E331: FY22 Progress and Plans for FY23

Neural network based tuning to exploit machine-wide sensitivities in pursuit of high beam quality

---

Auralee Edelen on behalf of E331 / SLAC National Accelerator Laboratory

FACET-II PAC Meeting, 26 October 2022, SLAC

FACET-II

SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University

U.S. DEPARTMENT OF ENERGY

E331 Science Motivation

Major limitations in the way accelerator tuning is done:

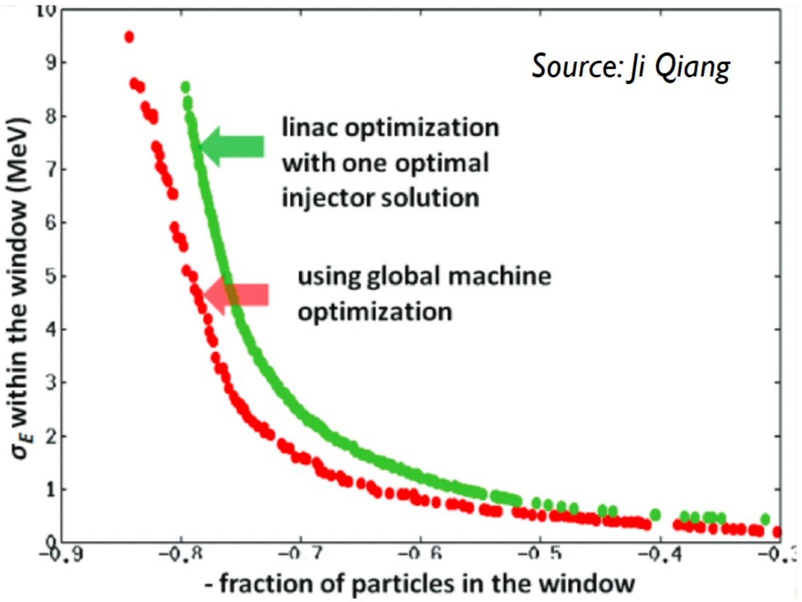
- Piecemeal tuning of subsystems *(known to be sub-optimal)*
- Indirect use of high-dimensional diagnostics *(e.g. images)*
- Often a lack of accurate online models

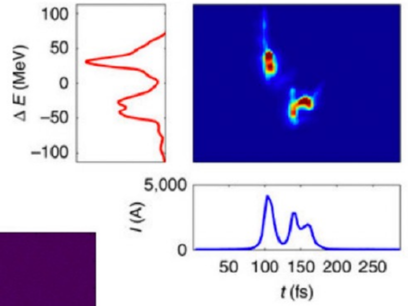
→ *Potentially limiting factors in control of extreme beams*

More global view can enable better control:

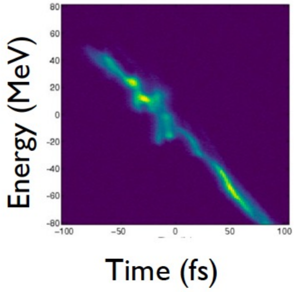
- Fully exploit unknown system-wide sensitivities + nonlinearities
- Faster switching between setups *(if using global representation of machine)*
- Better handling of parameter tradeoffs *(e.g. jitter, matching, longitudinal phase space)*

Comprehensive, system-wide control is likely to be a key factor in improving custom control of extreme beams, but this is a difficult task





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

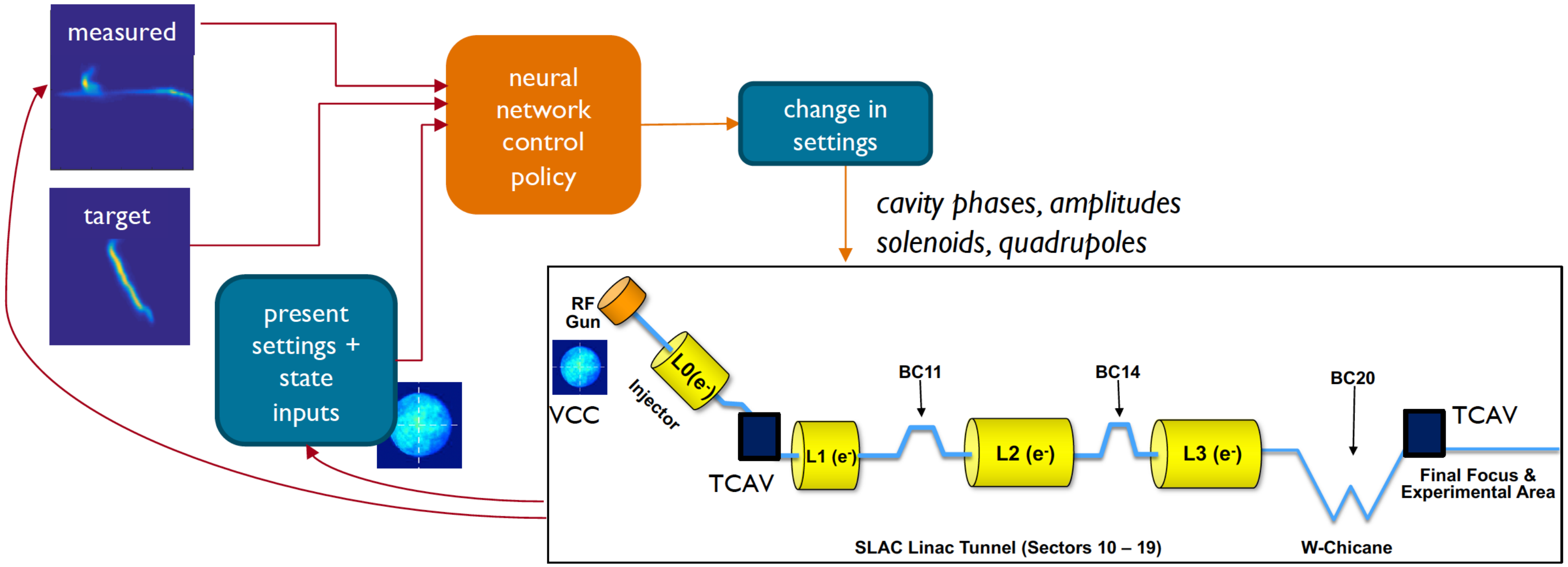


A. Marinelli, IPAC'18

E331 Science/Technical Goals

Main goal: develop and demonstrate methods to leverage global learned system responses to aid fast, high-quality tuning of beams under challenging conditions

(build up incrementally to machine-wide neural network-based reinforcement learning)

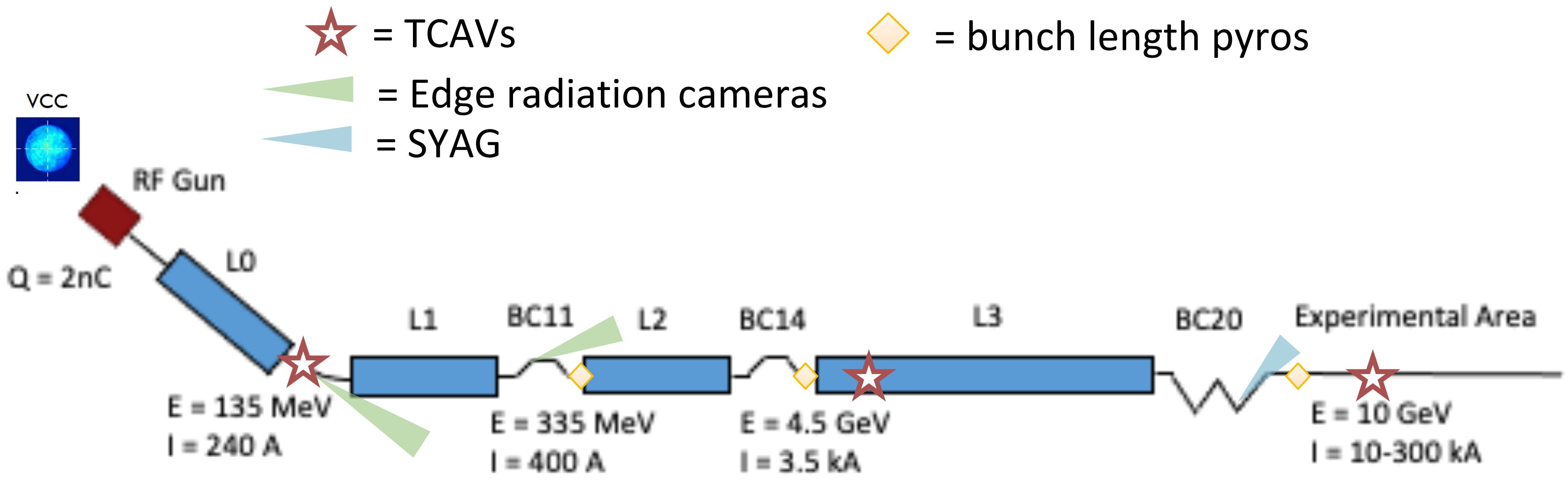


Science/Technical Goal	Target Time	Definition of Success
Evaluate methods for high-dimensional, high-quality control over beams using learned responses, starting with small-scale problems + single-bunch mode	1-2 years	Automated tuning of transverse emittance and longitudinal phase space: faster, higher-quality tuning than standard methods, new capabilities in control
High-quality control over extreme beams and plasma experiments	2-3 years	Same as above but for more challenging setups/target beams
Deliver algorithms and interfaces for regular operation	continual	Tools incorporated into regular use + transitioned to operations

Staged approach gradually increases complexity
Success defined by performance (new capabilities) and transition to operations

SLAC

E331 Diagnostic and Observables



Similar diagnostic needs to E327

- LPS diagnostics (e.g. injector + downstream TCAVs)
- Emittance measurements, x-y beam sizes from wires, transverse phase space from screens
- Upstream inputs: virtual cathode camera, QE map once available, laser diagnostics
- Readbacks from settings (gun solenoid, gun and linac phases/amplitudes etc)
- DAQ: ~150 scalar diagnostics (e.g. BPMs, toroids, RF readbacks, BLEN pyros) and multiple image diagnostics (SYAG, EOS, TCAV)

→ Flexibility in E331 enables adaptation to installation / commissioning schedule for different diagnostics

Numerous diagnostics can be used to inform tuning or be used as tuning targets

FY22 Progress to date - shift timeline

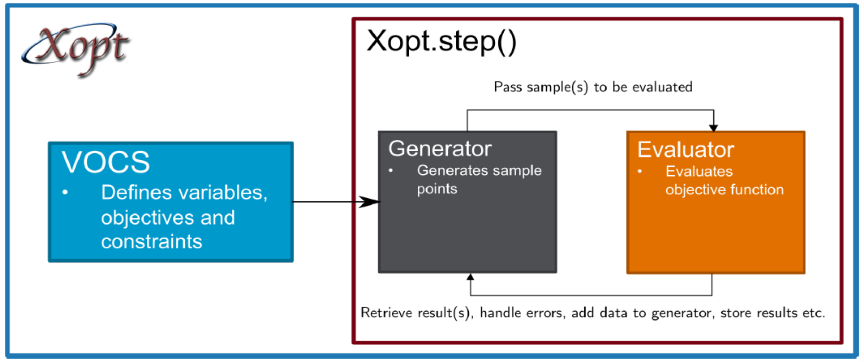
Shift Summary Date	Experiment Num	Shift Start Time	Shift End Time	Useful Beam Time	Accelerator Downtime	User Downtime	Brief Summary	
11/17/21	E327	11/17/21 11:44	11/17/21 17:44	6	0	0	Tested software. Ran ND scan characterizing injector emittance vs sol,buck, cq,sq	
11/20/21	E327	11/19/21 20:12	11/20/21	12	0	0	Gathered training data for ML optimiization of injector emittance. Tested software.	
11/29/21	E327	11/28/21 18:17	11/29/21	8	0	0	Test Bayes Exp for injector emittance	
12/4/21	E331	12/3/21 12:00	12/4/21 0:00	12	0	0	Ran Bayes Exp on emittance + bmag	
12/11/21	E331	12/10/21 20:25	12/11/21 12:25	16	0	0	Ran Bayes Exp on emittance + bmag	
12/17/21	E327	12/16/21 20:13	12/17/21 8:13	12	0	0	TCAV measurements scanning L2 phase data gathered. Inj opt data gathered with match	TCAV
2/27/22	E331	2/27/22 11:59	2/27/22 23:59	12	0	0	Compared opt methods for injector emittance + match at new laser wavelength of 253nm with 266 nm prior data	
5/14/22	E331	5/13/22 18:02	5/14/22 5:02	11	0	0	Characterize emittance at 1.8 nC with Bayes Ex. Optimize with BO + other methods and gather comparative data	
8/22/22	E331	8/21/22 14:30	8/22/22 2:30	6	6	0	Ran Bayesian Optimization on Sextupole movers. Gathered TCAV, EOS and wire scanner data at different sextupole mover positions	TCAV
			Sum Total Hrs	95	6	0		

- Shared beam time with E327 (see Claudio’s talk for details)
- Deployed initial software tools for measurements and optimization
- Characterized injector under different charge settings and laser parameters (1.8nC, 700 pC)
- Tested new ML algorithms for efficient characterization and tuning (applied to injector emittance and IP spot size tuning)
- Next steps: continue scaling up + use data gathered to move toward more comprehensive model-based approaches; incorporate TCAVs in tuning (once they are fully operationally ready)

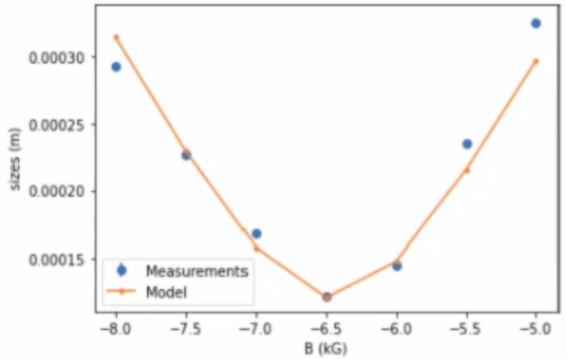
First results demonstrate utility of ML optimization tools → data gathered will be used in next phases of project

E331 Progress: Practicalities and Infrastructure

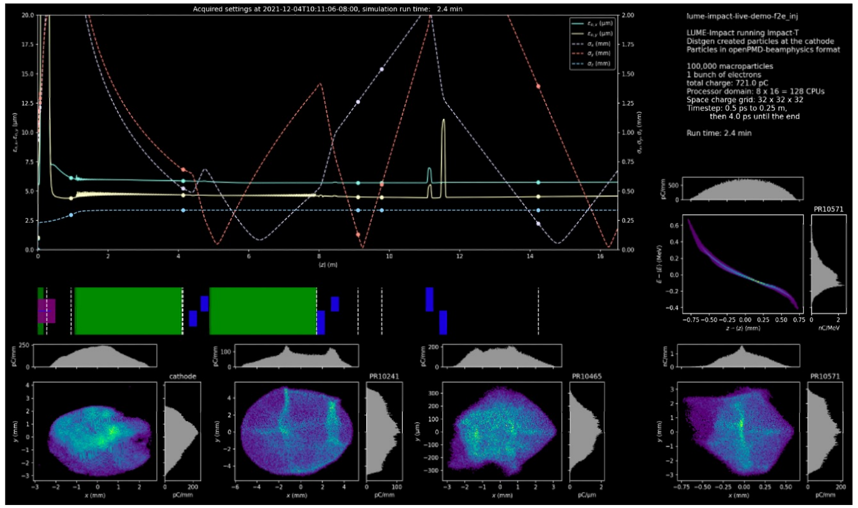
- Thoroughly vetted **adaptive emittance measurement method** for use in automated emittance optimization (PyEmittance) https://github.com/slaclab/PyEmittance
- Integrated DAQ for beam synchronous acquisition into python code for emittance measurements
- Obtained access to a **Rhel7-compatible machine** for control system → *necessary for cutting-edge algorithm testing*
- Integrated **Xopt into FACET-II control system** → *aids algorithm transfer between systems and will make it easy to test new algorithms on FACET-II*
- **Deployed online LUME-IMPACT model of injector** (live reading from machine and making predictions) → *particle-in-cell code includes space charge, uses VCC image to automatically create initial particle distribution*



Xopt running on FACET-II for easy ML algorithm deployment on different tuning problems



Adaptive quad scan emittance measurement deployed for robust measurements



FACET-II Injector model running online using LUME-IMPACT

https://www.lume.science/

E331 Progress: ML for Efficient Characterization

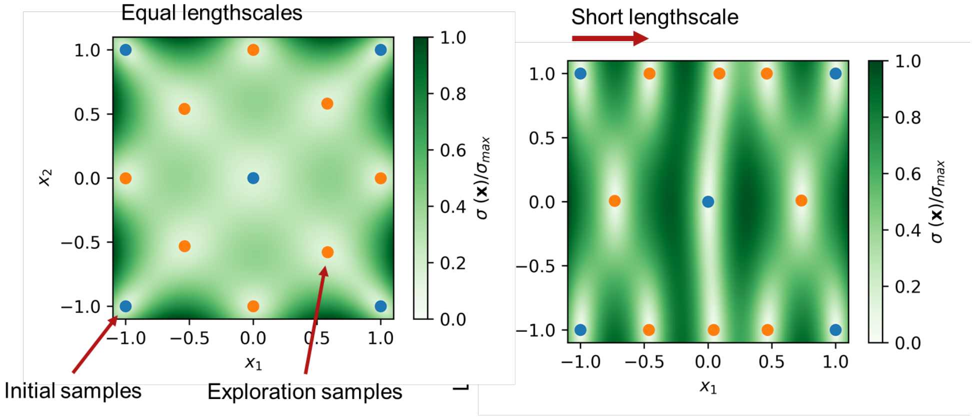
Better Data Sampling: Bayesian Exploration

$$\alpha(\boldsymbol{x}) = \sigma(\boldsymbol{x}) \prod_{i=1}^{N} p_i(g_i(\boldsymbol{x}) \geq h_i)\, \Psi(\boldsymbol{x}, \boldsymbol{x_0})$$

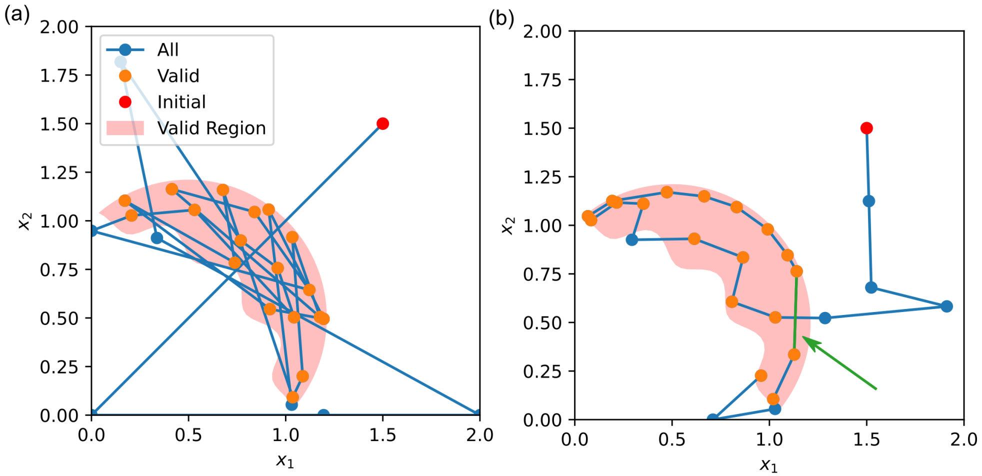
R. Roussel et. al.
Nat. Comm. **2021**

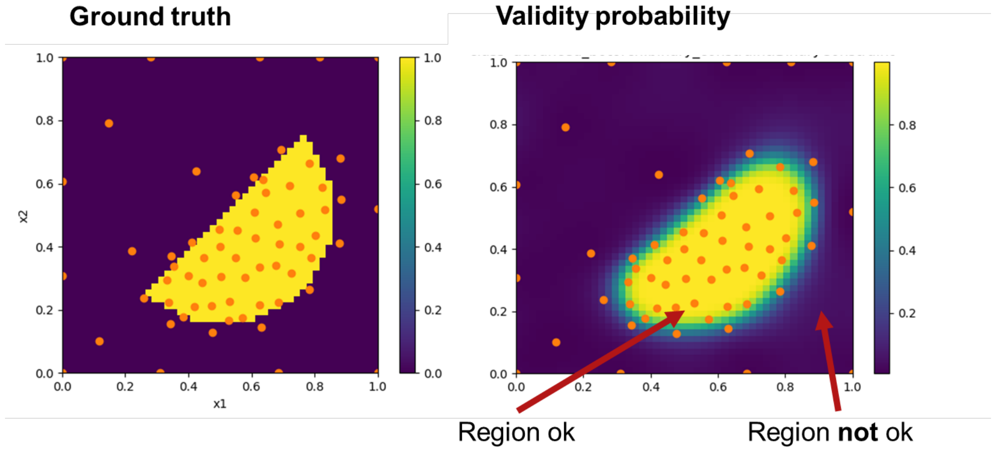
proximal biasing

adaptive sampling



learning constraints





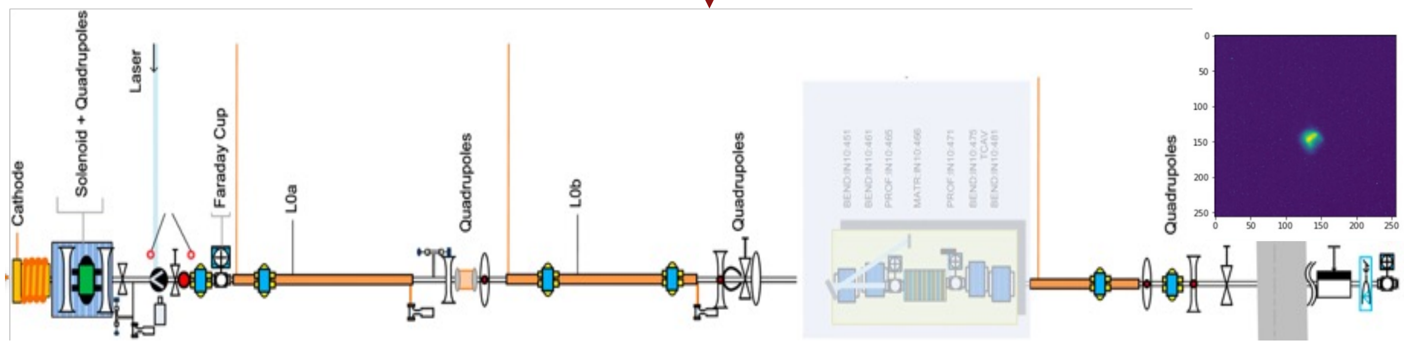
Enables sample-efficient characterization of high-dimensional spaces, while respecting both input and output constraints

E331 Progress: ML for Efficient Characterization

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

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

ML Models of Injector

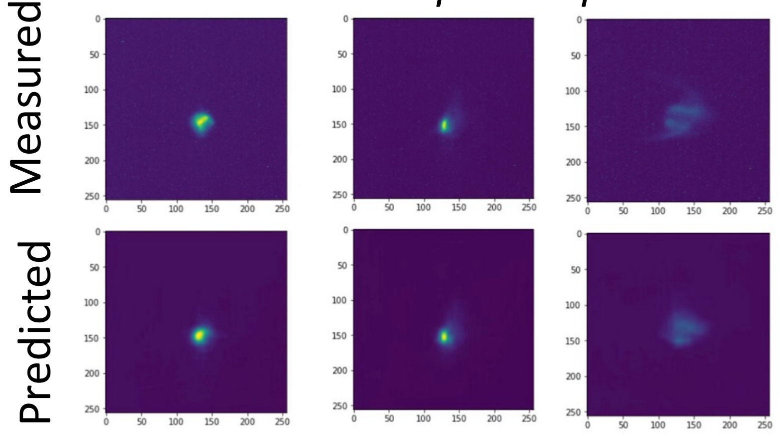


FACET-II Injector

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**

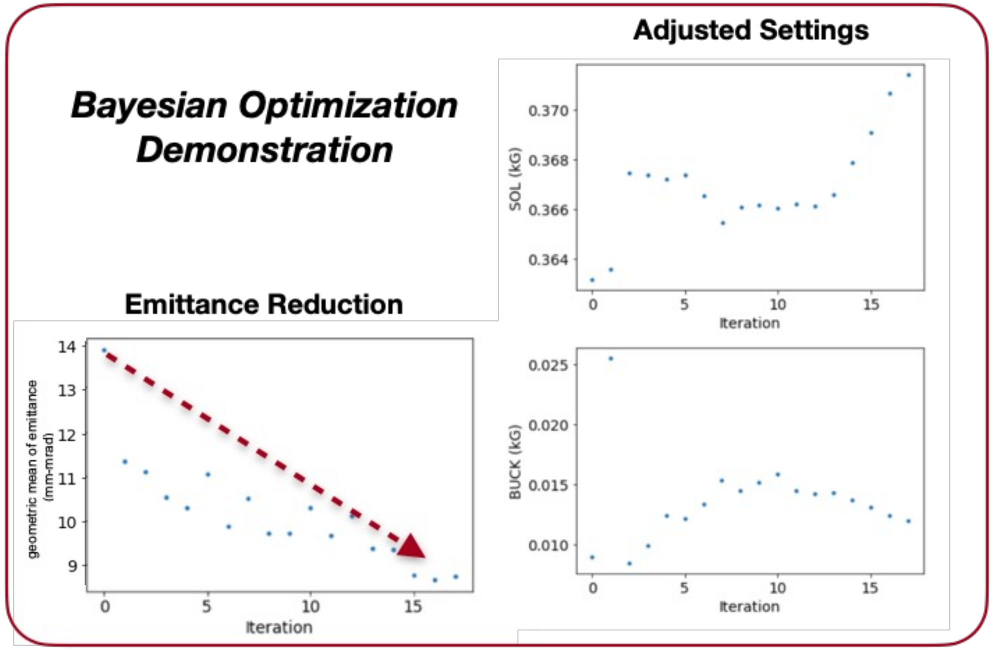
transverse phase space

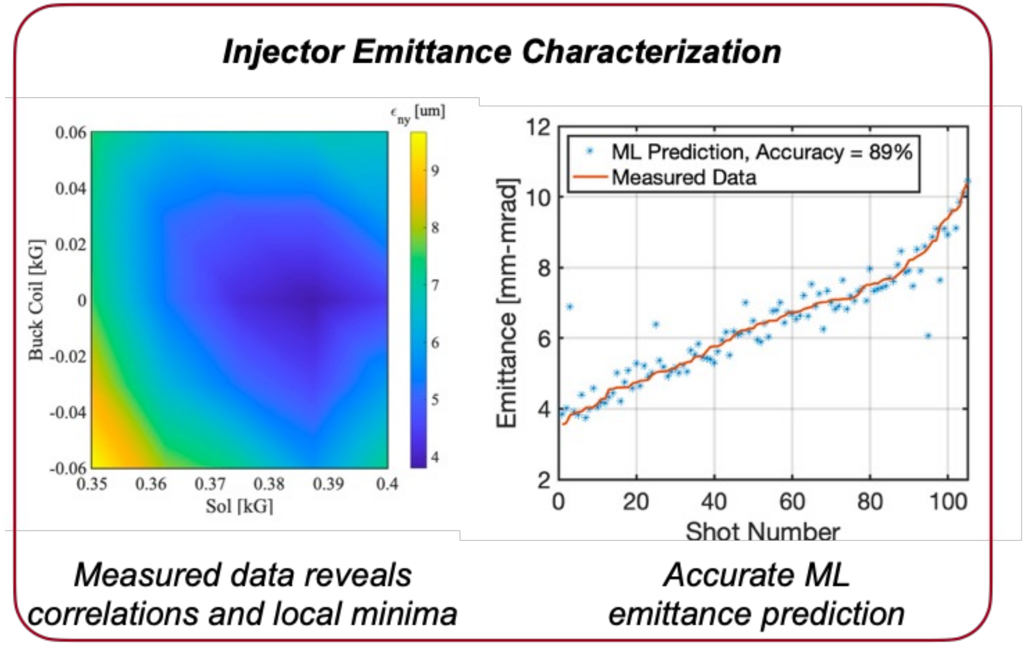


Use of Bayesian exploration to generate training data was sample-efficient, reduced burden of data cleaning, and resulted in a well-balanced distribution for the training data set over the input space. ML models were immediately useful for optimization.

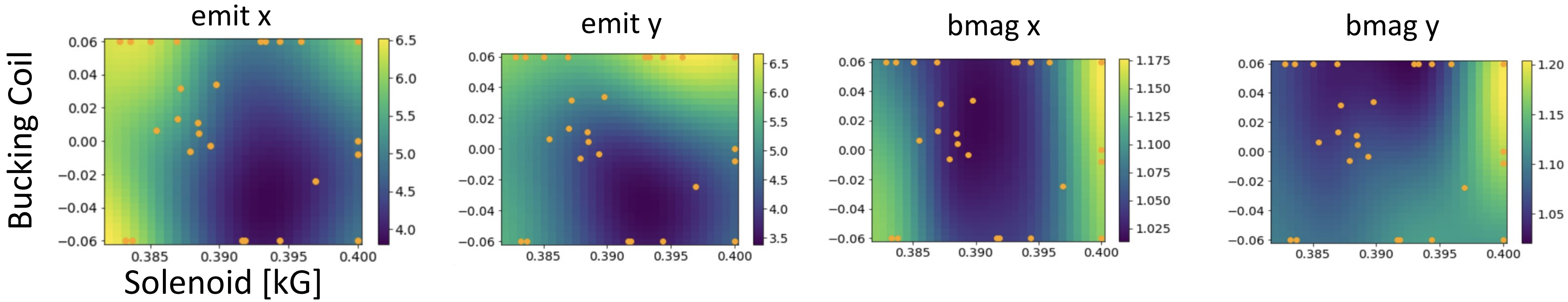
E331 Progress: Bayesian Optimization and Characterization of Injector

- Demonstrations of Bayesian optimization on the injector with up to 10 variables
- Extensive data obtained from characterization studies at 2nC and 700pC
- ML models from data give insight into machine behavior → still exploring this extensively





1.8 nC

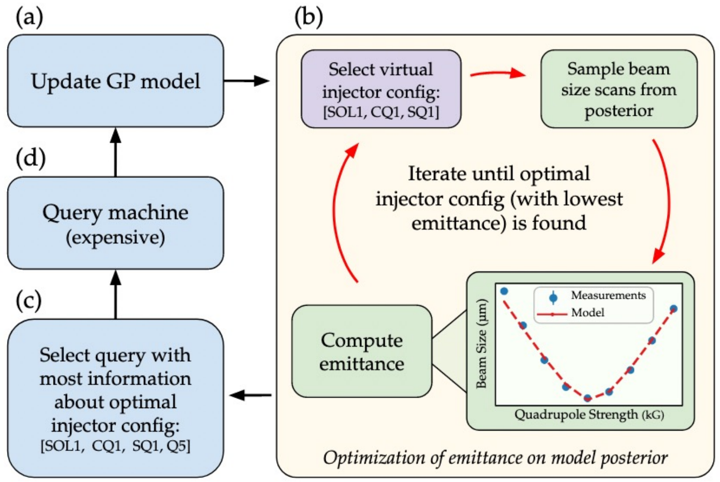


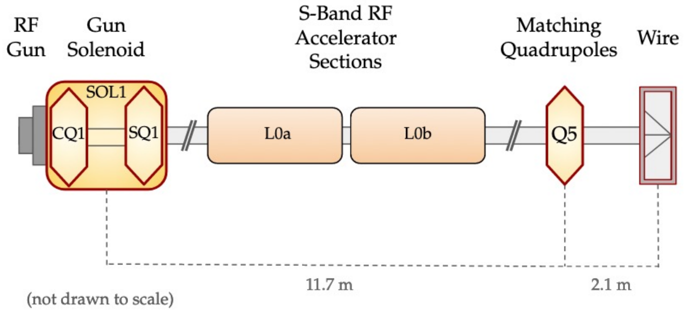
ML model predictions and new sample locations

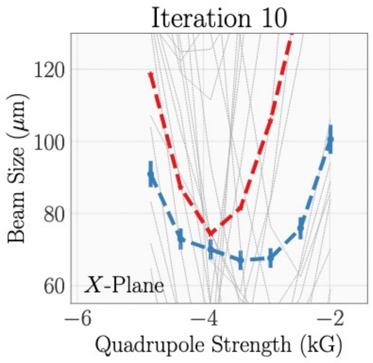
(learning to balance tradeoffs between outputs)

E331 Progress:
Efficient Emittance Optimization with Partial Measurements

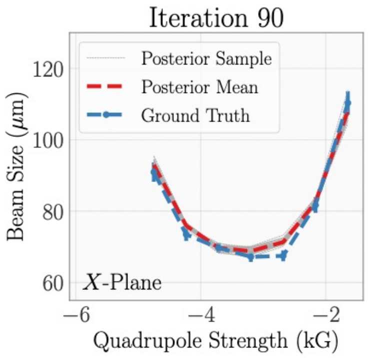
- **Instead of tuning on costly emittance measurements directly, learn a fast-executing model online for beam size while optimizing**
- **Demonstrated new algorithmic paradigm leveraging "Bayesian Algorithm Execution" (BAX) for 20x speedup in tuning** → *learn on direct observables (e.g. beam size); do inferred "measurements" (e.g. emittance) much more quickly on the model than would be possible on the machine*

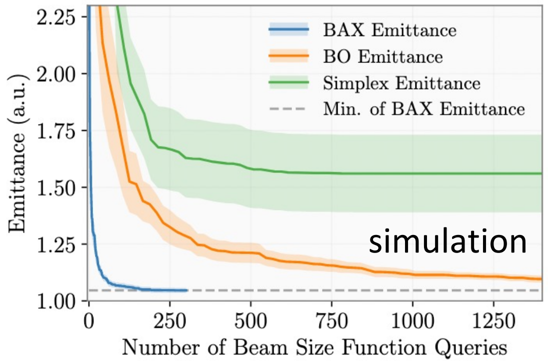


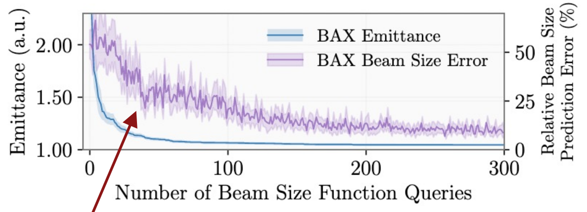




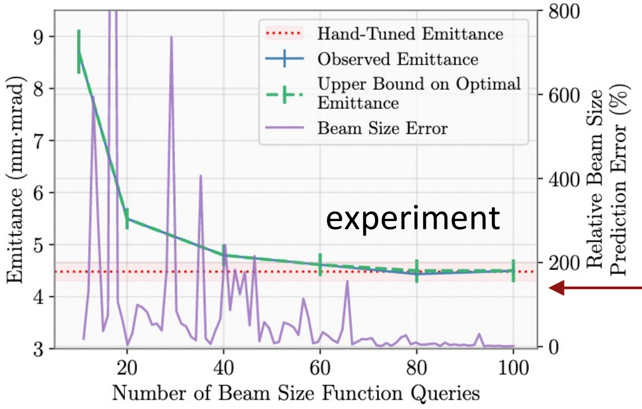
model is learned on-the-fly







Convergence of beam size prediction error gives practical indicator of optimization convergence *(no need to do direct emittance measurement until the end)*



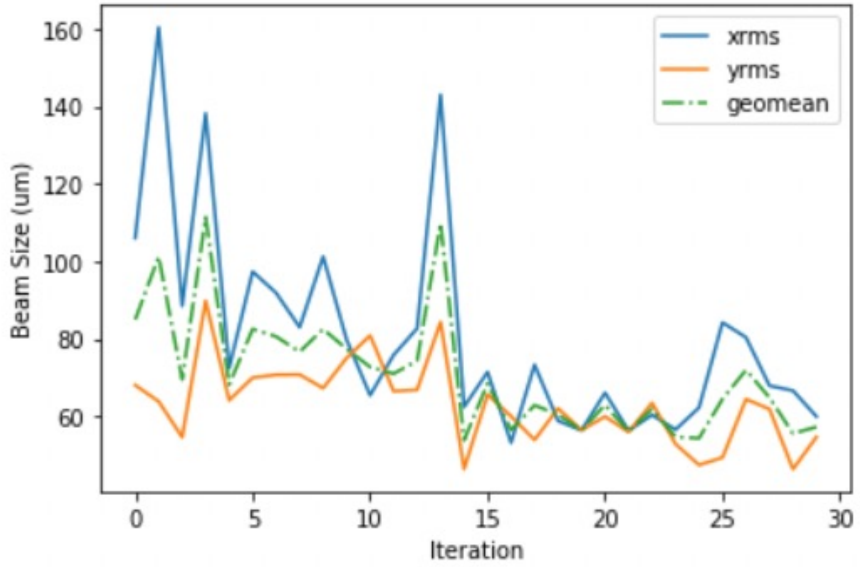
Found equivalent quality to hand-tuning in about 70 iterations *(just a few minutes with computationally optimized routine)*

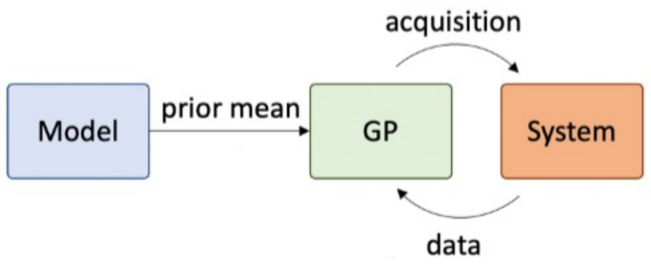
https://arxiv.org/abs/2209.04587

New method demonstrated at FACET-II has 20x speed improvement over standard emittance optimization method. Paradigm shift in how tuning on indirectly computed beam measurements (such as emittance) is done.

E331 Progress: Optimization of Sextupoles for Performance at IP

- Ran constrained Bayesian optimization on the sextupole movers (8 variables total) to minimize spot size as measured on the wires in S20
- Recorded auxiliary data (TCAV and EOS, BSA)
- First step toward more comprehensive tuning in S20
- Used software, Xopt, established for previous runs with little need for adjustment to this specific problem → *nice demonstration of extensibility*





Next:
- Want to use on both IPs (with multi-objective optimization) and use greater number of variables
- Use data to inform faster subsequent optimization (demonstrated in simulation on a different problem)

Automatically tuned for a small, round beam at the IP using sextupole movers. Ready for next steps in tuning both IPs and with broader set of variables.

Potential evolution of experiment

- Next steps:
  - Continue scaling up for combined injector + downstream tuning
  - Use data gathered to move toward more comprehensive model-based approaches (neural network prior mean, reinforcement learning)
  - Incorporate TCAVs in tuning for longitudinal phase space (once fully operationally ready); initial demos in sim and on other systems done

- New task: simultaneous optimization of the beam spot at both IPs (adjusting sextupole movers and other variables in S20)

- Farther future:
  - PWFA optimization
    - Reduction of beam jitter (synergy with E325 + E327)
    - LPS tuning/control in conjunction with PWFA diagnostics (synergy with E325 + E327)
    - Can leverage virtual diagnostic from E327 as additional tuning output
  - ML aided LPS shaping with the laser heater (synergy with E325 + E327)

Desired facility upgrades

- TCAV desired upgrades same as for E327:
  - S14 and S10 TCAVs operational
  - S20 TCAV resolution optimized (increased voltage, optimized optics)

- Computing
  - Rhel7 upgrade to control system (right now using a single test box)
  - GPU integration into compute resources

- Laser heater → would like to have screens rather than needing to rely on wires (full distribution, faster measurements)

Synergies between ML experiments (as of last PAC)

Machine Control and Understanding

Diagnostics (Need information to make decisions)		Control (How to make decisions)	
Edge Radiation Emittance Diagnostics (E326)	Virtual TCAV Predictive Diagnostics (E327)	Adaptive Feedback (E325)	Learned Control (Reinforcement Learning, New proposal)
Non-destructive, single shot continuous monitoring of emittance of high-current beams	Longitudinal phase space diagnostics, always on, and for extremely short bunches	Stable, high-quality beams through control of unmodeled accelerator behavior	Fast, high-quality control of extreme beams by exploiting learned FACET-II responses

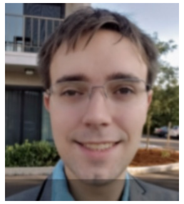
Synergistic experiments, individual success enhances all research

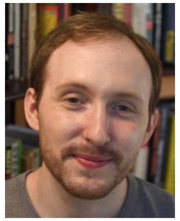
Backup Slides

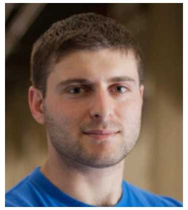
Collaboration

- SLAC: A. Edelen, C. Emma, R. Roussel, B. O'Shea, S. Miskovich, W. Neiswanger, G. White, S. Gessner, C. Mayes, D. Ratner
- LANL: A. Scheinker











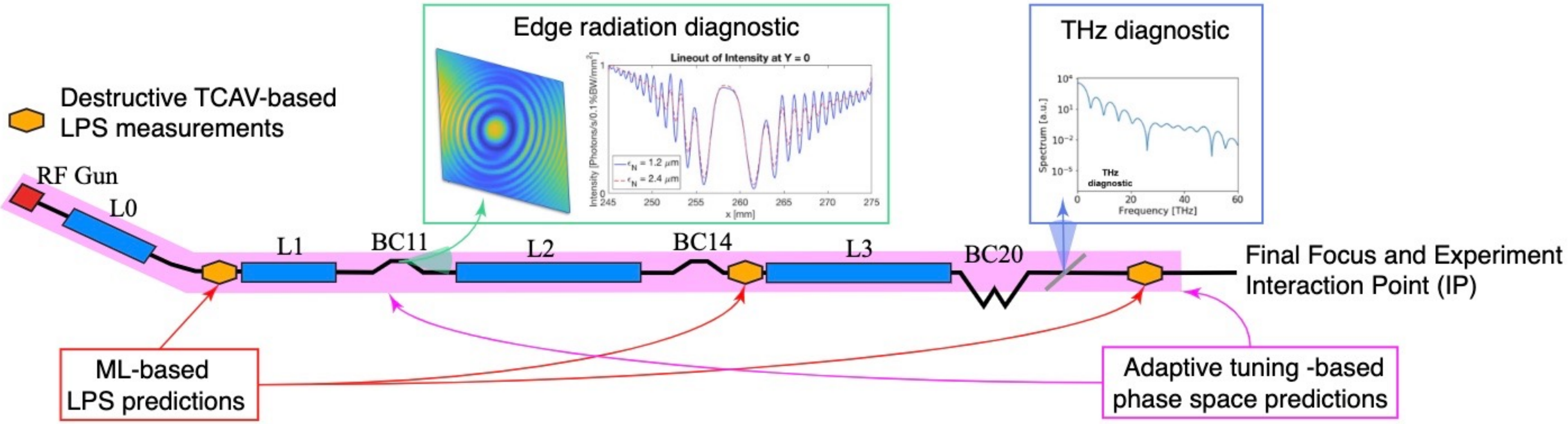
Publications

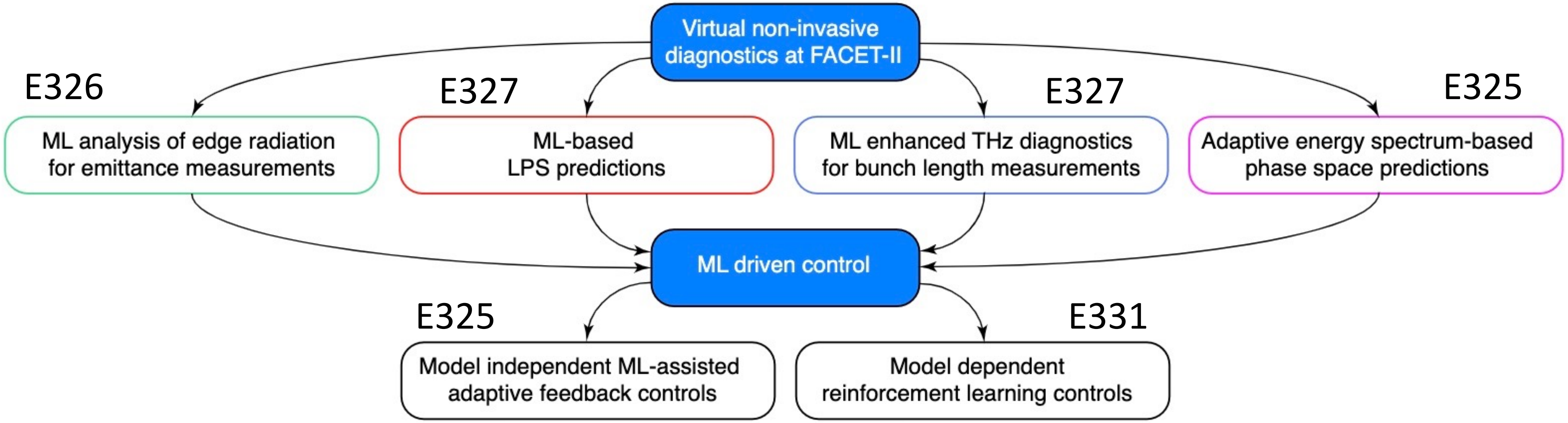
1. C. Emma, A. Edelen, A. Hanuka, B. O'Shea, A. Scheinker, "Virtual diagnostic suite for electron beam prediction and control at FACET-II" *Information* 2021, 12(2), 61; https://doi.org/10.3390/info12020061

2. S.A. Miskovich, A.L. Edelen, and C.E. Mayes, "PyEmittance: A General Python Package for Particle Beam Emittance Measurements with Adaptive Quadrupole Scans", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 1003-1005. doi:10.18429/JACoW-IPAC2022-TUPOST059, https://accelconf.web.cern.ch/ipac2022/papers/tupost059.pdf

3. S. Miskovich, W. Neiswanger, W. Colocho, C. Emma, J. Garrahan, T. Maxwell, S. Ermon, A. Edelen, D. Ratner, "Bayesian Algorithm Execution for Tuning Particle Accelerator Emittance with Partial Measurements," under review https://arxiv.org/abs/2209.04587

Landscape of AI/ML Activities at FACET-II





Synergistic experiments, individual success enhances all research + facility operation

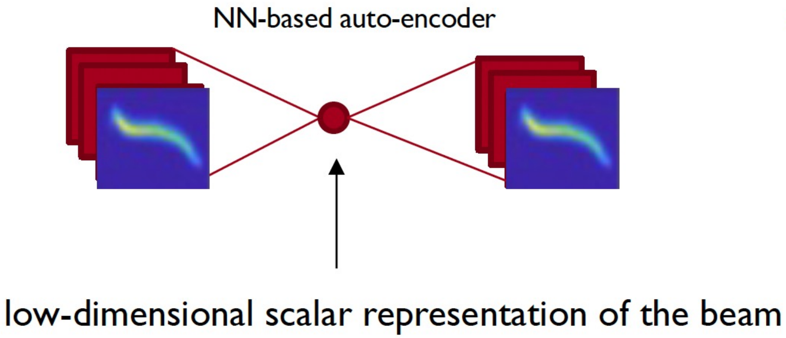
Fine-tuning with a learned metric

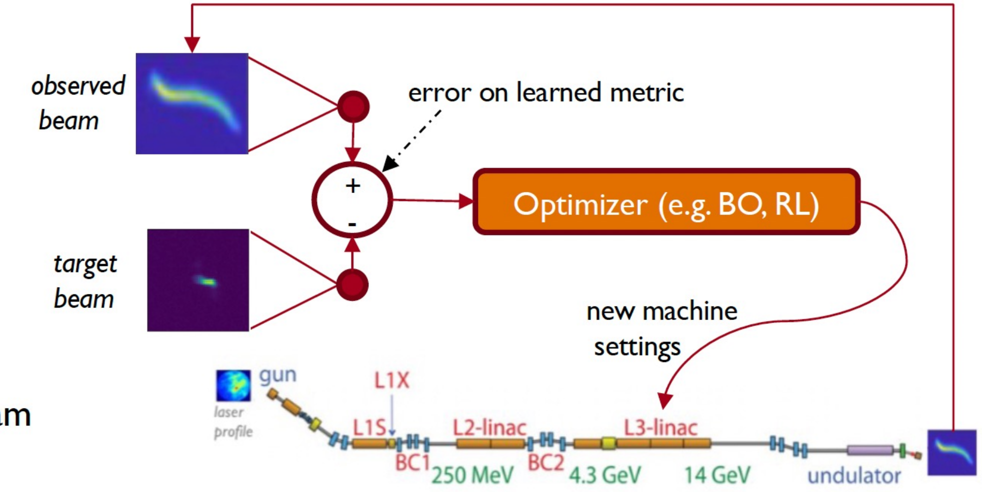
Could a learned metric:

(1) ignore irrelevant features/noise?

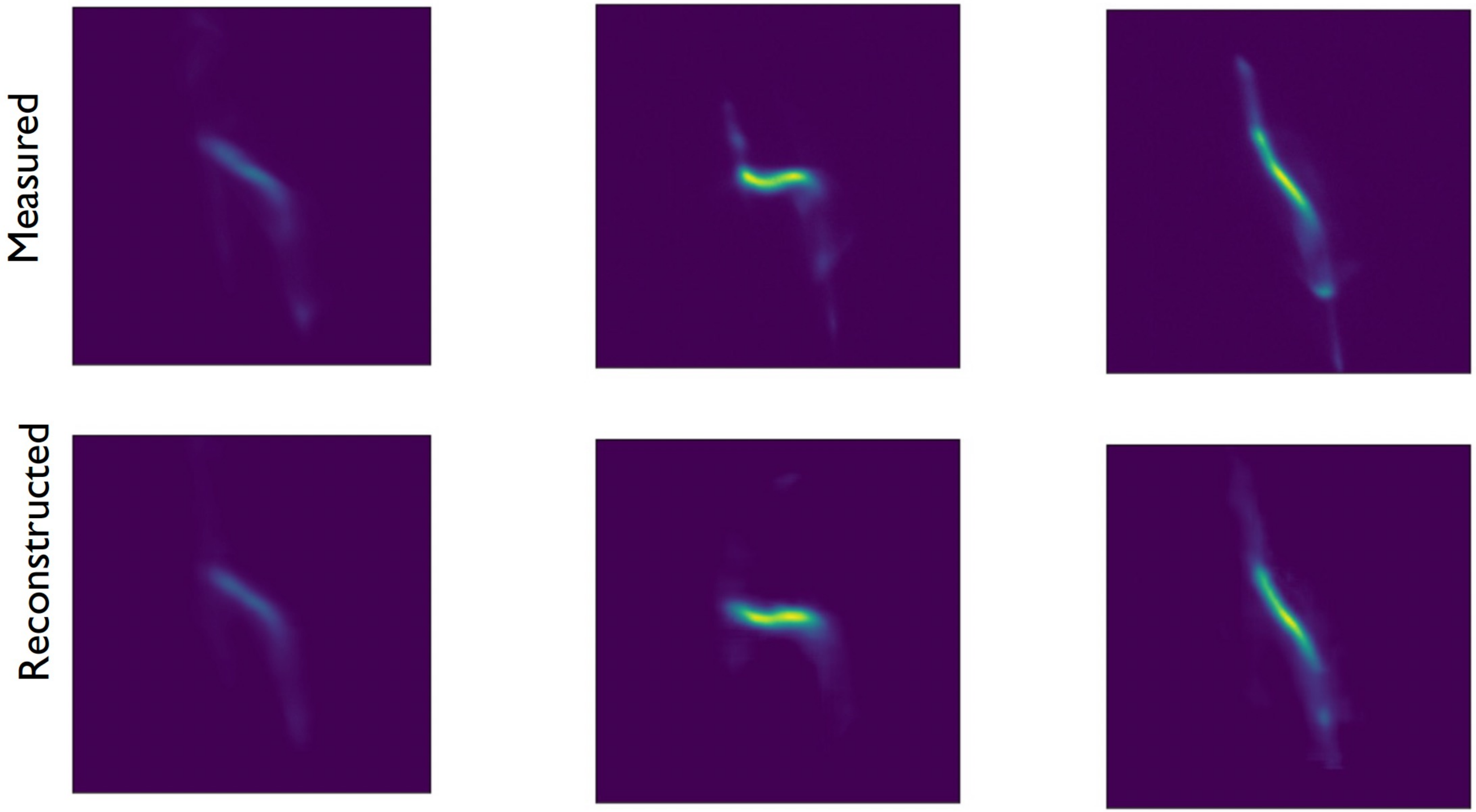
(2) be more robust to cropping/alignment issues?

(3) enable combination of metrics to be distilled into a few scalar values (e.g. photon-side metrics + LPS, multiple bunches) for simultaneous tuning?





Reconstruction examples



Two-dimensional latent space, 256x256 pixels

Target latent reconstruction

Target and final latent reconstruction

Target and final raw

Example from machine in SXR