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

(\textit{build up incrementally to machine-wide neural network-based reinforcement learning})

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

Staged approach gradually increases complexity
Success defined by performance (new capabilities) and transition to operations
E331 Diagnostic and Observables

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

- 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)
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/
Better Data Sampling: Bayesian Exploration

\[
\alpha(x) = \sigma(x) \prod_{i=1}^{N} \mathbb{P}(g_i(x) \geq h_i) \Psi(x, x_0)
\]

Enables sample-efficient characterization of high-dimensional spaces, while respecting both input and output constraints.
E331 Progress: ML for Efficient Characterization

• 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

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
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

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.

https://arxiv.org/abs/2209.04587
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

<table>
<thead>
<tr>
<th>Diagnostics</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge Radiation Emittance Diagnostics (E326)</td>
<td>Adaptive Feedback (E325)</td>
</tr>
<tr>
<td>Virtual TCAV Predictive Diagnostics (E327)</td>
<td>Learned Control (Reinforcement Learning, New proposal)</td>
</tr>
<tr>
<td>Non-destructive, single shot continuous monitoring of emittance of high-current beams</td>
<td>Stable, high-quality beams through control of unmodeled accelerator behavior</td>
</tr>
<tr>
<td>Longitudinal phase space diagnostics, always on, and for extremely short bunches</td>
<td>Fast, high-quality control of extreme beams by exploiting learned FACET-II responses</td>
</tr>
</tbody>
</table>

Synergistic experiments, individual success enhances all research
Backup Slides
Collaboration


• 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


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

Measured

Reconstructed

Two-dimensional latent space, 256x256 pixels
Example from machine in SXR

Target latent reconstruction

Target and final latent reconstruction

Target and final raw