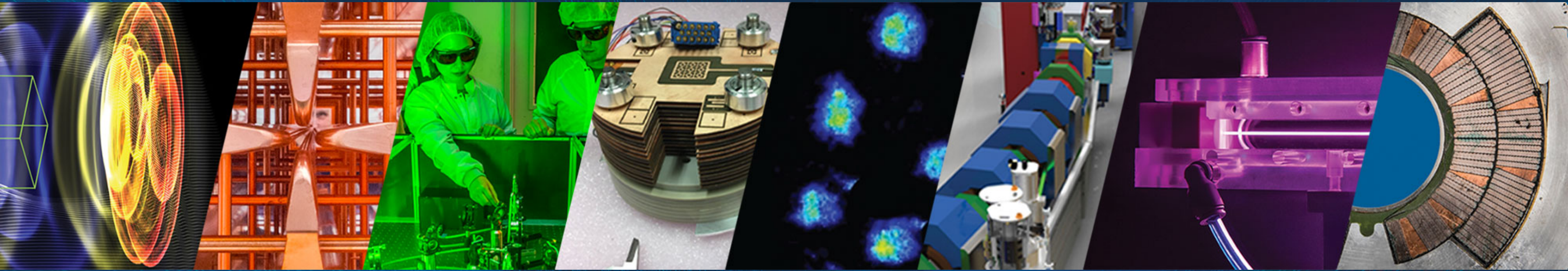
Machine Learning Applications for ALS & ALS-U

Simon C. Leemann, ALS Accelerator Physics Program Deputy, for the ALS-AP ML Team
Accelerator Technology & Applied Physics Division



Director's Review
May 16, 2024

BERKELEY LAB
ACCELERATOR TECHNOLOGY & APPLIED PHYSICS DIVISION ATAP
U.S. DEPARTMENT OF ENERGY | Office of Science

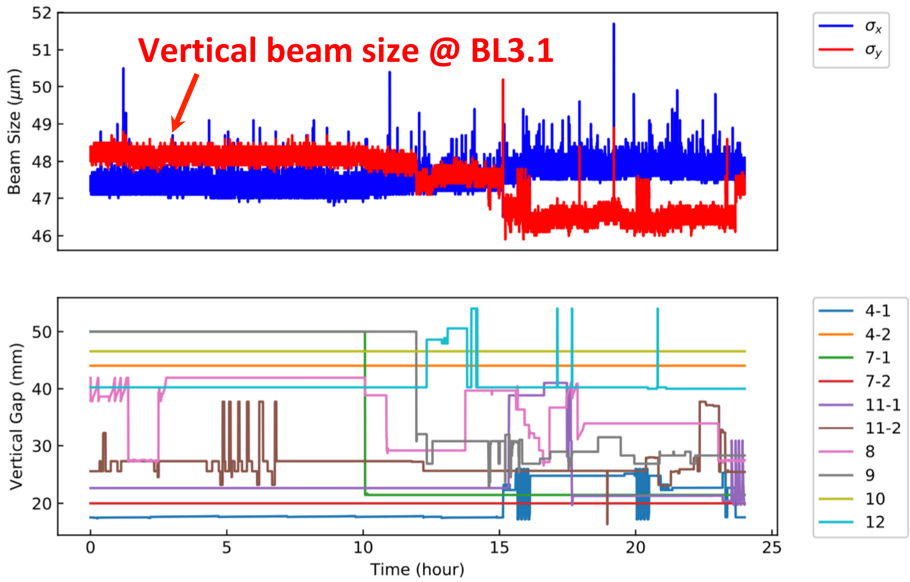
Intro: Machine Learning (ML) at the ALS

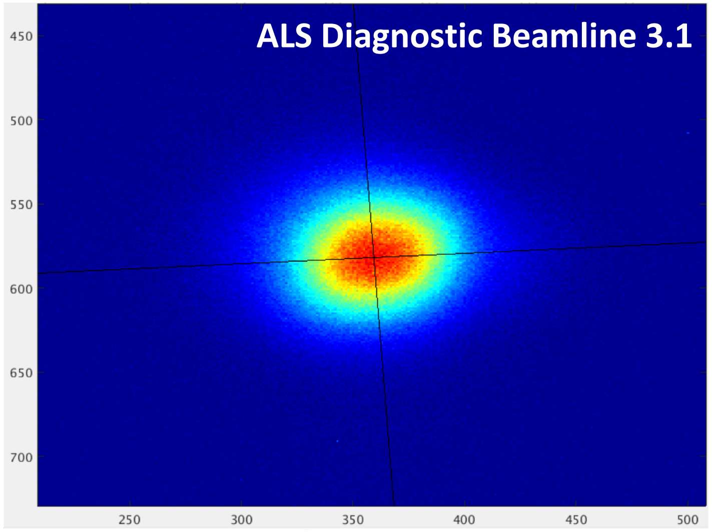
- **ALS** has been kept at the forefront of soft x-ray light sources for 3 decades by continuous **upgrades** and **R&D effort**
- **ML** presents excellent new opportunity for accelerator R&D to extend **ALS leadership**
- ALS ML efforts have so far been enabled by a 3-year grant funded jointly by **DOE BES ADRP & ASCR** as well as by **ALS operations funds**
- Initial **ALS ML R&D effort:** use ML as a powerful "new" tool to solve "old" accelerator problems:
  - **Accelerator operations:** automated tuning, replace feedback approaches, virtual diagnostics ➔ *Project #1*
  - **Accelerator development:** improve physics understanding, augment/extend lattice optimization, accelerate multi-objective optimization (e.g. MOGA) ➔ *Project #2*

#1 ML for Acc Ops: Stabilizing Beam Size at ALS

- State-of-the-art light sources **achieve excellent stability** in terms of beam position/angle & current (orbit feedbacks, top-off injection)
- In spite of *extensive* correction efforts, **beam size** is still perturbed by **insertion device (ID)** config changes ➔ can affect experimental resolution
- Problem is nonlinear, complex, and non-stationary
- Previous solutions relied on approximations & required extensive **dedicated machine time** for frequent recalibration (feed-forward tables)
- Resulting level of performance has started to become a limitation at most demanding experiments (STXM, XPCS, ptychography, …)

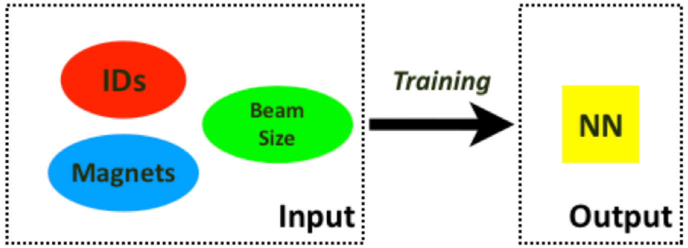
➢ Expected to become a **serious issue** in next-generation light sources (diffraction-limited storage rings, eg. APS-U, ALS-U, …)

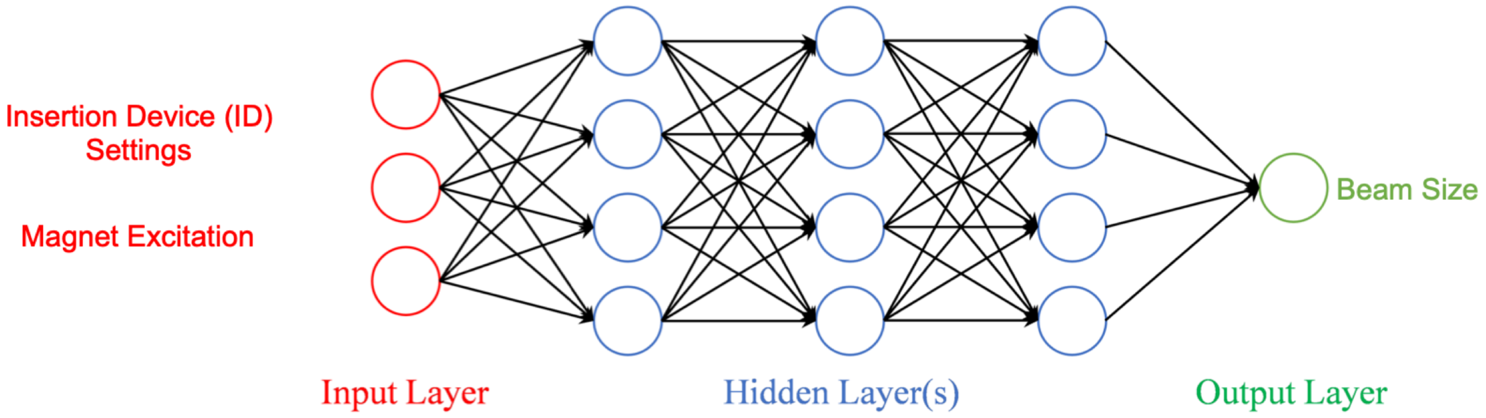




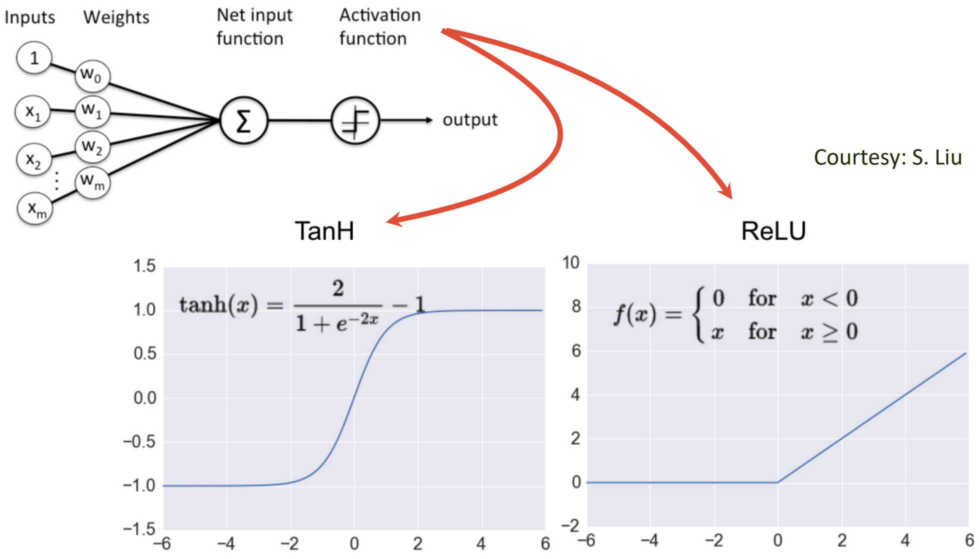
Developing a Solution Based on Machine Learning

- **Machine Learning** can exploit large amounts of data that are already collected during routine operations ➔ ***"training"***
- Once trained, **neural network (NN)** provides **predictions** for beam size changes that result from ID config changes & skew quad corrections (V disp. wave)





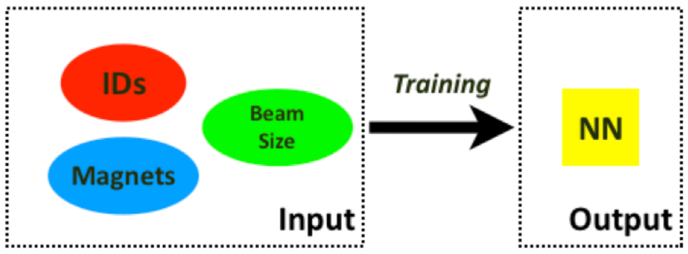
Fully-connected 5-layer NN (29-128-64-32-1)



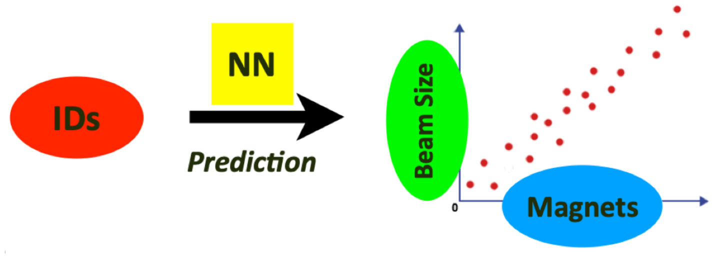
Courtesy: S. Liu

Developing a Solution Based on Machine Learning

- **Machine Learning** can exploit large amounts of data that are already collected during routine operations ➔ ***"training"***
- Once trained, **neural network (NN)** provides **predictions** for beam size changes that result from ID config changes & skew quad corrections (V disp. wave)



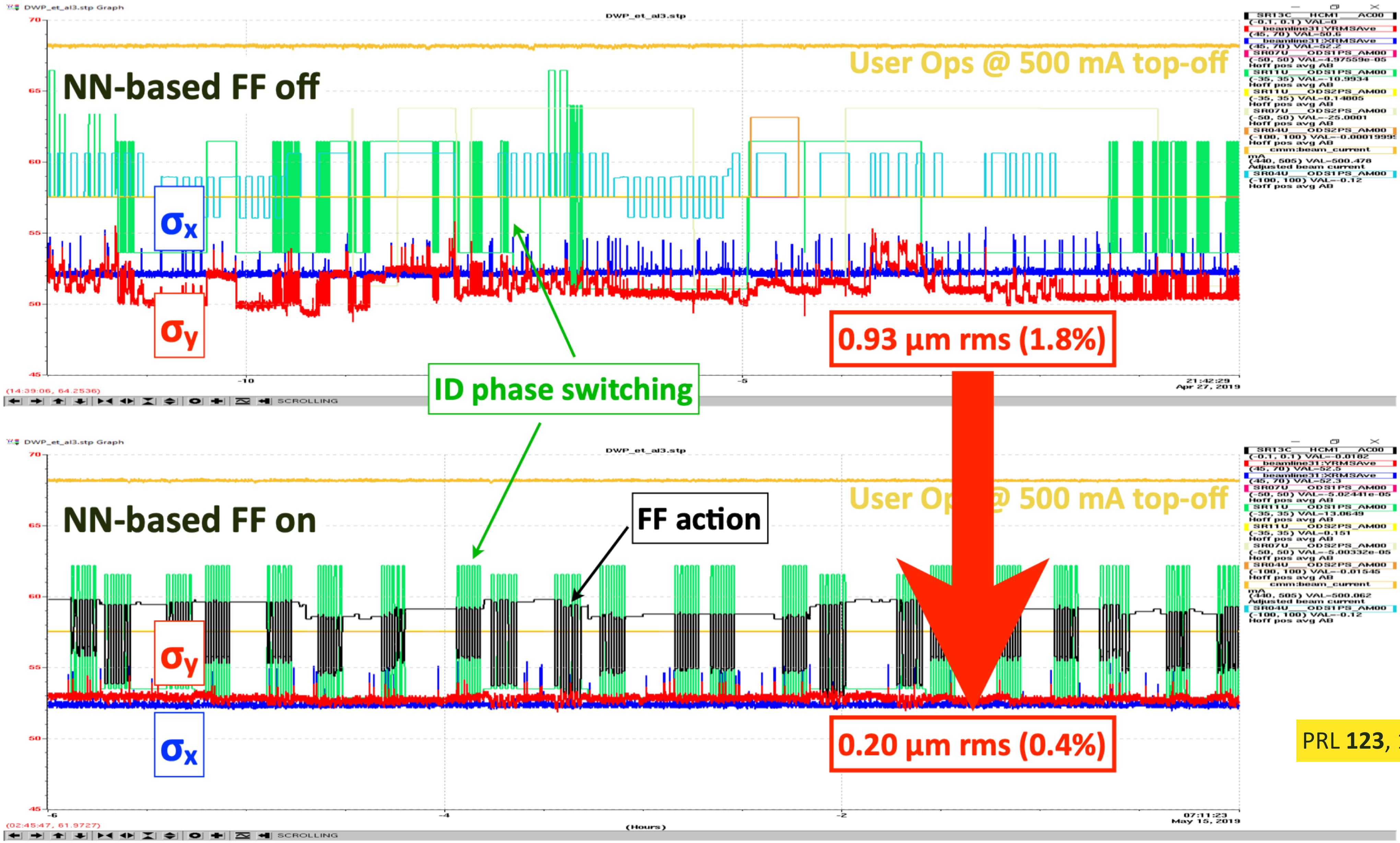
- These predictions can serve as a **dynamic lookup** ➔ which skew quad correction required to compensate for changes resulting from currently applied ID config?



- If such a lookup is incorporated into the accelerator control system as a **feed forward (FF)**, we can stabilize the storage ring over prolonged periods of time & **online retraining** can mitigate drift

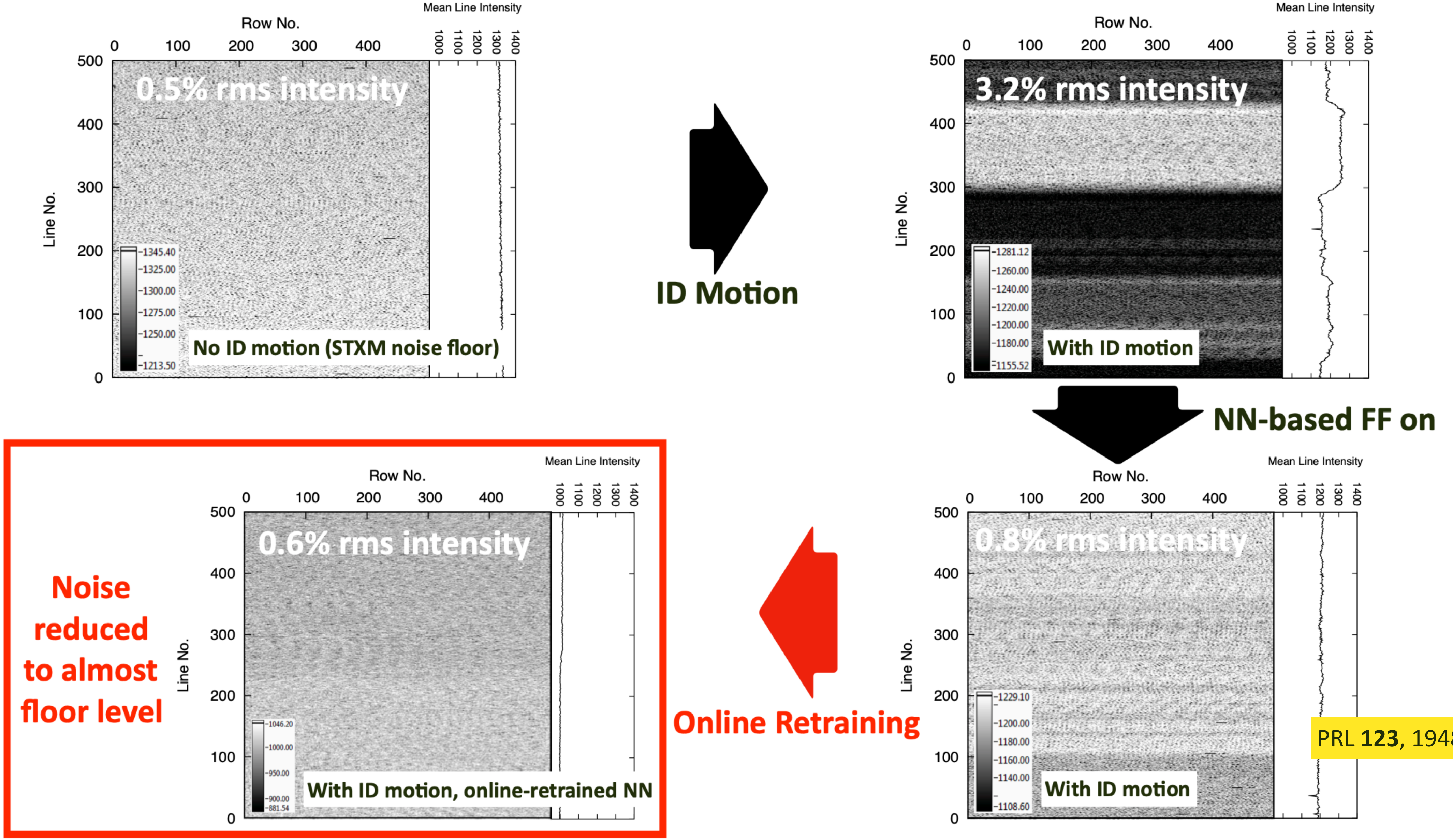
PRL **123**, 194801 (2019)

Results: NN-based FF Off vs. On During User Ops



PRL **123**, 194801 (2019)

Stabilization Confirmed at STXM @ ALS Beamline 5.3.2.2

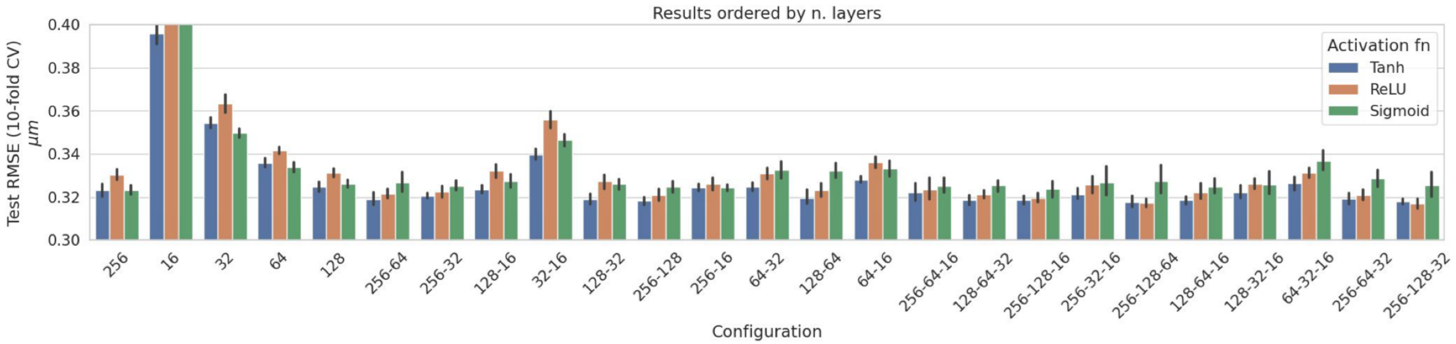


PRL **123**, 194801 (2019)

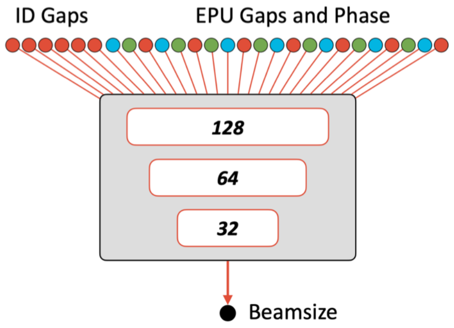
Recently: Improved Model With Online Fine-tuning & Implementation

- Split up **DNN** into 2 models
- Revised primary DNN based on broad hyper-parameter optimization
- Revised & accelerated training DAQ: ID scanning based on one prior year of user ops
- Online retraining ➔ **online fine-tuning** anchored DNN using sliding window buffer ➔ measured performance now at diagnostic BL noise floor
- Robustness & integration ➔ new high-level interface with event logging & monitoring

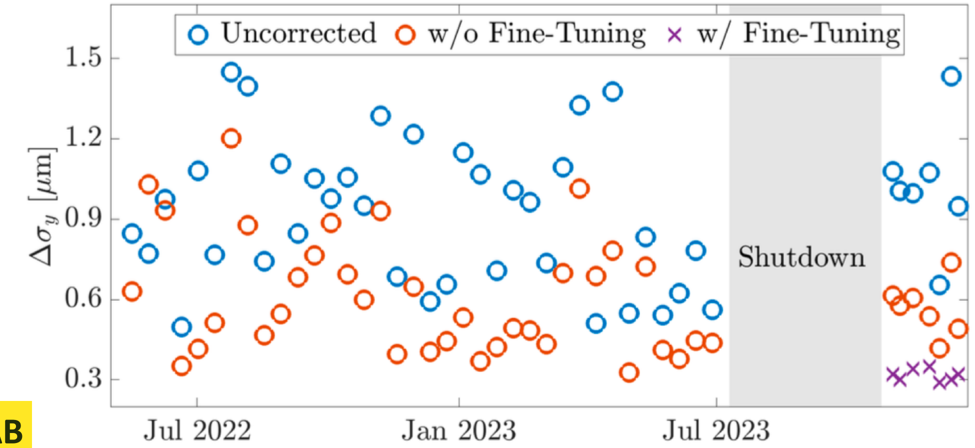
➢ One-click operation provides for becoming standard part of everyday user ops



Courtesy: A. Pollastro



Courtesy: T. Hellert

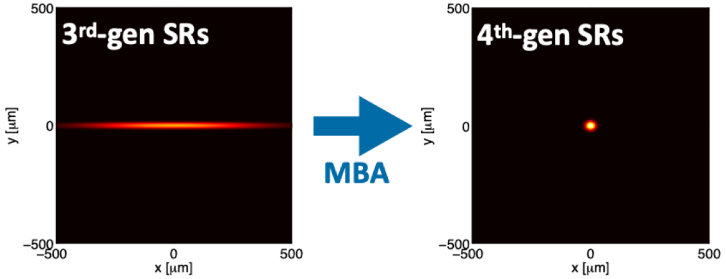


submitted to PRAB

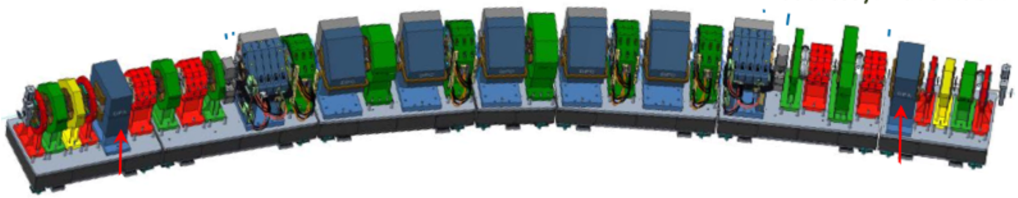
#2 ML for Acc Dev: Improving Multi-Objective Optimization

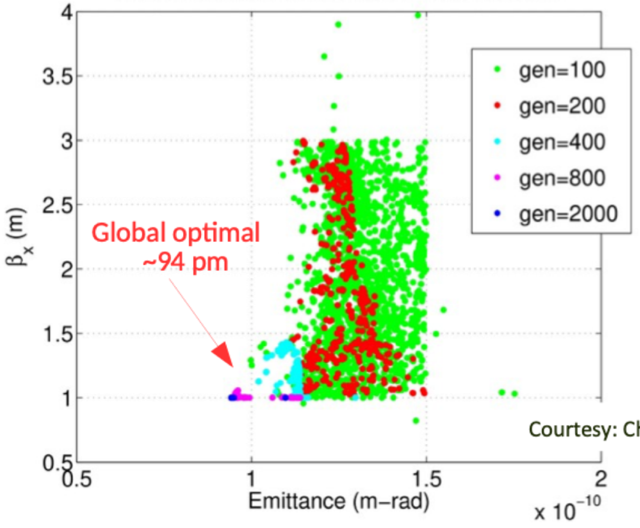
- **4th-generation storage rings (4GSRs)** leverage multi-bend achromat (MBA) lattices to deliver ultra-high brightness & large coherent fraction
- But **MBA lattices** are very challenging: dense & exploit very strong focusing ➔ drives large chromatic terms & higher-order corrections
- Solutions not only highly nonlinear but involve many degrees of freedom (DoF) ➔ **demanding optimization**:
  - tough objectives, many often in direct competition
  - large number of parameters, many boundary constraints

➢ **Multi-objective genetic algorithms (MOGA)** are highly successful at such optimization & have become tool of choice among community



Courtesy: Dave Robin





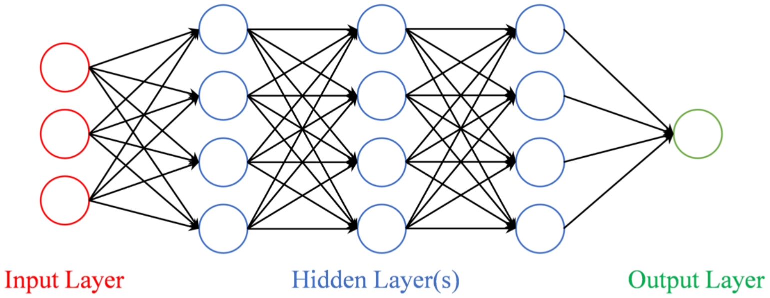
Courtesy: Changchun Sun

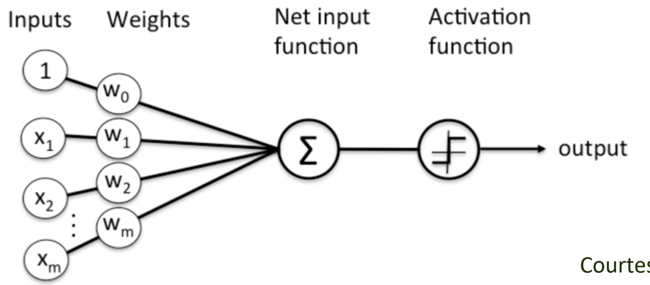
Improving MOGA: ML to the Rescue

- But MOGA's **stochastic nature** is inherent weakness ➔ need to evaluate vast number of lattice candidates, most ultimately rejected
- Do *not* want to artificially limit DoF, search ranges, or make many initial assumptions about attractive solutions ➔ so what can we do?
- ML can be employed to render **deep neural networks (DNNs)** ➔ **surrogate models** used in place of computationally expensive evaluation (many-turn nonlinear tracking)

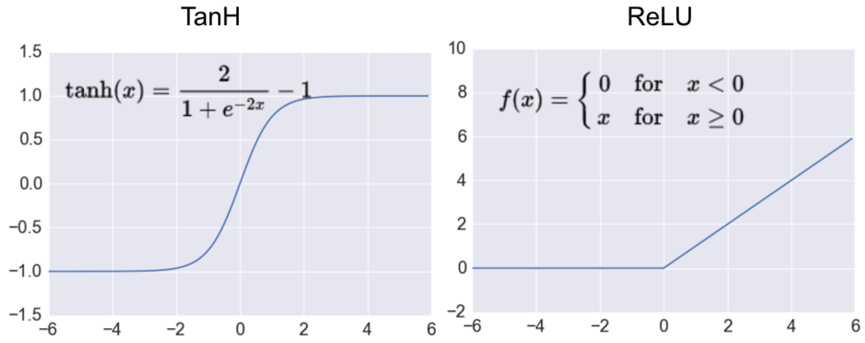
➢ Lattice candidate evaluation becomes near **instantaneous**

- And ideally, we'll want to target:
  - speed up MOGA without modifying MOGA/tracking tools or existing workflows & without sacrificing physics fidelity
  - direct optimization of relevant physics quantities (ε_0, DA, MA)
  - combined linear/nonlinear optimization employing all free parameters (quadrupoles & sextupoles & …)



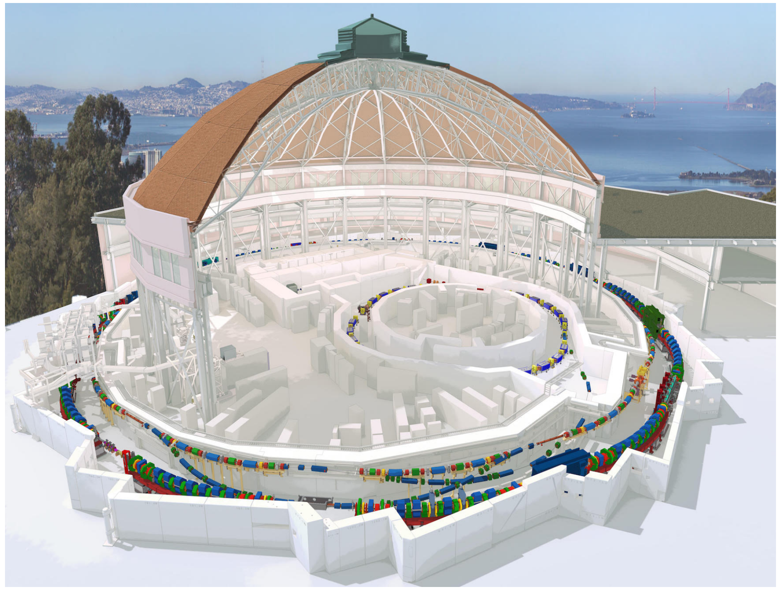


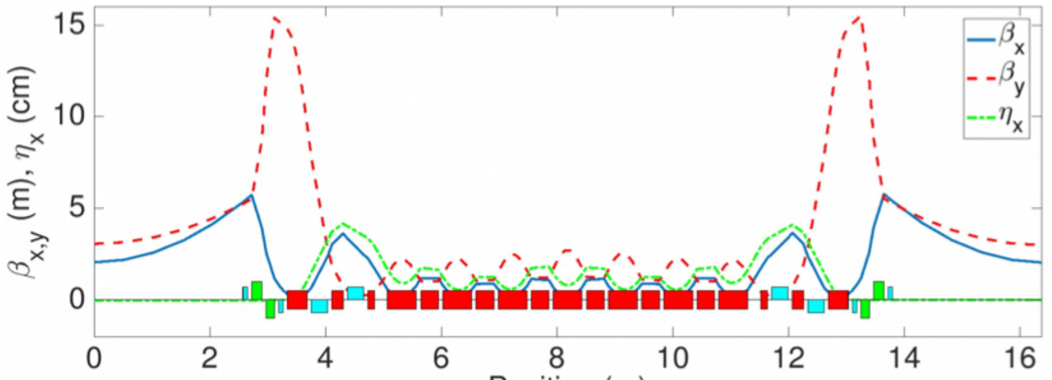
Courtesy: S. Liu



ALS-U Optimization as a Test Case for ML

- **ALS-U storage ring** calls for challenging 9BA lattice to achieve ≈75 pm rad (round beam) @ 2 GeV in <200 m ➔ dense, strong focusing, **very strained optics**

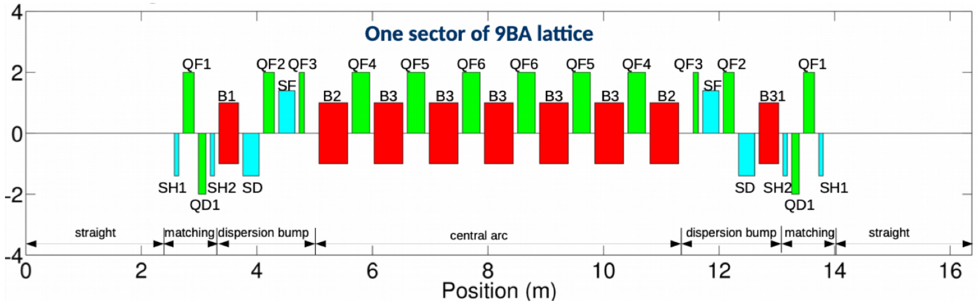




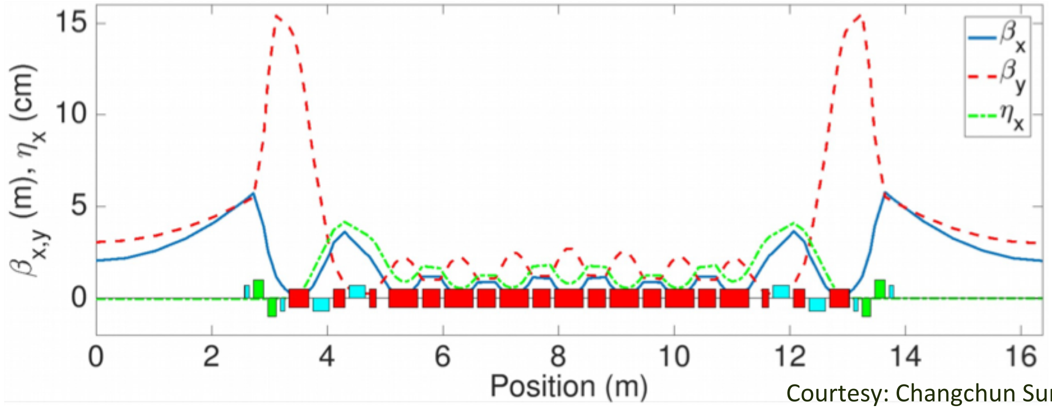
Courtesy: Changchun Sun

ALS-U Optimization as a Test Case for ML

- **ALS-U storage ring** calls for challenging 9BA lattice to achieve ≈75 pm rad (round beam) @ 2 GeV in <200 m ➔ dense, strong focusing, **very strained optics**
- Initial optimization: 9 quadrupoles, 4 sextupoles ➔ **11 free knobs** (later: include reverse bending & superbends)
  - Roughly a dozen magnet/lattice **constraints** on top of pre-determined quadrupole ranges
  - **Objectives**: ε_0, MA, and on-momentum DA (modeled as integrated diffusion rate)



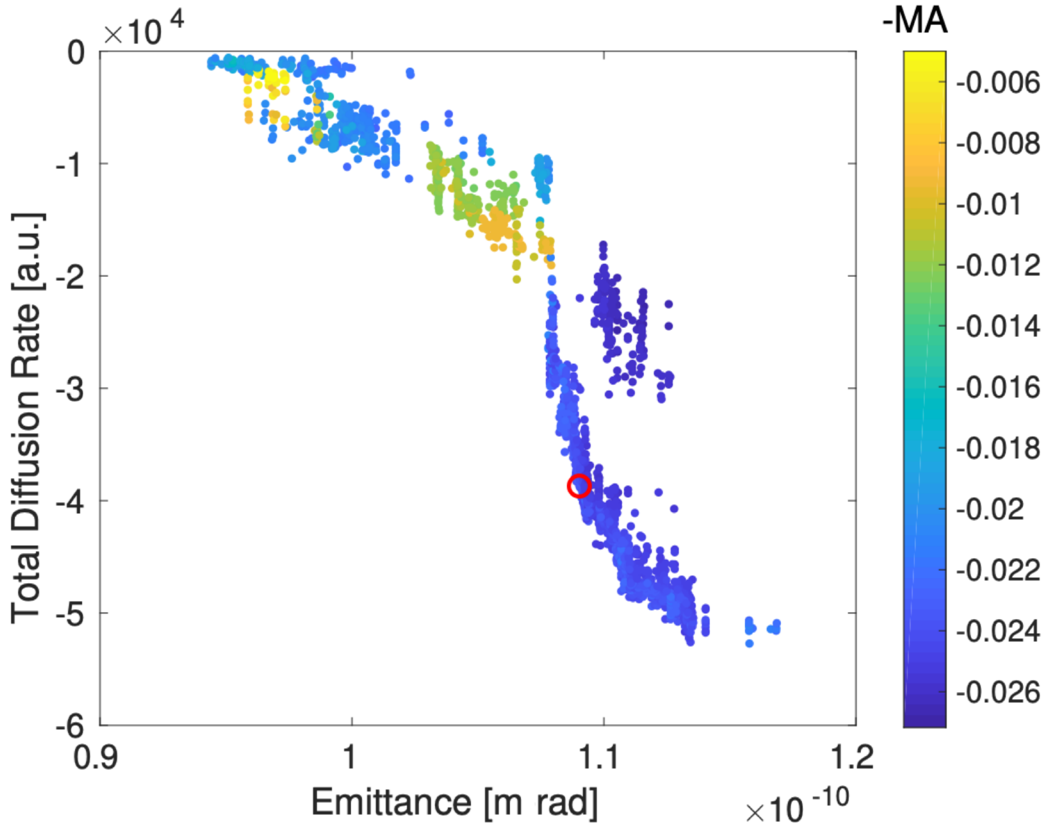
Natural emittance	$\varepsilon_0 < 155\,\mathrm{pm\,rad}$
Maximum beta	$\beta_{x,y} < 30\,\mathrm{m}$
Maximum dispersion	$\eta_x < 15\,\mathrm{cm}$
Fractional tunes	$0.1 < \nu_{x,y} < 0.4$
Dispersion at center of straight	$\lvert\eta_x^*\rvert < 1\,\mathrm{mm}$
Beta at center of straight	$1\,\mathrm{m}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



Courtesy: Changchun Sun

ALS-U Optimization as a Test Case for ML

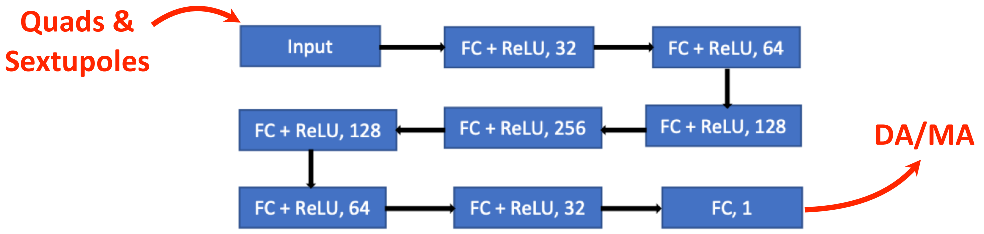
- **ALS-U storage ring** calls for challenging 9BA lattice to achieve ≈75 pm rad (round beam) @ 2 GeV in <200 m ➔ dense, strong focusing, **very strained optics**
- Initial optimization: 9 quadrupoles, 4 sextupoles ➔ **11 free knobs** (later: include reverse bending & superbends)
  - Roughly a dozen magnet/lattice **constraints** on top of pre-determined quadrupole ranges
  - **Objectives**: ε_0, MA, and on-momentum DA (modeled as integrated diffusion rate)
- Ultimately, a highly **staged MOGA approach** resulted in
  - ±1 mm DA (compatible with on-axis swap-out & AR)
  - ≈1 hr overall lifetime (including x4 boost from 3HCs)
- …but required ***months* of CPU time** on large clusters



Courtesy: Changchun Sun

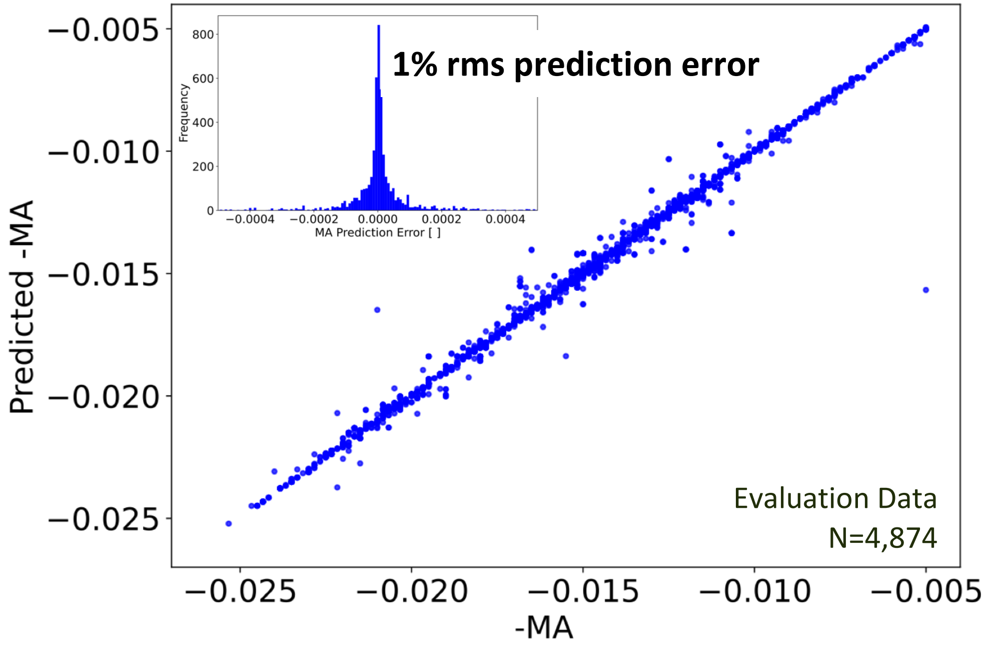
ML for Full Linear & Nonlinear ALS-U Optimization

- **Training data** for 11D problem cannot be acquired through systematic sampling of input space
- But do not want to make too many assumptions or "wise choices" ➔ retain generality of approach
- Instead: use first few generations of conventional **MOGA** as training data for deep neural networks (DNNs)
- **Two 8-layer DNNs** used in MOGA instead of calls to Tracy for DA and MA (via many-turn tracking)
  - Traditional MOGA requires about **640 gen** (5000 children/gen) ➔ ≈8 days on 1000-core cluster
  - Training 2 DNNs to get DA/MA predictions ≈1% rms requires data from about **10 gen**
- Once DNNs trained (≈30 min on desktop CPU) ➔ quasi-instantaneous lookup (16 ms) vs. DA/MA tracking (88 sec)



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

Courtesy: Yuping Lu



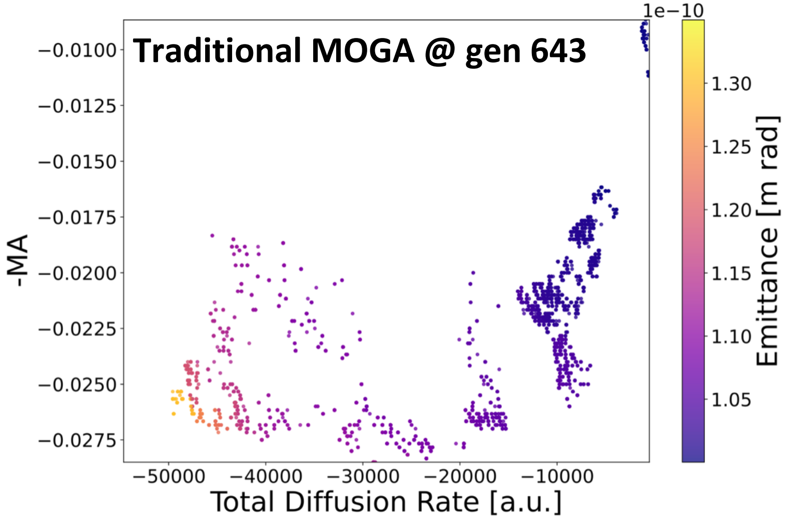
NIM-A **1050**, 168192 (2023)

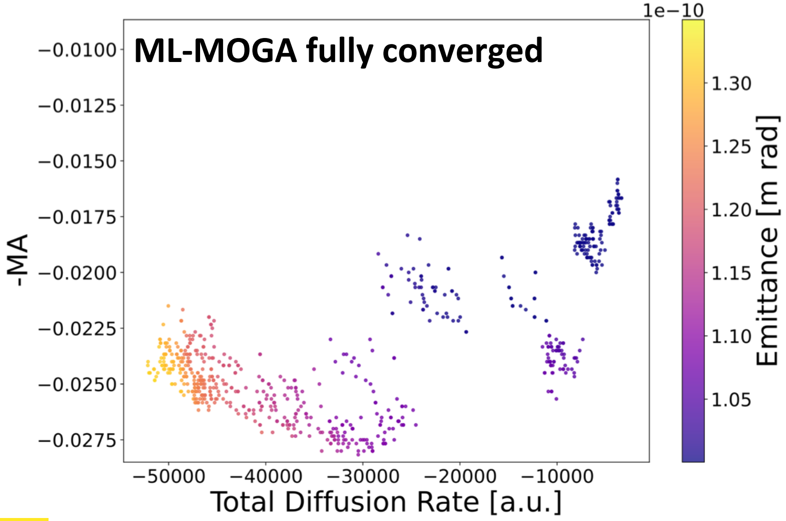
Results: ML-MOGA Successful & 40× Faster

- **ML predictions** are not 100% accurate (training based on early data)
- Initial ML-MOGA solutions show disagreement compared to **tracking validation** ➔ but can **retrain DNNs** with data from validation step
- Iterate cycles of *validation–retraining–ML-MOGA* using **model-independent distance metrics** to determine **convergence**

➢ ML-MOGA very quickly converges (6-8 iterations) toward true **Pareto-optimal front** ➔ overall **speedup ≈ 40×** (incl. training effort)

- Once fully converged, **ML-MOGA inputs & objectives match** those of traditional MOGA to within "noise floor"
- **Flexible**: can be adapted to other lattice optimization problems as long as can provide reasonably accurate DNNs
- Potential to **fully automate** entire optimization campaign & **optimize in parallel** from the start for many error lattices is highly attractive ➔ derive truly ***global*** optimum





NIM-A **1050**, 168192 (2023)

Success, Outlook & Opportunities for Collaboration

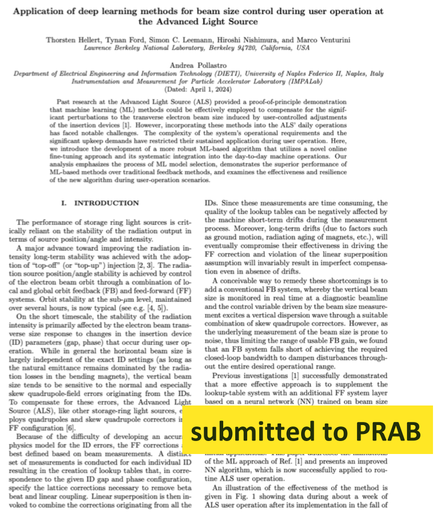
- Project success: ML stabilization now always running during user ops @ ALS & 1x **PRL**, 1x **NIM-A**, 1x submitted **PRAB**, ***2021 Klaus Halbach Prize for Innovative Instrumentation***

- ML applied to accelerators shows vast potential to enable
  - new **more aggressive designs**, but also
  - ***exploit full performance*** of existing & soon to be commissioned rings

➢ These are highly relevant issues in both present (ALS) & future **4th-generation storage rings** (ALS-U)

➢ We have plenty of ideas for future ML applications (White Paper submitted to E. Lessner)

- ATAP & LBL continue to **foster great collaboration** on **ML for accelerators**
  - ➢ https://atap.lbl.gov/research/crosscutting-endeavors/artificial-intelligence-and-machine-learning/
  - ➢ https://ml4sci.lbl.gov/projects



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 Contract No. DEAC02-05CH11231

BERKELEY LAB

ACCELERATOR TECHNOLOGY & APPLIED PHYSICS DIVISION ATAP

U.S. DEPARTMENT OF ENERGY | Office of Science