Virtual Diagnostics for High Brightness Beams

Physics and Applications of High Brightness Beams Workshop
June 22, 2023

Auralee Edelen, Brendan O'Shea, Claudio Emma, Ryan Roussel,
Robbie Watt, Daniel Ratner, Chris Mayes *(SLAC)*
J.P. Gonzalez-Aguilera *(U. Chicago)*
John Power, Eri Wisniewski, Seongyeol Kim *(ANL)*

edelen@slac.stanford.edu

Motivation: Better Diagnostics Inherently Linked to Better Beams

Two major categories of need motivate machine learning (ML) enhanced tools in accelerators:

- **New fundamental capabilities in beam production and control:** attaining unprecedented beam parameters, finely-detailed customization and characterization of beams
- **Facility operations:** efficiency of tuning and quality of beam delivery for scientific users → increase science output, reduce time-to-discovery

General Accelerator R&D Program

Accelerator and Beam Physics Roadmap

DOE Accelerator Beam Physics Roadmap Workshop

September 6–8, 2022

Available here

ABP Grand Challenges

Intensity – *"How do we increase beam intensity by orders of magnitude?"*

Quality – *"How do we increase the beam phase space density by orders of magnitude?"*

Control – *"How do we measure and control the beam distribution down to the individual particle level?"*

Prediction – *"How do we develop predictive 'virtual particle accelerators'"*

Need advanced methods to obtain information about the beam at unprecedented levels of detail and speed

*→ **important for fine customization/control, user analysis, and improving physics models***

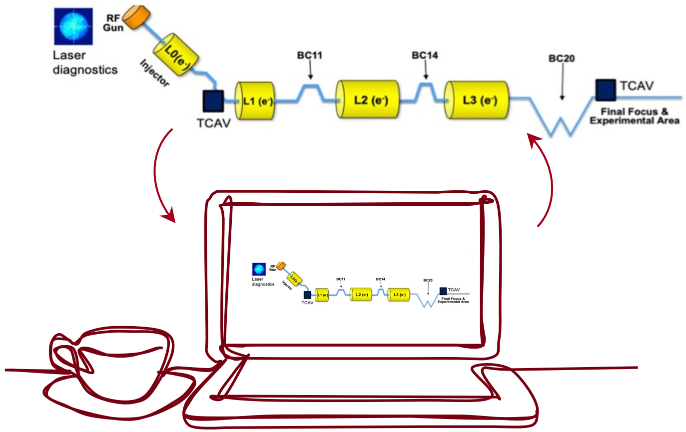
Virtual Diagnostics ←→ Virtual Accelerators

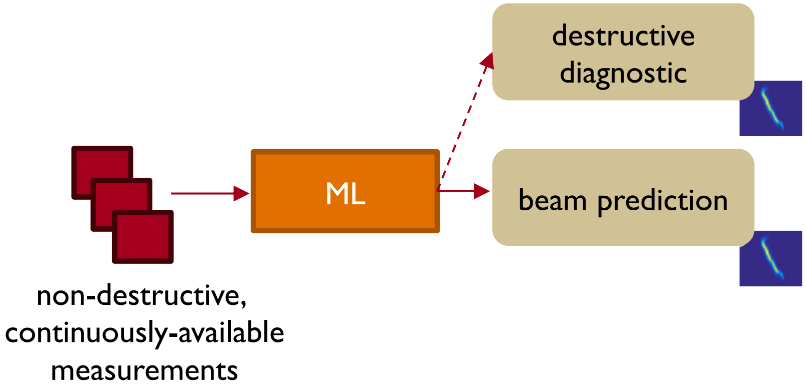
Many long-standing efforts to make **"virtual accelerators"** that closely match machine behavior

- Predict machine behavior that isn't directly accessible
- Related to the idea of "digital twins" when combined with tracking/adapting to changes in the system

A **"virtual diagnostic"** is an extension of this concept

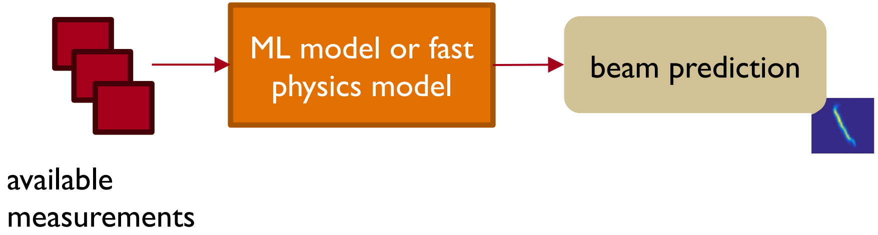
- Predict beam output in cases where a diagnostic does not exist, is destructive, or updates more slowly than desired
- Machine learning enables new capabilities in prediction → *do not need a physics model, just need sufficiently well-correlated measurements*





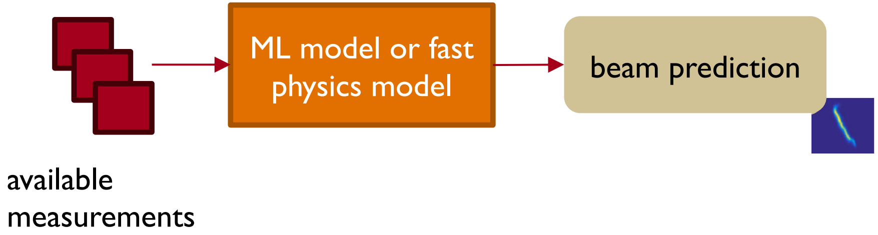
Virtual Diagnostics ←→ ML-Enhanced Diagnostics

Fast, detailed predictions of quantities that aren't continuously available

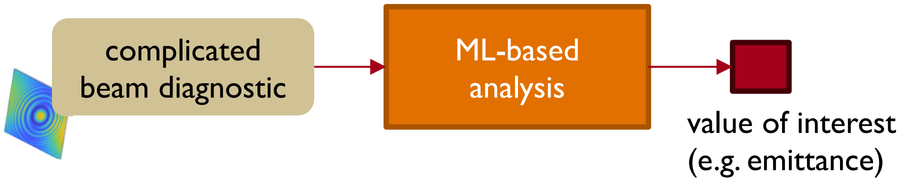


Virtual Diagnostics ←→ ML-Enhanced Diagnostics

Fast, detailed predictions of quantities that aren't continuously available

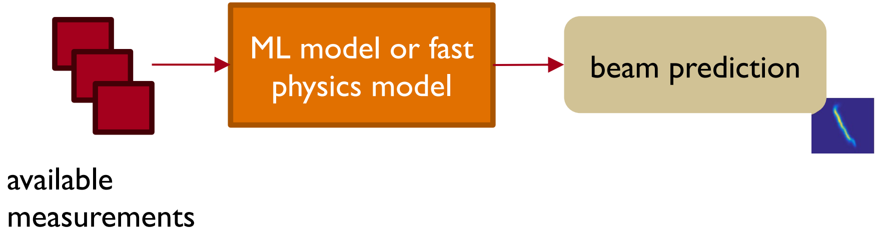


Fast analysis of complicated diagnostic output

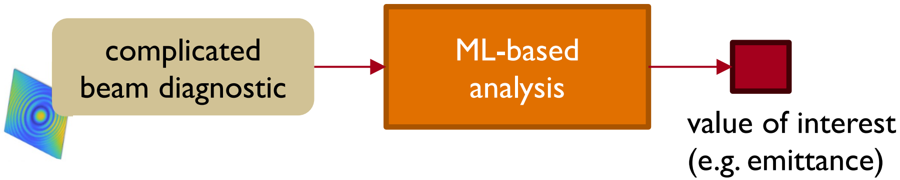


Virtual Diagnostics ←→ ML-Enhanced Diagnostics

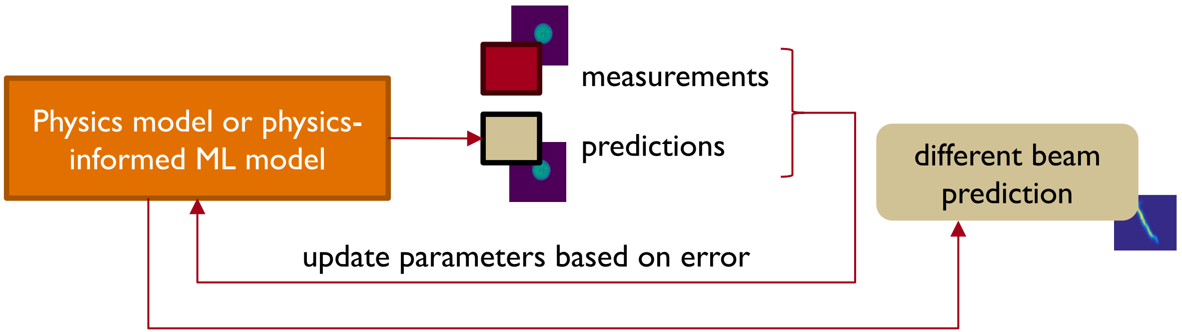
Fast, detailed predictions of quantities that aren’t continuously available



Fast analysis of complicated diagnostic output

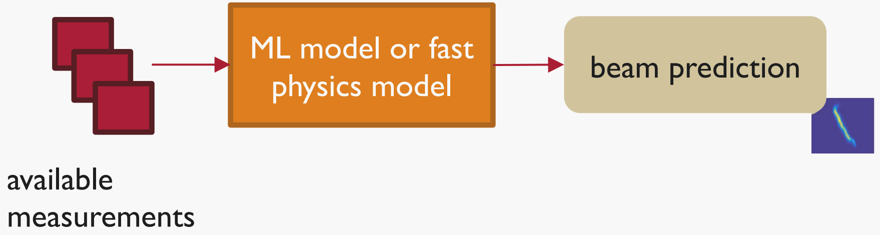


Active tuning of system models to infer unseen variables or beam behavior

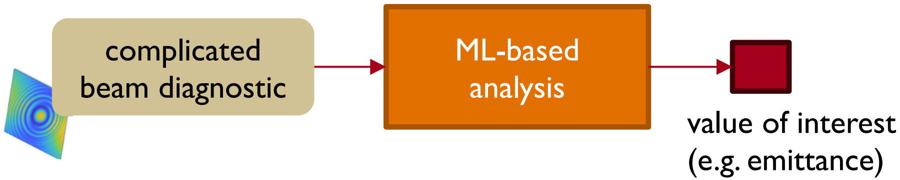


Virtual Diagnostics ←→ ML-Enhanced Diagnostics

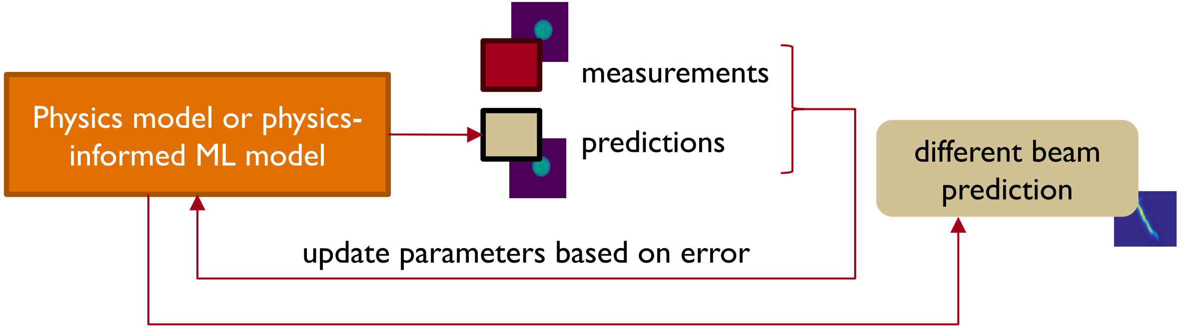
Fast, detailed predictions of quantities that aren’t continuously available



Fast analysis of complicated diagnostic output



Active tuning of system models to infer unseen variables or beam behavior

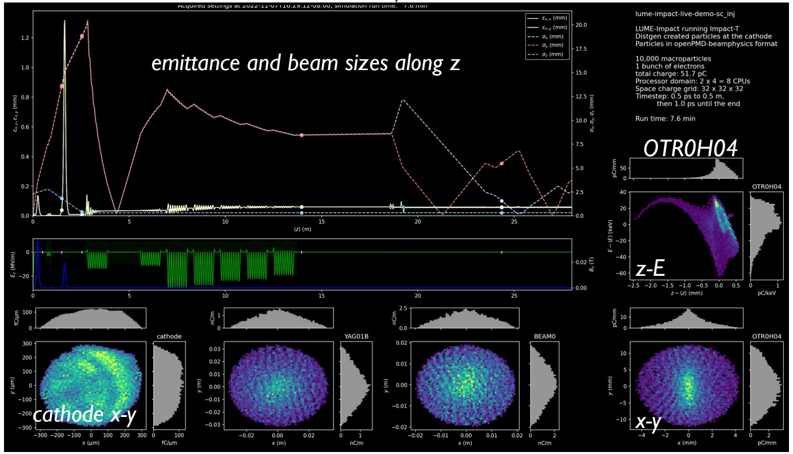


Online Models / Virtual Accelerators as a type of "Virtual Diagnostic"

Used combination of online physics simulation and Bayesian optimization algorithms to aid LCLS-II injector commissioning

Readings from machine via EPICS

injector settings, laser profile from VCC image

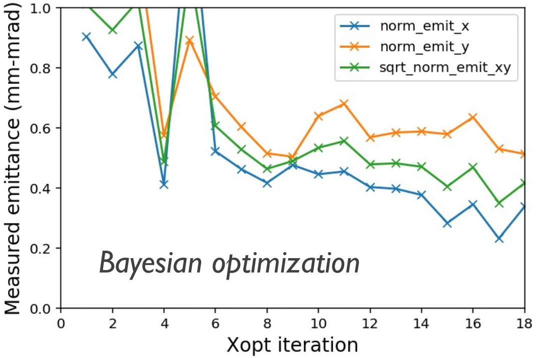


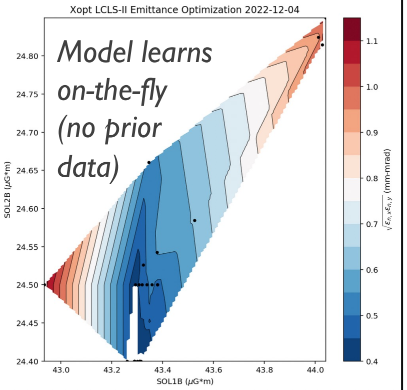
LCLS-II live sim: run on HPC and display in control room

Updates every 3-8 mins, space charge included, uses LUME-IMPACT

Adjust settings / ranges with insight from predictions

Hand over to ML-based optimization for fine tuning





Best emittance yet obtained during LCLS-II injector commissioning

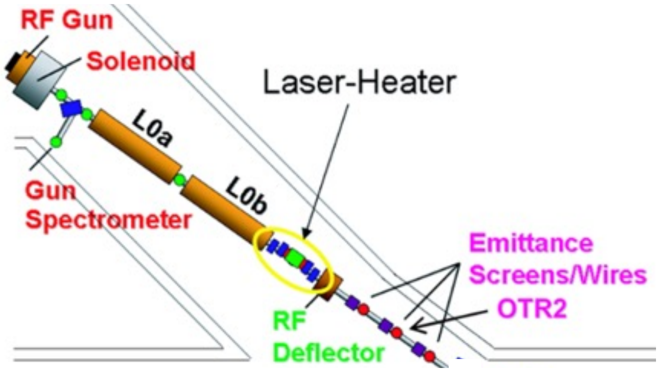
despite extensive previous hand-tuning

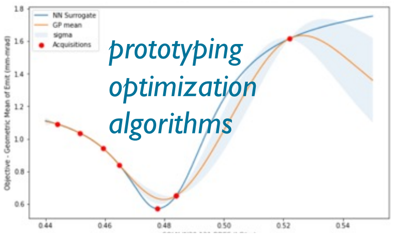
Physicists' intuition aided by detailed online physics model → simple example of how a "virtual accelerator" can aid tuning

HPC enables fundamentally new capabilities in what can be realistically simulated online

Example of Faster Execution with ML: LCLS Injector

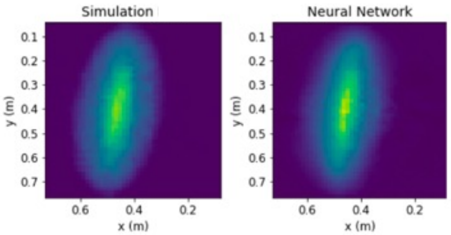
- ML models trained on detailed IMPACT simulations over entire valid range of injector settings and drive laser settings
- Several models with different combinations of output tailored to specific need *(phase space prediction, emittance/match, beam sizes, etc.)*
- Using to develop/prototype new algorithms before testing online (e.g. 20x speedup in emittance tuning: *https://arxiv.org/abs/2209.04587*)
- Will be deployed online for prediction of beam phase space and Twiss parameters

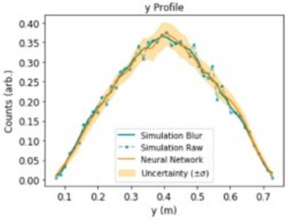




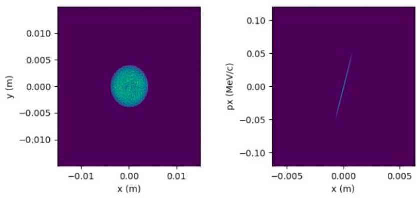
prototyping optimization algorithms

ML model matches simulation under interpolation

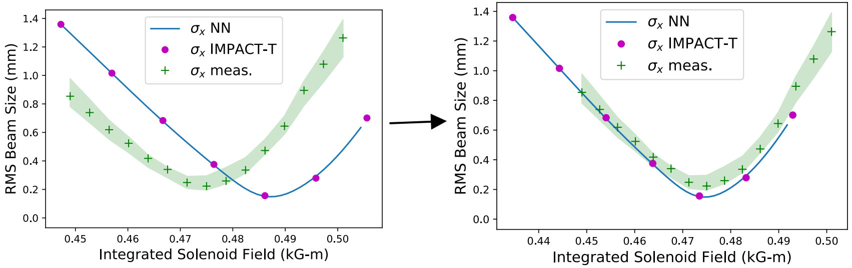




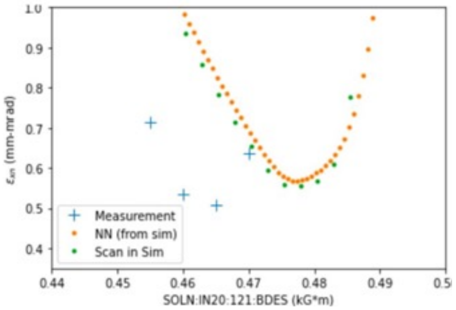
Simulation and ML model trained on it are qualitatively similar to measurements under interpolation



interactive model widget and visualization tools



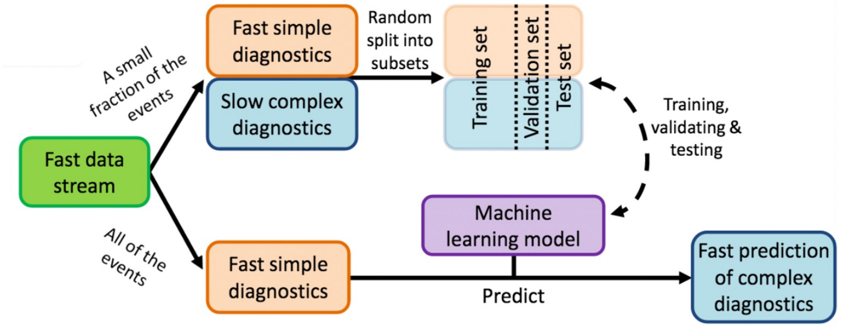
Automatic adaptation of models and identification of sources of deviation between simulations and as-built machine

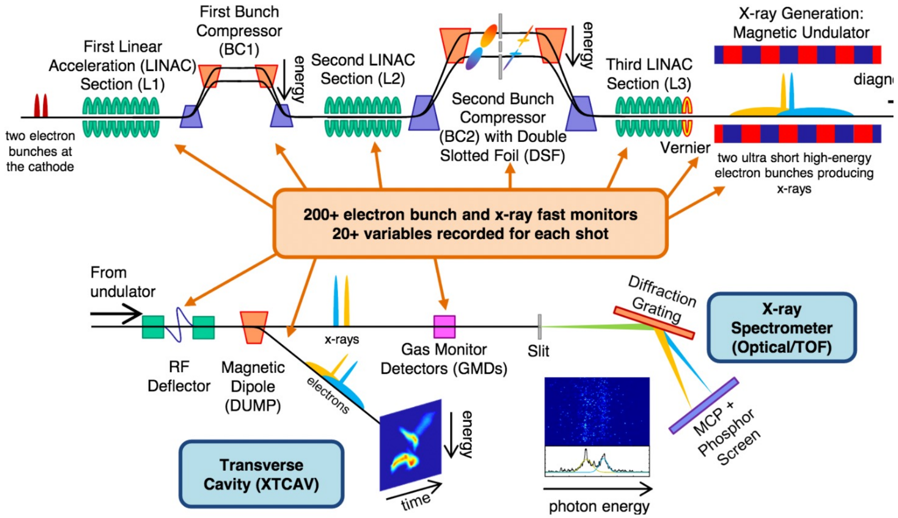


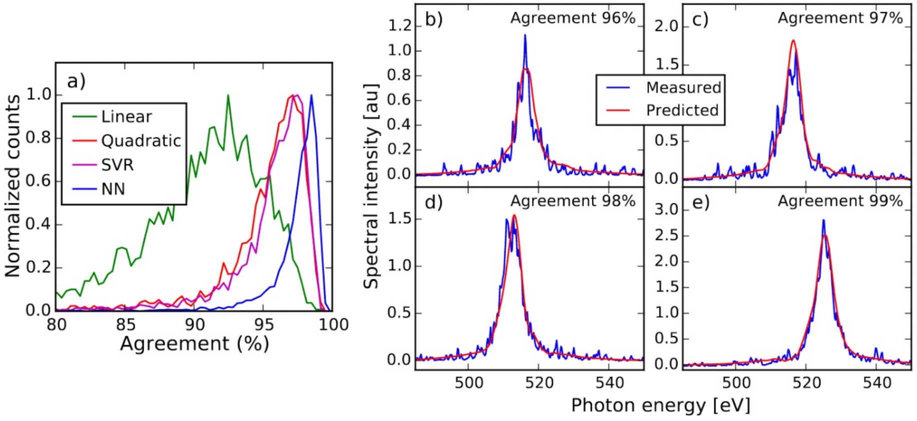
ML models trained on simulations and measurements have enabled fast prototyping of new optimization algorithms, facilitated rapid model adaptation under new conditions, and can directly aid online tuning and operator decision making

LCLS Virtual Diagnostic Example: Fill in Missing Shots

- Used correlations with simpler/faster diagnostics to fill in predictions for more complex, slowly-updating diagnostics
- Good agreement between predicted and measured spectra for neural network
- For LCLS-II, going to much higher repetition rates (MHz beam rate) → ongoing work in edge ML for fast data reduction and reconstruction



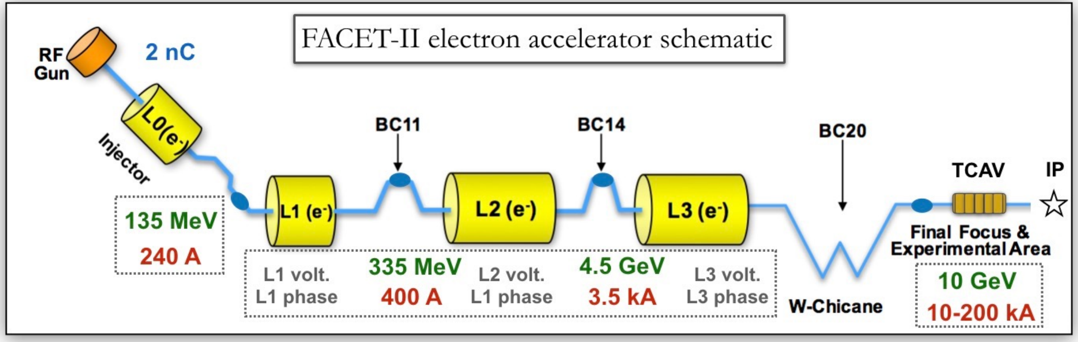




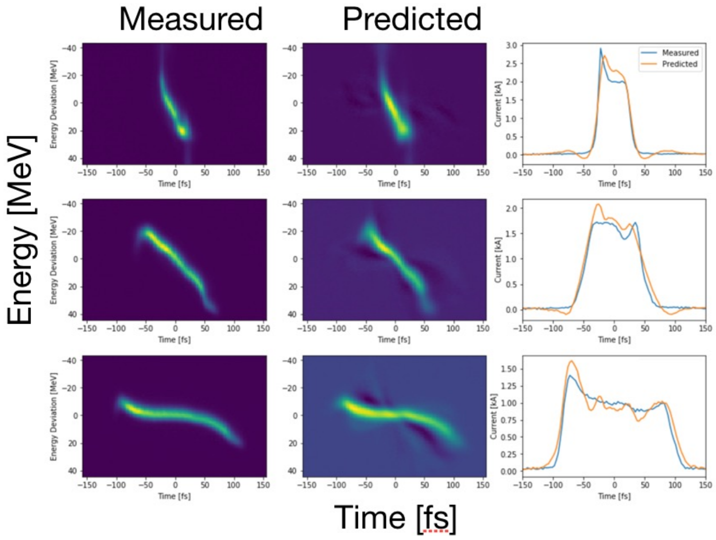
A. Sanchez-Gonzalez, et al., Nature Communications, (2017)

C. Emma

FACET-II Longitudinal Phase Space Virtual Diagnostic

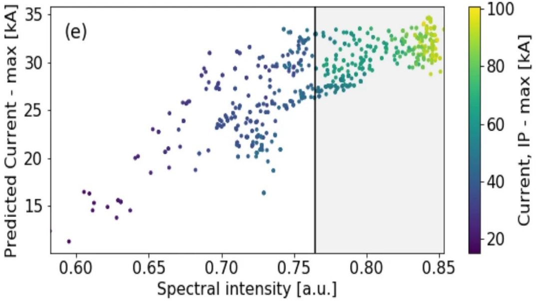


Initial experimental demonstration of LPS prediction with data from LCLS



*C. Emma, et al. – PRAB **21**, 112802 (2018)*

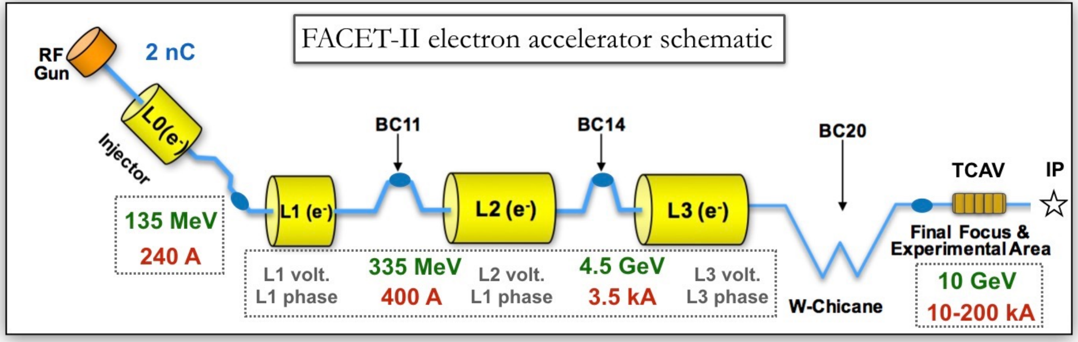
- LPS tuning critical for FACET science program
- LPS measurement is destructive → cannot be measured with TCAV and simultaneously used for experiments
- 'Virtual TCAV' gives non-destructive prediction of beam LPS in experimental area
- Simulations indicate incorporation of non-destructive spectral information could enable prediction beyond the TCAV resolution



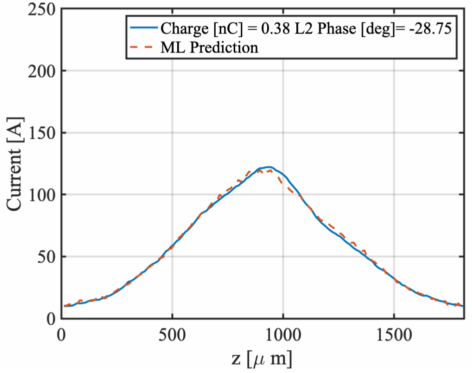
A. Hanuka et al, Sci. Reports 11, 2945 (2021)

C. Emma

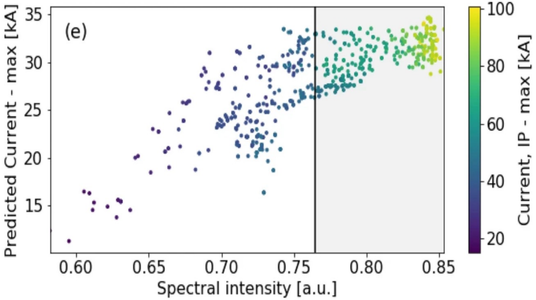
FACET-II Longitudinal Phase Space Virtual Diagnostic

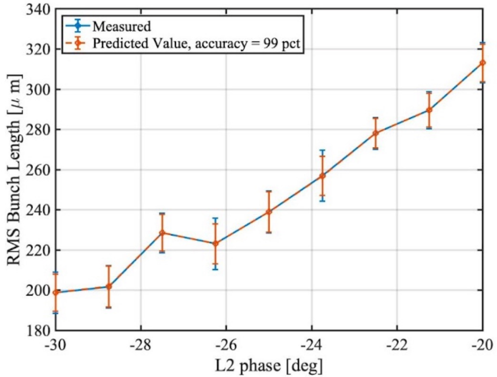


FACET-II experimental demonstration at low charge (0.5 nC)



- LPS tuning critical for FACET science program
- LPS measurement is destructive → cannot be measured with TCAV and simultaneously used for experiments
- 'Virtual TCAV' gives non-destructive prediction of beam LPS in experimental area
- Simulations indicate incorporation of non-destructive spectral information could enable prediction beyond the TCAV resolution





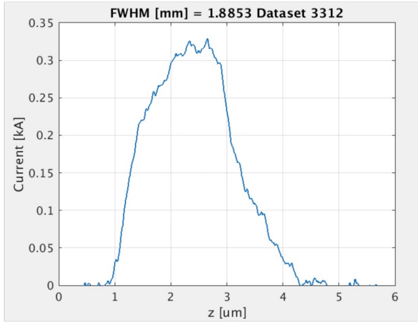
First experimental results demonstrate ML-based current profile + bunch length prediction at FACET-II

LPS Predictions at FACET-II Injector

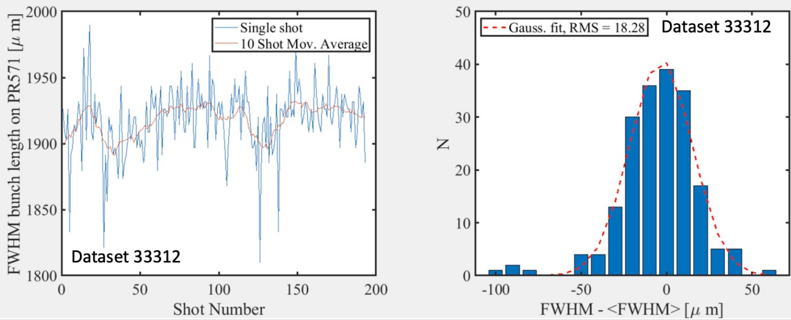
C. Emma

- TCAV used to measure current profile and characterize shot-to-shot current/bunch length variations in the injector
- Bunch length variations correlated with injector RF, magnet, laser parameters
- ML model used to predict changes in the bunch length from non-destructive inputs
- Non-destructive LPS diagnostic to be used for tuning/data analysis in upcoming runs

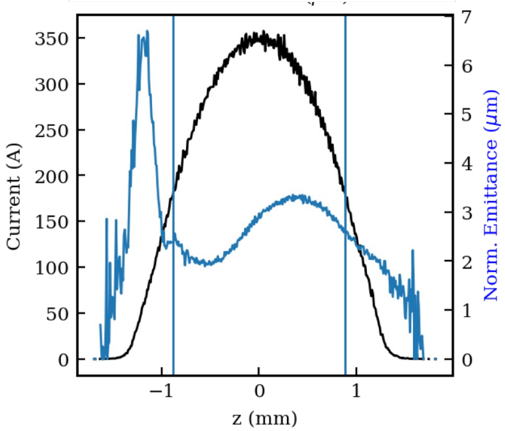
Current profile (measured)

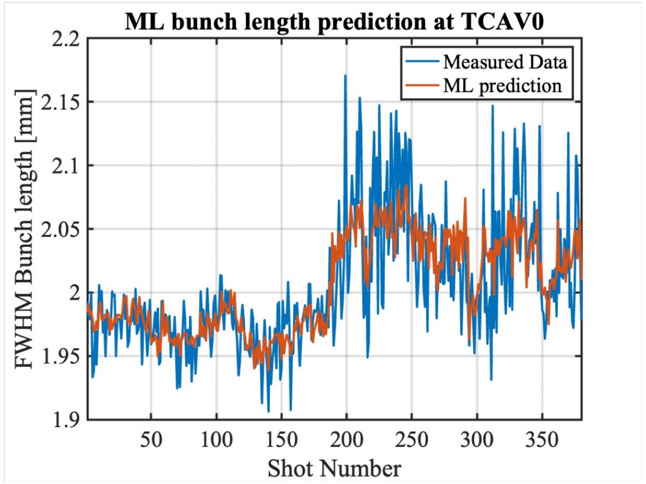


Bunch length variation (measured)



Current profile (simulation)

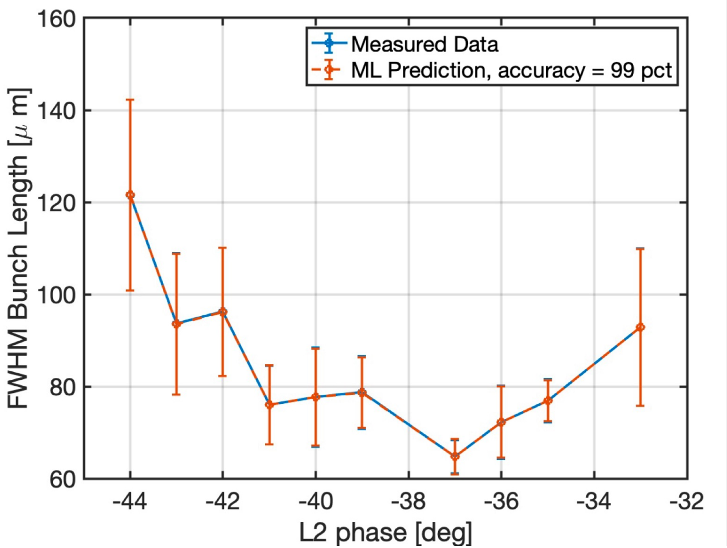


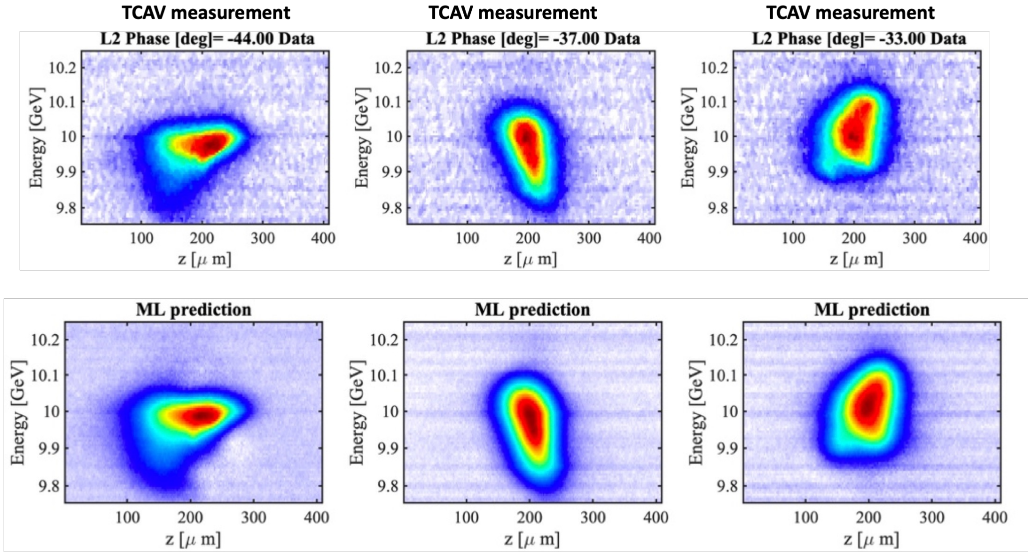


ML based diagnostic successfully predicts bunch length at the injector exit. Extension to 2D LPS to follow this year.

C. Emma

Neural network prediction of FWHM bunch length and longitudinal phase space in FACET-II experimental area

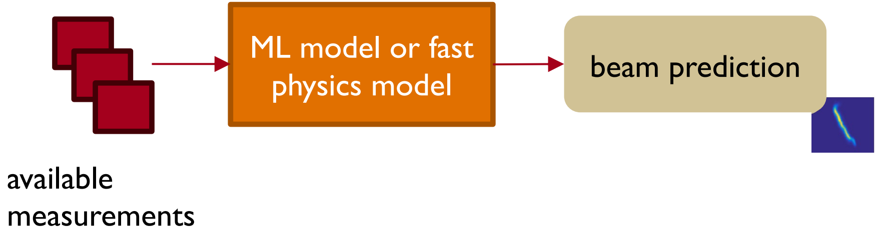




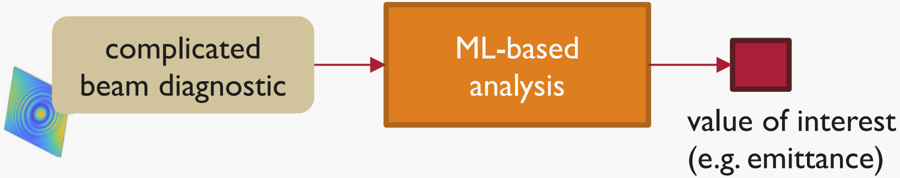
ML based LPS diagnostic feasibility demonstrated at FACET-II. Upcoming work focused on robustness + multiple locations/beam configurations.

Virtual Diagnostics ←→ ML-Enhanced Diagnostics

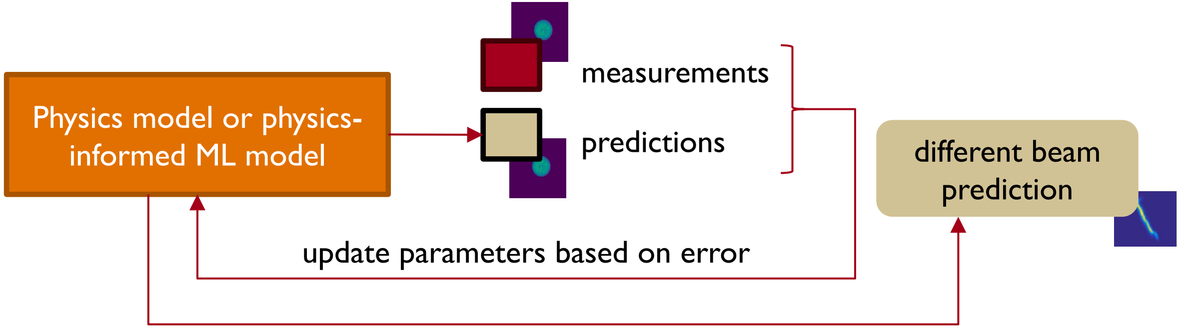
Fast, detailed predictions of quantities that aren't continuously available



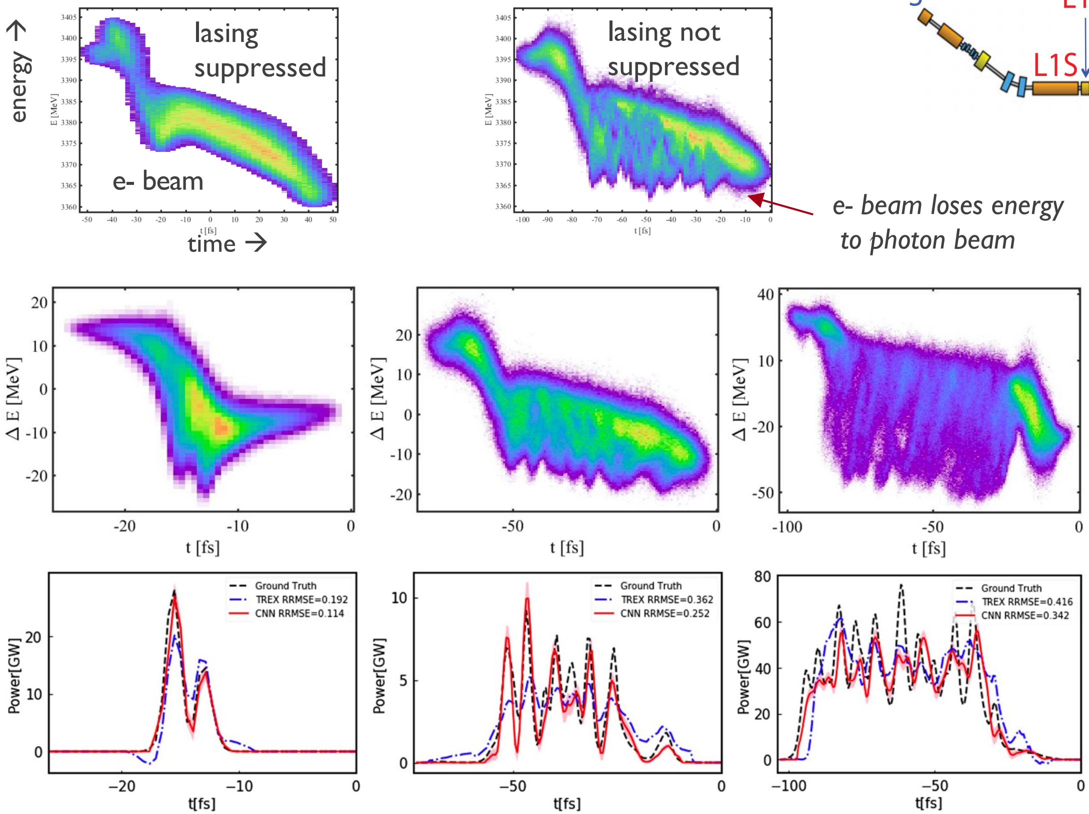
Fast analysis of complicated diagnostic output



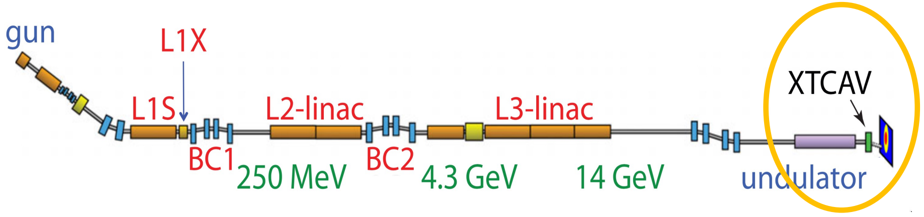
Active tuning of system models to infer unseen variables or beam behavior



ML-Based Analysis of XTCAV Images for X-ray Power Profile



X. Ren et al., PRAB 2020



At LCLS routinely use XTCAV images to predict unmeasured photon beam power profile

Standard method uses estimated energy loss:

- *Slow/iterative and doesn't work well into saturation* *(uses vertical slices of image)*
- *Needs an associated lasing-off image*

Convolutional neural network (CNN) analysis:

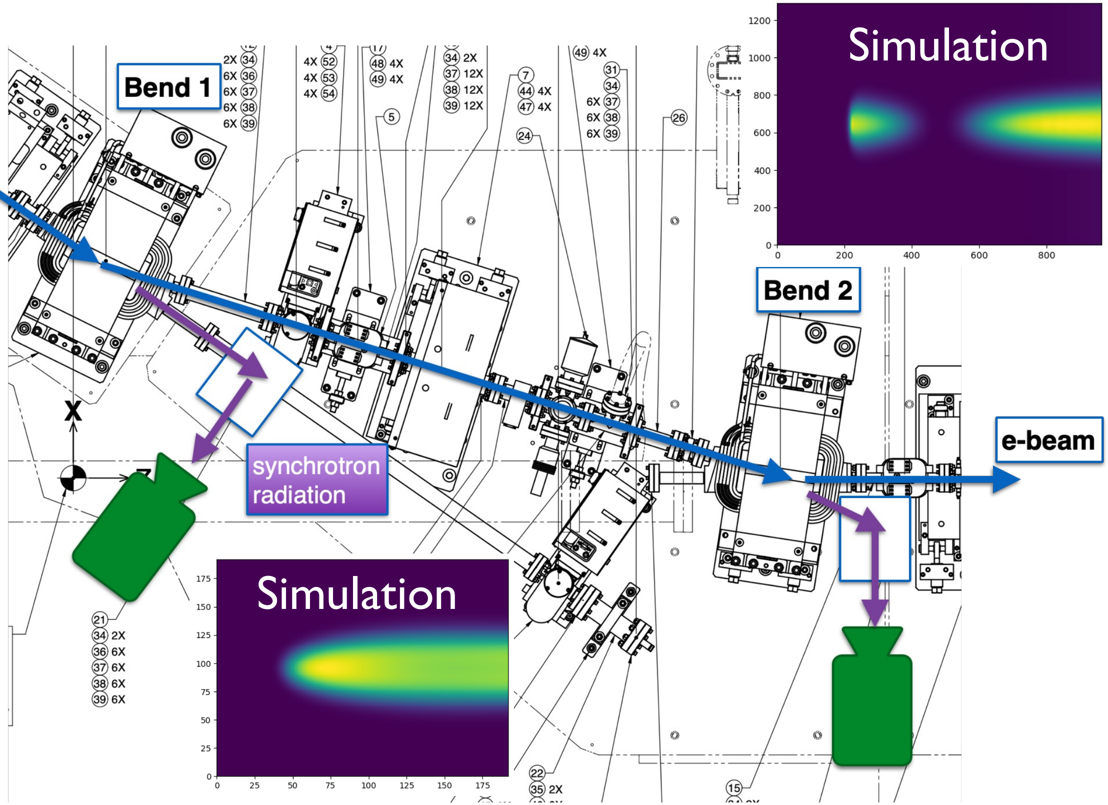
- *Uses the whole image*
- *Does not require a lasing-off reference*
- *Faster / more accurate than standard reconstruction technique*
- *Can be used into saturation regime*

Simple example of ML-enhanced analysis for photon beam power profile → needs more work to ensure robustness

Robbie Watt, Brendan O'Shea, Doug Storey, Carsten Hast

Single-Shot Emittance Monitoring

- Want a single-shot, non-destructive diagnostic that is "always on"
- Radiation from dogleg: theory suggests can be sensitive to x and y emittance, energy spread
- On-the-fly ML-based image analysis extracts beam emittance, mismatch from radiation pattern
- Push analysis to "the edge" to get rapid (10 Hz) beam quality quantification
- Working on single-image emittance measurement in BC11 at FACET-II; exploring applications to LCLS-II

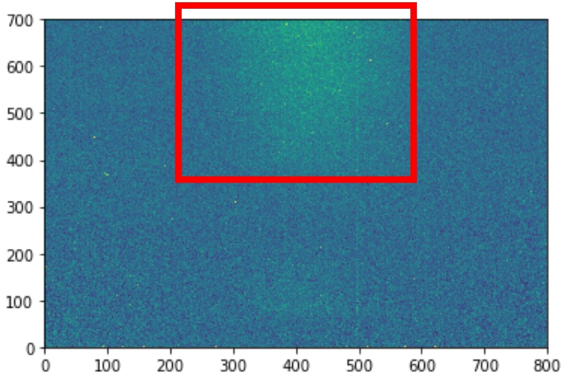


Non-destructive single shot diagnostic will use ML-analysis to give emittance evolution along accelerator

R. Watt, B. O'Shea

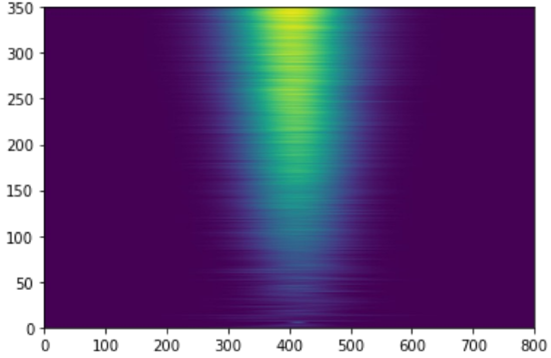
Injector emittance diagnostic – first results

Data from edge rad camera 4/19/22

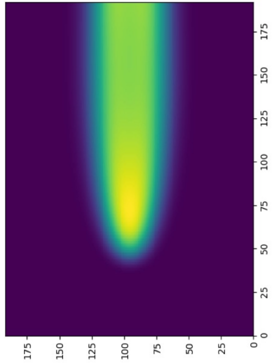


ML-based noise reduction

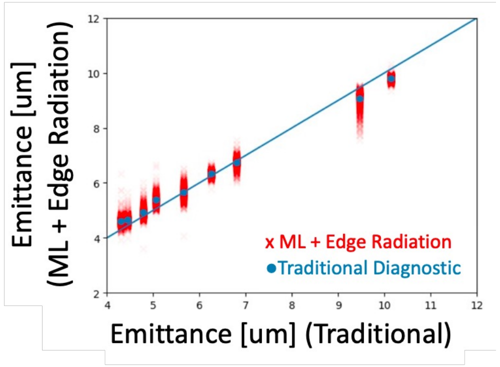
Noise filtered image



Simulation



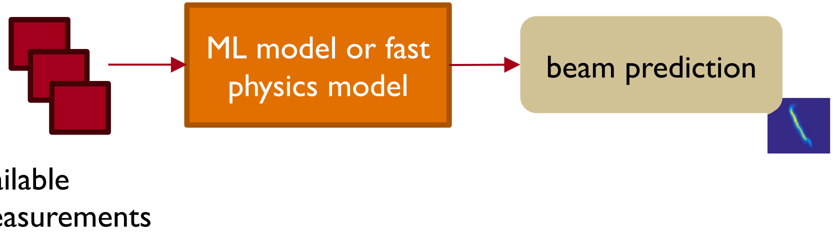
- Initial measurement data taken at injector dog-leg
- ML-based methods extract SR pattern from noisy data
- Noise filtering improves fitting and reduces uncertainty in extracted beam parameters
- First results of ML-based image analysis reproduce expected results from simulation and traditional diagnostic measurements



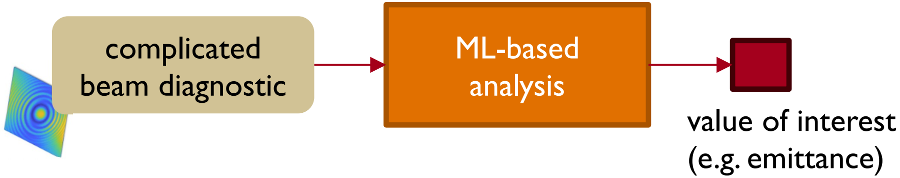
ML-based image analysis reproduces expected results from e-beam tracking simulations and traditional diagnostic measurements.

Virtual Diagnostics ←→ ML-Enhanced Diagnostics

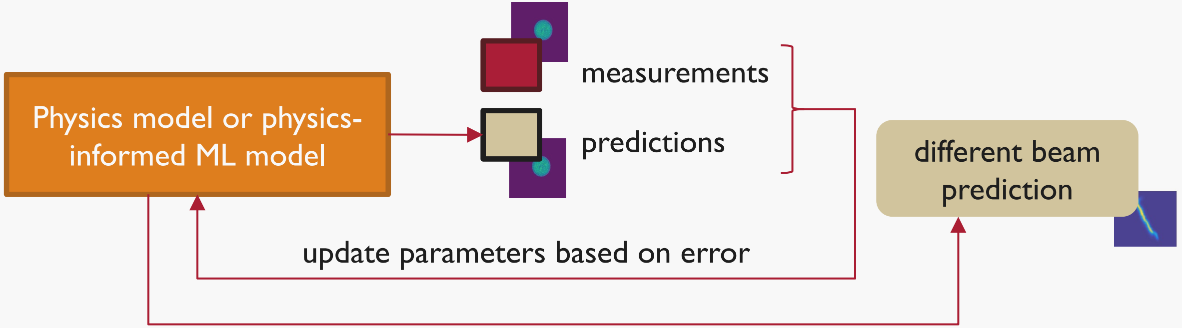
Fast, detailed predictions of quantities that aren't continuously available



Fast analysis of complicated diagnostic output



Active tuning of system models to infer unseen variables or beam behavior



Finding Sources of Error Between Simulations and Measurements

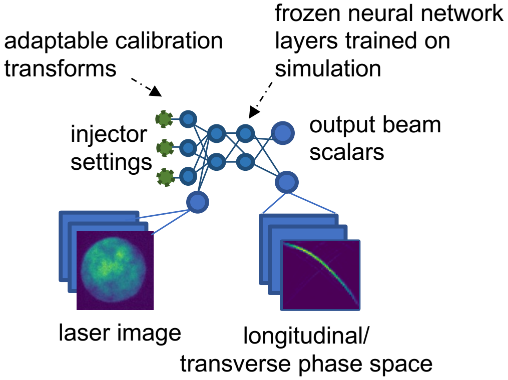
Many non-idealities not included in physics simulations:

- **static error sources** (e.g. magnetic field nonlinearities, physical offsets)
- **time-varying changes** (e.g. temperature-induced phase calibrations)

Want to identify these to get better understanding of machine performance

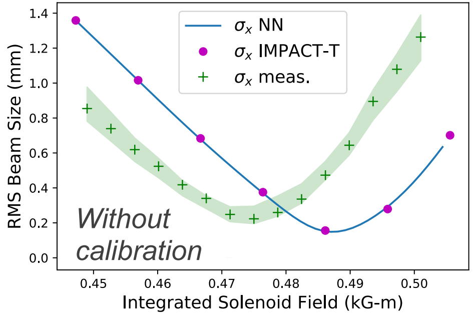
→ ***ML model allows fast / automatic exploration of error sources in high dimension***

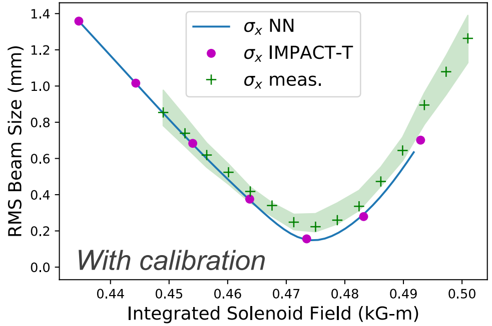
Example: calibration offset in injector solenoid strength found automatically with neural network model (trained first in simulation, then calibrated to machine)



Inputs
Laser radius
Laser spot sizes
Pulse length
Charge
Solenoid
L0A phase
L0B phase
SQ quad
CQ quad
6 matching quads

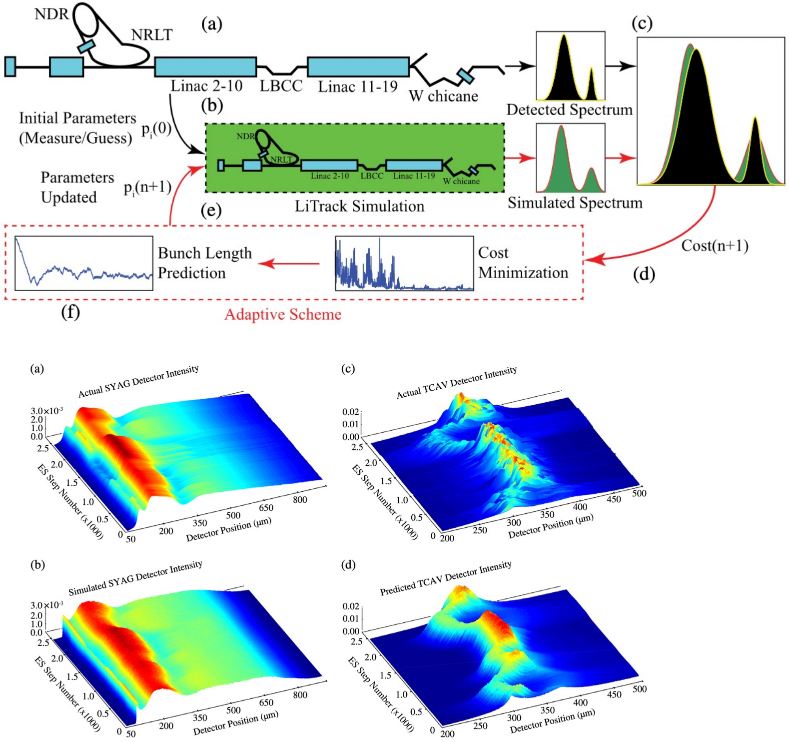
Outputs
Beam size (x,y)
Emittance (x,y)
Bunch length

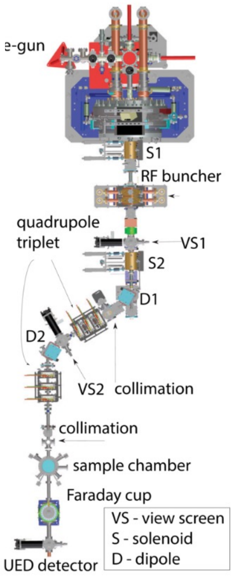


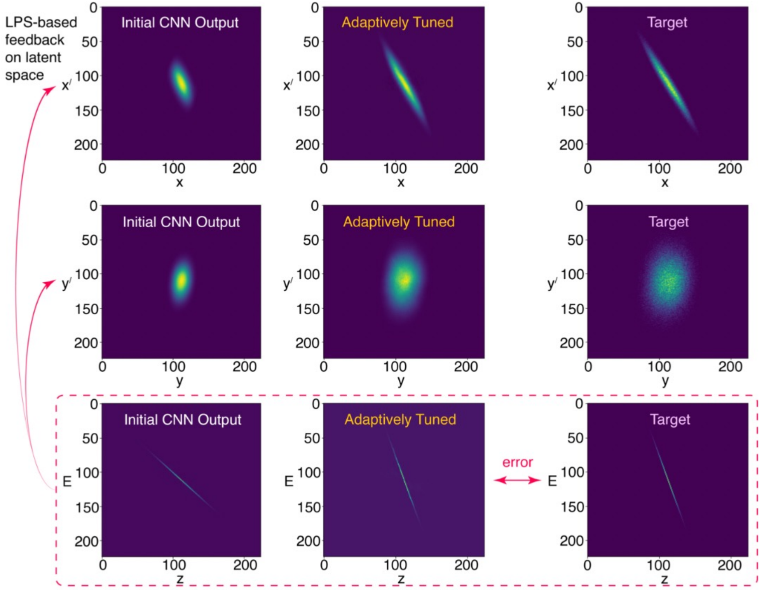


Speed of ML models enables rapid identification of error sources between idealized physics simulations and real machine
→ **path toward gaining insights into machine performance**

Continuous Feedback to Adaptively Tune Models







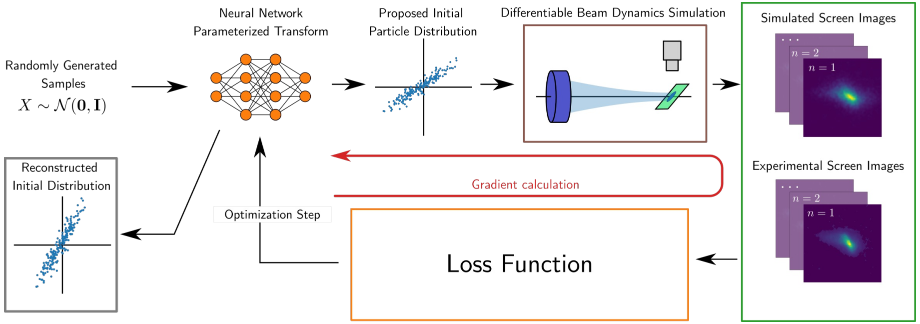
A. Scheinker and S. Gessner, PRAB 18, 102801 (2015)

A. Scheinker, JINST **16** P10008 (2021)

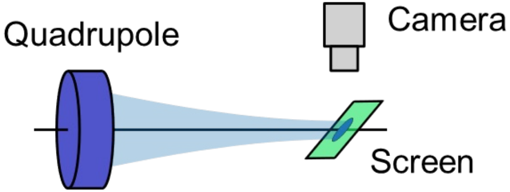
Phase Space Reconstruction with Differentiable Tracking Simulations

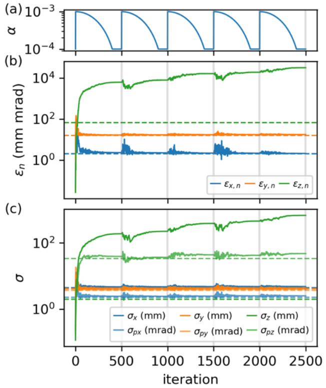
See J.P.'s talk from Tuesday

Differentiable pipeline for reconstructing 6D phase space distribution using neural network parameterization

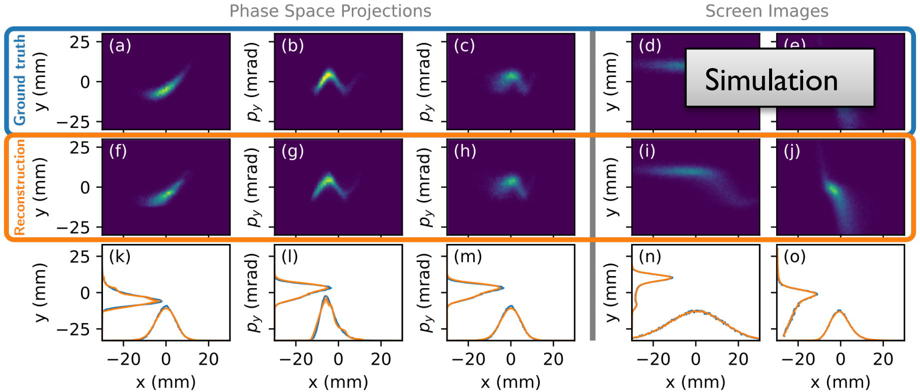


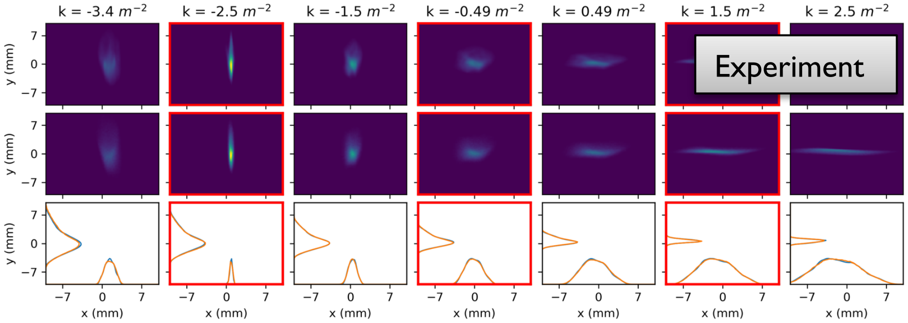
Reconstruct 4D phase space distribution + approx. energy spread from simple beamline diagnostic and 10 measurements





Confidence estimates





ML combined with differentiable simulations opens up a new paradigm for constructing detailed phase space diagnostics in a way that is computationally-efficient and sample-efficient

Summary/Conclusions

ML and HPC enable wide array of "virtual diagnostic" capabilities, including detailed online physics simulations tied to the control system

- Used online IMPACT model for LCLS-II commissioning (3-8 min execution)
- Neural network system models: ms execution and adaptive tuning of models

ML combined with differentiable simulation techniques enables beam characterization with a high level of detail and minimal data

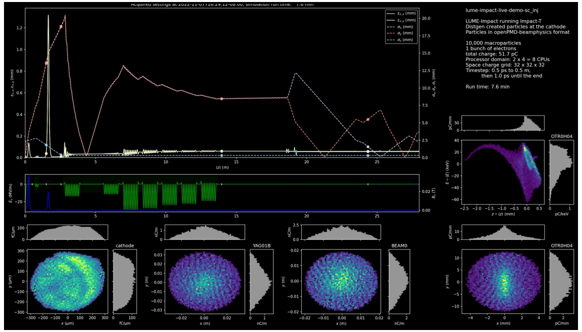
- Have demonstrated unprecedented capabilities in detailed phase space reconstruction for simple quadrupole scan measurement
- Opens up new way of thinking for what a diagnostic can be

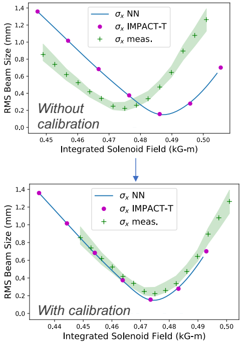
Shot-to-shot ML-enhanced diagnostics at FACET-II will provide continuous predictions of beam quality (longitudinal and transverse)

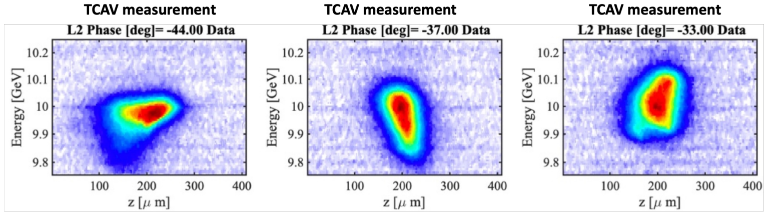
- Will aid both beam control and user analysis for experiments
- Good progress on experimental results so far → full steam ahead!

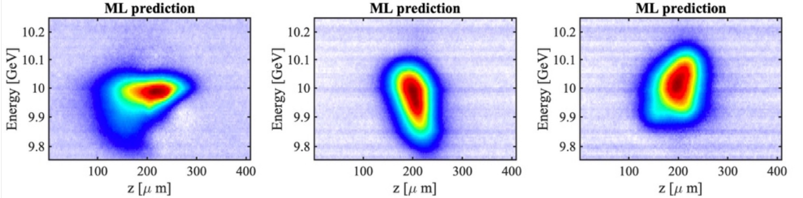
Still much work ahead for ML-enhanced diagnostics in practice

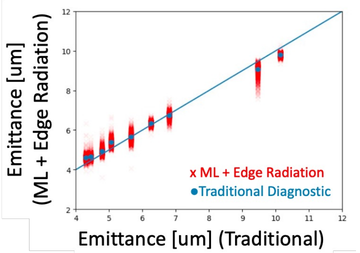
- Ensuring reliability of predictions and uncertainty estimates under changing conditions (i.e. distribution shift)
- MLOps and related software infrastructure for regular deployment/maintenance (track performance of and update ML components as needed)











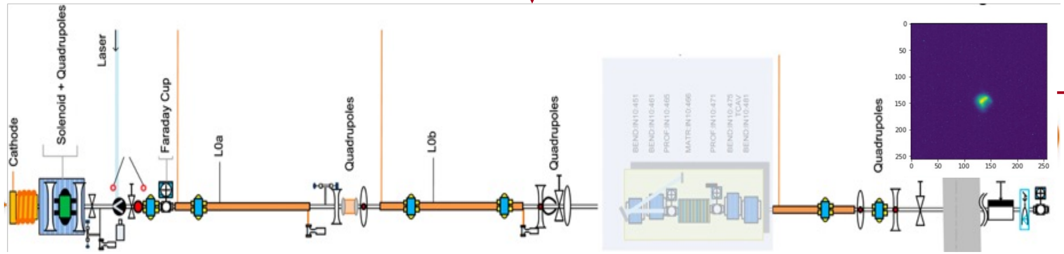
Backups

Efficient Characterization of FACET-II Injector

Automatic Exploration
(constrained to useful values of emittance and match)

Comprehensive ML Models of Injector

Setting changes on 10 variables (solenoid, bucking coil, corrector quads and matching quads)



FACET-II Injector

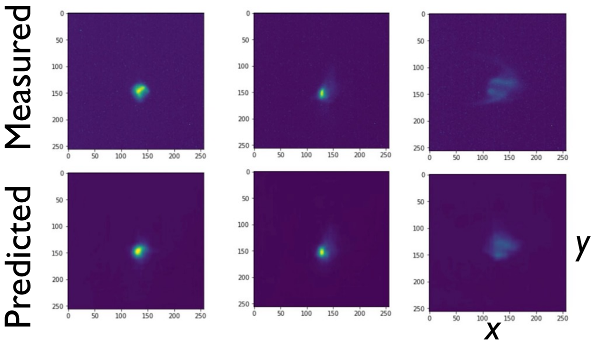
x-y emit, match, and beam images

- Used Bayesian Exploration for efficient high-dimensional characterization (10 variables) of emittance and match at 700pC: **2 hrs for 10 variables compared to 5 hrs for 4 variables with N-D parameter scan**

- Data was used to train neural network model of injector response predicting x-y beam images. GP ML model from exploration predicts emittance and match.

- **Example of integrated cycle between characterization, modeling, and optimization → now want to extend to larger system sections and new setups**

transverse phase space

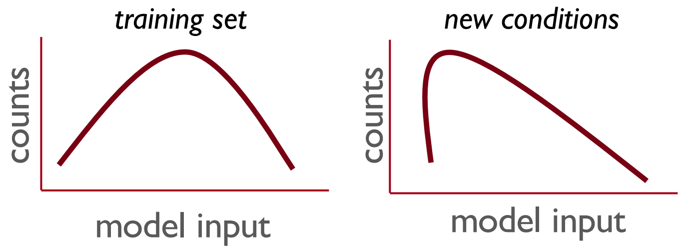


Use of Bayesian exploration to generate training data was sample-efficient, reduced burden of data cleaning, and resulted in a well-balanced distribution for the training data set over the input space. ML models were immediately useful for optimization.

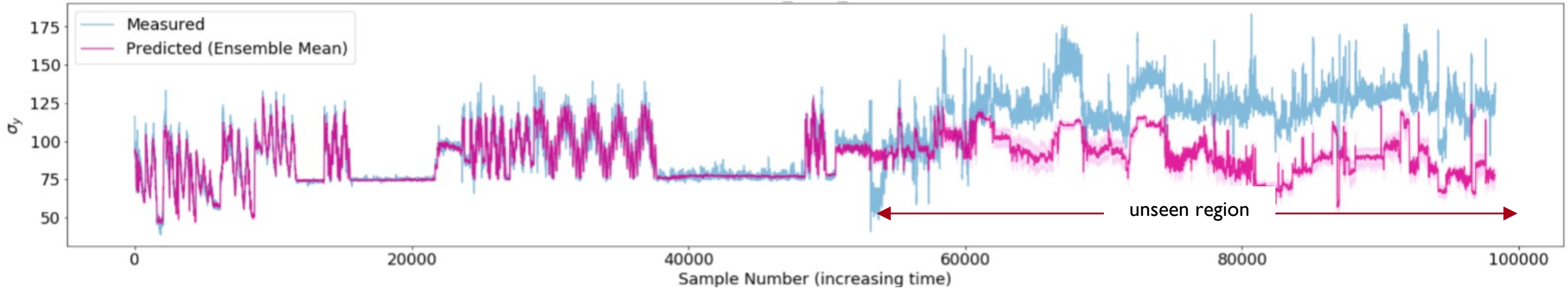
Uncertainty Quantification / Robust Modeling / Model Adaptation

Major area of AI/ML research: statistical distribution shift between training and test data degrades prediction

Distribution shift is extremely common in accelerators, due to both deliberate changes in beam configuration and uncontrolled or hidden variables



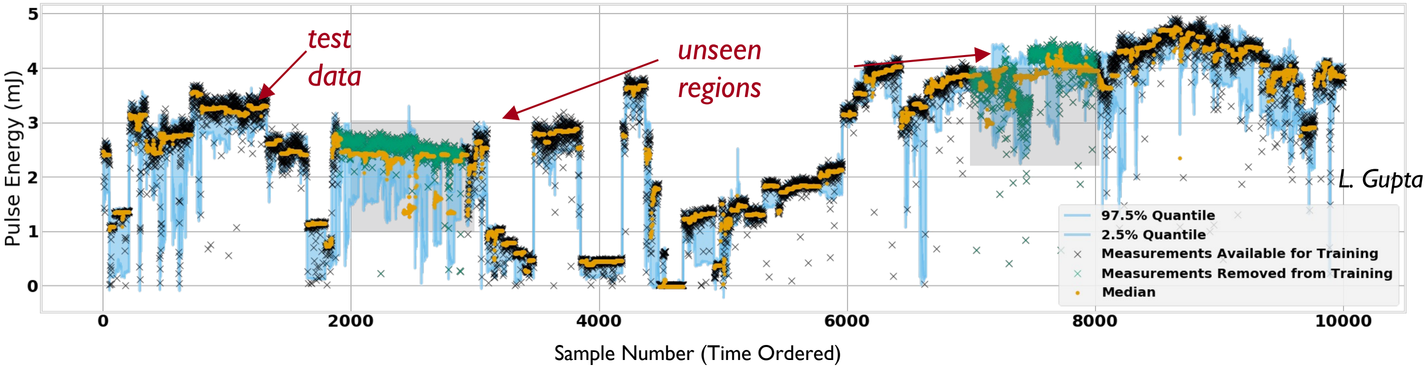
Example: beam size prediction and uncertainty estimates under drift from a neural network
Uncertainty estimate from neural network ensemble does not cover prediction error, but does give a qualitative metric for uncertainty



Reliable uncertainty estimates and model adaptation methods are key for putting online models to use operationally

Uncertainty Quantification / Robust Modeling

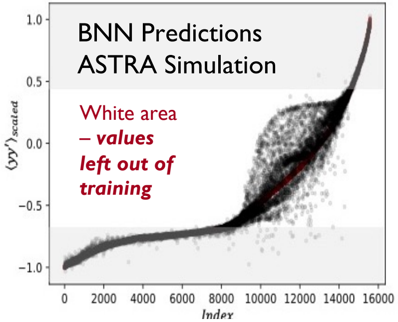
Essential for decision making under uncertainty (e.g. safe opt., intelligent sampling, virtual diagnostics)



Current approaches

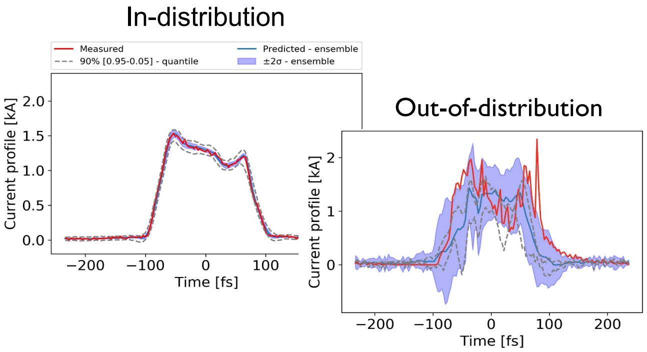
- Ensembles
- Gaussian Processes
- Bayesian NNs
- Quantile Regression

Neural network with quantile regression predicting FEL pulse energy at LCLS



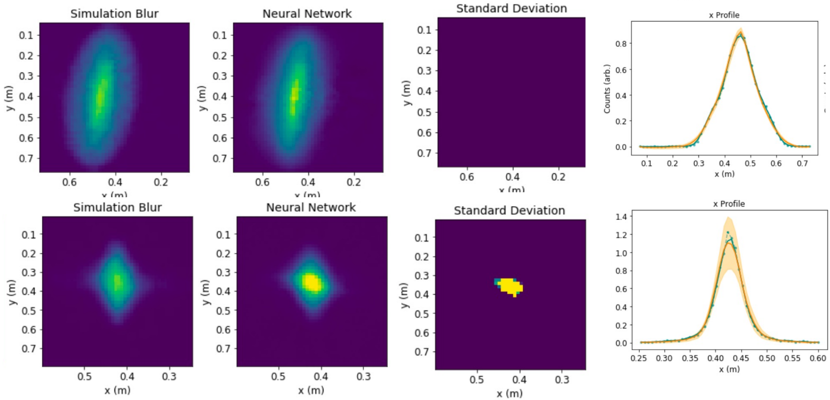
Scalar parameters for the LCLS-II injector (Bayesian neural network)

A. Mishra et. al., PRAB, 2021



longitudinal phase space (quantile regression + ensemble)

O. Convery, et al., PRAB, 2021

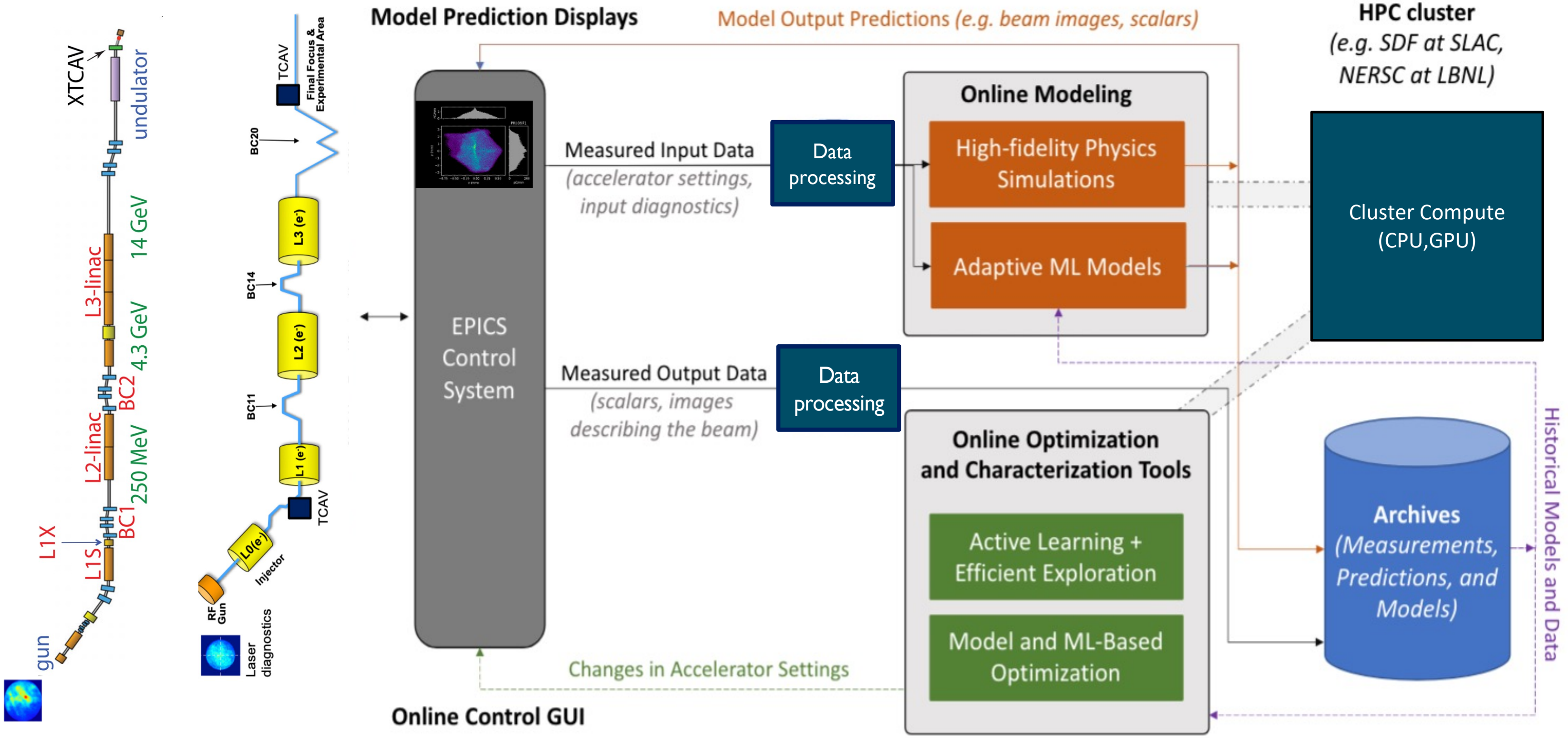


LCLS injector transverse phase space (ensemble)

Goal: Full Integration of AI/ML Optimization, Data-Driven Modeling, and Physics Simulations

*Working on a **facility-agnostic*** ecosystem for online simulation, ML modeling, and AI/ML driven characterization/optimization

Will enable system-wide application to aid operations, and help drive AI/ML development *(e.g. higher dimensionality, robustness, combining algorithms efficiently)*

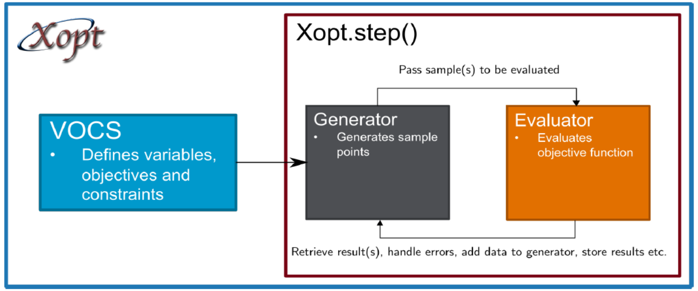


Making good progress toward this vision with open-source, modular software tools

Modular, Open-Source Software Development

Community development of **re-usable, reliable, flexible software tools** for AI/ML workflows has been essential to maximize return on investment and ensure transferability between systems

Modularity has been key: separating different parts of the workflow + using shared standards




```
vocs:
  name: TNK_test
  variables:
    x1: [0, 3.14159]
    x2: [0, 3.14159]
  objectives: {y1: MINIMIZE}
  constraints:
    c1: [GREATER_THAN, 0]
    c2: ['LESS_THAN', 0.5]
```

```
algorithm:
  name: bayesian_exploration
  options:
      n_initial_samples: 5
      n_steps: 25
      generator_options:
          batch_size: 1
          #sigma: [[0.01, 0.0],
          use_gpu: False
```

Different software for different tasks:

Optimization algorithm driver *(e.g. Xopt)*

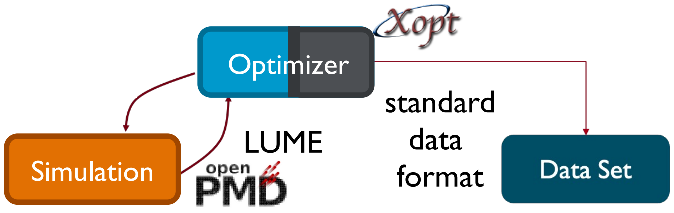
Visual control room interface *(e.g. Badger)*

Simulation drivers *(e.g. LUME)*

Standards model descriptions, data formats, and software interfaces *(e.g. openPMD)*

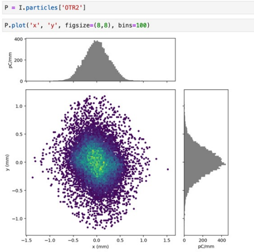
Online model deployment *(LUME-services)*

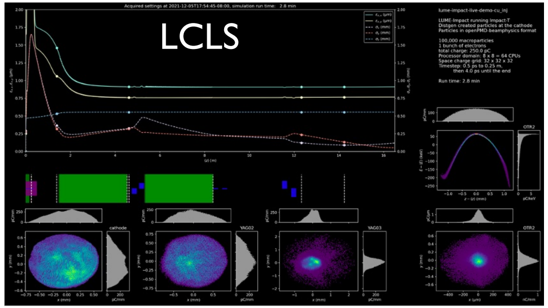
More details at https://www.lume.science/

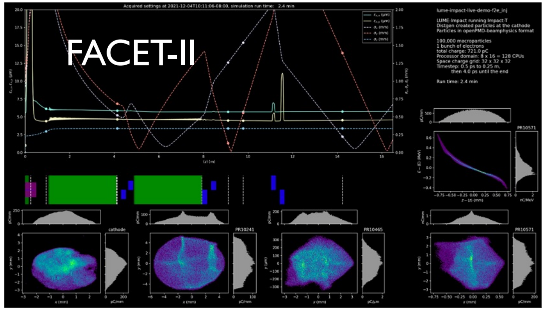


Impact
ASTRA
GPT
Bmad
Genesis
SRW

gen_1.json
root:
variables:
generation: 1
vocs:
error: [] 1241 items
inputs: [] 1241 items
outputs: [] 1241 items





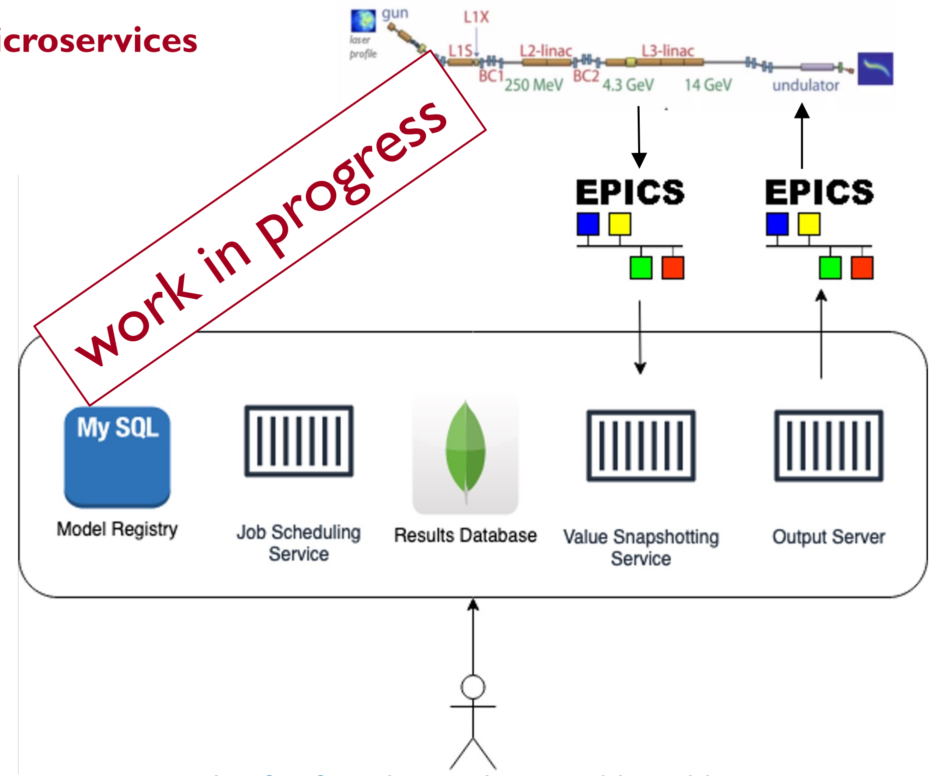


Online Impact-T simulation and live display; trivial to get running on FACET-II using same software tools as the LCLS injector

Modular open-source software has been essential for our work. We welcome new users and contributors.

LUME-services: An online modeling service built on microservices

- LUME-services is a Python package providing data APIs for inter-service interactions and user tooling
- Models are pip-installable Python packages and templates may be auto-generated using the LUME-services tools
- Models run in containers when a user schedules a workflow run
- The template provides Continuous Integration (CI) tools (e.g. GitHub actions) for users to use for testing and deployment
- Have demoed for a variety of physics sims and ML models at SLAC
  → now testing / improving for new cases
- Have not yet integrated MLOps components (e.g. continuous/triggered automated model adaptation)
- Resources:
  - lume-services https://slaclab.github.io/lume-services/demo/
  - lume-model https://slaclab.github.io/lume-model/
  - lume-epics https://slaclab.github.io/lume-epics/
  - distgen https://github.com/ColwynGulliford/distgen



Interface for packaging arbitrary models, model registry

Enforcement of minimal metadata (model descript, owner, model type, PVs)

Ability to scale to arbitrary number of models and clients

Result storage + programmatic IOC for model results

Infrastructure for reliable, continuous online model deployment and model version tracking / updating

Aimed for transferrable design between platforms → we welcome collaborators!