Machine Learning-based Beam Size Stabilization at ALS

Simon C. Leemann
ALS Accelerator Physics, ATAP & ALS Divisions, Lawrence Berkeley National Laboratory
Feb 9, 2021

OWLE ML Seminar
The Advanced Light Source at Berkeley Lab
The Advanced Light Source at Berkeley Lab

184" cyclotron yoke (construction 1940)
The Advanced Light Source at Berkeley Lab

Triple-bend achromat lattice (12 sectors, 3 with superbends, 10 ID straights)
The Advanced Light Source at Berkeley Lab

≈40 beamlines, ≈5000 hrs/y, ≈2000 users/y
Many Successful Efforts to Stabilize Electron Beams

• **Top-off** keeps ALS stored current variation <0.2%

Courtesy: C. Steier, PAC’09
Many Successful Efforts to Stabilize Electron Beams

- **Top-off** keeps ALS stored current variation <0.2%
- At low energy, ALS strongly affected by insertion device (ID) imperfections & continuously changing EPU gaps/phases

![Graph with beam intensity over time](Courtesy: C. Steier, PAC'09)
Many Successful Efforts to Stabilize Electron Beams

- **Top-off** keeps ALS stored current variation <0.2%
- At low energy, ALS strongly affected by insertion device (ID) imperfections & continuously changing EPU gaps/phases
  - **Orbit feedback** and ID feed-forwards stabilize source positions/angles to **sub-micron** level at many tens of Hz

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 9/39
Many Successful Efforts to Stabilize Electron Beams

- **Top-off** keeps ALS stored current variation <0.2%
- At low energy, ALS strongly affected by insertion device (ID) imperfections & continuously changing EPU gaps/ phases
  - **Orbit feedback** and ID feed-forwards stabilize source positions/angles to sub-micron level at many tens of Hz
  - **ID feed-forwards** & tune feedback stabilize optics at source points
  - **ID skew feed-forwards** stabilize source size
    - require recording lookup tables (time consuming)
    - tables are imperfect and **machine drifts** over time

![Graphs and diagrams](Courtesy: C. Steier, PAC'09)

Thermal, Ground, Water Table, etc.

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 10/39
The Problem: Beam Size vs. ID Motion

• Nevertheless, during routine user ops observe vertical source size variations when ID configurations change

≈4% source size variation

PRL 123, 194801 (2019)

• Traditionally, 3rd-gen. sources considered <10% acceptable, but...

ALS Diagnostic Beamline 3.1

SR from 1st arc dipole ("round beam") → KB mirrors → C filter → 1-3 keV x-rays → LYSO scintillator crystal → visible → CCD

How this Problem Affects Sensitive Experiments

• Vertical source size fluctuations show up as intensity variations at highly sensitive beamlines, such as the STXM at ALS beamline 5.3.2.2
  – STXM zone plate focal length ≈1 mm → no independent & reliable $I_0$ measurement
  – Very small spot size in focus (>20 nm → scan >10×10 μm²)
  – Fast raster scanning for differential measurements → no averaging (≈1 ms/pixel, 1 s/line, 6 min/scan)
  – Monochromator plane is H → V source size fluctuations directly affect experimental noise floor

• 4th-gen. rings such as ALS-U will be equipped with many more such highly sensitive beamlines: STXM, XPCS, ptychography, etc.

PRL 123, 194801 (2019)
Need to Solve This Problem at the Source

• Why use **Machine Learning (ML)** to attack this issue?
  – ML can model highly nonlinear processes and is extremely flexible
  – ML does not require a priori understanding underlying physics (e.g. machine drift) → but allows extracting valuable system information a posteriori
  – ML can substantially outperform conventional fitting (polynomial regression)

• ML requires reproducible events → confirmed in experiments

• ML ideally requires large data sets for training → ALS digital control system can provide that
Need to Solve This Problem at the Source

- Why use **Machine Learning (ML)** to attack this issue?

- **Machine Learning** can model highly nonlinear processes and is extremely flexible.
- **ML** does not require a priori understanding underlying physics (e.g., machine drif)
  - but allows extracting valuable system information a posteriori.
- ML can substantially outperform conventional fitting (polynomial regression).
- **ML** requires reproducible events confirmed in experiments.
- **ML** ideally requires large datasets for training.
- **ALS** digital control system can provide that.

- First example: offline analysis of user ops data
  - 26 ID parameters ("input") \( \rightarrow \) predict V beam size @ BL3.1 ("output")
  - Recorded 8 Msamples @ 10 Hz \( \rightarrow \) 6 Msamples used for training, 2 Msamples for validation \( \rightarrow \) training took 30 min on powerful GPU.

The fitting result using model with three fully connected layers \([128, 64, 32]\) gives decent YEARMS prediction:

- **MSE**: 0.0230

There are 6000000 data points in the training dataset and 1000000 in evaluation.

- **MSE**: 0.0232

Courtesy: S. Liu

Prediction within 0.3% of measured beam size.
Need to Solve This Problem at the Source

- Why use Machine Learning (ML)?
  - ML can model highly nonlinear processes and is extremely flexible.
  - ML does not require a priori understanding of underlying physics (e.g., machine drift) but allows extraction of valuable system information a posteriori.
  - ML can substantially outperform conventional fitting (polynomial regression).
- ML requires reproducible events confirmed in experiments.
- ML ideally requires large data sets for training. ALS digital control system can provide that.

First example: offline analysis of user ops data

- 26 ID parameters ("input") → predict V beam size @ BL3.1 ("output")
- Recorded 8 Msamples @ 10 Hz → 6 Msamples used for training, 2 Msamples for validation → training took 30 min on powerful GPU.

ML prediction clearly outperforms polynomial regression.

PRL 123, 194801 (2019)
From Prediction to Correction

• Introduced "dispersion wave parameter" (DWP) to modify standard ALS dispersion wave (skew quadrupole excitation pattern) \(\rightarrow\) allows adjusting vertical emittance (global conserved quantity)

\[
\vec{K} = \vec{K}_0 + (\chi_0 + \chi) \Delta \vec{K}, \quad \vec{K} \in \mathbb{R}^{16+16}
\]

LOCO & Setup \(\rightarrow\) DWP

Dispersion Wave

SQSF \(\downarrow\) SQSD

Current ALS Optics

"Dispersion" Function (122 BPMs)

EPAC 2000, TUP3A17, p.1098
From Prediction to Correction (cont.)

- Introduced local skew correction via non-standard skew quadrupole adjustment.

\[ \vec{K} = \vec{K}_0 + (0 + 0) \sim \vec{K}, \sim \vec{K}_0^2 \]

LOCO & Setup

Dispersion Wave

Local Skew Correction

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 17/39
From Prediction to Correction (cont.)

• Introduced "dispersion wave parameter" (DWP) to modify standard ALS dispersion wave (skew quadrupole excitation pattern) → allows adjusting vertical emittance (global conserved quantity)

\[ \vec{K} = \vec{K}_0 + (\chi_0 + \chi)\Delta \vec{K}, \quad \vec{K} \in \mathbb{R}^{16+16} \]

• Observed varying ID configurations affect primarily vertical dispersion \( \varepsilon_y \)
• Can therefore stabilize beam size globally by adjusting DWP

EPAC 2000, TUP3A17, p.1098

Current ALS Optics
Building a NN-based ID Feed-Forward

- Training: measure beam sizes while scanning DWP & various ID configurations → acquire data at 10 Hz → input for **training** of NN (DL)

Deep Learning

\[ \text{ID}s \quad \sigma_y \quad \text{Skews} \quad \text{Input} \rightarrow \text{Training} \rightarrow \text{NN} \quad \text{Output} \]

- Requires only large amounts of data & reproducibility
Building a NN-based ID Feed-Forward

• Training: measure beam sizes while scanning DWP & various ID configurations → acquire data at 10 Hz → input for training of NN (DL)

• Result of DL is prediction for DWP required to keep beam size constant for arbitrary ID configurations → run as NN-based ID FF

Deep Learning

<table>
<thead>
<tr>
<th>Input</th>
<th>Training</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDs</td>
<td>σ_y</td>
<td>NN</td>
</tr>
<tr>
<td>Skews</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Application during ops

<table>
<thead>
<tr>
<th>Input</th>
<th>NN</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDs</td>
<td>σ_y</td>
<td>Skews</td>
</tr>
<tr>
<td></td>
<td>Prediction</td>
<td>Set Target σ_y</td>
</tr>
</tbody>
</table>

• Requires only large amounts of data & reproducibility

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 20/39
How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works

How a Neural Network (NN) Works
Deep Learning: How we Trained the NN

Input Layer: ID settings (22-35 Dimension) and DWP (1 Dimension)
Three Hidden Fully Connected Layers:
128, 64, 32 neurons in each layer
Output Layer: Vertical Beam Size (1 Dimension)

Regularization: L2 regularizer with $\lambda = 10^{-4}$
Optimization: Adam Optimizer with learning rate $\alpha = 10^{-3}$

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Training MSE</th>
<th>Evaluation MSE</th>
<th>Training MSE</th>
<th>Evaluation MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>128-64</td>
<td>0.0265</td>
<td>0.0268</td>
<td>0.0257</td>
<td>0.0260</td>
</tr>
<tr>
<td>256-64</td>
<td>0.0243</td>
<td>0.0245</td>
<td>0.0259</td>
<td>0.0262</td>
</tr>
<tr>
<td>512-128</td>
<td>0.0243</td>
<td>0.0247</td>
<td>0.0243</td>
<td>0.0247</td>
</tr>
<tr>
<td>128-64-32</td>
<td>0.0238</td>
<td>0.0242</td>
<td>0.0243</td>
<td>0.0245</td>
</tr>
<tr>
<td>256-128-64</td>
<td>0.0236</td>
<td>0.0240</td>
<td>0.0240</td>
<td>0.0246</td>
</tr>
<tr>
<td>256-128-64-32</td>
<td>0.0245</td>
<td>0.0249</td>
<td>0.0245</td>
<td>0.0248</td>
</tr>
</tbody>
</table>

PRL 123, 194801 (2019)
Resulting NN Enables ID Feed-Forward at ≈3 Hz

**Proposed DWP**

<table>
<thead>
<tr>
<th>Proposed DWPs</th>
<th>Predicted Beam Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.06</td>
<td>50.3</td>
</tr>
<tr>
<td>0</td>
<td>52.1</td>
</tr>
<tr>
<td>0.06</td>
<td>54.0</td>
</tr>
</tbody>
</table>

Compare with Target Beam Size

Choose proper DWP

Insertion Device (ID) Settings

Neural Network

Predicted Beam Size

**ID**s

Prediction

Skews

Set Target \( \sigma_y \)

PRL 123, 194801 (2019)
Physics Shift: Data Collection for NN Training

Example from 11 h training @ 500 mA top-off

DWP

±0.06 units

Scanning ID gaps (and shifts)

48 μm

±4.75 μm
Physics Shift: Running NN-based ID Feed-Forward

Evaluation @ 500 mA top-off

Training required ≈15 min on single core

7.5 μm p-p (15%)
1.5 μm rms (3%)

1.9 μm p-p (4%)
0.2 μm rms (0.4%)

DWP

σγ

FF off

FF on

48 μm

Training required ≈15 min on single core

7.5 μm p-p (15%)
1.5 μm rms (3%)

1.9 μm p-p (4%)
0.2 μm rms (0.4%)

DWP

σγ

FF off

FF on

48 μm
Physics Shift: Running NN-based ID Feed-Forward

Evaluation @ 500 mA top-off

STXM images BL 5.3.2.2 (D. Shapiro & M. Marcus)

- 3.2% variation
- 0.5% variation
- 0.8% variation

DWP

FF off

FF on

σ_y

7.5 μm p-p (15%)
1.9 μm p-p (4%)
1.5 μm rms (3%)
0.2 μm rms (0.4%)
0.5% variation

FF off

FF on

Evaluated on @ 500 mA top-off
Physics Shift: Running NN-based ID Feed-Forward

Evaluation @ 500 mA top-off

STXM images BL 5.3.2.2 (D. Shapiro & M. Marcus)

Evaluation @ 500 mA top-off

Physics Shift: Running NN-based ID Feed-Forward

Evaluation @ 500 mA top-off

Physics Shift: Running NN-based ID Feed-Forward

Physics Shift: Running NN-based ID Feed-Forward
Physics Shift: Running NN-based ID Feed-Forward

Evaluation @ 500 mA top-off

STXM images BL 5.3.2.2 (D. Shapiro & M. Marcus)

- No ID motion
- FF off
- 12 dB

- FF on

No FF Baseline
No FF, single ID move
No FF, multi ID move
FF, multi ID move

3.2% variation
0.5% variation
0.8% variation

Simulation

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 28/39
Towards First Experiments During User Ops

• Use machine shift to acquire training data by scanning operational IDs in a quasi-randomized fashion (favoring operational gap range) → train NN
• Put this NN into FF operation during user ops and evaluate
Stabilization Confirmed During First User Ops Trial

User Ops @ 500 mA top-off

DWP = FF action on dispersion wave

0.42 μm rms (0.8%)

0.18 μm rms (0.3%)

ID gaps & shifts moving during user ops
(only subset shown here)
Stabilization Confirmed During First User Ops Trial

User Ops @ 500 mA top-off

DWP = FF action on dispersion wave

Stepping related to EPU not included in original training

\[ \sigma_x = 3.5 \mu m \ (7\%) \]

\[ \sigma_y = 0.42 \mu m \ rms \ (0.8\%) \]

\[ DWP = FF \ action \ on \ dispersion \ wave \]

ID gaps & shifts moving during user ops

(only subset shown here)
Stabilization Confirmed During First User Ops Trial

User Ops @ 500 mA top-off

DWP = FF action on dispersion wave

Online Retraining?

ID gaps & shifts moving during user ops (only subset shown here)

0.42 μm rms (0.8%)

0.18 μm rms (0.3%)

PRL 123, 194801 (2019)
Online Retraining: Improve NN with User Ops Data

So far: "Conventional" Machine Learning

Scanning during Acc Phys Shift → Neural Network (Predictive Model) → Feed-Forward → User Ops

Online Retraining: Improve NN with User Ops Data

Online Retraining: apply user ops data to improve NN → swap NN used for ID FF on the fly

NN can be continuously online retrained during user ops to improve FF performance (exploiting huge amounts of data acquired during user ops)

PRL 123, 194801 (2019)
Substantial Improvement After Online Retraining

User Ops @ 500 mA top-off

Online Retrained NN in FF Ops

NN-based FF off

0.16 μm rms (0.3%)

0.20 μm rms (0.4%)

ID gaps & shifts moving during user ops (only subset shown here)

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 35/39
Substantial Improvement After Online Retraining

User Ops @ 500 mA top-off

Online Retrained NN in FF Ops

NN-based FF on gaps & shifts moving during user ops (only subset shown here)

0.6% variation

DWP

0.16 μm rms (0.3%)

0.20 μm rms (0.4%)

Simon C. Leemann • Machine Learning-based Beam Size Stabilization at ALS
OWLE ML Seminar • Feb 9, 2021 • 36/39

PRL 123, 194801 (2019)
Results: NN-based FF Off vs. On During User Ops

NN-based FF off

ID phase switching

\[ \sigma_x \]

\[ \sigma_y \]

0.93 \( \mu \text{m} \) rms (1.8%)

NN-based FF on

FF action

\[ \sigma_x \]

\[ \sigma_y \]

0.20 \( \mu \text{m} \) rms (0.4%)

User Ops @ 500 mA top-off

PRL 123, 194801 (2019)
Stabilization Confirmed at Experiment

ALS Beamline 5.3.2.2

Noise reduced to almost floor level

Online Retraining

With ID motion, online-retrained NN

0.6% rms intensity

0.5% rms intensity

No ID motion (STXM noise floor)

3.2% rms intensity

With ID motion

NN-based FF on

PRL 123, 194801 (2019)
Thank You!

Acknowledgments:
Shuai Liu, Hiroshi Nishimura, Matthew A. Marcus, David Shapiro, Changchun Sun, Nathan Melton, Alex Hexemer, Dani Ushizima, Mike Ehrlichman, Gregg Penn, Thorsten Hellert, Yuping Lu, Erik Wallen, Warren Byrne, Fernando Sannibale, Marco Venturini, Andreas Scholl