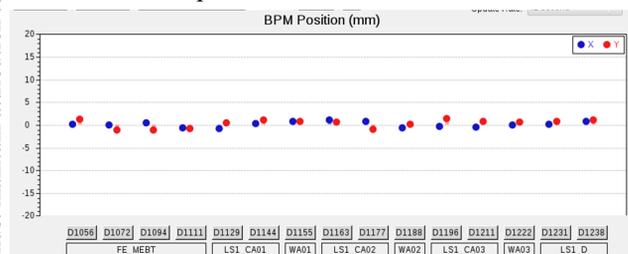


Figure 7: BPM readings. Beam center deviation is within  $\pm 1.5$  mm.

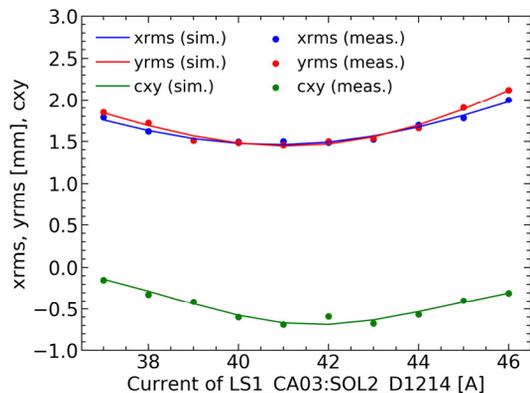


Figure 8: Beam rms sizes and XY coupling term as a function of the quadrupole current in MEBT.

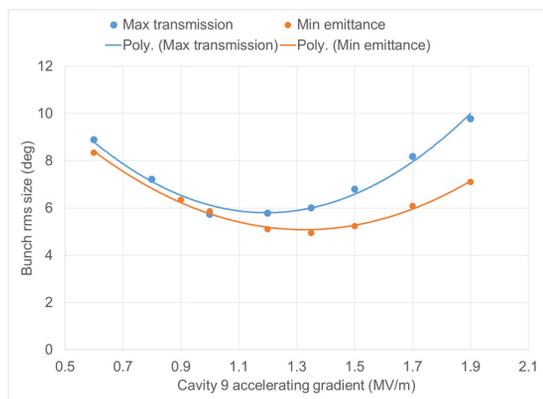


Figure 9: Beam longitudinal rms size as a function of the cavity accelerating gradient for two cases of the MHB tuning: (1) maximum transmission (blue) and (2) minimum longitudinal emittance (red).

## ECR ION SOURCE SIMULATIONS

Due to complexity of physical processes, there is no fully self-consistent model of ECR ion sources (ECRIS). The existing computer models of ECRIS are based on various simplifications and use some empirical parameters to reproduce experimental data. The most comprehensive review of the current status of ECRIS simulations is given in [10].

We have decided to build a CST Particle Studio [11] model for the room temperature ECRIS in order to include two major factors strongly effecting the beam dynamics such as magnetization and space charge effects in the multi-component ion beam extracted from the ECRIS. The model includes 3D fields of the ion source but does not include plasma processes and stripping of ions. The simulation starts by generating a distribution of various ion charge states inside the resonance region of the ECR plasma. Then, the ions are tracked to the location of the extraction aperture. The analysis of ions distribution in the extraction aperture shows that rms beam parameters in the LEBT are mostly defined by the geometry of the extraction system and does not strongly depend upon the method how the ions are generated inside the plasma. Multiple ion charge states and ion species are extracted from the ECR assuming a flat plasma meniscus. The further tracking is performed in the presence of beam space charge and an external solenoidal field. These simulations show that a hollow beam structure in the real space (see Fig. 10) is formed due to the different focal length of the solenoid focusing for different ions and the presence of space charge. In addition, due to the large beam size in the solenoid, there is an effect of spherical aberrations.

The ion beam in the ECRIS is magnetized as a consequence there is a strong correlation term in the  $x$ - $y'$  and  $y$ - $x'$  phase planes after the beam extraction as shown in Fig. 11. We have also applied TRACK code for the 4D beam dynamics simulations of multi-component, multi-charge ion beam in the LEBT. Figure 12 shows the measured beam images along the LEBT together with simulated beam images in the same locations. Overall, the TRACK code reproduces the particle distribution in the real space.

So far, beam intensity in the LEBT was low,  $\sim 50$   $\mu\text{A}$ . We expect to face more complicated beam dynamics issues when the beam current of a single ion specie will be higher by an order of magnitude. More advanced ECRIS models would be necessary to optimize ECRIS operational mode and beam transport with low emittance growth.

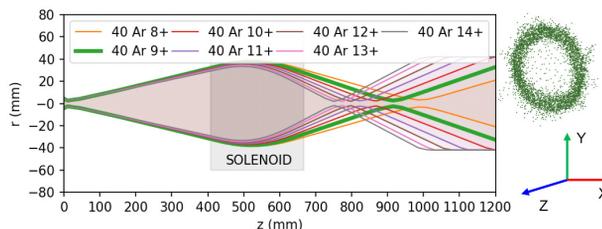


Figure 10: Envelopes of several charge states of argon after the extraction from ECRIS (left) and  $\text{Ar}^{9+}$  beam cross section in the focal plane of the solenoid (right).

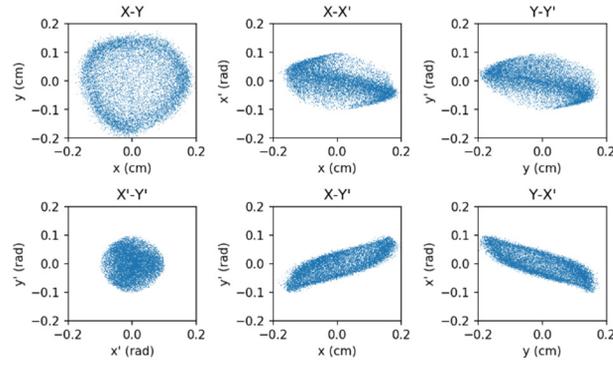


Figure 11: Simulated phase space plots in the Cartesian coordinates at the solenoid focal plane for  $^{40}\text{Ar}^{9+}$ .

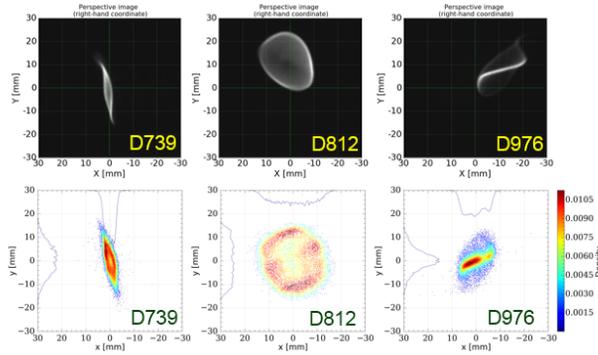


Figure 12: Measured (top) and simulated (bottom) beam images along the LEBT for  $^{40}\text{Ar}^{9+}$ .

## BAYESIAN STATISTICS FOR MACHINE TUNING

We also aim on using statistical methods to gain information from the measured data. Below is an example of the application of Bayesian inference of the profile measurement to infer the unknown linearized beam distribution at the exit of the ECR source. The measurement data used in the inference is recorded using a beam viewer, located downstream of the first three electrostatic quadrupoles in the LEBT. The transverse beam size and the correlations vector  $\sigma_i = (\sigma_x, \sigma_y, \sigma_{xy})$  in  $i^{\text{th}}$  measurement with the voltage setting  $V_i = (V_i^1, V_i^2, V_i^3)$  is available. The linearized 4D distribution is to be inferred as 10 parameters:  $\theta = (\epsilon_x, \beta_x, \alpha_x, \epsilon_y, \beta_y, \alpha_y, c_{xy}, c_{xyy}, c_{xyy}, c_{xyy})$ . We can use FLAME model to predict the measurement as  $\sigma_{\text{model},i} = f(V_i, \theta)$  and assume that the measurement only differs from the prediction by a Gaussian random number  $\xi_i$  with an amplitude  $\delta = (\delta_x, \delta_y, \delta_{xy})$ , which reads:

$$\sigma_{\text{measure},i} = \sigma_{\text{model},i} + \delta \xi_i.$$

Using the Bayesian formula as:

$$P(\theta, \delta | (\sigma_1, V_1), \dots, (\sigma_i, V_i), \dots) = \frac{P((\sigma_1, V_1), \dots, (\sigma_i, V_i), \dots | \theta, \delta) P(\theta, \delta)}{P((\sigma_1, V_1), \dots, (\sigma_i, V_i), \dots)}$$

Since we assume the difference of the measurement and the model is Gaussian, the likelihood is written as

$$P((\sigma_1, V_1), \dots, (\sigma_i, V_i), \dots | \theta, \delta) \sim \prod_i \frac{1}{\delta_x \delta_y \delta_{xy}} e^{-\frac{(\sigma_{\text{measure},i} - \sigma_{\text{model},i})^2}{2\delta^2}}$$

We can use Markov Chain Monte Carlo (MCMC) method to conduct the Bayesian inference [12] and reach saturation as shown in Fig. 13 for one of the parameters, the horizontal emittance. When the iteration reaches saturation, the result fit well with the experimental data. A similar behaviour is observed for all other 9 parameters of the beam distribution. In addition, we observed that the result of Bayesian inference fits better than the optimizer results, when comparing the fitting of the transverse correlation with standard beam optics methods.

We plan to continue using Bayesian method for machine tuning and expect that it will provide statistics information on reliability of beam parameters deduced from the measurements, better scaling to high dimensional problem, less local minimum problem and suggest the future experiments.

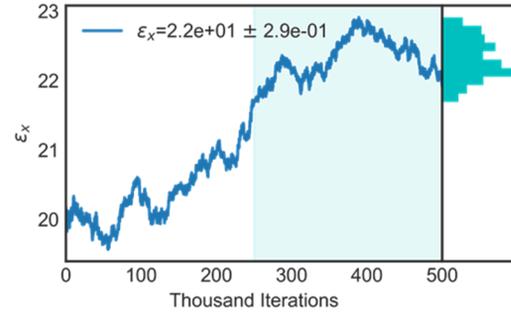


Figure 13: The saturation of the inference after 500K iteration of MCMC.

## RARE ISOTOPE BEAMS

The layout of the FRIB target and fragment separators is shown in Fig. 14. Two software packages, COSY INFINITY [13] and LISE<sup>++</sup> [14] have been heavily used for the design and optimization of the primary beam interaction with the target and transport of rare isotope beams. Due to the large beam emittance after the target, large aperture magnets and large momentum acceptance of the fragment separators, the 5th order optics in COSY INFINITY package has been applied for the design of fragment separators. The design has been verified with extensive Monte Carlo simulations using LISE<sup>++</sup> code with embedded COSY INFINITY transport maps [15]. The simulations show that separation of isotopes on the selection slits is very small, a few mm. During the initial set-up of the fragment separators very large-scale simulations would be necessary to identify a specific isotope of interest among many unwanted products. The isotope of interest can be at very low intensity therefore parallel version of COSY and significant improvement of the LISE<sup>++</sup> [14,16] are necessary for quick tuning of the transport and selection of isotopes for the experiments. This is especially important due to FRIB being a single user facility.

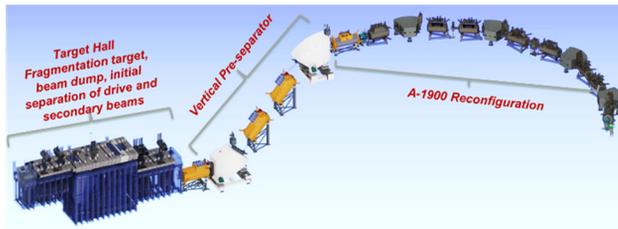


Figure 14: Layout of the beam FRIB experimental systems.

### Preparation of Rare Isotopes for Post Acceleration

There are many nuclear physics experiments that require selection of particular isotopes and re-acceleration to ~3-12 MeV/u. Figure 15 shows current layout of the selection of isotopes in the fragment separator, stoppage in the helium gas cell, bunching in the RFQ cooler-buncher (RFQ CB) and charge breeding in the Electron Beam Ion Trap (EBIT), extraction from EBIT and injection to the post accelerator at 12 keV/u. The intensities of rare isotope beams produced by FRIB will be 4-5 orders of magnitude higher than currently available from the NSCL cyclotron. While there are several codes available for study and optimized design of the helium gas cell, RFQ CB and EBIT, there is no computer model that fully represents all processes in these devices such as 3D electromagnetic fields, interaction with gas atoms, charge-exchange reactions and most importantly space charge of ions. The space charge effects become crucial for the optimal design and operation of these devices with high intensity of isotope beams from FRIB. Therefore, development of such codes is critical for the FRIB science program.

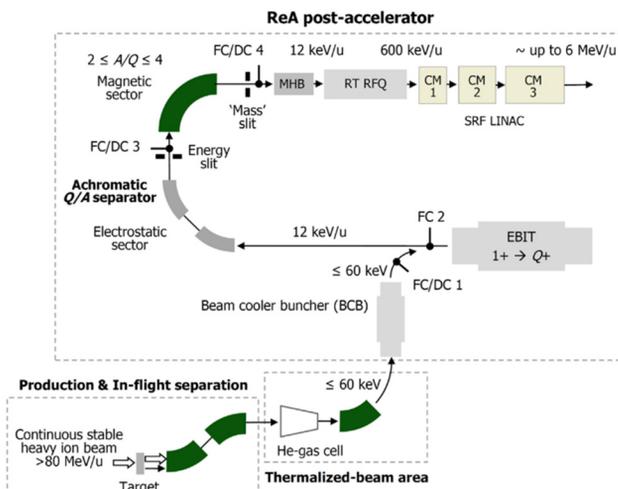


Figure 15: Preparation of secondary beams for injection into the post-accelerator.

### CONCLUSION

Several well-established optimization and simulation codes were available for the design of FRIB accelerator and experimental systems. All these codes are being used for refining FRIB systems and transition to operation. Currently we are focused on the development of on-line physics applications for tuning of the driver linac and update of

the virtual accelerator model. The latter is being performed primarily to include the 3D maps of various electromagnetic devices and misalignment data of the accelerator components.

New parallel codes for large scale Monte Carlo simulation would be necessary for the quick setup of fragment separators and experiments with rare isotope beams. The intensities of rare isotope beams will be 4-5 orders of magnitude higher than in it is available now. Therefore, computer models of the devices for preparation of stopped rare isotope beams for post-acceleration should be updated to include 3D electromagnetic fields, interaction with gas atoms, charge-exchange reactions and, most importantly, space charge forces.

### ACKNOWLEDGMENTS

We thank Drs. A. Lapierre, M. Hausmann, M. Portillo and O. Tarasov for help in description of FRIB experimental systems.

### REFERENCES

- [1] J. Wei *et al.*, "The FRIB Superconducting Linac - Status and Plans", in *Proc. LINAC'16*, East Lansing, MI, USA, Sep 2016. doi:10.18429/JACoW-LINAC2016-M01A01
- [2] E. Pozdeyev *et al.*, "FRIB Front End Design Status", in *Proc. LINAC'12*, Tel-Aviv, Israel, Sep 2012. <http://jacow.org/LINAC2012/papers/thpb097.pdf>
- [3] Q. Zhao *et al.*, "FRIB Accelerator Beam Dynamics Design and Challenges", in *Proc. HB'12*, Beijing, China, Sep 2012, <http://jacow.org/HB2012/papers/weo3b01.pdf>
- [4] J. Qiang *et al.*, *J. Comput. Phys.*, vol. 163, pp. 434, 2000.
- [5] P. N. Ostroumov and K.W. Shepard, *Phys. Rev. ST Accel. Beams*, vol. 11, pp- 030101, 2001.
- [6] G. Machicoane *et al.*, "Design Status of ECR Ion Sources and LEBT for FRIB", in *Proc. ECRIS'12*, Beijing, China, Sep 2012. <http://jacow.org/ECRIS2012/papers/thyo03.pdf>
- [7] P.N. Ostroumov *et al.*, "Accelerator Physics Advances in FRIB (Facility for Rare Isotope Beams)", in *Proc. IPAC'18*, doi:10.18429/JACoW-IPAC2018-THYGBF4
- [8] Z. He *et al.*, *Phys. Rev. ST Accel. Beams*, vol 17, pp. 034001, March 2014.
- [9] <https://pypi.org/project/PyQt5/>
- [10] V. Mironov, *ICFA Beam Dynamics Newsletter*, no. 73, pp. 65-73, ed. G. Machicoane and P. N. Ostroumov, <http://icfa-bd.kek.jp/Newsletter73.pdf>
- [11] <https://www.cst.com/>
- [12] W. K. Hastings, *Biometrika*, vol. 57, no. 1, Apr 1970, pp. 97-109, doi:10.1093/biomet/57.1.97
- [13] K. Makino and M. Berz, *NIMA*, vol. 585, pp. 346-350, 2006. [http://bt.pa.msu.edu/index\\_cosy.htm](http://bt.pa.msu.edu/index_cosy.htm)
- [14] O.B. Tarasov and D. Bazin, *NIM B*, vol. 376, pp. 185, 2016, <http://lise.nsl.msu.edu>
- [15] M. Portillo and M. Hausmann, "Simulation of Rare Isotope Beam Production, Separation, and Delivery", FRIB internal report, FRIB-T40301-TD-000822-R001, 2015.
- [16] M.P. Kuchera *et al.*, *NIM B*, vol. 376, pp. 168, 2016.
- [17] A. Lapierre *et al.*, *Phys. Rev. Accel. Beams*, vol. 21, pp. 053401, 2018.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

# NOVEL, FAST, OPEN-SOURCE CODE FOR SYNCHROTRON RADIATION COMPUTATION ON ARBITRARY 3D GEOMETRIES

Dean Andrew Hidas\*, Brookhaven National Laboratory, Upton, NY, USA

## Abstract

Open Source Code for Advanced Radiation Simulation (OSCARS) is an open-source project (<https://oscars.bnl.gov>) developed at Brookhaven National Laboratory for the computation of synchrotron radiation from arbitrary charged particle beams in arbitrary and time-dependent magnetic and electric fields on arbitrary geometries in 3D. Computational speed is significantly increased with the use of built-in multi-GPU and multi-threaded techniques which are suitable for both small scale and large scale computing infrastructures. OSCARS is capable of computing spectra, flux, and power densities on simple surfaces as well as on objects imported from common CAD software. It is additionally applicable in the regime of high-field acceleration. The methodology behind OSCARS calculations will be discussed along with practical examples and applications to modern accelerators and light sources.

## INTRODUCTION

OSCARS [1] is an open source software developed at Brookhaven National Laboratory (BNL). OSCARS is a general purpose code for the calculation of radiation from charged particles in motion. Primary uses are for synchrotron and accelerator facilities where photon density distributions and heat loads on accelerator and beam-line equipment is of interest. This software allows for the calculation of these properties on arbitrary shaped surfaces in 3 dimensions. Recently added features include the ability to import surface models directly in the common Computer Aided Design (CAD) STL format and the implementation of time dependent magnetic and electric source fields.

The core code is written in modern C++ for speed and has a python extension which serves as the main application programming interface (API). The API was written in python for ease of use and integrability by the larger scientific community. Currently OSCARS is available for Python 2.7 and Python 3+, for Linux, OS X, and Windows operating systems.

Throughout this paper the 2 example undulators referred to (U49 and EPU49) have a period of 49 mm, 55 periods plus terminating fields, and a peak magnetic field of 1 Tesla. U49 is a planar undulator with the only non-zero component of the magnetic field  $\vec{B}$  being a sinusoidally varying  $B_v$  (vertical component). EPU49 is an elliptically polarizing undulator where  $B_h$  (horizontal component) and  $B_v$  are sinusoidally varying with the same peak magnitude, but phase shifted by  $\pi/2$  relative to each other while  $B_l$  (longitudinal component) remains zero. The beam parameters used in these simulations are that of the NSLS-II 6.6 m straight sections with

a beam energy  $E = 3$  GeV, energy spread  $\Delta E/E = 0.001$ , emittance  $\epsilon_{h,v} = [0.9, 0.008]$  nm rad, and beta function in the center  $\beta_{h,v} = [1.5, 0.8]$  m.

## PARTICLE BEAMS

Particle beams in OSCARS are defined by particle mass, energy, current, and their position and direction. Optionally one can include the emittance ( $\epsilon_{h,v}$ ), beta function ( $\beta_{h,v}$ ), and energy spread ( $\Delta E/E$ ). In multi-particle simulations the energy, initial position, and initial momentum are sampled accordingly assuming a Gaussian distribution of the position ( $\sigma$ ) and momentum ( $\sigma'$ ). OSCARS also allows for the definition of multiple beams which are sampled randomly according to their relative weights given. Any beam direction is valid in OSCARS. For convenience several pre-defined beams exist in OSCARS, for example in this case "NSLSII-ShortStraight".

## MAGNETIC AND ELECTRIC FIELDS

Several field types (applicable to both magnetic and electric fields) are available in OSCARS. These source fields may be time dependent, static, or any combination therein. Simple parameters exist for configuring any static field as a time-dependent sinusoidal resonant field with a phase offset. Some common built-in fields include uniform fields in a given range, Gaussian fields, sinusoidal undulator (wiggler) fields with terminating fields, user input python functions, and 1D, 2D, and 3D discrete field data. In the case of the 1D discrete field data, the field points do not need to be uniformly distributed in space as OSCARS will regularize it by interpolation for internal use and fast internal field lookup. OSCARS also has an interpolating tool (using a cubic spline method) in the case that you have field data measured at several parameter points (such as undulator gap, but not restricted to this) but wish to use the field at an intermediate point. Several input data formats are implemented for convenience.

## CALCULATION OF ELECTRIC FIELD

Particle trajectories are calculated in 3D according to the relativistic Lorentz equation given in Eq. (1) using a fourth-order Runge-Kutta (RK4) algorithm or an optional adaptive step RK4 method. Care is taken in the RK4 method implementation to avoid  $\beta \geq 1$  by an iterative step-halving method until the criteria  $\beta < 1$  is satisfied in for the trajectory propagation of Eq. (1):

$$\frac{d\vec{p}}{dt} = q(\vec{E} + c\vec{\beta} \times \vec{B}). \quad (1)$$

\* dhidas@bnl.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

The source electric and magnetic fields in Eq. (1) need not be static fields, *e.g.*  $\vec{E} = \vec{E}(t)$  and  $\vec{B} = \vec{B}(t)$ . The electric field in the frequency domain is given in Eq. (2) which is derived by taking a Fourier transform of the Liéard-Wiechert potentials and given in many texts, for example [2–4]. Equation (2) is valid for relativistic and non-relativistic particles as well as in the near-field:

$$\vec{E}(\vec{x}, \omega)/C = \frac{1}{\gamma^2} \int_{-\infty}^{+\infty} \frac{\hat{n} - \vec{\beta}}{D^2(1 - \hat{n} \cdot \vec{\beta})^2} e^{i\omega(\tau+D/c)} d\tau \quad (2)$$

$$+ \int_{-\infty}^{+\infty} \frac{\hat{n} \times (\hat{n} - \vec{\beta}) \times \dot{\vec{\beta}}}{R(1 - \hat{n} \cdot \vec{\beta})^2} e^{i\omega(\tau+D/c)} d\tau.$$

Here,  $C = \frac{2I}{\hbar q \mu_0 c}$ ,  $\vec{p}$  is the particle momentum,  $\tau$  the laboratory time,  $q$  particle charge,  $\vec{E}$  electric field,  $c$  speed of light in vacuum,  $\vec{\beta}$  the particle velocity divided by  $c$ ,  $\vec{B}$  the magnetic field, and  $\vec{x}$  the observation point.  $\omega$  is the photon angular frequency of interest,  $\epsilon_0$  the permittivity of free space,  $\hat{n}$  a unit vector in the direction from the particle to the observation point, and  $D$  the distance between the particle and the observation point.

For the calculation of spectra and flux OSCARS performs a numerical integration of Eq. (2) once the trajectory has been calculated using Eq. (1). This is done either on the Central Processing Unit (CPU) or GPU depending on which mode is selected. The spectrum in Fig. 1 and flux densities in Fig. 2 use this method. An interpolating method is used to calculate trajectory points on a regularized grid from the results of the RK4 propagation. Each successive integration step (for a flux, spectrum, or power density calculation) uses  $2N + 1$  trajectory integration points until the desired relative precision is reached (typically 1%, but the user is free to specify any alternative).

## POWER DENSITY

Once the trajectory is calculated, the power density can be calculated from Eq. (3):

$$P(\vec{x}) = \frac{qI}{16\pi^2\epsilon_0 c} \int_{-\infty}^{\infty} \frac{\vec{n} \times ((\vec{n} - \vec{\beta}) \times \frac{\vec{a}}{c})}{(1 - \vec{\beta} \cdot \vec{n})^5} \frac{1}{D^2} (\hat{n} \cdot \hat{S}) d\tau \quad (3)$$

where  $I$  is the beam current,  $\vec{a}$  the particle acceleration and  $\hat{S}$  a unit normal vector for the surface point at position  $\vec{x}$ . Here again, the numerical integration is performed either on the CPU or the GPU if desired.

The power density distribution in the  $x$ - $y$  plane 30 m downstream from the example U49 is shown in Fig. 3. OSCARS allows for the calculation of power density distributions on arbitrary objects in 3D. Several examples of this are shown in Fig. 4. These were calculated using the GPU option available in OSCARS, but can as well easily be computed on a single-core, in multi-threaded mode, or using grid/cloud computing.

Power density distributions can also be calculated on 3D surfaces designed in common CAD software. OSCARS accepts the STL (stereo lithography) file format. Ray-tracing

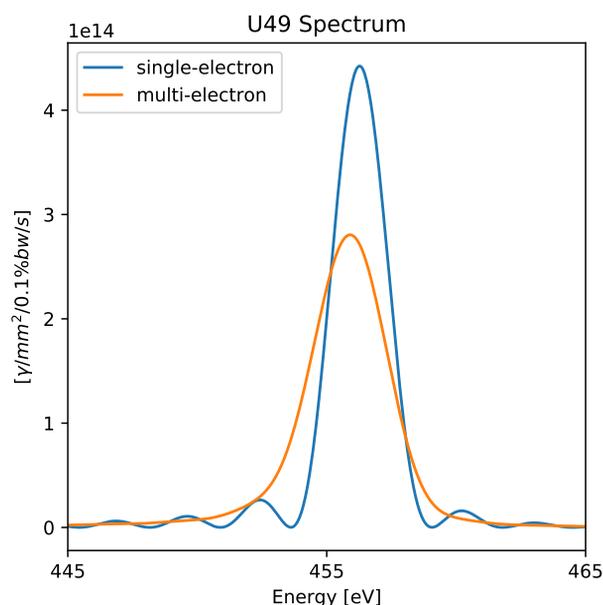


Figure 1: On-axis spectrum for the simulated U49 undulator 30 m downstream showing the 3rd harmonic for a filament beam (single-electron), and realistic beam using the NSLS-II design parameters for a 6.6 m straight section (multi-electron).

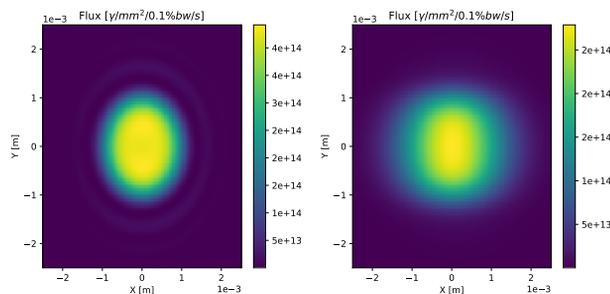


Figure 2: Flux for the simulated U49 undulator 30 m downstream showing the 3rd harmonic for a filament beam (left) and realistic beam (right) using the NSLS-II design parameters for a 6.6 m straight section.

is performed in the far-field approximation in such a way that the radiation will be blocked at the first surface it encounters (*i.e.* *shadow effect*). This will lend itself well to the investigation of very complex objects within reach of high intensity beams from synchrotron radiation. A simple example of this effect is shown in Fig. 5 where two spheres of different sizes are shown with the U49 source 30 meters upstream.

As an example of OSCARS time dependent source field calculations Fig. 6 shows the  $\beta_z$  response of a point-charge bunch in a strong longitudinal electric field. The power density 1 m downstream of this is shown in Fig. 7. Although of limited interest at most accelerators these quantities (spectra, flux, power) can be calculated for extremely high field gradients with time dependence.

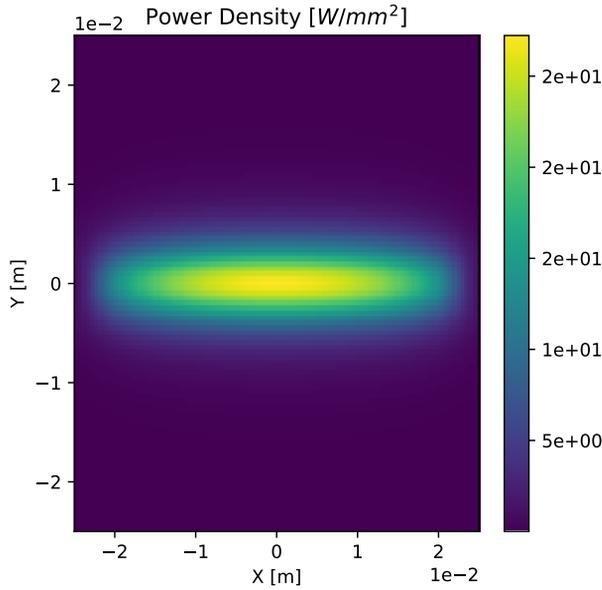


Figure 3: Power density from U49 as seen 30 m downstream from the source.

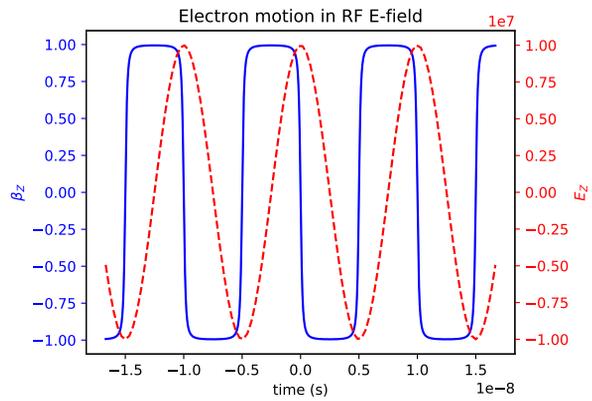


Figure 6: High gradient acceleration of a point-charge bunch.

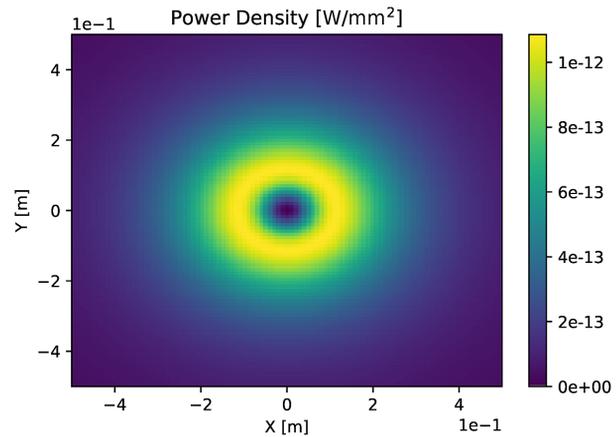


Figure 7: Power density distribution from the motion described in Fig. 6.

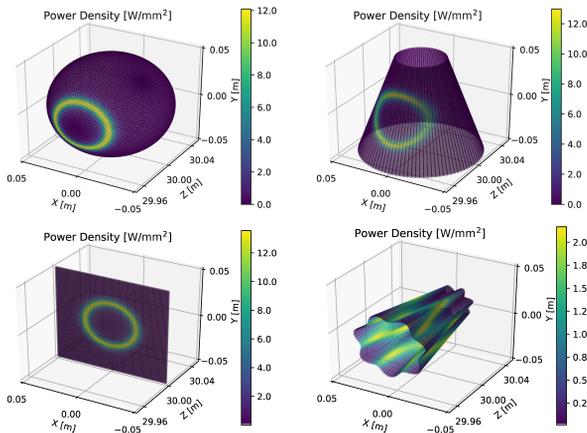


Figure 4: Power density on various 3D shapes for the simulated EPU49 undulator 30 m downstream.

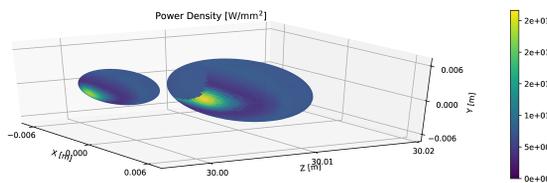


Figure 5: Power density on 2 spheres produced with CAD software and imported into OSCARS in the STL file format.

## GPU, MULTI-THREADING, AND GRID COMPUTING

OSCARS was designed as a multi-threaded application which is also capable of using multiple GPUs simultaneously making use of direct GPU-GPU cascading memory transfers. Native multi-threading is available by default for all calculations on all platforms. The GPU routines are currently written using CUDA [5] for compatible NVidia GPUs and are also available for all calculations. For both multi-threaded and GPU calculations one typically has a set of points in 1D (like a spectrum), 2D (for instance flux maps or power density distributions), or 3D (arbitrary shapes and surfaces in 3D). Given a trajectory, these points are then distributed to the specified number of threads in the case of multi-threading, or sent to the GPU in large thread-blocks until complete. When trajectory interpolations are done to reach the desired level of calculation accuracy they are done using mutex locking when CPU calculations are enabled to avoid thread memory collisions, but are done on-the-fly by each GPU thread as needed.

On a modern GPU the time for a calculation similar to the single-particle flux in Fig. 2 for a  $300 \times 300$  grid is significantly reduced going from 1 core of an Intel® Xeon® CPU

E5-2630 v3 @ 2.40GHz to using the GPU option in OSCARS on a modest 1344 core Quadro® K4200 GPU (greater than a factor of 15). On a modern high performance GPU one should easily achieve over a factor of 100 in performance improvement.

For extremely large-scale simulations OSCARS was designed to be run on modern grid/cloud computing infrastructures such as the Open Science Grid (OSG) [6]. The multi-particle spectra and flux shown in Fig. 1 and Fig. 2 respectively were performed on OSG and contain 100,000 particles sampled from the beam distributions. Tools are also available for the commonly used Message-Passing Interface (MPI) [7] which may be beneficial when one has a cluster which utilizes MPI, however on a typical workstation the multi-threaded utility will outperform MPI since there is very little overhead in the straightforward implementation used as compared to MPI.

## CONCLUSION

A new and modern simulation code for advanced radiation simulation has been developed which is fast, powerful, flexible, and open source. Notably, this new simulation is capable of calculating power densities on arbitrary geometries in 3D and utilizes native multi-threading and GPU computing infrastructure for significantly increased performance for large-scale simulations.

## ACKNOWLEDGMENTS

This work has been supported by the U.S. Department of Energy, under contract DE-SC0012704.

## REFERENCES

- [1] <https://oscars.bnl.gov>
- [2] J. D. Jackson, "Classical Electrodynamics," Berkeley, CA, USA, 1981,
- [3] A. Hoffman, "The Physics of Synchrotron Radiation," Cambridge, UK, 2004,
- [4] H. Onuki and P. Elleaume, "Undulators, Wigglers and Their Applications," London, UK, 2003
- [5] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable Parallel Programming with CUDA," *ACM Queue*, vol. 6, no. 2, pp. 39–53, March/April 2008.
- [6] R. Pordes *et al.*, "The Open Science Grid," *J. Phys. Conf. Ser.*, vol. 78, pp. 012057, 2007. doi:10.1088/1742-6596/78/1/012057
- [7] L. Dalcín, R. Paz, and M. Storti, "MPI for Python," *Journal of Parallel and Distributed Computing*, vol. 65, no. 9, 2005.

# MAIN AND FRINGE FIELD COMPUTATIONS FOR THE ELECTROSTATIC QUADRUPOLES OF THE MUON $g$ -2 EXPERIMENT STORAGE RING\*

Eremey Valetov<sup>1†</sup> and Martin Berz<sup>‡</sup>, Michigan State University, East Lansing, MI 48824, USA  
<sup>1</sup>also at Lancaster University and the Cockcroft Institute, UK

## Abstract

We consider semi-infinite electrostatic deflectors with plates of different thickness, including plates with rounded edges, and we calculate their electrostatic potential and field using conformal mappings. To validate the calculations, we compare the fringe fields of these electrostatic deflectors with fringe fields of finite electrostatic capacitors, and we extend the study to fringe fields of adjacent electrostatic deflectors with consideration of electrostatic induction, where field falloffs of semi-infinite electrostatic deflectors are slower than exponential and thus behave differently from most magnetic fringe fields. Building on the success with electrostatic deflectors, we develop a highly accurate and fully Maxwellian conformal mappings method for calculation of main fields of electrostatic particle optical elements. A remarkable advantage of this method is the possibility of rapid recalculations with geometric asymmetries and mispowered plates. We use this conformal mappings method to calculate the multipole terms of the high voltage quadrupole used in the storage ring of the Muon  $g$ -2 Experiment (FNAL-E-0989). Completing the methodological framework, we present a method for extracting multipole strength falloffs of a particle optical element from a set of Fourier mode falloffs. We calculate the quadrupole strength falloff and its effective field boundary (EFB) for the Muon  $g$ -2 quadrupole, which has explained the experimentally measured tunes, while simple estimates based on a linear model exhibited discrepancies up to 2%.

## INTRODUCTION

Methods for measurement of anomalous magnetic dipole moment (MDM) and electric dipole moment (EDM) using a storage ring rely on electrostatic particle optical elements, including the Muon  $g$ -2 Experiment's storage ring at FNAL, which uses electrostatic quadrupoles with a curved reference orbit. Accordingly, it is necessary to accurately model main and fringe fields of electrostatic elements. In particular, inaccurate treatment of fringe fields of electrostatic elements provides a mechanism for energy conservation violation.

In this research, we address the problem of accurate representation for fringe fields of electrostatic deflectors, as well as for main and fringe fields of electrostatic quadrupoles with the specific case of the Muon  $g$ -2 quadrupole [1] considered. Our model of the main field of the Muon  $g$ -2

quadrupole allows rapid recalculations with geometric asymmetries and mispowered plates, the latter being useful, *inter alia*, for simulations of RF scraping and the effects of damaged quadrupole resistors. For the fringe field of the Muon  $g$ -2 quadrupole, we calculated the field map and extracted the falloff of the quadrupole strength, which was a basis for achieving a good agreement of calculated tunes with experimentally measured tunes.

## CONFORMAL MAPPING METHODS

We used conformal mappings for calculation of fringe fields of electrostatic deflectors and the main field on the Muon  $g$ -2 quadrupole [2, App. A]. A conformal mapping (or conformal map) is a transformation  $f : \mathbb{C} \rightarrow \mathbb{C}$  that is locally angle preserving. A conformal mapping satisfies Cauchy–Riemann equations and, therefore, its real and imaginary parts satisfy Laplace's equation:  $\Delta \mathcal{R}(f) = 0$  and  $\Delta \mathcal{I}(f) = 0$ . Conformal mappings automatically provide the electrostatic potential in cases where the electrostatic element's geometry can be represented by a polygon, possibly with some vertices at the infinity. The domain of a conformal mapping is called the canonical domain, and the image of a conformal mapping is called the physical domain. A Schwarz–Christoffel mapping is a conformal mapping from the upper half-plane as the canonical domain to the interior of a polygon as the physical domain.

The electrostatic potential for a cross section or a longitudinal section modeled by a generalized polygon may be found by obtaining a conformal mapping  $f$  from a suitable canonical domain to the polygon. A bi-infinite strip is a suitable canonical domain if the polygon comprises two groups of consecutive sides characterized by the same constant Dirichlet boundary condition, with two constant values in total. A rectangular part of a bi-infinite strip is a suitable canonical domain when the physical domain is a logical (or generalized) quadrilateral.

If the solution of the Laplace equation in the canonical domain is  $\phi$ , the solution of the Laplace equation in the physical domain is  $\varphi = \phi \circ f^{-1}$ . In practice, the electrostatic potential is usually the appropriately selected, shifted, and scaled real or imaginary part of  $f^{-1}$ .

The solution for the electrostatic potential obtained this way is fully Maxwellian in the sense that the analytic formula for  $f$  or  $f'$  results in the solution for the potential satisfying the Laplace equation.

As described in [3,4], inverse conformal mapping  $g = f^{-1}$  may be obtained by

\* Fermilab report FERMILAB-CONF-18-571-PPD.

† Email: evaletov@fnal.gov. ORCID: 0000-0003-4341-0379.

‡ ORCID: 0000-0001-6141-8230.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

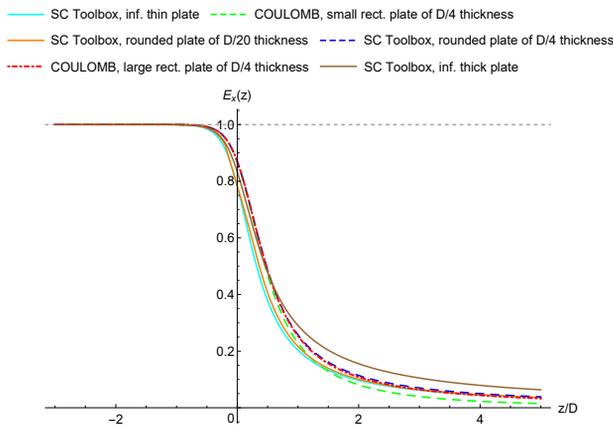


Figure 1: The electrostatic field falloffs  $E_x(z)$  in cases 2 (solid cyan), 3 (solid brown), 5 (solid orange), 6 (dashed blue), 7 (dashed green), and 8 (dot-dashed red).

1. solving the equation

$$f(g(w)) - w = 0 \quad (1)$$

using the Newton–Raphson or another numerical method; or

2. solving the ODE [2, App. A]

$$\frac{dg(w)}{dw} = \frac{1}{f'(g(w))}, \quad g(w_0) = z_0. \quad (2)$$

## DIFFERENTIAL ALGEBRA

The tracking code *COSY INFINITY* [5] has a rigorous and efficient implementation of differential algebra (DA) [6, pp. 86–102]. *Mathematica* also has a DA implementation, where DA values are encoded in *SeriesData* objects. One of the advantages of DA computations is that Taylor series expansions can be produced automatically for a highly complex analytic function by substituting its comprising functions with their DA values. Another advantage is the DA fixed-point algorithms, e.g., for the inversion of a Taylor series expansion, that complete in finitely many steps.

## FRINGE FIELDS OF ELECTROSTATIC DEFLECTORS

Using conformal mappings, we obtained electrostatic field falloffs for semi-infinite capacitors with infinitely thin, infinitely thick, and finitely thick plates [2, Ch. 2]. For finitely thick plates, we considered plate thicknesses  $D/4$  and  $D/20$  in terms of the aperture  $D$ , as well as cases with rectangular and realistically rounded plate edges. The rounded edges were approximated by piecewise linear curves with 42 line segments, which is a good approximation.

In each case, we obtained a conformal mapping from the bi-infinite strip  $0 \leq \Im(w) \leq 1$  to the physical domain of the

form [3, p. 46]

$$f(w) = f(w_0) + c \int_{w_0}^w \exp\left[\frac{\pi}{2}(\alpha_- - \alpha_+)\zeta\right] \prod_{j=1}^n \left[\sinh\frac{\pi}{2}(\zeta - w_j)\right]^{\alpha_j - 1} d\zeta,$$

where the number of vertices  $n$  and angles  $\pi\alpha$  are parameters of the polygonal model and the parameters  $w$  were found using the *MATLAB* package *Schwarz–Christoffel Toolbox* (*SC Toolbox*) [7].

To validate the fringe fields of semi-infinite electrostatic capacitors obtained using conformal mappings, we compared the results with the electrostatic field of finite rectangular electrostatic capacitors with finitely thick plates and different plate sizes, which were computed by Helmut Soltner (FZ Jülich) using *COULOMB*'s [8] boundary element method (BEM) field solver. This comparison shows good agreement.

Overall, the fringe fields for the following cases of one electrostatic capacitor were calculated:

1. Semi-infinite capacitor with infinitely thin plates, manually obtained conformal mapping;
2. Semi-infinite capacitor with infinitely thin plates, *SC Toolbox* calculations;
3. Semi-infinite capacitor with infinitely thick plates, *SC Toolbox* calculations;
4. Semi-infinite capacitor with plates of  $D/20$  thickness, *SC Toolbox* calculations;
5. Semi-infinite capacitor with plates of  $D/20$  thickness and rounded edges, *SC Toolbox* calculations;
6. Semi-infinite capacitor with plates of  $D/4$  thickness and rounded edges, *SC Toolbox* calculations;
7. Large rectangular capacitor with plates of  $D/4$  thickness, *COULOMB* calculations; and
8. Small rectangular capacitor with plates of  $D/4$  thickness, *COULOMB* calculations.

A comparison of the fringe fields of electrostatic capacitors in cases 2–8 is shown in Fig. 1.

Because of electrostatic induction, the field of two adjacent electrostatic deflectors differs from the superposition of the fields of these deflectors as they would be in empty space. To study this effect, we modeled fringe fields of two adjacent semi-infinite capacitors with finitely thick plates and symmetric, antisymmetric, and different voltages (see Figs. 2 and 3). For adjacent electrostatic capacitors with symmetric and antisymmetric voltages, making use of the symmetries, we obtained the field using a conformal mapping from the bi-infinite strip the same way as for one electrostatic capacitor. In the case of adjacent electrostatic capacitors of different voltages, we used a composite conformal mapping that is suitable for solving the Laplace equation with multiple Dirichlet boundary conditions as detailed in [3, pp. 77–83].

We found that, unlike fringe fields of most magnetic elements, fringe fields of electrostatic deflectors fall off

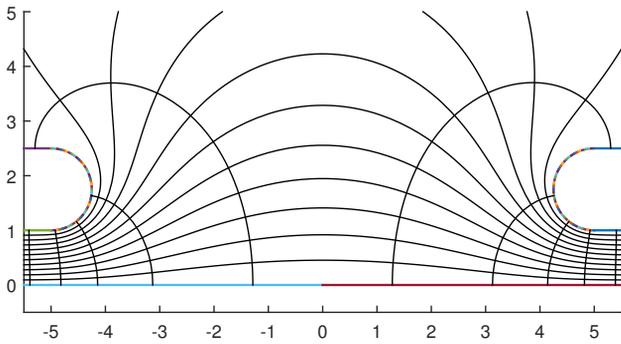


Figure 2: The electrostatic field and equipotential lines of two adjacent semi-infinite capacitors with plates of  $3D/4$  thickness, symmetric voltages, and rounded edges.

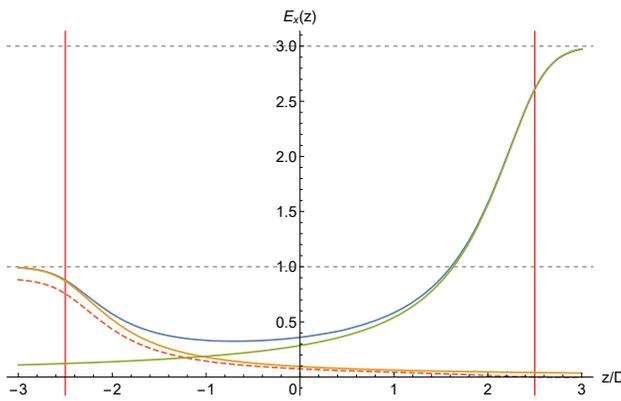


Figure 3: The electrostatic field  $E_{A\&B}(z)$  (blue) of two adjacent semi-infinite capacitors with plates of  $D/2$  thickness and different voltages  $V_A = 1$  and  $V_B = 3$ , individual fields  $E_A(z)$  (orange) and  $E_B(z)$  (green) of each capacitor as in empty space, and the difference  $E_{A\&B}(z) - E_B(z)$  (dashed red) that would be equal to  $E_A(z)$  without electrostatic induction.

slower than exponentially. Thus, Enge functions of the form  $F_N(z) = \left[ 1 + \exp\left(\sum_{j=1}^N a_j \left(\frac{z}{D}\right)^{j-1}\right) \right]^{-1}$ , where  $z$  is the longitudinal coordinate and  $D$  is the aperture, are not suitable for accurate modeling of the asymptotic behavior of field falloffs of electrostatic deflectors, as Fig. 4 illustrates. However, a piecewise function consisting of an Enge function at  $z/D < c$  and  $G(z) = \left[ \sum_{j=1}^{N_2} b_j \left(\frac{z}{D}\right)^{j-1} \right]^{-1}$  at  $z/D \geq c$  for some  $c > 0$ , smoothly glued as

$$H(z) = \frac{1}{1 + \exp\left[\sum_{j=1}^{N_1} a_j \left(\frac{z}{D}\right)^{j-1}\right]} \frac{1}{1 + \exp\left[\left(\frac{z}{D} - c\right)^2\right]} + \frac{1}{\sum_{j=1}^{N_2} b_j \left(\frac{z}{D}\right)^{j-1}} \frac{1}{1 + \exp\left[-\left(\frac{z}{D} - c\right)^2\right]},$$

models field falloffs of electrostatic deflectors accurately [2, Ch. 2], as shown in Fig. 5.

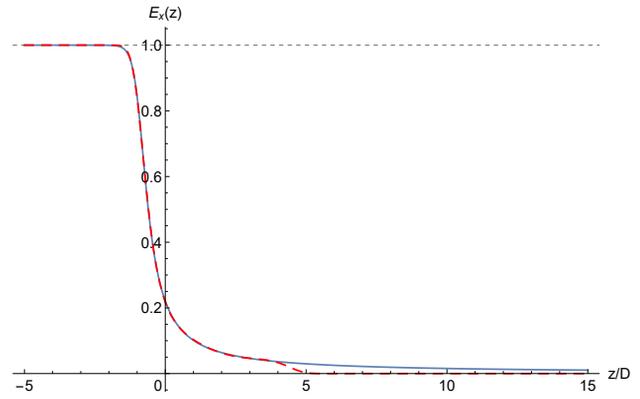


Figure 4: The plot on the left shows a 5-parameter Enge function  $F_5(z)$  (dashed red), fitted to the  $x$  component of the electrostatic field falloff  $E_x(z)$  (solid blue) of a semi-infinite capacitor with infinitely thin plates.

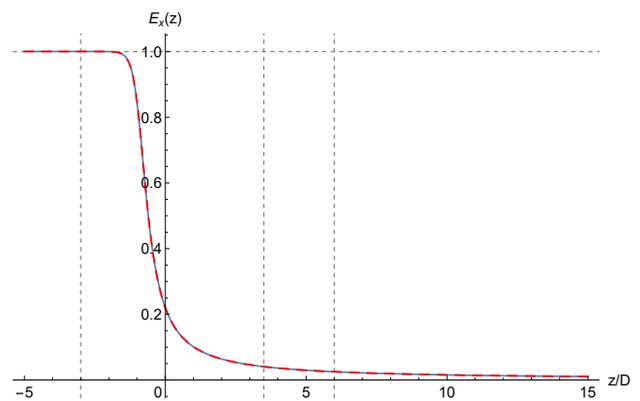


Figure 5: The plot on the right shows a function (dashed red) of the alternative form  $H(z)$ , fitted to  $E_x(z)$  (solid blue) and enhanced in the interval  $-3.5 \leq z/D \leq 6.5$  by adding a Fourier exponential series expansion of the difference  $E_x(z) - H(z)$ .

## MAIN FIELD OF THE MUON $g-2$ COLLABORATION QUADRUPOLE

The main field of an electrostatic element such as the Muon  $g-2$  collaboration quadrupole can be obtained using the following general method [2, 9–13]:

1. Calculate the electrostatic potential using conformal mapping methods with one plate at 1 V and the other Dirichlet boundary conditions (the remaining plates, the rectangular enclosure, and the trolley rails) of 0 V.
2. Apply plate distance errors as perturbations to four copies of the potential, each copy corresponding to one plate at 1 V and the other Dirichlet boundary conditions of 0 V.
3. Apply appropriate rotations to these four copies of the potential, scale the copies (e.g., by  $\pm 2.4 \times 10^4$  or with misspelled values), and use their superposition.

We considered two polygonal models of the cross section: (1) the nominal case with symmetric voltages and no ge-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

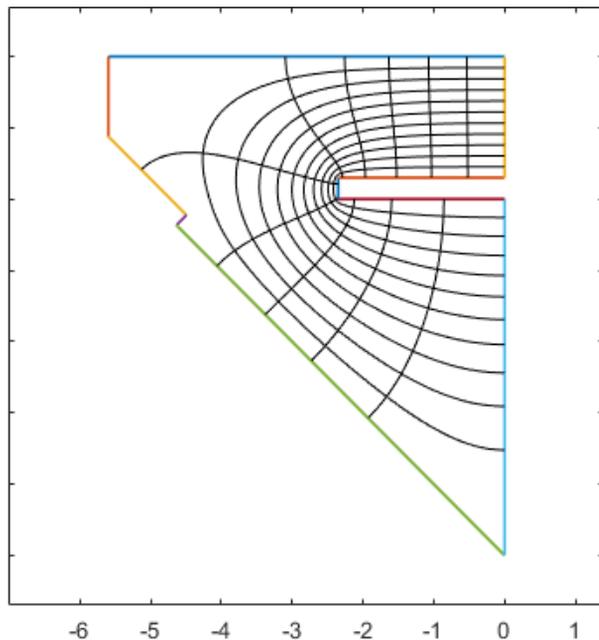


Figure 6: The polygonal model of the Muon  $g-2$  quadrupole in the SM case.

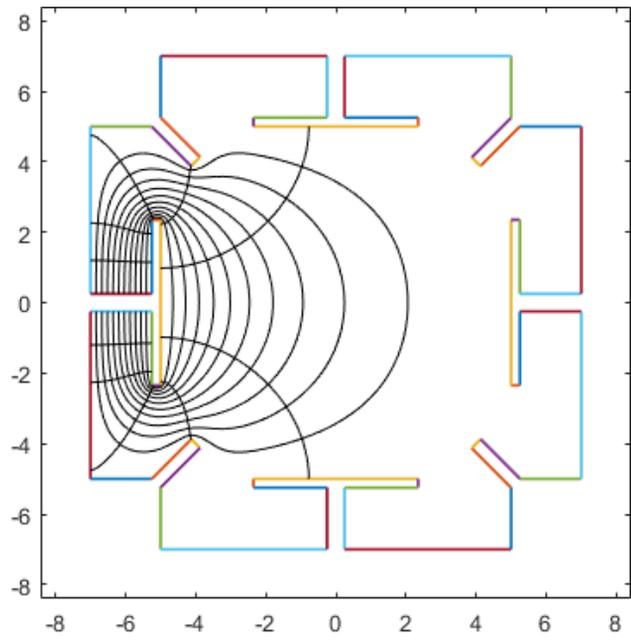


Figure 7: The polygonal model of the Muon  $g-2$  quadrupole in the NSM case.

ometric asymmetries (“SM”), and (2) the general case of mispowered plates and geometric asymmetries (“NSM”). In the former case, we simplified the polygonal model using the four-fold rotational symmetry and the four mirror symmetries, as Fig. 6 shows. The conformal mapping theory for physical domains as  $n$ -connected regions for  $n \geq 2$  is quite challenging or restricted (see, e.g., [14] and [3, pp. 64–70]) compared to simply-connected regions. We approximated the cross-sectional geometry in the NSM case by a simply-connected region shown in Fig. 7 using connecting rods between the rectangular enclosure and the four plates, which were placed in the middle of the back side of each plate to minimize their impact on the multipole terms.

In both cases, considering two constant Dirichlet boundary conditions are interposed by two von Neumann boundary conditions, forming a logical quadrilateral, we used a conformal mapping from a rectangular part of a bi-infinite strip. The derivative of the conformal map  $f$  from the canonical domain to the physical domain was [3, p. 49]

$$f'(z) = c \operatorname{cn}(z|m) \operatorname{dn}(z|m) \prod_{j=1}^n (\operatorname{sn}(z|m) - \operatorname{sn}(x_j + iy_j|m))^{\alpha_j - 1},$$

where  $\operatorname{sn}$ ,  $\operatorname{cn}$ , and  $\operatorname{dn}$  are the Jacobi elliptic functions<sup>1</sup>,  $K$  is the complete elliptic integral of the first kind<sup>2</sup>, the number of vertices  $n$  and angles  $\pi\alpha$  are parameters of the polygonal model, and the parameters  $x$ ,  $y$ ,  $m$ , and  $c$  were found using the *SC Toolbox*.

Knowing the analytic expression for a derivative  $f'$  of a conformal mapping  $f$  and the constant part (the scalar

value)

$$g_0 = \operatorname{cons}(g(0)) \in \mathbb{C}$$

of the DA value of  $g = f^{-1}$  at the origin  $w = 0$ , we can obtain the DA inverse  $g(0)$  at the origin as

$$g(0) = g_0 + \left(\partial^{-1} f'(g_0)\right)^{-1}.$$

In the SM case, the derivative  $f'$  of the conformal mapping has a branch point at the preimage of the origin, which corresponds to the reference orbit. This presents certain difficulties in the analysis. For example, it is not possible to obtain the electrostatic potential multipole terms by obtaining  $f$  via a Taylor series expansion of  $f'$  and then calculating the inverse series. The same applies to the calculation of DA values of  $f$  at point  $z = g_0$ .

In view of this, for the SM case, we obtained the multipole expansion for the Muon  $g-2$  quadrupole up to order 24 in the form

$$\varphi(r, \theta) = \frac{A_0}{2} + \sum_{j=1}^N r^j (A_j \cos(j\theta) + B_j \sin(j\theta)) + O(r^{N+1})$$

of the electrostatic potential by solving the restriction of the ODE from Eq. (2) to the vertical edge of the polygonal model, using the solution as a boundary condition in solving the Cauchy–Riemann PDE in *Mathematica*, and performing Fourier analysis.

We also calculated the multipole expansion in *MATLAB* in the SM case up to order 24 by computing the inverse values of the conformal mapping object  $\mathbf{f}$  at an equidistant discretization of a circle of radius  $R$  into  $N = 1001$  arc intervals of length  $\Delta t = 2\pi R/N$  as

$$u = \left( \mathbf{f}^{-1} \left( R \cos(j\Delta t), \mathbf{f}^{-1} \left( R \sin(j\Delta t) \right) \right) \right)_{j=0}^{N-1},$$

<sup>1</sup>Definitions of the Jacobi elliptic functions can be found at <http://mathworld.wolfram.com/JacobiEllipticFunctions.html>.  
<sup>2</sup>The complete elliptic integral of the first kind is defined at <http://mathworld.wolfram.com/CompleteEllipticIntegraloftheFirstKind.html>.

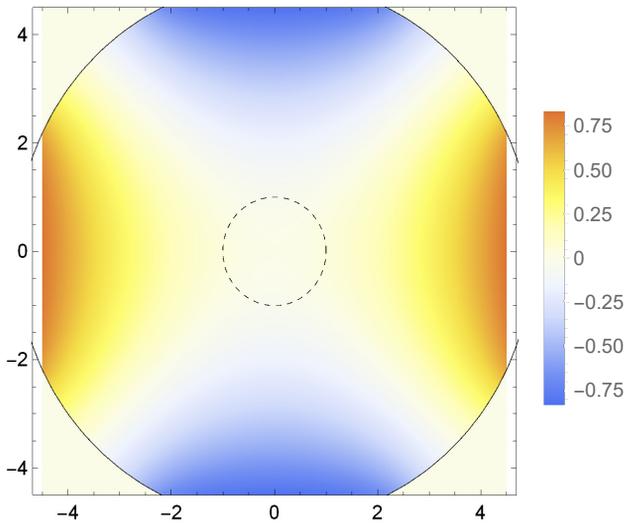


Figure 8: A heatmap plot of the multipole expansion of the electrostatic potential in the NSM case, up to order 24.

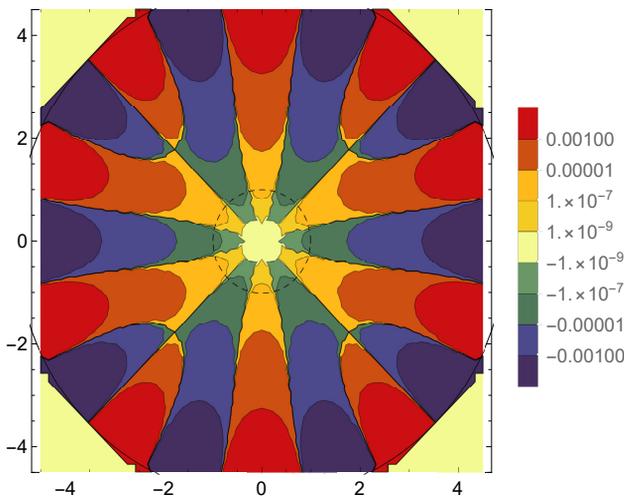


Figure 9: A contour plot of the multipole expansion of the electrostatic potential in the NSM case, orders 3 to 24.

taking the discrete Fourier transform (DFT) of the electrostatic potential  $\varphi$  around the circle, and obtaining the Fourier modes using the Hermitian symmetry.

For the NSM case, we obtained the multipole expansion up to order 24 using the DA inverse of the conformal mapping using *COSY INFINITY* (see Figs. 8 and 9) and *Mathematica*, as well as using Fourier analysis applied to its conventional inverse in *MATLAB*. In particular, we produced a *COSY INFINITY* program called *mterms*, as well as its variant written in *Python*, that calculates the multipole terms of the Muon  $g-2$  quadrupole for a given set of mispowered plates and plate misalignments.

Our *a posteriori* error analysis indicates that the DA method is accurate. This application of the conformal mappings method was near its limit in terms of the complexity of geometry due to the crowding phenomenon [15], where crowding refers to a close collocation of preimages of the

polygon vertices. However, the method can be expanded to significantly more complex geometries using the cross ratios of the Delaunay triangulation (CRDT) algorithm [3, 16, 17]. The conformal mappings method has the advantage of an analytic, fully Maxwellian formula and allows rapid recalculations with adjustments to the geometry and mispowered plates [2, Ch. 3].

The multipole terms we calculated for the Muon  $g-2$  quadrupole were used to study the effects of an unpowered plate [18] and for RF scraping studies [19]. We used these multipole terms to explain the so-called oomph effect [20], which refers to the power of Muon  $g-2$  quadrupoles being apparently 4% higher than the voltage to which they are set. More recently, the multipole terms were used to study the effect of damaged resistors with affected RC time constants on beam dynamics variables such as the beta function, tunes, and coherent betatron oscillation (CBO) frequencies in the Muon  $g-2$  storage ring [21].

## FRINGE FIELD OF THE MUON $g-2$ COLLABORATION QUADRUPOLE

In the fringe field of an electrostatic particle optical element, due to the dependence of the field on the longitudinal component  $s$ , the expansion of the electrostatic field takes the general Taylor–Fourier form [2, 22, 23]

$$\varphi(r, \theta, s) = \sum_{k=0}^{+\infty} \sum_{l=0}^{+\infty} M_{k,l}(s) \cos(l\theta + \theta_{k,l}) r^k.$$

Multipole terms  $M_{k,l}(s)$  vanish for  $k < l$  and  $k = l + 1, l + 3, \dots$  in the general case.

We compare it with the Fourier expansion of the electrostatic potential

$$\varphi(r, \theta, s) = \frac{a_0(r, s)}{2} + \sum_{l=0}^{+\infty} a_l(r, s) \cos(l\theta + \theta_l),$$

where we assume the constant part  $a_0(r, s)/2$  to be zero, considering the gauge invariance of the electrostatic potential.

Thus, for a set of radii  $r_j$  for  $j = 1, 2, \dots, N$ ,

$$a_l(r_j) = \sum_{m=0}^{+\infty} M_{l+2m,l} r_j^{l+2m}, \quad (3)$$

and we can extract an approximation of the  $2l$ -pole strength  $M_{l,l}$  from Fourier modes  $a_l$  by solving a matrix equation [23].

We developed a *Python* program called *STEP File Generator* (or *stepfg*). This program produces 3D STEP (ISO 10303-242 [24]) files from polygonal models specified by vertices. Compared to performing this process manually in CAD software, our software has workflow efficiency advantages. The resulting STEP file can be used in many general 3D, including BEM or finite element method (FEM) solver, programs.

Using the *STEP File Generator*, we effectively extruded a polygonal model of a  $90^\circ$  section of the full cross section of the Muon  $g-2$  quadrupole. Due to the curvature radius  $R = 711.2$  cm being relatively large compared to the half-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

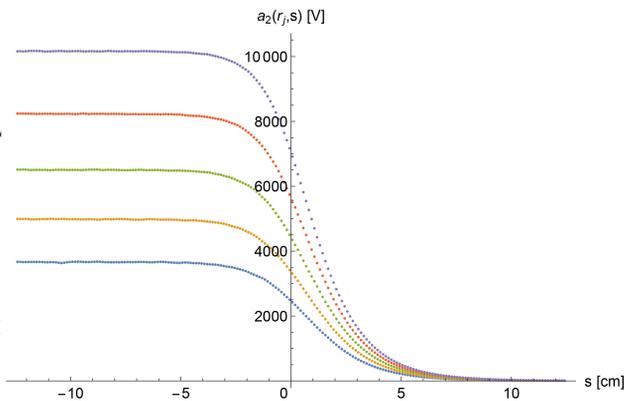


Figure 10: Falloffs of 2nd order Fourier modes  $a_2(r_j)$  calculated at radii  $r = 1.8, 2.1, 2.4, 2.7, 3.0$  cm from Wu's field data. Curves with larger magnitudes correspond to larger radii.

aperture  $d = 5$  cm, the approximation of a straight reference orbit is quite accurate for the purpose of calculating the fringe field.

The electrostatic potential was calculated by Helmut Soltner using *COULOMB*'s BEM field solver at a grid-point set of coordinates at a set of radii

$$r = 1.8, 2.1, 2.4, 2.7, 3.0 \text{ cm} \quad (4)$$

and longitudinal coordinates with the discretization size ranging from  $\Delta z = 0.625$  cm generally to  $\Delta z = 0.078$  cm near the edge of the quadrupole, where the field falloff is the steepest.

From this field data, we calculated the Fourier mode falloffs using the DFT and the Hermitian symmetry at the set of radii listed in Eq. (4). These Fourier mode falloffs are shown in Figs. 10 and 11. We extracted the quadrupole strength falloff and its effective field boundary (EFB)  $z_{\text{EFB}} = 1.2195$  cm for the Muon  $g-2$  quadrupole by applying the method outlined above. We also fitted Enge function coefficients using the Levenberg–Marquardt Gauss–Newton method to the falloff of the quadrupole strength [2, Ch. 3].

For a comparison, we applied [25] the same method of calculating multipole strengths to the electrostatic field data that was obtained for the Muon  $g-2$  quadrupole using *Opera-3d*'s [26] FEM field solver by Wanwei Wu (University of Mississippi), which accounts for the curvature of the reference orbit. In particular, we interpolated Wu's field data and applied the multipole terms extraction method. Additionally, we fitted a nonlinear model defined as the multipole expansion to the raw field data in each cross section and obtained similar results. The field falloffs and the EFBs obtained from Soltner–Valetov and Wu field data are in good agreement, as Fig. 12 shows.

The discrepancies of simple estimates of the tunes based on linear models (Methods 1–3 in [27, 28]) were 1% to 2% in the case of vertical tunes, while for horizontal tunes they were 0.1% to 0.2%. In both cases, the differences between the calculated and the measured tunes exceeded the data

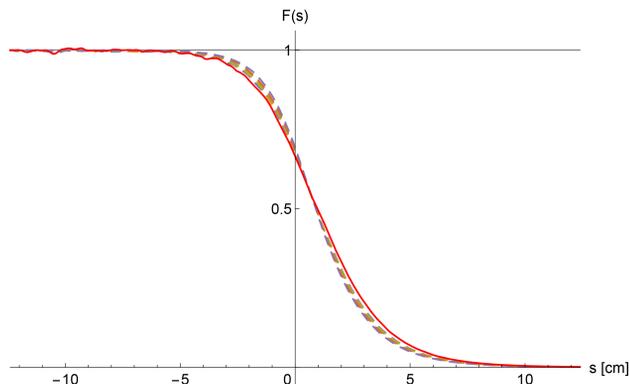


Figure 11: The Fourier modes  $a_2(r_j)$  (dashed plot style, scaled to 1 well inside the quadrupole) alone fall off more quickly than the true quadrupole strength  $M_{2,2}$  (solid red). This is because the second derivative of  $M_{2,2}(s)$  is negative in the beginning of the fringe field and positive on the outside, impacting the additional terms based on the second derivative of  $M_{2,2}$  in Eq. (3).

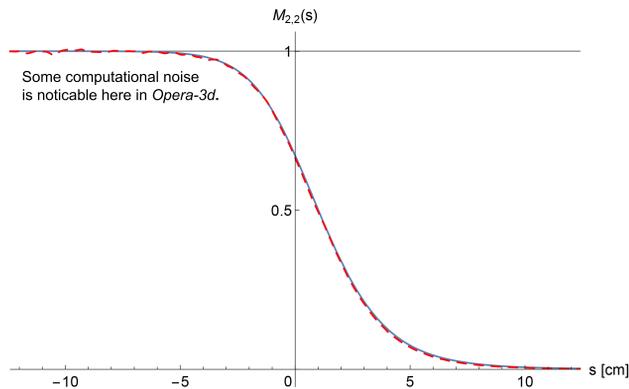


Figure 12: The falloff of the multipole term  $M_{2,2}$  agrees well between calculations based on Soltner–Valetov field data ( $z_{\text{EFB}} = 1.2195$  cm; solid blue) and field data by Wu ( $z_{\text{EFB}} = 1.1233$  cm; dashed red).

errors of the measured tunes. The consideration of fringe fields or EFBs (Methods 5 and 4 in [27, 28], respectively) significantly improved the agreement between the calculated and the experimentally measured tunes for each quadrupole voltage. A consistency between horizontal tunes that were calculated with consideration of the fringe fields and experimentally measured tunes was reported in [29].

## ACKNOWLEDGMENTS

The authors would like to acknowledge the Muon  $g-2$  collaboration, in the context of which the work on the field of the Muon  $g-2$  quadrupole was performed. This work was supported by the U.S. Department of Energy under Contract No. DE-FG02-08ER41546. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

## REFERENCES

- [1] Y. K. Semertzidis *et al.*, “The Brookhaven Muon ( $g-2$ ) Storage Ring High Voltage Quadrupoles,” *Nucl. Instr. Meth. Phys. Res. A*, vol. 503, no. 3, pp. 458–484, 2003.
- [2] E. Valetov, “Field Modeling, Symplectic Tracking, and Spin Decoherence for EDM and Muon  $g-2$  Lattices,” Michigan State University, East Lansing, MI 48824, Ph.D. thesis, 2017, Fermilab report FERMILAB-THESIS-2017-21.
- [3] T. A. Driscoll and L. N. Trefethen, *Schwarz–Christoffel Mapping*, ser. Cambridge Monographs on Applied and Computational Mathematics. Cambridge, UK: Cambridge University Press, 2002.
- [4] L. N. Trefethen, “Numerical Computation of the Schwarz–Christoffel Transformation,” *SIAM J. Sci. and Stat. Comput.*, vol. 1, no. 1, p. 82, 1980.
- [5] K. Makino and M. Berz, “COSY INFINITY Version 9,” *Nucl. Instr. Meth. Phys. Res. A*, vol. 558, no. 1, pp. 346–350, 2006. <http://www.sciencedirect.com/science/article/pii/S0168900205021522>
- [6] M. Berz, *Modern Map Methods in Particle Beam Physics*, ser. Advances in Imaging and Electron Physics. San Diego, CA: Academic Press, 1999. [Online]. Available: <https://cds.cern.ch/record/427002>
- [7] T. A. Driscoll, “Algorithm 756: A MATLAB Toolbox for Schwarz–Christoffel Mapping,” *ACM Trans. Math. Softw.*, vol. 22, no. 2, pp. 168–186, Jun. 1996. [Online]. Available: <http://doi.acm.org/10.1145/229473.229475>
- [8] INTEGRATED Engineering Software. *COULOMB*. <https://www.integratedsoft.com/product/coulomb/>.
- [9] E. Valetov and M. Berz, “E989 Note 144: Main Field of the Muon  $g-2$  Storage Ring High Voltage Quadrupole,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-11655, 2018.
- [10] E. Valetov and M. Berz, “E989 Note 100: Main Field of the Muon  $g-2$  Collaboration Quadrupole in the Case of Mismatched Plates,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-4628, 2016.
- [11] E. Valetov and M. Berz, “Analytic Quad Fields Under Mismatching and Geometric Errors: March 2017 Update,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-5552, 2017.
- [12] E. Valetov and M. Berz, “Potential Multipoles of the Brookhaven Muon ( $g-2$ ) Storage Ring High Voltage Quadrupoles,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-4280, 2016.
- [13] E. Valetov and M. Berz, “Quadrupole Multipoles from Conformal Maps, 3D Fringe Fields, and on to Full Ring Dynamics,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-3723, 2016.
- [14] D. Crowdy, “The Schwarz–Christoffel Mapping to Bounded Multiply Connected Polygonal Domains,” *Proc. R. Soc. A*, vol. 461, no. 2061, pp. 2653–2678, 2005.
- [15] L. Banjai, Private communication.
- [16] T. A. Driscoll and S. A. Vavasis, “Numerical Conformal Mapping Using Cross-Ratios and Delaunay Triangulation,” *SIAM J. Sci. Comput.*, vol. 19, no. 6, pp. 1783–1803, 1998.
- [17] L. Banjai, “Revisiting the Crowding Phenomenon in Schwarz–Christoffel Mapping,” *SIAM J. Sci. Comput.*, vol. 30, no. 2, pp. 618–636, 2008.
- [18] D. Tarazona, E. Valetov, M. Berz, and K. Makino, “Quad Plate off Scenario in Storage Ring,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-7481, 2017.
- [19] D. Tarazona, M. Berz, and K. Makino, “Muon Losses - Scraping Studies COSY Update,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-8882, 2017.
- [20] M. Berz, K. Makino, E. Valetov, and D. Tarazona, “Quad Voltage / Quad Voltage = Oomph: A Quantitative Oomph Theory,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-12110, 2018.
- [21] D. Tarazona, M. Berz, and K. Makino, “Influence of Bad Resistors on BD Variables,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-14313, 2018.
- [22] E. Valetov and M. Berz, “Fringe Field of the Muon  $g-2$  Storage Ring High Voltage Quadrupole,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-12119, 2018.
- [23] R. Jagasia, Private communication.
- [24] ISO/IEC, *IS 10303-242:2014 Industrial Automation Systems and Integration – Product Data Representation and Exchange – Part 242: Application Protocol: Managed Model-Based 3D Engineering*. Geneva, Switzerland: International Organization for Standardization, 2014.
- [25] E. Valetov and M. Berz, “Comparison of Field Falloffs Obtained From Two Field Datasets for the Muon ( $g-2$ ) Collaboration High Voltage Quadrupole,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-7478, 2017.
- [26] *OPERA-3d User Guide*. Oxford, England: Vector Fields Limited, 2004.
- [27] K. Makino, E. Valetov, and M. Berz, “E989 Note 104: Detailed Linear Optics and Tunes of the  $g-2$  Ring,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-5715, Mar. 2017.
- [28] K. Makino, E. Valetov, and M. Berz, “E989 Note 110: Linear Tune Tables for the  $g-2$  Ring,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-6791, June 2017.
- [29] D. Tarazona, M. Berz, and K. Makino, “Tune/Resonances,” Muon  $g-2$  Collaboration, Fermi National Accelerator Laboratory, Batavia, IL, G-2 Experiment Document GM2-doc-12921, 2018.

# A COMPACT PERMANENT MAGNET SPECTROMETER FOR CILEX

M. Khojayan<sup>†</sup>, A. Cauchois, J. Prudent, A. Specka

LLR (Laboratoire Leprince-Ringuet), CNRS and Ecole Polytechnique, Palaiseau UMR7638, France

## Abstract

Laser Wakefield acceleration experiments make extensive use of small permanent magnets or magnet assemblies for analyzing and focusing electron beams produced in plasma accelerators. Besides being compact, these magnets have to have a large angular acceptance for the divergent laser and electron beams which imposes constraint of the gap size. We will present the optimized design and characterization of a 100 mm long, 2.1 Tesla permanent magnet dipole. Furthermore, we will present the performance of such a magnet as a spectrometer in the CILEX/APOLLON 10 PW laser facility in France.

## INTRODUCTION

CILEX (Centre Interdisciplinaire de la Lumiere Extrême / Interdisciplinary Center for Extreme Light) is a research center which aims at using the Apollon-10P laser for exploring laser-matter interaction at extremely high laser intensities ( $\sim 10^{22}$  W/cm<sup>2</sup>). The long focal length area of CILEX will be used to investigate plasma acceleration and radiation generation. It will be equipped with two interaction chambers able to accommodate laser focal lengths ranging from 3 m to 30 m. The spectrometer magnet presented here is designed to be compatible with the mentioned laser parameters. Sketch of the laser-plasma interaction chamber is shown in Fig. 1. The total volume of interaction chamber is about 3 m<sup>3</sup> and it is practical to use permanent magnets for characterization of electron beam. Permanent magnets, as compared to electromagnets, do not need power supplies (and consequently cooling system) and can be more compact due to smaller apertures. In our case, the apertures of the magnets will be limited by the envelope of the laser, which according to our requirements, should pass through the magnets unobstructed. The paper

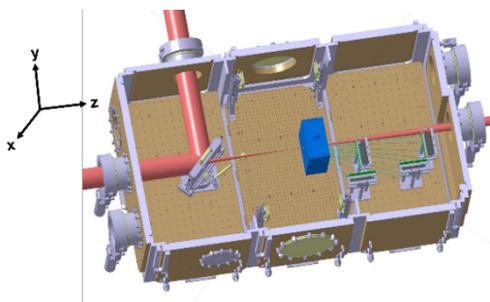


Figure 1: Schematic of interaction chamber at CILEX with laser envelope shown in purple and dipole magnet in blue colors. Measurement screens and reference electron trajectories (green) are shown as well.

is organized as follows. First, the design considerations of a permanent magnetic dipole are introduced together with

<sup>†</sup> martin.khojayan@llr.in2p3.fr

analytical and computational field estimations. Next, construction of a magnet and field measurements are presented. Finally, the result of beam dynamics simulations implementing 3D magnetic field data are presented by applying the magnet as a spectrometer inside the interaction chamber of the CILEX facility.

## MOTIVATION

Let us consider a simple C shaped dipole (Fig. 2). The field in the gap of height  $h_g$  is driven by a permanent magnet of the same height  $h_{pm} = h_g$  and of width  $w_{pm}$ . Assuming constant field in the gap (no horizontal field component) it is straightforward to calculate the magnetic flux density using Ampere's and flux conservation laws:

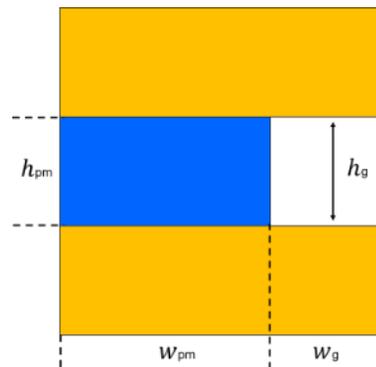


Figure 2: 2D view of a C shaped dipole with a permanent magnet shown in blue and steel/iron in yellow colors.

$$B_g = -\frac{B_r w_{pm}}{w_g + \frac{\mu_r h_g w_{pm}}{h_{pm}}} \quad (1)$$

with  $B_r$  and  $\mu_r$  being the remnant field and relative permeability of the permanent magnet. From Eq. (1) it follows that the maximum achievable field in the gap cannot exceed the remnant induction field of the magnet independent of how small the gap height ( $h_g$ ) is for reasonably large magnet width ( $w_g$ ). It becomes clear that for a permanent magnet (PM) dipole, to reach fields higher than remnant field of an individual magnet, a special arrangement of magnetic sub-materials is necessary. Common structures for generating very strong fields are Halbach [1] and Stelter [2] configurations. Nowadays, permanent magnet dipoles are widely used in magnetic resonance imaging applications [3], in facilities such as third [4] and next [5] generation light sources. Neodymium iron Boron (Nd-Fe-B) magnets are the desired candidates for generating strong fields due to their high remnant induction and the highest up to date BH energy product [6]. Moreover, almost linear behaviour of the demagnetization curve [7] and relative permeability close to unity makes the analytical design of such systems relatively straight forward.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## DESIGN OF A PM DIPOLE

### Analytic Estimation of Magnetic Field

For design considerations of a magnetic dipole we assume that the magnet is H shaped with a 2D schematic view (upper half) shown in Fig. 3. The parts shown in yellow color represent iron material: pole tip as well as the yokes for the flux circulation. The pole tip is surrounded by neodymium iron boron magnets with arrows indicating magnetization direction of each magnet. In the same scheme  $h_g$  is half the gap height and  $w/h$  signify the width / height of each surrounding magnet. The main idea of using such geometry is the collection of the flux from the surrounding permanent magnets and concentration into the magnet gap. The strength of the magnet built in this way can therefore exceed the residual field ( $B_r$ ) of the permanent magnet materials and, in principle, reach the saturation field of pole material which can be more than 2 Tesla. Next, applying flux conservation and Ampere's laws [8] for the above magnetic circuit (neglecting fringe field effects), one can obtain the following expression for magnetic flux density in the gap:

$$B_g = -\frac{1}{h_g} \frac{(w_1 + w_2 + w_3)B_r}{\left(\frac{w_g}{h_g} + \frac{\mu_r w_1}{h_1} + \frac{\mu_r w_2}{h_2} + \frac{\mu_r w_3}{h_3}\right)} \quad (2)$$

It is possible to further optimize the latter expression assuming a square shape magnet poles and similar sizes for surrounding magnets. Since the gain in magnetic field over magnet weight is an important measure in designing the magnet, Fig. 4 illustrates such a dependence for 10 mm gap height. The choice of the gap value is dictated from the magnet exit position compatible with the laser envelope at that position. The indicated field / weight value (red asterisk) on the plot corresponds to a pole size of 51 mm which was chosen according to the commercially available sizes of Nd-Fe-B 40 grade magnets and not from the beam dynamics requirements.

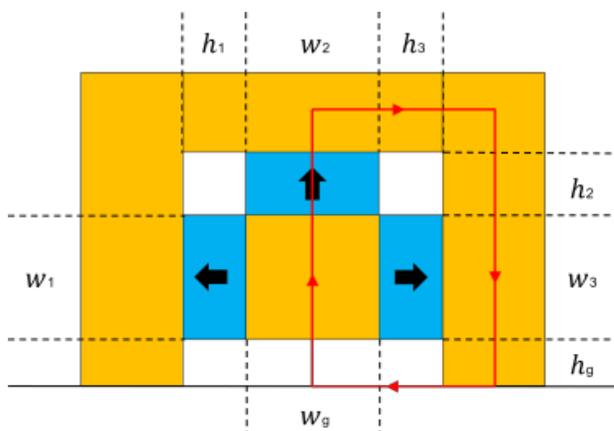


Figure 3: Schematic (2D) view of top half of H shaped magnet with a red flux line following Ampere's law. Iron is shown in yellow with the pole tip surrounded by Neodymium N40 grade magnets expressed in light blue.

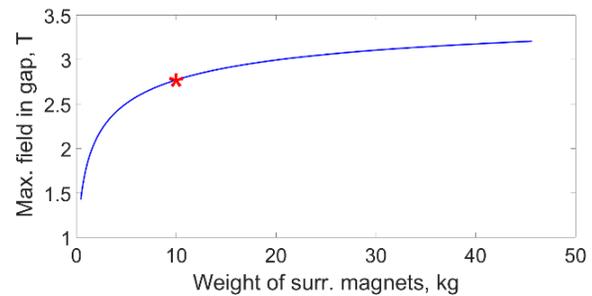


Figure 4: Peak magnetic field vs. total weight of the surrounding magnets shown in Fig. 2. Red asterisk represents the values assuming 51 mm size of each square magnet pole. 10 mm gap size/height is assumed in the plot.

### Computation of Magnetic Field

Magnet geometry of Fig. 3 has been studied using three-dimensional TOSCA software [9]. Following dimensions have been used in the calculations:  $h_1 = h_3 = 5.08$  cm,  $w_1 = w_3 = 7.08$  cm,  $h_2 = w_2 = 5.08$  cm. For the case of neodymium magnet, by taking advantage of linear BH behaviour in the second quadrant of BH curve, the slope  $\mu_r$  can be calculated depending on the grade of the magnet (1.0691 in the case of the Nd-Fe-B 40 grade magnets). Non-linear iron properties have been used in TOSCA for induction versus field data. Due to the symmetry, only one eighth of the magnet geometry was created in 3D TOSCA modeller. A block having a scale factor 3 in all dimensions has been applied as a region of background. The magnetization direction of each magnet was defined by means of Euler angles. 2.5 mm step for meshing yielded over 1.5 million node points for calculation. The aim was to reach magnetic field at the gap as close as possible to the saturation field (2.26 Tesla in this case) of the pole material. A fixed ratio of 20 for magnet gap height over length was imposed. Figure 5 illustrates the expected peak magnetic flux density at the magnet gap for different gap heights obtained from TOSCA simulations (red circles). In the same plot analytical estimation of the field is shown applying an empirical scale factor  $2.26 / 2.96$ , where 2.96 Tesla corresponds to the analytical peak field value for 10 mm gap allowing infinitely high saturation field for the iron. It can be seen from the plot that for the case of 10 mm gap, the value of peak magnetic field in the gap is close to the saturation field of the iron pole and, in the same time, for larger gap sizes the considered geometry is not optimum anymore.

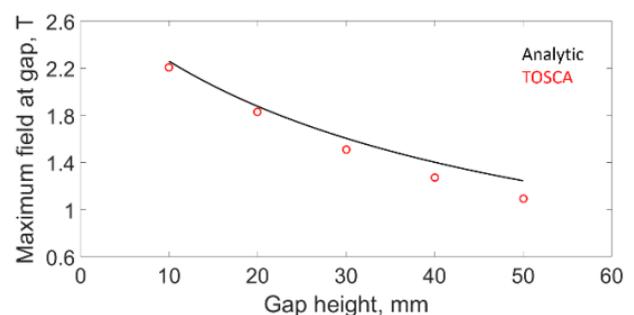


Figure 5: Magnetic induction in the gap for various gap sizes. Red dots: TOSCA calculations. Black curve: analytical estimation applying a factor 0.7635.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## Construction and Measurements

The PM dipole has been constructed at LLR (Laboratoire Leprince-Ringuet) using Nd-Fe-B 40 grade magnets as surrounding magnet material and allied pure iron (2.18 T saturation field) as pole and yoke material. A dedicated mechanical tooling and careful adjustment were the key points during the magnet assembly due to strong force existent between the magnets and corrosion risks that are not negligible for neodymium magnets. Furthermore, special arrangement of 56 magnets was needed during the assembly to balance the field imperfections of individual magnets (Fig. 6). In Fig. 6, the dimensions ( $w \times h \times l$ ) of the magnets are shown in legends. However, it is believed, that geometry using iron poles will smoothen the effects of field imperfections and defects in magnet once the pole material

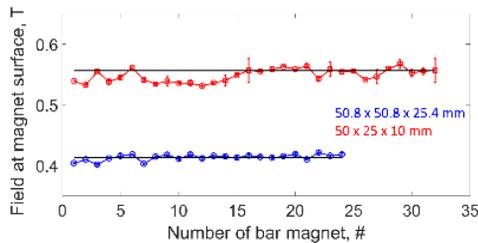


Figure 6: Measured magnetic induction at the surface of two types of neodymium magnets. For each case, black lines represent analytically estimated field values assuming 0.5 mm offset of Hall probe sensor from the magnet surface.

is saturated. The assembled dipole and corresponding TOSCA model of the magnet are presented in Fig. 7. Magnets shown in blue have been added to strengthen the flux on both sides of the poles. The length of the dipole was 100 mm. The assembled magnet was measured at the measurement bench of LLR laboratory. The position of Hall probe was aligned to the magnet using an optical system with high magnification camera (50  $\mu\text{m}$  pixel resolution) in the field of view of 5 cm diameter centered on the pole gap at the magnet exit face (Fig. 8). The Hall probe was mounted on linear translation stages which, according to manufacturer, have 0.07 microradians angular slope and 2 micrometers bidirectional repeatability [10]. Measured and simulated field profiles are plotted in Fig. 9.

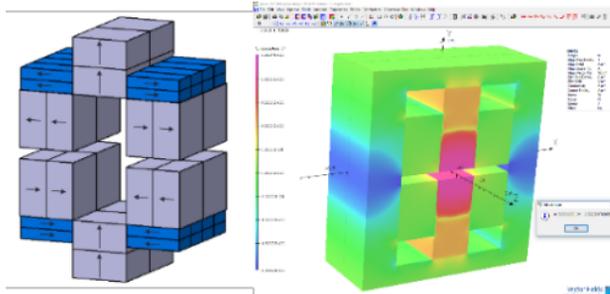


Figure 7: Left: overview of different block magnets used to assemble the dipole. Arrows on the blocks illustrate the magnetization directions. Right: TOSCA model of 100 mm long magnet yielding 2.092 T peak field value in the gap. The main component of induction field is shown in color map from blue (lowest value) to red (highest value).

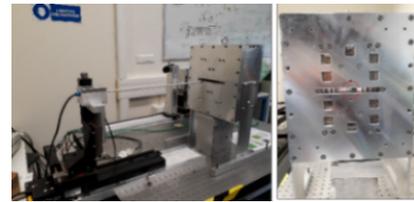


Figure 8: Left: LLR measurement bench with a Hall probe mounted on linear transition stages. Right: defined area of interest for hall probe alignment using optical system.

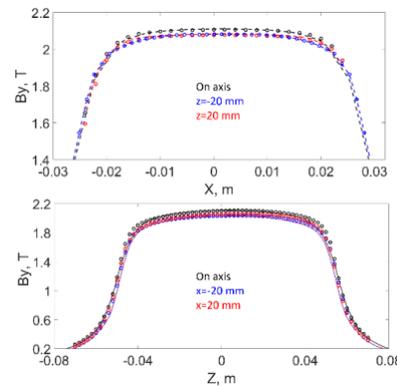


Figure 9: On and off-axis main component field profiles along horizontal (upper plot) and longitudinal (lower plot) directions. For each case dots correspond to the measured and curves to the simulated field values. A scaling factor ( $\sim 1\%$ ) is applied to match the measured and simulated field values.

Table 1: Summary of Magnet Characteristics

Parameter	Value
Gap height, mm	10
Peak field at gap center, T	2.092
Horizontal good field region ( $\sim 10\%$ inhomogeneity), mm	$\sim 50$
Horizontal full gap region, mm	$\sim 150$
Length, mm	100
Height, mm	260
Width, mm	240
Weight, kg	$\sim 50$

The result is remarkable yielding to less than 1% disagreement between predicted and measured field values. The characteristics of the dipole are given in Table 1. No field weakening has been observed as a result of demagnetization effects [11] due to a very strong field at the specific corners of individual magnets (see TOSCA simulation of the magnet on the right side of the Fig. 7).

## PM DIPOLE AS A SPECTROMETER

### Implementing Magnetic Field Data into ASTRA

ASTRA [12] simulation software has been applied to study the dynamics of different energy electrons through the designed magnet. The magnetic field data has been implemented into ASTRA by means of a Cartesian grid of

2 mm × 2 mm × 5 mm ( $H \times W \times L$ ) step size with corresponding values of field components at each grid point. An idealized electron bunch having Gaussian distribution in all sub spaces was assumed in simulations. Electron beam parameters are summarized in Table 2. The considered energy range of [56–2350] MeV was defined as the lowest energy exiting the magnet from the side of the mechanical gap and the highest energy which is measurable out of 50 milliradians divergent laser cone within the interaction chamber. A 3D space charge routine [13] was applied in ASTRA by using 0.5 million macro particles in a bunch. Electron reference trajectories and various screens (marked as S#) are illustrated in Fig. 10.

Table 2: Electron Beam Parameters used in ASTRA

LPA beam parameter	Value
Charge, pC	10
Peak current, kA	1.2
Energy, MeV	56–2350
Transverse rms size, $\mu\text{m}$	2.5
Transverse rms divergence, mrad	2

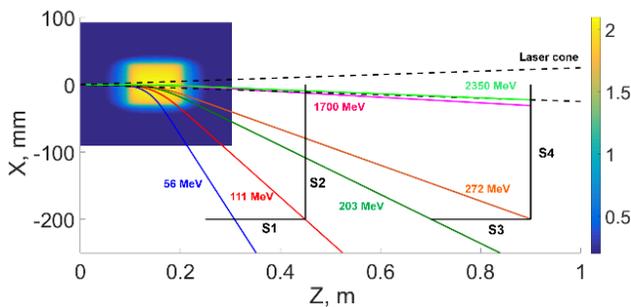


Figure 10: Range of electron trajectories and corresponding screen positions/orientations for energy measurement inside the interaction chamber of CILEX. The laser opening is exemplified by dashed black lines.

### Energy Resolution

For each the above-mentioned energy (assuming negligible energy spread) the rms resolution at each screen position was estimated as the ratio of rms beam size in the horizontal plane and the mean beam position. The orientations of the screens have been taken into account in the simulations. Figure 11 plots the beam envelopes at the positions of S1 and S2. In the plot the black line indicates the separation of the screens. Electrons of energies below 200 MeV are strongly affected by the fringe field focusing / defocusing effects. In addition, the fringe field effects may be further enhanced / diminished by adjusting the magnet laterally. The energy resolution at various screens and at different energies is summarized in Fig. 12 yielding less than 8% rms value within the whole energy range for 2 milliradians divergence of electron beam.

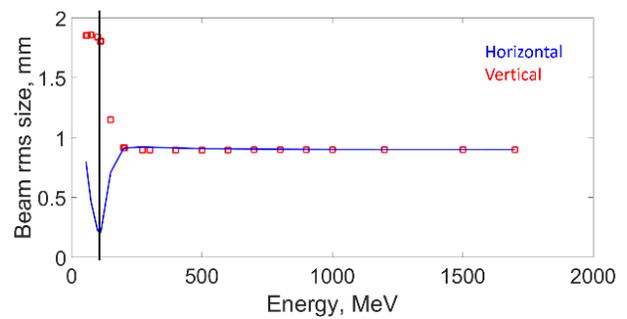


Figure 11: Beam rms transverse sizes at screen positions of S1 and S2. Blue line: horizontal, red rectangles: vertical.

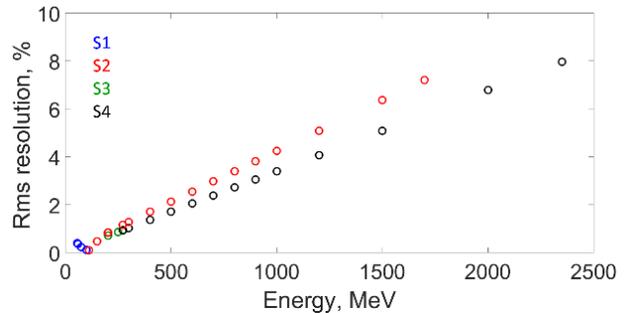


Figure 12: RMS energy resolution estimated at different screens for 2 milliradians beam divergence. Different colors symbolize the outcome at different screen positions.

## SUMMARY

A 2.1 Tesla permanent magnet dipole has been designed, constructed, measured and characterized as a spectrometer for CILEX. Very good agreement ( $\sim 1\%$ ) between predicted and measured magnetic field values has been obtained. A precise electron beam tracking has been done afterwards by applying the 3D field map of the magnet into ASTRA simulation program. According to tracking results the magnet will enable to measure electrons from 56 MeV to 2.35 GeV with rms resolution of below 8 percent at full energy range. Besides, for smaller laser divergence the magnet can be shifted downstream and the energy resolution may be improved by adding a focusing element into the configuration.

## ACKNOWLEDGMENTS

The work is partially supported by the European Union's Horizon 2020 Research and innovation program under grant agreement No 730871.

## REFERENCES

- [1] K. Halbach, "Application of permanent magnets in accelerators and electron storage rings", *Journal of Applied Physics*, vol. 57, pp. 3605, 1985.
- [2] R. E. Stelter, "Dipole permanent magnet structure", United States patent, 1997.
- [3] K. Halbach, "Design of permanent multipole magnets with oriented rare earth cobalt material", *Nuclear Instruments and Methods*, vol. 169, 1980.

- Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.
- [4] C. Li, “Efficiency of permanent magnet assemblies for MRI devices”, *IEEE Transactions on Magnetics*, vol. 41, pp. 10, 2005.
  - [5] G. Tosin *et al.*, “Bending magnets made with permanent magnets for LNLS-2 electron storage ring”, in *Proc. PAC’09*, Vancouver, Canada, May 2009, paper MO6PFP001.
  - [6] S.C. Gottschalk *et al.*, “Permanent magnet systems for free-electron lasers”, *Nuclear Instruments and Methods in Physics Research A*, vol. 507, pp. 181-185, 2003.
  - [7] J. F. Herbst *et al.*, “Structural and magnetic properties of Nd<sub>2</sub>Fe<sub>14</sub>B”, *J. Appl. Phys.*, vol. 57, pp. 4086, 1985.
  - [8] B. C. Brown, “Design formulas for the strength, compensation and trimming of hybrid permanent magnets”, FERMILAB-Conf-96/273, 1996.
  - [9] TOSCA, Opera-3d, Cobham Technical Services, Software for electromagnetic design, Oxford, UK <http://operafea.com/>
  - [10] Thorlabs: Linear Translation Stages, [https://www.thorlabs.com/newgrouppage9.cfm?objectgroup\\_id=7652](https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=7652)
  - [11] M. Katter, “Angular dependence of the demagnetization stability of sintered Nd-Fe-B magnets”, *IEEE Transactions on Magnetics*, vol. 41, pp. 10, 2005.
  - [12] K. Floettmann, ASTRA, <http://www.desy.de/mpyf10>
  - [13] G. Poplau *et al.*, “3D space charge calculations for bunches in the tracking code ASTRA”, in *Proc. EPAC’06*, Edinburgh, UK, Jun 2006, paper WEPC121.

# SIMULATIONS OF BEAM CHOPPING FOR POTENTIAL UPGRADES OF THE SNS LEBT CHOPPER\*

B. X. Han<sup>†</sup>, V. V. Pevlov, R.F. Welton, R. B. Saethre, S.N. Murray Jr., T. R. Pennisi, C.M. Stinson, M. P. Stockli, SNS, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

## Abstract

The LEBT chopper is a critical element of the SNS accelerator system. In this work, the benefit of increasing the chopping voltage amplitude for the present chopping pattern is shown with beam simulations, and an ongoing hardware upgrade of the chopper pulser units is discussed. In addition, with the prospect of higher voltage capability of the new pulser design, two alternative chopping patterns which reduce the switching frequency of the chopper pulsers down to  $\frac{1}{2}$  or  $\frac{1}{4}$  of the present chopping pattern, are also explored with beam simulations.

## INTRODUCTION

The Spallation Neutron Source (SNS) accelerator system consists of a 65 keV H<sup>-</sup> injector, a 2.5 MeV RFQ, a 1 GeV linac chain (DTL-CCL-SCL), and a proton accumulator ring. The H<sup>-</sup> injector is made of a RF-driven, Cs enhanced H<sup>-</sup> ion source and a two-lens electrostatic low energy beam transport (LEBT). The injector feeds the RFQ accelerator with 1 ms long H<sup>-</sup> beam pulses at 60 Hz. Figure 1 shows a cross-section view of the SNS H<sup>-</sup> injector.

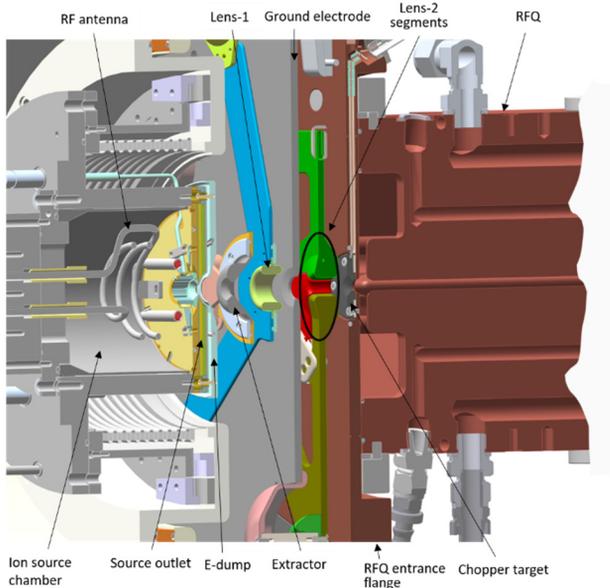


Figure 1: A cross-section view of the SNS H<sup>-</sup> injector.

To facilitate the multi-turn beam stacking in the accumulator ring and to create gaps for clean beam extraction from the ring, the 1 ms H<sup>-</sup> pulses are chopped  $\sim 300$  ns at the ring revolution frequency ( $\sim 1$  MHz) in the LEBT in front of the RFQ. The second lens electrode of the LEBT is

\* This work was performed at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, under contract number DE-AC05-00OR22725 for the United States Department of Energy.

<sup>†</sup> hanb@ornl.gov

azimuthally split into four segments to allow applications of various transverse electric fields on top of the lens voltage for beam steering (misalignment correction), chopping or blanking. A donut-shape TZM plate surrounding the RFQ entrance aperture serves as chopper target as shown in Figure 1 [1, 2].

## THE PRESENT CHOPPING PATTERN AND BEAM SIMULATION

### The Present Chopping Pattern

The four segments of the lens-2 are driven independently by four bipolar high voltage pulsers for beam chopping. With the present chopping pattern, a pair of neighbouring two segments are driven to negative potential and the other two are driven to positive potential, i.e. the lens-2 segments are driven as two opposing pairs with opposite potentials, to generate the transverse field needed for beam chopping. The waveforms of the four pulsers are configured in a manner as shown in Figure 2 so that the beam deflection is sequentially rotated to four different directions to avoid sputtering and heat loading at a single spot on the chopper target. The lens-2 is oriented in a way that the beam deflection directions coincide with the gaps between the neighbouring vanes of the RFQ to minimize ions impacting the vanes if the beam is not completely intercepted at the chopper target during beam chopping [3].

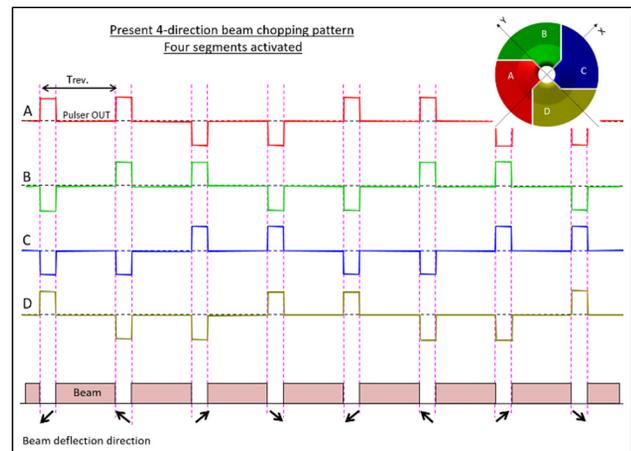


Figure 2: Bipolar chopping voltages applied sequentially on opposing pairs of the lens-2 segments.

### Beam Simulations

The SIMION 8.1 code [4] was used to track the ions starting from the ion source outlet aperture to the RFQ vanes. A total of 10000 H<sup>-</sup> ion macro-particles were launched with a total charge of  $1.5 \times 10^{-8}$  coulomb over  $0.25 \mu\text{s}$  to simulate the space charge effect of a 60 mA

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

beam. Typical operational voltage settings on the ion source and LEBT electrodes were used for beam transport, and different chopping voltages were superimposed on the lens-2 segments to examine the beam chopping performance. The tips of the RFQ vanes were included in the geometry, but the RFQ fields were not modeled. Shown in Figure 3 is a simulation of beam transport through the LEBT and injection into the RFQ. In Figure 4, the X-X' phase space distribution output from the SIMION simulation is plotted at the RFQ injection reference plane along with an output from PBGUNS (a 2-D code capable of plasma emission model) [5] simulation and a real emittance measured on an injector test stand. The RFQ acceptance ellipse with 4x normalized rms emittance,  $4\epsilon_{n,rms} = 1.4 \pi \text{ mm-mrad}$ , and  $\alpha = 1.6$ , and  $\beta = 0.06 \text{ mm/mrad}$  is overlaid on the phase-space plots. The SIMION output agrees reasonably well with PBGUNS output and (more importantly) with the real measurement data in terms of overall beam size and angle.

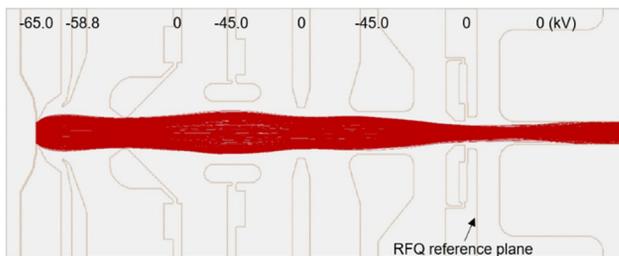


Figure 3: SIMION simulation of beam transport through the LEBT and injection into the RFQ.

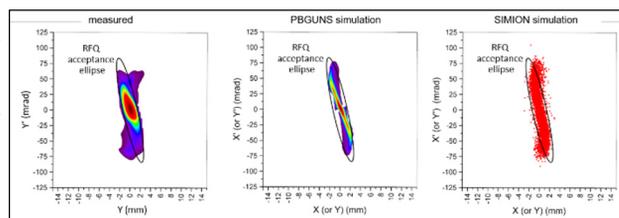


Figure 4: X-X' plots of the injected beam at the RFQ injection reference plane for a measured emittance and outputs from simulations with PBGUNS and SIMION.

Figure 5 shows a case of simulation of beam chopping. Bipolar voltages  $\pm 2.5 \text{ kV}$ , the maximum voltage amplitude available for routine operation, were applied on opposing two pairs of the lens-2 segments representing one of the four chopping cycles illustrated in Figure 2. This figure shows a cross-section view on the plane of beam deflection and its isometric view is shown at the bottom.

To evaluate beam chopper performance in simulations, first of all, we examine the primary goal of beam chopping, i.e. deflecting the beam out of the RFQ acceptance either in real space or in position-angle phase space; meanwhile we also check the beam spot profile on the chopper target and the behavior of the fraction of the beam that enters the RFQ cavity if the beam was not fully intercepted at the chopper target. Figure 6 shows the deflected beam (in red dots) at three longitudinal locations, 1) at the front face of

the chopper target, plotted vs. the RFQ entrance aperture and the injected beam (green dots) in real space, 2) at the RFQ reference plane, plotted vs. the RFQ acceptance ellipse and injected beam in X-X' phase space (X axis is aligned with the beam deflection direction), and 3) at the tips of the RFQ vanes. Simulations were conducted for higher voltage amplitudes with  $\pm 3.0 \text{ kV}$  and  $\pm 3.5 \text{ kV}$  in addition to the case of  $\pm 2.5 \text{ kV}$ , and the results are shown in Figure 6.

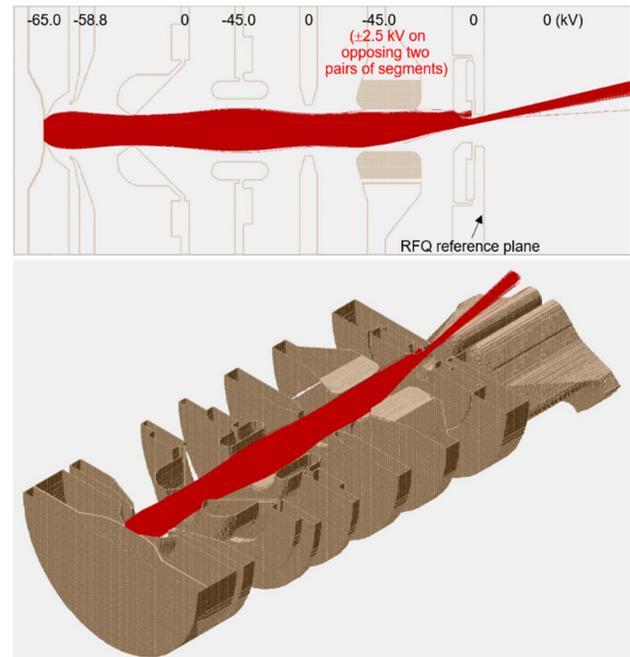


Figure 5: SIMION simulation of beam chopping with  $\pm 2.5 \text{ kV}$  on the opposing two pairs of the lens-2 segments.

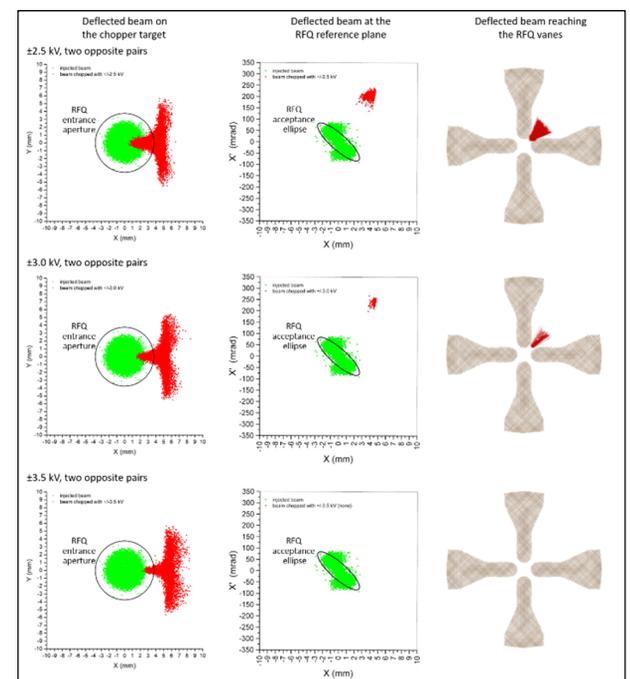


Figure 6: Simulation outputs for chopping with  $\pm 2.5$ ,  $\pm 3.0$ , and  $\pm 3.5 \text{ kV}$  on opposing pairs of lens-2 segments.

It can be seen from Figure 6 that chopping with  $\pm 2.5$  kV deflects the beam adequately out of the RFQ acceptance ellipse. In practice, post-RFQ beam current waveforms show clean gaps created with  $\pm 2.5$  kV chopping, as shown in Figure 7.

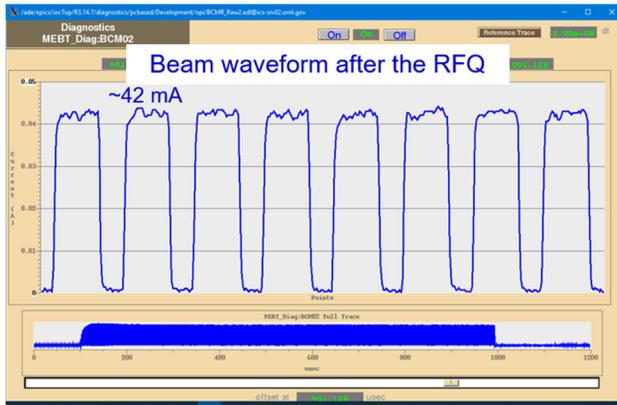


Figure 7: Chopped post-RFQ beam waveform.

However, the simulation indicates that a fraction of the beam is able to reach and impact on the RFQ vane tips. Erosion damage has been observed on the vane tips of the old RFQ which was replaced with a new one after  $\sim 16$  years operation. While this is not the only scenario of ions impacting on the vane tips (others include blanking a short beam at the start and end of the 1 ms pulses, and inadequate deflection of ions during the unavoidable rise and fall time of chopper pulsers [6]), it is important to mitigate the effect as much as possible. Increasing the chopping voltage amplitude is desired to reduce the amount of beam entering the cavity and reaching the vane tips during chopping. The simulation suggests, chopping with  $\pm 3.0$  kV further separates the deflected beam from the RFQ acceptance ellipse and seems to also clear the beam from RFQ vanes. Increasing the voltage to  $\pm 3.5$  kV appears to completely deflect the beam out of the RFQ entrance aperture and eliminate beam entering into the RFQ cavity during chopping.

## THE CHOPPER PULSER HARDWARE UPGRADE EFFORT

The existing high voltage pulsers of the SNS LEBT chopper have become obsolete in terms of spare parts. Furthermore, they are functionally limited to maximum voltages of about  $\pm 2.5$  kV due to arcing and power dissipation issues. A new bipolar pulser has been designed, and a prototype has been built and tested on a test-bench. The pulser uses fast push-pull transistor BEHLKE switch modules. A schematic of the new pulser unit is shown in Figure 8. According to test-bench results, the new pulser is capable of driving  $\pm 3.5$  kV and up to  $\pm 5.0$  kV is expected with further development. The new pulser unit also features reduced rise/fall time, improved timing stability, minimized power losses and better cooling [7].

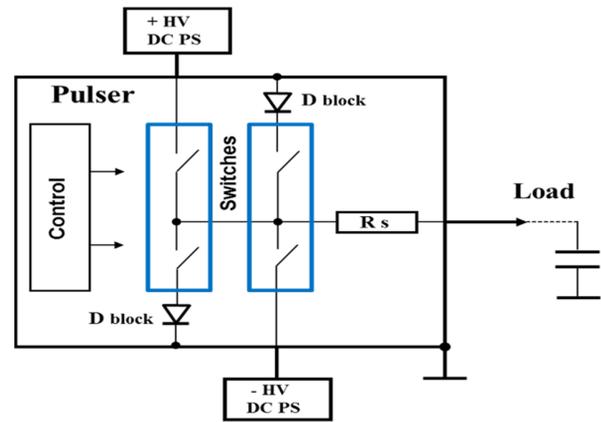


Figure 8: A schematic of the new design pulser unit.

## EXPLORING POSSIBLE ALTERNATIVE CHOPPING PATTERNS

### Proposed New Chopping Patterns

With the prospect of increased high voltage capability of the chopper pulser hardware, there are possibilities of alternative new chopping patterns. These involve activation of only two or even just one segment of the lens-2 at a time during beam chopping, and thus reducing the switching frequency of each of the four pulsers down to  $\frac{1}{2}$  or  $\frac{1}{4}$  of the present chopping pattern. With reduced switching frequencies, the switches have more time to recover from charging and discharging cycles. The average currents flowing through the switches are also reduced. Less stress and less heat dissipation are desirable for operational reliability of fast HV switches.

Figures 9 and 10 are voltage waveforms of the four individual segments of lens-2 for the two alternative chopping patterns being considered. To deflect the beam in the directions of gaps between the RFQ vanes in the same way as the present chopping pattern, the lens-2 needs to be rotated azimuthally by  $45^\circ$ . The chopping pattern of Figure 9 was proposed in [6] with preliminary beam simulation results.

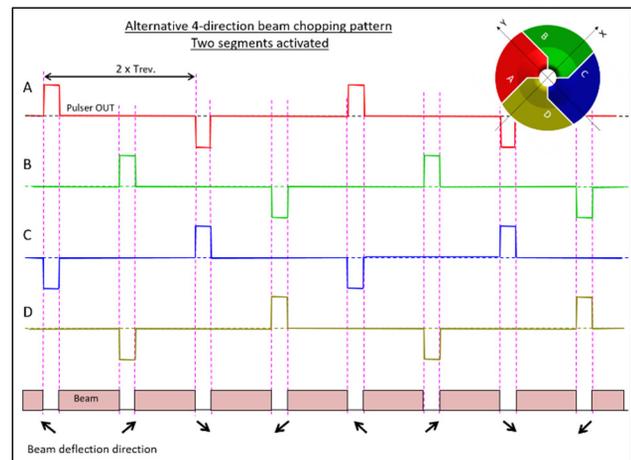


Figure 9: Bipolar chopping voltages applied sequentially on opposing segments of lens-2.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

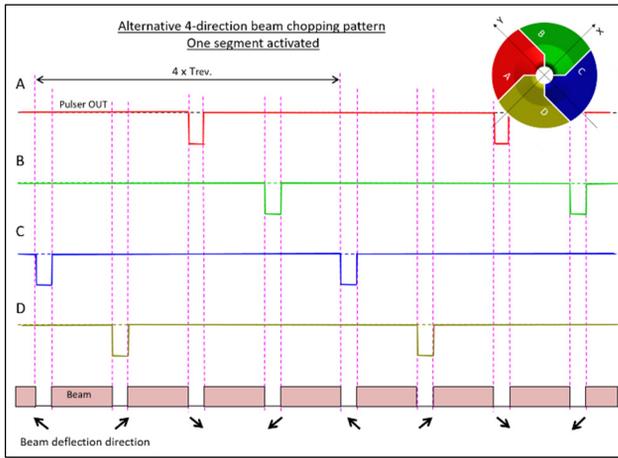


Figure 10: A negative chopping voltage applied sequentially on segments of lens-2.

### Simulations of Beam Chopping

Simulations were conducted for the proposed new chopping patterns with different voltage amplitudes for the opposing two segments starting from  $\pm 2.5$  kV up to  $\pm 5.0$  kV and  $-8.0$  kV for the single segment case. Figure 11 shows visualization of beam deflection for the  $\pm 2.5$  kV case, as in Figure 5. Figure 12 summarizes the outputs of all simulation runs for the new chopping patterns in the same way as in Figure 6.

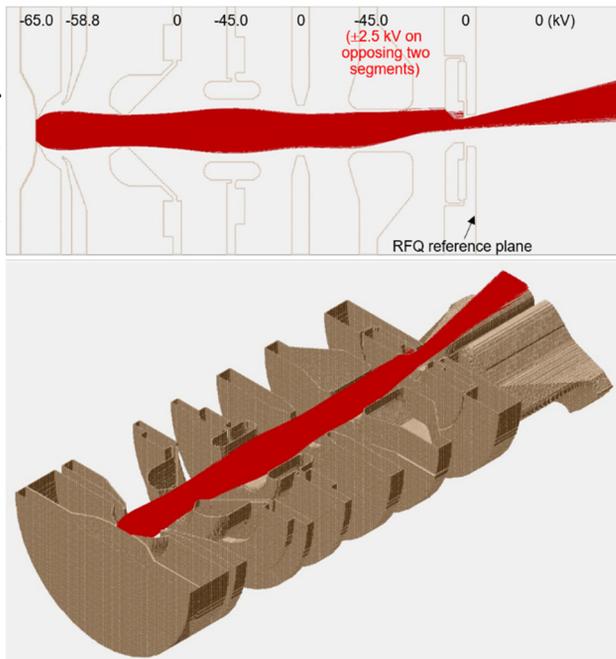


Figure 11: SIMION simulation of beam chopping with  $\pm 2.5$  kV on the opposing two segments of lens-2.

The simulation output suggests that the proposed new chopping patterns require higher voltage, such as  $\pm 4.0$  kV ( $-8.0$  kV for single segment case) or more to achieve separation of the beam from the RFQ acceptance ellipse with reasonable margin. Clearing the RFQ vane tips from impacting ions would require even higher voltage - beyond

the capability of the chopper pulser under development. For the single segment chopping case, it is a serious challenge at this point to achieve the required voltage amplitude even with the new chopper pulser.

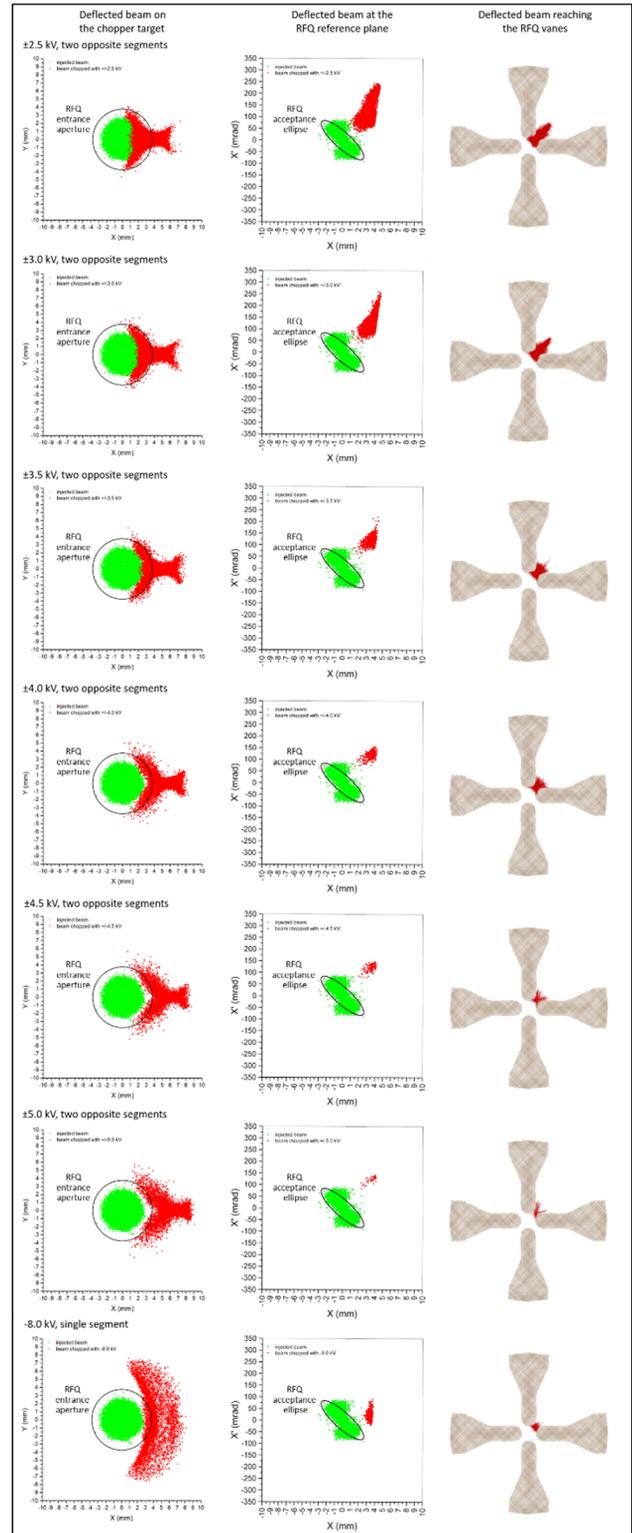


Figure 12: Simulation outputs for chopping with  $\pm 2.5$ ,  $\pm 3.0$ ,  $\pm 3.5$ ,  $\pm 4.0$ ,  $\pm 4.5$ ,  $\pm 5.0$  kV on opposing segments or  $-8.0$  kV on a segment of lens-2.

### Concern with Beam Scraping on lens-2

As shown in Figure 13, for the same voltage difference, the Ex field inside the lens-2 aperture is substantially lower for the case of 2-segment chopping pattern compared to the two-pair pattern, especially in the outer region along the Y axis. So, much higher voltage difference is needed to achieve the deflection for the outer ions, but the ions in the center of the beam will start to scrape the lens-2 electrode if the voltage amplitude is too high, e.g.  $\pm 4.5$  kV as shown in Figure 14.

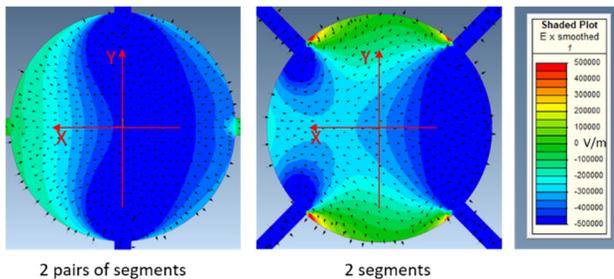


Figure 13: Comparison of Ex field distributions of the two-pair and two-segment chopping.

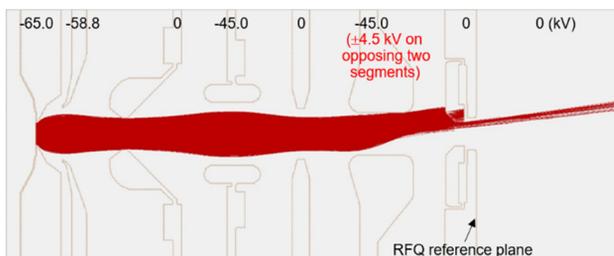


Figure 14: Beam scraping on the lens-2 electrode.

### SUMMARY

A description of the SNS H<sup>-</sup> injector with LEPT beam chopper was given. Beam simulations were conducted for the present beam chopping pattern, which involves activation of all four HV pulsers at the same time. Simulations

indicate voltage amplitude higher than the presently limited 2.5 kV is desired to minimize ions impacting on the RFQ vane tips. An ongoing effort to upgrade the chopper pulser hardware to enable higher voltage capability was discussed,  $\pm 3.5$  kV has been achieved on a test-bench and  $\pm 5.0$  kV is expected with further development. With the prospect of higher voltage capability by new design pulsers, possible alternative beam chopping patterns which involve activation of only two or even just one pulser at a time were explored. The proposed new chopping patterns will significantly reduce the stress on the HV switches due to reduced switching frequencies. But, beam simulations suggest higher voltage amplitude,  $\pm 4.0$  kV or above, is required to deflect the beam out of the RFQ acceptance ellipse with reasonable separation. Clearing the RFQ vane from impacting ions requires even higher voltage amplitude, but beam scraping on lens-2 will become a concern.

### REFERENCES

- [1] R. Keller, R. Thomae, M. Stockli, and R. Welton, "Design, Operational Experiences and Beam Results Obtained with the SNS H<sup>-</sup> Ion Source and LEPT at Berkeley Lab" in *AIP Conf. Proc.*, vol. 639, pp. 47, 2002.
- [2] Martin P. Stockli, Robert F. Welton, and Baoxi Han, "Record productions establish RF-driven sources as the standard for generating high-duty-factor, high-current H<sup>-</sup> beams for accelerators", *Rev. Sci. Instrum.*, vol. 89, pp. 052202, 2018. doi:10.1063/1.5025328
- [3] J. W. Staples, J. J. Ayers, D. W. Cheng, J. B. Greer, M. D. Hoff, and A. Ratti, "The SNS Four-phase LEPT Chopper", in *Proc. PAC'99*, New York, NY, USA, Mar-Apr 1999, pp. 1961-1963.
- [4] SIMION, <https://simion.com/>
- [5] PBGUNS, <https://www.far-tech.com/pbguns/>
- [6] V. Peplov, B. Han, R. Saethre, and M. Stockli, "Alternative configuration and timing control for beam chopping system at the SNS linac", in *Proceedings of the IEEE 21st International Conference on Pulsed Power (PPC)*, Brighton, UK, 2017
- [7] Vladimir V. Peplov, and Robert B. Saethre, "Performance of High Voltage Bipolar Pulse Generator with Fast Push-Pull Solid-State Switch Modules", to be published.

# LONGITUDINAL BEAM DYNAMICS IN FRIB AND REA LINACS

A. S. Plastun<sup>†</sup>, P. N. Ostroumov, A. C. C. Villari, Q. Zhao

Facility for Rare Isotope Beams, Michigan State University, 48824, East Lansing, Michigan, USA

## Abstract

The Front-End and first three cryomodules of the Facility for Rare Isotope Beam (FRIB) at Michigan State University (MSU) commissioned in July, 2018. The paper describes the online tuning procedures of the longitudinal beam dynamics through the FRIB linac. These procedures include tuning of the accelerating field phases and amplitudes in the cavities. We developed an automated simulation-based tuning procedure for the multi-harmonic buncher. In order to tune the radio-frequency quadrupole (RFQ) we measured and calculated its threshold voltage and scanned its longitudinal acceptance. Tuning of the rebunchers and superconducting accelerating cavities is performed by means of the phase scans and Time-Of-Flight (TOF) beam energy measurements with beam position and phase monitors.

While FRIB is being commissioned, the re-accelerator (ReA3) for rare isotope beams (RIBs) is being upgraded. We redesigned the ReA3 RFQ to improve its cooling system and provide reliable operation with 16.1 MHz pre-bunched ion beams with  $A/Q = 5$ . In order to provide matching of any ReA3 beam both to the following upgrade cryomodules and physics experiments' requirements, room temperature rebuncher/debuncher is being designed. The design procedure includes the beam dynamics, electromagnetic, thermal and mechanical simulations and optimizations.

## INTRODUCTION

The Facility for Rare Isotope Beams (FRIB) [1] is being built to provide 400 kW ion beams up to uranium for the rare isotope production. The beam will be accelerated by the RFQ and 316 superconducting RF (SRF) cavities to the energy of 200 MeV/u.

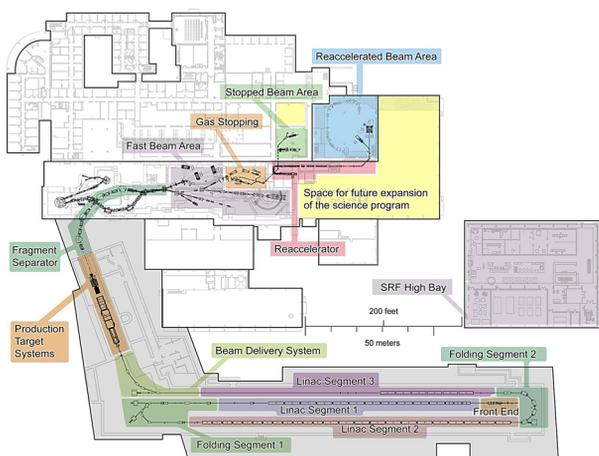


Figure 1: FRIB Layout.

The FRIB linac consists of the Front-end and three linac segments (see Fig. 1). The DC beam, created in the ECR ion sources is accelerated to 500 keV/u and injected into the cryomodules of the linac segment 1 (LS1) as shown in Fig. 2 [2]. The 12 keV/u DC beam is pre-bunched by a multi-harmonic buncher to form a small longitudinal emittance. The RFQ is followed by medium energy beam transport (MEBT) to the cryomodules. The MEBT has two room-temperature quarter-wave (QWR) rebunchers to match the beam to the following SRF cavities in the longitudinal phase space.

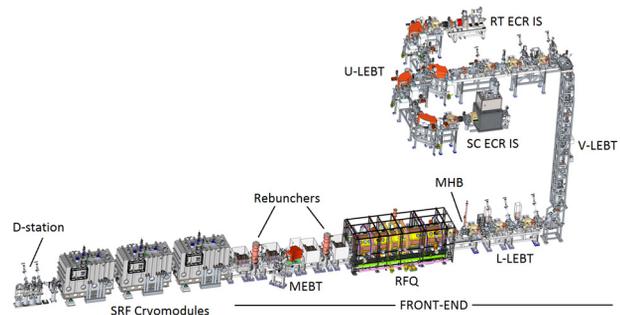


Figure 2: FRIB Front-end and first three cryomodules.

The detailed description of the FRIB linac components is given elsewhere [3-6].

FRIB is designed to accelerate two uranium charge states (33+ and 34+) in LS1 simultaneously, strip them in the liquid lithium stripper and continue acceleration of all five charge states (76+ through 80+) in one RF bucket [4, 5]. Proper tune of field amplitudes and phases of all FRIB cavities, including the bunching ones, is an essential procedure required to provide low losses and low emittance growth of the multi-charge-state beam [6].

## MULTI-HARMONIC BUNCHER

The main purpose of the multi-harmonic buncher (MHB) is to form bunches with a small longitudinal emittance, which is required to maintain low particle losses in the SRF part of FRIB in multi-charge acceleration mode [6]. Also, MHB provides the longitudinal matching of the beam with the RFQ acceptance to maintain small emittance along the RFQ.

The FRIB MHB consists of two resonant quarter-wave lines attached to a pair of conical drift tubes [7]. The long line resonates at frequencies of 40.25 MHz and 120.75 MHz. The short line resonates at 80.5 MHz. Tuning of the MHB is an optimal choice of harmonics amplitudes and phases to provide small longitudinal beam emittance, proper matching with the RFQ and ~80% transmission of the accelerated beam.

<sup>†</sup> email address: plastun@frib.msu.edu

At the time of the Front-end commissioning we didn't have any diagnostic tools to characterize the longitudinal beam emittance. For this reason, we developed a simulation-based tuning approach.

### Simulation-based Tuning of the MHB

The idea of this approach is to find the MHB tune, which could be closely simulated. And then just rescale the amplitudes to provide minimum longitudinal emittance according to the simulation results. Since we could reliably measure the beam transmission through the RFQ, the selected tune was the "maximum transmission tune".

First, we created the electrostatic model of the FRIB RFQ vane tips in CST [8] which is based on the point-to-point tip geometry provided by PARMTEQ [9]. The latter was used to generate the vane tip modulation. Figure 3 shows the 3D model of the vanes. Since the RFQ has voltage increase along the resonator from 60 kV to 112 kV, the 3D-field distribution has been multiplied by the linear  $V(z)$ -law and exported cell-by-cell to use in the beam dynamics simulations. All beam dynamics simulations presented in this paper have been performed with TRACK code [10].

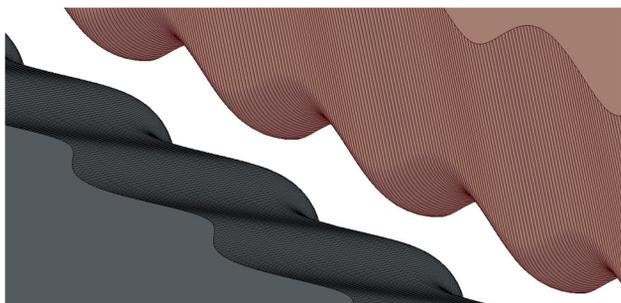


Figure 3: Vane tips of the FRIB RFQ 3D model.

Second, we created the RF and electrostatic models for the MHB in CST. Beam dynamics simulations in the MHB showed that combined 3D electromagnetic field distribution of three harmonics can be replaced by a sequence of three gaps with electrostatic field of charged drift tubes resonating at MHB frequencies, as presented in Fig. 4. The focusing solenoid in front of the RFQ was also added into the TRACK model.

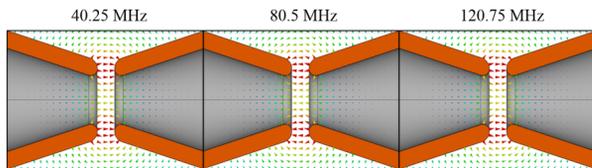


Figure 4: MHB model for beam dynamics simulation.

Finally, we developed the python-based interface to TRACK to enhance optimizing capabilities. Python script using the Nelder-Mead simplex method automatically finds amplitudes of MHB harmonics corresponding to the maximum beam transmission. The MHB settings for "minimum ratio of longitudinal emittance to transmission" and "minimum longitudinal emittance" were found as well. Phase-space plots at the entrance of the RFQ are presented

in Fig. 5. Parameters of the beam exiting the RFQ are summarized in Table 1.

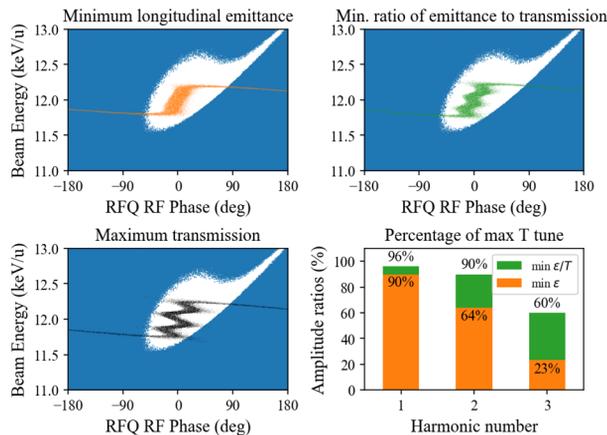


Figure 5: Phase-space plots for MHB tunes and their harmonic amplitudes. White area represents the RFQ longitudinal acceptance.

Table 1: Beam Parameters After the RFQ

MHB Tune	Long. emittance, rms ( $\pi \cdot \text{keV}/u \cdot \text{ns}$ )	RFQ+MEBT Transmission (%)
Max T	0.103	83.8
Min $\epsilon/T$	0.079	81.6
Min $\epsilon$	0.074	76.3

FRIB control system is built on Experimental Physics and Industrial Control System (EPICS) [11]. We widely use python *epics* package to get data and control the accelerator. In order to tune the real MHB for maximum transmission we developed the python script scanning the amplitudes and phases of harmonics. The highest transmission can be achieved in two-three iterations of phase-amplitude sweep. Other MHB tunes are set by scaling the amplitudes as shown in Fig. 5.

During the commissioning of the first three cryomodules we performed the SRF cavity amplitude scan and measure the bunch length at each field level. Scan result were used to reconstruct the longitudinal beam emittance after the acceleration by seven SRF cavities. The measured (reconstructed) and simulated emittances are presented in Table 2.

Table 2: Longitudinal rms Beam Emittance

MHB Tune	Simulated emittance ( $\pi \cdot \text{keV}/u \cdot \text{ns}$ )	Measured emittance ( $\pi \cdot \text{keV}/u \cdot \text{ns}$ )
Max T	0.14	0.19
Min $\epsilon$	0.12	0.14

## RADIO-FREQUENCY QUADRUPOLE

The first faraday cup (FC) downstream the FRIB RFQ is located behind the third quadrupole triplet of the MEBT. The non-accelerated particles are lost upstream of the FC.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

This fact allows us to measure the RFQ threshold voltage [12, 13]. Comparing the calculated and measured values for the threshold voltage we can set the design field level in the RFQ. The relationship of the design  $V_0$  and threshold  $V_{th}$  voltages is:

$$\frac{V_{th}}{V_0} = \cos\varphi_s, \quad (1)$$

here  $\varphi_s$  is the synchronous phase in the regular part of the RFQ, which is equal to  $-25^\circ$  for the FRIB RFQ. It results in  $V_{th} = 63$  kV at  $V_0 = 69.5$  kV for  $^{40}\text{Ar}^{9+}$  used for the measurements of the threshold curve (see Fig. 6).

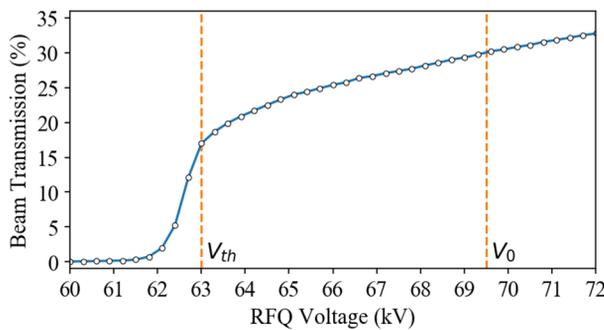


Figure 6: Threshold curve of the FRIB RFQ. Horizontal axis represents the maximum RFQ voltage along the resonator.

This procedure for the accelerating field calibration is widely applied to drift tube linacs (DTLs) with magnetic focusing, such as an Alvarez-DTL. But it cannot be used with an RFQ if one measures the output beam current right at the RFQ exit, because fraction of non-accelerated beam is usually high.

Another way to check the field setting of the RFQ is the energy scan of the longitudinal acceptance. We created a script, which sweeps the DC beam energy, scales the LEPT optics according to equations below and measures the RFQ+MEBT beam transmission:

$$\frac{V}{V_{des}} = \frac{W}{W_{des}}, \quad \frac{I}{I_{des}} = \frac{v}{v_{des}}, \quad (2)$$

here  $V$  are voltages of electrostatic quadrupoles and e-bends and  $I$  are currents of magnets,  $W$  is the beam energy,  $v$  is the beam velocity. Index *des* means design value corresponding to the energy of 12 keV/u. The result of the energy scan is presented in Fig. 7. In the simulation we used 4D-Gaussian distribution with rms emittance of  $0.12 \pi \cdot \text{mm} \cdot \text{mrad}$ . Since the simulated and measured energy profiles agree very well, we believe that the RFQ voltage is set correctly at  $V_0 = 69.5$  kV.

RF Phase of the RFQ relative to the FRIB global timing system is always equal to  $0^\circ$ . All other cavities, including the MHB, have to be properly phased relative to the beam.

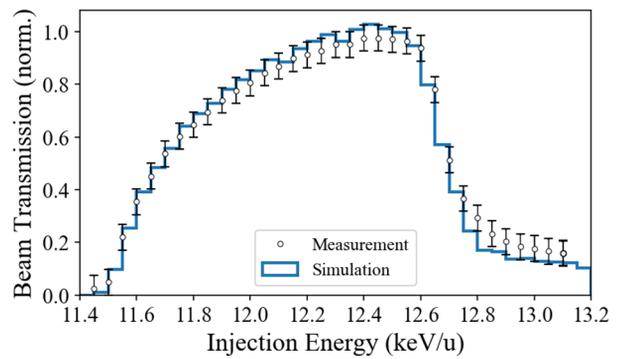


Figure 7: Energy scan of the RFQ longitudinal acceptance.

## CAVITY PHASE SCAN

Phase scan of the cavities includes measurement of the phase and amplitude of the signal from the beam position monitors (BPMs) next to the cavity. Phase of the signal is a function of the time of flight from the cavity to the BPM. The lowest phase corresponds to the highest beam velocity, i.e. to the cavity phase for maximum acceleration. The design accelerating phase is calculated from the phase of maximum acceleration. Figure 8 shows the result of a cavity phase scan. Measured points are interpolated with cubic splines and maximum acceleration phase is calculated ( $124^\circ$  in the Fig. 8). The accelerating phase is equal to  $94^\circ$  in this plot and corresponds to  $-30^\circ$  synchronous phase.

The cavity accelerating fields are adjusted to match the design energy gain. Beam energy is automatically calculated online from the measured phases of BPMs' signals.

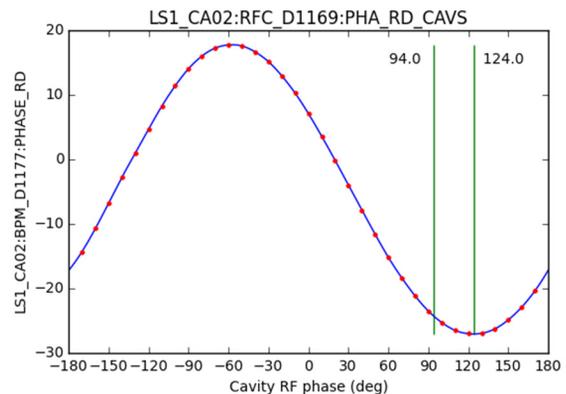


Figure 8: Example of the cavity phase scan result. Vertical axis is a phase of the signal from the BPM next to the cavity.

During the first three cryomodule commissioning phase scans were performed at 20% of the design cavity field level to avoid strong RF defocusing and steering of the misaligned beam. A phase scan procedure takes about 1 minute, beam trajectory correction takes 1 – 10 minutes (longer for the first several cavities) and about 3 minutes are required to ramp up the cavity field to the design amplitude. The LS1 commissioning scheduled for the spring 2019 will include phase scans of 88 SRF cavities. We are developing the python application to scan the cavities one-

by-one automatically including turn-on procedures and field ramp-up. Our current goal is 60 seconds per cavity.

### Rebuncher Phase Scan

If the cavity is a rebuncher the operating phase is then defined as a cross-point of two curves measured at two different field levels as shown in Fig. 9. BPM magnitude signal is used to figure out which of two cross-points is bunching (i.e.  $-90^\circ$  from the maximum acceleration phase) and which is debunching (i.e.  $+90^\circ$  from the maximum acceleration phase). Buncher voltage is calibrated by energy measurement at the phase of maximum acceleration.

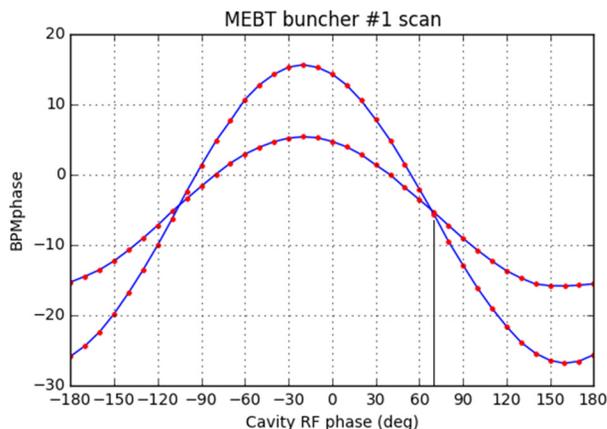


Figure 9: Example of the buncher phase scan result.

## RE-ACCELERATOR UPGRADE

While FRIB is being commissioned, the MSU re-accelerator (ReA) for rare isotope beams (RIBs) is being upgraded. ReA was commissioned as ReA3 in 2015 [14] and currently accelerates rare isotope beams with charge-to-mass ratio from 0.25 to 0.5 at the energy range from 0.3 to 6 MeV/u. The ReA3 upgrade includes: (a) replacement of the ReA3 RFQ electrodes to improve the cooling and to provide higher capture efficiency for 16.1 MHz bunches with charge-to-mass ratio down to 0.2, (b) adding another three cryomodules after the ReA3, (c) installation of the new electron beam ion trap (EBIS), (d) new RF controllers. Upgraded ReA options called ReA6 and ReA12 will significantly extend the scientific program allowing to reach beam energies up to 24 MeV/u (see Fig. 10).

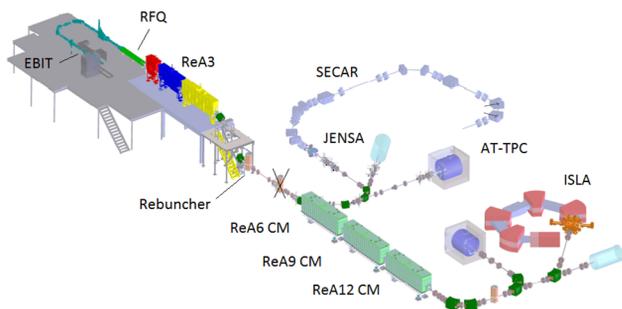


Figure 10: MSU Re-accelerator layout.

In order to match any ReA3 beam to the following upgrade cryomodules and meet the requirements of low energy nuclear physics experiments', a room temperature cw rebuncher/debuncher will be developed and built.

### ReA3 RFQ Upgrade

To provide the reliable cw operation of the ReA3 RFQ the electrodes were re-designed to reduce the inter-vane voltage from 86.5 kV to 70 kV, peak fields from 1.6 to 1.45 Kilpatrick units, RF power consumption from 160 kW to 100 kW. The RFQ resonator and, consequently, the length of the electrodes remains the same. To gain more energy at reduced voltage we implemented the trapezoidal modulation of the electrodes in the acceleration section of the structure. Currently, no RFQ-design codes capable of designing the electrodes with both sinusoidal and trapezoidal modulation are available. To deal with it we developed the practical design approach for RFQs [15]. It is based on a VBA Macro running in CST STUDIO and performing the cell-by-cell construction of the 3D CAD model for the modulated electrodes. The macro automatically adjusts the cell lengths to match the desired synchronous phase law. Each iteration consists of electrostatic simulation of the CAD model, 1D beam dynamics simulation of the reference particle, evaluation of the RF phase in the cell center and adjustment of the cell length if needed. At the end of the design procedure we generated the CAD model which was used for the electrodes' fabrication. The photo of the machined electrodes is shown in Fig. 11. The installation is scheduled for 2019.



Figure 11: Manufactured electrodes for ReA3 RFQ Upgrade (provided by Kress GmbH, Germany). #1 – trapezoidal part, #2 and #4 – output radial matcher, producing round beam, #3 – input radial matcher and sinusoidal part.

### Rebuncher

The main purpose of the rebuncher is to match ReA3 beams with the energy of  $12 \cdot Q/A$  MeV/u to the ReA6 cryomodule, here  $Q$  is an ion charge state and  $A$  is an ion mass number. Another function of the rebuncher is to debunch 0.3 – 6.0 MeV/u beams for ReA3 users. The requirement for the bunch length is  $\pm 1$  ns and for the energy spread is  $\pm 1$  keV/u. We have studied several options: 80.5 MHz or

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

161 MHz cavity, single rebuncher in two different locations or two rebunchers, double-gap QWR and multi-gap IH structure, etc. The main criteria were low construction and operational costs. Figure 12 shows the basic design of the rebuncher cavity. Its parameters are presented in Table 3.

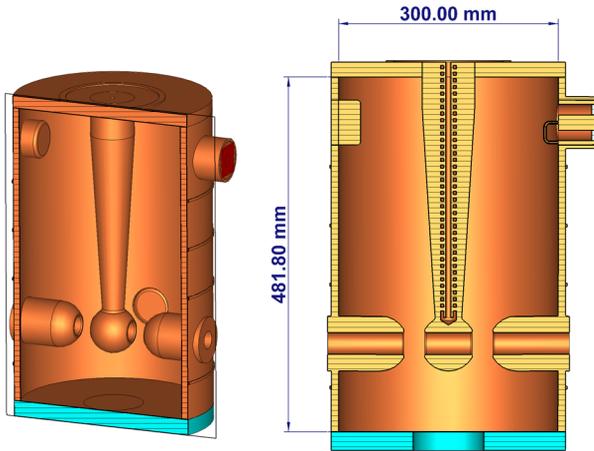


Figure 12: ReA rebuncher design.

Design of cw cavities usually includes coupled RF, thermal and mechanical simulations due to high heat load from RF losses. Multiphysics design of the cavity is fully completed with a combination of CST STUDIO and SolidWorks Flow Simulation Tool. The latter is mostly used to calculate the heat transfer coefficients for water cooling channels.

Table 3: ReA Rebuncher Parameters

Parameter	Value	Unit
Frequency $f_0$	161	MHz
$\beta_G$	0.1	
Voltage $V_0$	200	kV
RF power $P_0$	4	kW
Quality factor $Q_0$	13,000	
Peak surface field $E_{peak}$	0.6	Kilpatrick units

## ACKNOWLEDGMENTS

Authors would like to thank our colleagues for their help with the experiments: T. Maruta, T. Yoshimoto, H. Ren, J. Stetson, S. Renteria, G. Machicoane, E. Pozdeyev, C. Morton, S. Cogan, S. Lidia, S-H. Kim, S. Zhao, and many other physicists and engineers of FRIB. Authors thank B. Mustapha from ANL for the support with TRACK code.

This material is based upon work supported by the National Science Foundation under Grant No. PHY-1565546.

## REFERENCES

[1] J. Wei *et al.*, “The FRIB Superconducting Linac: Status and Plans”, in *Proc. LINAC’16 Conf.*, East Lansing, MI, USA, Sep. 2016, paper MO1A01, pp. 1.

[2] E. Pozdeyev, “FRIB Front End construction and commissioning”, in *Proc. IPAC’18 Conf.*, Vancouver, Canada, May 2018, paper MOZGBF1, pp. 58.

[3] T. Xu *et al.*, “Progress of FRIB SRF Production”, in *Proc. SRF’17 Conf.*, Lanzhou, China, July 2017, paper TUXAA03, pp. 345.

[4] P.N. Ostroumov and K.W. Shepard, *Phys. Rev. ST Accel. Beams*, vol. 3, pp. 030101 (2000).

[5] R. C. York *et al.*, “Beam dynamics in the FRIB linac”, in *Proc. HB’10 Conf.*, Morschach, Switzerland, Sept. 2010, paper TUO1B05, pp. 319.

[6] Q. Zhao *et al.*, “FRIB accelerator beam dynamics design and challenges”, in *Proc. HB’12 Conf.*, Beijing, China, Sept. 2012, paper WEO3B01, pp. 404.

[7] J. Holzbauer *et al.*, “Electromagnetic design of a multi-harmonic buncher for the FRIB driver linac”, in *Proc. PAC’11 Conf.*, New York, NY, USA, Mar. 2011, paper TUP091, pp. 1000.

[8] CST Studio Suite, <http://www.cst.com>

[9] K. R. Crandall *et al.*, “PARMTEQ-RFQ Design Codes”. Los Alamos National Laboratory report LA-UR-96-1836.

[10] The beam dynamics code TRACK. <http://www.phy.anl.gov/atlas/TRACK>

[11] EPICS. <http://www.aps.anl.gov/epics/>

[12] W. Panofsky, “Linear accelerator beam dynamics”, Lawrence Rad. Lab. Rep., UCRL report 1216, 1951.

[13] I.M. Kapchinskiy, *Theory of resonance linear accelerators*, Harwood, 1985.

[14] A. Villari *et al.*, “Commissioning and first accelerated beams in the reaccelerator (ReA3) of the National Superconducting Cyclotron Laboratory, MSU”, in *Proc. IPAC’16 Conf.*, Busan, Korea, Mar. 2016, paper TUPMR024, pp. 1287.

[15] A. S. Plastun and P. N. Ostroumov, *Phys. Rev. Accel. Beams*, vol. 21, pp. 030102, 2018.

# ANALYSIS OF EMITTANCE GROWTH IN A GRIDLESS SPECTRAL POISSON SOLVER FOR FULLY SYMPLECTIC MULTIPARTICLE TRACKING

C. E. Mitchell\*, Ji Qiang, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

## Abstract

Gridless spectral methods for self-consistent symplectic space charge modeling possess several advantages over traditional momentum-conserving particle-in-cell methods, including the absence of numerical grid heating and the presence of an underlying multi-particle Hamiltonian. Nevertheless, evidence of collisional particle noise remains. For a class of such 1D and 2D algorithms, we provide analytical models of the numerical field error, the optimal choice of spectral modes, and the numerical emittance growth per time step. We compare these results with the emittance growth models of Struckmeier, Hoffman, Kesting, and others.

## INTRODUCTION

Distinguishing between physical and numerical emittance growth observed in long-term tracking of beams with space charge is critical to understanding beam performance in high-intensity proton rings. Numerical emittance growth has been modeled as a collisional increase of the beam phase space volume driven by random noise caused by the use of a small number of macroparticles [1–4]. Recently, several authors have developed methods for multiparticle tracking (in plasmas or beams) using variational or explicitly symplectic algorithms designed to preserve the geometric properties of the self-consistent equations of motion [5–7]. In this paper, we address the problem of numerical emittance growth generated by the multi-particle symplectic algorithm described in [7]. Due to its relative simplicity, this algorithm can be used as a test-bed for explicit probabilistic models of numerical errors in the computed field and numerical emittance growth.

## SYMPLECTIC SPECTRAL ALGORITHM

We apply the algorithm described in Section III of [7] to treat the Poisson equation in a general bounded domain  $\Omega \subset \mathbb{R}^d$  ( $d \leq 2$ ) with conducting boundary  $\partial\Omega$ . The symplectic map describing a numerical step in the path length coordinate  $s$  is performed by applying second-order operator splitting to the following multi-particle Hamiltonian:

$$H = \sum_{j=1}^{N_p} H_{\text{ext}}(\vec{r}_j, \vec{p}_j, s) - \frac{n}{N_p} \frac{1}{2} \sum_{j=1}^{N_p} \sum_{k=1}^{N_p} \sum_{l=1}^{N_l} \frac{1}{\lambda_l} e_l(\vec{r}_j) e_l(\vec{r}_k).$$

Here  $H_{\text{ext}}$  is the single-particle Hamiltonian in the external applied fields,  $N_p$  denotes the number of simulation particles,  $N_l$  denotes the number of computed modes, and  $n$  is a space charge intensity parameter. The smooth functions  $e_l$

form an orthonormal basis for the space of square-integrable functions on the domain  $\Omega$ , and satisfy

$$\nabla^2 e_l = \lambda_l e_l, \quad e_l|_{\partial\Omega} = 0, \quad (\lambda_l < 0). \quad (1)$$

It follows from  $H$  that each particle moves in response to the smooth space charge force  $\vec{F} = -\nabla U$ , where

$$U(\vec{r}) = -\frac{n}{N_p} \sum_{l=1}^{N_l} \sum_{j=1}^{N_p} \frac{1}{\lambda_l} e_l(\vec{r}_j) e_l(\vec{r}). \quad (2)$$

The space charge potential  $U$  satisfies the Poisson equation  $\nabla^2 U = -\rho$  and  $U|_{\partial\Omega} = 0$ , where  $\rho$  is a particle-based approximation to the beam density, given in terms of the modes  $e_l$  ( $l = 1, 2, \dots$ ) by:

$$\rho = \sum_{l=1}^{N_l} \rho^l e_l, \quad \rho^l = \frac{n}{N_p} \sum_{j=1}^{N_p} e_l(\vec{r}_j). \quad (3)$$

The set of functions

$$\vec{e}_l = \frac{1}{\sqrt{-\lambda_l}} \nabla e_l \quad (l = 1, 2, \dots) \quad (4)$$

is orthonormal and can be extended to an orthonormal basis for the space of square-integrable vector-valued functions on  $\Omega$ . The relationships between the corresponding modes of  $\rho$ ,  $U$ , and  $\vec{F}$  are then given simply by:

$$U^l = \rho^l / \lambda_l, \quad F^l = -\sqrt{-\lambda_l} U^l. \quad (5)$$

By appropriately grouping the sums appearing in the space charge kick, the complexity of a single numerical step using the Hamiltonian  $H$  scales as  $O(N_p N_l)$  [7].

## PROBABILISTIC MODEL

Assume that particle coordinates  $(\vec{r}_j, \vec{p}_j)$ ,  $j = 1, \dots, N_p$  are sampled from the joint probability density  $P_N$  given by:

$$P_N(\vec{r}_1, \vec{p}_1, \dots, \vec{r}_{N_p}, \vec{p}_{N_p}) = \prod_{j=1}^{N_p} P(\vec{r}_j, \vec{p}_j), \quad (6)$$

where  $P$  is the probability density on the single-particle phase space describing an ideal (smooth) beam distribution. If  $a$  denotes any function on the single-particle phase space, we denote its beam-based average by

$$\langle a \rangle = \frac{1}{N_p} \sum_{j=1}^{N_p} a(\vec{r}_j, \vec{p}_j), \quad \Delta a = a - \langle a \rangle. \quad (7)$$

\* ChadMitchell@lbl.gov

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Likewise, if  $F, G$  denote random variables on the multi-particle phase space, we denote their expected value and covariance as

$$E[F] = \int F dP_N, \quad \text{Cov}[F, G] = E[FG] - E[F]E[G].$$

If  $a_1$  and  $a_2$  denote two functions on the single-particle phase space, then  $\langle a_1 \rangle$  and  $\langle a_2 \rangle$  define random variables on the multi-particle phase space, and it follows from the fact that distinct particles are independent and identically distributed, see Eq. (6), that ( $j = 1, 2$ )

$$E[\langle a_j \rangle] = E[a_j], \quad \text{Cov}[\langle a_1 \rangle, \langle a_2 \rangle] = \frac{1}{N_p} \text{Cov}[a_1, a_2].$$

A generalization of this result is provided in the Appendix.

## FIELD ERROR AND OPTIMAL MODES

### Field Error

The space charge potential of an ideal beam with density  $\rho_{\text{exact}}$  is given by the exact solution of the Poisson equation:

$$\nabla^2 U_{\text{exact}} = -\rho_{\text{exact}}, \quad U_{\text{exact}}|_{\partial\Omega} = 0, \quad (8)$$

where  $\rho_{\text{exact}}$  is the spatial projection of the phase space density  $nP$  appearing in Eq. (6). It follows that

$$\rho_{\text{exact}}^l = \int_{\Omega} \rho_{\text{exact}}(\vec{r}) e_l(\vec{r}) d\vec{r} = n E[e_l]. \quad (9)$$

Let  $\delta\rho = \rho - \rho_{\text{exact}}$  and  $\delta\vec{F} = \vec{F} - \vec{F}_{\text{exact}}$ , where  $\rho$  and  $\vec{F}$  are the numerically computed quantities obtained using Eqs. (2–5). It follows that for all modes with  $l, m \leq N_l$ ,

$$E[\delta\rho^l] = 0, \quad \text{Cov}[\delta\rho^l, \delta\rho^m] = \frac{n^2}{N_p} \text{Cov}[e_l, e_m]. \quad (10)$$

Using the relationships in Eq. (5) gives:

$$E[\delta F^l \delta F^m] = \begin{cases} \frac{1}{N_p} \frac{n^2}{\sqrt{\lambda_l \lambda_m}} \text{Cov}[e_l, e_m] & (l, m \leq N_l) \\ \frac{n^2}{\sqrt{\lambda_l \lambda_m}} E[e_l] E[e_m] & (l, m > N_l) \end{cases}. \quad (11)$$

The mean-squared value of the computed field error at any point  $\vec{r} \in \Omega$  is then given by:

$$E[|\delta\vec{F}(\vec{r})|^2] = \sum_{l,m=1}^{\infty} E[\delta F^l \delta F^m] \vec{e}_l(\vec{r}) \cdot \vec{e}_m(\vec{r}). \quad (12)$$

Define the  $L^2$  norm of the error in the computed field by:

$$\|\delta\vec{F}\|^2 = \int_{\Omega} |\delta\vec{F}(\vec{r})|^2 d\vec{r} = \sum_{l=1}^{\infty} (\delta F^l)^2. \quad (13)$$

Taking the expected value of Eq. (13) using Eq. (11) gives:

$$E[\|\delta\vec{F}\|^2] = -\frac{1}{N_p} \sum_{l=1}^{N_l} \frac{n^2}{\lambda_l} \text{Var}[e_l] - \sum_{l=N_l+1}^{\infty} \frac{n^2}{\lambda_l} E[e_l]^2. \quad (14)$$

The quantity Eq. (14) splits cleanly into contributions due to particle noise (leftmost sum) and mode truncation (rightmost sum).

### Numerical 1D Example

Consider a 1D domain  $\Omega = (0, a)$  containing a beam with an ideal beam distribution  $P$  with parabolic spatial profile:

$$P(x, p) = \frac{3}{4h} \left\{ 1 - \frac{(x-d)^2}{h^2} \right\} \delta(p), \quad |x-d| \leq h. \quad (15)$$

Figure 1 illustrates the predicted rms error in the computed field (black), together with the statistically computed rms error obtained by averaging over 200 distinct random seeds (red dashed). We see good agreement with the analytical model Eqs. (11–12). The error is largest near the beam core, with Gibbs ringing near the beam edges. Figure 2

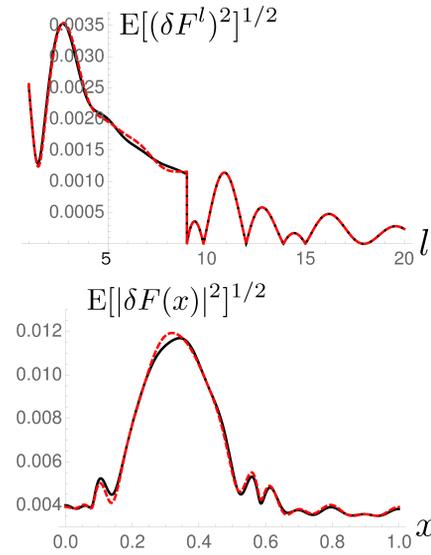


Figure 1: Error in the computed field obtained using  $N_p = 1000$ ,  $N_l = 9$ ,  $n = 1$  corresponding to the density Eq. (15) with  $a = 1$ ,  $d = 1/3$ , and  $h = 1/4$ . (Upper) RMS error in the coefficient of mode  $l$ . (Lower) RMS error in the field at each position  $x$ .

illustrates the expected norm of the field error as a function of the number of particles and the number of modes. For fixed  $N_l = 9$ , the error decreases monotonically with  $N_p$ , approaching a nonzero limit. However, for fixed  $N_p = 1000$ , the error attains a minimum near  $N_l = 9$ . The problem of choosing an optimal mode cutoff is addressed in the next section.

### Optimal Mode Set

For an ideal density  $\rho_{\text{exact}}$ , we can determine the optimal set of modes  $S$  that must be computed to minimize the expected total error Eq. (14). Since  $\lambda_l < 0$ , every term in this sum is nonnegative. Since every mode with index  $l$  must contribute to either the leftmost sum or the rightmost sum in Eq. (14), the quantity Eq. (14) is globally minimized when we enforce the condition that  $l \in S$  if and only if:

$$\frac{E[(\delta F^l)^2]}{(F_{\text{exact}}^l)^2} = \frac{\text{Var}[\delta\rho^l]}{(\rho_{\text{exact}}^l)^2} = \frac{1}{N_p} \frac{\text{Var}[e_l]}{E[e_l]^2} \leq 1. \quad (16)$$

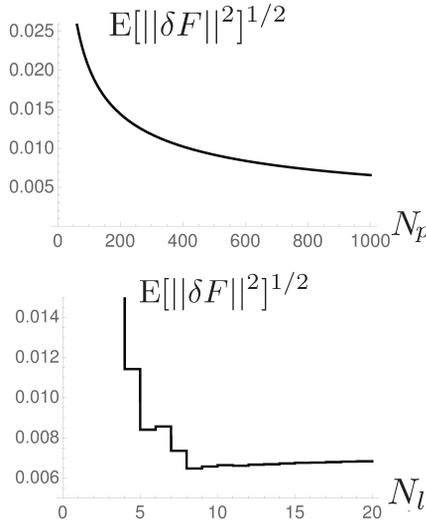


Figure 2: Norm of the error in the computed field corresponding to the beam density Eq. (15) with  $a = 1$ ,  $d = 1/3$ , and  $h = 1/4$ . (Upper) Shown vs. number of particles for fixed mode cutoff  $N_l = 9$ . (Lower) Shown vs. number of modes for fixed  $N_p = 1000$ . In both cases  $n = 1$ .

That is,  $l \in S$  if and only if the rms size of fluctuations in the coefficient  $\rho^l$  resulting from particle noise is less than or equal to the value of  $\rho^l$  for the ideal (smooth) density. Because  $E[e_l]^2$  generally decays more rapidly than  $\text{Var}[e_l]$  with increasing index  $l$ , the set  $S$  is generally finite, and it follows from Eq. (16) that the optimal number of modes increases monotonically with the number of simulation particles  $N_p$ . Note that even in 1D, the optimal set  $S$  does not in general take the form of a consecutive set of indices  $S = \{1, 2, 3, \dots, N_l\}$ .

## ANALYSIS OF EMITTANCE GROWTH

### Emittance Growth on a Single Step

Consider the case of a beam propagating with space charge in a set of linear external focusing fields. The change in  $x$ -emittance under a space charge kick  $(x, p) \mapsto (x, p + \tau F)$  of step size  $\tau$  is given by:

$$\epsilon^2 - \epsilon_0^2 = 2\tau A + \tau^2 B, \quad (17)$$

where the terms  $A$  and  $B$  take the forms:

$$A = \langle \Delta x^2 \rangle \langle \Delta p \Delta F \rangle - \langle \Delta x \Delta p \rangle \langle \Delta x \Delta F \rangle, \quad (18)$$

$$B = \langle \Delta x^2 \rangle \langle \Delta F^2 \rangle - \langle \Delta x \Delta F \rangle^2. \quad (19)$$

Note that Eq. (18) may have variable sign, while Eq. (19) is always nonnegative. Both  $A$  and  $B$  are invariant under the transformation:

$$x \rightarrow x + c, \quad p \rightarrow p + ax + b, \quad F \rightarrow F + gx + h \quad (20)$$

for any constants  $a, b, c, g$ , and  $h$ . This allows us to remove all linear correlations with  $x$ , writing

$$x = E[x] + x_u, \quad (21)$$

$$p = E[p] + \frac{\text{Cov}[x, p]}{\text{Var}[x]}(x - E[x]) + p_u, \quad (22)$$

$$e'_l = E[e'_l] + \frac{\text{Cov}[x, e'_l]}{\text{Var}[x]}(x - E[x]) + e'_{l,u}. \quad (23)$$

Here  $e'_l = \partial e_l / \partial x$ . Replacing  $x, p$ , and  $e'_l$  with  $x_u, p_u$ , and  $e'_{l,u}$  if necessary, we may therefore assume that  $E[x] = 0$ ,  $E[p] = 0$ ,  $E[e'_l] = 0$ ,  $\text{Cov}[x, p] = 0$ , and  $\text{Cov}[x, e'_l] = 0$ . Applying our probabilistic model to the random variables  $A$  and  $B$  using the results of the Appendix gives a decomposition into modes of the form:

$$E[A] = \sum_{l=1}^{N_l} \frac{n}{\lambda_l} A^l, \quad E[B] = \sum_{l,m=1}^{N_l} \frac{n^2}{\lambda_l \lambda_m} B^{lm}, \quad (24)$$

$$\text{Var}[A] = \sum_{l,m=1}^{N_l} \frac{n^2}{\lambda_l \lambda_m} A^{lm}, \quad (25)$$

$$\text{Var}[B] = \sum_{l,m,l',m'=1}^{N_l} \frac{n^4}{\lambda_l \lambda_m \lambda_{l'} \lambda_{m'}} B^{lm'l'm'}. \quad (26)$$

In general, the mode coefficients are complicated and must be evaluated using computer algebra. However, in the smooth beam limit  $N_p \rightarrow \infty$ , we have:

$$\lim_{N_p \rightarrow \infty} A^l = \text{Var}[x] \text{Cov}[p, e'_l] E[e_l], \quad (27)$$

$$\lim_{N_p \rightarrow \infty} B^{lm} = \text{Var}[x] \text{Cov}[e'_l, e'_m] E[e_l] E[e_m]. \quad (28)$$

Note also that  $\text{Var}[A]$  and  $\text{Var}[B]$  are each  $O(1/N_p)$ . If  $P$  is chosen such that  $x$  and  $p$  are independent (aside from possible linear correlation), we also have the following results, accurate through first order in  $1/N_p$ :

$$E[A] = 0, \quad \text{Var}[A] = \frac{1}{N_p} \text{Var}[x] \text{Var}[p] E[B]. \quad (29)$$

Finally, we may evaluate the coefficients appearing in  $E[B]$  through first order in  $1/N_p$  to give:

$$B^{lm} = \lim_{N_p \rightarrow \infty} B^{lm} + \frac{1}{2N_p} (T^{lm} + T^{ml}), \quad (30)$$

where

$$\begin{aligned} T^{lm} = & \text{Var}[x] \text{Cov}[e'_l, e'_m] \text{Cov}[e_l, e_m] \\ & - 3 \text{Var}[x] \text{Cov}[e'_l, e'_m] E[e_l] E[e_m] \\ & + 2 \text{Cov}[x^2, e_l] \text{Cov}[e'_l, e'_m] E[e_m] \\ & + 2 \text{Var}[x] \text{Cov}[e'_l, e'_m, e_l] E[e_m]. \end{aligned} \quad (31)$$

These results can be compared to the model of emittance growth on a single step described in Section IV of [4]. That model is equivalent to treating the kick  $F$  as a random field

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

with  $E[F(x)] = 0$  and spatially-varying  $\text{Var}[F(x)]$ , and neglecting contributions to the emittance due to statistical fluctuations in the beam moments and nonlinear correlations between  $F$  and  $x$ . Applying this model to Eq. (18–19) gives:

$$E[A] = 0, \quad E[B] = \text{Var}[x] E[\text{Var}[F(x)]]. \quad (32)$$

Using the model of numerical field error given in the previous section Eq. (12), we find (in the 1D case) that:

$$\text{Var}[F(x)] = \frac{1}{N_p} \sum_{l,m=1}^{N_l} \frac{n^2}{\lambda_l \lambda_m} \text{Cov}[e_l, e_m] e'_l(x) e'_m(x). \quad (33)$$

It follows from Eq. (33) that

$$E[B] = \frac{1}{N_p} \sum_{l,m=1}^{N_l} \frac{n^2}{\lambda_l \lambda_m} \text{Var}[x] \text{Cov}[e_l, e_m] \text{Cov}[e'_l, e'_m]. \quad (34)$$

Comparing with Eq. (31), we see that this approximation is equivalent to assuming that the emittance growth vanishes in the limit  $N_p \rightarrow \infty$  and neglecting all but the first term of Eq. (31).

### Numerical 1D Example

Consider a 1D domain  $\Omega = (0, a)$ , using an ideal beam distribution  $P$  of the form:

$$P(x, p) = \frac{1}{2\pi\sigma_p\sigma_x} \exp\left(-\frac{p^2}{2\sigma_p^2}\right) \exp\left(-\frac{(x-a/2)^2}{2\sigma_x^2}\right). \quad (35)$$

A statistical test was performed as follows. We randomly generated a beam consisting of  $N_p$  particles  $(x, p)$  by sampling from the density Eq. (35). The space charge force  $F(x)$  was computed at all particle locations using the symplectic spectral algorithm with  $N_l = 15$ ,  $n = 1$ . Terms  $A$  and  $B$  of Eqs. (18–19) were computed, and this procedure was repeated for 1M distinct random seeds. Fig. 3 provides histograms of the results, illustrating the probability density of the random variables  $A$  and  $B$ . In each figure, the quantity on the horizontal axis is shown after subtracting the expected value obtained in the smooth beam limit  $N_p \rightarrow \infty$ , given by using Eqs. (27–28) in Eq. (24). Here  $\lim_{N_p \rightarrow \infty} E[A] = 0$  and  $\lim_{N_p \rightarrow \infty} E[B] = 3.75 \times 10^{-3}$ . The results become more sharply peaked around the predicted smooth limiting value as  $N_p \rightarrow \infty$ , with a standard deviation that scales as  $O(1/\sqrt{N_p})$ , as predicted. Table 1 provides the mean and standard deviation of  $A$  and  $B$ . Comparing the computed mean  $\mu$  and standard deviation  $\sigma$  of  $A$  with the prediction Eq. (29), we see that  $\mu$  deviates from  $E[A]$  by  $<10^{-6}$ , and  $\sigma$  is in agreement with  $\text{Var}[A]^{1/2}$ .

### Emittance Growth in a FODO Channel

We modeled a 1 GeV proton beam with 100 A current in a FODO lattice of period 1 m, using a 2D rectangular domain of size  $6.5 \times 6.5$  mm. The lattice zero current phase advance per period is 87 degrees, and the depressed phase advance is 74 degrees. Figure 4 shows the emittance evolution of

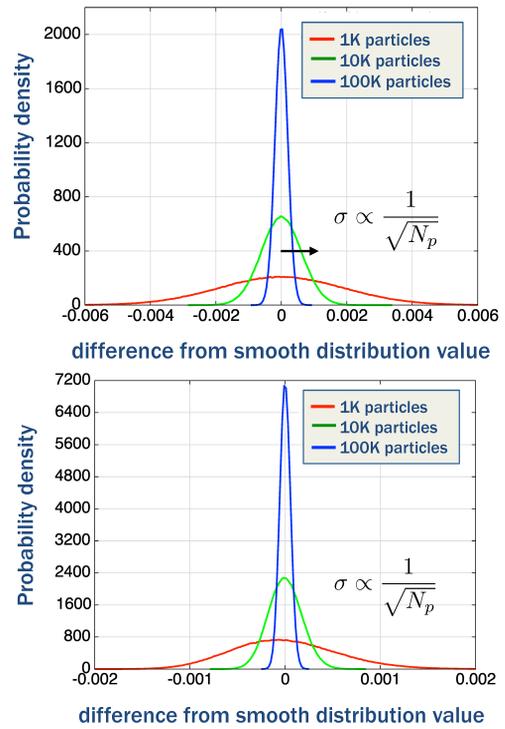


Figure 3: Probability density for the emittance contributions  $A$  and  $B$  of Eqs. (18–19) for the case  $N_l = 15$ ,  $n = 1$ , obtained by sampling beams with varying  $N_p$  using 1M random seeds. (Upper) Term  $A$ . (Lower) Term  $B$ . The values of  $A$  and  $B$  are statistically uncorrelated.

Table 1: Emittance Contributions on a Single Step

$N_p$	term A		term B	
	$\mu$	$\sigma$	$\mu$	$\sigma$
$10^3$	$-9.9 \times 10^{-7}$	$1.9 \times 10^{-3}$	$2.2 \times 10^{-6}$	$5.5 \times 10^{-4}$
$10^4$	$-3.3 \times 10^{-7}$	$6.1 \times 10^{-4}$	$4.1 \times 10^{-7}$	$1.8 \times 10^{-4}$
$10^5$	$+2.9 \times 10^{-7}$	$1.9 \times 10^{-4}$	$1.0 \times 10^{-7}$	$5.6 \times 10^{-5}$

an initially matched KV beam with  $\epsilon_{x,n} = \epsilon_{y,n} = 1 \mu\text{m}$  over 100,000 periods using 15 horizontal and 15 vertical modes (so  $N_l = 15 \times 15$ ). Despite the small number of modes and the small number of particles, the emittance is preserved within 0.4%. The emittance evolution is dominated by period-period fluctuations, and the rms amplitude of these fluctuations scales as  $N_p^{-\beta}$ , with a best fit exponent of  $\beta = 0.57$ , approximately consistent with  $O(1/\sqrt{N_p})$ .

By contrast, Fig. 5 shows the emittance evolution of an initially matched Gaussian beam with  $\epsilon_{x,n} = \epsilon_{y,n} = 1 \mu\text{m}$  using  $32 \times 32$  modes, illustrating linear emittance growth. A least-squares fit to determine the emittance growth rate was performed for each value of  $N_p$ , and the resulting emittance growth rate data scales as  $N_p^{-\alpha}$  with  $\alpha = 0.996$ , consistent with  $O(1/N_p)$ . The rms amplitude of residual fluctuations after removing the linear fit indicates that these fluctuations

scale as  $N_p^{-\beta}$ , with  $\beta = 0.58$ , similar to the KV case and nearly consistent with  $O(1/\sqrt{N_p})$ .

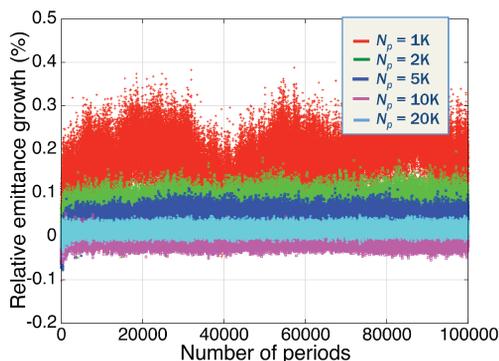


Figure 4: Evolution of the 4D emittance  $\sqrt{\epsilon_x \epsilon_y}$  for a matched KV beam propagating in a FODO lattice using  $15 \times 15$  spectral modes for several values of  $N_p$ .

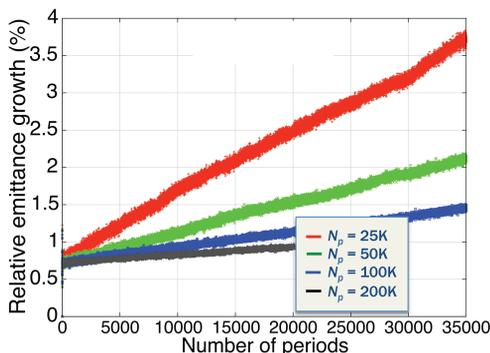


Figure 5: Evolution of the 4D emittance  $\sqrt{\epsilon_x \epsilon_y}$  for an initially Gaussian beam propagating in the same FODO lattice using  $32 \times 32$  spectral modes for several values of  $N_p$ .

Letting  $\epsilon_n$  denote the emittance after the  $n$ th numerical step, we have the expected emittance change at step  $n$ :

$$E[\epsilon_n^2 - \epsilon_{n-1}^2] = 2\tau E[A] + \tau^2 E[B], \quad (36)$$

where  $A$  and  $B$  denote the contributions Eqs. (18–19) at step  $n$ . Dividing by  $\tau$ , assuming that the ideal (smooth) beam distribution evolves slowly from step to step (a questionable assumption), and assuming a relative emittance growth  $\ll 1$ , we obtain an approximate expression for the emittance growth rate given by (compare the result in [4]):

$$E \left[ \frac{d\epsilon}{d\tau} \right] \approx \frac{1}{2\epsilon_0} \tau E[B]. \quad (37)$$

Similarly, noting that:

$$\text{Var}[\epsilon_n^2 - \epsilon_{n-1}^2] = 4\tau^2 \text{Var}[A] + 4\tau^3 \text{Cov}[A, B] + \tau^4 \text{Var}[B],$$

keeping to leading order in the stepsize  $\tau$ , and making the same assumptions as above gives an approximate expression for the rms fluctuations of the emittance from step to step:

$$\sigma_{\Delta\epsilon} \approx \frac{\tau}{\epsilon_0} \text{Var}[A]^{1/2}. \quad (38)$$

Using Eqs. (29) and (34), we see that the two quantities Eqs. (37) and (38) are expected to scale as  $O(1/N_p)$  and  $O(1/\sqrt{N_p})$ , respectively, as observed.

The approximate scaling analysis above effectively neglects horizontal-vertical coupling. A probabilistic treatment of the dynamical emittance growth driven by particle noise in the presence of space charge was previously proposed using a moment analysis of the Vlasov-Fokker-Planck equation [1], [8], where the role of collisional heat exchange between the degrees of freedom is emphasized. A connection could be made by relating the friction and diffusion coefficients in this model to the explicit model of field and single-step emittance growth described here.

## CONCLUSION

We developed probabilistic models of the computed field error and numerical emittance growth on a single step for a fully symplectic spectral space charge tracking algorithm [7]. The models described here are independent of the detailed geometry of the domain, which appears indirectly through the set of modes  $e_l$  and eigenvalues  $\lambda_l$ . The model of the computed field error, including both particle noise and errors due to the use of a finite mode cutoff, is relatively simple, and can be used to select an optimal set of computed modes that minimizes the expected norm of the computed field error for a given particle number.

A complete probabilistic model of dynamical emittance growth is difficult, but insight can be drawn from a model of emittance growth on a single numerical step. This emittance growth is driven by the two terms Eqs. (18) and (19). Term  $A$  in general has negligible expected value, but  $\text{Var}[A] \sim O(1/N_p)$ , and this variance drives fluctuations in the emittance from step to step, which scale approximately as  $O(1/\sqrt{N_p})$ . Term  $B$  is always nonnegative. Its expected value contains a term that is independent of  $N_p$ , which drives emittance growth in the smooth-beam limit, and a term due to particle noise, that drives additional emittance growth scaling as  $\sim O(1/N_p)$ . Statistical moments of these terms can be decomposed into contributions from various modes, and evaluated for a given beam distribution function. We observe a scaling of emittance growth rates and emittance fluctuations consistent with this model for a beam propagating in a FODO channel.

Additional work is underway to investigate the validity of this probabilistic model by evaluating the role played by statistical correlations between successive numerical steps. Finally, while it appears that symplecticity alone is insufficient to eliminate the diffusive effects of numerical noise, it is suspected that these effects can be further suppressed by using higher-order particle shapes (such as those described in [6]), which serve to additionally filter high-frequency components of the computed space charge fields. A discussion of this approach and additional strategies for noise filtering in space charge tracking is provided in [9].

## ACKNOWLEDGEMENTS

This work was supported by the Director, Office of Science, Office of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and made use of computer resources at the National Energy Research Scientific Computing Center.

## APPENDIX

Let  $a_j$  ( $j = 1, \dots, N$ ) and  $b_k$  ( $k = 1, \dots, M$ ) denote functions defined on the single-particle phase space. Under the probability model Eq. (6), we may evaluate the expected value and variance of any polynomial in the beam averages  $\langle a_j \rangle$ ,  $\langle b_k \rangle$ . This includes, for example, Eqs. (18–19). It follows from the fact that distinct particles are independent and identically distributed that:

$$\begin{aligned} \mathbb{E} \left[ \prod_{j=1}^N \langle a_j \rangle \right] &= \prod_{j=1}^N \mathbb{E}[a_j] + \\ &\frac{1}{N_p} \sum_{\substack{j,k=1 \\ j < k}}^N \text{Cov}[a_j, a_k] \prod_{\substack{n \neq j \\ n \neq k}}^N \mathbb{E}[a_n] + O \left( \frac{1}{N_p^2} \right), \end{aligned} \quad (39)$$

and also:

$$\begin{aligned} \text{Cov} \left[ \prod_{j=1}^N \langle a_j \rangle, \prod_{k=1}^M \langle b_k \rangle \right] &= \\ \frac{1}{N_p} \sum_{j=1}^N \sum_{k=1}^M \prod_{r \neq j}^N \mathbb{E}[a_r] \prod_{s \neq k}^M \mathbb{E}[b_s] \text{Cov}[a_j, b_k] &+ O \left( \frac{1}{N_p^2} \right). \end{aligned} \quad (40)$$

Throughout this paper, we evaluate all quantities to first order in  $1/N_p$ . We may use the linearity of  $\mathbb{E}[\cdot]$  and the bilinearity of  $\text{Cov}[\cdot, \cdot]$  to extend this result to polynomials in  $\langle a_j \rangle$ ,  $\langle b_k \rangle$ .

## REFERENCES

- [1] J. Struckmeier, *Phys. Rev. ST Accel. Beams*, vol. 3, pp. 034202, 2000.
- [2] O. Boine-Frankenheim, I. Hofmann, J. Struckmeier, and S. Appel, *Nucl. Instrum. Methods Phys. Res. Sect. A*, vol. 770, pp. 164, 2015.
- [3] I. Hofmann and O. Boine-Frankenheim, *Phys. Rev. ST Accel. Beams*, vol. 17, pp. 124201, 2014.
- [4] F. Kesting and G. Franchetti, *Phys. Rev. ST Accel. Beams*, vol. 18, pp. 114201, 2015.
- [5] B. A. Shadwick, A. B. Stamm, and E. G. Evstatiev, *Physics of Plasmas*, vol. 21, pp. 055708, 2014.
- [6] S. D. Webb, *Plasma Phys. Control. Fusion*, vol. 58, pp. 034007, 2016.
- [7] Ji Qiang, *Phys. Rev. Accel. Beams*, vol. 20, pp. 014203, 2017.
- [8] J. Struckmeier, *Phys. Rev. E*, vol. 5, no. 4, pp. 830, 1996.
- [9] Ji Qiang, "Long-term simulation of space-charge effects," submitted to *Nucl. Instrum. and Methods in Phys. Res. A*, 2018.

# THEORETICAL AND COMPUTATIONAL MODELING OF A PLASMA WAKEFIELD BBU INSTABILITY\*

S. D. Webb<sup>†</sup>, N. M. Cook, D. L. Bruhwiler  
 RadiaSoft, LLC, Boulder, CO 80301

A. Burov, V. Lebedev, S. Nagaitsev  
 Fermi National Accelerator Laboratory, Batavia, IL 60510

## Abstract

Plasma wakefield accelerators achieve accelerating gradients on the order of the wave-breaking limit,  $mc^2 k_p / e$ , so that higher accelerating gradients correspond to shorter plasma wavelengths. Small-scale accelerating structures, such as plasma and dielectric wakefields, are susceptible to the beam break-up instability (BBU), which can be understood from the Panofsky-Wenzel theorem: if the fundamental accelerating mode scales as  $b^{-1}$  for a structure radius  $b$ , then the dipole mode must scale as  $b^{-4}$ , meaning that high accelerating gradients necessarily come with strong dipole wake fields. Because of this relationship, any plasma-accelerator-based future collider will require detailed study of the trade-offs between extracting the maximum energy from the driver and mitigating the beam break-up instability. Recent theoretical work predicts the tradeoff between the witness bunch stability and the amount of energy that can be extracted from the drive bunch, a so-called efficiency-instability relation. We will discuss the beam break-up instability and the efficiency-instability relation and the theoretical assumptions made in reaching this conclusion. We will also present preliminary particle-in-cell simulations of a beam-driven plasma wakefield accelerator used to test the domain of validity for the assumptions made in this model.

## INTRODUCTION

A future lepton collider will have to operate at center of mass energies near 10 TeV. Conventional warm copper or superconducting rf structures are limited to around  $50 \text{ MeV m}^{-1}$  accelerating gradients, meaning that a conventional rf linac would require hundreds of kilometers of accelerating structures. Smaller scale millimeter or THz structures can achieve gradients closer to  $300 \text{ MeV m}^{-1}$  bring the length of the linac to the scale of ten kilometers. The gradients in both of these structures are primarily limited by breakdown phenomena. Much higher accelerating gradients are possible in accelerating structures that are already broken down, that is to say plasmas. The accelerating gradients available to plasma-based accelerators – beam-driven plasma wakefield accelerators (PWFAs) or laser-driven laser plasma accelerators (LPAs) – have accelerating gradients limited by the wave-breaking limit of the plasma

$E_{WB} = mc^2 k_p / e \approx 96 \cdot \sqrt{n_{pe} [\text{cm}^{-3}]} [\text{V/m}]$  which, for labo-

ratory plasma densities around  $1 \times 10^{16} \text{ cm}^{-3}$  corresponds to accelerating gradients near  $10 \text{ GV m}^{-1}$ .

This accelerating gradient comes at the cost of small-scale structures, since decreasing the plasma wavelength  $\lambda_p = 2\pi/k_p$  increases the wave-breaking limit. At the same time, this shrinks the size of the accelerating plasma wave, which makes the witness bunch more prone to various transverse instabilities, such as the hosing instability [1], which is similar to the transverse beam break-up (BBU) instability [2] in traditional linear accelerators. In the case of conventional accelerator, BBU is dictated by the transverse size of the beam pipe, while the accelerating gradient depends on the longitudinal length scales of the accelerating cavity, two independent parameters. In plasma accelerators, the “beam pipe” and “accelerating cavity” are both the plasma wave, and the scales are not independent.

Because these scales are related, there exists an *efficiency-instability relationship* [3] predicted to limit how much energy can be extracted from the driver of a plasma accelerator before the BBU instability makes the witness bunch unusable for collider applications. Our ongoing work is to study the domain of validity of this theory, and ways to avoid this limit, to design a high-efficiency plasma accelerator based TeV lepton collider.

## BEAM BREAK-UP AND HOSING INSTABILITIES

The dipole beam break-up instability occurs when a beam is off-center from the beam pipe. This excites a transverse dipole field, where betatron oscillations of the head of the bunch can drive the tail of the bunch to larger betatron amplitudes. A description of this instability in conventional linear accelerators is discussed in detail in Chapter 3 of Chao [4], and is illustrated in Fig. 1.

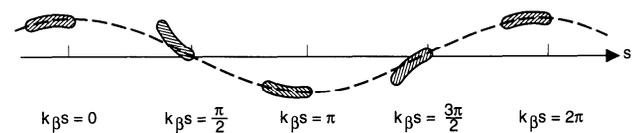


Figure 1: Illustration of the beam break-up instability in Chao [4].

In the beam break-up instability, the head of the bunch begins with some transverse offset. This excites a dipole wake field, which oscillates at the betatron wavelength. The oscillating dipole wake acts as a resonant driving term for the

\* Work supported by Department of Energy Office of Science, Office of High Energy Physics under award no. DE-SC0018718

<sup>†</sup> swebb@radiasoft.net

tail of the bunch. A two-macroparticle model that includes acceleration has the set of ODEs for a head and tail particle:

$$\frac{d}{ds} \left[ \gamma(s) \frac{dy_1}{ds} \right] + k_\beta^2 \gamma(s) y_1 = 0, \quad (1a)$$

$$\frac{d}{ds} \left[ \gamma(s) \frac{dy_2}{ds} \right] + k_\beta^2 \gamma(s) y_2 = -\frac{Nr_e W}{2\gamma(s)L} y_1, \quad (1b)$$

where here  $W$  is the dipole wake function, which scales with the structure radius as  $b^{-4}$ . The acceleration mitigates the instability to some extent, but asymptotically for slow acceleration the growth factor

$$\Upsilon = -\frac{Nr_e W_1(z) L_0}{4k_\beta \gamma_i L} \ln \frac{\gamma_f}{\gamma_i} \quad (2)$$

leads to unacceptably large growth in the tail emittance for a collider without mitigating the instability. The hosing instability in plasma accelerators described by Whittum *et al.* [1] has a similar structure to the beam break-up instability for conventional linear accelerators.

One established approach to damping this instability in conventional linacs is to use BNS damping [5]. In BNS damping, a correlated energy spread across the bunch, combined with the chromaticity of the lattice, tunes the betatron frequency of the tail away from the resonance. The instability is completely stabilized if

$$\frac{\Delta k_\beta}{k_\beta} = \frac{\Upsilon}{k_\beta L}. \quad (3)$$

Driving this frequency spread with linear chromaticity,  $\Delta k_\beta = \xi \Delta p/p$ , for a plasma accelerator would require correlated energy spread on the order of 10%, far too large for a final focus in a linear collider.

Another possibility for mitigating the instability is to use the octupolar focusing that comes from ion motion in the plasma channel [6]. The space charge fields of the drive and witness bunch can cause the background ions to move, which can in some cases lead to “ion collapse” and the complete spoiling of the witness bunch emittance [7]. However, if this is carefully balanced so that the ion motion is only slightly perturbed, the perturbed ion density can introduce sufficient nonlinear focusing to introduce a tune shift with amplitude that detunes the resonance without the need for a large correlated energy spread. The nonlinear transverse focusing will increase the emittance in the tail of the witness bunch, but the emittance growth could be acceptable for collider applications if the nonlinear ion focusing is sufficient to prevent the instability.

## EFFICIENCY-INSTABILITY RELATIONSHIP

Without these mitigation schemes, the efficiency of a single stage of a plasma accelerator is related to the same parameters of the nonlinear wake as the instability growth rate. Under a handful of assumptions, this leads to an efficiency-instability relationship [3].

We can define the efficiency of an accelerator stage as the ratio of the beam power transferred to the witness bunch to the power in the drive bunch,  $\eta_t = P_t/P_b$ . The strength of the instability can be described as the ratio of the transverse focusing force to the strength of the forcing term,  $\eta_p = -F_t/F_r$ . Because the focusing and the acceleration are both directly related to the plasma wake parameters, the efficiency and instability are related approximately as

$$\eta_t \approx \frac{\eta_p}{4(1 - \eta_p)}, \quad (4)$$

so, from this relationship, achieving high efficiency makes the witness bunch particularly susceptible to the beam break-up instability. If this conclusion holds, and the instability cannot be mitigated, it would put a hard limit on the efficiency available for a plasma accelerator for collider applications.

There are a handful of assumptions that go into this relationship:

1. The sheath of the plasma wake is a good conductor;
2. The transverse witness bunch wake in the variable-radius plasma bubble will satisfy a particular relationship, described in [3]; and
3. The plasma wake is not strongly deformed by the witness bunch.

It is important to determine when these assumptions are correct, to determine what range of parameters the efficiency-instability relationship applies to, how the wake field picture is modified when one or more of these assumptions is not valid, and if we can develop a modified picture for transverse beam break-up instabilities when the assumptions are valid.

Because the blowout regime in plasma accelerators is so complicated, it is unlikely we will be able to probe these limits using an analytical or semi-analytical theory. We will therefore have to resort to computing wake functions from self-consistent electromagnetic particle-in-cell simulations.

## COMPUTING WAKE FUNCTIONS IN PLASMA ACCELERATORS

It is important to note that much of the preceding analysis—the beam break-up/hosing instability, the need for BNS damping, and the efficiency-instability relationship these imply—rely on the assumption that the wake fields generated by the witness bunch are well-described by a linear response function. However, the domain of validity for this approximation in a plasma accelerator in the blowout regime has not been established. To determine whether the linear response is valid, we are using FBPIC to run particle-in-cell simulations of first the hollow channel.

FBPIC [8] uses the quasi-cylindrical spectral Lifschitz algorithm for electromagnetic computations. The algorithm decomposes the fields into  $e^{im\theta}$  azimuthal modes, which makes it ideal for isolating individual wake field components,

since for example simulating only the  $m = 0$  and  $m = 1$  modes will preclude quadrupolar and higher-order wake functions. Because FBPIC uses a Fourier-Bessel PSATD algorithm, it is less susceptible to various numerical instabilities, making it easier to interpret the data without having to filter high-frequency numerical Čerenkov and other artifacts.

The theoretical treatment of the beam break-up instability and the conclusions that are applied to plasma accelerators rely on the existence of a wake function, that is that the transverse and longitudinal forces can be described as a linear response to the witness bunch current:

$$\mathbf{F} \propto \int d\zeta' \mathbf{W}(\zeta - \zeta') I(\zeta'). \quad (5)$$

We will call this relationship, as applied to plasma accelerators, the “wake ansatz” or, in the frequency domain, the “impedance ansatz”. This is a linear response approximation; in a conventional accelerator with a stainless steel or other conducting beam pipe, it is the assumption that the material obeys Ohm’s law.

To determine if and when the “impedance ansatz” is valid, we are running self-consistent particle-in-cell simulations of plasma-based accelerator structures with the FBPIC code. From these simulations, we will attempt to extract wake functions for a given configuration of drive bunch charge, and vary the parameters of the drive bunch to see if those wake functions are predictive for a range of parameters and not just the specific parameters of one scenario.

Our benchmark problem is the hollow plasma channel, which has analytic solutions for its wake functions (see Schroeder et al. [9]). This provides a benchmark for short lengths behind the drive bunch, after which surface plasma waves at the hollow channel edge distort the wake fields [10]. However, in the limit that the surface plasma waves are linear, a wake function should still exist.

To test the existence of a meaningful wake function, we will compute a wake function from the response to one drive bunch in an FBPIC simulation, and compare the fields predicted by that wake function for a different set of parameters for the drive bunch. To compute the wake function, we follow a four-step process: (1) compute  $\mathbf{F}$  along the axis from the simulations; (2) take the Fourier transform of  $\mathbf{F}$  and then (3) divide out the Fourier transform of the beam current to give the impedance; (4) take the inverse Fourier transform of the impedance to compute the wake function. In this way, we can determine the domain of validity for the wake function model. We illustrate this process with some preliminary results for the hollow channel below.

For simplicity in demonstrating the concept, we are focusing our efforts on the accelerating fields. Once we have benchmarked this approach, we will then turn our attention to studying the dipole wake fields, which are of critical interest for the beam break-up instability.

First, we compute the electric field for an axisymmetric drive bunch using FBPIC. In Fig. 2 we can see the wake amplitude collapsing due to the surface plasma waves in the hollow channel edge. If this is the result of a purely linear

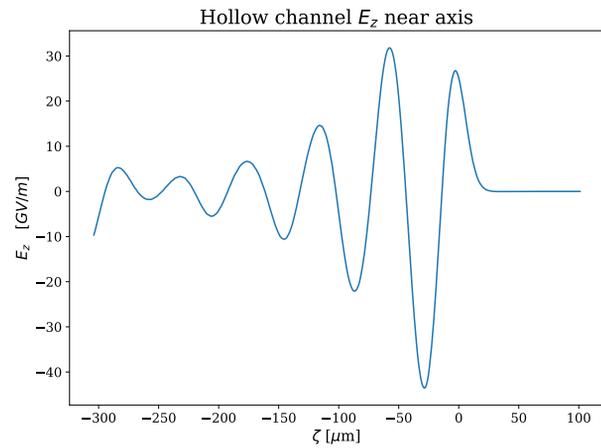


Figure 2: The trailing electric field in the plasma hollow channel, measured from a reference location.

response, we should be able to capture that effect using an impedance model.

To deconvolve the bunch current from the wake function, we take a Fourier transform of the fields, Fig. 3, and then

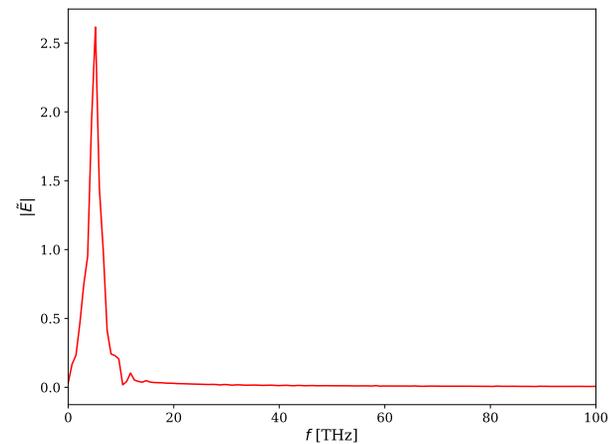


Figure 3: The Fourier transform of the electric field generated by the drive bunch.

divide by the Fourier transform of the bunch current density to obtain the impedance. The numerically computed impedance has several clear spikes with some line width, and some residual lower-amplitude features that are likely the result of noise in the particle-in-cell simulation.

It remains to determine the best way to extract the wake function from this data. The computed impedance in Fig. 4 has a considerable amount of noise that must be filtered, as well as a handful of smaller peaks that may or may not be physical. The inverse Fourier transform of this will provide the wake function as a Fourier series.

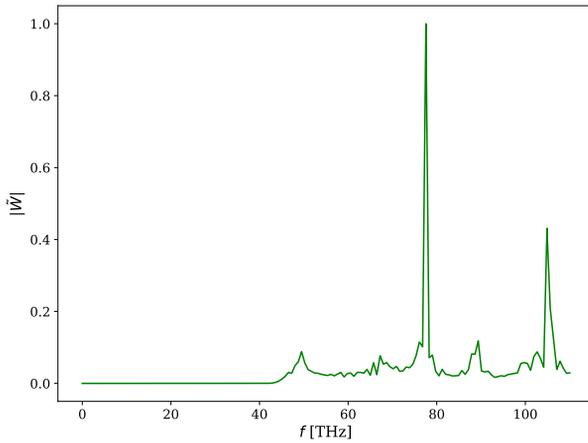


Figure 4: The numerically computed impedance.

## FUTURE WORK

We are currently developing and benchmarking an approach to extracting wake functions from particle-in-cell simulations of plasma accelerating structures. In the future, we will apply these techniques to a blowout plasma wakefield accelerator with a loaded wake, such as depicted in Fig. 5, and determine where this approach is valid for describing the difference between a loaded and unloaded wake, such as in Fig. 6.

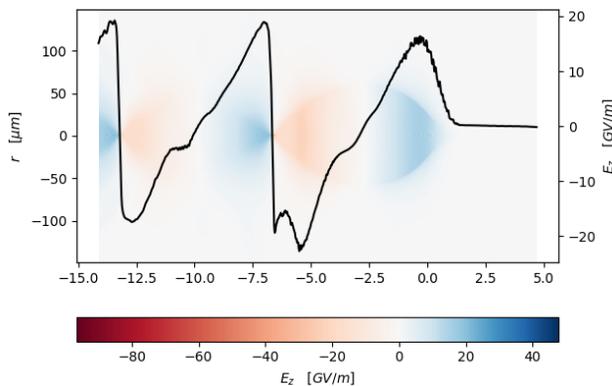


Figure 5: Blowout PWFA with a loaded wake.

Once we have computed the longitudinal and transverse dipole wake functions for a blowout PWFA with a short, low-charge witness bunch, we will be able to compare predictions for the fields as a function of witness bunch charge to determine when the impedance ansatz breaks down. This will provide critical insights into the domain of validity of the efficiency-instability relationship, the exact nature of the beam break-up instability in loaded blowout plasma wakefield accelerators, and the details required for a mitigation technique for the instability.

## ACKNOWLEDGEMENTS

We would like to thank the FBPIC development team, and in particular Remi Lehe (LBNL), for their help and support.

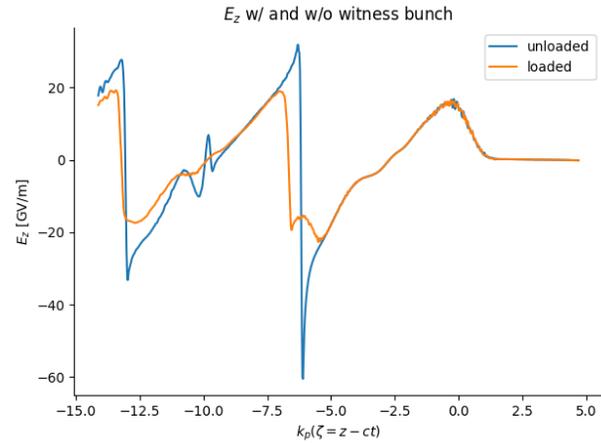


Figure 6: A comparison of a loaded and unloaded blowout PWFA wake.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of High Energy Physics under award no. DE-SC0018718.

## REFERENCES

- [1] D. H. Whittum, W. M. Sharp, S. S. Yu, M. Lampe, and G. Joyce, "Electron-hose instability in the ion focused regime", *Phys. Rev. Lett.* **67**, 991 (1991).
- [2] R. B. Neal and W. K. H. Panofsky, *Science* **152**, 1353 (1966); W. K. H. Panofsky and M. Bander, *Rev. Sci. Instr.* **39**, 206 (1968).
- [3] V. Lebedev, A. Burov, and S. Nagaitsev, "Efficiency versus instability in plasma accelerators", *Phys. Rev. Acc. Beams* **20**, 121301 (2017).
- [4] A. W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators*, J. Wiley & Sons, New York (1993).
- [5] V. Balakin, A. Novokhatsky, and V. Smirnov, *Proc. 12th Int'l. Conf. High Energy Accel.* Fermilab (1983).
- [6] A. Burov, S. Nagaitsev, and V. Lebedev, "Beam Breakup Mitigation by Ion Mobility in Plasma Acceleration", arXiv:1808.03860 [physics.acc-ph].
- [7] J. B. Rosenzweig, A. M. Cook, A. Scott, M. C. Thompson, and R. B. Yoder, "Effects of Ion Motion in Intense Beam-Driven Plasma Wakefield Accelerators", *Phys. Rev. Lett.* **95**, 195002 (2005).
- [8] R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, "A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm", *Comp. Phys. Comm.* **203** pp. 66-82 (2015).
- [9] C. B. Schroeder, D. H. Wittum, and J. S. Wurtele, "Multimode Analysis of the Hollow Plasma Channel Wakefield Accelerator", *Phys. Rev. Lett.* **82**, 1177 (1999).
- [10] C. A. Lindström *et al.*, "Measurement of Transverse Wakefields Induced by a Misaligned Positron Bunch in a Hollow Channel Plasma Accelerator", *Phys. Rev. Lett.* **120**, 124802 (2018).

# REVIEW OF SPECTRAL MAXWELL SOLVERS FOR ELECTROMAGNETIC PARTICLE-IN-CELL: ALGORITHMS AND ADVANTAGES

R. Lehe\*, J.-L. Vay, Lawrence Berkeley National Laboratory, CA 94720 Berkeley, U.S.A

## Abstract

Electromagnetic Particle-In-Cell codes have been used to simulate both radio-frequency accelerators and plasma-based accelerators. In this context, the Particle-In-Cell algorithm often uses the finite-difference method in order to solve Maxwell's equations. However, while this method is simple to implement and scales well to multiple processors, it is liable to a number of numerical artifacts that can be particularly serious for simulations of accelerators.

An alternative to the finite-difference method is the use of spectral solvers, which are typically less prone to numerical artifacts. The present article reviews recent progress in the use of spectral solvers for simulations of plasma-based accelerators. This includes techniques to scale those solvers to large number of processors, extensions to cylindrical geometry, and adaptations to specific problems such as boosted-frame simulations.

## INTRODUCTION

Particle-In-Cell (PIC) codes [1, 2] are widely used in various fields of physics, and in particular in accelerator physics. For many accelerator-related problems, electrostatic PIC codes are usually sufficient to capture the physics at stake. However, some applications do require full electromagnetic PIC codes. This includes for instance accelerators based on laser-plasma interactions [3–7], where e.g. the self-consistent evolution of the laser driver needs to be captured by the PIC algorithm.

For these applications that require an electromagnetic Particle-In-Cell code, the Finite-Difference-Time-Domain (FDTD) method (e.g. [8]) has been the most commonly-used approach for solving Maxwell's equations. However, due to some of the limitations of the FDTD method, other methods are increasingly being used—and this includes spectral solvers.

This paper focuses on spectral solvers for PIC codes and their advantages—with an emphasis on their application to laser-plasma interactions. Note that, for the sake of conciseness, the present paper is restricted to Particle-In-Cell codes that do solve Maxwell's equations on a grid (in which case spectral solvers are sometimes referred to as *pseudo-spectral*), and does not discuss the set of *gridless* spectral electromagnetic algorithms that have been recently developed, in the context of accelerator simulations (e.g. [9–11]).

\* rlehe@lbl.gov

## SPECTRAL SOLVERS, AND DIFFERENCE WITH FDTD SOLVERS

In order to summarize the principle of the spectral solvers, let us contrast them with the FDTD algorithm. In the standard Yee FDTD algorithm [8], Maxwell's equations

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}, \quad (1)$$

$$\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \mu_0 \mathbf{j} \quad (2)$$

are discretized in two ways:

- Spatial derivatives are approximated by a finite difference between neighboring points on a staggered grid. For instance,

$$\partial_x E_x|_{i,j,k}^n = \frac{E_x|_{i+1/2,j,k}^n - E_x|_{i-1/2,j,k}^n}{\Delta x}.$$

- Time derivatives are approximated by a finite difference between consecutive time steps. For instance,

$$\partial_t E_x|_{i+1/2,j,k}^{n+1/2} = \frac{E_x|_{i+1/2,j,k}^{n+1} - E_x|_{i+1/2,j,k}^n}{\Delta t}.$$

where we adopted standard notations whereby superscripts represent the index of the time step whereas subscripts represent positions on a staggered grid. As a consequence of the above simple space and time discretizations, the discretized Maxwell equations can easily be rewritten as a set of explicit update equations for the  $\mathbf{E}$  and  $\mathbf{B}$  fields.

While the above approximations allow fast execution and efficient parallelization, they also introduce numerical artifacts. One of these numerical artifacts is *spurious numerical dispersion*, i.e. the fact that the phase velocity of simulated electromagnetic waves (in vacuum) differs from the speed of light  $c$ , and depends on their wavelength and propagation angle. Spurious numerical dispersion can have a very serious impact in realistic simulations, and can lead to unphysical results. For instance, in the context of laser-plasma interactions, numerical dispersion can lead to spurious early dephasing in laser-driven accelerator [12], unphysical growth of emittance [13], and erroneous angle-frequency correlations in high-harmonics generation [14].

One of the main motivation for spectral solvers is to mitigate numerical dispersion. This is done by overcoming the approximations of FDTD schemes in two ways:

- Spatial derivatives are approximated by a high-order expression involving many grid points. These derivatives are typically evaluated in Fourier space for efficiency. Algorithms that use this feature but retain a

finite-difference approximation in time are typically referred to as *Pseudo-Spectral Time Domain* algorithms (PSTD) [15, 16].

- Instead of using a finite-difference approximation of the time derivatives, Maxwell's equation can be integrated analytically over one time step, in spectral space. Algorithms that use this additional feature are referred to as *Pseudo-Spectral Analytical Time Domain* algorithms (PSATD) [17, 18].

The above two points are explained in more details in the next paragraphs.

The high-order approximation of the spatial derivative in PSTD and PSATD schemes are of the form

$$\partial_x E_x|_{i,j,k}^n = \sum_{\ell=0}^{p/2-1} c_{\ell,p} \frac{E_x|_{i+1/2+\ell,j,k}^n - E_x|_{i-1/2-\ell,j,k}^n}{\Delta x},$$

where  $p$  is the order of the approximation and  $c_{\ell,p}$  are coefficients that are given in [19] (For  $p = 2$ , this reduces to the standard finite-difference expression of the Yee solver). It is common to use large values for  $p$  in practice (e.g.  $p = 32$  or  $p = 64$ ). Since these large values entail a significant computational cost, these derivatives are often more efficiently evaluated in Fourier space, by using the convolution theorem, i.e.

$$\mathcal{F}[\partial_x E_x] = i[k_x]_p \hat{E}_x,$$

$$[k_x]_p = \sum_{\ell=0}^{p/2-1} c_{\ell,p} \frac{e^{ik_x(\ell+1/2)\Delta x} - e^{-ik_x(\ell+1/2)\Delta x}}{i\Delta x},$$

where  $\mathcal{F}$  represents the Fourier transform and  $\hat{E}_x \equiv \mathcal{F}[E_x]$ . Note that, when  $p$  goes to infinity,  $[k_x]_p$  goes to  $k_x$ , and thus the spatial derivatives are evaluated with full spectral accuracy. However, the case  $p = \infty$  is rarely used in practice due to parallelization issues (as explained in the next section).

Using high-order spatial derivatives (i.e. high  $p$ ) makes the numerical dispersion relation less anisotropic. However, it does not mitigate spurious numerical dispersion altogether, unless the treatment of the time derivative is made more accurate too. One way to achieve this is to retain a finite-difference approximation in time, but use a small time step (much smaller than the Courant-Friedrichs-Lewy limit)—but this comes at a high computational cost. Another possible approach is to integrate Maxwell's equations analytically in spectral space.

When evaluated in spectral space, the spatially-discretized Maxwell equations indeed reduce to a simple set of ordinary differential equations with constant coefficients:

$$\frac{\partial \hat{\mathcal{B}}(\mathbf{k}, t)}{\partial t} = -i[\mathbf{k}]_p \times \hat{\mathcal{E}}(\mathbf{k}, t) \quad (3)$$

$$\frac{1}{c^2} \frac{\partial \hat{\mathcal{E}}(\mathbf{k}, t)}{\partial t} = i[\mathbf{k}]_p \times \hat{\mathcal{B}}(\mathbf{k}, t) - \mu_0 \hat{\mathcal{J}}(\mathbf{k}, t). \quad (4)$$

As such, these equations can be integrated analytically from  $t = n\Delta t$  to  $t = (n+1)\Delta t$ , under the assumption that  $\hat{\mathcal{J}}(\mathbf{k}, t)$

is constant over one time step [17, 18]. By taking  $\hat{\mathcal{E}}^n$  and  $\hat{\mathcal{B}}^n$  as initial conditions, this analytical integration can yield  $\hat{\mathcal{E}}^{n+1}$  and  $\hat{\mathcal{B}}^{n+1}$ , i.e. the updated fields at the next time step:

$$\hat{\mathcal{E}}^{n+1} = C\hat{\mathcal{E}}^n + iS \frac{[\mathbf{k}]_p}{[k]_p} \times \hat{\mathcal{B}}^n - \frac{S}{[k]_p} \hat{\mathcal{J}}^{n+1/2} + (1-C) \frac{[\mathbf{k}]_p([\mathbf{k}]_p \cdot \hat{\mathcal{E}}^n)}{[k]_p^2} + \frac{[\mathbf{k}]_p([\mathbf{k}]_p \cdot \hat{\mathcal{J}}^{n+1/2})}{[k]_p^2} \left( \frac{S}{[k]_p} - \Delta t \right), \quad (5)$$

$$\hat{\mathcal{B}}^{n+1} = C\hat{\mathcal{B}}^n - iS \frac{[\mathbf{k}]_p \times \hat{\mathcal{E}}^n}{[k]_p} + i \frac{1-C}{[k]_p^2} [\mathbf{k}]_p \times \hat{\mathcal{J}}^{n+1/2}, \quad (6)$$

where  $C = \cos([k]_p c \Delta t)$ ,  $S = \sin([k]_p c \Delta t)$ , and  $[k]_p = \sqrt{[k]_p^2}$ .

In summary, the update of the  $E$  and  $B$  fields (from time step  $n$  to time step  $n+1$ ) in the PSATD scheme consists of three steps:

1. Apply a forward Fourier transform, in order to obtain the fields in spectral space at time step  $n$  ( $\hat{\mathcal{E}}^n$ ,  $\hat{\mathcal{B}}^n$ ) from the fields in real space ( $E^n$ ,  $B^n$ ).
2. Apply Eqs. (5)-(6) to obtain the fields in spectral space at time step  $n+1$  ( $\hat{\mathcal{E}}^{n+1}$ ,  $\hat{\mathcal{B}}^{n+1}$ ).
3. Apply an inverse Fourier transform, in order to obtain the fields in real space at time step  $n+1$  ( $E^{n+1}$ ,  $B^{n+1}$ ) from the fields in spectral space.

Using this scheme and a high spatial order  $p$ , spurious numerical dispersion can generally be mitigated to negligible levels.

## RECENT DEVELOPMENTS

### Parallelization

One of the major issues with spectral solvers has been their scaling across many computing nodes. In principle, the PSTD and PSATD schemes require a global Fourier transform across the whole computational domain. When using a standard domain decomposition technique for parallelization, this Fourier transform can be implemented by a global, distributed Fast Fourier Transform (FFT). However, distributed FFTs do not scale well to many compute nodes, due to the large amount of inter-node communications that they involve.

In the case of Maxwell's equations, one alternative to the global, distributed FFT is to use a local FFT in each sub-domain, along with guard cells [18]. In principle, using a local FFT instead of a global one introduces errors. But, because the Maxwell update Eqs. (1)–(2) are a hyperbolic set of equations, the errors remains confined to the guard

cells (provided that there are enough of them) [18,20]. These errors can thus be eliminated by copying the data from the valid regions of neighboring sub-domains into the guard cells of the local sub-domain, immediately after the Maxwell field update.

The minimal number of guard cells that eliminates these errors can be calculated for any given order  $p$  [20]. This number increases with  $p$ ; and for typical orders (e.g.  $p = 32$ ,  $p = 64$ ), tens of guard cells are required. While this number is large in comparison with FDTD (only one or two guard cells required), its impact on performance can be mitigated by using relatively large sub-domains, with several cores working on the same sub-domain through shared-memory programming paradigms.

By using local FFTs and guard cell exchanges between neighboring sub-domains, a much more favorable scaling can be obtained than with global, distributed FFTs. For instance, it was possible to reach nearly-linear strong scaling on a few hundreds of thousands of cores in [21], whereas global FFTs become prohibitively expensive at this scale.

### *Cylindrical Geometry*

For problems with nearly-cylindrical symmetry, a full 3D Cartesian mesh is not always optimal. In fact, by using an azimuthal expansion and a few 2D  $r$ - $z$  grids (one per azimuthal mode), computational costs can be very substantially reduced [22].

PIC codes using the azimuthal expansion were first implemented using the FDTD approach [22–24]. A hybrid PSTD-type algorithm was later implemented by retaining a finite-difference approach in the radial direction while using a spectral approach for the longitudinal derivatives [25]. Finally, a fully spectral (in  $r$  and  $z$ ) PSATD algorithm was derived and implemented [26].

In cylindrical geometry, the PSATD algorithm relies on a Fourier transform along the  $z$  direction, and a Hankel transform along the  $r$  direction. It turns out that the Fourier-Hankel representation of Maxwell's equations has a similar structure as the 3D Cartesian Fourier representation (Eqs. (3)–(4)). Therefore, the Fourier-Hankel representation of Maxwell's equations can also be integrated analytically over one time step, and, as a result, also mitigates spurious numerical dispersion [26].

## **SOME ADDITIONAL ADVANTAGES OF SPECTRAL SOLVERS**

While, in the previous sections, the spectral solvers were mainly motivated by the mitigation of the spurious numerical dispersion, they do have additional advantages. Two of these advantages are presented in the next subsections.

### *Accurate Evaluation of the Lorentz Force*

A common shortcoming of FDTD PIC codes is the inaccuracy of the calculated Lorentz force in cases where the  $\mathbf{E}$  and  $\mathbf{v} \times \mathbf{B}$  cancel very closely. This situation arises for instance when a relativistic bunch of electrons co-propagates

with a laser: in this case, the term  $\mathbf{E} + \mathbf{v} \times \mathbf{B}$  is on the order of  $\mathbf{E}/\gamma^2$ , where  $\gamma$  is the Lorentz factor of the bunch.

This tight cancelation of the  $\mathbf{E}$  and  $\mathbf{v} \times \mathbf{B}$  terms is difficult to capture in an FDTD PIC code. This is because the fields  $\mathbf{E}$  and  $\mathbf{B}$  are staggered in time and space, and are therefore not evaluated exactly in the same way, when interpolated onto the macroparticles. (For instance,  $\mathbf{B}$  is typically averaged in time, before being interpolated to the macroparticles, whereas  $\mathbf{E}$  is not.) These slight differences in the evaluation of  $\mathbf{E}$  and  $\mathbf{B}$  introduce small numerical errors that can dominate the term  $\mathbf{E} + \mathbf{v} \times \mathbf{B}$  in the case where the two terms cancel closely [27].

On the other hand, in the PSATD algorithm, the fields  $\mathbf{E}$  and  $\mathbf{B}$  are both defined at the same time (and it is also possible to define them on the same points in space). As a result, the fields are evaluated in a similar way, and can cancel appropriately. This was confirmed for instance in test simulations of an electron co-propagating with a laser [26].

### *Mitigation of the Numerical Cherenkov Instability*

Some types of PIC simulations can become numerically much cheaper when using a different reference frame than that of the laboratory [28]. This is the case for instance in simulations of laser-wakefield acceleration, where the discrepancy between the laser wavelength and the length of the accelerating plasma stage can be greatly reduced in an appropriate Lorentz frame (commonly known as the “boosted frame”) [28]. Yet, when using the boosted frame, the bulk of the plasma stage moves relativistically with respect to the grid. In PIC codes, this gives rise to a numerical instability, known as the Numerical Cherenkov Instability (NCI) [29,30].

Various methods have been used to mitigate the NCI, so as to carry out robust boosted-frame simulations. Some of these methods can be applied to the FDTD or PSTD approach [31–38]. However, the PSATD approach is the only one that is compatible with a particularly robust and elegant mitigation technique: the Galilean technique [39,40].

In the Galilean technique, Maxwell's equations are solved in a Galilean system of coordinates that moves with the bulk of the plasma. It turns out that solving Maxwell's equations with the PSATD scheme in these coordinates suppresses the NCI—without the need for any further numerical correction [39,40]. Additionally, this method naturally generalizes to cylindrical coordinates, when using the above mentioned Fourier-Hankel representation.

## **CONCLUSION**

Simulations of accelerators can be numerically challenging when using the electromagnetic PIC algorithm. Spectral solvers can alleviate some of these numerical challenges, by using a more accurate discretization of Maxwell's equations. In addition, recent developments allow to efficiently parallelize these solvers across many nodes, and to port them to cylindrical geometry.

Yet, it is important to keep in mind that spectral solvers are one possible solution among others, for many of the above challenges. For instance, spurious numerical dispersion can be mitigated to some degree by non-standard finite-difference methods (e.g. [13,41–43]). Similarly, the inaccuracy in the Lorentz force can be partially mitigated by using higher-order interpolation in time [27]

## ACKNOWLEDGEMENTS

This work was partly supported by the Director, Office of Science, Office of High Energy Physics, U.S. Dept. of Energy under Contract No. DE-AC02-05CH11231, and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, in support of the nation's exascale computing imperative.

This document was prepared as an account of work sponsored in part by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, nor the authors makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

## REFERENCES

- [1] R. Hockney and J. Eastwood, *Computer Simulation Using Particles*. Taylor & Francis, 1988.
- [2] C. Birdsall and A. Langdon, *Plasma Physics via Computer Simulation, Appendix E*. Series in Plasma Physics, Taylor & Francis, 2004.
- [3] W. P. Leemans, A. J. Gonsalves, H.-S. Mao, K. Nakamura, C. Benedetti, C. B. Schroeder, C. Tóth, J. Daniels, D. E. Mittelberger, S. S. Bulanov, J.-L. Vay, C. G. R. Geddes, and E. Esarey, “Multi-gev electron beams from capillary-discharge-guided subpetawatt laser pulses in the self-trapping regime,” *Phys. Rev. Lett.*, vol. 113, p. 245002, Dec 2014.
- [4] E. Esarey, C. B. Schroeder, and W. P. Leemans *Rev. Mod. Phys.*, vol. 81, pp. 1229–1285, Aug 2009.
- [5] S. P. D. Mangles, C. D. Murphy, Z. Najmudin, A. G. R. Thomas, J. L. Collier, A. E. Dangor, E. J. Divall, P. S. Foster, J. G. Gallacher, C. J. Hooker, D. A. Jaroszynski, A. J.

Langley, W. B. Mori, P. A. Norreys, F. S. Tsung, R. Viskup, B. R. Walton, and K. Krushelnick, “Monoenergetic beams of relativistic electrons from intense laser-plasma interactions,” *Nature*, vol. 431, pp. 535–538, Sep 2004.

- [6] J. Faure, Y. Glinec, A. Pukhov, S. Kiselev, S. Gordienko, E. Lefebvre, J.-P. Rousseau, F. Burgy, and V. Malka, “A laser-plasma accelerator producing monoenergetic electron beams,” *Nature*, vol. 431, pp. 541–544, Sep 2004.
- [7] C. G. R. Geddes, C. Toth, J. van Tilborg, E. Esarey, C. B. Schroeder, D. Bruhwiler, C. Nieter, J. Cary, and W. P. Leemans, “High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding,” *Nature*, vol. 431, pp. 538–541, Sep 2004.
- [8] K. Yee, “Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media,” *Antennas and Propagation, IEEE Transactions on*, vol. 14, pp. 302–307, may 1966.
- [9] S. D. Webb, D. T. Abell, N. M. Cook, and D. L. Bruhwiler, “A Spectral Symplectic Algorithm for Cylindrical Electromagnetic Plasma Simulations,” *ArXiv e-prints*, Sept. 2016.
- [10] D. T. Abell, N. M. Cook, and S. D. Webb, “Symplectic modeling of beam loading in electromagnetic cavities,” *Phys. Rev. Accel. Beams*, vol. 20, p. 052002, May 2017.
- [11] I. Andriyash, R. Lehe, and V. Malka, “A spectral unaveraged algorithm for free electron laser simulations,” *Journal of Computational Physics*, vol. 282, pp. 397–409, 2015.
- [12] B. M. Cowan, D. L. Bruhwiler, J. R. Cary, E. Cormier-Michel, and C. G. R. Geddes *Phys. Rev. ST Accel. Beams*, vol. 16, p. 041303, Apr 2013.
- [13] R. Lehe, A. Lifschitz, C. Thaury, V. Malka, and X. Davoine *Phys. Rev. ST Accel. Beams*, vol. 16, p. 021301, Feb 2013.
- [14] G. Blaclard, H. Vincenti, R. Lehe, and J. L. Vay, “Pseudospectral maxwell solvers for an accurate modeling of doppler harmonic generation on plasma mirrors with particle-in-cell codes,” *Phys. Rev. E*, vol. 96, p. 033305, Sep 2017.
- [15] J. M. Dawson *Rev. Mod. Phys.*, vol. 55, pp. 403–447, Apr 1983.
- [16] Q. H. Liu *Microwave and Optical Technology Letters*, vol. 15, no. 3, pp. 158–165, 1997.
- [17] I. Haber, R. Lee, H. Klein, and J. Boris. 1973.
- [18] J.-L. Vay, I. Haber, and B. B. Godfrey *Journal of Computational Physics*, vol. 243, pp. 260–268, 2013.
- [19] B. Fornberg, “High-order finite differences and the pseudospectral method on staggered grids,” *SIAM Journal on Numerical Analysis*, vol. 27, no. 4, pp. 904–918, 1990.
- [20] H. Vincenti and J.-L. Vay, “Detailed analysis of the effects of stencil spatial variations with arbitrary high-order finite-difference maxwell solver,” *Computer Physics Communications*, vol. 200, pp. 147–167, 2016.
- [21] H. Vincenti and J.-L. Vay, “Ultra-high-order maxwell solver with extreme scalability for electromagnetic pic simulations of plasmas,” *Computer Physics Communications*, vol. 228, pp. 22–29, 2018.
- [22] A. F. Lifschitz, X. Davoine, E. Lefebvre, J. Faure, C. Rechatin, and V. Malka *J. Comput. Phys.*, vol. 228, pp. 1803–1814, Mar. 2009.

- [23] A. Davidson, A. Tableman, W. An, F. Tsung, W. Lu, J. Vieira, R. Fonseca, L. Silva, and W. Mori, "Implementation of a hybrid particle code with a pic description in r-z and a gridless description in  $\phi$  into osiris," *Journal of Computational Physics*, vol. 281, pp. 1063–1077, 2015.
- [24] P. Lee, T. Audet, R. Lehe, J.-L. Vay, G. Maynard, and B. Cros, "Modeling laser-driven electron acceleration using warp with fourier decomposition," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 829, pp. 358–362, 2016. 2nd European Advanced Accelerator Concepts Workshop–EAAC 2015.
- [25] P. Yu, X. Xu, A. Tableman, V. K. Decyk, F. S. Tsung, F. Fiuza, A. Davidson, J. Vieira, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, "Mitigation of numerical cerenkov radiation and instability using a hybrid finite difference-fft maxwell solver and a local charge conserving current deposit," *Computer Physics Communications*, vol. 197, 12 2015.
- [26] R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, "A spectral, quasi-cylindrical and dispersion-free particle-in-cell algorithm," *Computer Physics Communications*, vol. 203, pp. 66–82, 2016.
- [27] R. Lehe, C. Thauray, E. Guillaume, A. Lifschitz, and V. Malka *Phys. Rev. ST Accel. Beams*, vol. 17, p. 121301, Dec 2014.
- [28] J.-L. Vay *Phys. Rev. Lett.*, vol. 98, p. 130405, Mar 2007.
- [29] B. B. Godfrey *Journal of Computational Physics*, vol. 15, no. 4, pp. 504–521, 1974.
- [30] B. B. Godfrey, "Canonical momenta and numerical instabilities in particle codes," *Journal of Computational Physics*, vol. 19, no. 1, pp. 58–76, 1975.
- [31] B. B. Godfrey, J.-L. Vay, and I. Haber *Journal of Computational Physics*, vol. 258, pp. 689–704, 2014.
- [32] B. Godfrey, J.-L. Vay, and I. Haber *Plasma Science, IEEE Transactions on*, vol. 42, pp. 1339–1344, May 2014.
- [33] B. B. Godfrey and J.-L. Vay, "Suppressing the numerical cerenkov instability in {FDTD} {PIC} codes," *Journal of Computational Physics*, vol. 267, pp. 1–6, 2014.
- [34] B. B. Godfrey, "Review and recent advances in pic modeling of relativistic beams and plasmas," *AIP Conference Proceedings*, vol. 1777, no. 1, 2016.
- [35] B. B. Godfrey and J.-L. Vay *Computer Physics Communications*, 2015.
- [36] P. Yu, X. Xu, V. K. Decyk, F. Fiuza, J. Vieira, F. S. Tsung, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, "Elimination of the numerical cerenkov instability for spectral em-pic codes," *Computer Physics Communications*, vol. 192, pp. 32–47, 2015.
- [37] P. Yu, X. Xu, A. Tableman, V. K. Decyk, F. S. Tsung, F. Fiuza, A. Davidson, J. Vieira, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, "Mitigation of numerical cerenkov radiation and instability using a hybrid finite difference-fft maxwell solver and a local charge conserving current deposit," *Computer Physics Communications*, vol. 197, pp. 144–152, 2015.
- [38] F. Li, P. Yu, X. Xu, F. Fiuza, V. K. Decyk, T. Dalichaouch, A. Davidson, A. Tableman, W. An, F. S. Tsung, R. A. Fonseca, W. Lu, and W. B. Mori, "Controlling the numerical cerenkov instability in pic simulations using a customized finite difference maxwell solver and a local fft based current correction," *Computer Physics Communications*, vol. 214, 1 2017.
- [39] R. Lehe, M. Kirchen, B. B. Godfrey, A. R. Maier, and J.-L. Vay, "Elimination of Numerical Cherenkov Instability in flowing-plasma Particle-In-Cell simulations by using Galilean coordinates," *arXiv:1608.00227 [physics]*, July 2016.
- [40] M. Kirchen, R. Lehe, B. B. Godfrey, I. Dornmair, S. Jalas, K. Peters, J.-L. Vay, and A. R. Maier, "Stable discrete representation of relativistically drifting plasmas," *Physics of Plasmas*, vol. 23, no. 10, 2016.
- [41] M. Karkkainen, E. Gjonaj, T. Lau, and T. Weiland, vol. 1. 2006.
- [42] A. Pukhov *Journal of Plasma Physics*, vol. 61, no. 03, pp. 425–433, 1999.
- [43] R. Nuter, M. Grech, P. Gonzalez de Alaiza Martinez, G. Bonnaud, and E. d’Humières *The European Physical Journal D*, vol. 68, no. 6, 2014.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## List of Authors

**Bold** papercodes indicate primary authors; ~~crossed-out~~ papercodes indicate 'no submission'

### — A —

Abell, D.T.	<del>TUPAF04</del> , WEPAF03
Abramov, A.	TUPAF15
Acharya, D.	SUPAF12
Ackermann, W.	<b>TUPAG03</b>
Adelmann, A.	SUPAF11, SUPAF12, <b>MOPLG01</b> , <del>TUPAF07</del> , TUPAG05, <b>TUPAG14</b>
Aleksandrov, A.V.	<b>SUPAF10</b> , <b>TUPAG12</b>
Alimohamadi, M.	<b>SUPAG12</b>
Amorim, D.	<b>SUPAG01</b>
Andersson, J.	<b>TUPAF02</b>
André, K.	SUPAF03
Appel, S.	<b>SAPAF02</b>
Appelö, D.	<b>MOPLG03</b> , <b>MOPAF04</b>
Arbenz, P.	<b>TUPAG14</b>
Arduini, G.	<b>SUPAG01</b>
Aßmann, R.W.	<b>MOPLG01</b>
Avrakhov, P.V.	<b>TUPAF19</b>

### — B —

Bahls, C.R.	<b>TUPAG02</b>
Banerjee, D.	<b>SUPAG05</b>
Barber, D.P.	<b>MOPLG03</b> , <b>MOPAF04</b>
Bartnik, A.C.	<b>TUPAF10</b>
Bartosik, H.	<b>SUPAG01</b>
Bassi, G.	<b>TUPAF12</b>
Beaudoin, B.L.	<b>TUPAG10</b>
Beck, A.	<b>MOPAG02</b>
Beekman, I.B.	<del>TUPAF04</del>
Beltrán, J.R.	<b>SUPAF08</b>
Benedetto, E.	<b>SUPAG01</b>
Berg, J.S.	<b>TUPAF09</b> , <b>TUPAF10</b> , <b>TUPAF13</b>
Bernhard, J.	<b>SUPAG05</b>
Berz, M.	<b>MOPAF02</b> , <del>TUPAF11</del> , <b>TUPAG22</b>
Bettoni, S.	<b>MOPLG01</b>
Beznosov, O.	<b>MOPLG03</b> , <b>MOPAF04</b>
Bizzozero, D. A.	<b>MOPAG03</b>
Black, D.S.	<b>MOPLG01</b>
Böker, J.	<b>SUPAF05</b>
Boine-Frankenheim, O.	<b>SAPAG02</b> , <b>MOPLG01</b>
Boogert, S.T.	<b>SUPAG11</b> , <b>TUPAF15</b>
Boonpornprasert, P.	<b>TUPAF23</b>
Broaddus, P. N.	<b>MOPLG01</b>
Brooks, S.J.	<b>TUPAF08</b> , <b>TUPAF09</b> , <b>TUPAF13</b>
Brouwer, L.N.	<b>SUPAF01</b>
Brugger, M.	<b>SUPAG05</b>
Bruhwieler, D.L.	<del>WEPAF03</del> , <b>WEPLG03</b>
Burkhardt, H.	<b>SUPAG01</b> , <b>TUPAF01</b>
Burov, A.V.	<b>WEPLG03</b>
Byer, R.L.	<b>MOPLG01</b>
Büscher, M.	<b>SUPAF05</b>

### — C —

Caiafa, G.	<b>SUPAG03</b>
Calvi, M.	<b>MOPLG01</b>
Cankaya, H.	<b>MOPLG01</b>
Carlsson, J.A.	<del>WEPAF03</del>
Cathey, B.L.	<b>TUPAG12</b>
Cauchois, A.	<b>WEPAF01</b>
Ceballos, A.C.	<b>MOPLG01</b>
Cerfon, A.J.	<b>SUPAF09</b>
Cesar, D.B.	<b>MOPLG01</b>
Charitonidis, N.	<b>SUPAG05</b>
Chen, Q.S.	<b>SUPAF02</b>
Chen, S.	<b>SUPAF13</b>
Cheng, W.X.	<b>SAPAF01</b>
Chetvertkova, V.	<b>SAPAG02</b>
Collignon, V.	<b>SUPAF03</b>
Cook, N.M.	<b>SUPAF07</b> , <b>WEPLG03</b>
Corno, J.	<b>TUPAG08</b>
Cousineau, S.M.	<b>SUPAF10</b> , <b>TUPAG12</b>
Cowan, B.M.	<b>MOPLG01</b>
Crittenden, J.A.	<b>TUPAF08</b>

### — D —

De Gersem, H.	<b>MOPAG03</b> , <b>TUPAG03</b> , <b>TUPAG08</b> , <b>WEPLG02</b>
De Maria, R.	<b>TUPAF01</b> , <b>TUPAF02</b>
De Prisco, R.	<b>TUPAG11</b>
Dehler, M.M.	<b>MOPLG01</b>
Deng, H.	<b>MOPLG01</b>
Deng, H.X.	<b>SUPAF13</b> , <b>TUPAG19</b>
Deniau, L.	<b>TUPAF01</b>
Derouillat, J.	<b>MOPAG02</b>
Dobbins, J.	<b>TUPAF10</b>
Döbrich, B.	<b>SUPAG05</b>
Dorda, U.	<b>MOPLG01</b>
Dubus, A.	<b>SUPAG11</b>
Duxbury, P.M.	<b>TUPAF20</b>

### — E —

Edelen, A.L.	<b>SUPAF12</b>
Edelen, J.P.	<b>SUPAF07</b> , <b>TUPAG17</b>
Egenolf, T.	<b>MOPLG01</b>
Eidelman, Y.I.	<del>WEPAF03</del>
Ellison, J.A.	<b>MOPLG03</b> , <b>MOPAF04</b>
Engels, R.	<b>SUPAF05</b>
Engin, I.	<b>SUPAF05</b>
England, R.J.	<b>MOPLG01</b>
Erdelyi, B.	<b>SAPAG01</b>
Eshraqi, M.	<b>TUPAG11</b>

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

— F —

Fakhari, M.	MOPLG01
Fallahi, A.	MOPLG01
Fan, S.	MOPLG01
Fedurin, M.G.	TUPAG09
Feng, C.	SUPAF13
Feng, G.	SUPAF02
Ferrari, E.	MOPLG01
Feurer, T.	MOPLG01
Field, L.	TUPAF02
Fiscarelli, L.	SUPAG03
Flandroy, Q.	SUPAF03
Flisgen, T.	TUPAG01
Forton, E.	SUPAG09
Frei, F.	MOPLG01
Frey, M.	SUPAF11, SUPAF12, TUPAG05

— G —

Garcia Morales, H.	TUPAF15
Gatignon, L.	SUPAG05
Ge, L.	TUPAF05
Gebel, R.	SUPAF05
Geithner, W.	SAPAF02
Georg, N. G.	TUPAG08
Gerbershagen, A.	SUPAG05
Gerity, J.	WEPAF03
Gibson, S.M.	TUPAF15
Giovanozzi, M.	TUPAF01, TUPAF02
Gjonaj, E.	MOPLG02, MOPAG03, WEPLG02
Gläßle, T.G.	TUPAF01
Gnacadjia, E.	SUPAG11
Goldring, N.B.	TUPAG17
González, J.D.	SUPAF08
Gorgi Zadeh, S.	TUPAG08, TUPAG14
Grandsaert, T.J.	TUPAG11
Grech, M.	MOPAG02
Gu, Q.	SUPAF13
Gulliford, C.M.	TUPAF10

— H —

Haber, I.	TUPAG10
Hall, C.C.	SUPAF07, WEPAF03
Han, B.	WEPAF02
Han, W.J.	SUPAG10
Hao, Y.	SUPAG06, TUPAG20
Harris, J.S.	MOPLG01
Hartl, I.	MOPLG01
Hauenstein, D.	MOPLG01
Heinemann, K.A.	MOPLG03, MOPAF04
Heller, J.	TUPAG01, TUPAG02, TUPAG07
Hermann, B.	MOPLG01
Hermes, P.D.	TUPAF02
Hernalsteens, C.	SUPAF03, SUPAG11
Herregods, B.	SUPAF03
Hidas, D.A.	TUPAG21

Hiller, N.	MOPLG01
Hirano, T.	MOPLG01
Hoimyr, N.	TUPAF02
Hommelhoff, P.	MOPLG01
Huang, N. S.	TUPAG19
Huang, Y.-C.	MOPLG01
Huang, Z.	MOPLG01
Hudson, S. T. P.	SAPAF03
Huggins, A.M.	SUPAF01
Hughes, T.W.	MOPLG01
Hwang, K.	TUPAF22
Hützen, A.	SUPAF05

— I —

Iadarola, G.	TUPAF02
Iliakis, K.	TUPAF06
Illmer, J.	MOPLG01
Ischebeck, R.	MOPLG01
Ives, R.L.	SAPAF04
Iwasaki, Y.	TUPAF16

— J —

Jansson, A.	TUPAG11
Jensen, A.	SAPAF04
Jiang, Y.	MOPLG01
Jing, C.-J.	TUPAF19
Johnstone, C.	TUPAF11
Jowett, J.M.	TUPAF01
Jung, P. M.	SUPAF04, TUPAG13
Jungers, R.	SUPAF03

— K —

Kärtner, F.X.	MOPLG01
Keilman, M.V.	SUPAF07
Khabiboulline, T.N.	TUPAG04
Khojuyan, M.	WEPAF01
Kleeven, W.J.G.M.	SUPAG09
Kocevar, H.	TUPAG11
Koeth, T.W.	TUPAG10
Kornilov, V.	SAPAG02
Kostin, R.A.	TUPAF19
Kostoglou, S.	TUPAF02
Kramer, P.	SAPAG04
Kranjcevic, M.	TUPAG14
Krasilnikov, M.	TUPAF23
Kuroпка, W.	MOPLG01

— L —

Langenstein, T.	MOPLG01
Larson, J.M.	SAPAF03
Latina, A.	TUPAF01
Lebedev, V.A.	WEPLG03
Lee, Y.J.	MOPLG01
Leedle, K.J.	MOPLG01
Lehé, R.	WEPLG05

Lehrach, A. SUPAF05  
 Lemery, F. MOPLG01  
 Levinsen, Y. TUPAG11  
 Li, A. MOPLG01  
 Li, K. TUPAG19  
 Li, K.S.B. SUPAG01  
 Li, Y. SAPAF01  
 Li, Z. TUPAF05  
 Liebsch, M. SUPAG03  
 Litvinenko, V. SUPAF06  
 Liu, A. TUPAF19  
 Liu, B. SUPAF13  
 Liu, X. SUPAF02, SUPAG10  
 Lombosi, C. MOPLG01  
 Lou, W. TUPAF09  
 Lunin, A. TUPAG04  
 Luo, Y. SUPAG06

— M —

Ma, J. SUPAF06  
 Maclean, E.H. TUPAF02  
 Makino, K. MOPAF02, TUPAF11  
 Mandrillon, J. SUPAG09  
 Marchetti, B. MOPLG01  
 Marocchino, A. MOPAG01  
 Maruta, T. TUPAG20  
 Massimo, F. MOPAG02  
 Matthew, D.B. TUPAG10  
 Mayes, C.E. SUPAG04  
 Mayet, F. MOPLG01  
 McIntosh, E. TUPAF02  
 McIntyre, P.M. WEPAF03  
 Méot, F. MOPAF03, TUPAF04, TUPAF08,  
 TUPAF09, TUPAF13  
 Mereghetti, A. TUPAF02  
 Meseck, A. TUPAF14  
 Métral, E. SUPAG01  
 Miao, Y. MOPLG01  
 Midttun, Ø. TUPAG11  
 Milas, N. TUPAG11  
 Mitchell, C.E. WEPLG01  
 Mittelbach, A.K. MOPLG01  
 Miyamoto, R. TUPAG11  
 Moeller, P. SUPAF07  
 Molson, J. TUPAF02  
 Montbarbon, E. SUPAG05  
 Murray, S.N. WEPAF02  
 Musumeci, P. MOPLG01

— N —

Nagaitsev, S. WEPLG03  
 Nagler, R. SUPAF07  
 Naranjo, B. MOPLG01  
 Nash, B. TUPAG17  
 Neilson, J. SAPAF04  
 Nevay, L.J. SUPAG11, TUPAF15

Neveu, N.R. SUPAF12, SAPAF03  
 Ng, C.-K. TUPAF05  
 Niedermayer, U. MOPLG01  
 Nuhn, H.-D. TUPAF23  
 Nunez-delPrado, A. TUPAF10  
 Nuttens, V. SUPAG09

— O —

Oeftiger, A. SUPAG01  
 Olsen, V.K.B. TUPAF02  
 Orozco, E.A. SUPAF08, TUPAG23  
 Ostroumov, P.N. TUPAG20, WEPAF04  
 Otero Olarte, O. TUPAG23

— P —

Pauly, N. SUPAG11  
 Pellegrini, D. TUPAF02  
 Pennisi, T.R. WEPAF02  
 Peplov, V.V. WEPAF02  
 Pérez, F. MOPAG02  
 Persson, T. TUPAF01, TUPAF02  
 Petillo, J.J. SAPAF04  
 Petrone, C. SUPAG03  
 Pham-Xuan, V. TUPAG03  
 Pigott, A. MOPLG01  
 Pikhartova, H. TUPAF15  
 Planche, T. SUPAF04, TUPAG13  
 Plastun, A.S. TUPAG20, WEPAF04  
 Plostinar, D.C. TUPAG11  
 Pogorelov, I.V. WEPAF03  
 Pommerenke, H.W. TUPAG07  
 Ponton, A. TUPAG11  
 Prat, E. MOPLG01  
 Prudent, J. WEPAF01  
 Ptitsyn, V. TUPAF13  
 Pukhov, A.M. SUPAF05

— Q —

Qi, M. MOPLG01  
 Qiang, J. SAPLG01, SUPAG06, TUPAF22,  
 WEPLG01  
 Qin, B. SUPAF02, SUPAG10  
 Quatraro, D. SUPAG01

— R —

Rae, B. SUPAG05  
 Rainer, R.S. SAPAF01  
 Rakitzis, T. P. SUPAF05  
 Ranjbar, V.H. MOPAF05  
 Ratner, D.F. SUPAG08  
 Read, M.E. SAPAF04  
 Reiche, S. MOPLG01  
 Reimann, S. SAPAF02  
 Repond, J. TUPAF06  
 Ricketson, L.F. SUPAF09

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

Rivkin, L. **MOPLG01**  
 Roblin, Y. **SUPAG06**  
 Römer, U. **TUPAG08**  
 Rogacki, P. **SUPAG03**  
 Rosenthal, M.S. **SUPAG05**  
 Rosenzweig, J.B. **MOPLG01**  
 Rouson, D.W.I. **TUPAF04**  
 Ruisard, K.J. **TUPAG10**  
 Rumolo, G. **SUPAG01**  
 Russenschuck, S. **SUPAG03**  
 Ryne, R.D. **SUPLG01**

— S —

Saethre, R.B. **WEPAF02**  
 Sagan, D. **TUPAF10**  
 Salvant, B. **SUPAG01**  
 Sapinski, M. **SAPAF02**  
 Sapra, N. **MOPLG01**  
 Schmid, S. A. **WEPLG02**  
 Schmidt, F. **TUPAF01**  
 Schneidmiller, E. **TUPAF23**  
 Schönenberger, N. **MOPLG01**  
 Schöps, S. **TUPAG08**  
 Schwarz, M. **TUPAF06**  
 Schwinzerl, M. **TUPAF02**  
 Shaposhnikova, E.N. **TUPAF06**  
 Shea, T.J. **TUPAG11**  
 Shen, X. **MOPLG01**  
 Shields, W. **SUPAG11, TUPAF15**  
 Shiloh, R. **MOPLG01**  
 Shishlo, A.P. **TUPAF03**  
 Shreyber, I.S. **TUPAF01**  
 Simakov, E.I. **MOPLG01**  
 Singh, R. **SAPAF02**  
 Singh, S. **TUPAF02**  
 Sjobak, K.N. **TUPAF02**  
 Skär, E. **MOPLG01**  
 Skowroński, P.K. **TUPAF01**  
 Snopok, P. **TUPAF11**  
 Snuverink, J. **TUPAG05**  
 Sofikitis, D. **SUPAF05**  
 Solgaard, O. **MOPLG01**  
 Solyak, N. **TUPAG04**  
 Sorge, S. **SAPAG02**  
 Spadaro, T. **SUPAG05**  
 Specka, A. **MOPAG02, WEPAF01**  
 Spentzouris, L.K. **SAPAF03**  
 Stephan, F. **TUPAF23**  
 Stinson, C. **WEPAF02**  
 Stockli, M.P. **WEPAF02**  
 Stratakis, D. **MOPAF02**  
 Su, L. **MOPLG01**  
 Sukhanov, A.I. **TUPAG04**  
 Sulimov, A.A. **TUPAG08**  
 Sypfers, M.J. **MOPAF02, MOPAF01**

— T —

Tafel, A.D. **MOPLG01**  
 Tagger, J. **TUPAF12**  
 Tan, S. **MOPLG01**  
 Tarazona, D. **MOPAF02**  
 Tesse, R. **SUPAF03, SUPAG11**  
 Thomas, J. **SUPAF05**  
 Thomsen, H.D. **TUPAG11**  
 Trbojevic, D. **TUPAF08, TUPAF09, TUPAF13**  
 Tromp, R.M. **SUPLG02**  
 Tsoupas, N. **TUPAF08, TUPAF09, TUPAF13**

— V —

Valetov, E.V. **TUPAG22**  
 Valishev, A. **SAPLG02**  
 Van Beeumen, R. **TUPAG06**  
 van de Walle, J. **SUPAG09**  
 Van Der Kraaij, E. **SUPAG09**  
 Van Dijk, M.W.U. **SUPAG05**  
 van Rienen, U. **TUPAG01, TUPAG02, TUPAG07, TUPAG08, TUPAG14**  
 Vay, J.-L. **SUPAF07, WEPLG05**  
 Vélez, A.V. **TUPAG01**  
 Venturini, M. **SAPAG03**  
 Vergara, V.E. **SUPAF08**  
 Villari, A.C.C. **WEPAF04**  
 Vilsmeier, D.M. **SAPAF02**  
 Vollinger, C. **SAPAG04**  
 Volz, P.I. **TUPAF14**  
 Vuckovic, J. **MOPLG01**

— W —

Walker, S.D. **TUPAF15**  
 Wan, W. **SUPAF01, SUPAG07**  
 Wang, D. **SUPAF13, SUPAG07**  
 Wang, G. **SUPAF06**  
 Wang, Z. **SUPAF13**  
 Wang, Z. **SUPAF03**  
 Webb, S.D. **TUPAG17, WEPAF03, WEPLG03**  
 Welton, R.F. **WEPAF02**

— X —

Xiao, L. **TUPAF05**  
 Xuan, H. **MOPLG01**

— Y —

Yakovlev, V.P. **TUPAG04**  
 Yang, K. **MOPLG01**  
 Yoshimoto, T. **TUPAG20**  
 Yousefi, P. **MOPLG01**  
 Yu, L. **SAPAF01**  
 Yurkov, M.V. **TUPAF23**

— Z —

Zacharov, I.	TUPAF02	Zhang, Y.	WEPAF03
Zadeh, G.	TUPAG01	Zhang, Z.L.	<b>SUPAF10</b> , TUPAG12
Zannini, C.	SUPAG01	Zhao, Q.	TUPAG20, WEPAF04
Zemzemi, I.	MOPAG02	Zhao, Z.	MOPLG01
Zerbe, B.S.	<b>TUPAF20</b>	Zhao, Z.T.	SUPAF13
Zhang, H.	<b>SUPAG02</b> , WEPAF03	Zheng, D.	TUPAG02
Zhang, M.	SUPAF13	Zhou, K.S.	<b>TUPAF24</b>
Zhang, T.	TUPAG20	Zhu, J.	MOPLG01
		Zhukov, A.P.	TUPAG12

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2018). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

## *Institutes List*

### **ANL**

Argonne, Illinois, USA

- Hudson, S. T. P.
- Larson, J.M.
- Neveu, N.R.

### **AU**

Aarhus, Denmark

- Thomsen, H.D.

### **BNL**

Upton, Long Island, New York, USA

- Bassi, G.
- Berg, J.S.
- Brooks, S.J.
- Cheng, W.X.
- Fedurin, M.G.
- Hidas, D.A.
- Li, Y.
- Litvinenko, V.
- Luo, Y.
- Ma, J.
- Méot, F.
- Ptitsyn, V.
- Rainer, R.S.
- Ranjbar, V.H.
- Tagger, J.
- Trbojevic, D.
- Tsoupas, N.
- Wang, G.
- Yu, L.

### **CCR**

San Mateo, California, USA

- Ives, R.L.
- Read, M.E.

### **CERN**

Meyrin, Switzerland

- Amorim, D.
- Andersson, J.
- André, K.
- Arduini, G.
- Banerjee, D.
- Bartosik, H.
- Benedetto, E.
- Bernhard, J.
- Brugger, M.
- Burkhardt, H.
- Caiafa, G.
- Charitonidis, N.
- De Maria, R.
- Deniau, L.
- Döbrich, B.
- Field, L.
- Fiscarelli, L.

- Gatignon, L.
- Gerbershagen, A.
- Giovannozzi, M.
- Hermes, P.D.
- Hernalsteens, C.
- Hoimyr, N.
- Iadarola, G.
- Iliakis, K.
- Jowett, J.M.
- Kostoglou, S.
- Kramer, P.
- Latina, A.
- Li, K.S.B.
- Liebsch, M.
- Maclean, E.H.
- McIntosh, E.
- Mereghetti, A.
- Molson, J.
- Montbarbon, E.
- Métral, E.
- Oeftiger, A.
- Olsen, V.K.B.
- Pellegrini, D.
- Persson, T.
- Petrone, C.
- Quattraro, D.
- Rae, B.
- Repond, J.
- Rogacki, P.
- Rosenthal, M.S.
- Rumolo, G.
- Russenschuck, S.
- Salvant, B.
- Schmidt, F.
- Schwarz, M.
- Schwinzerl, M.
- Shaposhnikova, E.N.
- Shreyber, I.S.
- Sjobak, K.N.
- Skowroński, P.K.
- Van Dijk, M.W.U.
- Vollinger, C.
- Zannini, C.

### **CFEL**

Hamburg, Germany

- Cankaya, H.
- Fallahi, A.
- Kärtner, F.X.

### **Cornell University (CLASSE), Cornell Laboratory for Accelerator-Based Sciences and Education**

Ithaca, New York, USA

- Bartnik, A.C.
- Crittenden, J.A.
- Dobbins, J.
- Gulliford, C.M.

- Lou, W.
- Mayes, C.E.
- Sagan, D.

**Courant Institute of Mathematical Sciences, New York University**

New York, USA

- Cerfon, A.J.

**DESY**

Hamburg, Germany

- Aßmann, R.W.
- Barber, D.P.
- Dorda, U.
- Fakhari, M.
- Hartl, I.
- Kuroopka, W.
- Lemery, F.
- Marchetti, B.
- Mayet, F.
- Schneidmiller, E.
- Sulimov, A.A.
- Xuan, H.
- Yurkov, M.V.
- Zhu, J.

**DESY Zeuthen**

Zeuthen, Germany

- Boonpornprasert, P.
- Krasilnikov, M.
- Stephan, F.

**Deutsches Elektronen Synchrotron (DESY) and Center for Free Electron Science (CFEL)**

Hamburg, Germany

- Kärtner, F.X.

**EPFL**

Lausanne, Switzerland

- Ferrari, E.
- Rivkin, L.
- Zacharov, I.

**ESS**

Lund, Sweden

- De Prisco, R.
- Eshraqi, M.
- Grandsaert, T.J.
- Jansson, A.
- Kocevar, H.
- Levinsen, Y.
- Midttun, Ø.
- Milas, N.
- Miyamoto, R.
- Plostinar, D.C.
- Ponton, A.
- Shea, T.J.

**ETH**

Zurich, Switzerland

- Arbenz, P.
- Kranjcevic, M.

**Euclid Beamlabs LLC**

Bolingbrook, USA

- Jing, C.-J.

**Euclid TechLabs, LLC**

Solon, Ohio, USA

- Avrakhov, P.V.
- Kostin, R.A.
- Liu, A.

**Farhangian University**

Tehran, Iran

- Alimohamadi, M.

**Fermilab**

Batavia, Illinois, USA

- Burov, A.V.
- Johnstone, C.
- Khabiboulline, T.N.
- Lebedev, V.A.
- Lunin, A.
- Nagaitsev, S.
- Solyak, N.
- Stratakis, D.
- Sukhanov, A.I.
- Syphers, M.J.
- Valishev, A.
- Yakovlev, V.P.

**FRIB**

East Lansing, USA

- Hao, Y.
- Maruta, T.
- Ostroumov, P.N.
- Plastun, A.S.
- Villari, A.C.C.
- Yoshimoto, T.
- Zhang, T.
- Zhao, Q.

**Friedrich-Alexander Universität Erlangen-Nuernberg, University Erlangen-Nuernberg LFTE**

Erlangen, Germany

- Illmer, J.
- Mittelbach, A.K.

**FZJ**

Jülich, Germany

- Böker, J.
- Büscher, M.
- Engels, R.

- Engin, I.
- Gebel, R.
- Hützen, A.
- Lehrach, A.

#### **GSI**

Darmstadt, Germany

- Appel, S.
- Boine-Frankenheim, O.
- Chetvertkova, V.
- Geithner, W.
- Kornilov, V.
- Reimann, S.
- Sapinski, M.
- Singh, R.
- Sorge, S.
- Vilsmeier, D.M.

#### **HHU**

Duesseldorf, Germany

- Huggins, A.M.

#### **HHUD**

Dusseldorf, Germany

- Pukhov, A.M.
- Thomas, J.

#### **HIT**

Heidelberg, Germany

- Gläßle, T.G.

#### **Huazhong University of Science and Technology, State Key Laboratory of Advanced Electromagnetic Engineering and Technology,**

Hubei, People's Republic of China

- Han, W.J.

#### **HUST**

Wuhan, People's Republic of China

- Chen, Q.S.
- Feng, G.
- Liu, X.
- Qin, B.

#### **HZB**

Berlin, Germany

- Flisgen, T.
- Meseck, A.
- Vélez, A.V.
- Volz, P.I.

#### **IBA**

Louvain-la-Neuve, Belgium

- Collignon, V.
- Flandroy, Q.

- Forton, E.
- Hernalsteens, C.
- Herregods, B.
- Kleeven, W.J.G.M.
- Mandrillon, J.
- Nuttens, V.
- van de Walle, J.
- Van Der Kraaij, E.

#### **IBM T. J. Watson Center**

Yorktown Heights, New York, USA

- Tromp, R.M.

#### **IIT**

Chicago, Illinois, USA

- Neveu, N.R.

#### **Illinois Institute of Technology**

Chicago, Illinois, USA

- Snopok, P.
- Spentzouris, L.K.

#### **Indian Institute of Technology Madras**

Chennai, India

- Singh, S.

#### **INFN/LNF**

Frascati (Roma), Italy

- Marocchino, A.
- Spadaro, T.

#### **JAI**

Egham, Surrey, United Kingdom

- Abramov, A.
- Boogert, S.T.
- Garcia Morales, H.
- Gibson, S.M.
- Nevay, L.J.
- Pikhartova, H.
- Shields, W.
- Walker, S.D.

#### **JLab**

Newport News, Virginia, USA

- Roblin, Y.
- Zhang, H.
- Zhang, Y.

#### **LANL**

Los Alamos, New Mexico, USA

- Simakov, E.I.

**LBNL**

Berkeley, USA

- Brouwer, L.N.
- Hwang, K.
- Lehé, R.
- Mitchell, C.E.
- Qiang, J.
- Ryne, R.D.
- Van Beeumen, R.
- Vay, J.-L.
- Venturini, M.

**Leidos Corp**

Billerica, MA, USA

- Jensen, A.
- Petillo, J.J.

**LLNL**

Livermore, California, USA

- Ricketson, L.F.

**LLR**

Palaiseau, France

- Beck, A.
- Cauchois, A.
- Khojayan, M.
- Massimo, F.
- Prudent, J.
- Specka, A.
- Zenzemi, I.

**LULI**

Palaiseau, France

- Grech, M.
- Pérez, F.

**Maison de la Simulation, CEA**

Gif-sur-Yvette, France

- Derouillat, J.

**MSU**

East Lansing, Michigan, USA

- Berz, M.
- Duxbury, P.M.
- Makino, K.
- Tarazona, D.
- Valetov, E.V.
- Zerbe, B.S.

**Northern Illinois University**

DeKalb, Illinois, USA

- Erdelyi, B.
- Syphers, M.J.

**NTHU**

Hsinchu, Taiwan

- Huang, Y.-C.

**ORNL**

Oak Ridge, Tennessee, USA

- Aleksandrov, A.V.
- Cousineau, S.M.
- Han, B.
- Murray, S.N.
- Pennisi, T.R.
- Peplov, V.V.
- Shishlo, A.P.
- Stockli, M.P.
- Welton, R.F.
- Zhukov, A.P.

**ORNL RAD**

Oak Ridge, Tennessee, USA

- Cathey, B.L.
- Saethre, R.B.
- Stinson, C.
- Zhang, Z.L.

**ParaTools, Inc.**

Eugene, Oregon, USA

- Beekman, I.B.

**PSI**

Villigen PSI, Switzerland

- Acharya, D.
- Adelman, A.
- Bettoni, S.
- Calvi, M.
- Dehler, M.M.
- Ferrari, E.
- Frei, F.
- Frey, M.
- Hauenstein, D.
- Hermann, B.
- Hiller, N.
- Ischebeck, R.
- Lombosi, C.
- Prat, E.
- Reiche, S.
- Rivkin, L.
- Snuverink, J.

**Purdue University**

West Lafayette, Indiana, USA

- Lee, Y.J.
- Qi, M.

**RadiaSoft LLC**

Boulder, Colorado, USA

- Abell, D.T.
- Bruhwiler, D.L.

- Carlsson, J.A.
- Cook, N.M.
- Edelen, J.P.
- Eidelman, Y.I.
- Goldring, N.B.
- Hall, C.C.
- Keilman, M.V.
- Moeller, P.
- Nagler, R.
- Nash, B.
- Pogorelov, I.V.
- Webb, S.D.

**Rostock University, Faculty of Computer Science and Electrical Engineering**

Rostock, Germany

- Bahls, C.R.
- Gorgi Zadeh, S.
- Heller, J.
- Pommerenke, H.W.
- van Rienen, U.
- Zadeh, G.
- Zheng, D.

**SAGA**

Tosu, Japan

- Iwasaki, Y.

**ShanghaiTech University**

Shanghai, People's Republic of China

- Wan, W.

**SINAP**

Shanghai, People's Republic of China

- Chen, S.
- Deng, H.X.
- Feng, C.
- Gu, Q.
- Huang, N. S.
- Li, K.
- Liu, B.
- Wang, D.
- Wang, Z.
- Zhang, M.
- Zhao, Z.T.
- Zhou, K.S.

**SLAC**

Menlo Park, California, USA

- Edelen, A.L.
- England, R.J.
- Ge, L.
- Li, Z.
- Neilson, J.
- Ng, C.-K.
- Nuhn, H.-D.
- Ratner, D.F.
- Xiao, L.

**Sourcery Institute**

Oakland, California, USA

- Rouson, D.W.I.

**Stanford University**

Stanford, California, USA

- Black, D.S.
- Broadus, P. N.
- Byer, R.L.
- Ceballos, A.C.
- Deng, H.
- Fan, S.
- Harris, J.S.
- Hirano, T.
- Huang, Z.
- Hughes, T.W.
- Jiang, Y.
- Langenstein, T.
- Leedle, K.J.
- Miao, Y.
- Pigott, A.
- Sapra, N.
- Solgaard, O.
- Su, L.
- Tan, S.
- Vuckovic, J.
- Yang, K.
- Zhao, Z.

**Stony Brook University**

Stony Brook, USA

- Litvinenko, V.

**Tech-X**

Boulder, Colorado, USA

- Cowan, B.M.

**TEMF, TU Darmstadt**

Darmstadt, Germany

- Ackermann, W.
- Bizzozero, D. A.
- Boine-Frankenheim, O.
- Corno, J.
- De Gersem, H.
- Egenolf, T.
- Georg, N. G.
- Gjonaj, E.
- Niedermayer, U.
- Pham-Xuan, V.
- Römer, U.
- Schmid, S. A.
- Schöps, S.
- Skär, E.

**Texas A&M University**

College Station, USA

- Gerity, J.
- McIntyre, P.M.

#### **TRIUMF**

Vancouver, Canada

- Jung, P. M.
- Planche, T.

#### **UCF**

Orlando, USA

- Nunez-delPrado, A.

#### **UCL**

Louvain-la-Neuve, Belgium

- Jungers, R.
- Wang, Z.

#### **UCLA**

Los Angeles, USA

- Cesar, D.B.
- Musumeci, P.
- Naranjo, B.
- Rosenzweig, J.B.
- Shen, X.

#### **UIS**

Bucaramanga, Colombia

- Orozco, E.A.
- Otero Olarte, O.

#### **ULB - FSA - SMN**

Bruxelles, Belgium

- Dubus, A.
- Gnacadja, E.
- Hernalsteens, C.
- Pauly, N.
- Tesse, R.

#### **UMAG**

Santa Marta, Colombia

- Beltrán, J.R.
- González, J.D.
- Vergara, V.E.

#### **UMD**

College Park, Maryland, USA

- Beaudoin, B.L.
- Haber, I.
- Koeth, T.W.
- Matthew, D.B.
- Ruisard, K.J.

#### **University of Colorado at Boulder**

Boulder, USA

- Appelö, D.

#### **University of Crete**

Heraklion, Crete, Greece

- Rakitzis, T. P.
- Sofikitis, D.

#### **University of Erlangen-Nuremberg**

Erlangen, Germany

- Hommelhoff, P.
- Li, A.
- Schönenberger, N.
- Shiloh, R.
- Tafel, A.D.
- Yousefi, P.

#### **University of Hamburg, Institut für Experimentalphysik**

Hamburg, Germany

- Kuroпка, W.
- Mayet, F.

#### **University of Malta, Information and Communication Technology**

Msida, Malta

- Maclean, E.H.

#### **University of Oslo**

Oslo, Norway

- Sjobak, K.N.

#### **University of Rostock**

Rostock, Germany

- van Rienen, U.

#### **Universität Bern, Institute of Applied Physics**

Bern, Switzerland

- Feurer, T.

#### **UNM**

Albuquerque, New Mexico, USA

- Beznosov, O.
- Ellison, J.A.
- Heinemann, K.A.