Demonstrations of Machine Learning on Particle Accelerators

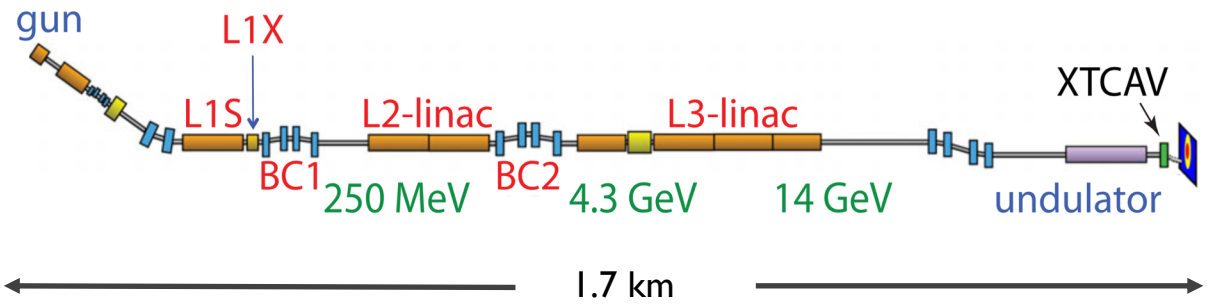
Auralee Edelen, Joe Duris, Claudio Emma, Adi Hanuka, Xiaobiao Huang, Dylan Kennedy, Chris Mayes, Nicole Neveu, Daniel Ratner, Xinyu Ren, Jane Shtalenkova, Faya Wang, Xiao Zhang (SLAC), Alex Scheinker (LANL), Andreas Adelmann, Yannick Huber, Matthias Frey (PSI), Jonathan Edelen (RadiaSoft), Sandra Biedron (UNM), Eric Cropp, Pietro Musumeci, Paul Denham (UCLA), Elena Fol (CERN)

NAPAC 2019
1-6 September, 2019
Lansing, MI

U.S. DEPARTMENT OF ENERGY
Office of Science

Fermilab
UCLA
Los Alamos NATIONAL LABORATORY EST. 1943
PAUL SCHERRER INSTITUT PSI
radiasoft
SLAC NATIONAL ACCELERATOR LABORATORY

Particle accelerators are difficult to model and control



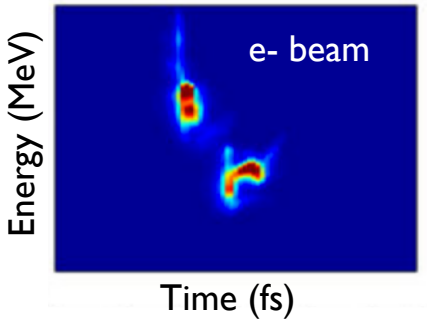
Interesting challenges for modeling / control

- Complex systems *(nonlinear, large parameter spaces)*
- Interacting subsystems
- Variety of diagnostics *(e.g. beam images)*
- Time-varying/ non-stationary behavior *("drift")*

Strong incentives for improving system understanding and control

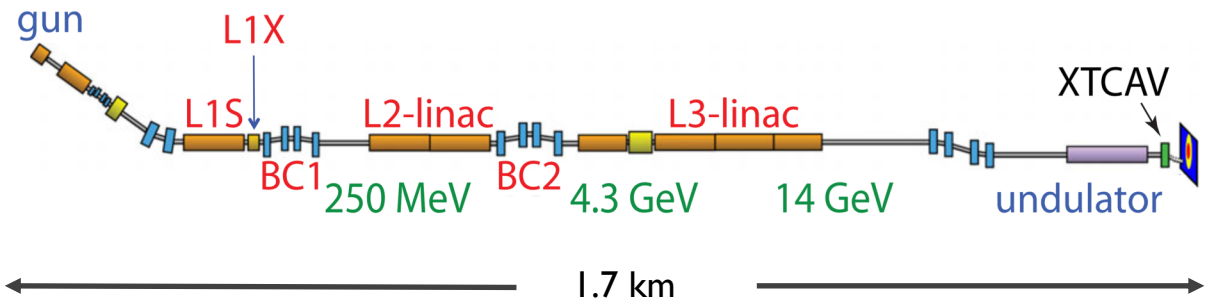
- High user demand → *want to switch between custom user requests quickly*
- High cost for unintended down-time → *user time, scientific output*
- Achieve challenging beam setups for new science goals

Bunches for two color FEL



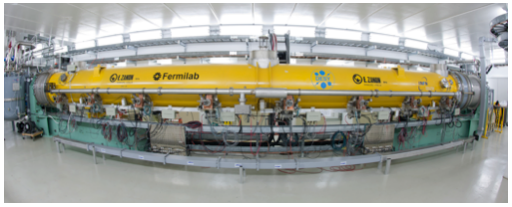
A. Marinelli, et al., Nat. Commun. **6**, 6369 (2015)

Particle accelerators are difficult to model and control



Interesting challenges for modeling / control

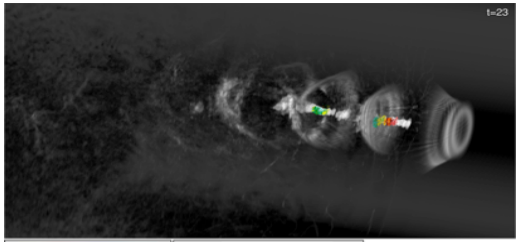
- Complex systems *(nonlinear, large parameter spaces)*
- Interacting subsystems
- Variety of diagnostics *(e.g. beam images)*
- Time-varying/ non-stationary behavior *("drift")*



Fermilab

Strong incentives for improving system understanding and control

- High user demand → *want to switch between custom user requests quickly*
- High cost for unintended down-time → *user time, scientific output*
- Achieve challenging beam setups for new science goals



LBNL Visualization Group

Even more challenging as we move to more complicated acceleration schemes (e.g. superconducting, plasma-based acceleration) and push to more extreme beam parameters

Major use cases for ML in particle accelerators

SLAC

- Detecting / predicting unwanted changes or failures
- Getting more useful information out of complicated machine signals *(e.g. images, waveforms)*
- System control and optimization
- Fast, accurate system models
- Facilitate improved physics understanding of machine behavior

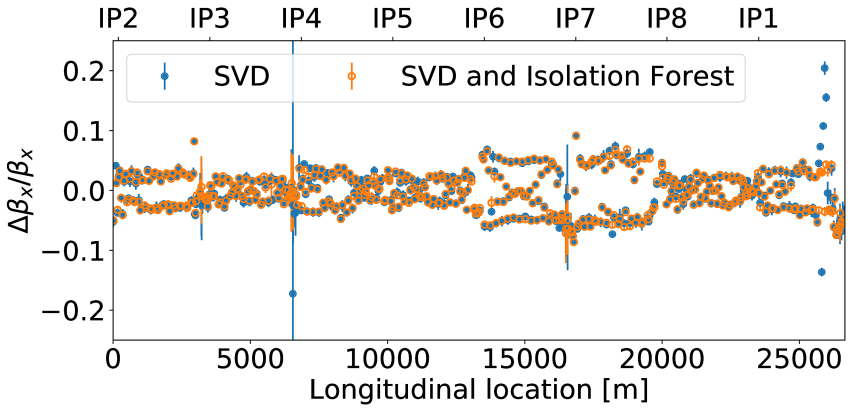
→ ***Will give examples of each of these***

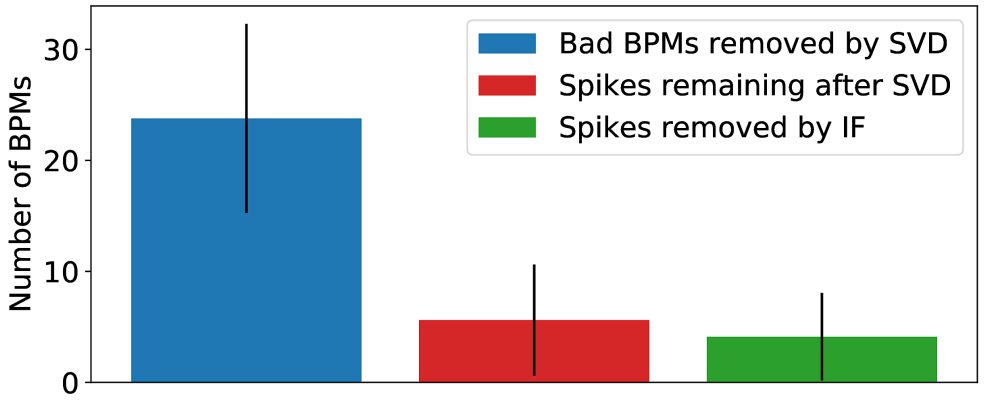
Anomaly Detection: identify bad BPM signals

SLAC

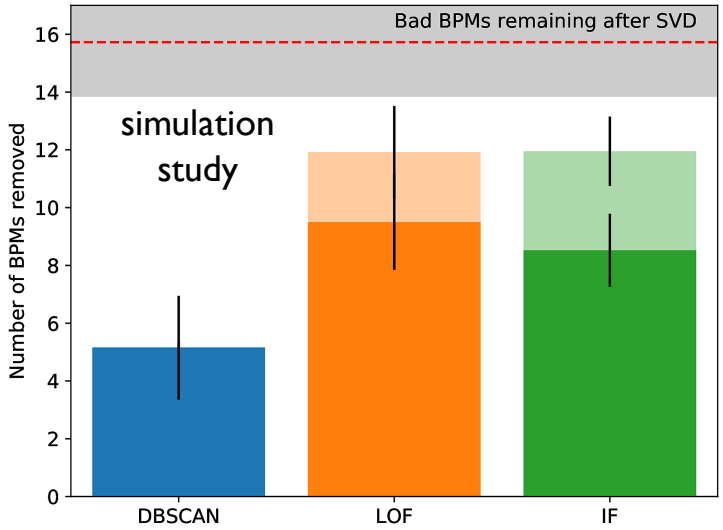
- *Don't want to use faulty BPM signals in optics measurement and correction!*
- Standard techniques (e.g. SVD) can remove most bad signals, but not all
- Various clustering algorithms have been applied to LHC BPMs *(DBSCAN, Isolation Forests, Local Outlier Factor)*

fewer outliers in beta-beating computation





Results from 2018 MD run



Lighter shade – total number removed
Darker shade – actual bad BPMs removed
Balance for different accelerator needs

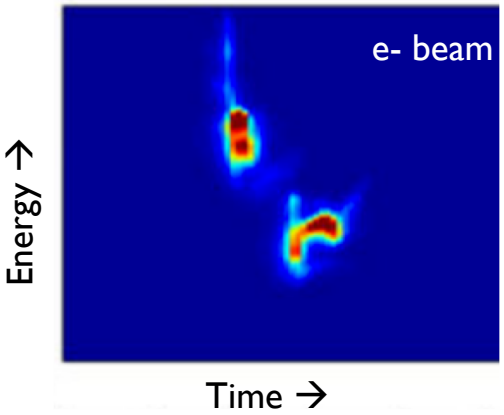
Diagnostic Analysis / Reconstruction

Complicated signals used in feedback control and experimental analysis

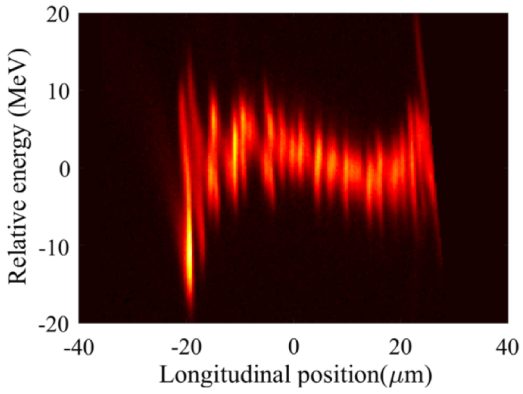
(e.g. beam images, rf waveforms)

→ *Can use ML to extract more useful information from these signals*

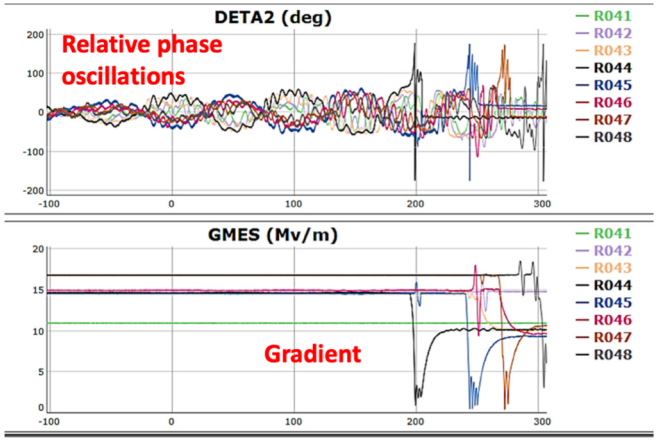
→ *NNs are particularly useful for this*



A. Marinelli, et al., Nat. Commun. **6**, 6369 (2015)

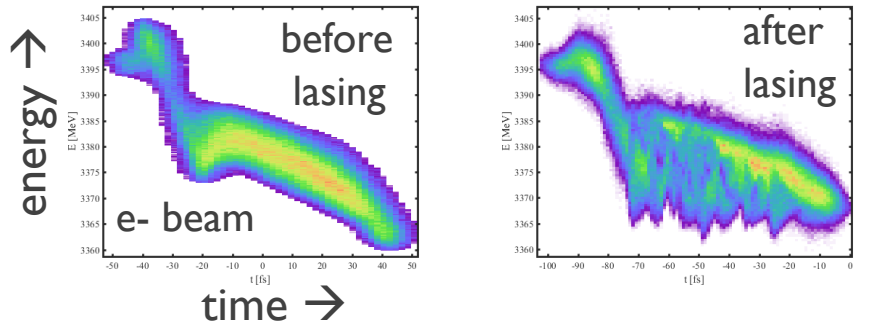


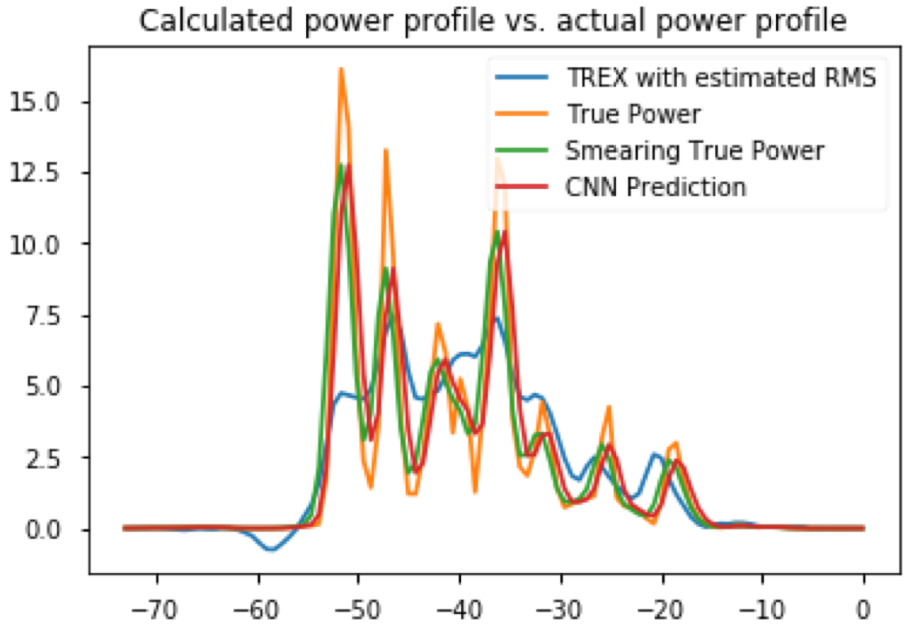
J. Qiang, et al., PRSTAB30, 054402, 2017



A. Solopova, IPAC'19

Diagnostic Analysis / Reconstruction: X-Ray power profile from e- beam image





Free Electron Laser: e- beam loses energy to photon beam

e- beam image before/after lasing process provides critical information to users about photon beam

- *relies on slow, iterative reconstruction algorithm to get X-ray power profile*
- *iterative method doesn't work well for all regimes (e.g. in saturation)*

Instead: use convolutional neural net to get accurate predictions quickly

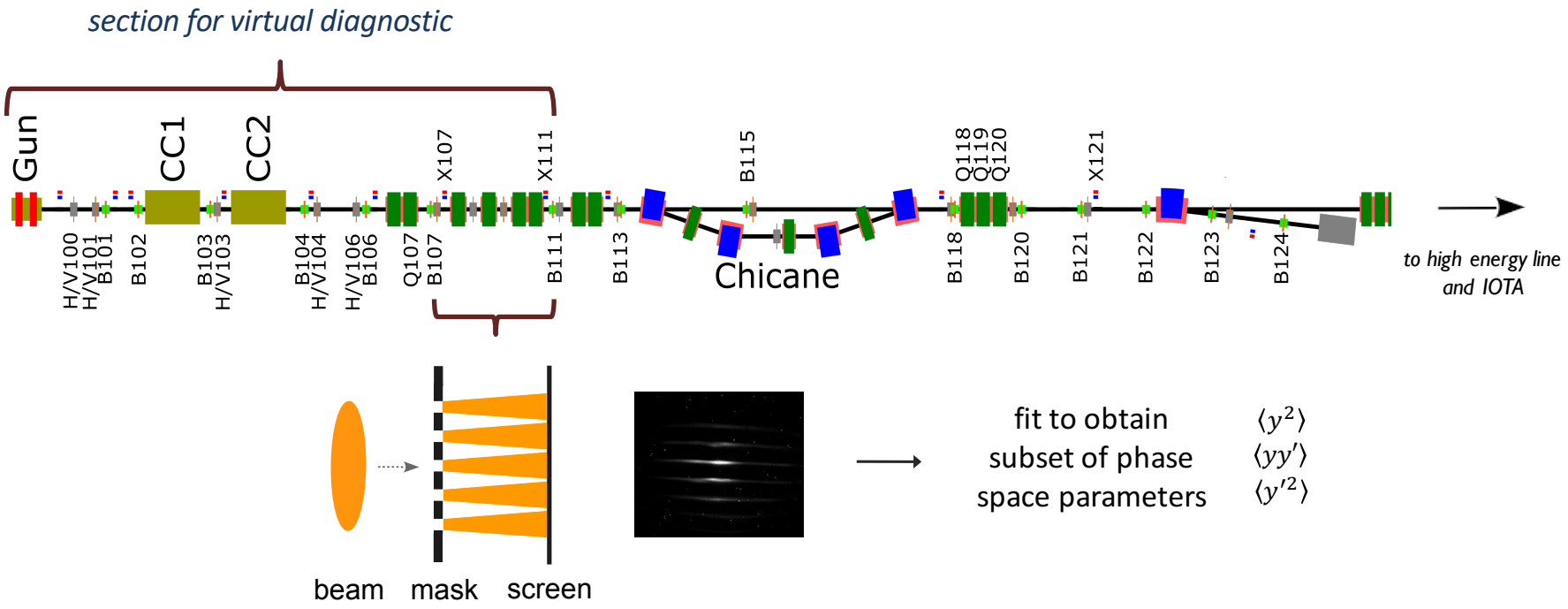
X. Ren, A. Edelen, D. Ratner, et al.

Reconstruction: Virtual Diagnostics

SLAC

Some diagnostic measurements are slow + destructive to the beam

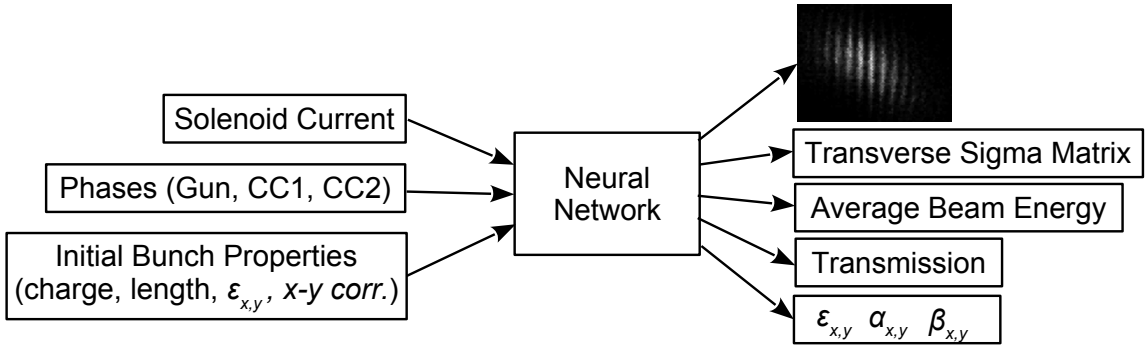
→ *can we use ML to get a non-destructive prediction of what these diagnostics would show?*

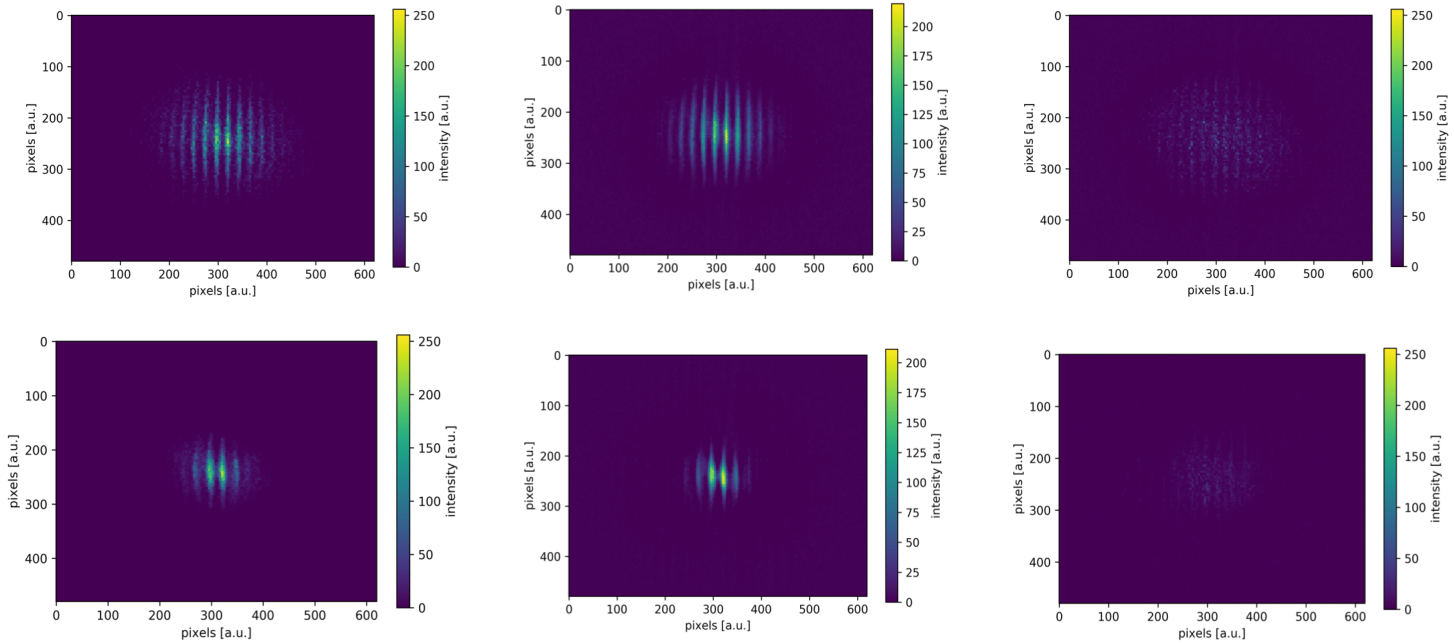


e.g. at FAST (Fermilab) multi-slit emittance measurements took 10-15 seconds in each plane

A. L. Edelen, et al., FNAL Accelerator Seminar, Jun. 2017: https://tinyurl.com/y3l32vbz
A. L. Edelen, J.P. Edelen, D. Edstrom, et al. NAPAC16, TUPOA51
A. L. Edelen, J.P. Edelen, et al. 1st ICFA ML Workshop, Feb. 2018
A. L. Edelen, J.P. Edelen, et al. IPAC18, WEPAF040

Reconstruction: Virtual Diagnostics





Simulated NN Predictions Difference

Reconstruction: Virtual Diagnostics

Real diagnostic not always available:

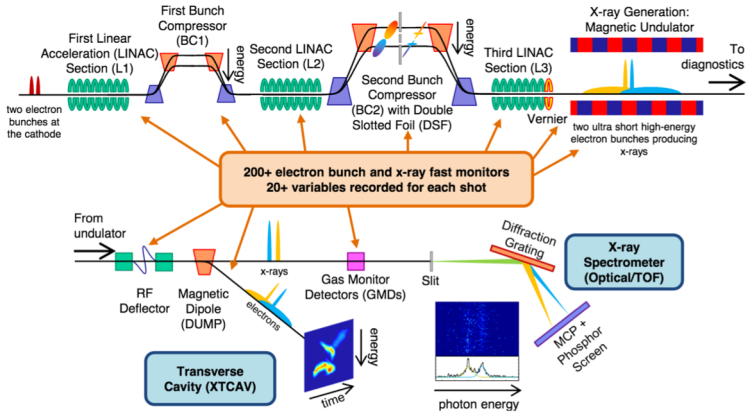
- *slower update rate than desired*
- *destructive, cannot use during user operations*
- *not sensitive in entire operating range*
- *moved to another location (e.g. cost constraints)*

→ *LCLS-II XTCAV*

→ *FACET-II XTCAV*

2016 study: can use archive data to learn correlation between fast and slow diagnostics

A. Sanchez-Gonzalez, et al.
https://arxiv.org/pdf/1610.03378.pdf



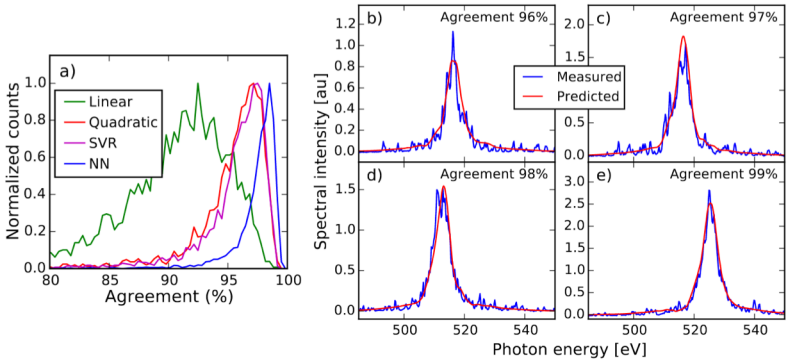


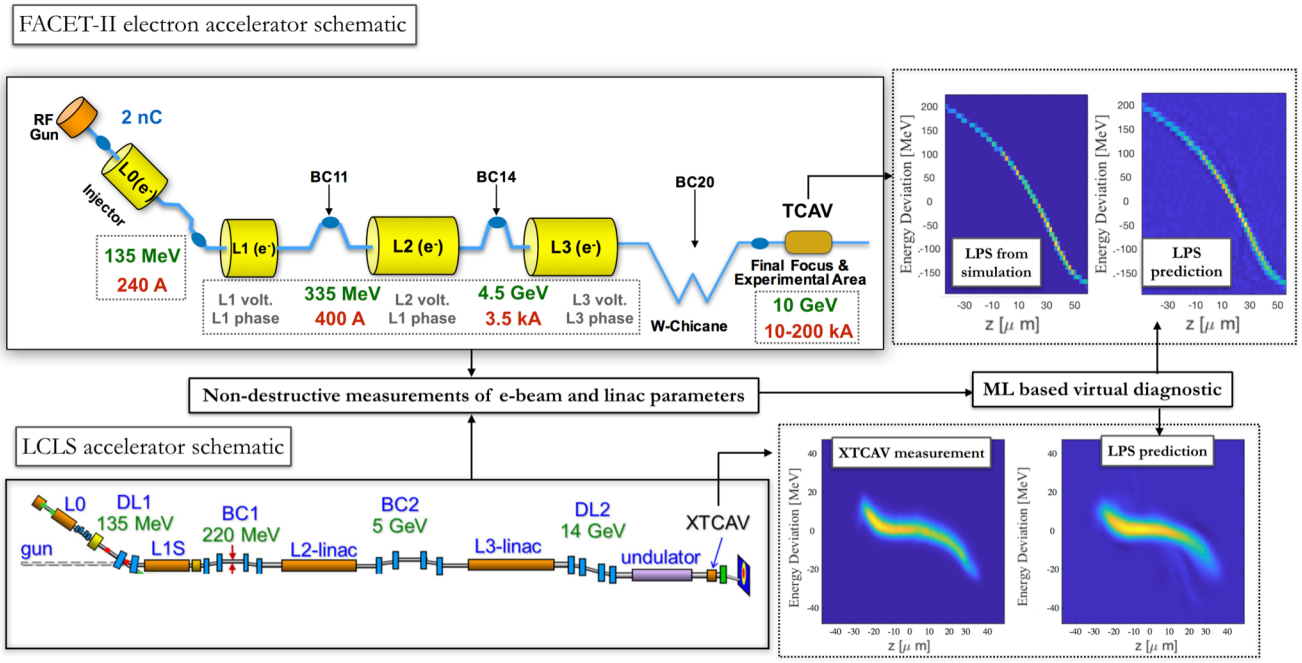
FIG. 4. Spectral shape prediction for a single pulse. (a) Histogram of agreements between the predicted and the measured spectra for the test set using the 4 different models. (b-e) Examples of the measured and the predicted spectra using a neural network to illustrate the accuracy for different agreement values.

Reconstruction: Virtual Diagnostics

SLAC

Real diagnostic not always available:

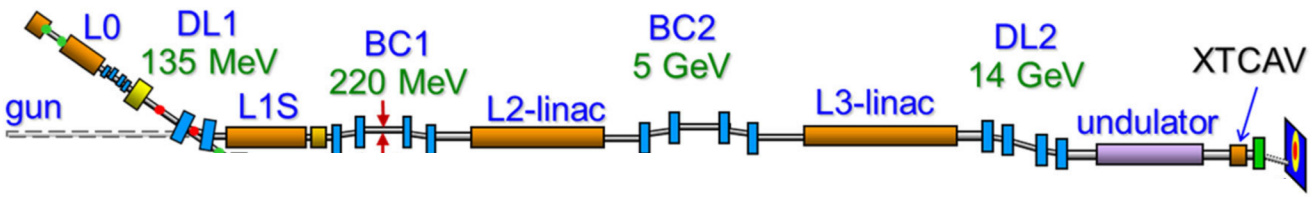
- *slower update rate than desired* → *LCLS-II XTCAV*
- *destructive, cannot use during user operations*
- *not sensitive in entire operating range* → *FACET-II XTCAV*
- *moved to another location (e.g. cost constraints)*



Preliminary simulation study for FACET-II and experimental study at LCLS looks encouraging

C. Emma, A. Edelen, et al., PRAB21, 112802 (2018)

Reconstruction: Virtual Diagnostics

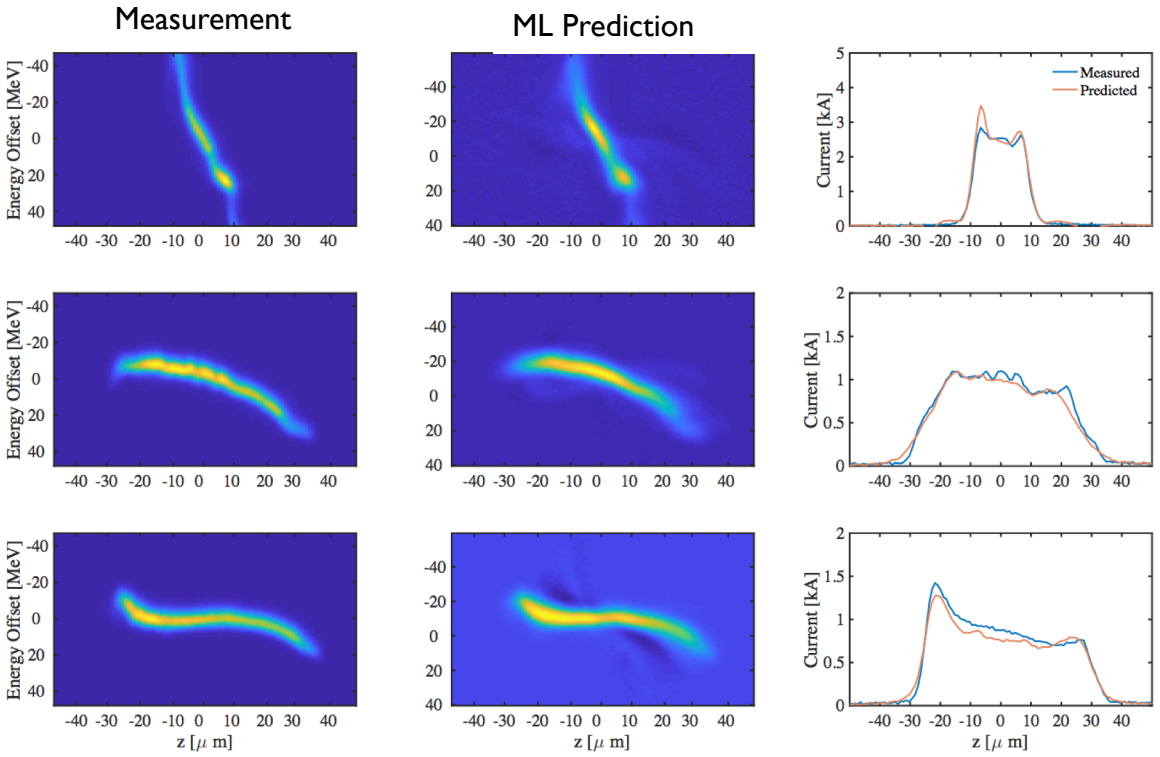


Two parameters scanned:

L1s phase from -21 to -27.8 deg

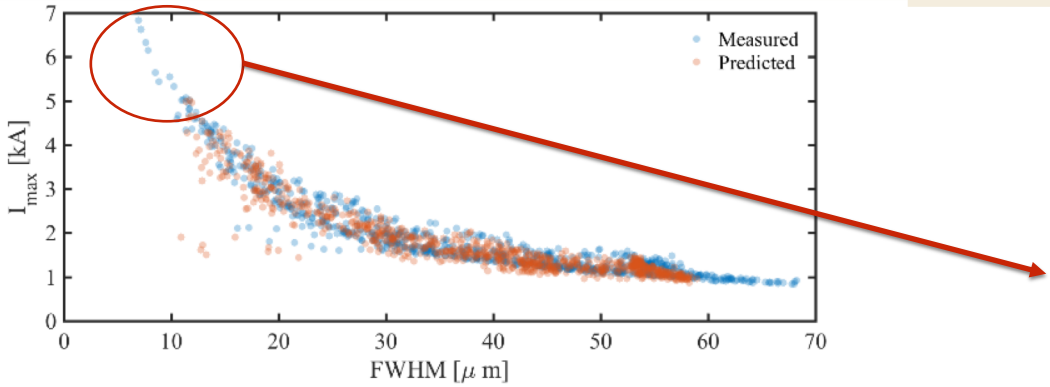
BC2 peak current from 1-7 kA

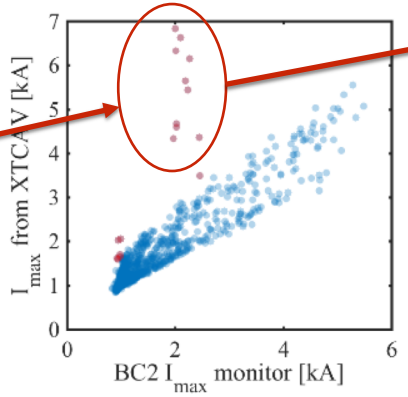
- 5 inputs
  - *L1S and LIX amplitude, LIS phase*
  - *BC1&2 peak current*
- Large sweep of phase space from 2D scan
- Ongoing dedicated effort in FACET-II to provide LPS information to their users



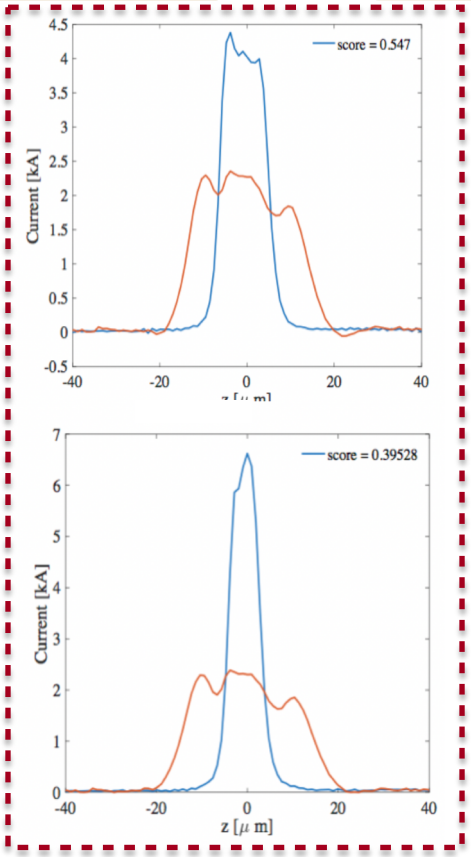
C. Emma, A. Edelen, et al., PRAB21, 112802 (2018)

Reconstruction: Virtual Diagnostics



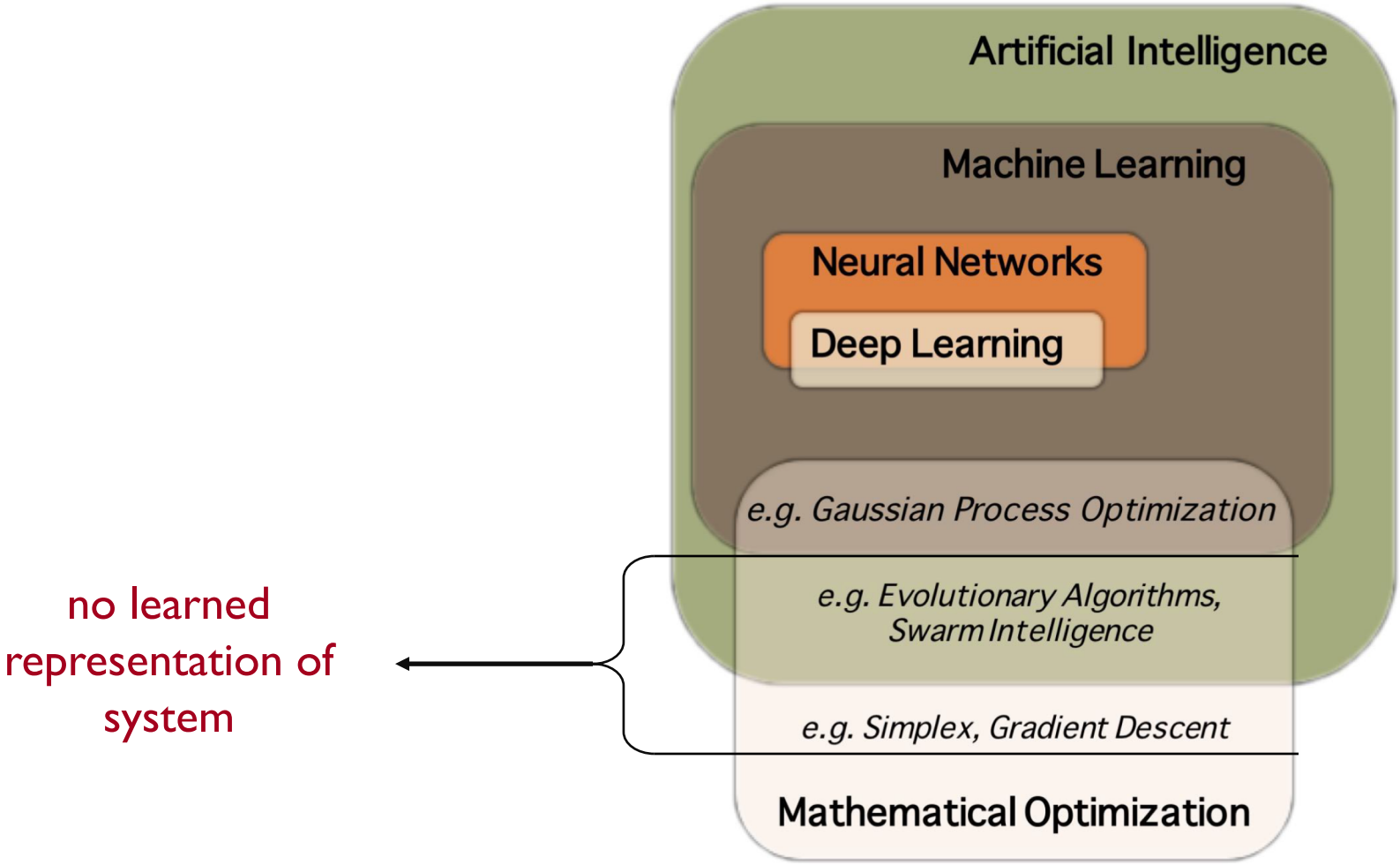


Shots with poor prediction



- For some shots XTCAV and BC2 current readings aren't consistent → *results in poor prediction*
  - → *ML model is only as good as the input it's given*
  - → *Flagging when to trust the prediction is important (e.g. tag bad shots)*

Control / Tuning

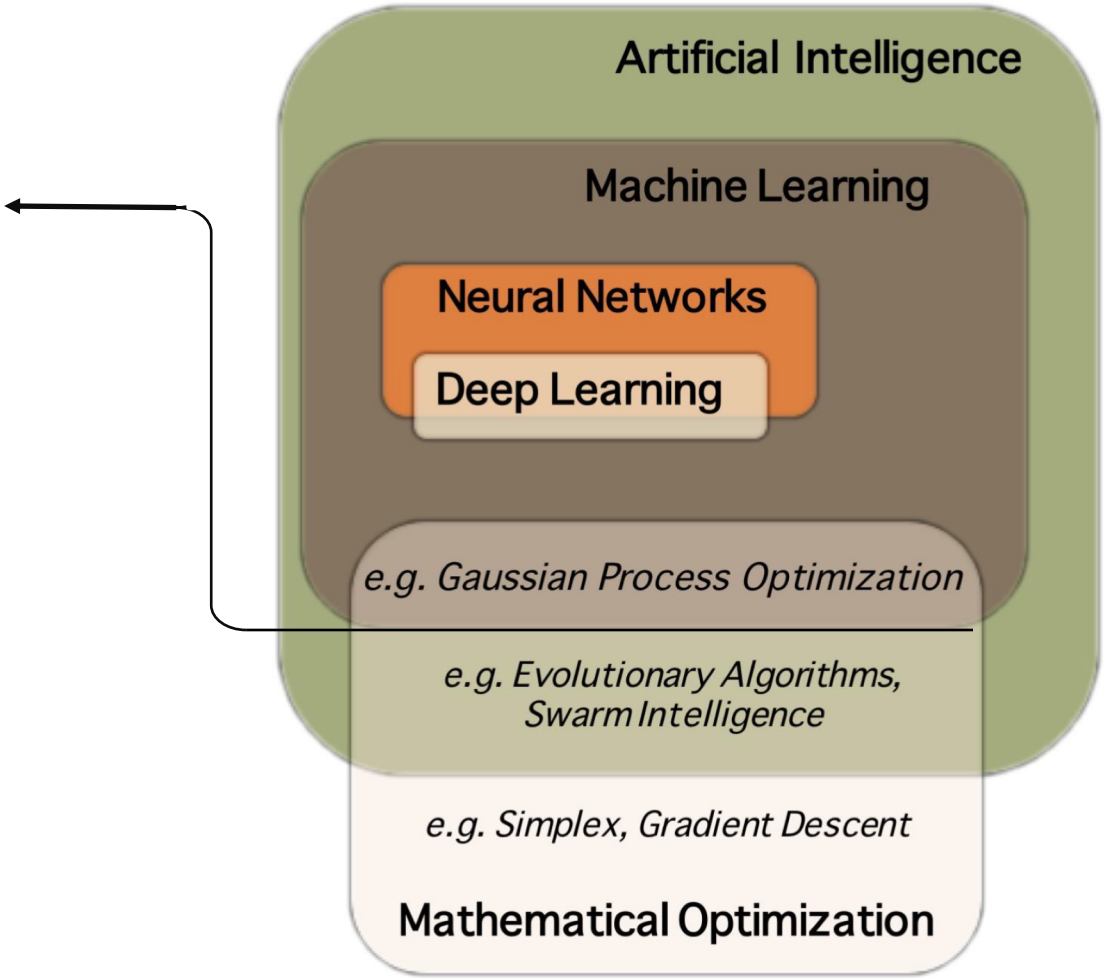


Control / Tuning

SLAC

Can also use ML to exploit learned representations of the system to inform the search

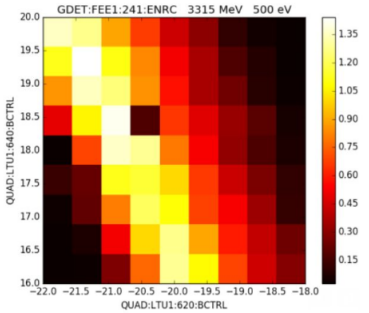
(model predictive control, Bayesian optimization, deep reinforcement learning, warm starts from ML models, inverse ML models)



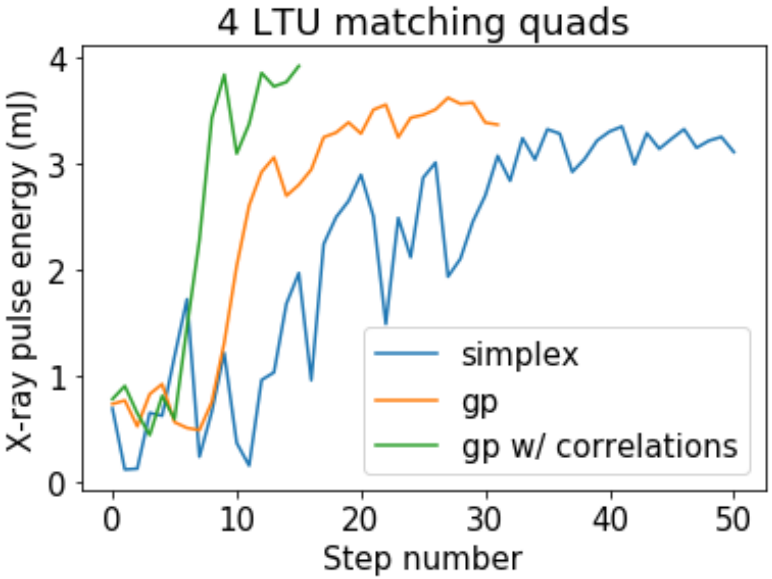
Control / Tuning: Bayesian Optimization

SLAC

- Bayesian Optimization with Gaussian Processes (GP) applied to FEL tuning at LCLS
  - → *tune quadrupole magnets to maximize FEL pulse energy*
- Incorporated physics correlations into kernel for GP model
- Recently the same tools were applied to SPEAR3 with minimal overhead (a few days of work)

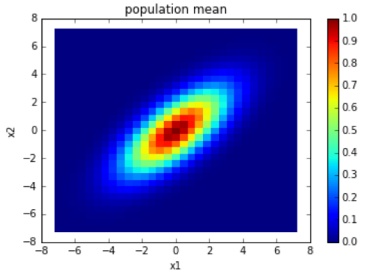


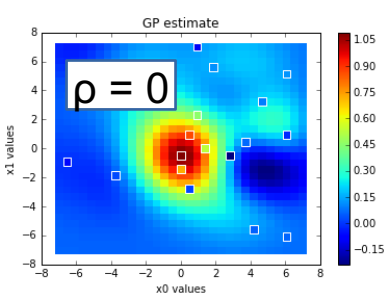
Measured FEL: adjacent quads are anticorrelated

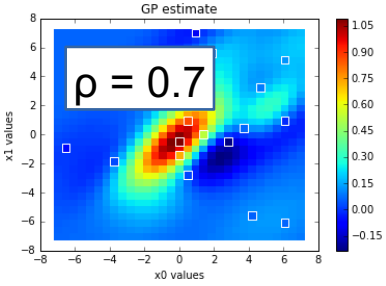


ground truth →

widths = 2
ρ = 0.7







*Kernel correlation improves regression on the **same samples***

figures: J. Duris, D. Kennedy

Control / Tuning: Safety Constraints

SLAC

Don't just want to maximize FEL energy → we have other requirements

- pulse energy briefly drops below certain level → *angry users!*
- beam losses go above a certain threshold → *damage machine!*

Add these requirements as safety constraints

Has been developed by ETH Zurich and tested experimentally at SwissFEL

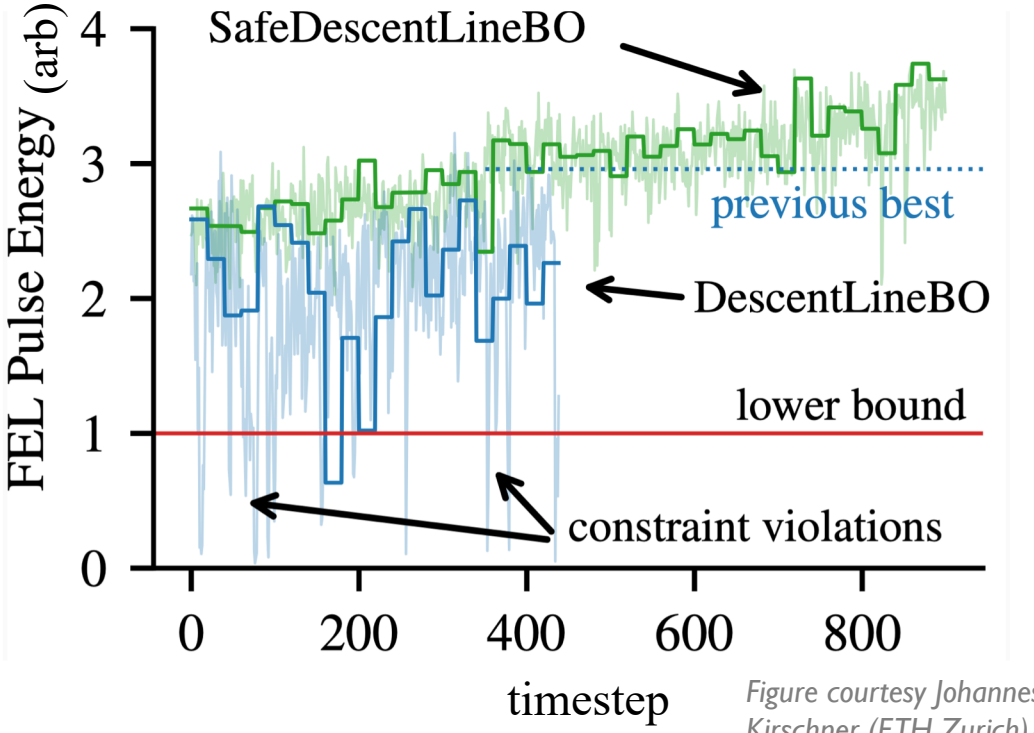


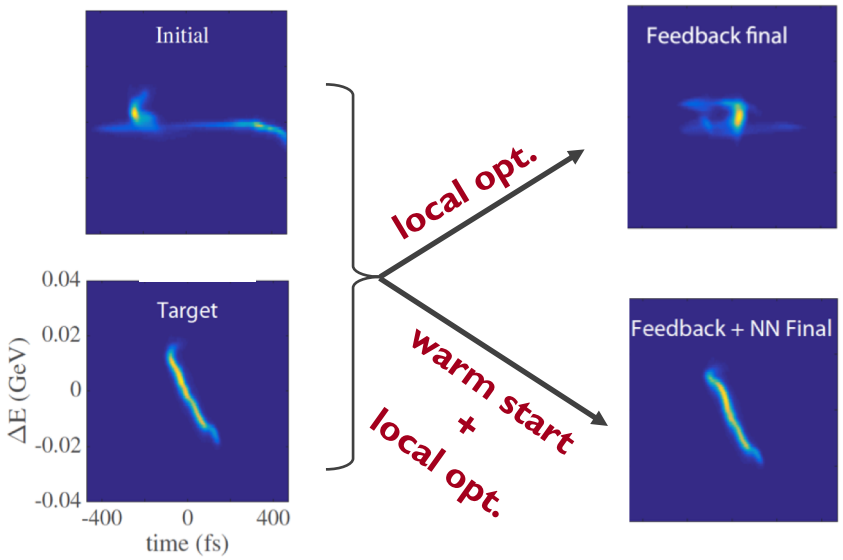
Figure courtesy Johannes Kirschner (ETH Zurich)

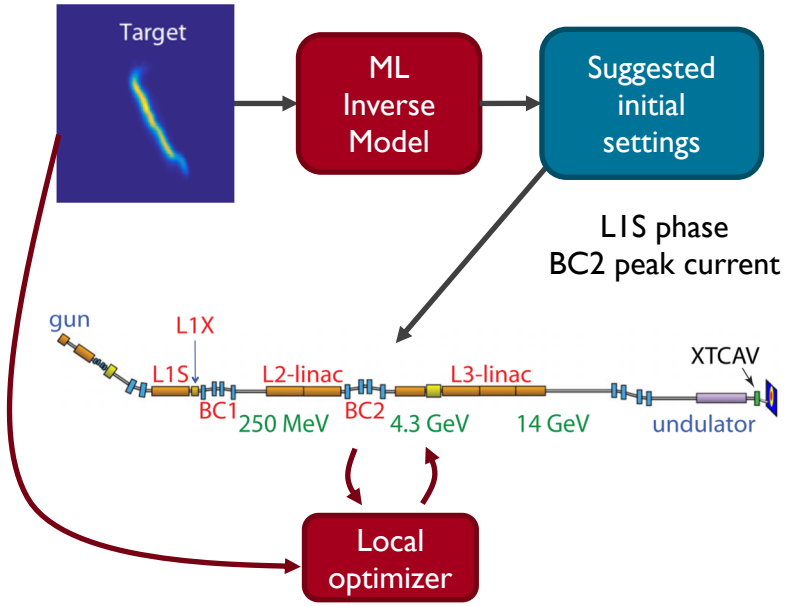
Control / Tuning: warm start with inverse models

A. Scheinker, A. Edelen, D. Bohler, C. Emma, A. Lutman., PRL 121, 044801 (2018)
earlier warm start work: A.L. Edelen, J.P. Edelen, et al., FEL '17

SLAC

- Often have to switch between user requests quickly
- Use inverse model to give rough suggested settings → then fine-tune with local optimizer
- Use a NN to map image to settings



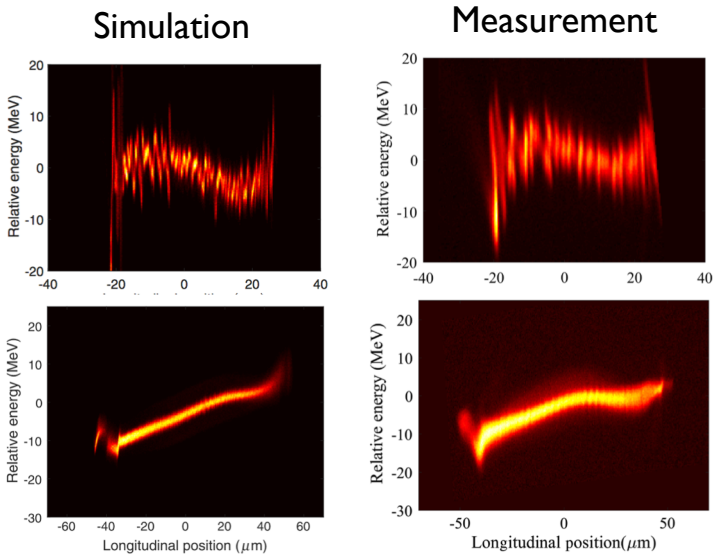


Local optimizer alone was unable to converge → **able to converge after initial settings from neural network**

Faster, more accurate machine models

SLAC

Accelerator simulations that include nonlinear + collective effects are powerful tools…



J. Qiang, et al., PRSTAB30, 054402, 2017

"10 hours on thousands of cores at the NERSC"

… but they are computationally expensive and don't always match the machine well

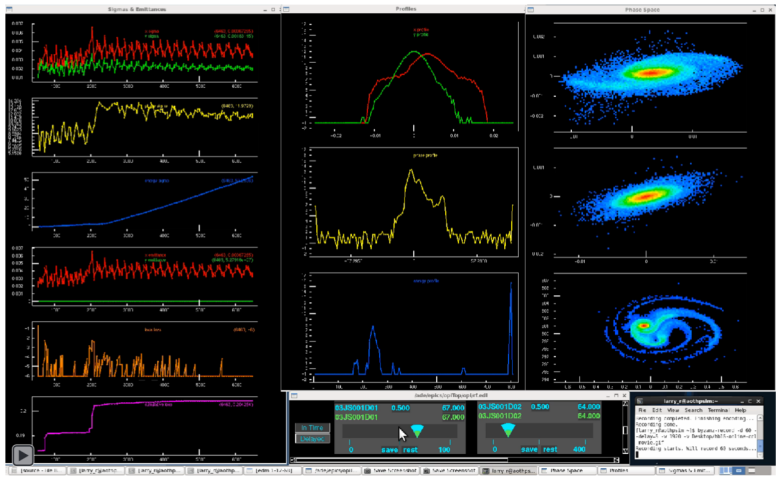
Prohibits use as an online model (e.g. diagnostic / control applications)
Impedes offline start-to-end optimization and control prototyping
Often takes much effort to replicate real machine behavior

Faster, more accurate machine models

SLAC

Accelerator simulations that include nonlinear + collective effects are powerful tools…

… but they are computationally expensive and don't always match the machine well



e.g. GPU-accelerated HPSim at LANSCE (based on PARMILA)

One approach: faster modeling codes

Simpler models (tradeoff with accuracy)

analytic calculations — *e. g. J. Galambos, et al., HPPA5, 2007*

Parallelization and GPU-acceleration of existing codes

HPSim/PARMILA — *X. Pang, PAC13, MOPMA13*

elegant — *I.V. Pogorelov, et al., IPAC15, MOPMA035*

Improvements to modeling algorithms

Lorentz-boosted frame — *J.-L. Vay, Phys. Rev. Lett. 98 (2007) 130405*

X. Pang, et al., PAC13, MOPMA13
L. Rybarcyk, HB2016, WEPM4Y01

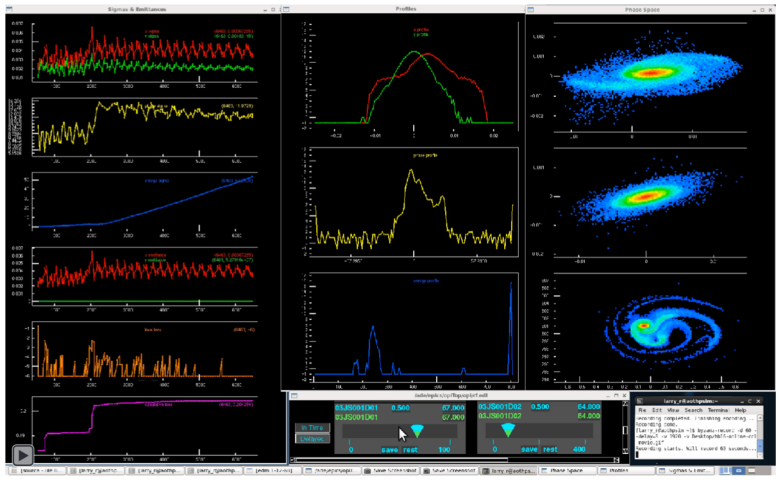
X. Pang, L. Rybarcyk, CPC185, is. 3 (2014)
X. Pang, IPAC15, WEXC2

Faster, more accurate machine models

SLAC

Accelerator simulations that include nonlinear + collective effects are powerful tools…

… but they are computationally expensive and don't always match the machine well



e.g. GPU-accelerated HPSim at LANSCE (based on PARMILA)

One approach: faster modeling codes

Execution often still isn't so fast (sec – mins)

Can require HPC resources

Still not easy to replicate machine behavior!

X. Pang, et al., PAC13, MOPMA13
L. Rybarcyk, HB2016, WEPM4Y01
X. Pang, L. Rybarcyk, CPC185, is. 3 (2014)
X. Pang, IPAC15, WEXC2

Faster, more accurate machine models

SLAC

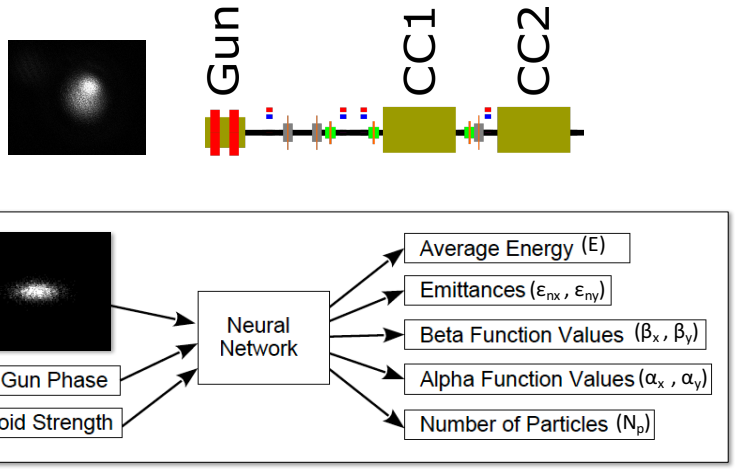
Accelerator simulations that include nonlinear + collective effects are powerful tools…

… but they are computationally expensive and don't always match the machine well

Initial examples from FAST injector at Fermilab:

PARMELA with 2-D space charge: ~ 20 minutes

Neural network: ~ a millisecond



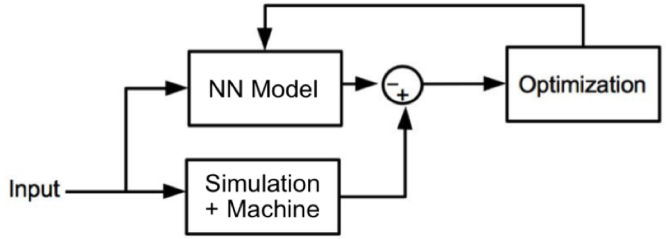
All mean absolute errors between 0.9% and 3.1% of the parameter ranges

Complementary approach: ML model

Once trained, neural networks can execute quickly

Train on sparse sample from high-fidelity simulations

Train on measured data

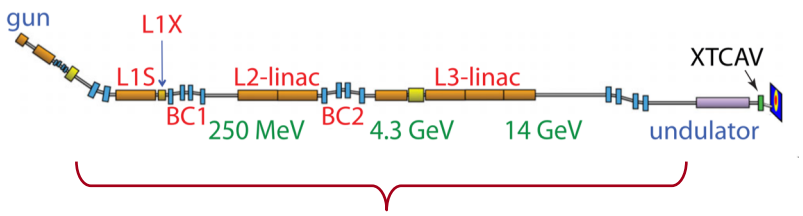


Faster, more accurate machine models: LCLS

SLAC

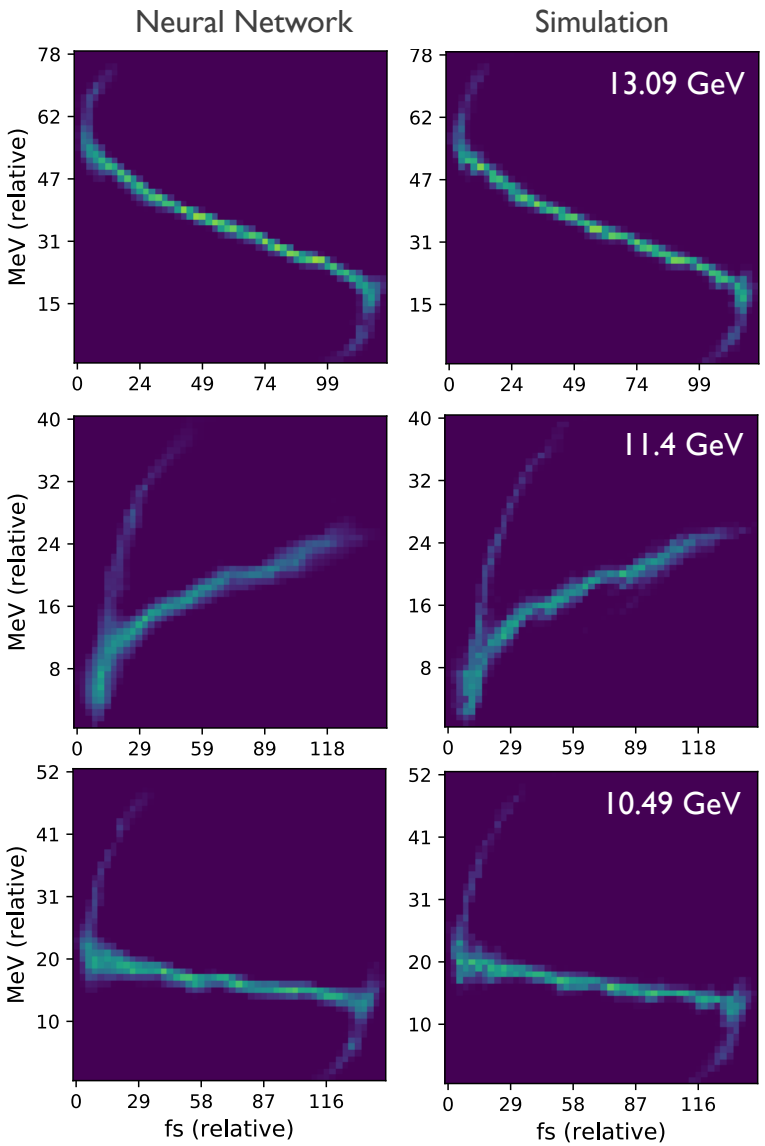
Wide scan of 6 settings for LCLS in Bmad

Variable	Min	Max	Nominal	Unit
L1 Phase	-40	-20	-25.1	deg
L2 Phase	-50	0	-41.4	deg
L3 Phase	-10	10	0	deg
L1 Voltage	50	110	100	percent
L2 Voltage	50	110	100	percent
L3 Voltage	50	110	100	percent



LCLS simulated with CSR, wakefields, space charge

- Trained neural network to predict 25 scalar outputs ($\sigma_{x,y,z}$ $\varepsilon_{x,y}$ $\sigma_{x',y'}$ σ_E *etc…*) and longitudinal phase space at the undulator entrance
- Good agreement with simulation (and **10^6x faster execution**)



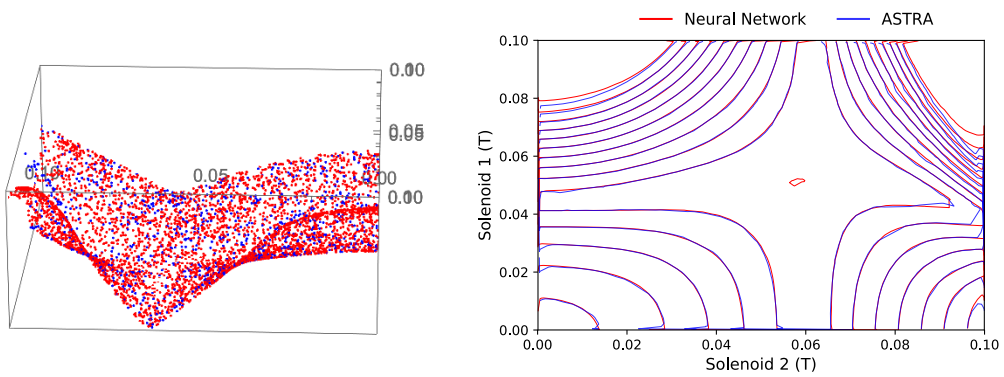
Faster, more accurate machine models: LCLS-II Injector

SLAC

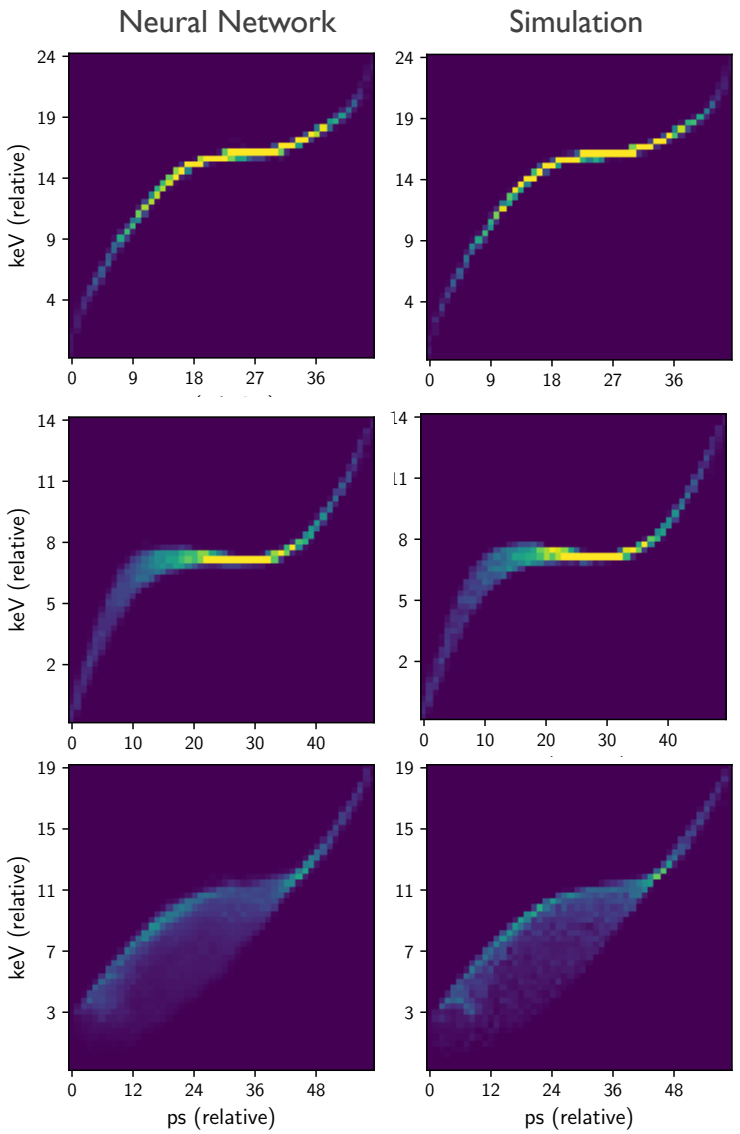
Wide scan of 5 settings in ASTRA

Variable	Min	Max	Nominal	Unit
Gun Phase	-10	10	-6.6	deg
Solenoid 1	0	0.1	0.06	T
Solenoid 2	0	0.1	0.03	T
Buncher Amplitude	0	2	1.80	MV/m
Buncher Phase	-100	-60	-80.3	deg

- Trained NN to predict 12 scalar outputs and longitudinal phase space (LPS)
- NN in good agreement with simulation (and **10^5x faster**)



Example σ_x surface from 2D scan, verified with ASTRA

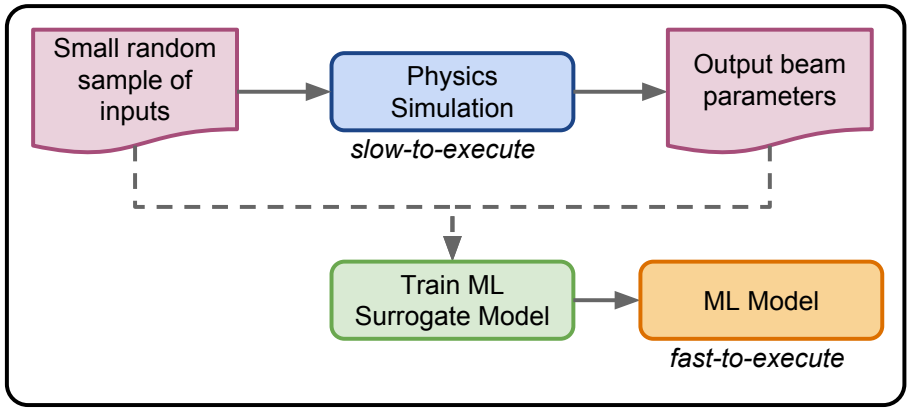


A. Edelen, N. Neveu, A. Adelmann, Y. Huber, M. Frey
https://arxiv.org/abs/1903.07759

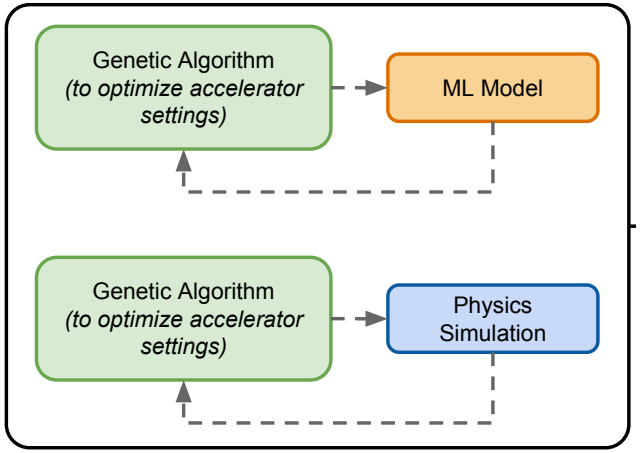
How reliably can we use these models with multi-objective optimization?

SLAC

Generate ML Model using Sparse Random Sample

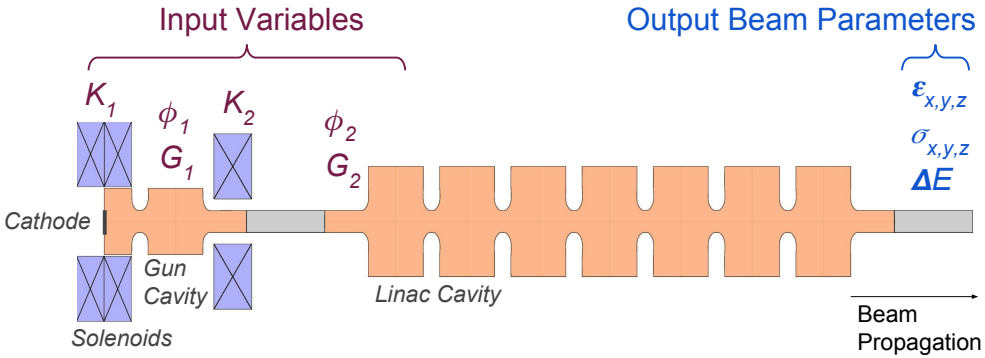


Run GA on ML Model and Physics Simulation

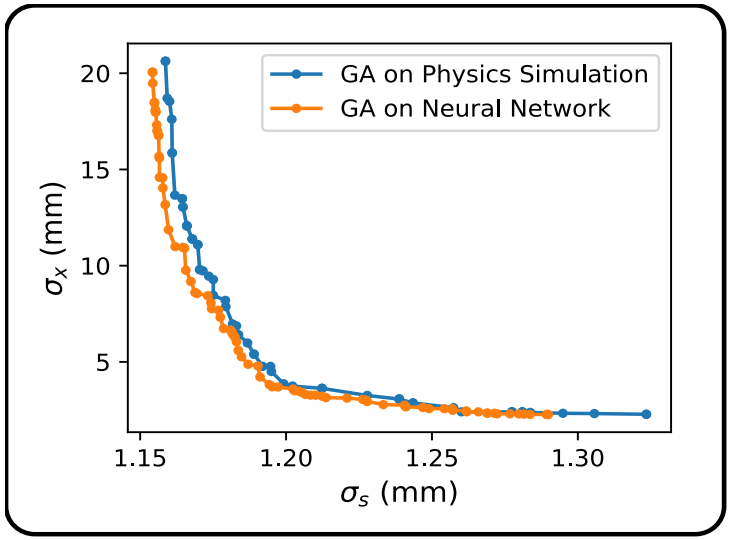


Test Case with Existing Data: Argonne Wakefield Accelerator Injector

OPAL simulation (PIC) : 3D space charge 3D field maps	NSGA-II for optimization: 200 generations ~600 individuals	500 random points for training



Compare Resulting Pareto Fronts



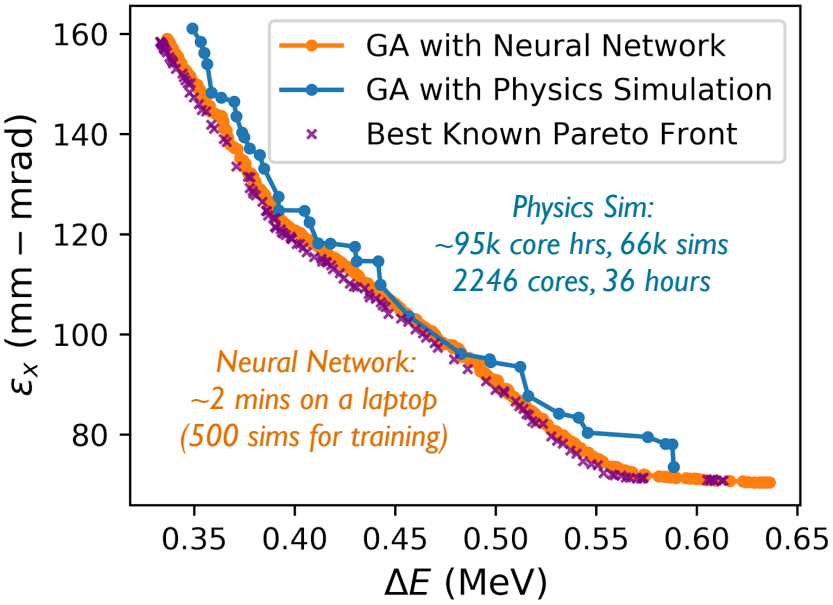
How reliably can we use these models with multi-objective optimization?

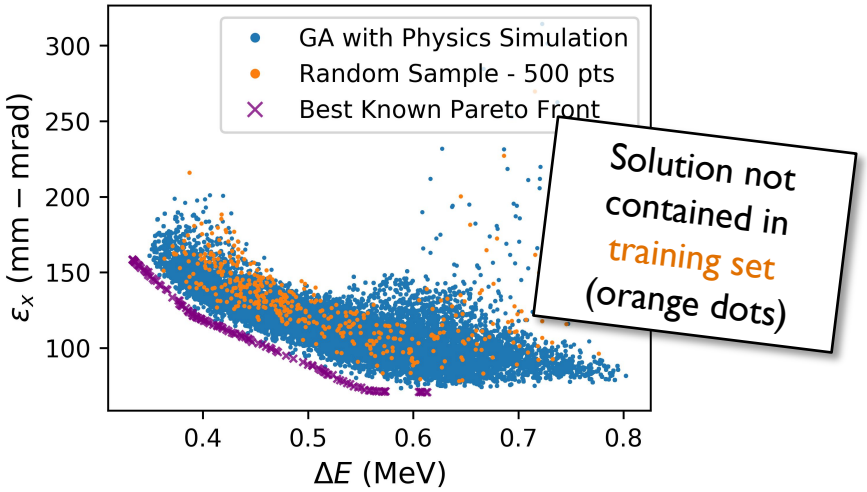
SLAC

Examined with PIC sims of the AWA injector:

MOGA solution with 6 inputs, 7 objectives required **~130x fewer simulation evaluations**

Surrogate model has **10^6 x faster execution**





In terms of time-to-solution:

~6.4 mins on 8 cores to make 500-point training data

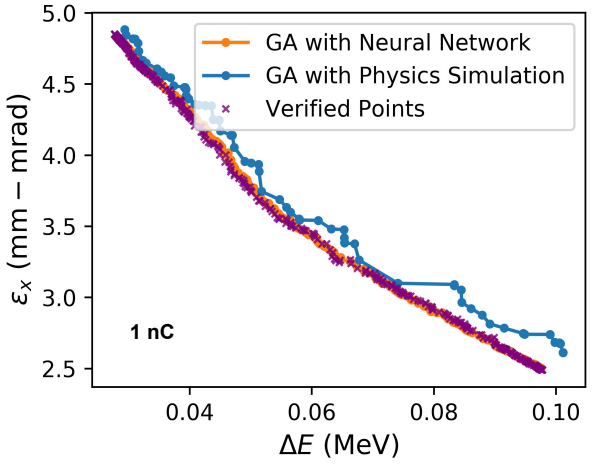
~10 minutes to train on a laptop

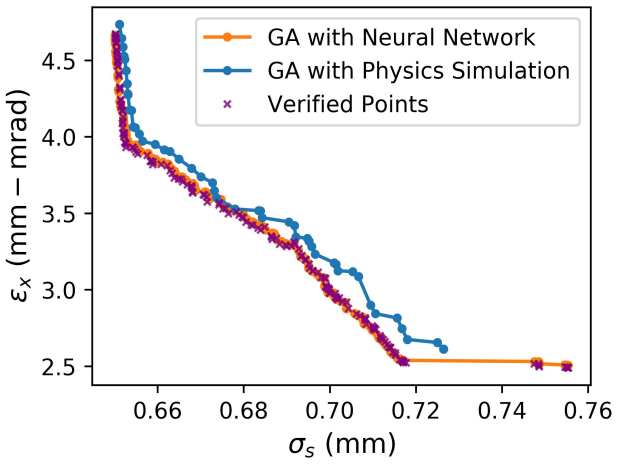
~2 minutes to do optimization on a laptop

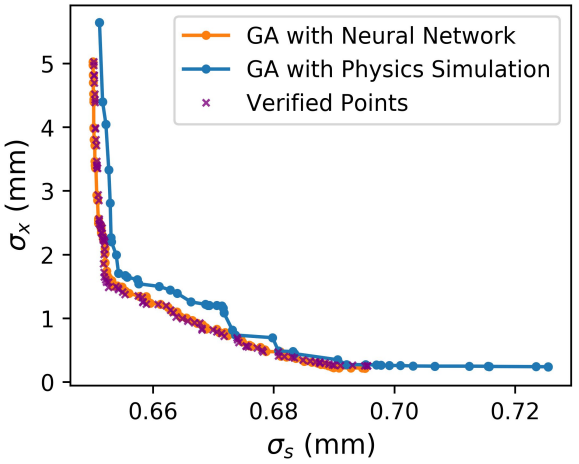
How reliably can we use these models with multi-objective optimization?

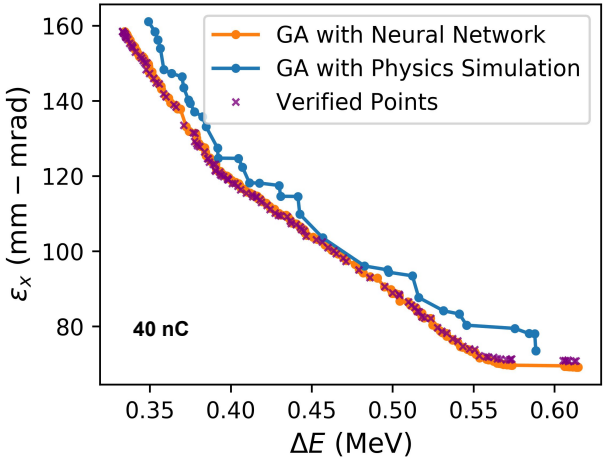
SLAC

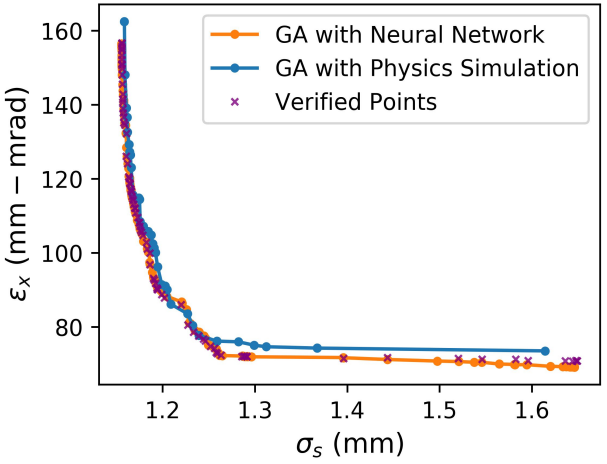
Similar results for other 2D Pareto fronts…

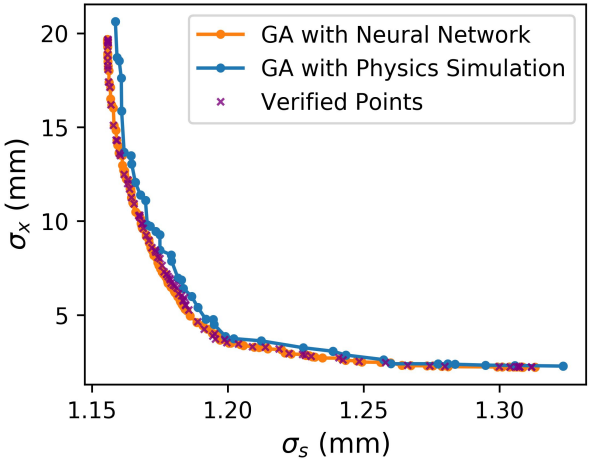












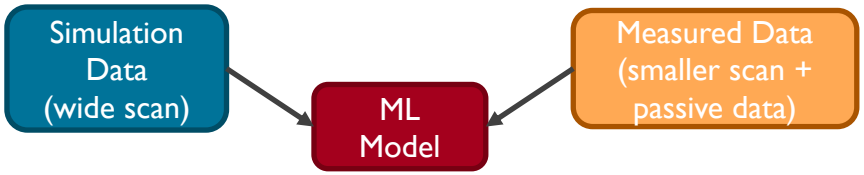
A. L. Edelen, J.P. Edelen, D. Edstrom, et al. IPAC18, WEPAF040
A. L. Edelen, J.P. Edelen, D. Edstrom, et al. NAPAC16, TUPOA51

Can we bridge the gap between our simulations and empirical machine behavior?

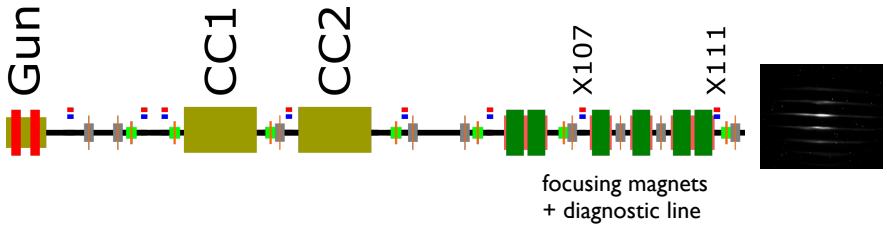
SLAC

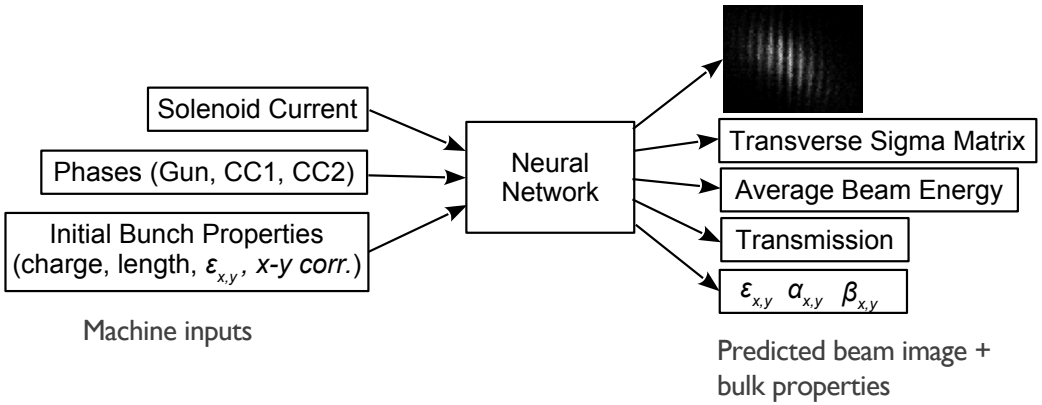
Poor agreement between physics simulation and measured data

Can't do a full parameter scan on machine (cost / time)



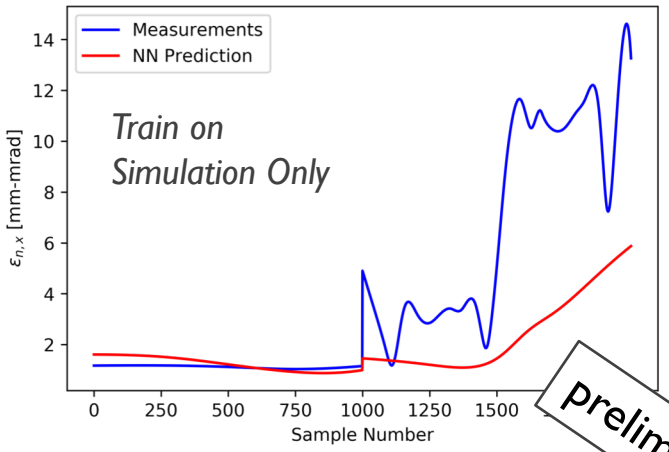
Can we pre-train in simulation and update with measured data?



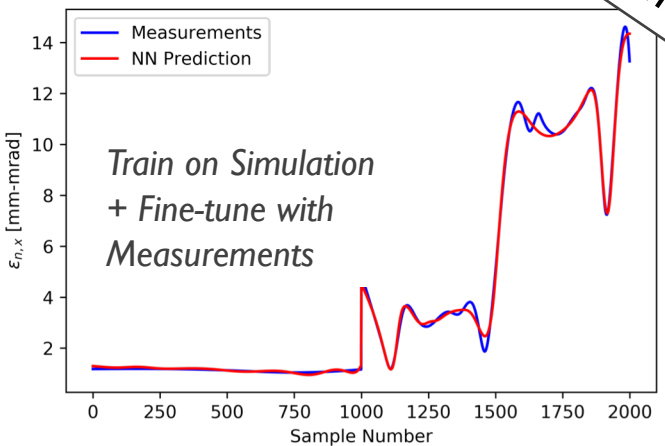


Initial results from study of injector systems look promising

→ *need to investigate strategies for doing this routinely and at larger scale*



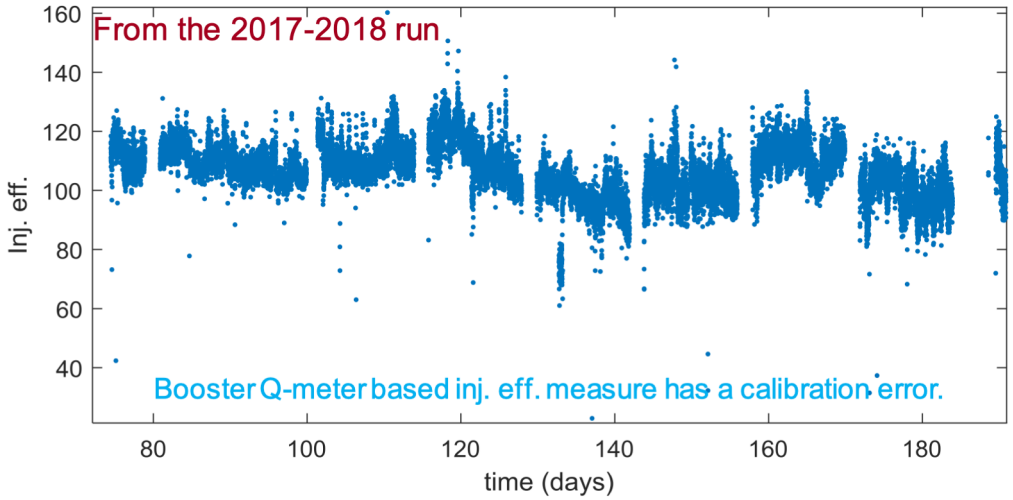
preliminary



Improve system understanding: learn about machine sensitivities

SLAC

- SPEAR3 storage ring injection efficiency varies → *trajectory feedback settings are frequently optimized to compensate*

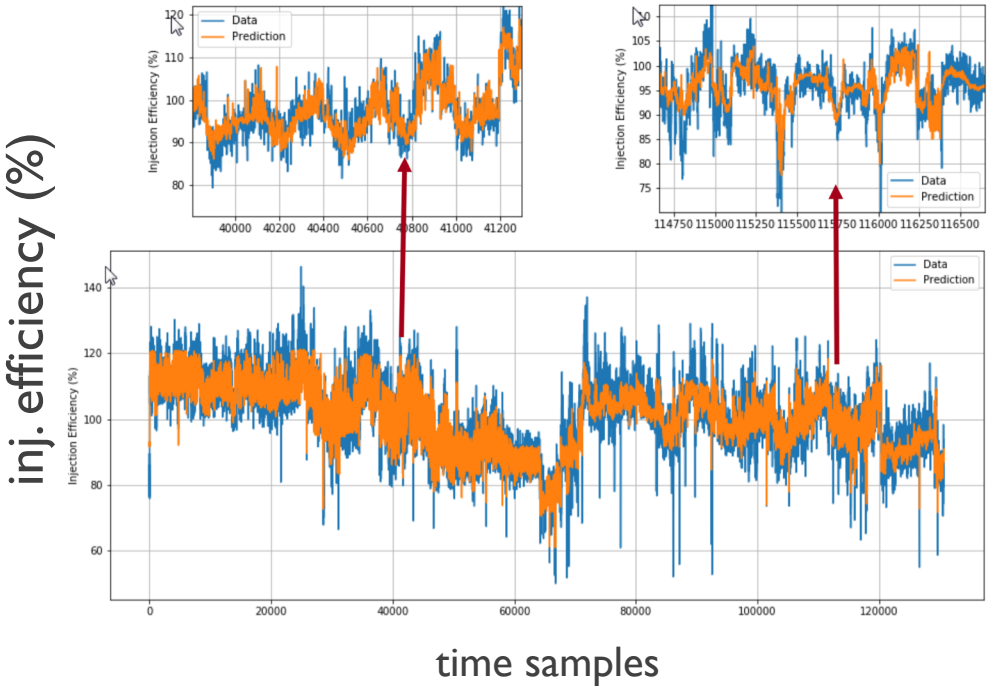


Improve system understanding: learn about machine sensitivities

SLAC

- SPEAR3 storage ring injection efficiency varies → *trajectory feedback settings are frequently optimized to compensate*

- Use NN model to discover what is driving the change *(i.e. find unanticipated parameter dependencies)*



Improve system understanding: learn about machine sensitivities

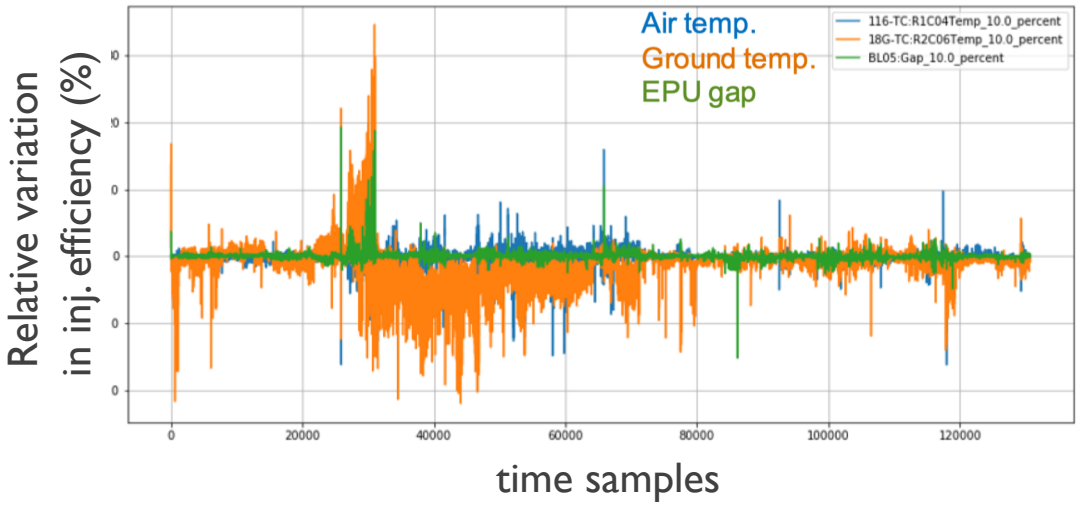
SLAC

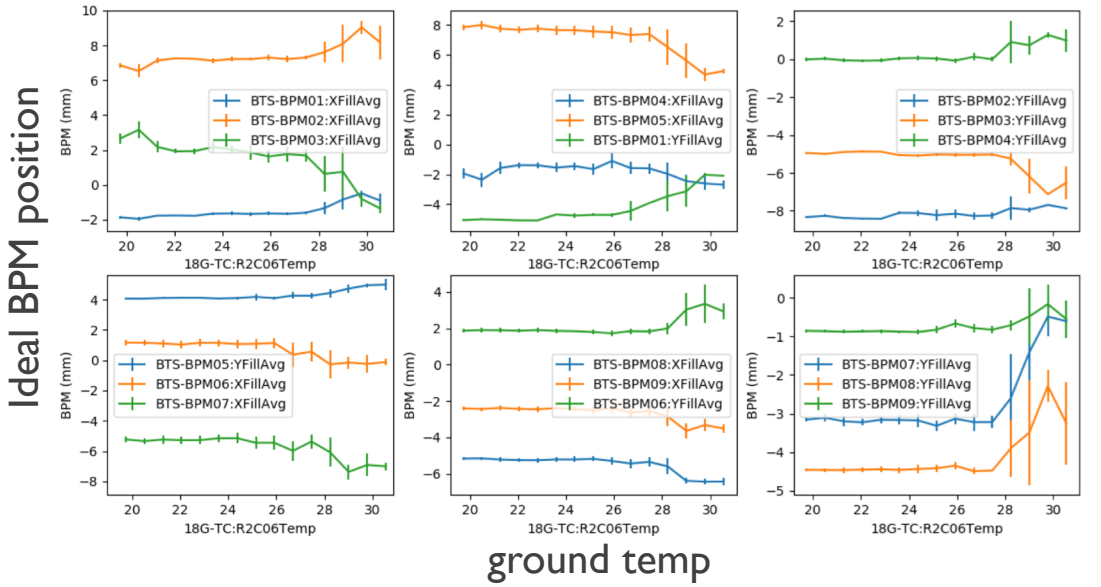
- SPEAR3 storage ring injection efficiency varies → *trajectory feedback settings are frequently optimized to compensate*

- Use NN model to discover what is driving the change *(i.e. find unanticipated parameter dependencies)*

→ *Found ground temperature was a significant factor*

→ *Can now use to predict ideal orbit given ground temperature*





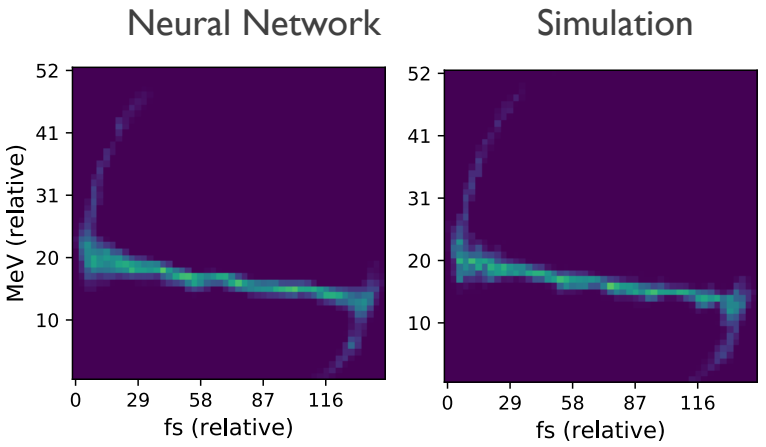
Major Use Cases for ML in Particle Accelerators

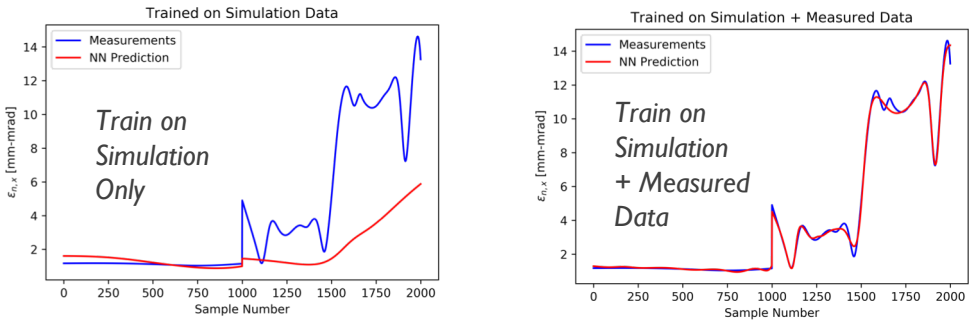
SLAC

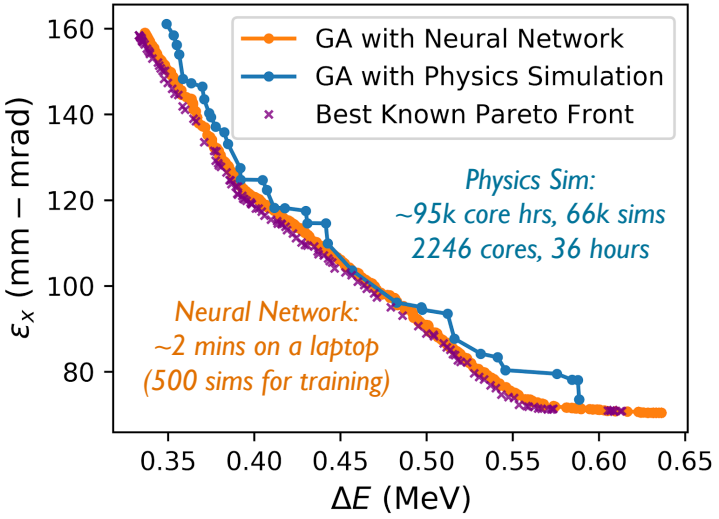
- Detecting / classifying / predicting unwanted changes or failures
  — *failing beam position monitors, cavity quenches*
- Getting more useful information out of complicated machine signals
  — *images, waveforms, etc.*
- System optimization and fast experiment setup
  — *need solutions for standard setups and previously unseen setups*
- System modeling for use in design, online modeling, and model-based control
- Facilitate improved understanding of factors that impact performance
  — *physics insight, machine sensitivities, hidden variables etc.*
- High throughput data analysis / rejection (e.g. LCLS user side)

Many Open Questions to Address…

- Robustness / model uncertainty e.g. *flag when not to trust ML algorithm – switch to physics models or standard control (e.g. outside training range, aberrant conditions)*
- Strategies for online retraining *adapt to new configurations / part replacements / drift*
- How best to combine simulation and measured data
- Scaling to higher dimension + problem complexity, wider range of conditions
- Which combinations of methods (ML and non-ML) will work best for different kinds of problems







Final Thoughts

- Growing community, three recent workshops:

Intelligent Controls for Particle Accelerators
Jan. 2018 at Daresbury Lab
Agenda/Talks: https://tinyurl.com/y9rg3uht

Machine Learning for Particle Accelerators
Feb. 2018 at SLAC, Feb. 2019 at PSI
Agenda/Talks '18: https://tinyurl.com/y988njbl
Agenda/Talks '19: https://tinyurl.com/y3u3vlpo
Tutorials '19: https://tinyurl.com/yypc9lq7

- ML is a complementary approach to existing techniques and is extremely flexible
- Many opportunities to use ML to improve accelerator performance
- Relatively simple methods can be readily put to use
- ***ML is not a panacea!*** *good workflows + data are essential, and many simpler techniques are not put to full use by the community (e.g. model predictive control with simple models, LiTrack virtual diagnostic by A. Scheinker and S. Gessner)*
- Still exploring the boundaries of usefulness/reliability and tradeoff with time investment

Acknowledgments to many colleagues / collaborators involved in ML examples shown!

Daniel Ratner (SLAC)

Xiaobiao Huang (SLAC)

Joe Duris (SLAC)

Claudio Emma (SLAC)

Nicole Neveu (SLAC)

Christopher Mayes (SLAC)

Xinyu Ren (SLAC)

Xiao Zhang (SLAC)

Jane Shtalenkova (SLAC)

Adi Hanuka (SLAC)

Hugo Slepicka (SLAC)

Faya Wang (SLAC)

Lauren Alsberg (SLAC)

Dorian Bohler (SLAC)

Glen White (SLAC)

Dylan Kennedy (SLAC)

Tim Maxwell (SLAC)

William Colocho (SLAC)

Matt Gibbs (SLAC)

Alex Halavanau (SLAC)

Eric Hemsing (SLAC)

Aashwin Mishra (SLAC)

Alan Heirich (SLAC)

Yuantao Ding (SLAC)

Gabriel Marcus (SLAC)

Alberto Lutman (SLAC)

Jonathan Edelen (RadiaSoft, LLC)

Dean "Chip" Edstrom Jr. (Fermilab)

Andreas Adelmann (PSI)

Matthias Frey (PSI)

Yannick Huber (PSI/ETH Zurich)

Raffaele Campanile (INFN)

Alex Scheinker (LANL)

Pietro Musumeci (UCLA)

Eric Cropp (UCLA)

Paul Denham (UCLA)

Sandra Biedron (UNM)

Jinhao Ruan (Fermilab)

Philippe Piot (Fermilab / NIU)

Presentation also included work from:

Anna Solopova (Jlab)

Elena Fol (CERN)

Johannes Kirschner (PSI/ETH Zurich)