Beam Based Optimization and Machine Learning for Synchrotrons

Simon C. Leemann & Alex Hexemer
ALS Accelerator Physics, ATAP & ALS Divisions, Lawrence Berkeley National Laboratory
April 30, 2019

ADRP Comparative Review, Germantown, MD, April 29 – May 1, 2019

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY | Office of Science

ALS ADVANCED LIGHT SOURCE

Project Info

- Project Title: ***Beam Based Optimization and Machine Learning for Synchrotrons***
- PI(s): ***Simon C. Leemann, Alex Hexemer***
- Institution: ***ATAP & ALS Divisions, LBNL***
- Collaborations: ***SSRL/SLAC (Xiaobiao Huang)***
- Begin Date: ***Sep 1, 2018***
- End Date: ***Aug 31, 2020***

- Presenter Name: ***Simon C. Leemann***

Project Goals and Accomplishments

Major goals:

- *Stabilize source size in ALS with ML-based feed-forward*
- *Use ML to assist stochastic optimization tools (e.g. MOGA)*

Accomplishments:

- *Trained NN to make predictions for vertical beam size as function of IDs & skews*
- *Used NN to run feed-forward stabilizing beam size in ALS*

Within schedule and budget? **Yes, under budget (still recruiting postdoc)**

1-2 highlights:

- *Vertical beam size stabilized during user ops to sub-percent rms level*
- *Observed ≈4-fold reduction of rms intensity fluctuations at STXM end station*

Project Background

- Collaboration with SSRL (X. Huang)
  - Share common theme: applying Machine Learning (ML) to the design, operation, and optimization of storage ring light sources
  - Both teams collaborate on topics of common interests while each team has its own emphasis related to its machine (ALS & SPEAR3)
- Funded Aug 2018 by BES (E. Lessner) & ASCR (T. Ndousse-Fetter) for 2 years
- So far, most studies applying ML to operational accelerators have been on FELs or colliders (high-energy physics)
- ML for storage ring light source operation has so far not been studied (to our knowledge) or published extensively
- Will leverage machine time available at ALS & SSRL as well as controls/ computing resources already available to the teams at both facilities

Common Proposal Theme

- Storage ring light sources presently undergoing a major transformation: multibend achromats (MAX IV), diffraction-limited storage rings, 4th-gen. storage rings, round beams, on-axis injection, Delta undulators, experiments exploiting high degrees of coherence, etc.
  (APS-U, ALS-U, and many international projects: SIRIUS, ESRF-EBS, HEPS, SLS-2, SOLEIL, Diamond-II, etc.)
- These rings
  - will require highly multivariate design optimization within an immense parameter space
  - will be much more complex to commission and optimize → require beam-based optimization (BBO)
  - will have tighter requirements for stability and field quality & will also be more sensitive to time-varying imperfections → diagnose & correct online
- **Goal: combine ML expertise with accelerator ops/dev expertise to enable and integrate novel solutions to meet these new challenges**

Common Proposal Theme in More Detail

- ML for Modeling & Lattice Design Optimization
  - Stochastic optimization (e.g. MOGA, MOPSO) powerful, but inefficient
  - Optimization often involves competing requirements & nonlinear tuning parameters
  - ➡ Supervised learning can recognize patterns in parameter space → apply to selection process or operation generating new solutions
  - ➡ ML can accelerate identification of elite solutions & reveal connections between optimization knobs and performance
- BBO and ML for Performance Tuning
  - ➡ Modify BBO algorithms and improve performance under noise & apply reinforcement learning techniques for online optimization
- ML for Compensation of Time-Varying Processes
  - ➡ Apply ML to compensate for ID motion without relying on static models based on time-consuming offline data collection (→ follow drifting machine)

Design Optimization

Ops Improvements

BERKELEY LAB

ALS

ALS Timeline & Deliverables

- First Year
  - Recruitment and onboarding of two new collaborators
  - Familiarization with ML & ML tools
  - Familiarization with ALS instrumentation, controls, and ops software integration (data acquisition)
  - Tuning of ML algorithms for predictions based on acquired machine data (training)
  - Study applying ML-based corrections through control system
  - Study of lattice design/optimization tools & ML modeling/evaluation tools (pattern recognition)
- Second Year
  - Integration of online corrections into ops software & performance validation
  - Study using ML to evaluate physics properties of candidate solutions (in e.g. MOGA)
  - Integration of ML tools into lattice design/optimization workflow
  - Validation of ML-enhanced design process
- Deliverables:
  - Research papers in peer-reviewed journals → *first being prepared for submission (also: Ph.D. Thesis UCB)*
  - Conference/workshop publications (talks & posters) → *first submitted to IPAC 2019, abstracts submitted for ICALEPCS 2019 & NA-PAC 2019; invited oral possible for IPAC 2020*
  - Detailed final report (Technical Report)
  - Repository of computer codes, simulation output data, and acquired experimental data (archived historical data) along with relevant documentation (formats, interfaces, configurations, etc.)

Ops

Dev

ALS Collaborators

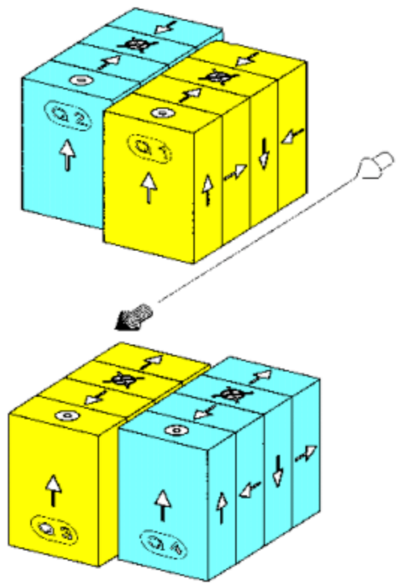
- 1 Postdoc (emphasis on ML expertise) funded by ASCR to work on this project 100% ➔ recruitment ongoing
- 1 Research/Staff Scientist (20%) to support ML for acc ops (Hiroshi Nishimura, now retired) ➔ search for successor ongoing
- 1 Research Scientist (10%) to support ML for design optimization (Changchun Sun)
- 1 Grad Student (Shuai Liu) & Postdoc (Nathan Melton) from ALS Computing Group (A. Hexemer) & LBNL CR Div (D. Ushizima) heavily involved in project while we recruit ASCR-funded Postdoc
- Involved also various other members of ALS AP Group & ATAP Div interested in ML applications at the ALS (F. Sannibale, M. Venturini, G. Penn, T. Hellert, M. Ehrlichman)

First Studies at ALS

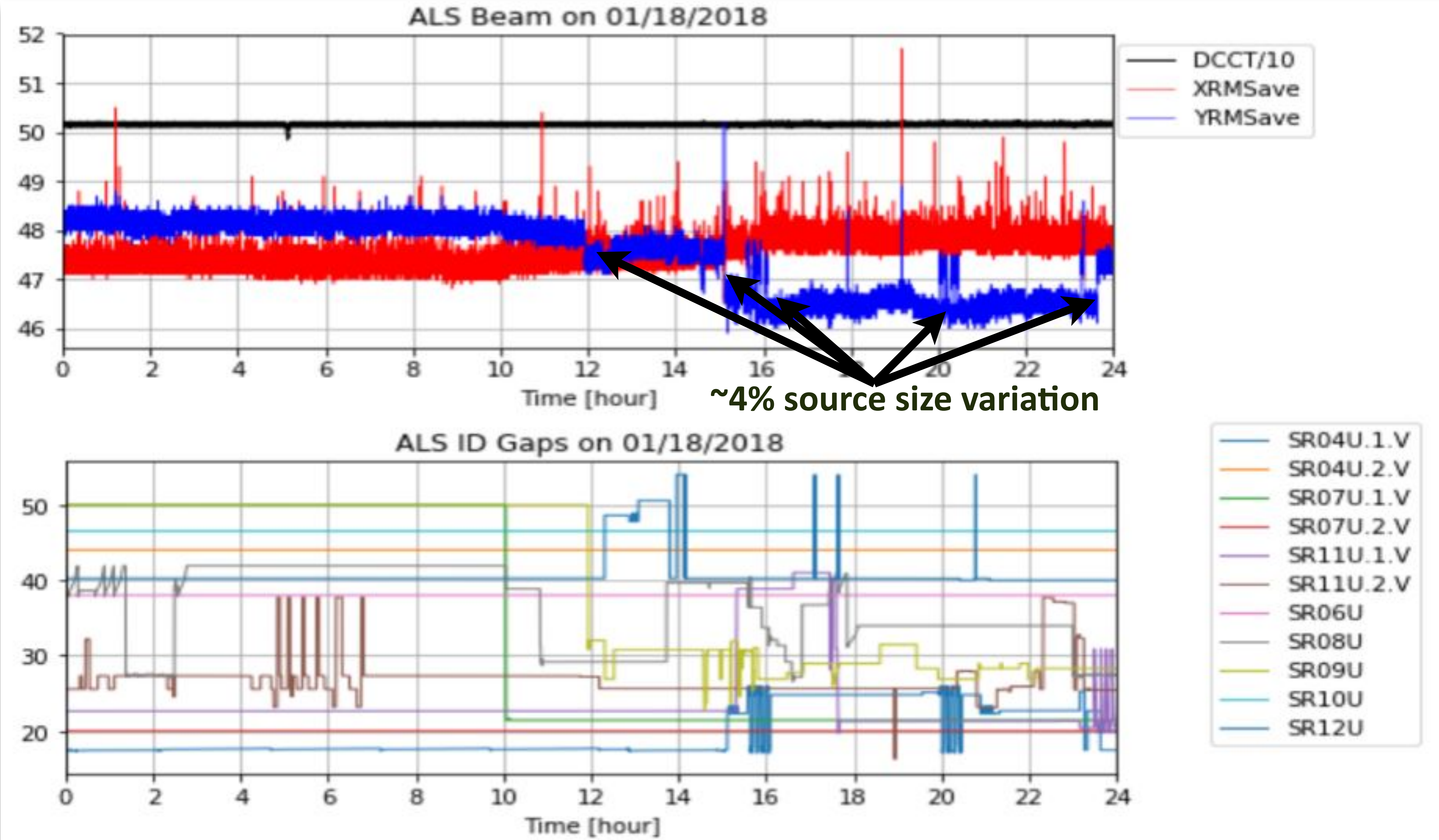
Using ML to Stabilize Source Size against ID Motion

The Problem: Beam Size vs. IDs

- ALS has various IDs with constantly changing gaps and phases
  - There are feed-forwards & feedbacks stabilizing the source positions/angles
  - There are feed-forwards & feedbacks stabilizing optics at the source points → local (beta) and global (tune)
  - There are feed-forwards to stabilize the source size
    - these require recording lookup tables
    - tables are imperfect and machine drifts over time
    - at low energy, ALS is strongly affected by EPUs & ID imperfections
- Nevertheless, during routine user ops source size changes @ BL3.1

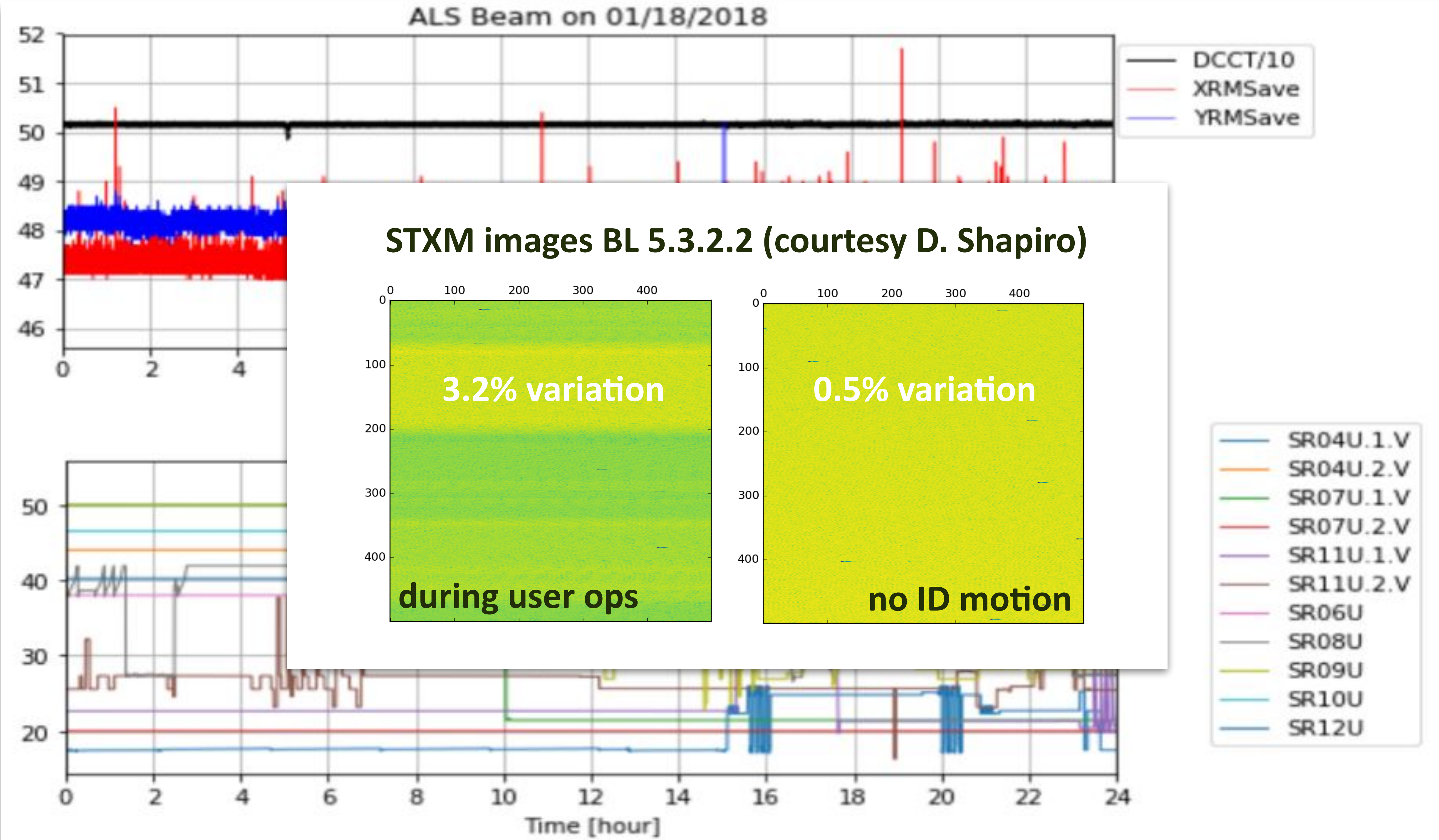


The Problem: Beam Size vs. IDs (cont.)



Courtesy: H. Nishimura

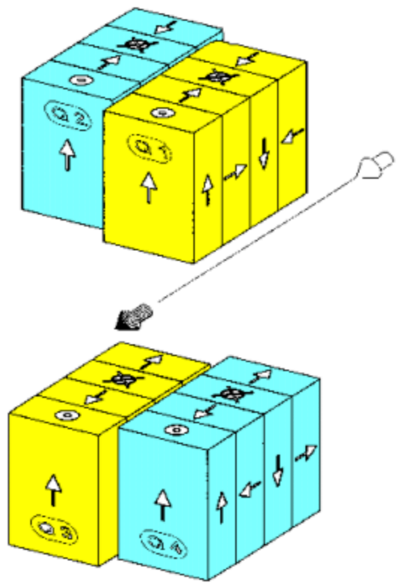
The Problem: Beam Size vs. IDs (cont.)



Courtesy: H. Nishimura

The Problem: Beam Size vs. IDs (cont.)

- ALS has various IDs with constantly changing gaps and phases
  - There are feed-forwards & feedbacks stabilizing the source positions/angles
  - There are feed-forwards & feedbacks stabilizing optics at the source points → local (beta) and global (tune)
  - There are feed-forwards to stabilize the source size
    - these require recording lookup tables
    - tables are imperfect and machine drifts over time
    - at low energy, ALS is strongly affected by EPUs & ID imperfections



- Nevertheless, during routine user ops source size changes @ BL3.1

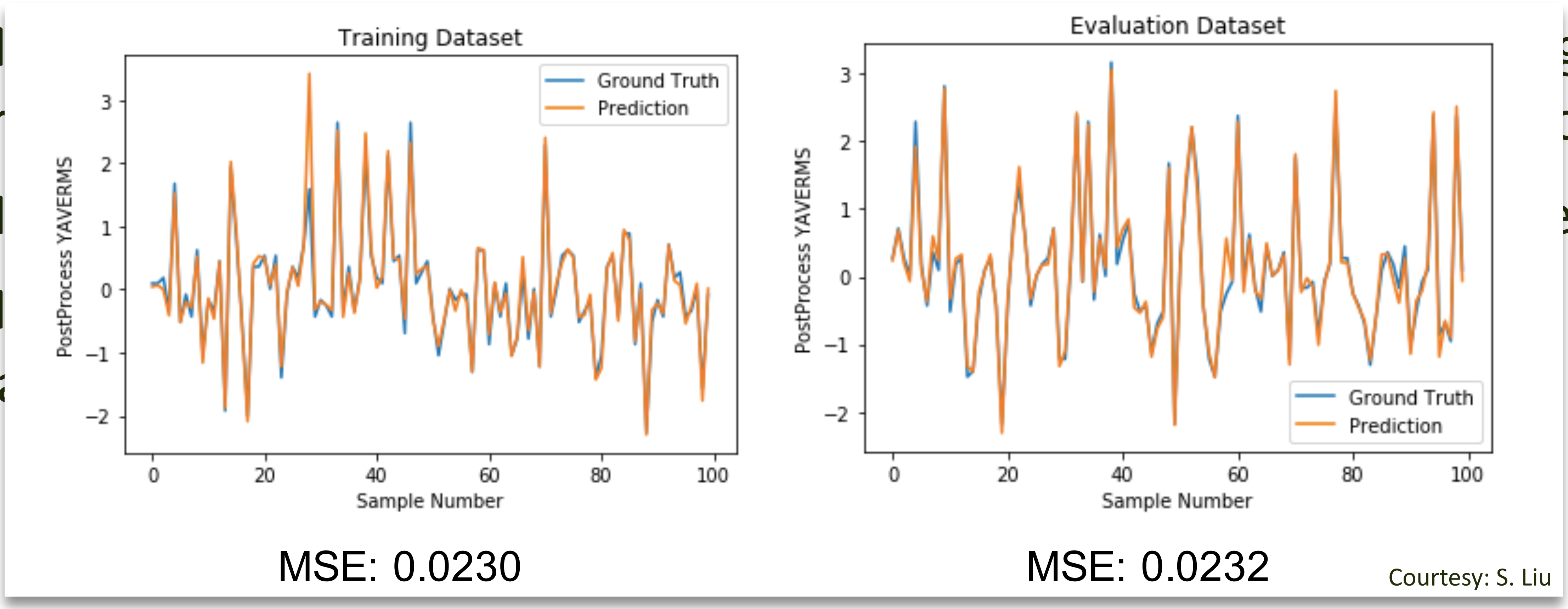
➡**Use ML to predict beam size changes as a function of arbitrary ID configuration → adjust skew quadrupoles to compensate**

Why Use ML?

- ML does not require a priori understanding underlying physics (e.g. drift), but allows extracting valuable system information a posteriori
- ML can model highly nonlinear processes and is extremely flexible
- ML requires reproducible events and ideally needs large data sets for training → ALS has huge amounts of data to offer

Why Use ML? (cont.)

- M … g. dr … ori
- M … e
- M … for tr…



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

From Prediction to Correction

- Introduced a scaling ("DWP") to standard ALS dispersion wave (skew quadrupole excitation pattern) ➔ allows adjusting vertical emittance (global conserved quantity)

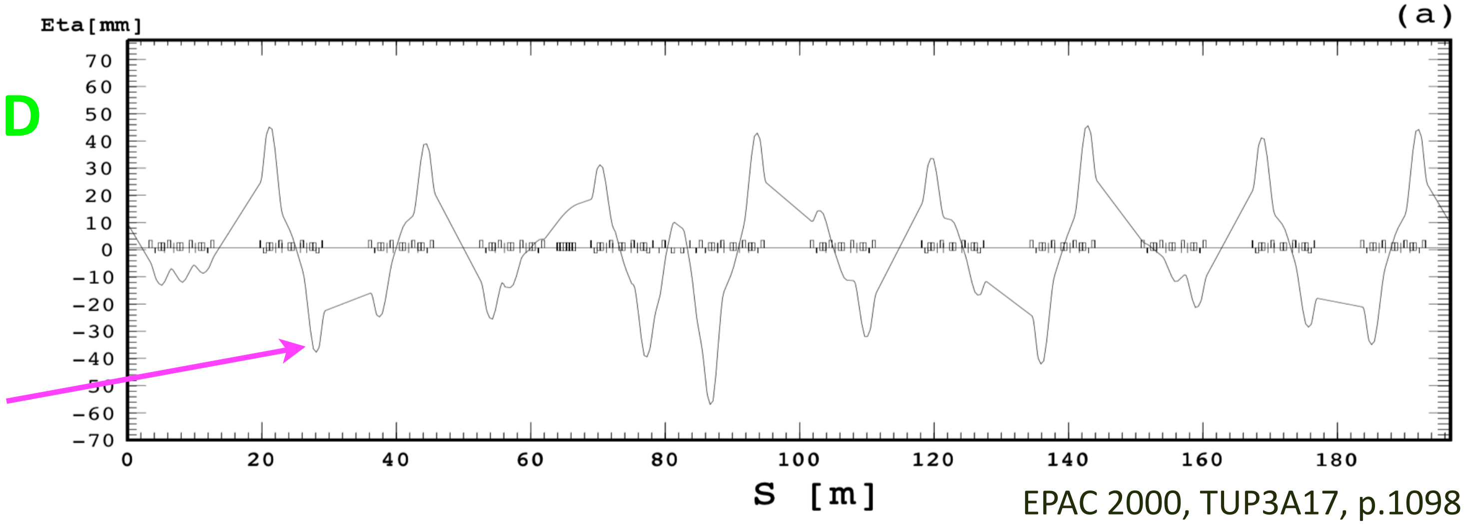
SQSF SQSD

$$\vec{K} = \underbrace{\vec{K}_0}_{\text{LOCO \& Setup}} + (\underbrace{\chi_0}_{\text{LOCO \& Setup}} + \underbrace{\chi}_{\text{DWP}})\underbrace{\Delta\vec{K}}_{\text{Dispersion Wave}}, \quad \vec{K} \in \mathcal{R}^{\overset{\text{SQSF}}{16}+\overset{\text{SQSD}}{16}}$$

LOCO & Setup

DWP

Dispersion Wave



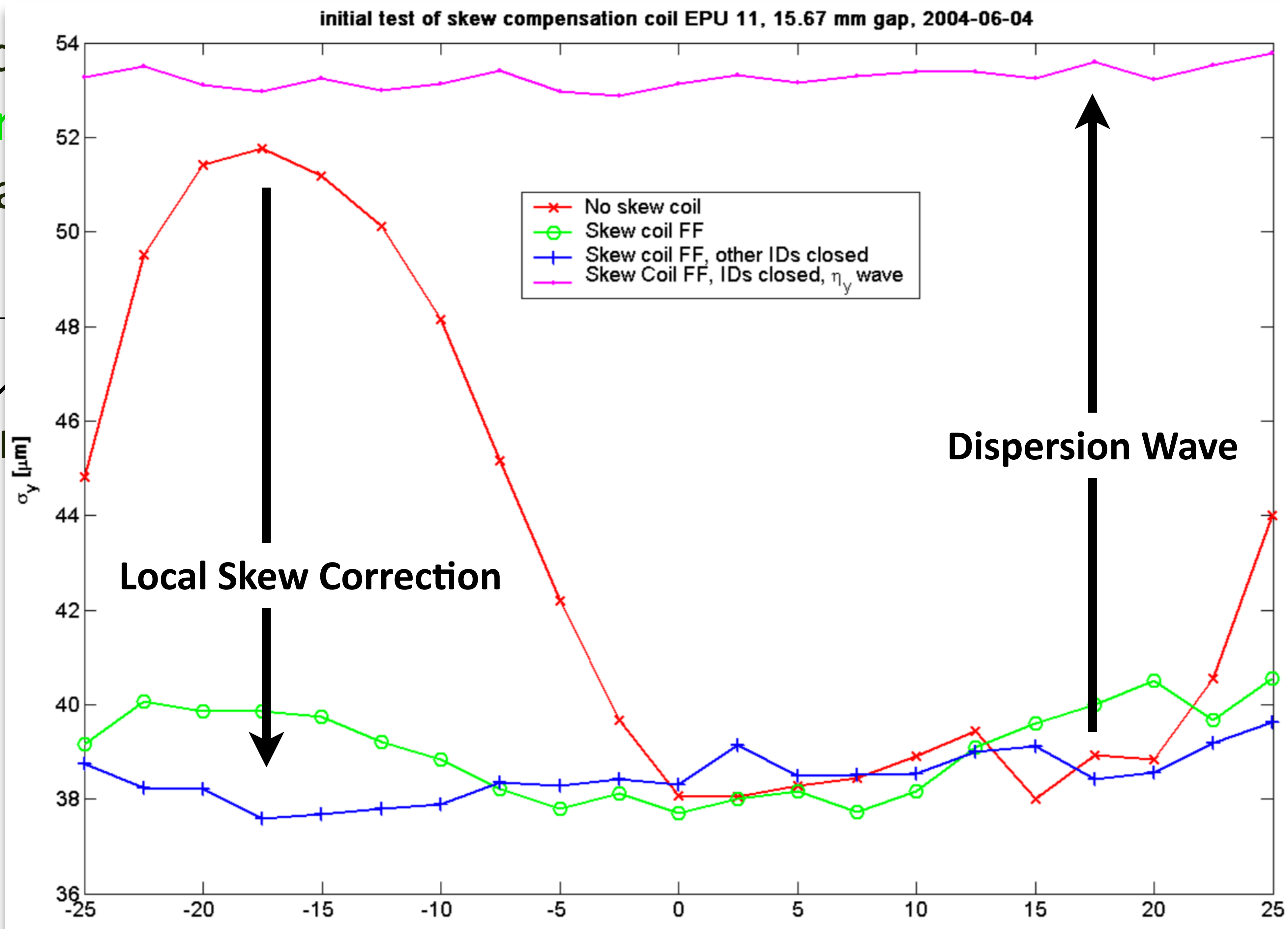
EPAC 2000, TUP3A17, p.1098

From Prediction to Correction (cont.)

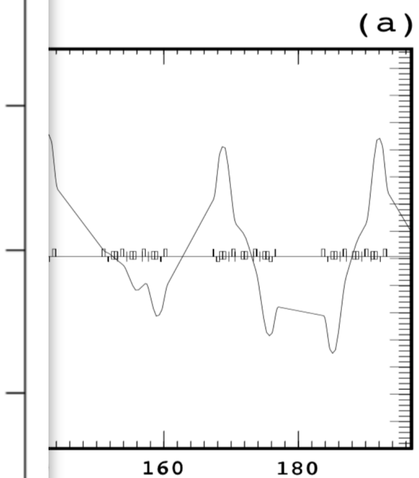
- Introc (skew quadr tance (globa

$\vec{K} = \vec{K}_0 +$

LOCO & Setu



EPAC 2004, MOPKF071, p.479



C 2000, TUP3A17, p.1098

From Prediction to Correction (cont.)

- Introduced a scaling ("DWP") to standard ALS dispersion wave (skew quadrupole excitation pattern) ➔ allows adjusting vertical emittance (global conserved quantity)

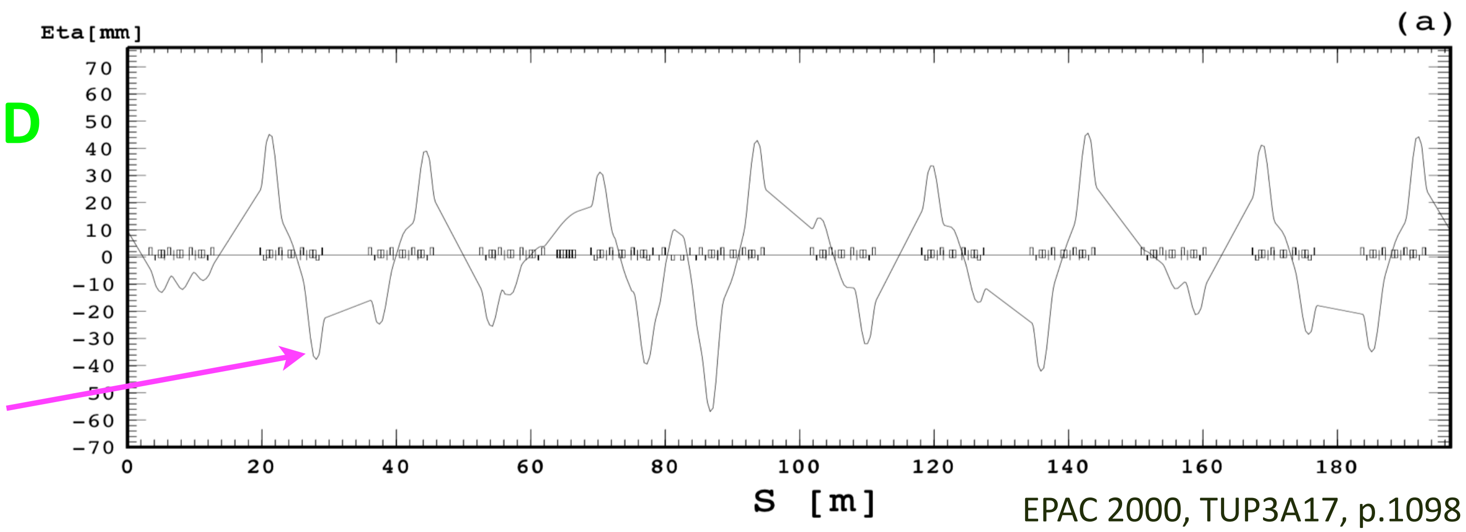
SQSF SQSD

$$\vec{K} = \vec{K}_0 + (\chi_0 + \chi)\Delta\vec{K}, \quad \vec{K} \in \mathcal{R}^{16+16}$$

LOCO & Setup

DWP

Dispersion Wave



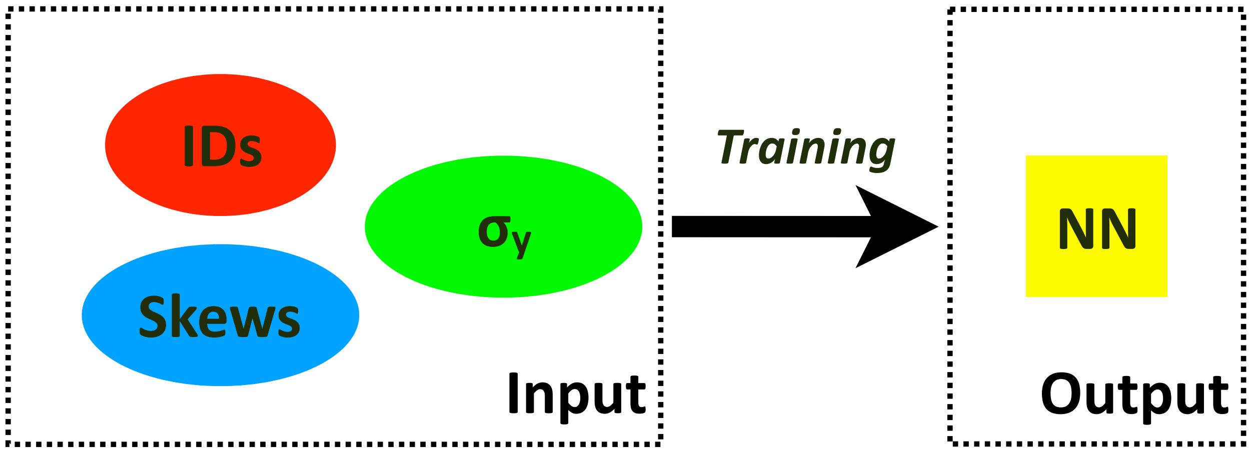
EPAC 2000, TUP3A17, p.1098

- Run simple PID feedback (FB) loop designed to stabilize source size as measured at BL3.1 by adjusting DWP ➔ scanning ID configurations & acquire data at 10 Hz ➔ input for training of NN
- Result is prediction for DWP required to keep beam size constant for arbitrary ID configurations ➔ works, but still have to overcome typical FB issues (transients, loop stability, etc.) ➔ prefer a feed-forward (FF)

A Better Idea: NN-based ID Feed-Forward

- Instead of dealing with FB issues, scan DWP while scanning ID configurations → acquire data at 10 Hz → input for training of NN (DL)

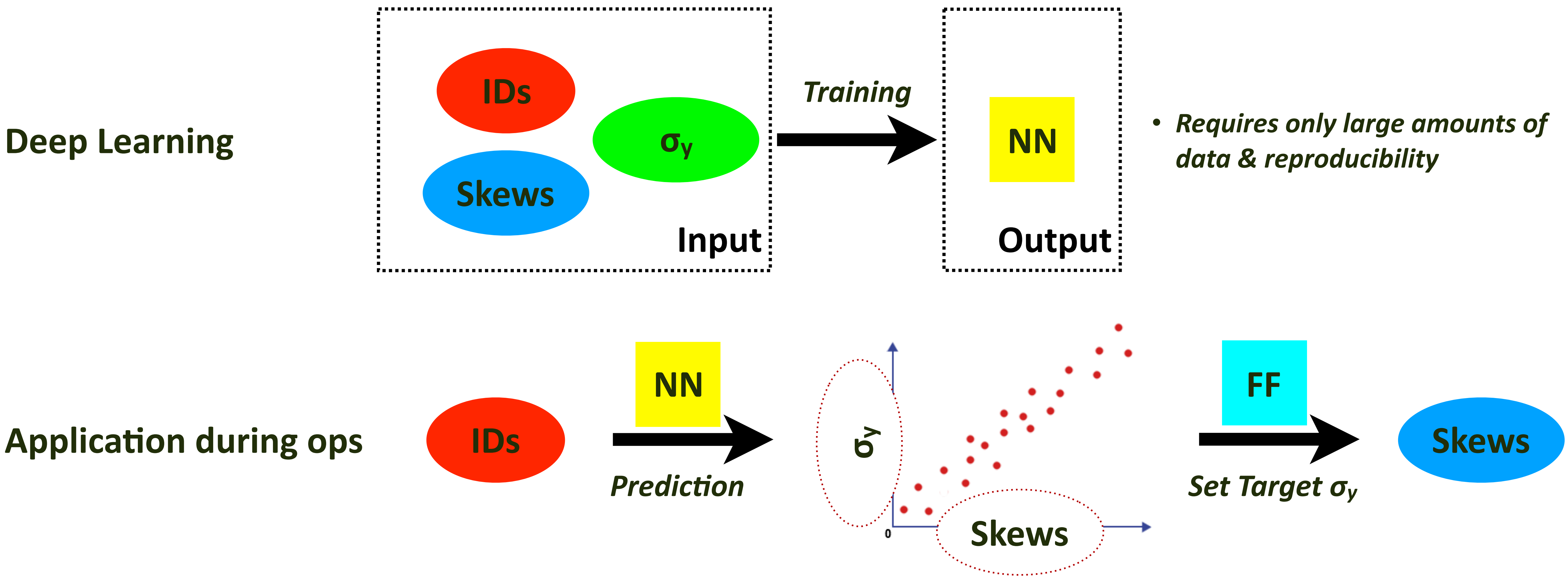
Deep Learning



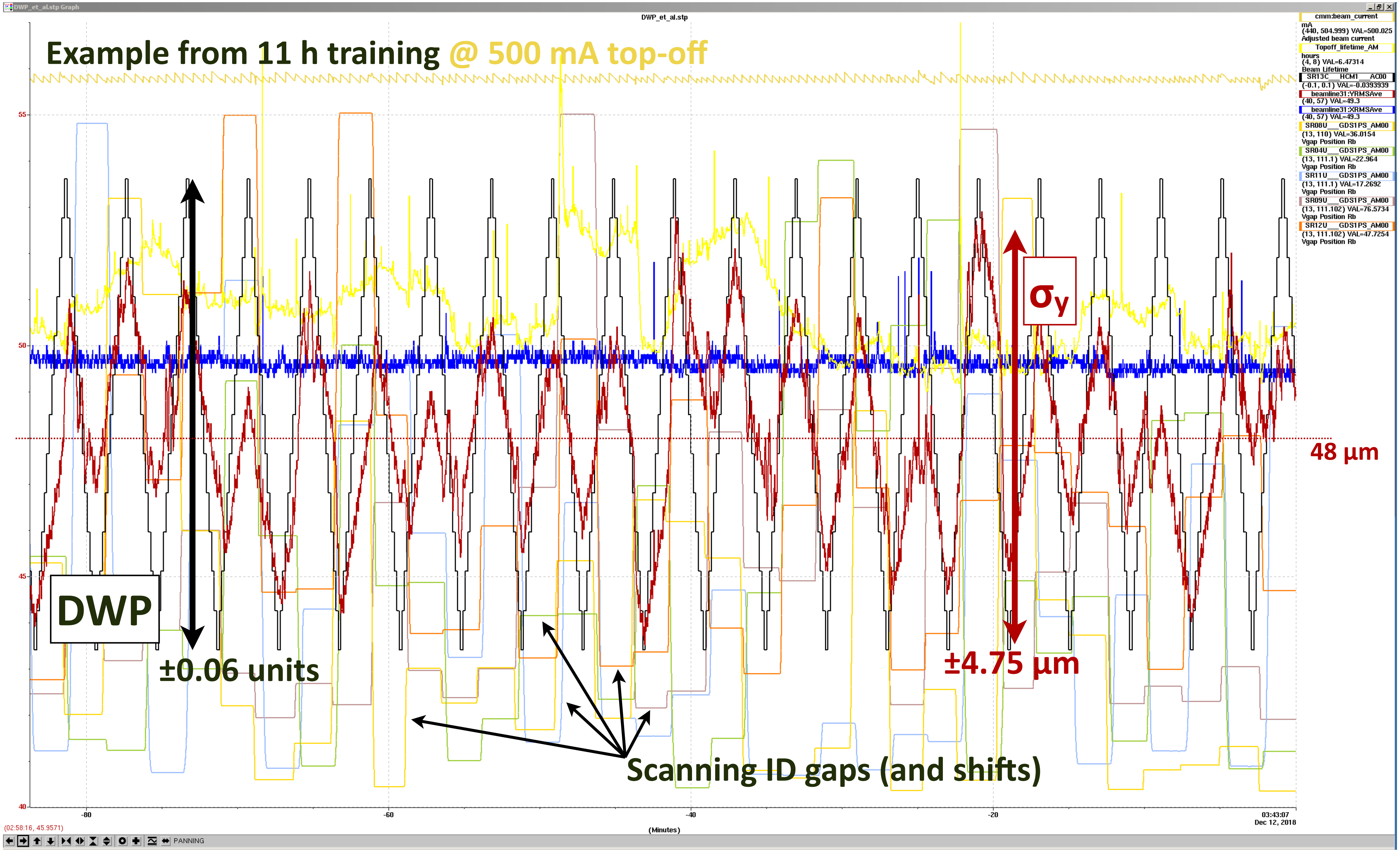
- ***Requires only large amounts of data & reproducibility***

A Better Idea: NN-based ID Feed-Forward

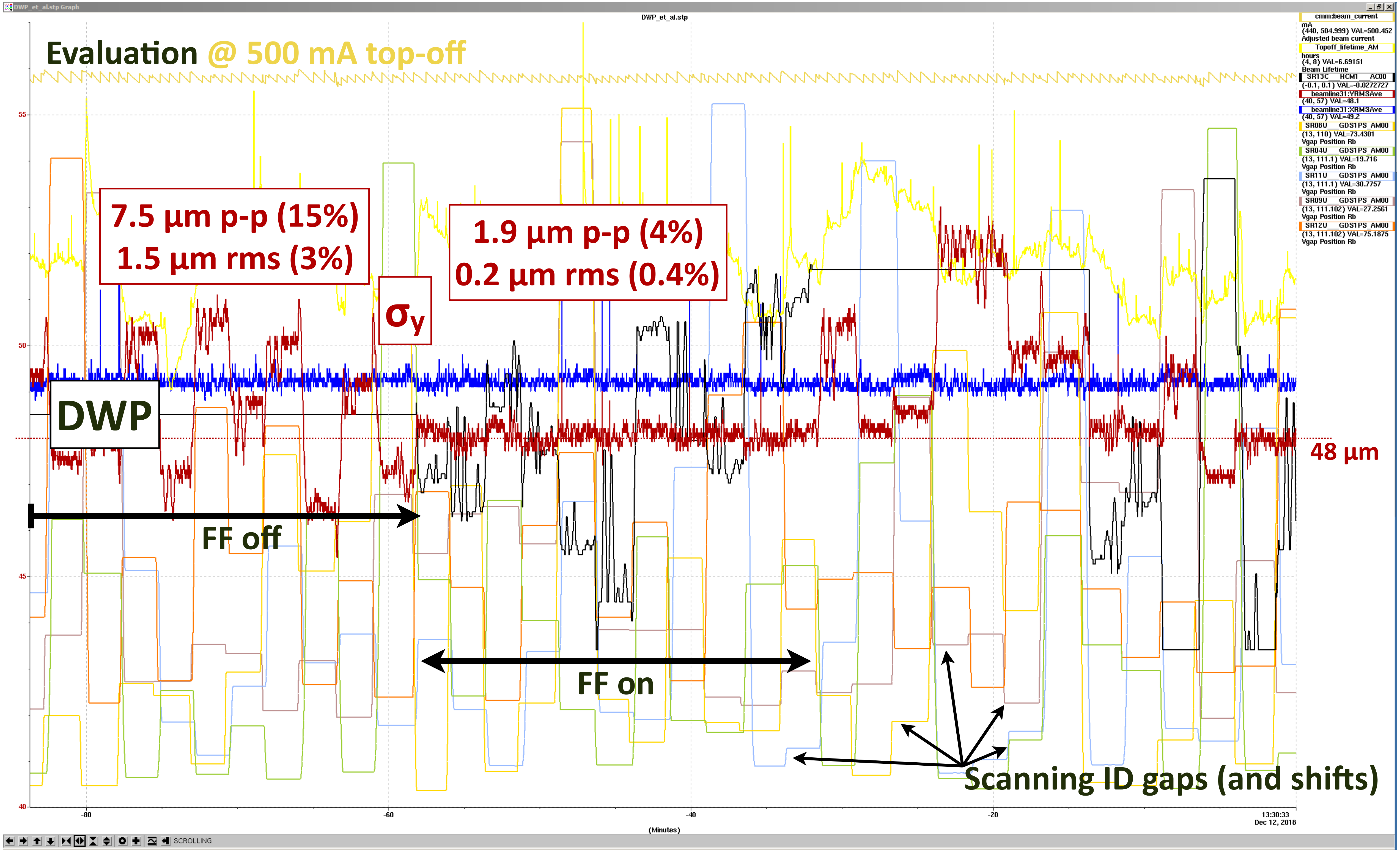
- Instead of dealing with FB issues, scan DWP while scanning ID configurations ➔ acquire data at 10 Hz ➔ input for training of NN (DL)
- Result of DL is prediction for DWP required to keep beam size constant for arbitrary ID configurations ➔ run as NN-based ID FF



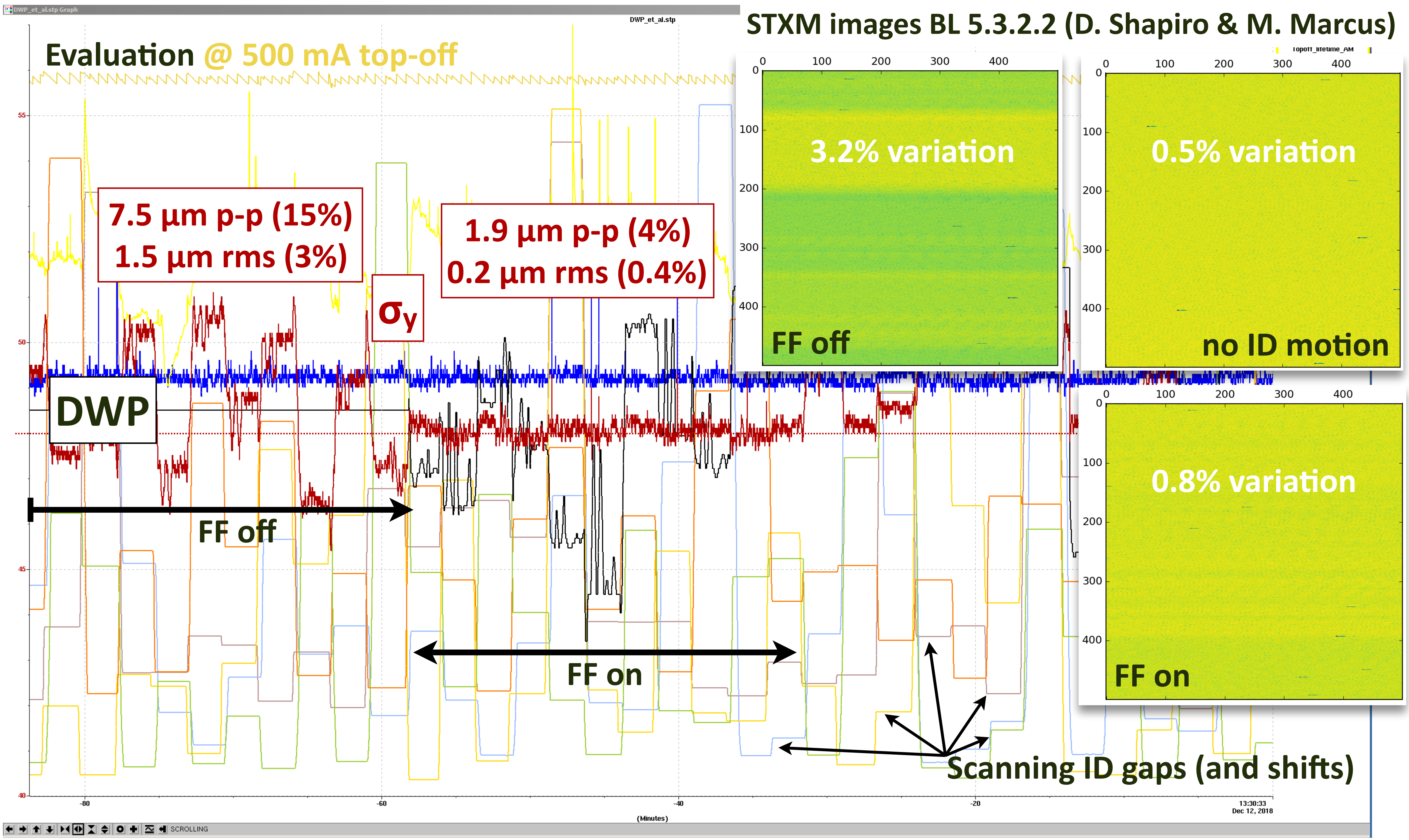
A Better Idea: NN-based ID Feed-Forward (cont.)



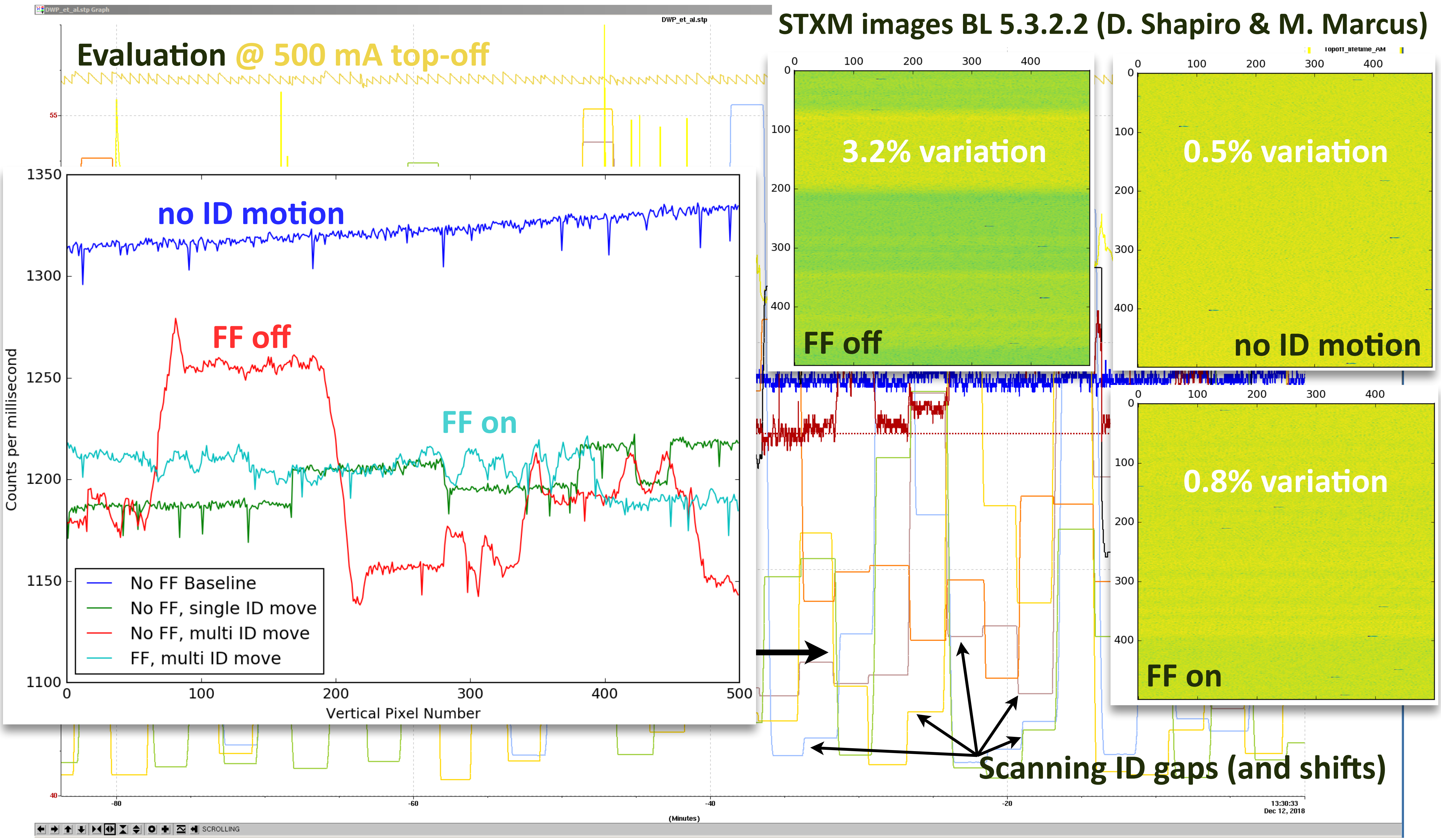
A Better Idea: NN-based ID Feed-Forward (cont.)



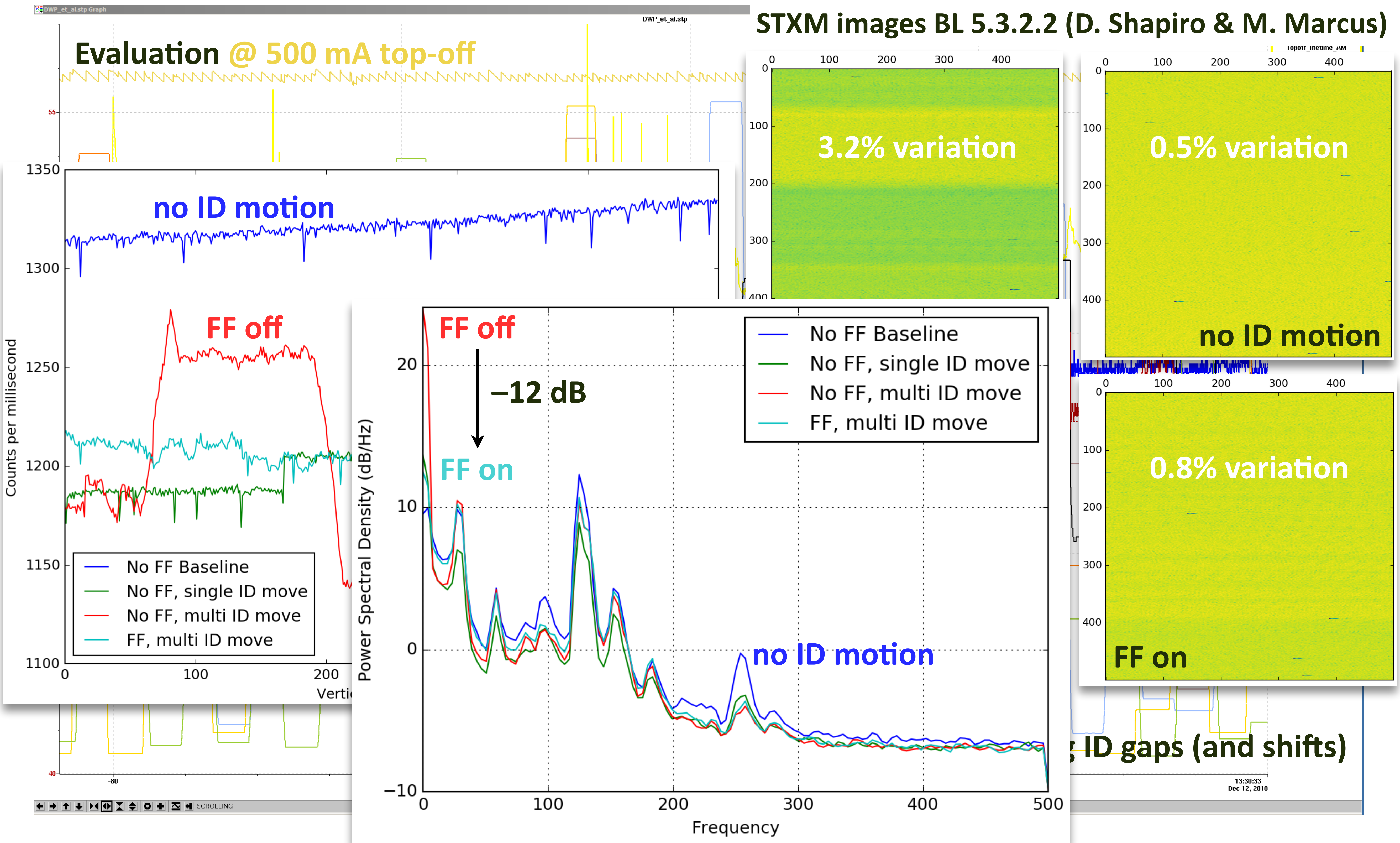
A Better Idea: NN-based ID Feed-Forward (cont.)



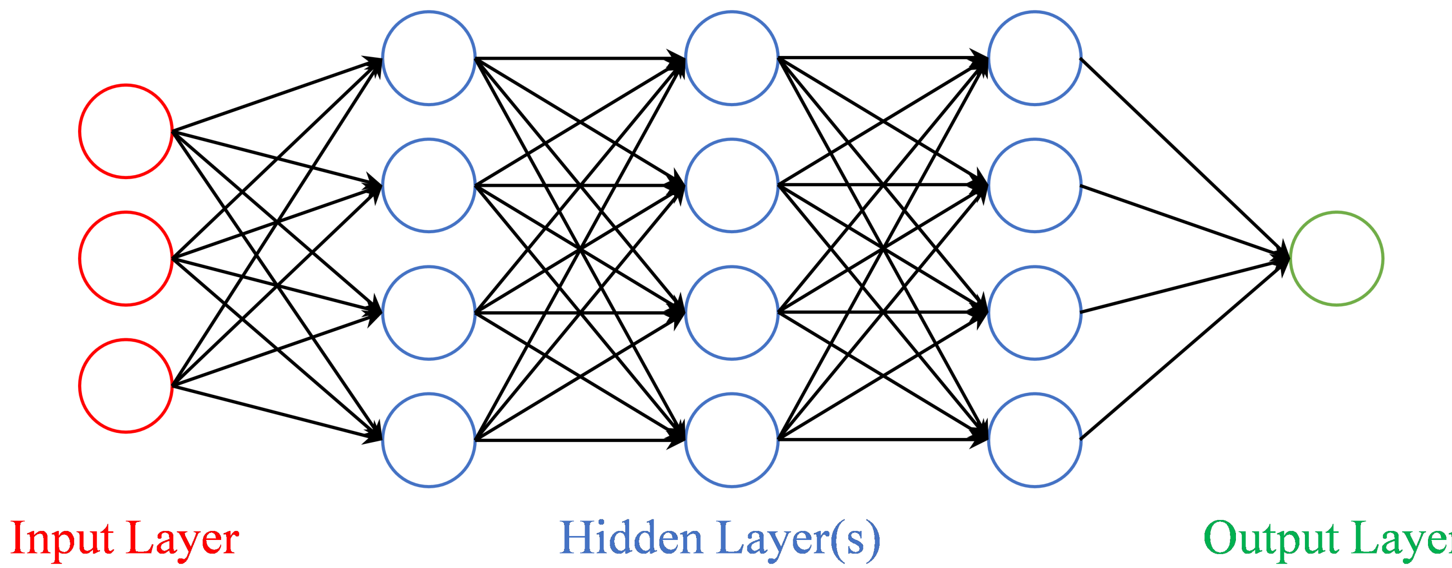
A Better Idea: NN-based ID Feed-Forward (cont.)



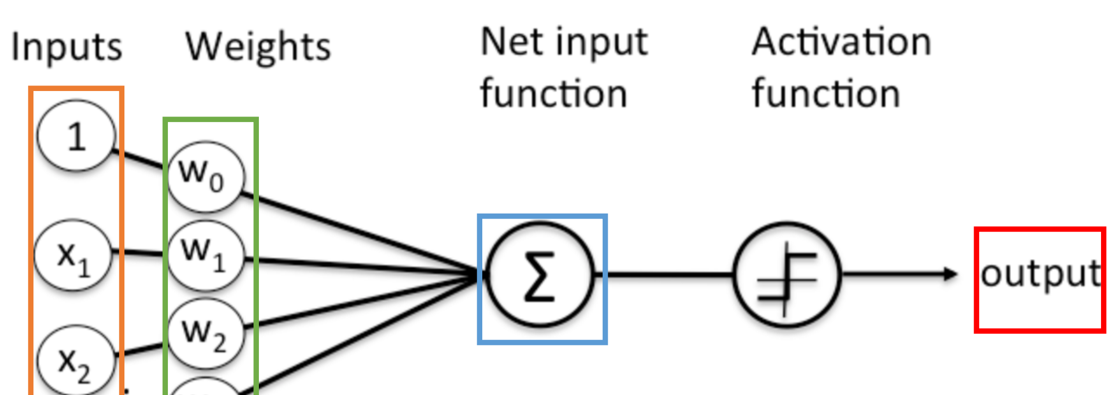
A Better Idea: NN-based ID Feed-Forward (cont.)



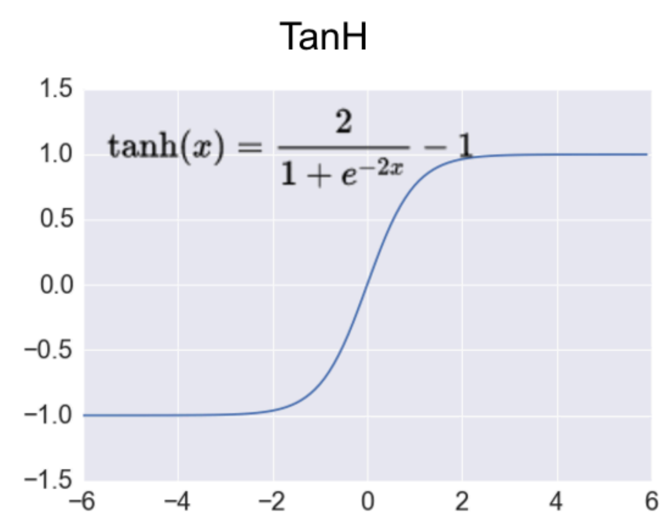
Neural Network — Details

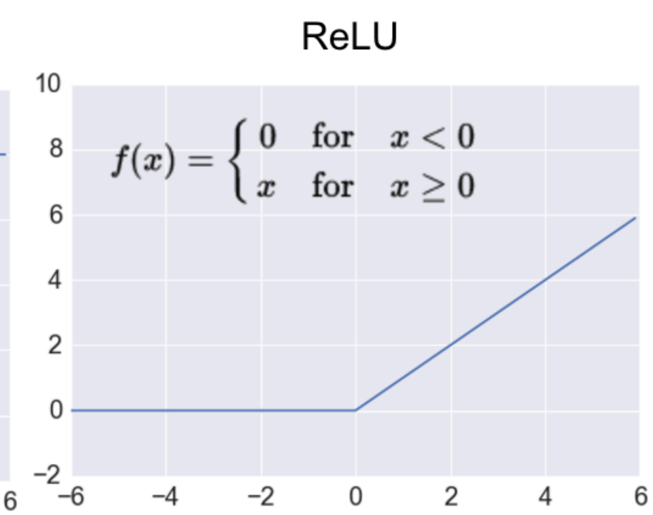


Courtesy: S. Liu

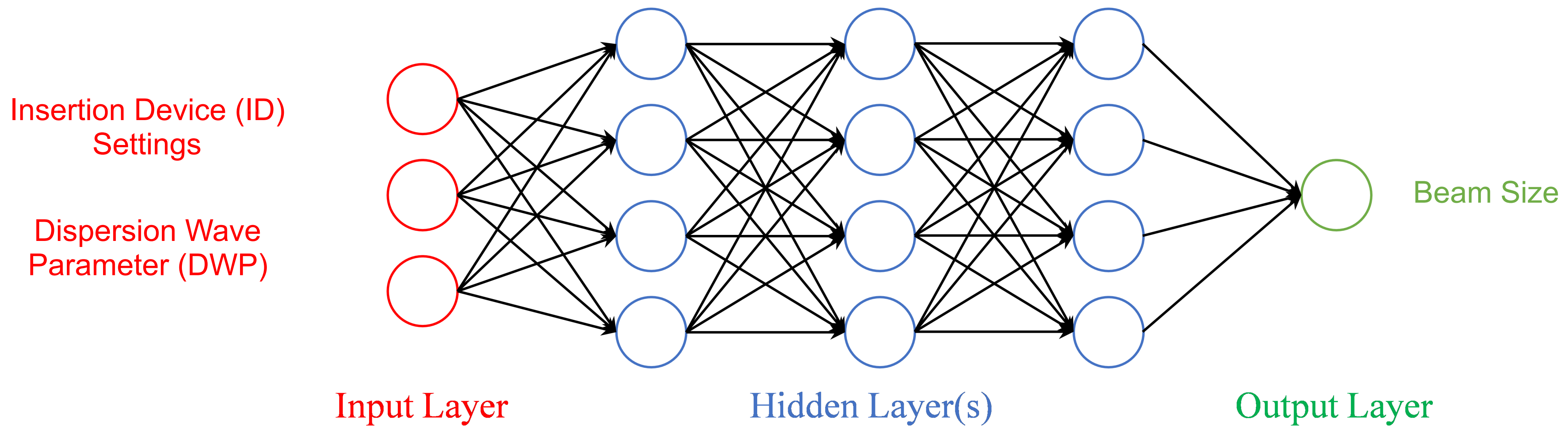


$$x_j^{l+1} = f\Big(\sum_j w_{ij} x_i^l\Big)$$





Neural Network — How is it Trained?



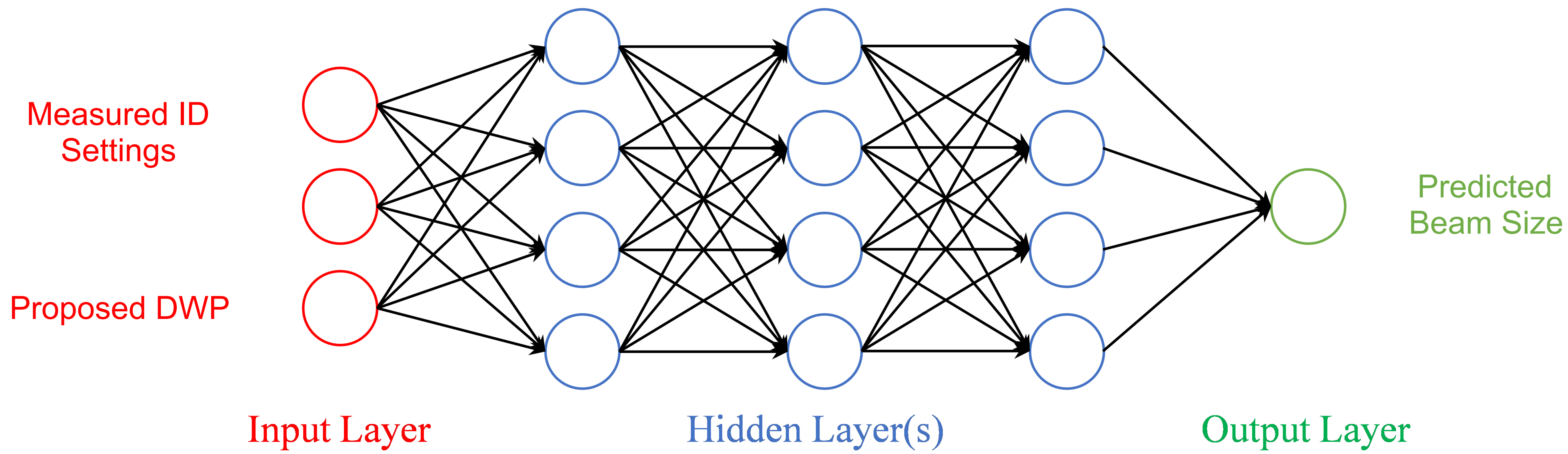
Courtesy: S. Liu

Input Layer: ID settings (22 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: L_2 regularizer with $\lambda = 10^{-4}$
Optimization: Adam Optimizer with learning rate $\alpha = 10^{-3}$

	Raw Data		With Square Features	
Architecture	Training MSE	Evaluation MSE	Training MSE	Evaluation MSE
128-64	0.0265	0.0268	0.0257	0.0260
256-64	0.0243	0.0245	0.0259	0.0262
512-128	0.0243	0.0247	0.0243	0.0247
128-64-32	0.0238	0.0242	0.0243	0.0245
256-128-64	0.0236	0.0240	0.0240	0.0246
256-128-64-32	0.0245	0.0249	0.0245	0.0248

Neural Network — Used in Feed-Forward



Courtesy: S. Liu

Proposed DWPs
-0.06
….
0
…
0.06

$\xrightarrow[\text{Neural Network}]{\text{Measured ID Settings}}$

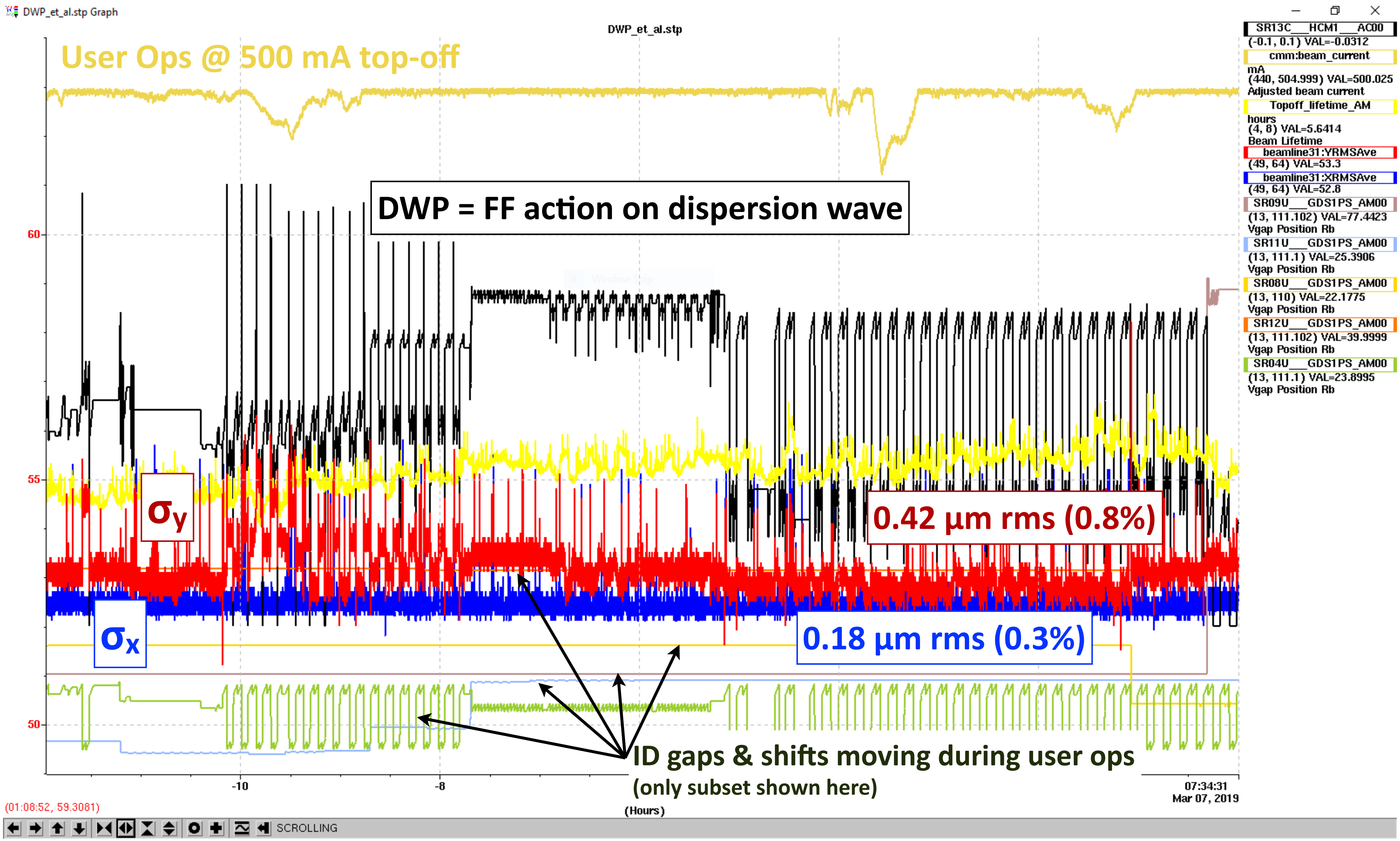
Predicted Beam Sizes
50.3
….
52.1
…
54.0

$\xrightarrow[\text{Beam Size}]{\text{Compare with Target}}$ Choose proper DWP

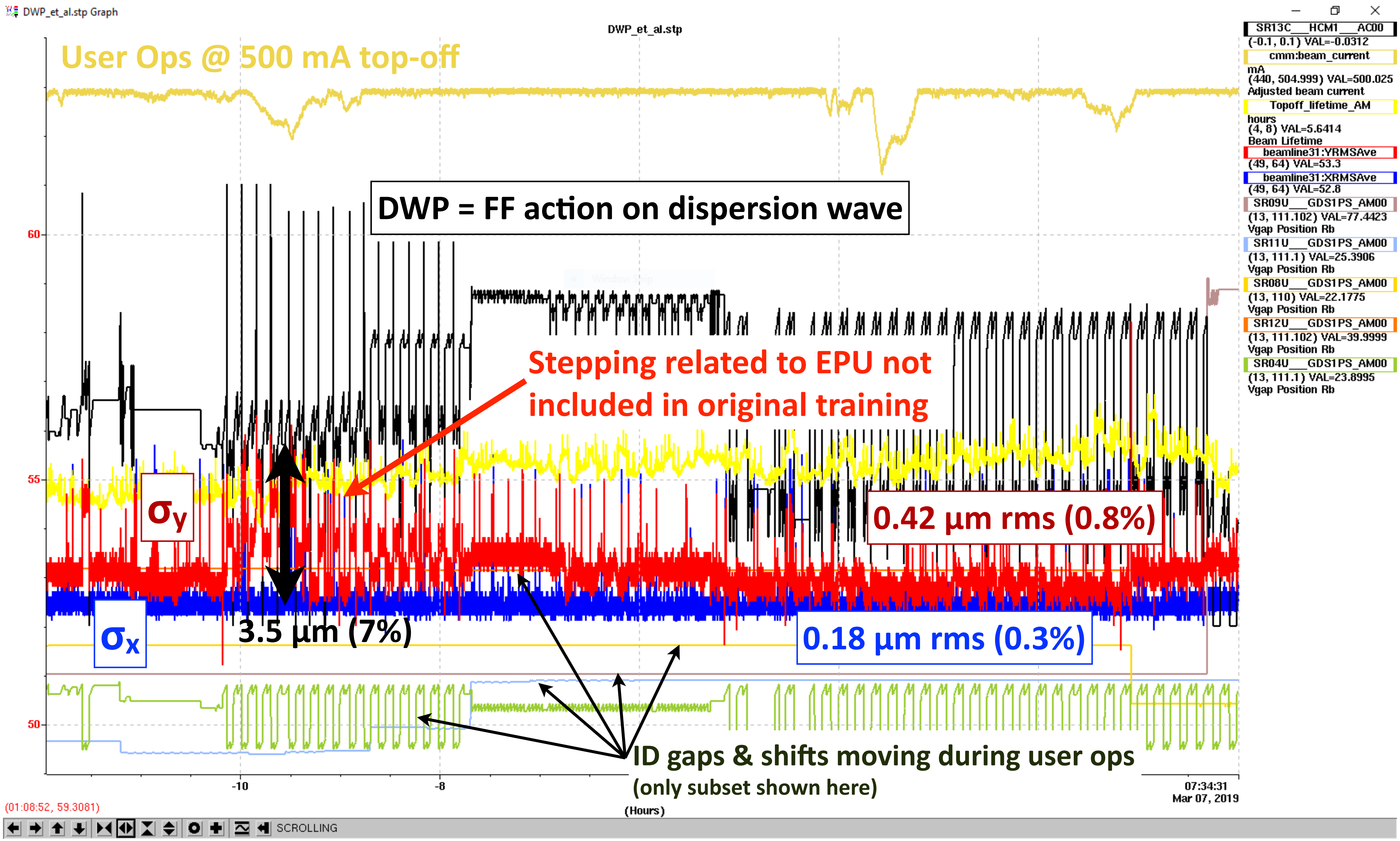
First Experiments During User Ops

- With the NN-based ID FF working well during machine shifts (using subset of IDs believed to have most effect) → move to next phase
- Use machine shift to acquire training data → train NN → put into FF operation during user ops
- Presents opportunity for *"Online Retraining"* exploiting huge amount of data acquired during user ops (and based on ID configuration space actually occupied by ALS users)

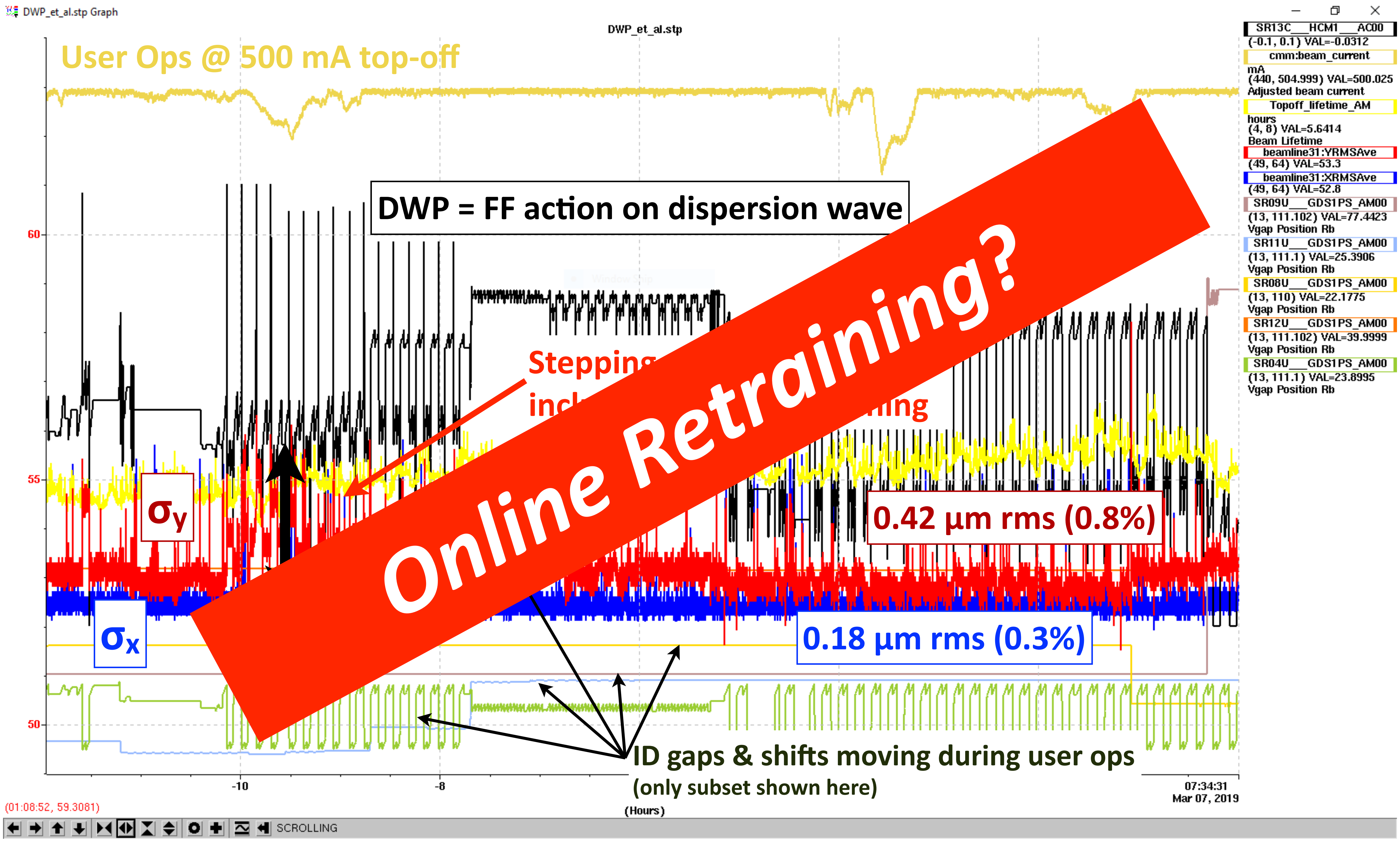
First Experiments During User Ops



First Experiments During User Ops

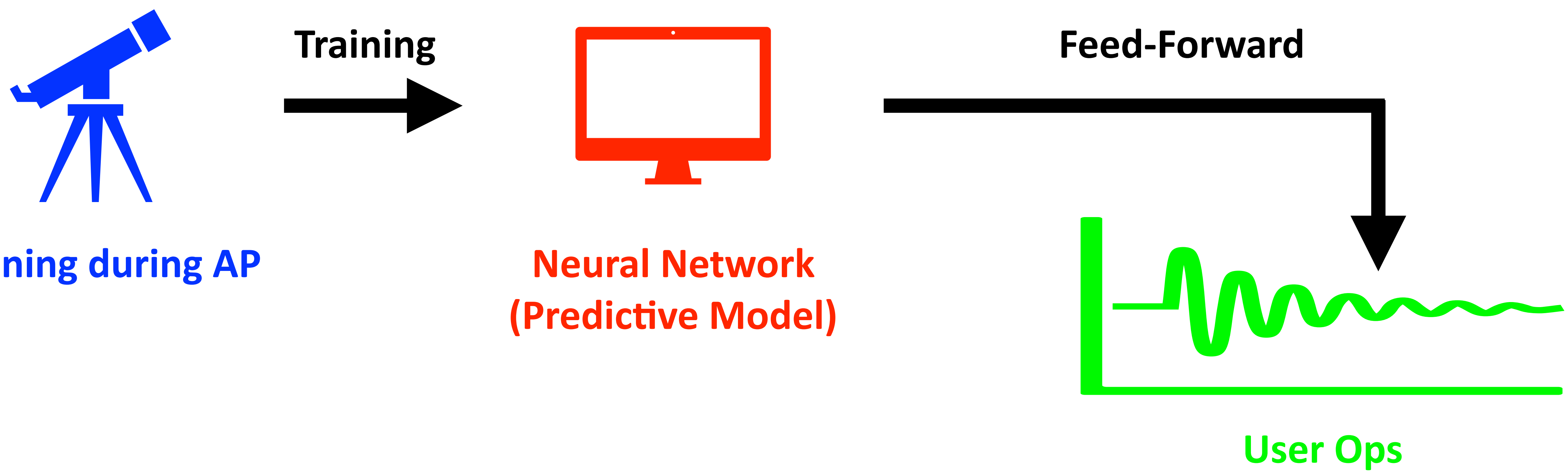


First Experiments During User Ops



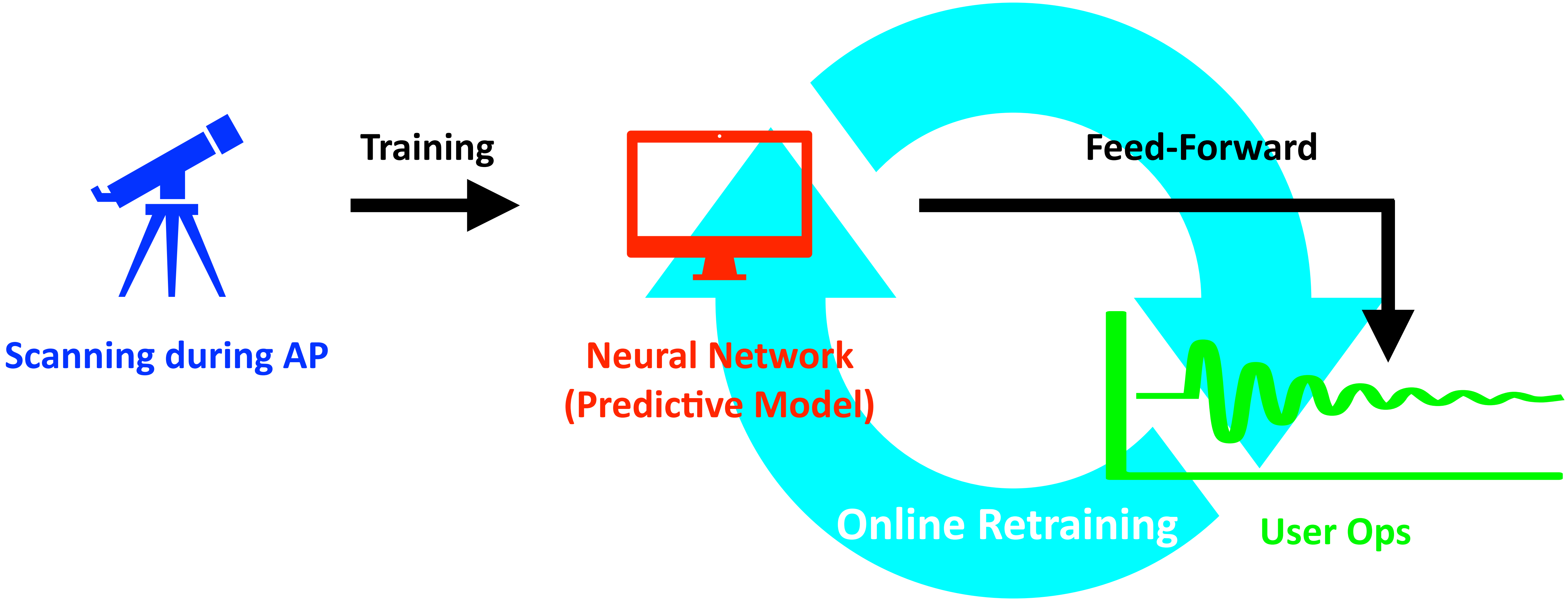
Ideas for Next Improvements

"Conventional" Machine Learning



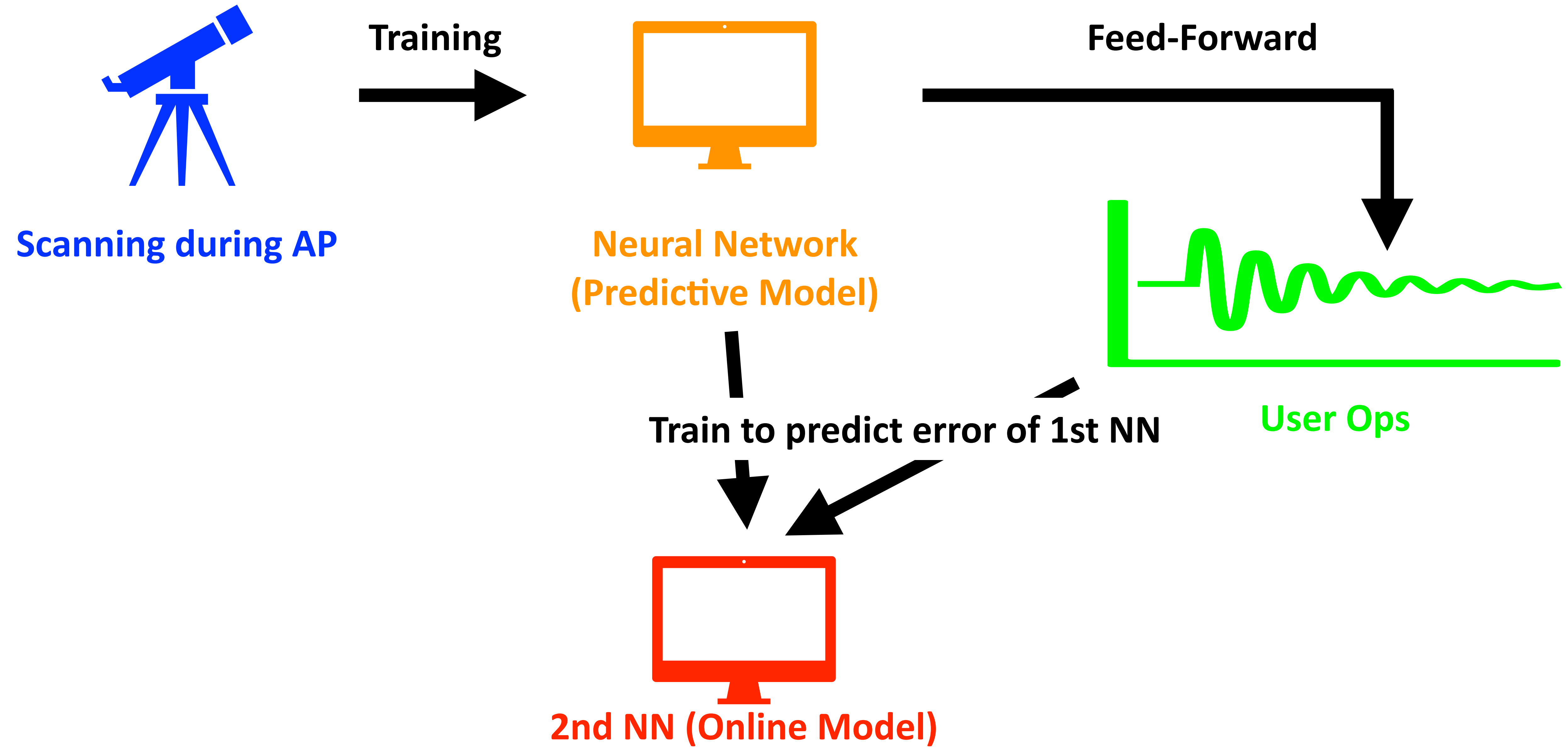
Ideas for Next Improvements (cont.)

Online Retraining (Simple Approach)



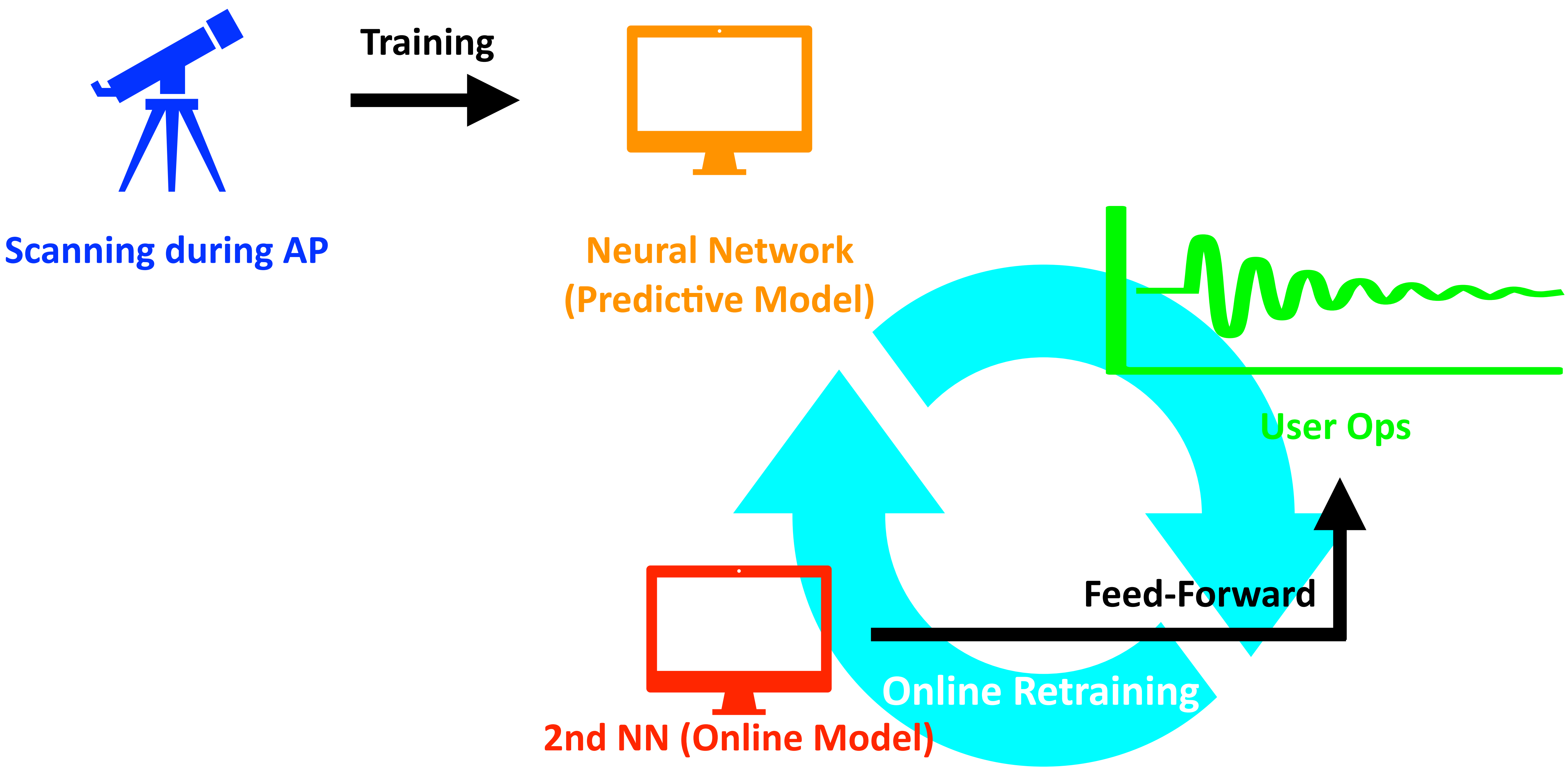
Ideas for Next Improvements (cont.)

Online Retraining (Alternate Approach), Step 1



Ideas for Next Improvements (cont.)

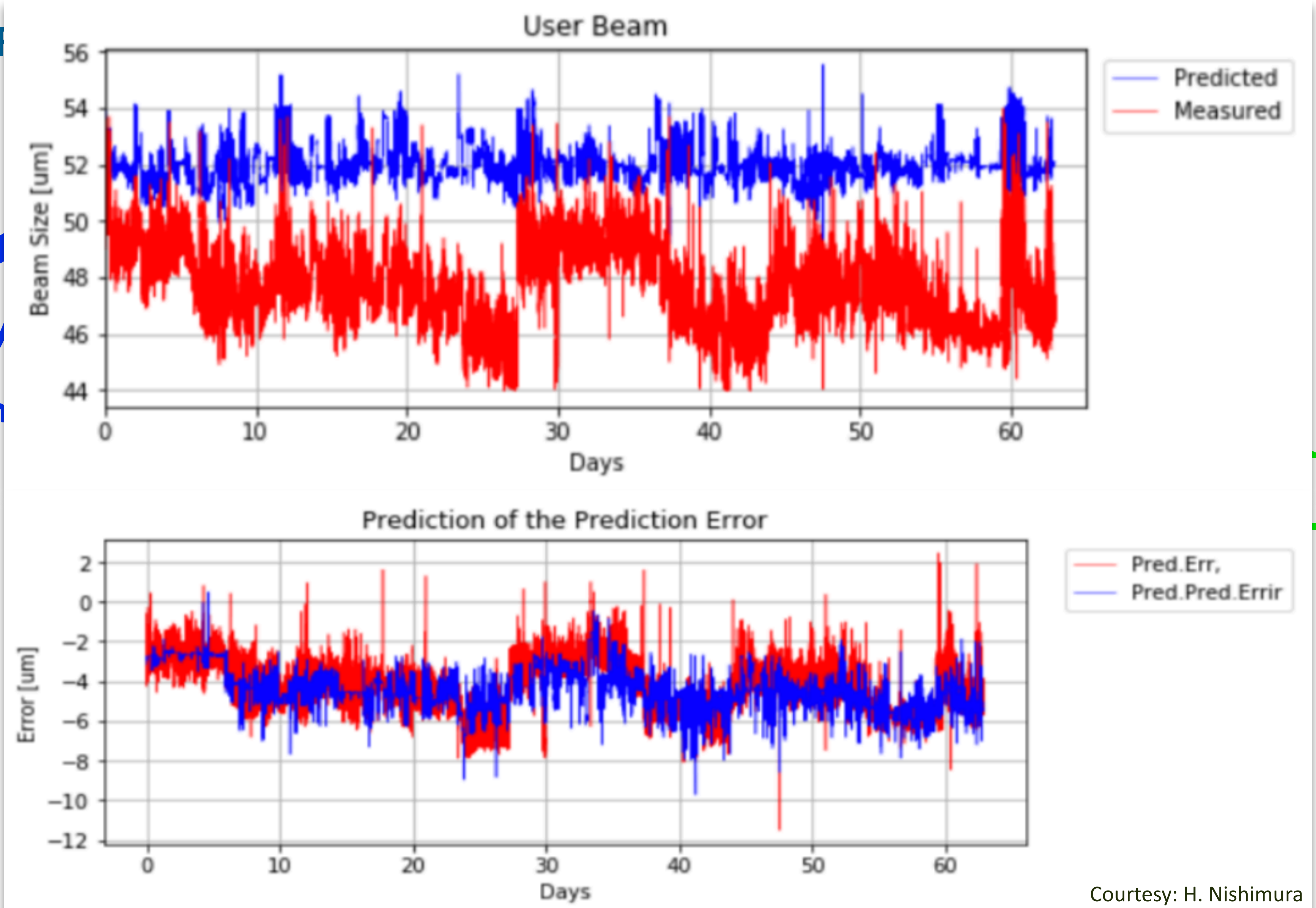
Online Retraining (Alternate Approach), Step 2



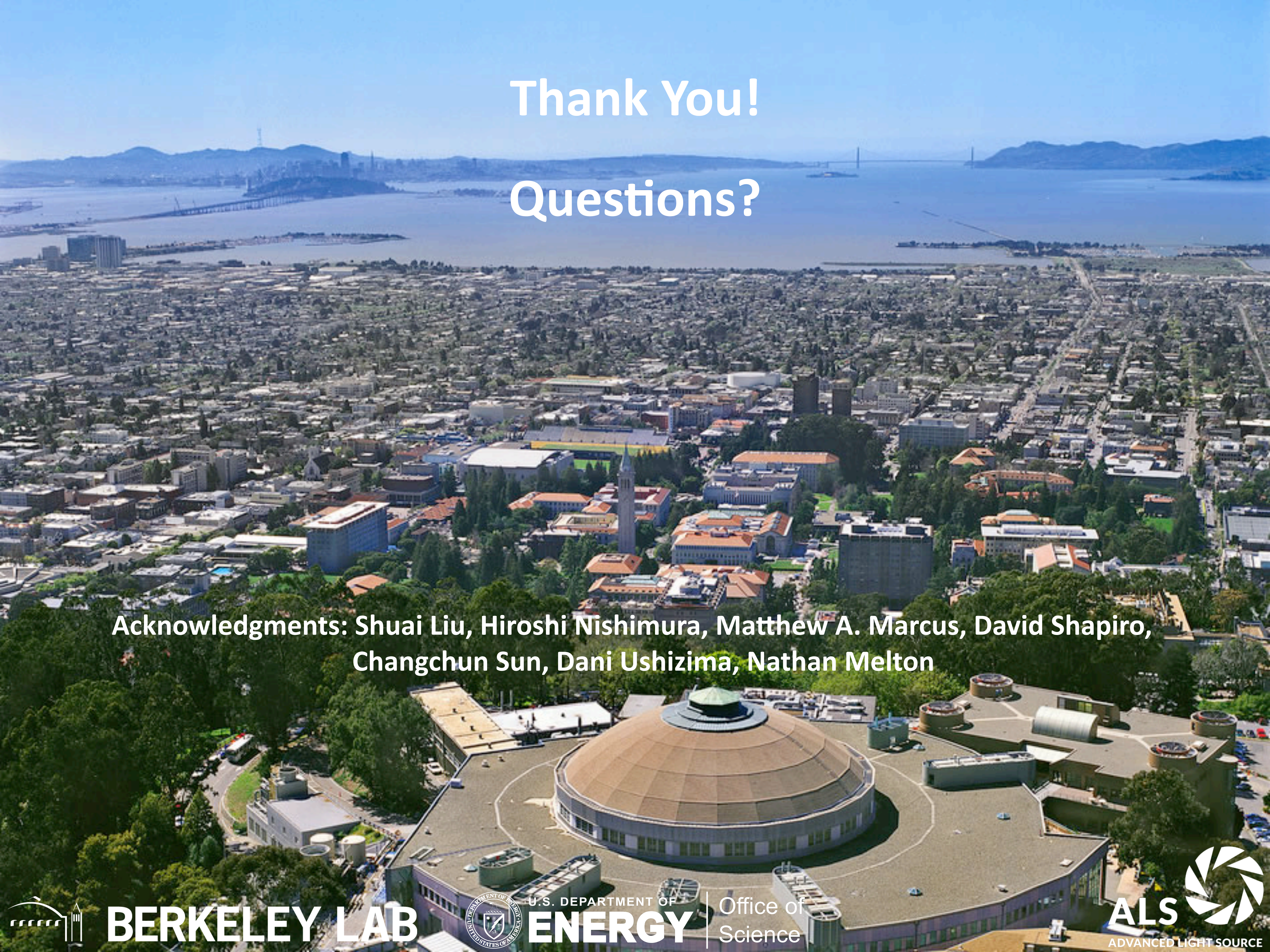
Not Tested Online yet, but Simulations Encouraging

Online I

Scannin



Courtesy: H. Nishimura



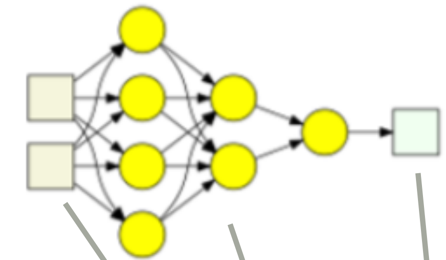
Thank You!

Questions?

Acknowledgments: Shuai Liu, Hiroshi Nishimura, Matthew A. Marcus, David Shapiro, Changchun Sun, Dani Ushizima, Nathan Melton

Backup Slides

Is ML any better than just fitting polynomials?

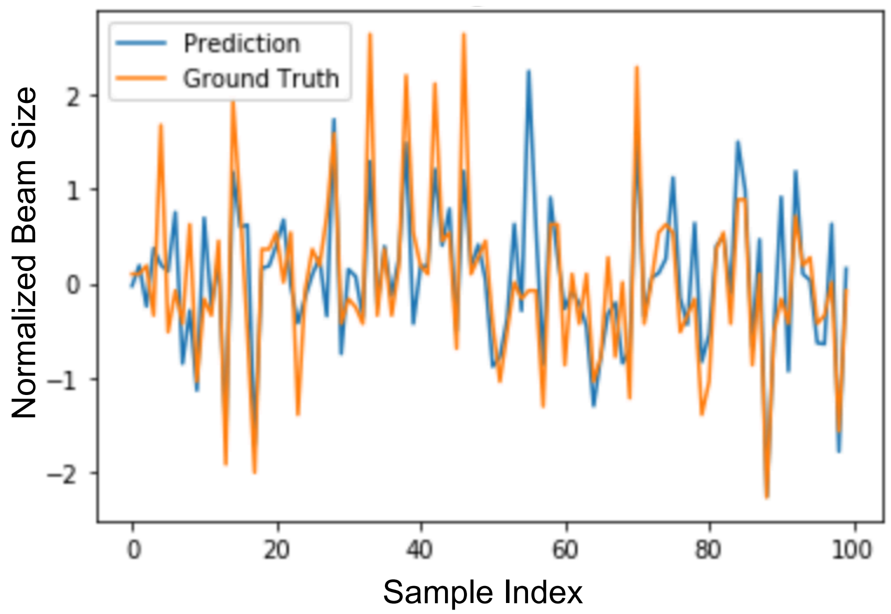


- Machine Learning (as detailed above)
  - training data RMSE 0.023
  - validation data RMSE 0.023

$$\mathrm{MSE} = \min_{w \in \mathcal{R}^m} \left[\sum_{i=1}^{n} (f(x_i, w) - y_i)^2 \right]$$

- 6th-order polynomial fit (applied to same data set)
  - training data RMSE 0.353
  - validation data RMSE 0.354

$$\mathrm{MSE} = \min_{w \in \mathcal{R}^{ml}} \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} \sum_{k=1}^{l} w_{jk} x_j^k + b)^2 \right]$$



MSE: 0.353

Courtesy: S. Liu

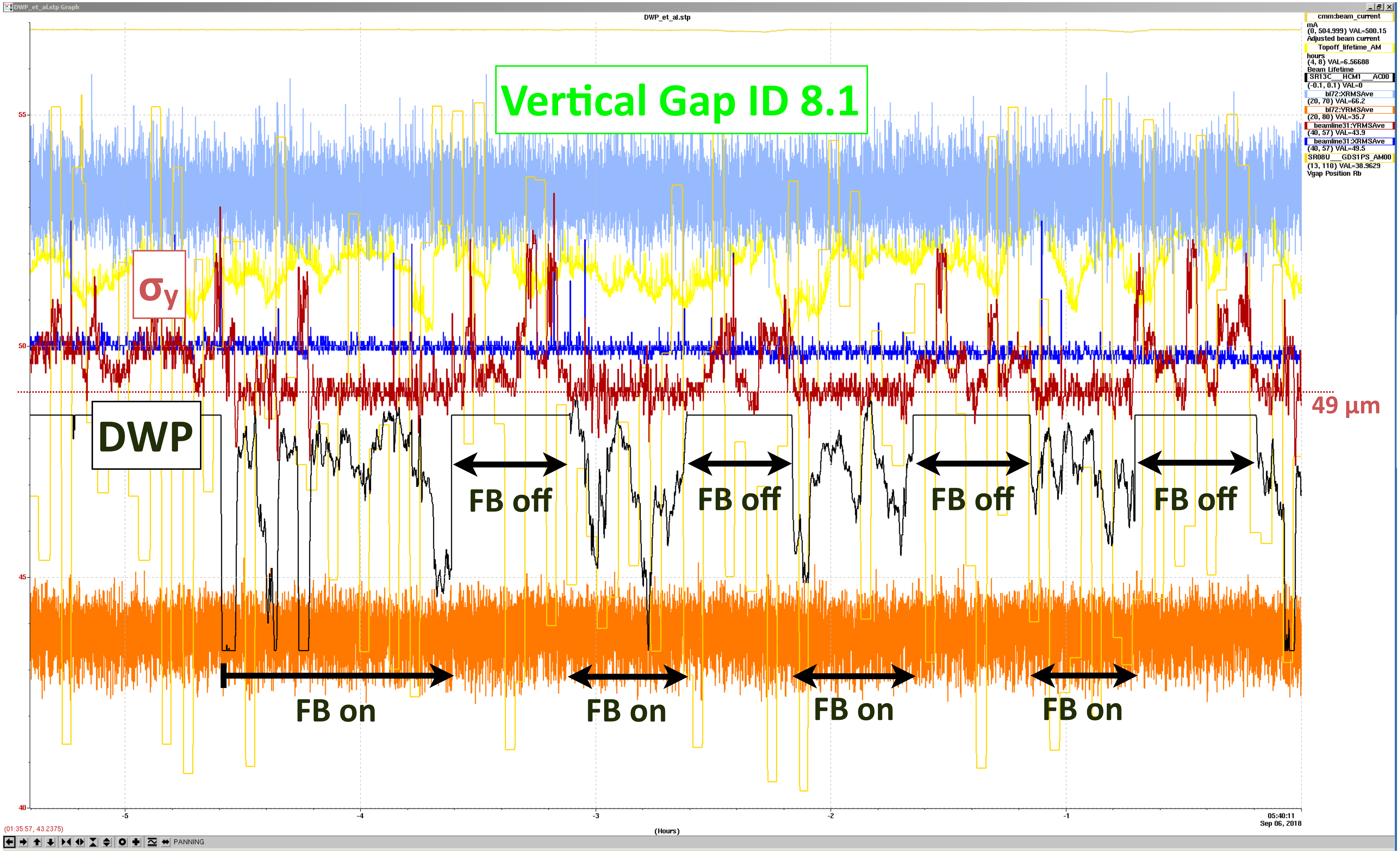
Polynomial Order	Training MSE	Evaluation MSE
1	0.476	0.478
2	0.428	0.429
3	0.381	0.383
4	0.367	0.368
5	0.357	0.359
6	0.353	0.354
Deep Learning	0.0230	0.0232

15-fold improvement

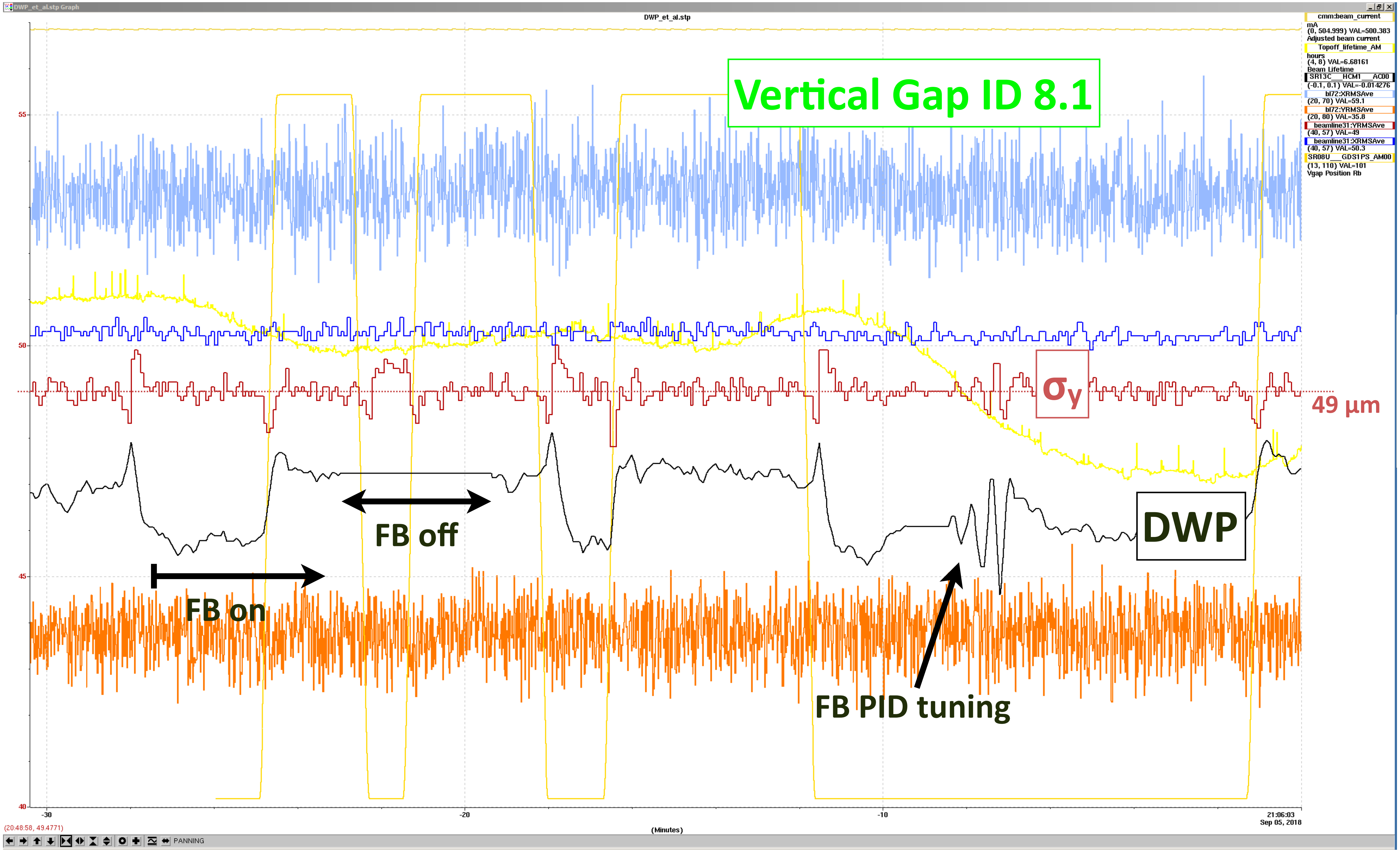
Which BLs are sensitive to beam size fluctuations?

- H beam size is highly stable due to beam physics in 3GLSs (flat machines, well corrected, low coupling → flat beam)
- V beam size can fluctuate significantly → BLs suffer from this if they
  - have entrance slits (apertures transform shape/size changes into intensity changes)
  - disperse in the H plane (monochromator)
  - rely on intensity measurement (I_0 difficult to measure properly)
  - use short acquisition time (→ no averaging), eg.
    - differential measurements (do not want to discard too many scans → acquisition time needs to be short compared to fluctuations)
    - raster scanning (STXM) & dynamics (XPCS) → lots of this @ ALS & ALS-U
    - want to operate at shot noise limit (3GLSs often heavily oversubscribed)
- When feature observed, want certainty it's sample and not source

Example of Initial Feedback Approach

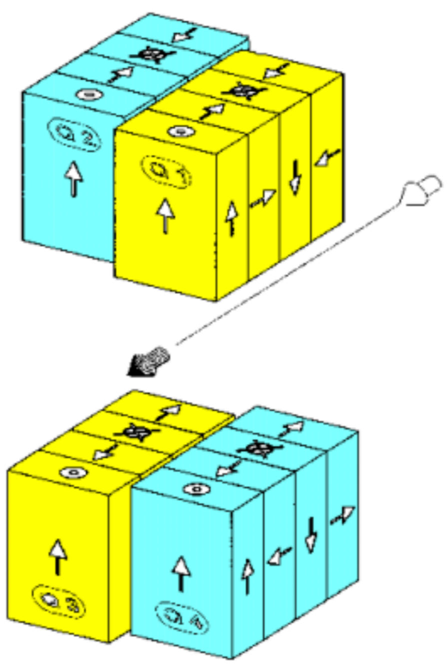


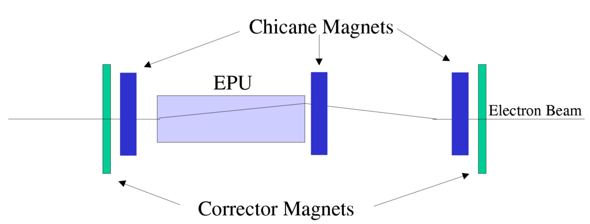
Example of Initial Feedback Approach (cont.)

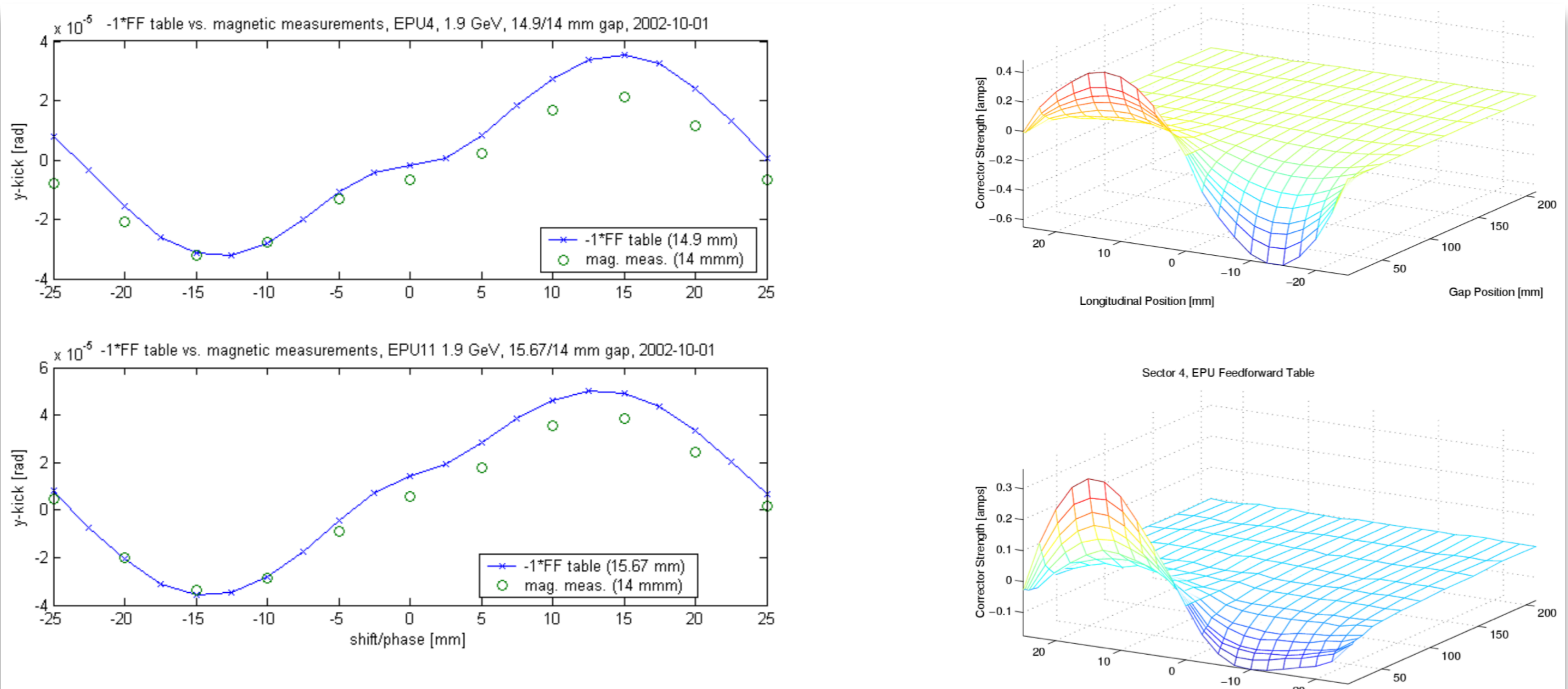


Errors Caused by IDs & Required Corrections

- Orbit distortions
  - caused by on-axis variation of field integrals (with gap or EPU phase)
    - corrected by shims (magic fingers) & local orbit correctors (FF, 200 Hz)
    - corrected by ring corrector magnets (FB, ≈1 Hz SOFB & 1.1 kHz FOFB)

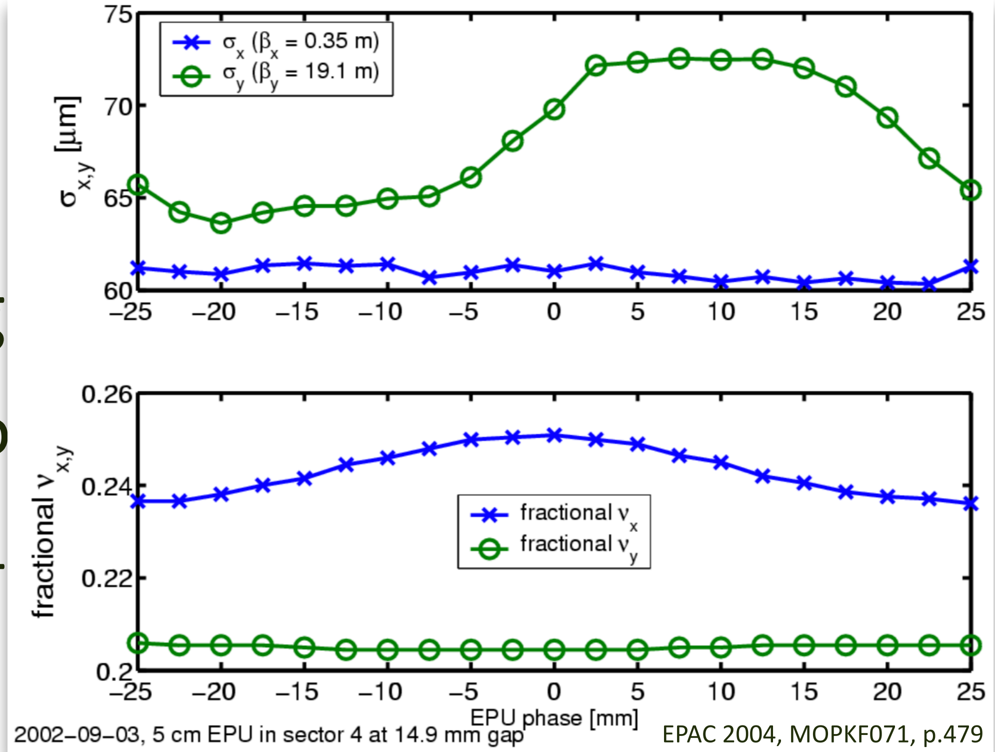






EPAC 2004, MOPKF071, p.479

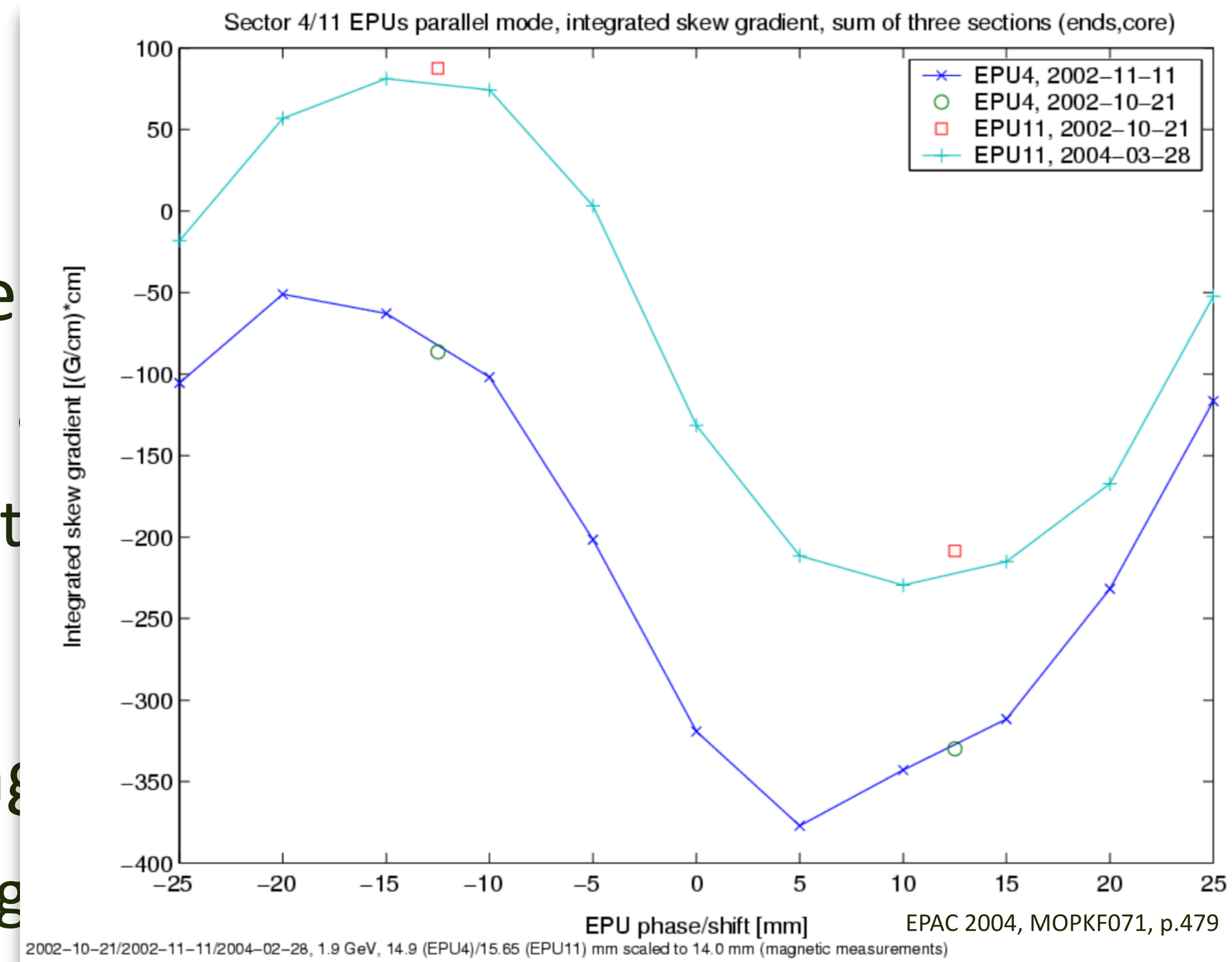
Errors Caused by IDs & Required Corrections



- Orbit distortions
  - caused by on-axis variation of field integ
    - corrected by shims (magic fingers) & local o
    - corrected by ring corrector magnets (FB, ≈1
- Beam size (primarily vertical)
  - caused by variation of ID focusing terms (with gap or EPU phase)
    - corrected by local quad trims and global quad adjustment (FF & FB)

Errors Caused by IDs & Required Corrections

- Orbit distortions
  - caused by on-axis variation of fie
    - corrected by shims (magic fingers)
    - corrected by ring corrector magnet
- Beam size (primarily vertical)
  - caused by variation of ID focusing
    - corrected by local quad trims and g
  - caused by variation of ID-induced coupling (usually with EPU phase)
    - corrected by local skew quad coils (FF)



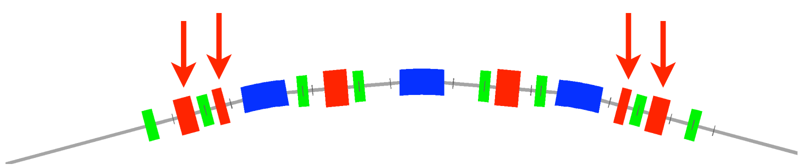
EPAC 2004, MOPKF071, p.479

Errors Caused by IDs & Required Corrections

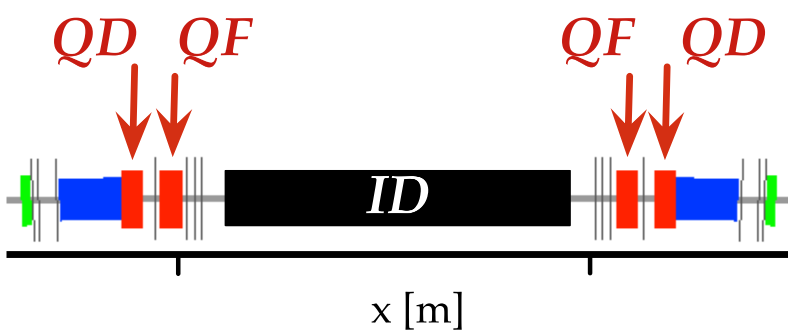
- Orbit distortions
  - caused by on-axis variation of field integrals (with gap or EPU phase)
    - corrected by shims (magic fingers) & local orbit correctors (FF, 200 Hz)
    - corrected by ring corrector magnets (FB, ≈1 Hz SOFB & 1.1 kHz FOFB)
- Beam size (primarily vertical)
  - caused by variation of ID focusing terms (with gap or EPU phase)
    - corrected by local quad trims and global quad adjustment (FF & FB)
  - caused by variation of ID-induced coupling (usually with EPU phase)
    - corrected by local skew quad coils (FF)
- Reduced injection efficiency & lifetime (nonlinear beam dynamics)
  - caused by higher-order ID effects (eg. field roll-off) → sets requirements for ID design and machine optics

ID Focusing Corrections Implemented in ALS ID FF

- Global Corrections
  - tunes (using lattice quads: 24 QF & 24 QD)
  - in addition: **tune FB** using same quads



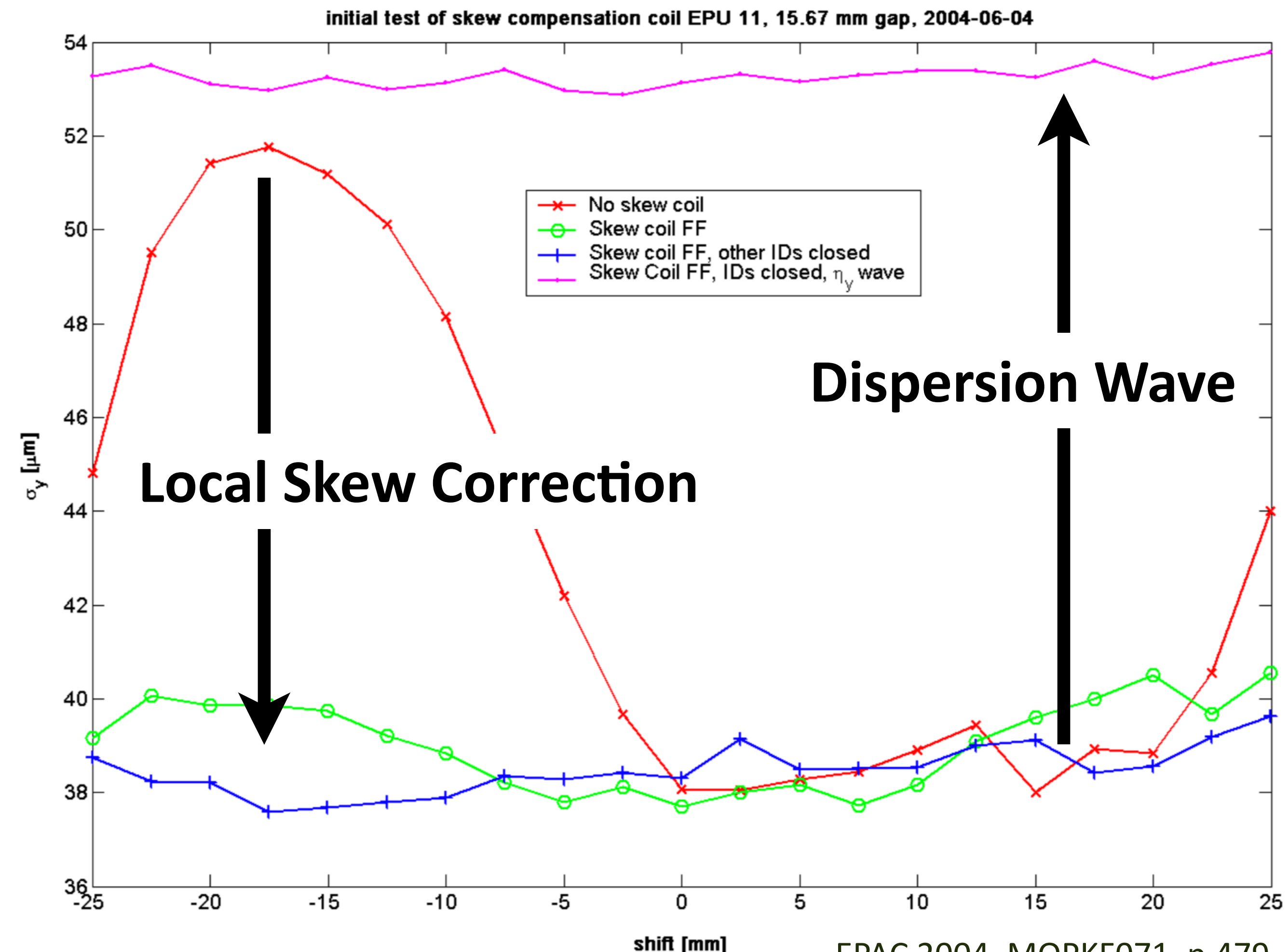
- Local Corrections for all IDs
  - β_y beat (using 2 QF & 2 QD locally)
    ➔ slightly increases $\Delta\nu_y$ ➔ can be removed by global tune correction



- Local Corrections for EPUs only
  - β_x beat (using 2 QF & 2 QD locally)
    ➔ locally also corrects $\Delta\nu_x$ since $\beta_x \approx 21$ m

Vertical Dispersion Wave Determines Effective ε_y

- Vertical source size is determined by
  - optics and coupling (local)
  - vertical emittance (global) consisting of
    - natural contribution (emission of SR is quantum process)
    - imperfections (unavoidable in real machines)
    - systematic η_y contributions (Dispersion Wave)



EPAC 2004, MOPKF071, p.479

For Accelerators Deep Learning is a Paradigm Shift

Traditional Approach

IDs
Skews
$\mathcal{M}$
Input

σ_y
Output

- *Requires physics/model knowledge*
- *Tends to rely on many (sometimes poor) assumptions*

Deep Learning

IDs
Skews
σ_y
Input

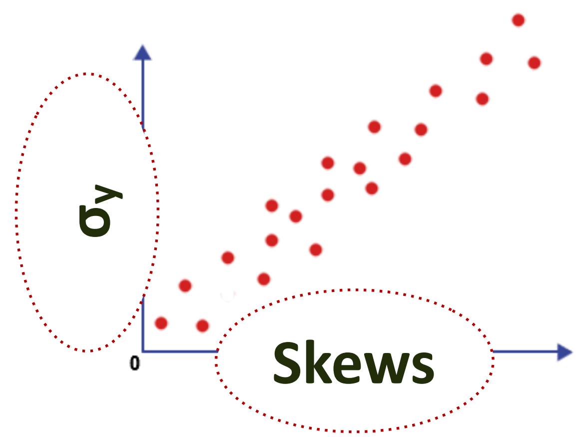
Training

NN
Output

- *Requires only large amounts of data & reproducibility*

Application during ops:

IDs
NN
Prediction



FF
Set Target σ_y
Skews

ALS/SSRL Collaboration Efforts on ML

- Regular collaboration meetings
- Common theme, different emphasis at each lab
  - Each lab has specific applications, which can generalize to applications at other labs
- Collaboration allows exploiting such efforts at both labs
  - Discuss potential new applications and/or solutions to common problems
  - Co-develop software, review and/or test each others' codes
  - Exploit and/or enhance solutions developed at other lab
- Collaboration meetings facilitate onboarding of postdocs, knowledge transfer, and stimulate new ideas
- Collaboration meetings support postdocs in getting involved with developments at other lab