

Machine Learning Applications for Performance Improvement and Developing Future Storage Ring Light Sources

Simon C. Leemann

ALS Accelerator Physics, ALS & ATAP Divisions, Lawrence Berkeley National Laboratory

67th ICFA Advanced Beam Dynamics Workshop on Future Light Sources: FLS 2023, Lucerne, Switzerland, Aug 27 – Sep 1, 2023

ACCELERATOR TECHNOLOGY& A TAPPLIED PHYSICS DIVISION







2 Recent Examples for Application of ML to Storage Ring Light Sources

Improving Operational Performance

PHYSICAL REVIEW LETTERS 123, 194801 (2019)

Featured in Physics

Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources

S.C. Leemann Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

S. Liu Department of Chemistry, University of California, Berkeley, California 94720, USA

A. Hexemer, M. A. Marcus, C. N. Melton, H. Nishimura, and C. Sun Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

(Received 16 May 2019; revised manuscript received 23 August 2019; published 6 November 2019

Synchrotron light sources, arguably among the most powerful tools of modern scientific discovery, are presently undergoing a major transformation to provide orders of magnitude higher brightness and transverse coherence enabling the most demanding experiments. In these experiments, overall source stability will soon be limited by achievable levels of electron beam size stability, presently on the order of several microns, which is still 1-2 orders of magnitude larger than already demonstrated stability of source position and current. Until now source size stabilization has been achieved through corrections based on a combination of static predetermined physics models and lengthy calibration measurements, periodically repeated to counteract drift in the accelerator and instrumentation. We now demonstrate for the first time how the application of machine learning allows for a physics- and model-independent stabilization of source size relying only on previously existing instrumentation. Such feed-forward correction based on a neural network that can be continuously online retrained achieves source size stability as low as 0.2 µm (0.4%) rms, which results in overall source stability approaching the subpercent noise floor of the most sensitive experiments.

DOI: 10.1103/PhysRevLett.123.194801

maps with unprecedented resolution and will lead to

deeper understanding of electrochemical systems such as

batteries and fuel cells. The measurement of dynamics

and kinetics to study chemical systems is another cat-

egory that will be directly impacted by the new sources.

An emerging technique to study this is x-ray photon

0031-9007/19/123(19)/194801(6)

correlation spectroscopy (XPCS) [10]. Ptychography as

Introduction.-Synchrotron radiation sources, specifi- well as XPCS rely heavily on high beam stability over cally third-generation storage ring light sources, have extended periods of time.

been tremendously successful tools of scientific discovery since the early 1990s [1]. As these facilities mature, a lies in their stability, resulting in constant position, angle new era of fourth-generation storage rings (4GSRs) based and intensity of radiation delivered at a tunable wavelength on diffraction-limited storage rings (DLSRs) [2-8] is with narrow width. In order to maintain constant intensity being ushered in. These sources will increase average a combination of top-off injection (maintaining constant brightness by 2-3 orders of magnitude while also delivbeam current) [11,12] and precise control over source ering high degrees of transverse coherence, for the first position and size is required. In third-generation light sources time even for x rays. High coherent flux will enable (3GLSs) the latter usually called for transverse beam size scientists to understand material compositions and stability within 10% of the rms electron beam size [13,14] dynamics ranging in length from microns to nanometers Now, however, first experiments at these sources are starting and in time from minutes to nanoseconds. The most to show limitations arising from such levels of source size notable and direct result of the new beam properties will control and it is evident that DLSRs, operating at much impact two techniques in particular. Ptychography [9] smaller source sizes, will call for significantly tighter control will take direct advantage of an increase in coherent flux over source size stability in order to exploit ultrahigh to decrease measurement times by orders of magnitude. brightness and transverse coherence. This will allow for the collection of complex 3D chemical State-of-the-art stabilization effort and its limitations -

To large extent the success of storage ring light sources

A typical example for the aforementioned source size

stabilization challenge is shown in Fig. 1. The vertical

electron beam size as measured at diagnostic beam line

3.1 [15] of Lawrence Berkelev National Laboratory's

Advanced Light Source (ALS) is displayed during a typical

user run. While the horizontal beam size remains constant

(spikes observed in both planes at the same time are

PRL 123, 194801 (2019)

Designing Future Storage Rings

Nuclear Inst. and Methods in Physics Research, A 1050 (2023) 16819



Full Length Article

Demonstration of machine learning-enhanced multi-objective optimization of ultrahigh-brightness lattices for 4th-generation synchrotron light sources

Y. Lu^a, S.C. Leemann^{a,*}, C. Sun^a, M.P. Ehrlichman^a, H. Nishimura^a, M. Venturini^a, T. Hellert^a

^a Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA ^b Deutsches Elektronen Synchrotron DESY, Notkestraße 85, 22607, Hambarg, Germany

ARTICLE INFO ABSTRACT

Storage ring Beam dynamics Lattice design

1. Introduction

* Corresponding author.

Available online 8 March 2023

E-mail address: SCLeemann@lbl.gov (S.C. Leemann).

Received 7 December 2022; Received in revised form 21 February 2023; Accepted 6 March 2023

0168-9002/Published by Elsevier B.V. This is an open access article under the CC BY license (http://

//doi.org/10.1016/j.nima.2023.16810

Fourth-generation storage rings enabled by multi-bend achromat lattices are being inaugurated worldwide and many more are planned for the next decade. These sources deliver stable ultra-high brightness radiation with unmatched levels of transverse coherence by virtue of their highly advanced magnetic lattices. Optimization of these challenging and strongly nonlinear lattices with many degrees of freedom bounded by extensive sets of constraints and multiple often conflicting optimization goals is highly demanding and requires application of the most advanced numerical tools available to the community. While multi-objective genetic algorithm ave been very successful in supporting these optimization efforts, the algorithms suffer from a fun limitation of their stochastic nature: an exceedingly yast number of candidate lattices, most of which eventual are rejected, has to be fully evaluated. This comes at immense computational cost and thus drives excessiv are rejected, has to be fully evaluated. This comes at minimize computational cost and three excession runtime despite use of large supercomputing clusters. We therefore propose to employ deep learning techniques and iterative retraining of neural networks to massively accelerate such lattice evaluation, thereby allowing lattice optimization to rely on far fewer a priori assumptions, open up to larger search ranges, and include righ rom the start and in parallel multiple error distributions to find truly global optima, all while completing a full optimization campaign in weeks rather than months. In this paper we present the neural network designs the deep learning approach, iterative retraining procedures, and demonstrate how these machine learning techniques can be incorporated into existing state-of-the-art optimization workflows with only minimal changes applied to the optimization pipeline itself and none at all to the employed tracking codes.

number of magnets that need to be tuned in a multi-variate and multiobjective optimization process. Apart from lattice design expertise, this Storage-ring based synchrotron light sources around the world are usually calls for the most advanced numerical and analytical resources presently undergoing a massive transformation. Pioneered in MAX available to the communit IV [1], the multi-bend achromat (MBA) [2] lattice has ushered in the Multi-objective genetic algorithms (MOGA) [3] have proven to be

one of the most successful and commonly used tools for the optimization of modern light source lattices [4-6]. Multiple variants of MOGA are available, among which the Pareto-based algorithm NSGA-II is the most popular [7,8]. Optimization of an MBA lattice with MOGA is highly non-trivial since ultra-high brightness, lifetime, and injection efficiency are usually in direct competition and a suitable trade-off needs to be carefully established, taking into account an exceedingly large number of constraints. While MOGA is extremely well equipped to undertake such optimization, it suffers from the fundamental limitation that-as a stochastic process-it requires a vast number of candidate lattices to be evaluated. Nonlinear lattice evaluation based on many turn particle tracking is very CPU-expensive and nevertheless, almost

era of 4th-generation storage rings (4GSRs); a class of ring-based light sources capable of delivering stable ultra-high brightness diffractionlimited synchrotron radiation with a high degree of transverse co-

herence simultaneously to dozens of beamlines. The MBA latticepresently foreseen by almost every new source and upgrade project-is composed of many small-aperture magnets with high field gradients capable of providing the strong focusing necessary to achieve ultralow emittance. This strong focusing reduces the dispersion and drives the natural chromaticity in the lattice. Combined, this calls for very strong sextupoles leading to highly nonlinear lattices exhibiting limited dynamic aperture (DA) and momentum aperture (MA) compared to those of 3rd-generation light sources. Apart from the many engineering difficulties in the design of a 4GSR, the beam physics and lattice optimization itself present a significant challenge due to the large all evaluated lattices are eventually rejected by MOGA. This weakness



Simon C. Leemann • Machine Learning Applications for Storage Ring Light Sources FLS 2023, Lucerne, Switzerland, Aug 27 – Sep 1, 2023



Part 1: ML Improving Performance of an Operational Storage Ring

 State-of-the-art storage ring light sources achieve excellent stability in terms of beam current (top-off injection) & beam position/angle (orbit feedbacks)







- State-of-the-art storage ring light sources achieve excellent stability in terms of beam current (top-off injection) & beam position/angle (orbit feedbacks)
- ID effects are countered with extensive correction efforts
- **ID feed forwards** rely on look-up tables that require **dedicated machine time** and periodic re-recording

EPU phase/shift [mm]

EPAC 2004, MOPKF071, p.479







20

-15

-10

- State-of-the-art storage ring light sources achieve excellent stability in terms of beam current (top-off injection) & beam position/angle (orbit feedbacks)
- ID effects are countered with extensive correction efforts
- ID feed forwards rely on look-up tables that require dedicated machine time and periodic re-recording
- Yet in spite of all these correction efforts, **beam size** is still perturbed by insertion device (ID) config changes







- State-of-the-art storage ring light sources achieve excellent stability in terms of beam current (top-off injection) & beam position/angle (orbit feedbacks)
- ID effects are countered with extensive correction efforts
- ID feed forwards rely on look-up tables that require dedicated machine time and periodic re-recording
- Yet in spite of all these correction efforts, **beam size** is still perturbed by insertion device (ID) config changes
- Resulting level of performance has started to become a **limitation** at most demanding experiments
- Expected to become a serious issue in 4th-generation sources, eg. **DLSRs** with **STXM**, **XPCS**, **ptychography**, etc.







Machine Learning to the Rescue

- Machine Learning (ML) can can model highly nonlinear processes, is extremely flexible
- ML can exploit large amounts of data that are already collected during routine operations → "training"





Machine Learning to the Rescue

- Machine Learning (ML) can can model highly nonlinear processes, is extremely flexible
- ML can exploit large amounts of data that are already collected during routine operations → "training"
- Once trained, neural network (NN) provides predictions for vertical beam size changes resulting from ID config changes & magnetic corrections
- Magnetic corrections implemented as excitation change to the 32 skew quadrupoles driving the vertical dispersion wave









Machine Learning to the Rescue

- Machine Learning (ML) can can model highly nonlinear processes, is extremely flexible
- ML can exploit large amounts of data that are already collected during routine operations → "training"
- Once trained, neural network (NN) provides predictions for vertical beam size changes resulting from ID config changes & magnetic corrections
- Magnetic corrections implemented as excitation change to the 32 skew quadrupoles driving the vertical dispersion wave
- These NN predictions can serve as a dynamic lookup
- If such a lookup is incorporated into the accelerator control system as a feed forward (FF), we can stabilize the electron beam source size over prolonged periods of time (online retraining exploited to mitigate machine drift)











Simulating NN-based ID FF During Physics Shift

- Scanned various ID configurations and skew excitations (DWP) to record initial training data
- Training fully-connected
 3-layer NN (128-64-32)
 required ≈15min on
 single core
- NN-based ID FF turned on while continuing to scan ID configurations → verify stabilization





NN-based ID FF Off vs. On During User Ops

- During user ops NN-based ID FF running at ≈3 Hz in addition to of all other conventional FBs and FFs
- Observed roughly 5-fold reduction of V rms source size motion at diagnostics BL
- Online retraining of the NN (using data acquired with FF engaged during user ops) → capture ID configurations not observed during initial training





Stabilization Confirmed at Most Sensitive Experiment

STXM @ ALS Beamline 5.3.2.2



ATAP





Part 2: ML Improving Design of Future Storage Ring Light Sources

Introduction: The Problem

• **4th-generation storage rings** (4GSRs) leverage MBA lattices to render ultra-high brightness with large coherent fraction









Introduction: The Problem

- **4th-generation storage rings** (4GSRs) leverage MBA lattices to render ultra-high brightness with large coherent fraction
- MBA lattices are very challenging: dense & exploit very strong focusing → drives strong chromatic & higher-order corrections
- Solutions often highly nonlinear & involve many degrees of freedom (DoF) → demanding optimization







Introduction: The Problem

- **4th-generation storage rings** (4GSRs) leverage MBA lattices to render ultra-high brightness with large coherent fraction
- MBA lattices are very challenging: dense & exploit very strong focusing → drives strong chromatic & higher-order corrections
- Solutions often highly nonlinear & involve many degrees of freedom (DoF) → demanding optimization
- Multi-objective genetic algorithms (MOGA) are highly successful at such optimization & have become tool of choice
- However, stochastic nature is inherent weakness
- Do not want to artificially limit DoF, search ranges, or make many initial assumptions about attractive solutions → so what can we do?







Improving MOGA: ML to the Rescue

- ML can be employed to render neural networks (NNs) → surrogate models used in lieu of computationally expensive evaluation
- Lattice candidate evaluation becomes near instantaneous
- Aim to speed up MOGA without modifying MOGA/tracking tools or existing workflows & <u>without</u> sacrificing **physics fidelity**
- Previous attempts [1-3] have focused on various aspects, but we set out with a different emphasis:
 - Direct optimization of relevant physics quantities (ϵ_0 , DA, MA)
 - Combined linear/nonlinear optimization involving all free parameters (quadrupoles & sextupoles)

M. Kranjčević, B. Riemann, A. Adelmann, A. Streun, PRAB **24** 014601, 2021.
 Y. Li, W. Cheng, L. Yu, R. Rainer, PRAB **21** 054601, 2018.
 J. Wan, P. Chu, Y. Jiao, PRAB **23** 081601, 2020.







ALS-U as a Test Case

- ALS-U storage ring (SR) calls for a challenging 9BA to achieve ≈75 pm rad (round beam) at 2 GeV in 200-m tunnel
 → diffraction limited @ 1.2 keV (1 nm)
- 9BA lattice becomes very dense & has strained optics



ATAP







ALS-U as a Test Case

- ALS-U storage ring (SR) calls for a challenging 9BA to achieve ≈75 pm rad (round beam) at 2 GeV in 200-m tunnel
 → diffraction limited @ 1.2 keV (1 nm)
- 9BA lattice becomes very dense & has strained optics
- MOGA (@ 2nd stage): 9 quads, 4 sextupoles → **11 free knobs**
- ≈15 magnet/lattice constraints on top of quadrupole ranges (from 1st stage)
- **Objectives:** ϵ_0 , MA, and on-momentum DA (total diff. rate)

Natural emittance	$\varepsilon_0 < 155 \mathrm{pmrad}$
Maximum beta	$\beta_{x,y} < 30 \mathrm{m}$
Maximum dispersion	$\eta_x < 15 \mathrm{cm}$
Fractional tunes	$0.1 < \nu_{x,y} < 0.4$
Dispersion at center of straight	$ \eta_x^* < 1\mathrm{mm}$
Beta at center of straight	$1\mathrm{m} < \beta^*_{x,y} < 5\mathrm{m}$
Beta in central arc bends (B3)	$\beta_{x,y}^{B3} < 4 \mathrm{m}$
Fractional tune difference	$ \nu_x - \nu_y < 0.01$
Chromatic sextupole strength (SF, SD)	$b_2 < 900 \mathrm{m}^{-3}$







ALS-U as a Test Case

- ALS-U storage ring (SR) calls for a challenging 9BA to achieve ≈75 pm rad (round beam) at 2 GeV in 200-m tunnel
 → diffraction limited @ 1.2 keV (1 nm)
- 9BA lattice becomes very dense & has strained optics
- MOGA (@ 2nd stage): 9 quads, 4 sextupoles → **11 free knobs**
- ≈15 magnet/lattice constraints on top of quadrupole ranges (from 1st stage)
- **Objectives:** ϵ_0 , MA, and on-momentum DA (total diff. rate)
- Highly staged **MOGA** approach resulted in
 - ±1 mm DA (on-axis swap-out injection with AR)
 - ≈ 1 hr lifetime (with 3HCs)
- ...but required months of CPU time on large clusters



Courtesy: Changchun Sun





ML for Full Linear & Nonlinear Lattice Optimization

- Training data for 11D problem can no longer be acquired through systematic sampling of input space...
- …but do not want to make too many assumptions → retain generality of approach
- Instead: use first few generations of **MOGA data** as training data for **deep neural networks (DNNs)**
- Two **8-layer DNNs** used in lieu of MOGA calls to TRACY for DA and MA



Fully-connected (FC) NN, using ReLU as activation function, # = node depth







ML for Full Linear & Nonlinear Lattice Optimization

- Training data for 11D problem can no longer be acquired through systematic sampling of input space...
- …but do not want to make too many assumptions → retain generality of approach
- Instead: use first few generations of **MOGA data** as training data for **deep neural networks (DNNs)**
- Two **8-layer DNNs** used in lieu of MOGA calls to TRACY for DA and MA
- Training 2 DNNs to get DA/MA predictions ≈1% rms requires about 50k lattices
- Compare: traditional MOGA requires about 640 gen
 (3.2M lattices evaluated) → 8 days on 1000-core cluster



Fully-connected (FC) NN, using ReLU as activation function, # = node depth







But of course it's a bit more complicated...

- ML predictions are not 100% accurate
- Training based on initial data only









But of course it's a bit more complicated...

- ML predictions are not 100% accurate
- Training based on initial data only
- Initial ML-MOGA solutions disagree with tracking validation & converged ML-predicted solutions not entirely non-dominated







But of course it's a bit more complicated...

- ML predictions are not 100% accurate
- Training based on initial data only
- Initial ML-MOGA solutions disagree with tracking validation & converged ML-predicted solutions not entirely non-dominated
- Want to **retrain DNNs** with an improved resampling of input space as in [5], ...
- ...but here for a many-dimensional input space without making <u>any</u> assumptions on smoothness of distributions

[5] A. Edelen, N. Neveu, M. Frey, et al., PRAB 23 044601, 2020.







Iterative Retraining Improves ML-MOGA

- Retraining DNNs with tracking validation data is computationally inexpensive & makes <u>no</u> assumptions on distributions
- Retrained DNN is used for next run starting with inputs from final gen of last run → Iterate this ML-validation-retraining process until ML-MOGA results reach the true Pareto-optimal front
- But when is that?
 - How do we know our predictions have become accurate enough and our ML-MOGA derived Pareto front is the actual Pareto front?
 - Minimizing no. of additional required iterations is crucial to maintaining **large overall speedup**







Distance Metrics & Convergence

- Introduce two distance metrics for input & objective space
- Euclidean norms normalized in each variable → single unit-free relative measure for movement of distribution in input/objective space
- Metrics inform us about:
 - MOGA can be considered truly converged once $\delta_{i,o}(m+1) \approx \delta_{i,o}(m)$
 - when there is no more added benefit from an additional retraining iteration, i.e. process fully converged once $\Delta_f \to 0$
- Model-independent metrics → full automation



Objective Space







Distance Metrics & Convergence (cont.)







ML-MOGA 1

ML-MOGA 2

Converged



• Retraining shows very **quick convergence** (6-8 iterations)







Results

- Retraining shows very quick convergence (6-8 iterations)
- Once fully converged, ML-MOGA inputs & objectives match those of traditional MOGA to within "noise floor" (MOGA stochastics)
- Overall speedup is roughly a factor 40 (incl. training & re-training effort)
- Only very minor modifications required to existing MOGA workflow/tools
- Convergence defined in model-independent way → can adapt to other optimization problems
- Potential to fully automate entire optimization campaign & optimize in parallel from the start for many error seeds is highly attractive → derive truly *global* optimum







Thank You!

Questions?

Acknowledgments: Shuai Liu, Nathan Melton, Yuping Lu, Hiroshi Nishimura, Changchun Sun, Matthew Marcus, David Shapiro, Alex Hexemer, Dani Ushizima, Mike Ehrlichman, Gregg Penn, Thorsten Hellert, Erik Wallen, Warren Byrne, Fernando Sannibale, Marco Venturini, Andrea Pollastro, Andreas Scholl, Rob Ryne, DOE Office of Science (BES ADRP & ASCR) Contract No. DEAC02-05CH11231







