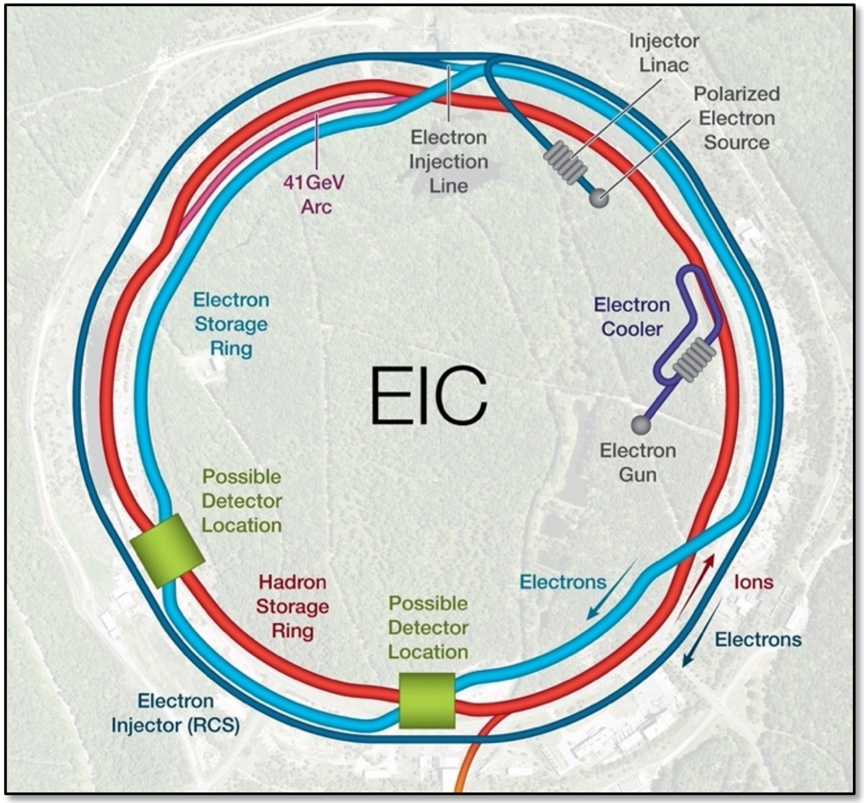
AI4EIC: AI/ML Thoughts on Accelerator Design/Modeling

- Staging the problem
  - No time for a full design lecture https://tinyurl.com/4z92u6hc
  - MOGA component optimization
- Strategy
  - Snowmass community white paper
    - Recommendations
  - Virtual twins in accelerator design
- Hot Off The Presses
  - Algorithmic improvements (even linear)
  - Nonlinear parameterization

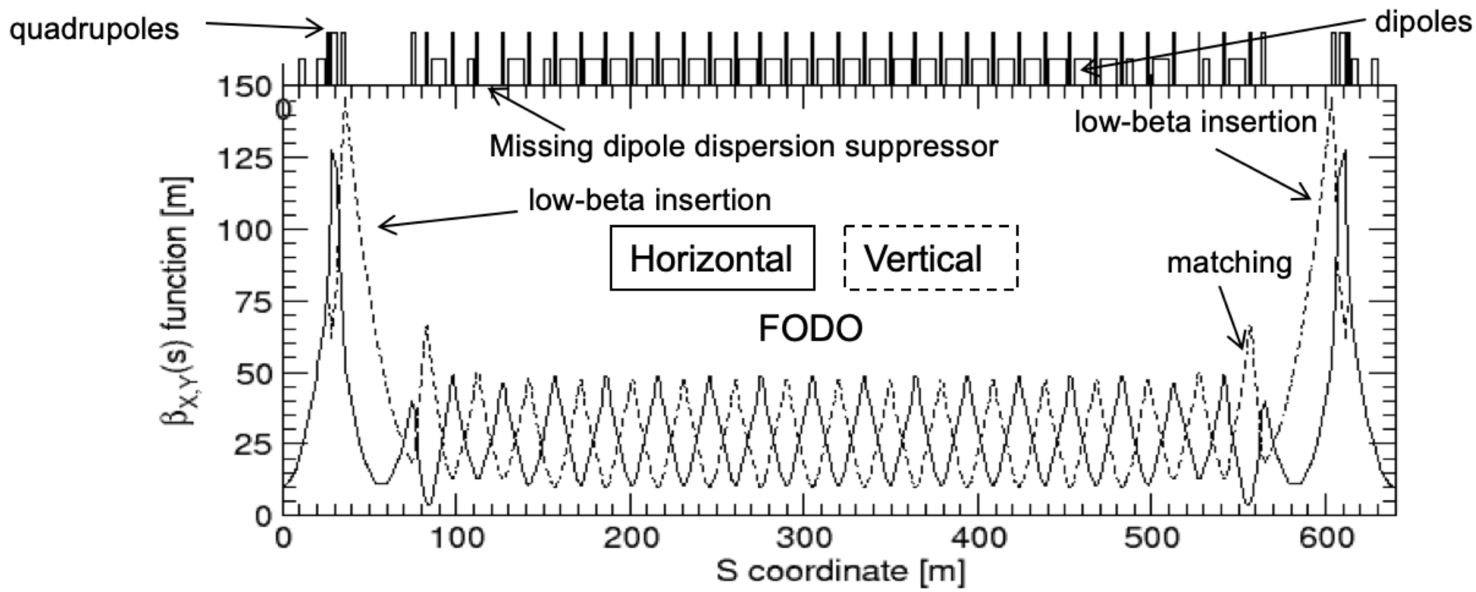


T. Satogata, Jefferson Lab

October 10, 2022

Caveat: Project is between CD-1/CD-2
Do NOT take any parameters shown here as gospel!

Staging the Problem: One-Sixth of RHIC (EIC HSR)



- Linear design "straightforward"
  - Then perturbative, "order-by-order" design
- Challenges driven by technology/engineering, cost/performance optimization/tradeoffs
- Many (nonlinearly) coupled parameters

- Why?
  - **Human managable** (complexity, risks)
  - **Human operator tunable**
  - Robust to small (technical) perturbations

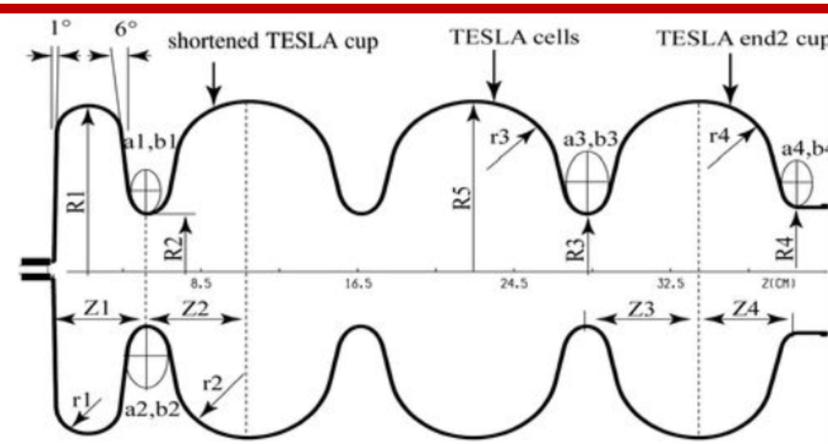
Jefferson Lab

Staging the Problem: Accelerator Physics Design Challenges

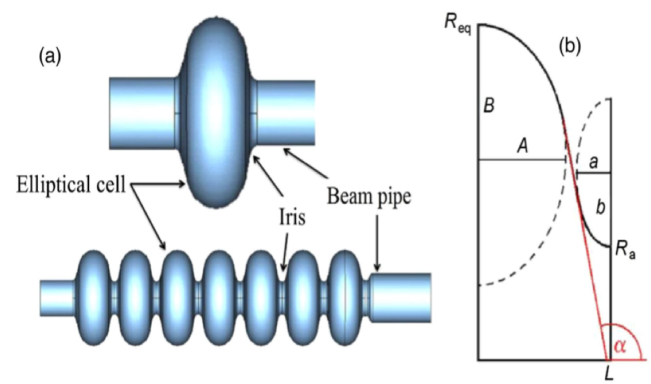
- Computer optimization methods have been used in accelerator physics "forever"
  - Modifications of iterative linear (and perturbative) methods
    - Orbit correction, MICADO, linear SVD methods; Optics parameter iterative design matching
  - Late 80's, early 90's: neural networks for control
    - e.g. orbit correction (Bozoki, Parseh, et al.)
- **Applications in AI/ML in accelerator physics has been developing over many years**
  - Complex adaptive control/operations, large data set fault classification
  - MOGA techniques started earlier for device design optimization
    - e.g. SRF electron guns, injector design optimization
- **Some challenges to AI/ML adoption in accelerator design space**
  - Simulations are **computationally complex**: o(10^{8-11}) particles for o(10^{6-10}) turns, fidelity/stability
    - Speed/scale may be improved o(x10) with symplectified maps (e.g. SLAC PEP-II in 1990s)
  - High intensity Coulomb interactions with **everything**: self-, colliding beam, environment
  - Most beam dynamics problems are **nonlinear** where they are interesting
  - Many cost/schedule/performance/technical tradeoffs: **"art of design"** and risk management
  - Many tradeoffs (field quality, construction tolerances/errors vs costs, R&D time, capability)
  - Environmental dynamical factors (ground motion/microphonics, temperatures, power stability)
- All of these challenges are applicable to future accelerator machine design (EIC?)

Staging the Problem: Component Design (SRF gun)

- **MOGA** is regularly used for **accelerator component design optimization**
  - Few dozen parameters (or fewer)
  - Pareto front optimization tradeoffs
  - Incorporate (some) engineering constraints
- Reflects design sim maturity/robustness
  - Translate input parameters to relevant optimization quantities
  - Consistent smoothness, convergence
  - Still require human intervention near **parametric singularities**
- ... but with simplifying assumptions
  - **Fidelity/scope/performance tradeoffs**
  - **MANY different codes/approaches**



ELBE SRF Photogun
1.3 GHz, 10 kW



Optimization of a traveling wave superconducting rf cavity for upgrading the International Linear Collider

V. Shemelin, H. Padamsee, and V. Yakovlev
FNAL, Batavia, Illinois 60510, USA

(Received 10 October 2021; accepted 24 January 2022; published 7 February 2022)

Also see A. Hofler
https://tinyurl.com/y6tsptk4

Advanced Accelerator Modeling and Simulation: Snowmass White Paper

- **Snowmass process**
  - 7-10 year community process
  - Sponsored by APS DPF
    - Large influence on accelerator R&D
  - Input to US HEPAP/P5
- A natural community focus for **accelerator modeling (design) priorities/strategies**
- Authorship included **many leaders of accelerator AI/ML community**
- Sets out recommendations for **next generation of accelerator modeling**
  - Likely the foundations of future advanced accelerator design approaches
  - Progress being made (NP/HEP/ASCR...)

Submitted to the Proceedings of the US Community Study
on the Future of Particle Physics (Snowmass 2021)

https://arxiv.org/abs/2203.08335

Snowmass21 Accelerator Modeling Community White Paper

by the Beam and Accelerator Modeling Interest Group (BAMIG)*

Authors (alphabetical): S. Biedron[13], L. Brouwer[1], D.L. Bruhwiler[7], N. M. Cook[7], A. L. Edelen[6], D. Filippetto[1], C.-K. Huang[9], A. Huebl[1], T. Katsouleas[15], N. Kuklev[4], R. Lehe[1], S. Lund[12], C. Messe[1], W. Mori[10], C.-K. Ng[6], D. Perez[9], P. Piot[4,5], J. Qiang[1], R. Roussel[6], D. Sagan[2], A. Sahai[11], A. Scheinker[9], E. Stern[14], M. Thévenet[16], F. Tsung[10], J.-L. Vay[1], D. Winklehner[8], and H. Zhang[3]

[1] *Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[2] *Cornell University, Ithaca, NY 14853, USA*
[3] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*
[4] *Argonne National Laboratory, Lemont, IL 60439, USA*
[5] *Northern Illinois University, DeKalb, IL 60115, USA*
[6] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[7] *RadiaSoft LLC, Boulder, CO 80301, USA*
[8] *Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*
[9] *Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[10] *University of California at Los Angeles, Los Angeles, CA 90095, USA*
[11] *University of Colorado Denver, Denver, CO 80204, USA*
[12] *Michigan State University, East Lansing, MI 48824, USA*
[13] *University of New Mexico, Albuquerque, NM 87106, USA*
[14] *Fermi National Accelerator Laboratory, Batavia, IL 60563, USA*
[15] *University of Connecticut, Storrs, CT 06269, USA*
[16] *Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany*

September 26, 2022

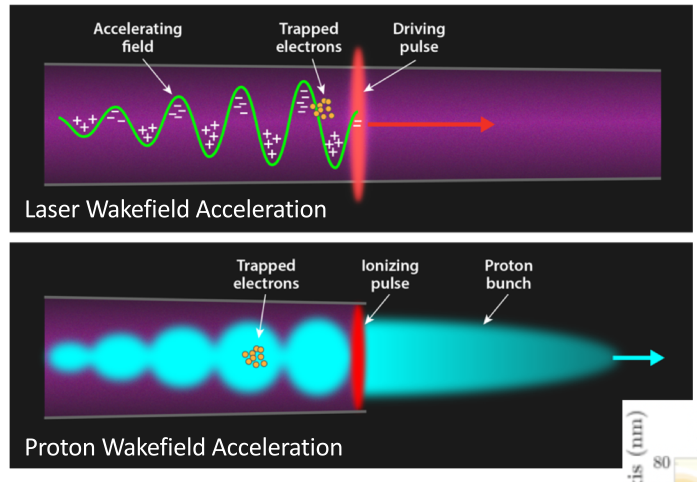
https://snowmass-compf2-accbeammodel.github.io

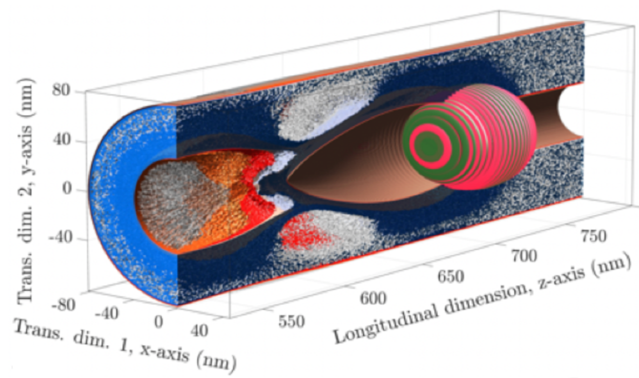
White Paper Recommendations

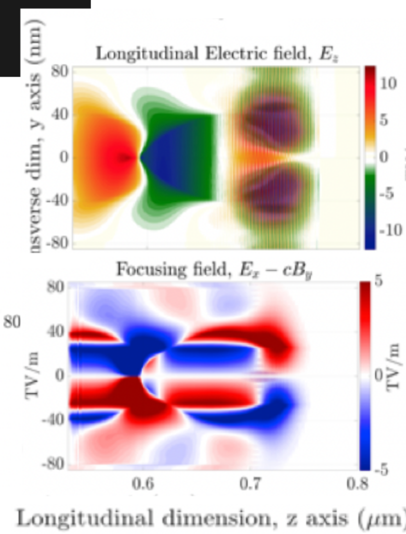
VERY nonlinear supercomputing cutting-edge examples

https://physics.aps.org/articles/v12/19

- Develop a comprehensive portfolio of particle accelerator and beam physics modeling tools in support of achieving Accelerator and Beam Physics Thrust Grand Challenges on **intensity**, quality, **control, and prediction**.
- Develop software infrastructure to enable **end-to-end virtual accelerator modeling and corresponding virtual twins** of particle accelerators.
- Develop **advanced algorithms and methods including AI/ML modalities and quantum computing technologies.**
- Develop **efficient and scalable software frameworks** and associated tools to effectively leverage next generation high performance and high-throughput computing hardware.
- Develop sustainable and reliable code maintenance practices, community benchmarking capabilities, and training opportunities to foster the cooperative application of accelerator software.
- Foster an open community that spans academia, national labs and industry to (a) develop software ecosystems, libraries, frameworks and standards, (b) curate data repositories, and (c) establish dedicated centers and distributed consortia with open governance models.







Nanoplasmonic channel acceleration
https://tinyurl.com/3mfzu7m4

Virtual/Digital Twins in Accelerator Design

- Community is “boostrapping” virtual/digital twins of many accelerators
  - Inexpensive training/testing for new AI/ML operations techniques
  - Natural extension of control room online modeling
  - (Also useful for operator training!)

- White paper:
  - These will enable the design and optimization of future accelerators at scale on supercomputers, with unprecedented levels of accuracy and speed. This capability will dramatically increase the **breadth of parameter space** that can be explored, thereby **enabling the design of particle accelerators that have not been possible before**.

DEVELOPING ROBUST DIGITAL TWINS AND REINFORCEMENT LEARNING FOR ACCELERATOR CONTROL SYSTEMS AT THE FERMILAB BOOSTER

D. Kafkes*, Fermi National Accelerator Laboratory, Batavia, USA
M. Schram, Thomas Jefferson National Accelerator Facility, Newport News, USA

https://doi.org/10.18429/JACoW-IPAC2021-TUPAB327

Abstract

We describe the offline machine learning (ML) development for an effort to precisely regulate the Gradient Magnet Power Supply (GMPS) at the Fermilab Booster accelerator complex via a Field-Programmable Gate Array (FPGA). As part of this effort, we created a digital twin of the Booster-GMPS control system by training a Long Short-Term Memory (LSTM) to capture its full dynamics. We outline the path we took to carefully validate our digital twin before deploying it as a reinforcement learning (RL) environment. Additionally, we demonstrate the use of a Deep Q-Network (DQN) policy model with the capability to regulate the GMPS against realistic time-varying perturbations.

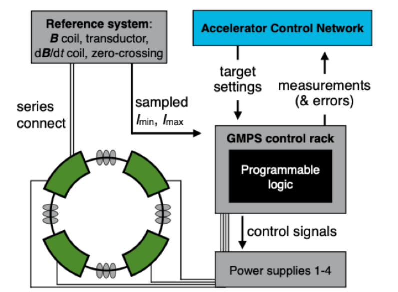


Figure 1: Overview of current PID-GMPS control system.

PHYSICAL REVIEW SPECIAL TOPICS - ACCELERATORS AND BEAMS **17,** 030701 (2014)

https://journals.aps.org/prab/pdf/10.1103/PhysRevSTAB.17.030701

Start-to-end simulation of x-ray radiation of a next generation light source using the real number of electrons

J. Qiang,* J. Corlett, C. E. Mitchell, C. F. Papadopoulos, G. Penn, M. Placidi, M. Reinsch, R. D. Ryne, F. Sannibale, C. Sun, and M. Venturini
Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

P. Emma
SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA

S. Reiche
Paul Scherrer Institute, 5232 Villigen PSI, Switzerland
(Received 12 December 2013; published 20 March 2014)

Jefferson Lab

Hot Off The Presses: Algorithm Improvements

- Community is paying close attention to fundamental algorithm AI/ML research
  - Some algorithm improvements will directly affect accelerator modeling bottlenecks
  - Even linear scaling advances improve computational reach, fidelity
- Recent advances even being made in computational linear algebra!
  - e.g. by AlphaTensor
  - “tensor decompositions within a finite factor space” parallels general control problem
    - accelerator physics design problem: decompose multidimensional observables to multidimensional knobs

Article

Discovering faster matrix multiplication algorithms with reinforcement learning

https://doi.org/10.1038/s41586-022-05172-4
Received: 2 October 2021
Accepted: 2 August 2022
Published online: 5 October 2022
Open access
Check for updates

Alhussein Fawzi[1,2,✉], Matej Balog[1,2], Aja Huang[1,2], Thomas Hubert[1,2], Bernardino Romera-Paredes[1,2], Mohammadamin Barekatain[1], Alexander Novikov[1], Francisco J. R. Ruiz[1], Julian Schrittwieser[1], Grzegorz Swirszcz[1], David Silver[1], Demis Hassabis[1] & Pushmeet Kohli[1]

Improving the efficiency of algorithms for fundamental computations can have a widespread impact, as it can affect the overall speed of a large amount of computations. Matrix multiplication is one such primitive task, occurring in many systems—from neural networks to scientific computing routines. The automatic discovery of algorithms using machine learning offers the prospect of reaching beyond human intuition and outperforming the current best human-designed algorithms. However, automating the algorithm discovery procedure is intricate, as the space of possible algorithms is enormous. Here we report a deep reinforcement learning approach based on AlphaZero[1] for discovering efficient and provably correct algorithms for the multiplication of arbitrary matrices. Our agent, AlphaTensor, is trained to play a single-player game where the objective is finding tensor decompositions within a finite factor space. AlphaTensor discovered algorithms that outperform the state-of-the-art complexity for many matrix sizes. Particularly relevant is the case of 4 × 4 matrices in a finite field, where AlphaTensor's algorithm improves on Strassen's two-level algorithm for the first time, to our knowledge, since its discovery 50 years ago[2]. We further showcase the flexibility of AlphaTensor through different use-cases: algorithms with state-of-the-art complexity for structured matrix multiplication and improved practical efficiency by optimizing matrix multiplication for runtime on specific hardware. Our results highlight AlphaTensor's ability to accelerate the process of algorithmic discovery on a range of problems, and to optimize for different criteria.

Hot Off The Presses: Chaotic System Parameterization

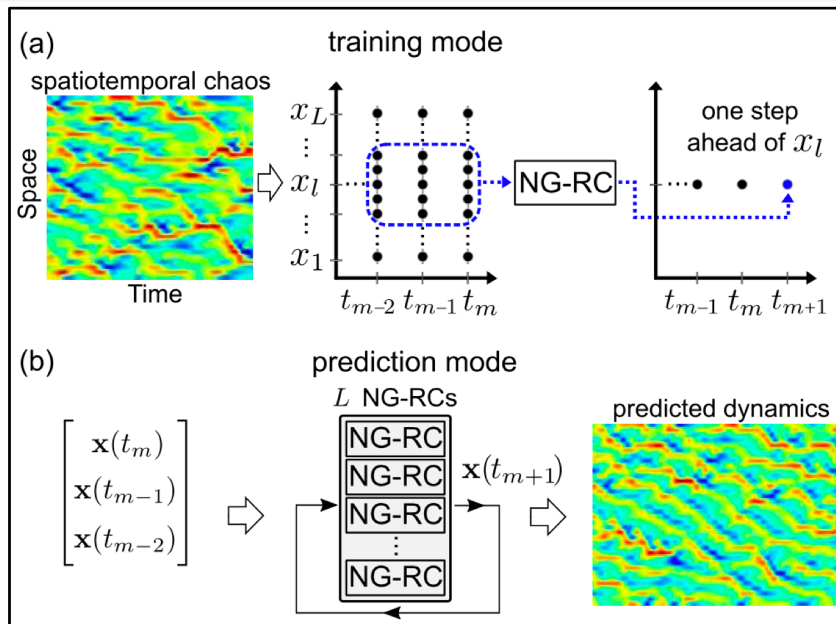
- Improvements in nonlinear/chaotic system forecasting
  - Surrogate model parameter reduction
  - Large improvements ($\sim 10^2$) in fidelity training speed
- Potential applicability to accelerator surrogate models
  - Especially interesting for initial nonlinear design (advantageous symmetries)
- Dimensionality reduction may also produce new accelerator tuning approaches
  - Near stability boundaries

Chaos

Learning spatiotemporal chaos using next-generation reservoir computing

Cite as: Chaos **32**, 093137 (2022); https://doi.org/10.1063/5.0098707
Submitted: 11 May 2022 • Accepted: 30 August 2022 • Published Online: 26 September 2022

Wendson A. S. Barbosa and Daniel J. Gauthier



Home | Registration | Agenda | Abstract Submission | For Participants | Contact Us

The deadline to submit an abstract is: **Wednesday, September 21, 2022**

Motivation

We are pleased to announce the 3rd ICFA Beam Dynamics Mini-Workshop on Machine Learning Applications for Particle Accelerators will be held in Chicago, Il, USA. This will be the third workshop in a series that began in 2018 at SLAC, CA, USA followed by a second workshop held at Villigen PSI, Switzerland in 2019. A third workshop had been planned to be held in Seoul, Korea in 2020, but unfortunately had to be canceled due to the COVID-19 pandemic.

The goal of this third workshop is to continue to work on building a world-wide community of researchers and engineers interested in applying artificial intelligence and machine learning technologies to particle accelerators.

The workshop will consist of four topics:

1. Tuning/optimization/control
2. Prognostics/alarm handling/anomaly-breakout detection
3. Data analysis
4. Simulations/modeling

54 registrants as of 10/10/22

Important Dates

August 26, 2022	Registration opens
November 4, 2022	General registration closes

Workshop Information

Dates: November 1–4, 2022

Event ID: 44128

Workshop Venue
Palmer House a Hilton Hotel
17 E Monroe Street
Chicago, IL 60603 USA

Meeting location and directions

Workshop Coordinator
Anna Petway

(631) 344-4776
(631) 344-5954
petway@bnl.gov

Closing Comments

- Accelerator design has been "traditional" for 60+ years
  - Modular, linear (and nonlinear perturbative) design
    - Nonlinear computational/conceptual complexity scales badly when improving broad fidelity
  - Driven by component technology advances, complexity reduction, risk aversion, operability
  - Modeling for design has been driven by these approaches
- Accelerator component improvements have benefitted from MOGA/AI/ML
  - e.g. parameterizable optimizable components like SRF cavities, most magnets
  - Very advanced components (crab SRF cavities, complex SC magnets) still bespoke design
- Community effort identified goals of next-generation accelerator modeling/design
  - Snowmass advanced accelerator modeling and simulation white paper
  - Recommendations on virtual twins, advanced algorithms, common frameworks
- Very active (understaffed) community
  - Exciting new developments from all of AI/ML (algorithms, learning, control, analysis)
  - 3rd ICFA beam dynamics mini-workshop: Chicago Nov 1-4