Machine Learning-based Beam Size Stabilization at ALS

Simon C. Leemann
ALS Accelerator Physics, ATAP & ALS Divisions, Lawrence Berkeley National Laboratory
Aug 7, 2020

SLAC AI Seminar

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY | Office of Science

ACCELERATOR TECHNOLOGY & APPLIED PHYSICS DIVISION ATAP

ALS ADVANCED LIGHT SOURCE

The Advanced Light Source at Berkeley Lab

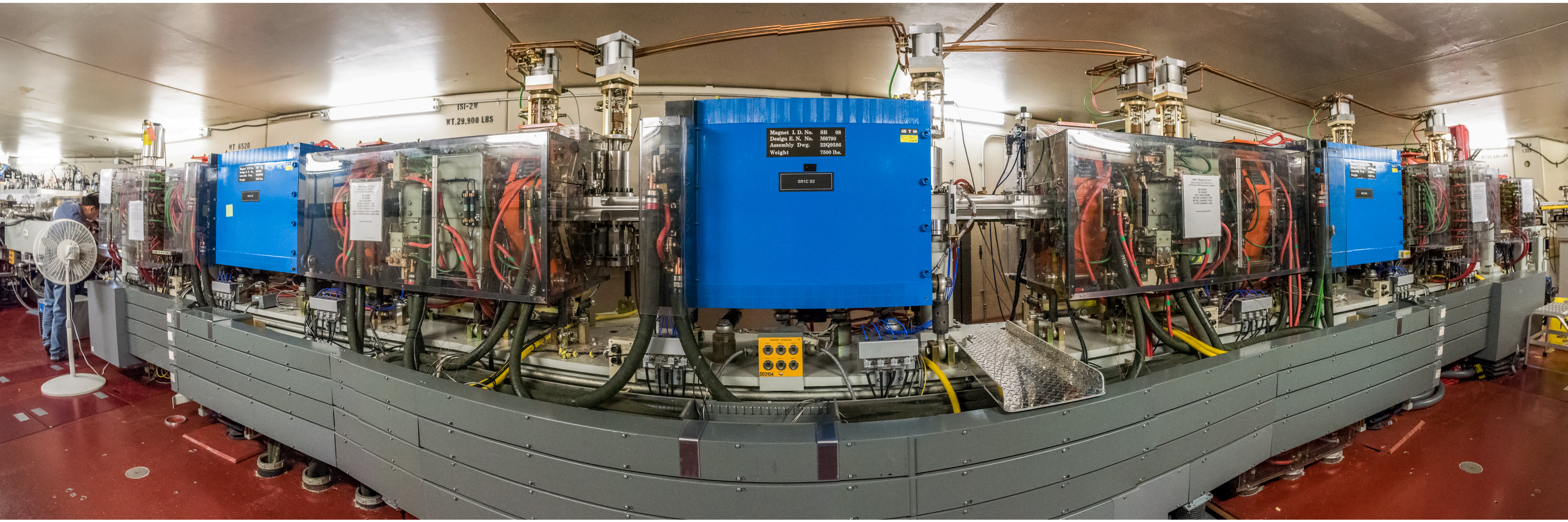
The Advanced Light Source at Berkeley Lab

184" cyclotron yoke (construction 1940)

The Advanced Light Source at Berkeley Lab

1.9 GeV Storage Ring, 196.8 m, 1993

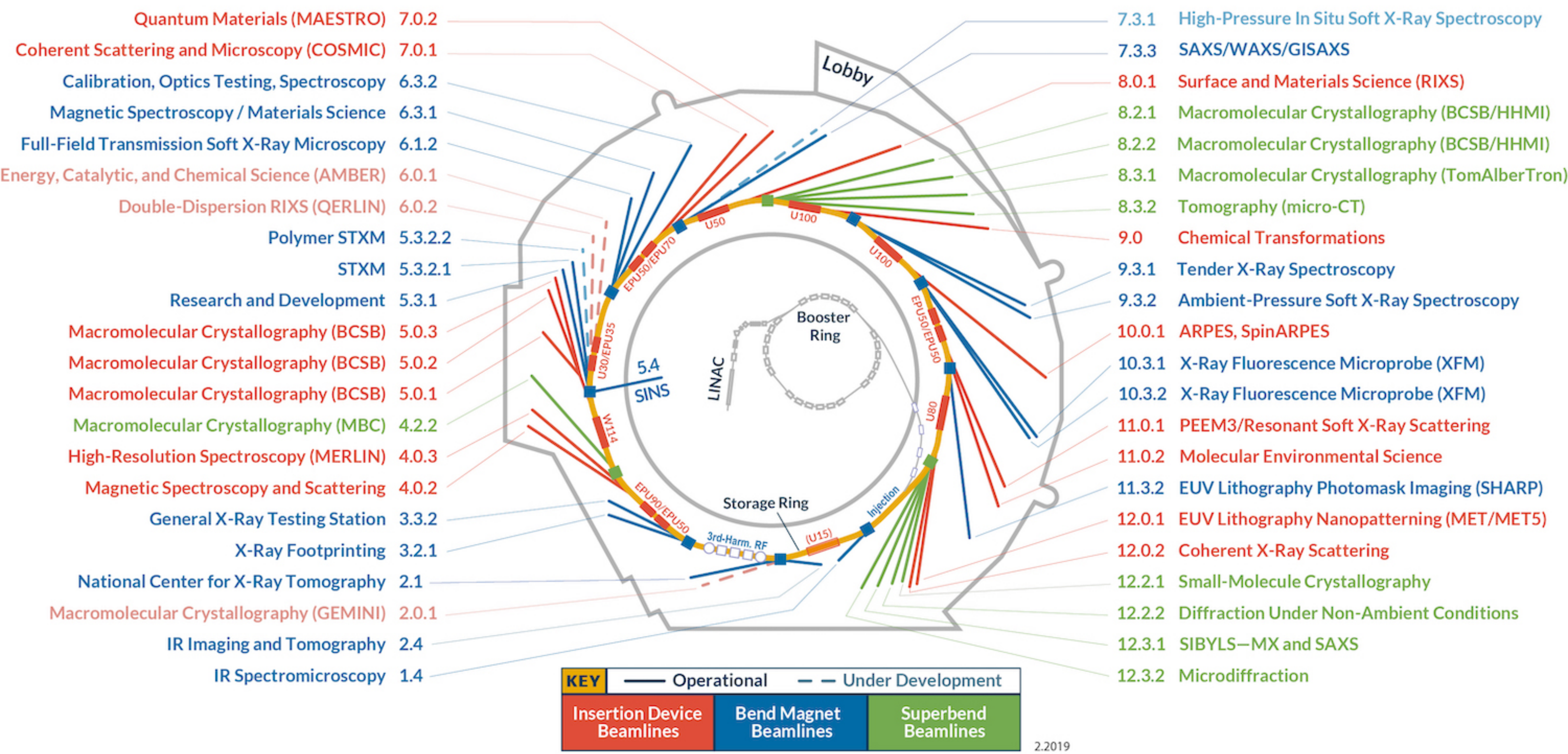
The Advanced Light Source at Berkeley Lab



Triple-bend achromat lattice (12 sectors)

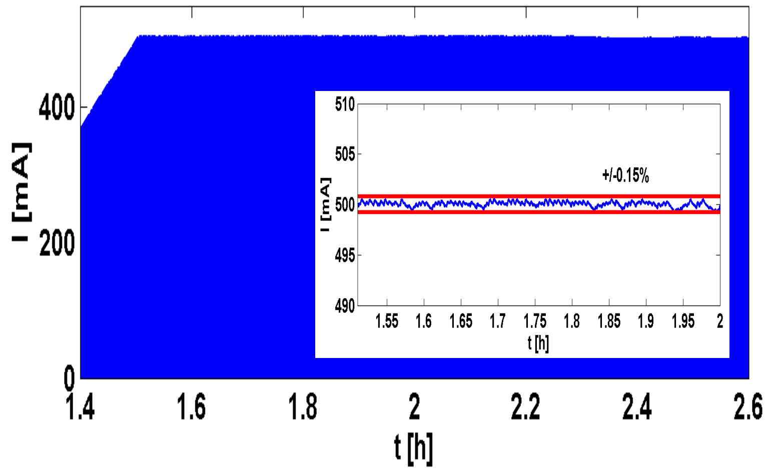
The Advanced Light Source at Berkeley Lab

≈40 beamlines, ≈5000 hrs/y, ≈2000 users/y



Many Successful Efforts to Stabilize Electron Beams

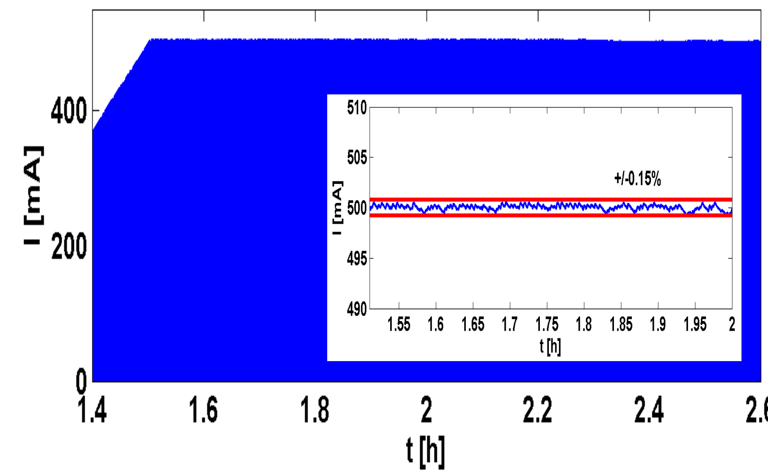
- **Top-off** keeps ALS stored current variation <0.2%



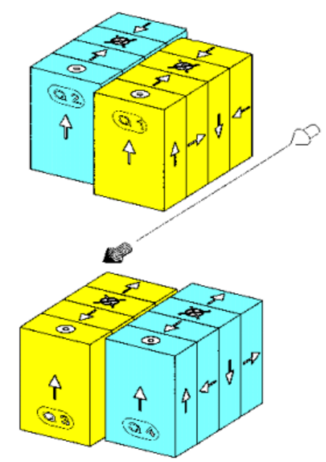
Courtesy: C. Steier, PAC'09

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

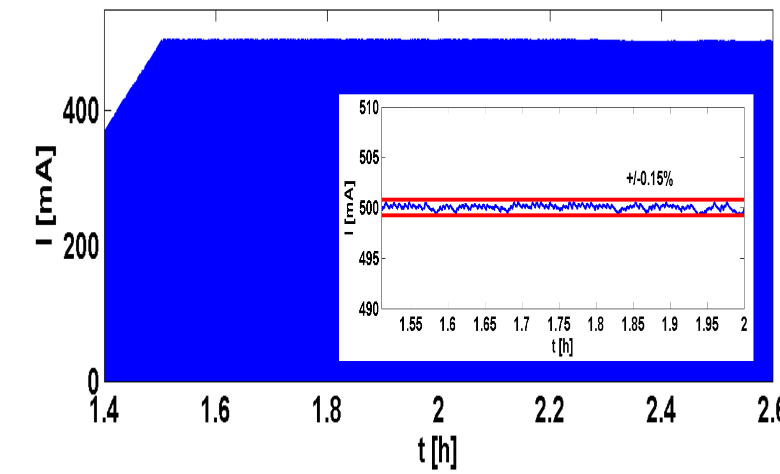


Courtesy: C. Steier, PAC'09

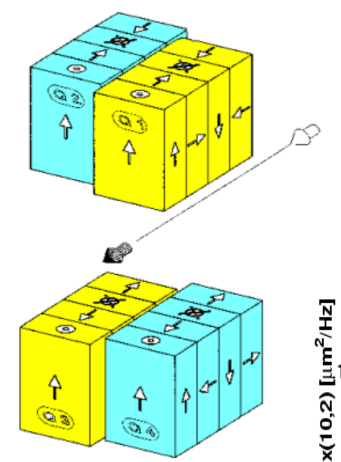


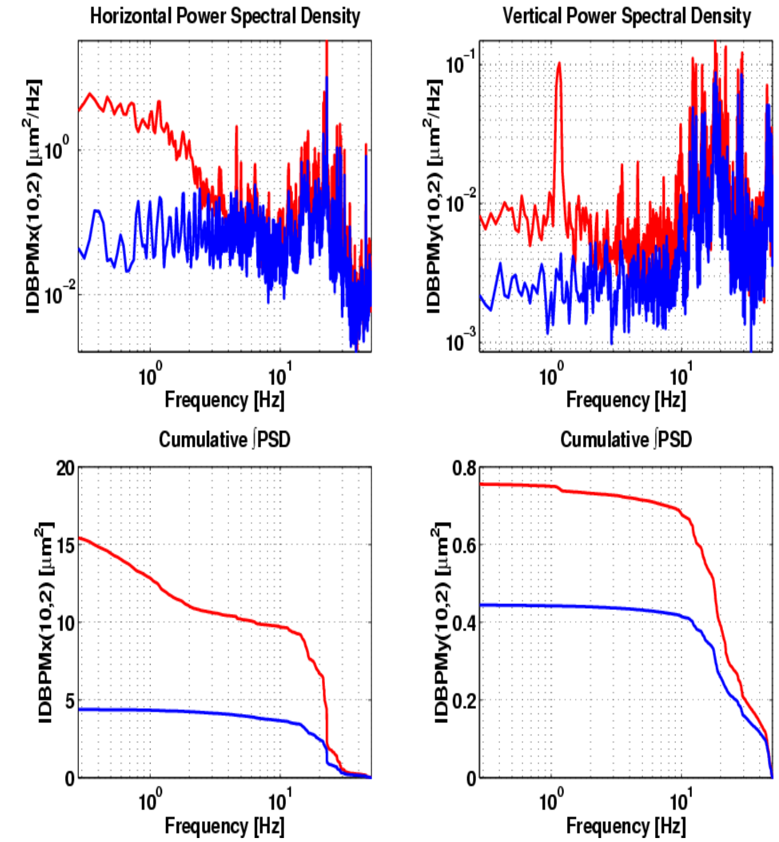
Many Successful Efforts to Stabilize Electron Beams

- **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



Courtesy: C. Steier, PAC'09

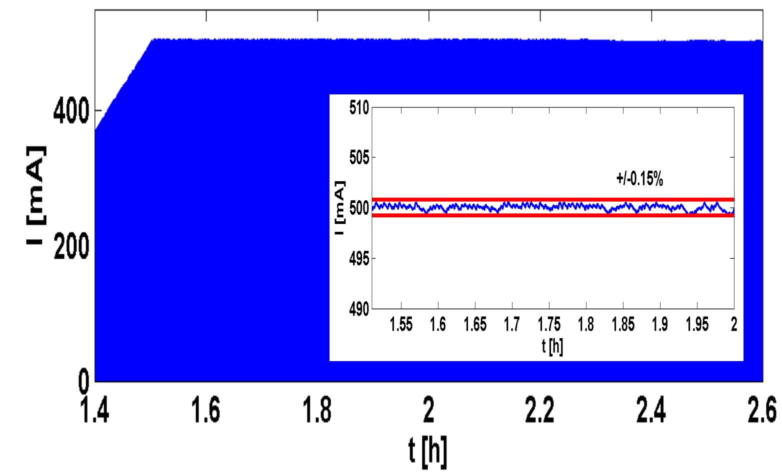




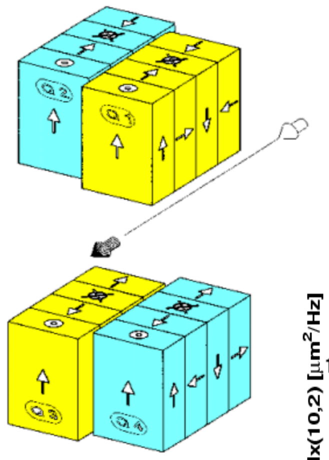
Many Successful Efforts to Stabilize Electron Beams

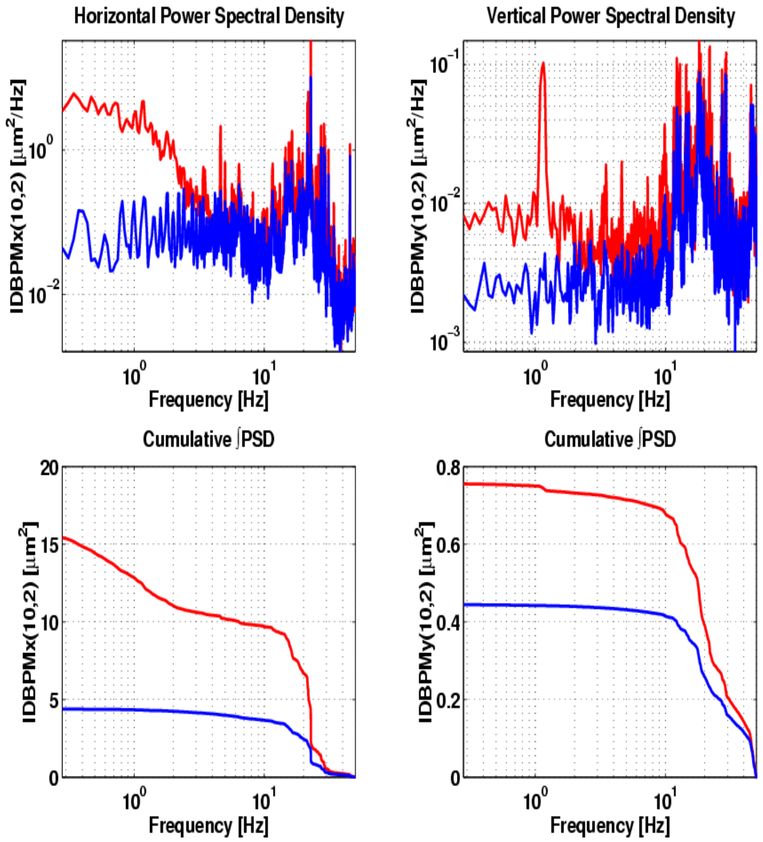
- **Top-off** keeps ALS stored current variation <0.2%
- At low energy, ALS strongly affected by insertion device (ID) imperfections & continuously changing EPU gaps/phases
  - **Orbit feedback** and ID feed-forwards stabilize source positions/angles to **sub-micron** level at many tens of Hz
  - **ID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

Thermal, Ground, Water Table, etc.



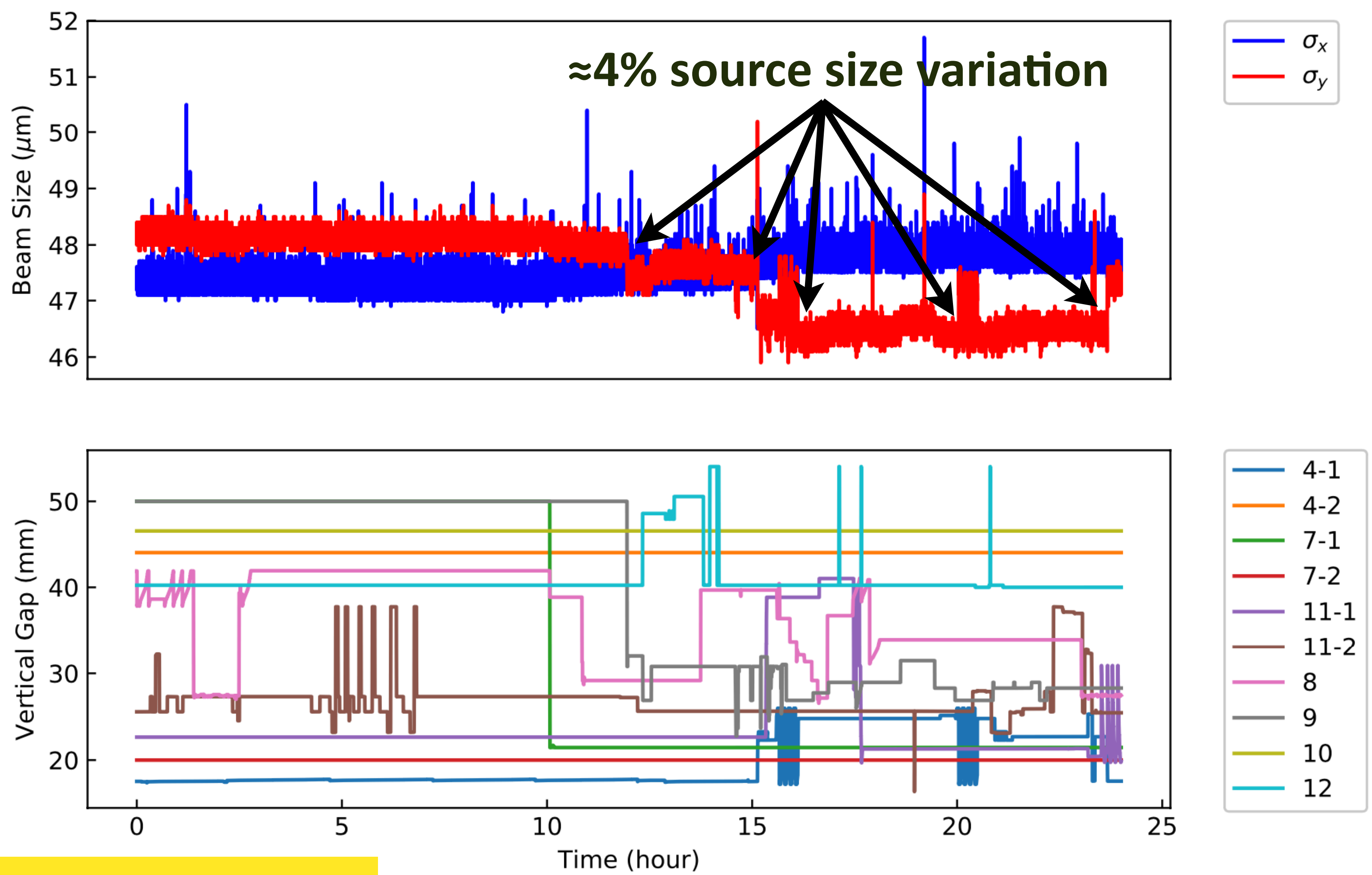
Courtesy: C. Steier, PAC'09





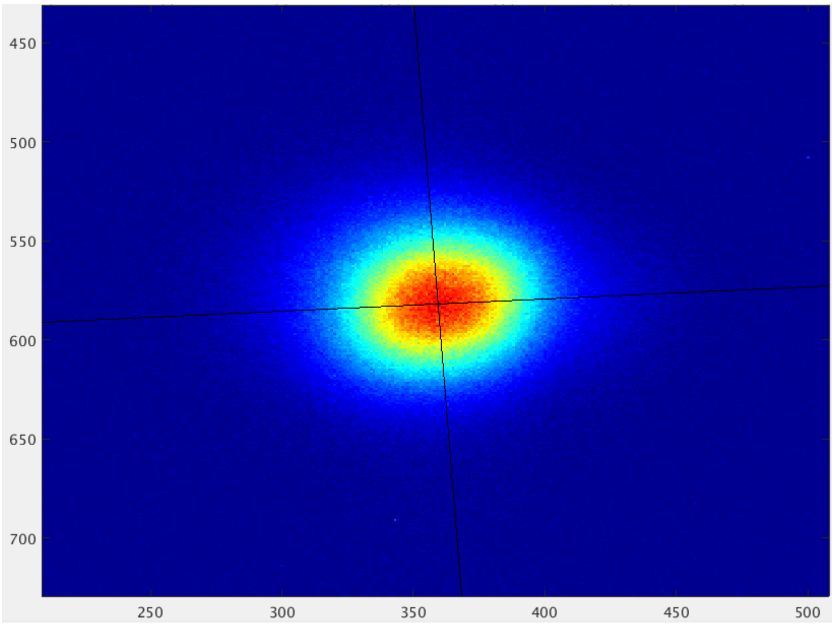
The Problem: Beam Size vs. ID Motion

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



PRL **123**, 194801 (2019)

ALS Diagnostic Beamline 3.1



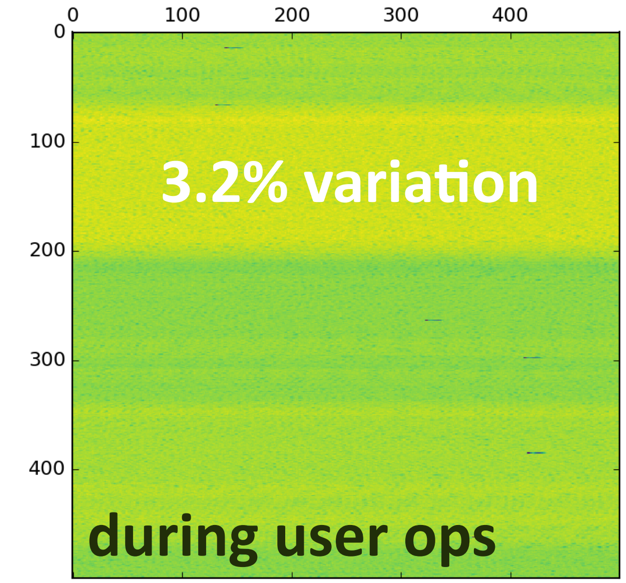
SR from 1st arc dipole ("round beam") → KB mirrors → C filter → 1-3 keV x-rays → LYSO scintillator crystal → visible → CCD

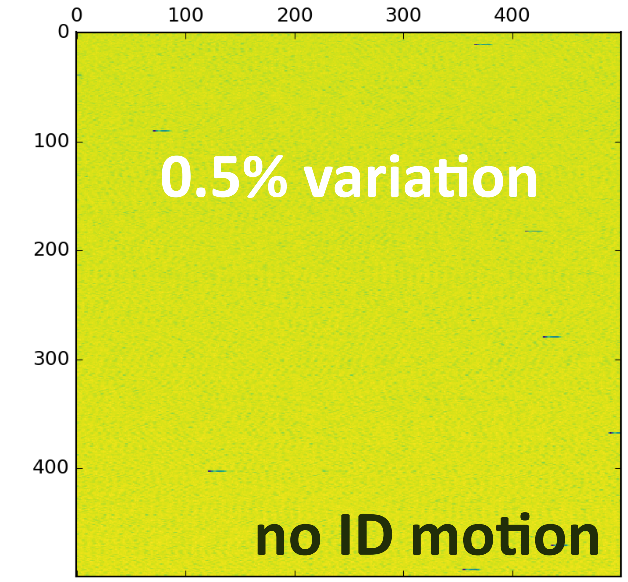
Rev. Sci. Instrum. **67**, 3368 (1996)

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

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.





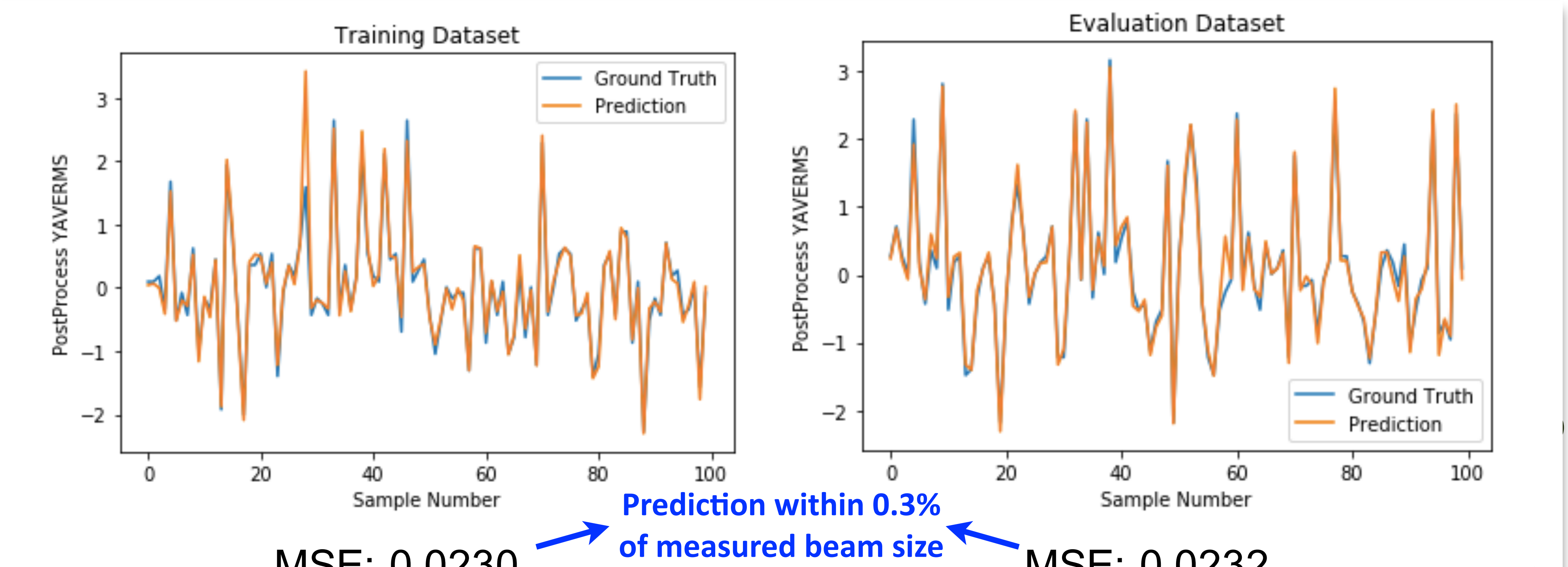
PRL **123**, 194801 (2019)

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?
  - –
  - –
  - –



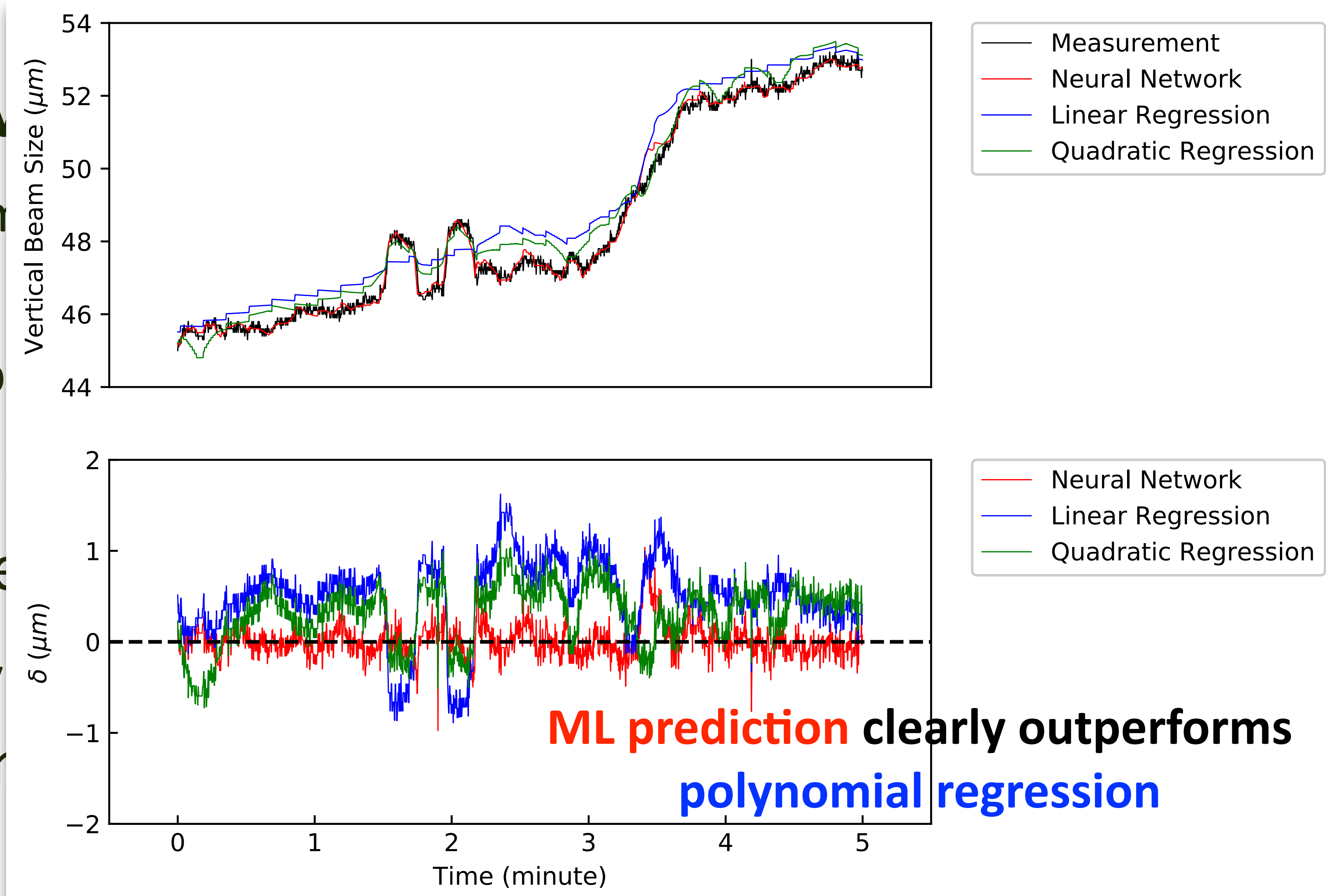
Courtesy: S. Liu

- M
- M … l
  sy

- 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

Need to Solve This Problem at the Source

- Why use M
  - ML can m
  - ML does ... machine drift) → b ... ori
  - ML can s ... egression)
- ML require ... ents
- ML ideally ... al control system car



- First example: offline analysis of user ops data

PRL **123**, 194801 (2019)

  - 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

From Prediction to Correction

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

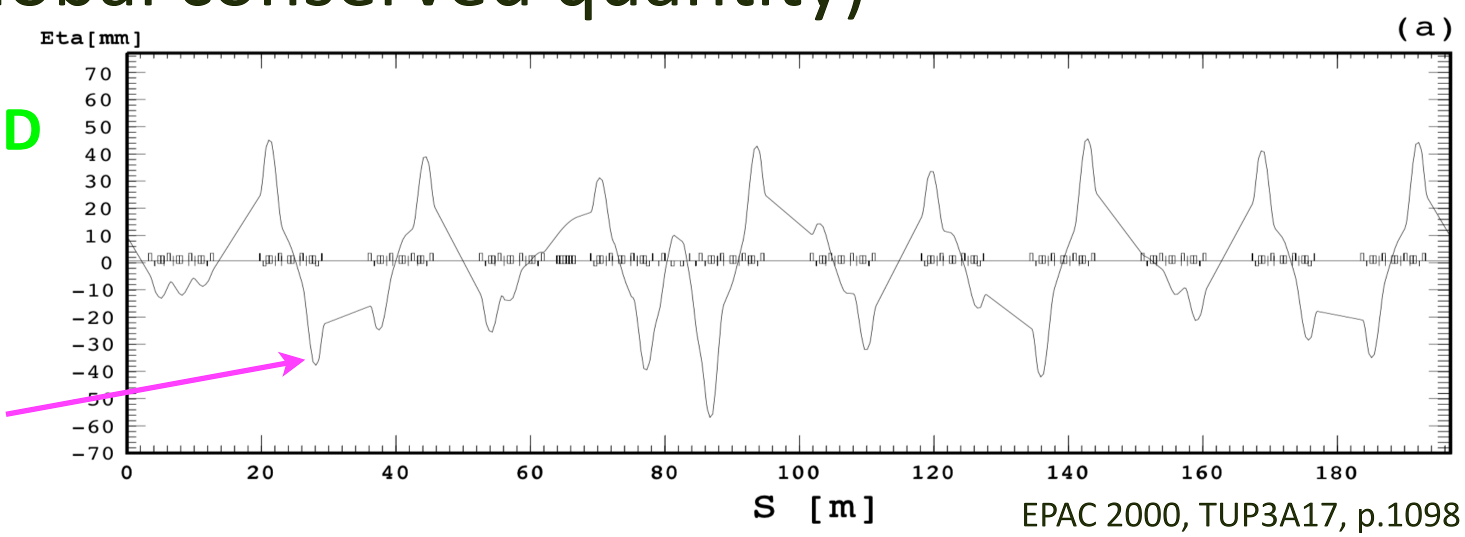
SQSF SQSD

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

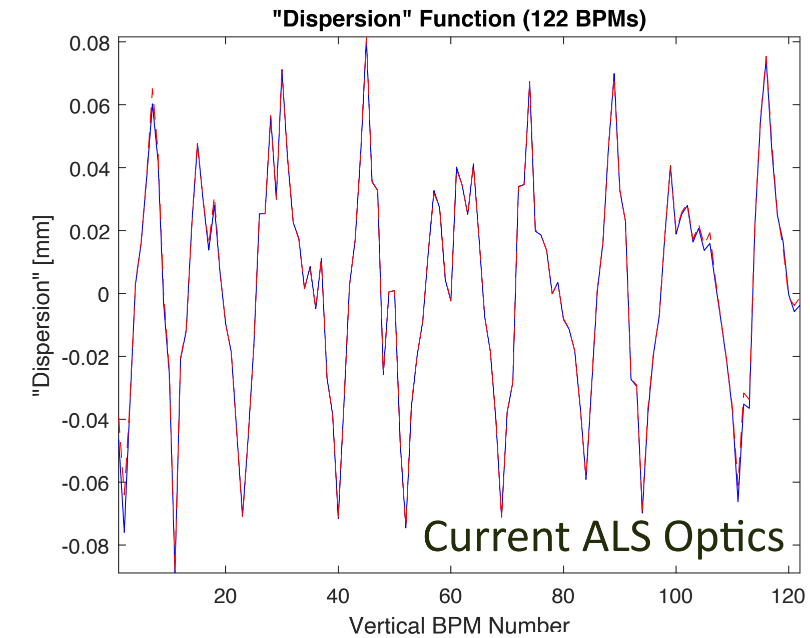
LOCO & Setup

DWP

Dispersion Wave



EPAC 2000, TUP3A17, p.1098

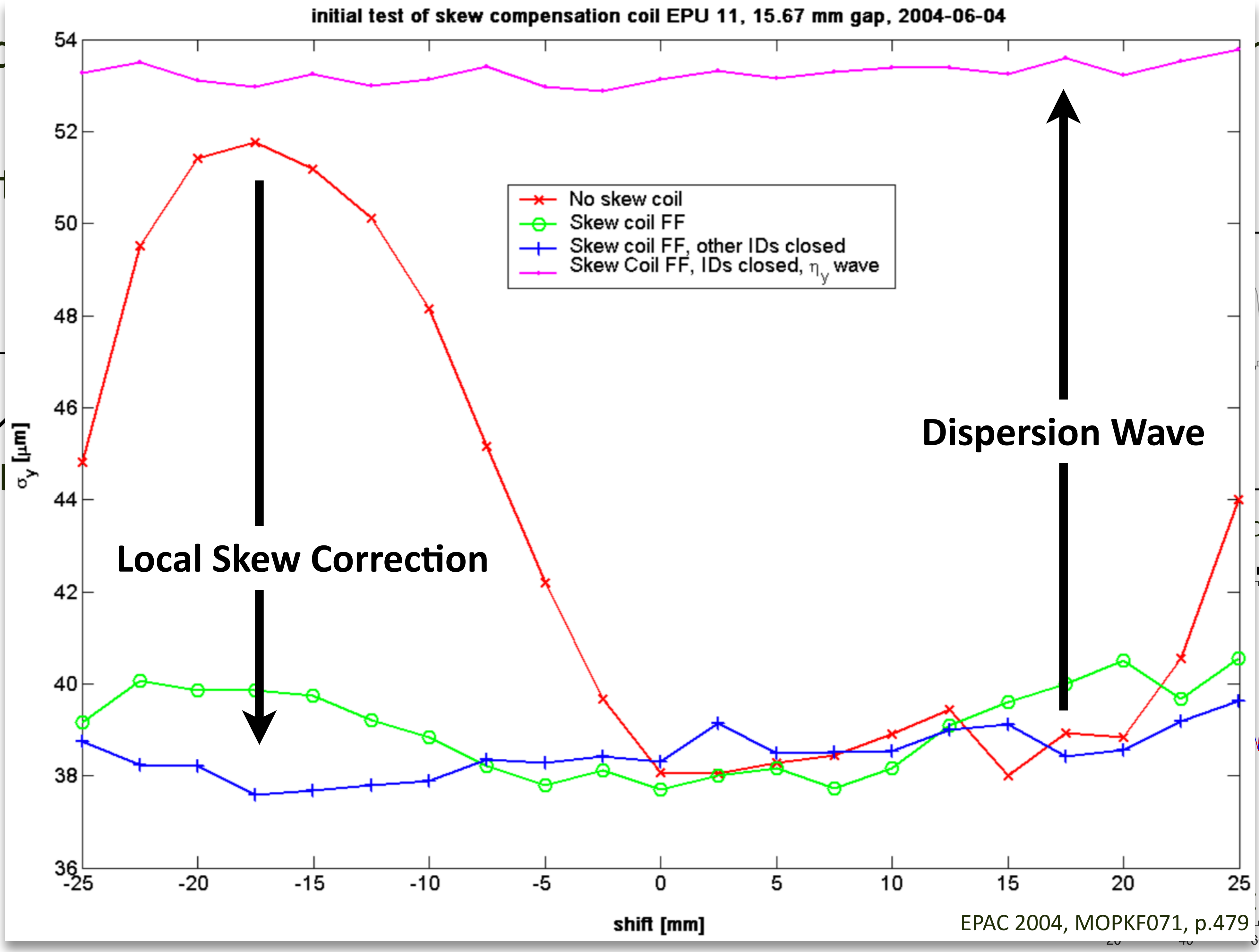


From Prediction to Correction (cont.)

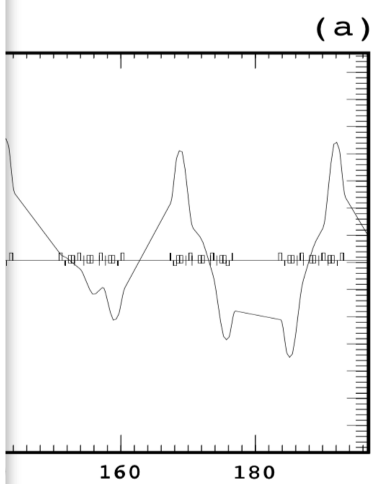
- Introd … ndard ALS d … lows adjust

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

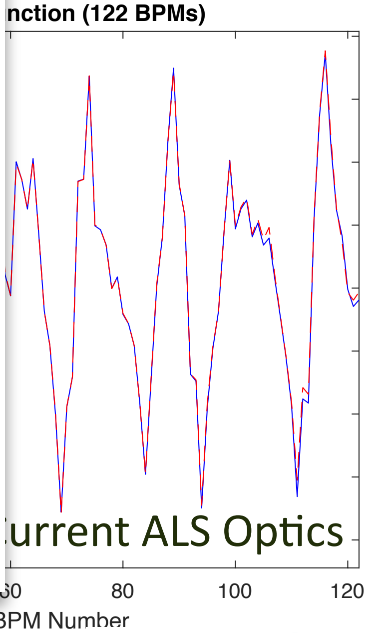
LOCO & Setu



EPAC 2004, MOPKF071, p.479



C 2000, TUP3A17, p.1098



BERKELEY LAB

ALS

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)

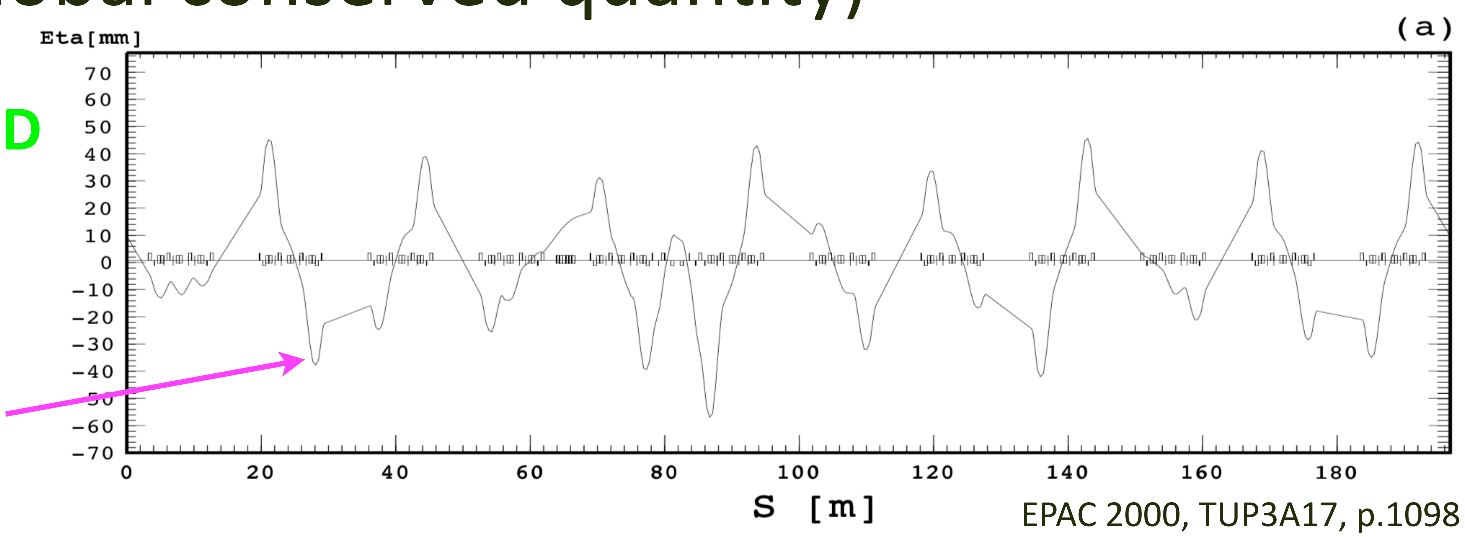
SQSF SQSD

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

LOCO & Setup

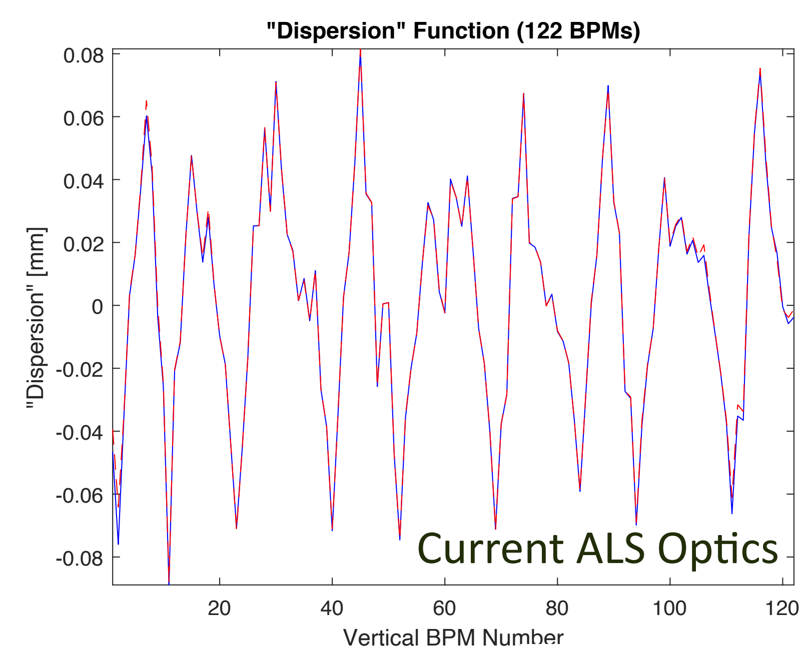
DWP

Dispersion Wave



EPAC 2000, TUP3A17, p.1098

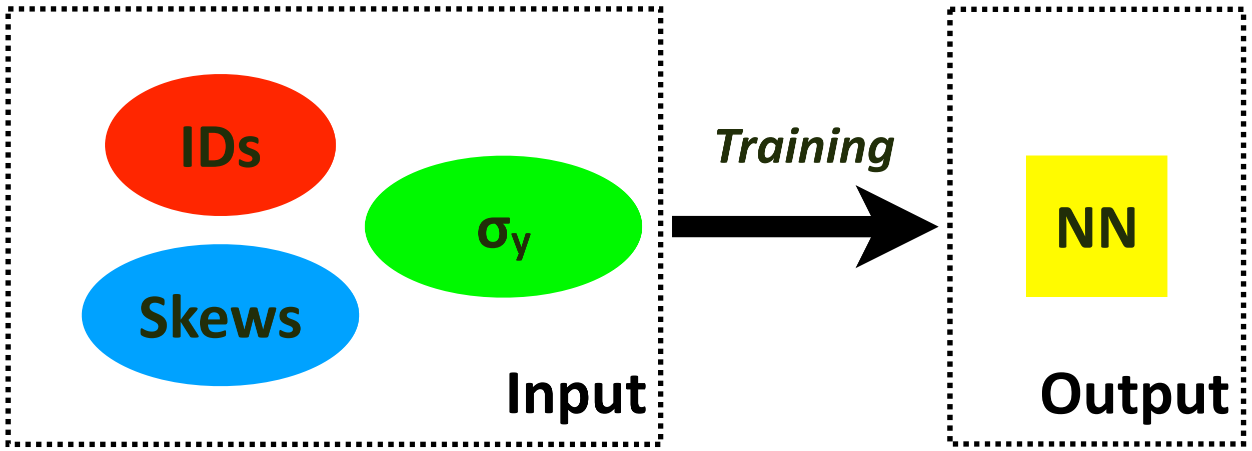
- Observed varying ID configurations affect primarily vertical dispersion → ε_y
- Can therefore stabilize beam size globally by adjusting DWP



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)

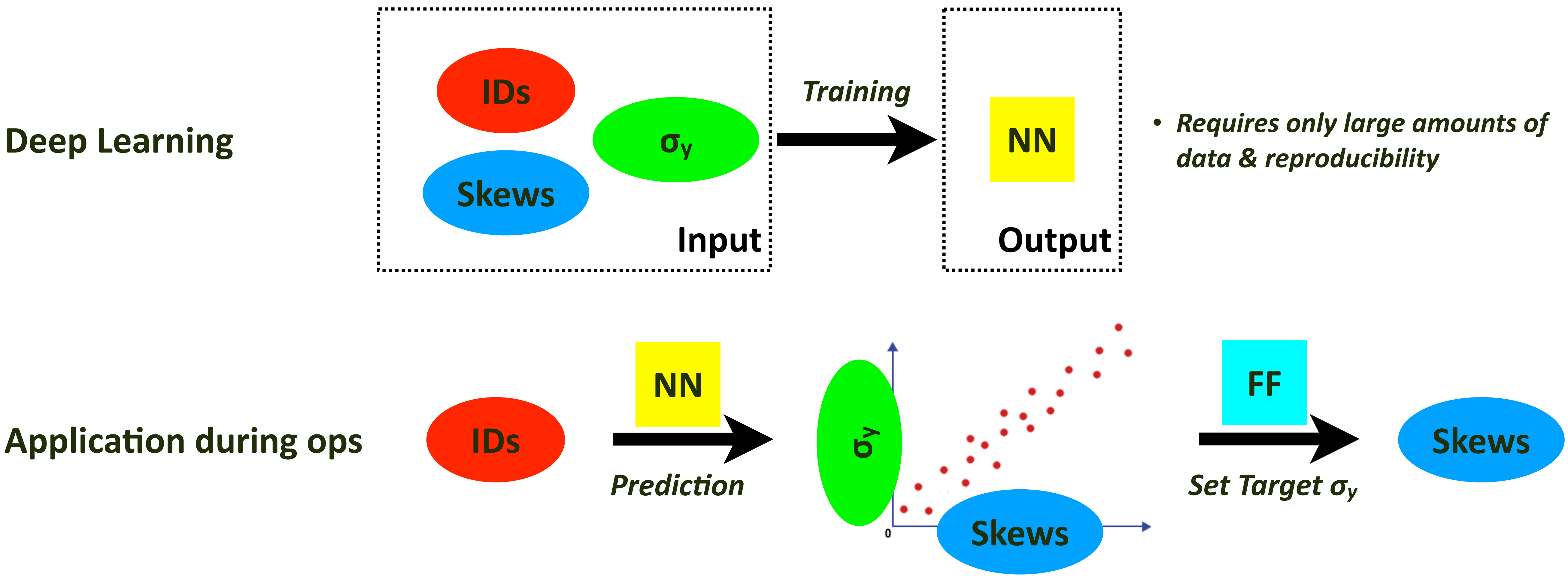
Deep Learning



- ***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

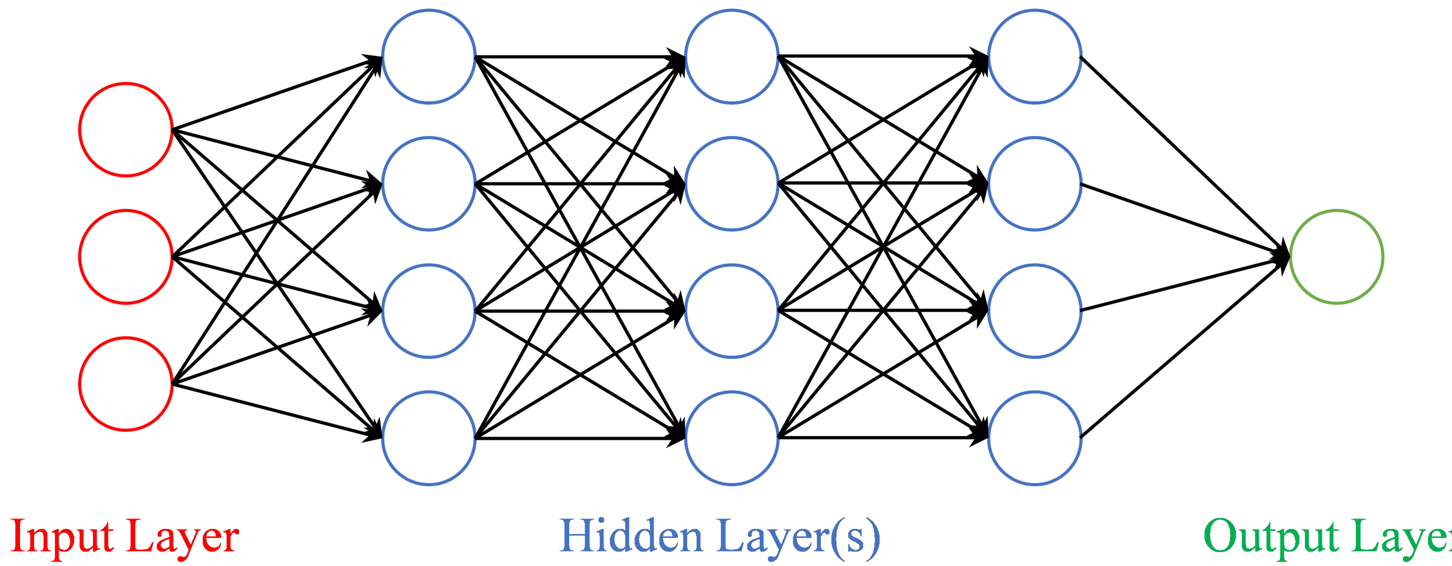


- ***Requires only large amounts of data & reproducibility***

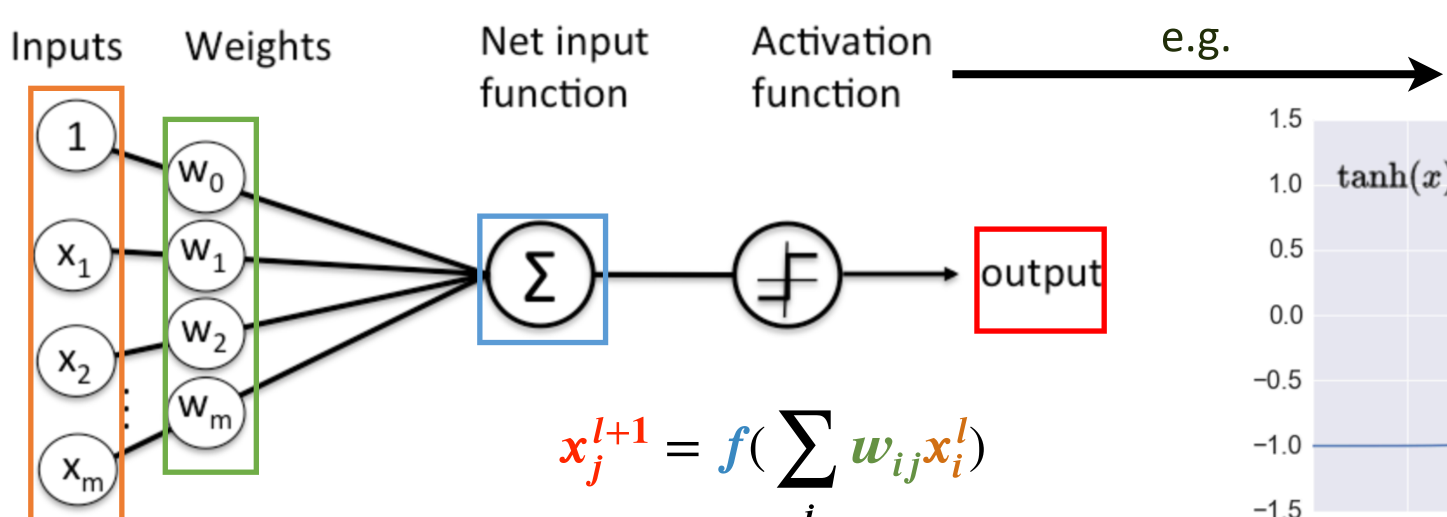
BERKELEY LAB

ALS

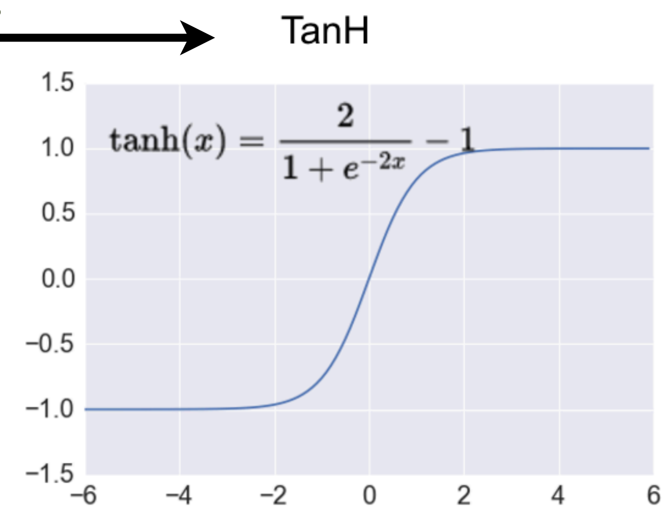
How a Neural Network (NN) Works

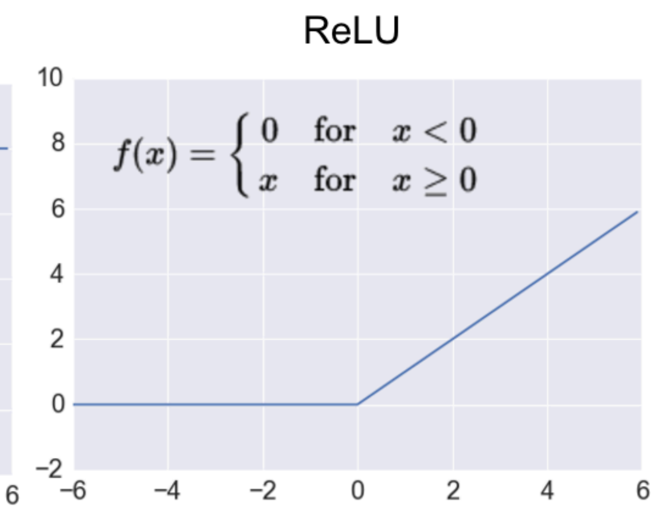


Courtesy: S. Liu

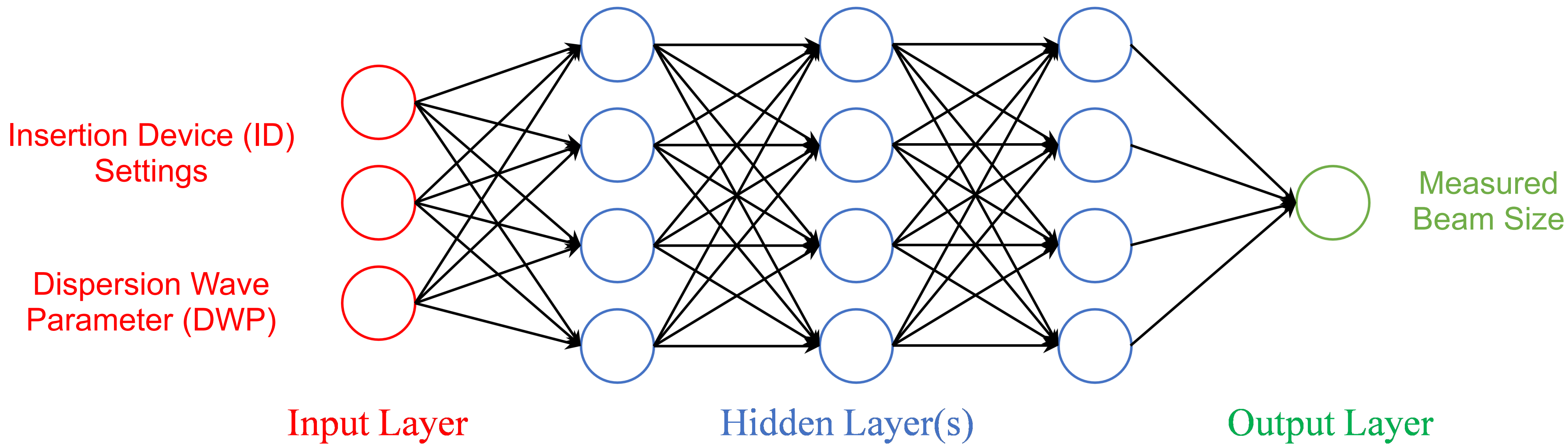


$$x_j^{l+1} = f\left(\sum_j w_{ij} x_i^l\right)$$



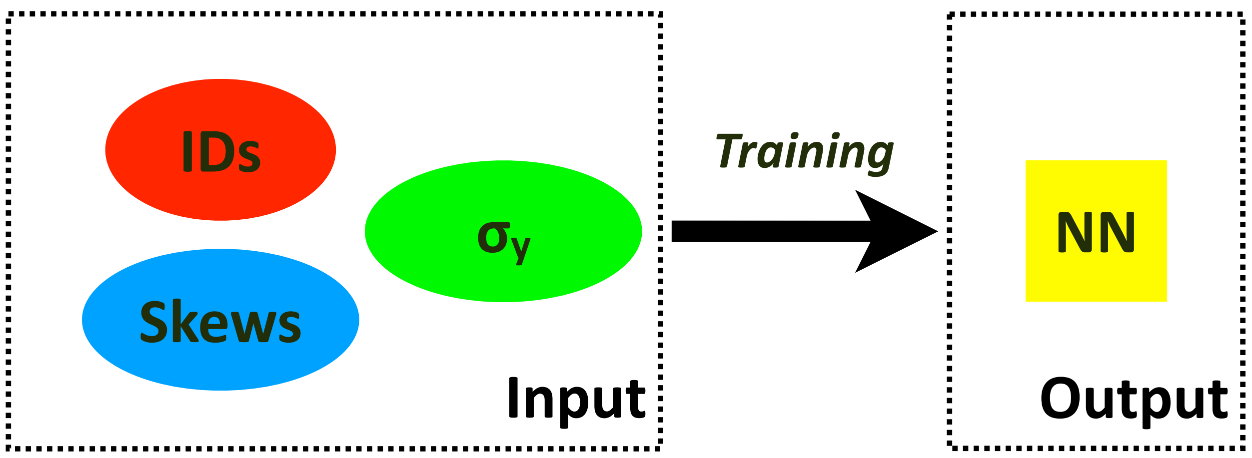


Deep Learning: How we Trained the NN



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)

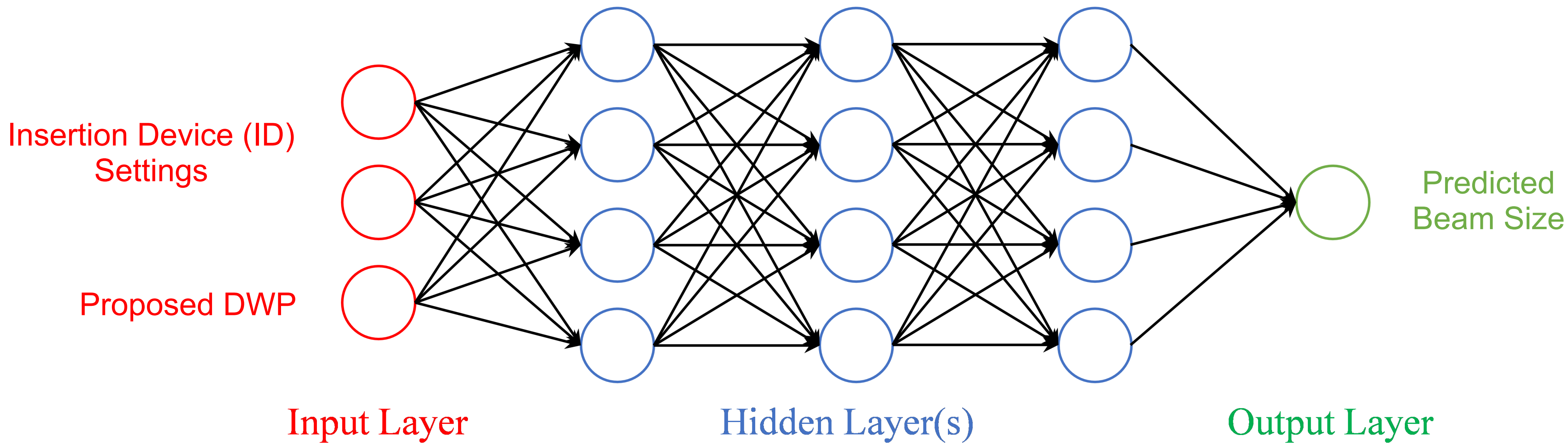
Regularization: L_2 regularizer with $\lambda = 10^{-4}$
Optimization: Adam Optimizer with learning rate $\alpha = 10^{-3}$



	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

PRL **123**, 194801 (2019)

Resulting NN Enables ID Feed-Forward at ≈3 Hz



Proposed DWPs

-0.06
….
0
…
0.06

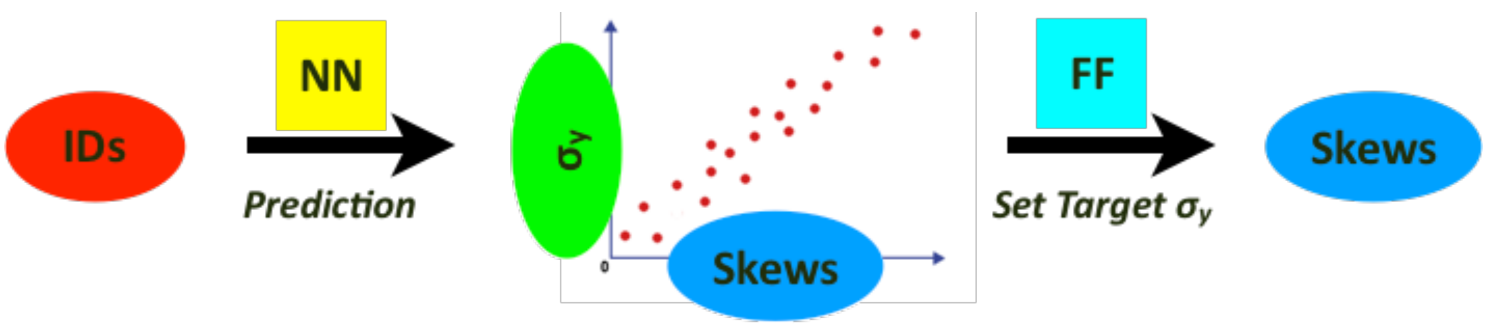
Measured ID Settings $\xrightarrow{\hspace{3em}}$ Neural Network

Predicted Beam Sizes

50.3
….
52.1
…
54.0

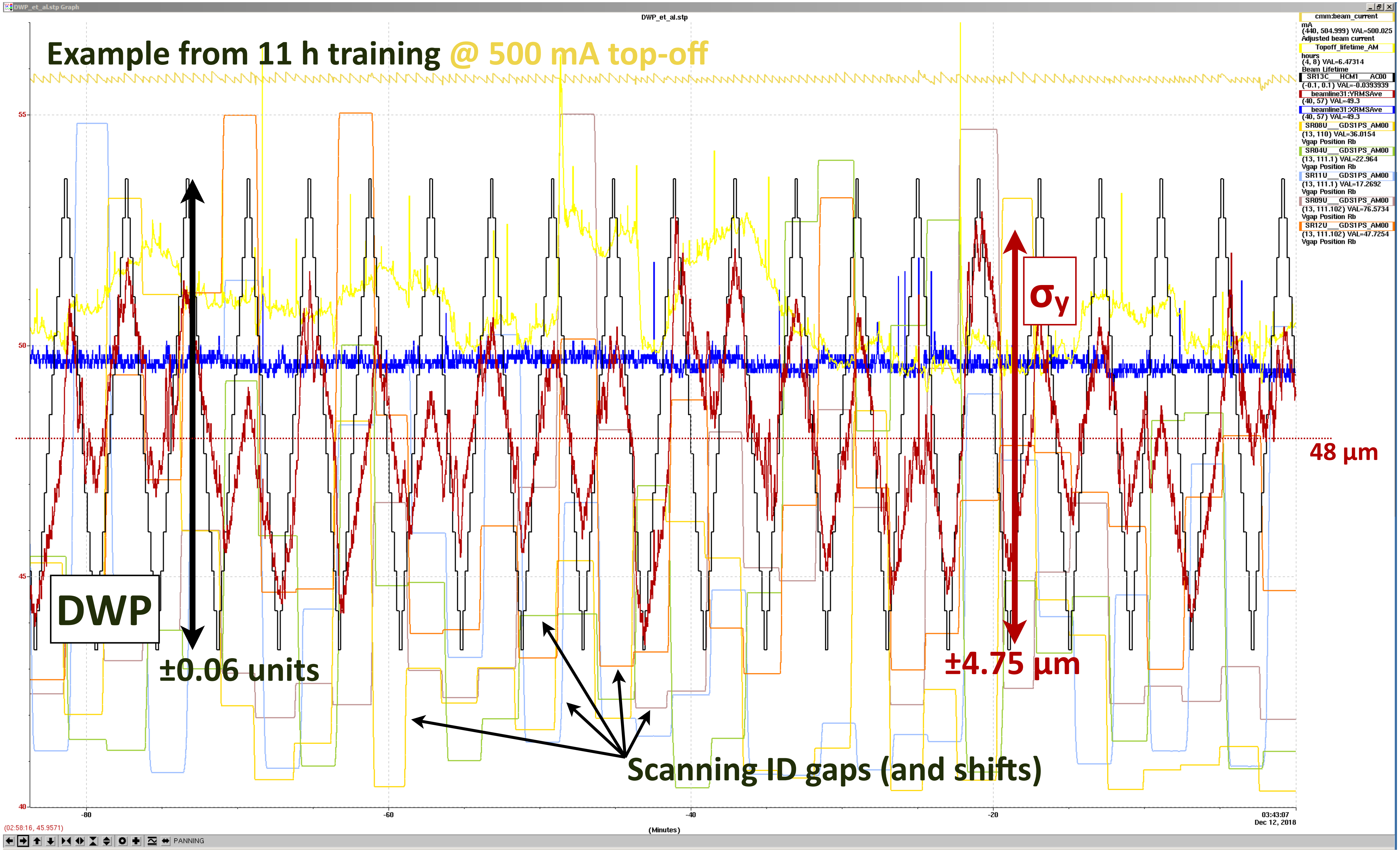
Compare with Target $\xrightarrow{\hspace{3em}}$ Beam Size

Choose proper DWP

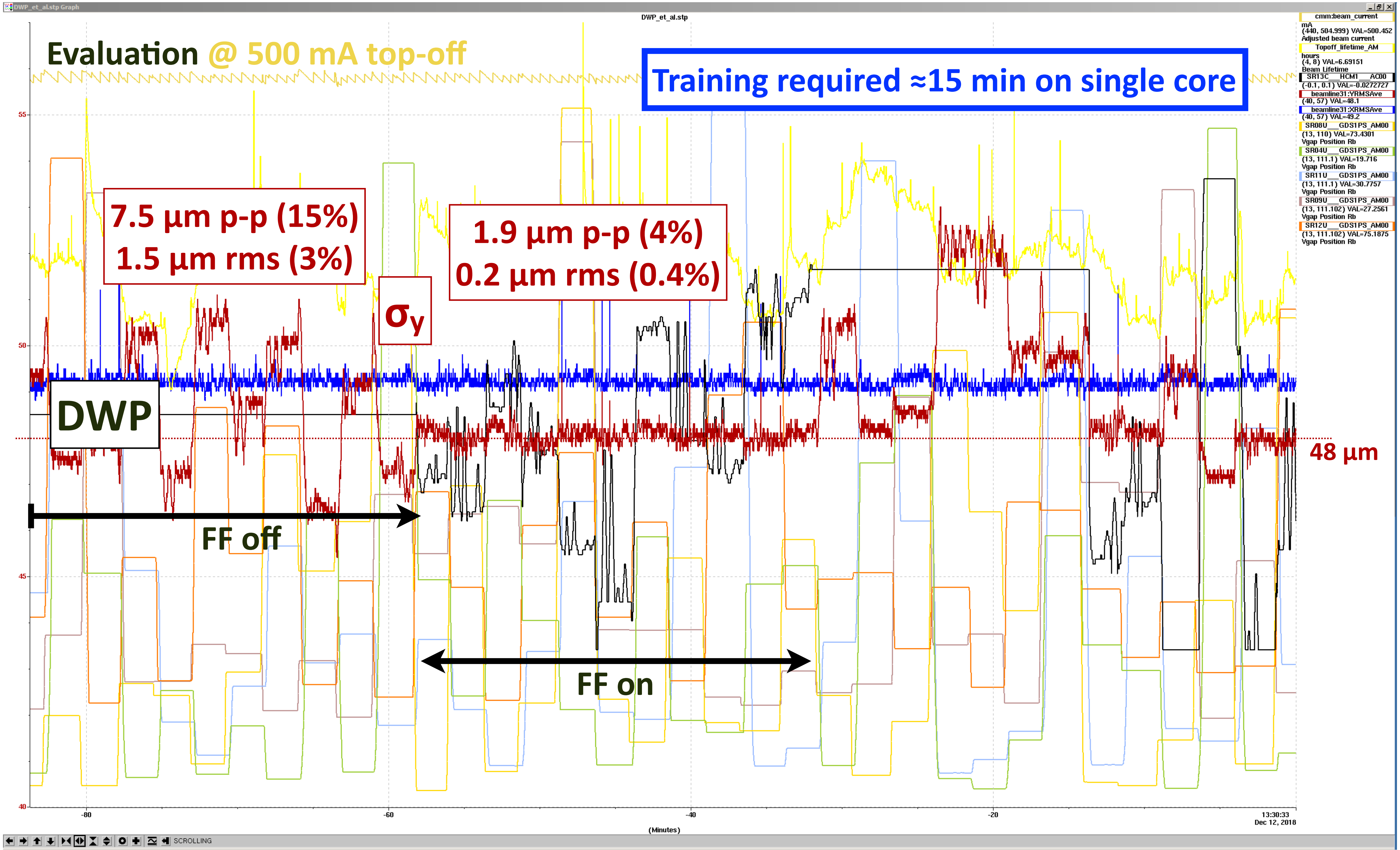


PRL **123**, 194801 (2019)

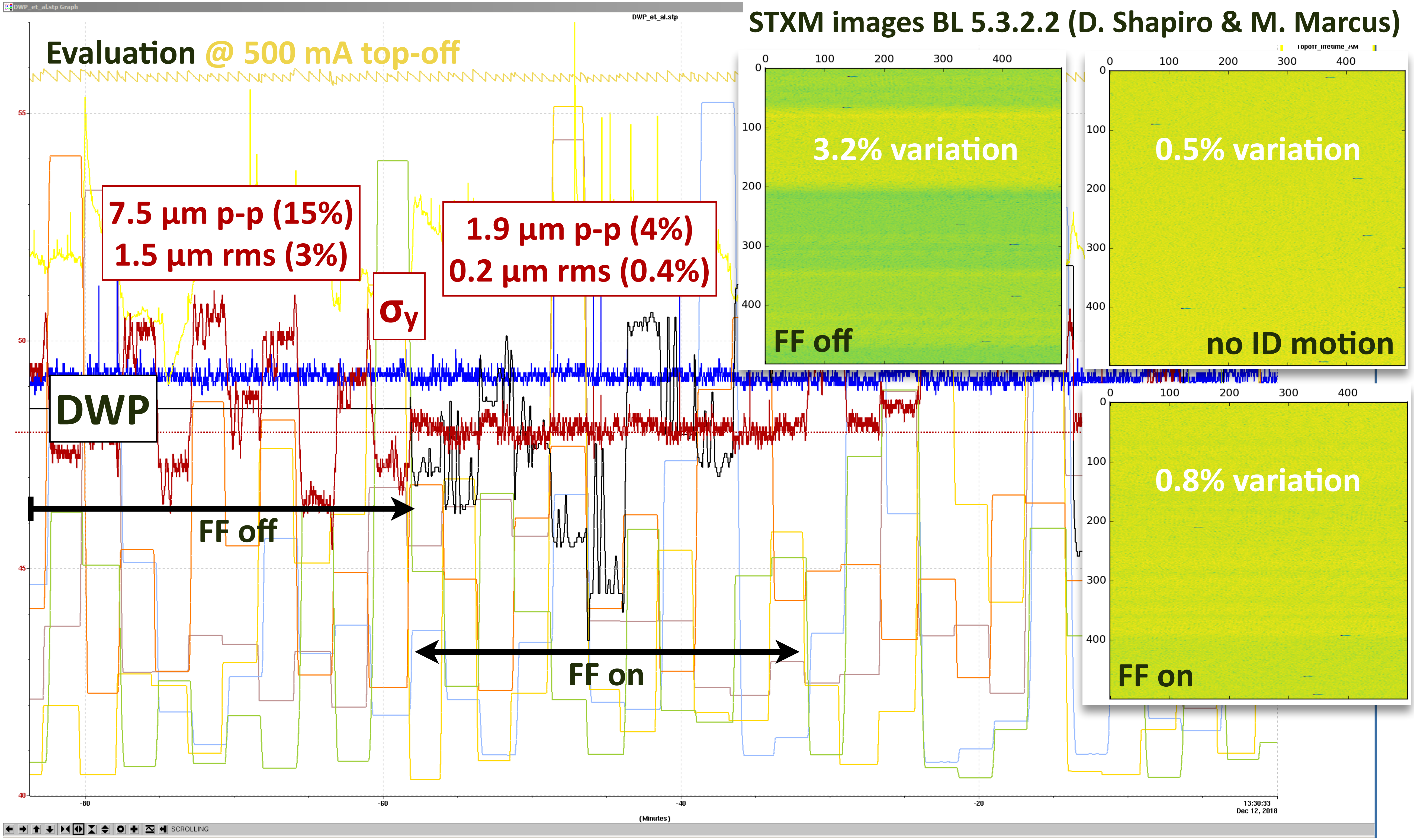
Physics Shift: Data Collection for NN Training



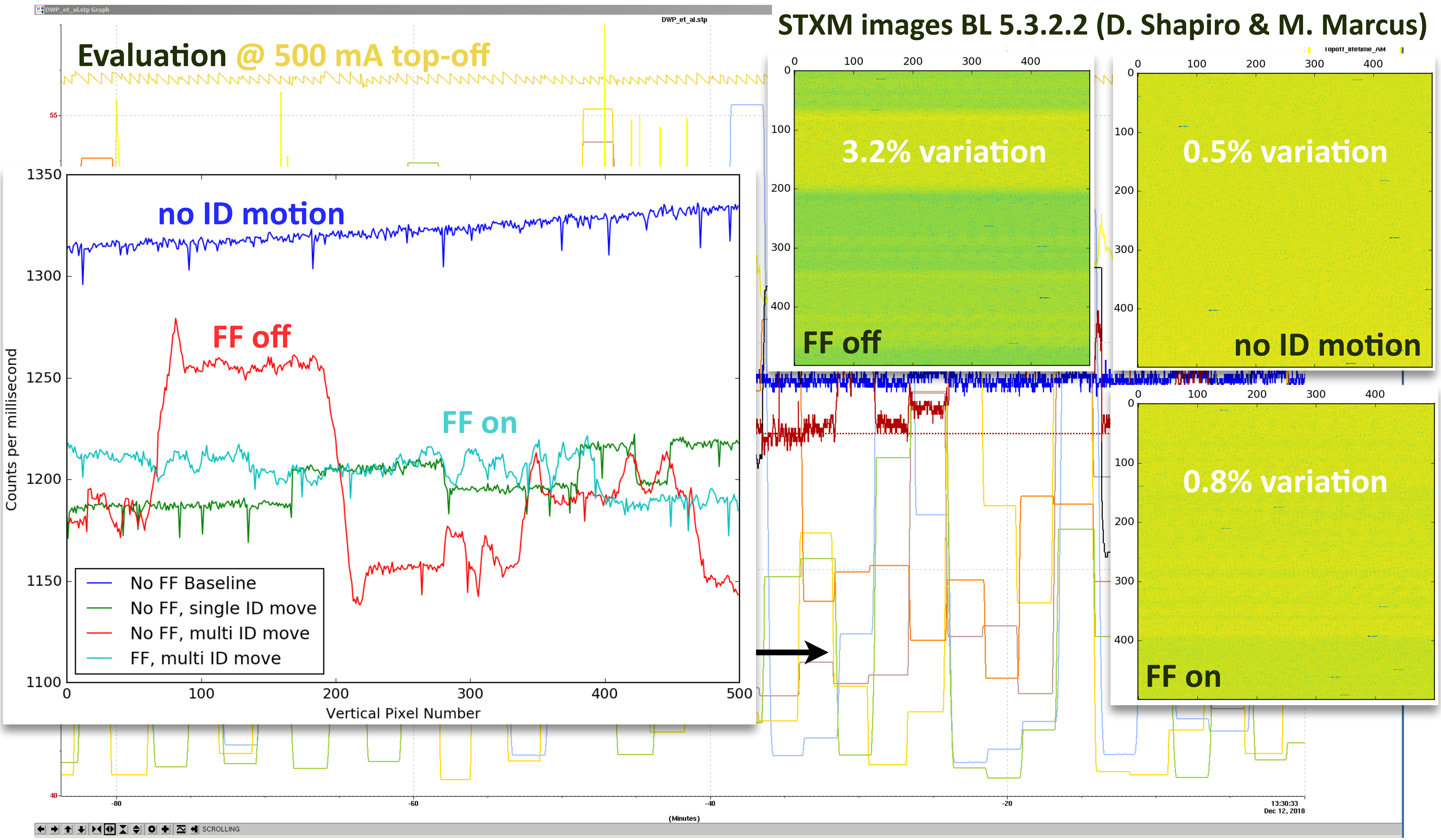
Physics Shift: Running NN-based ID Feed-Forward



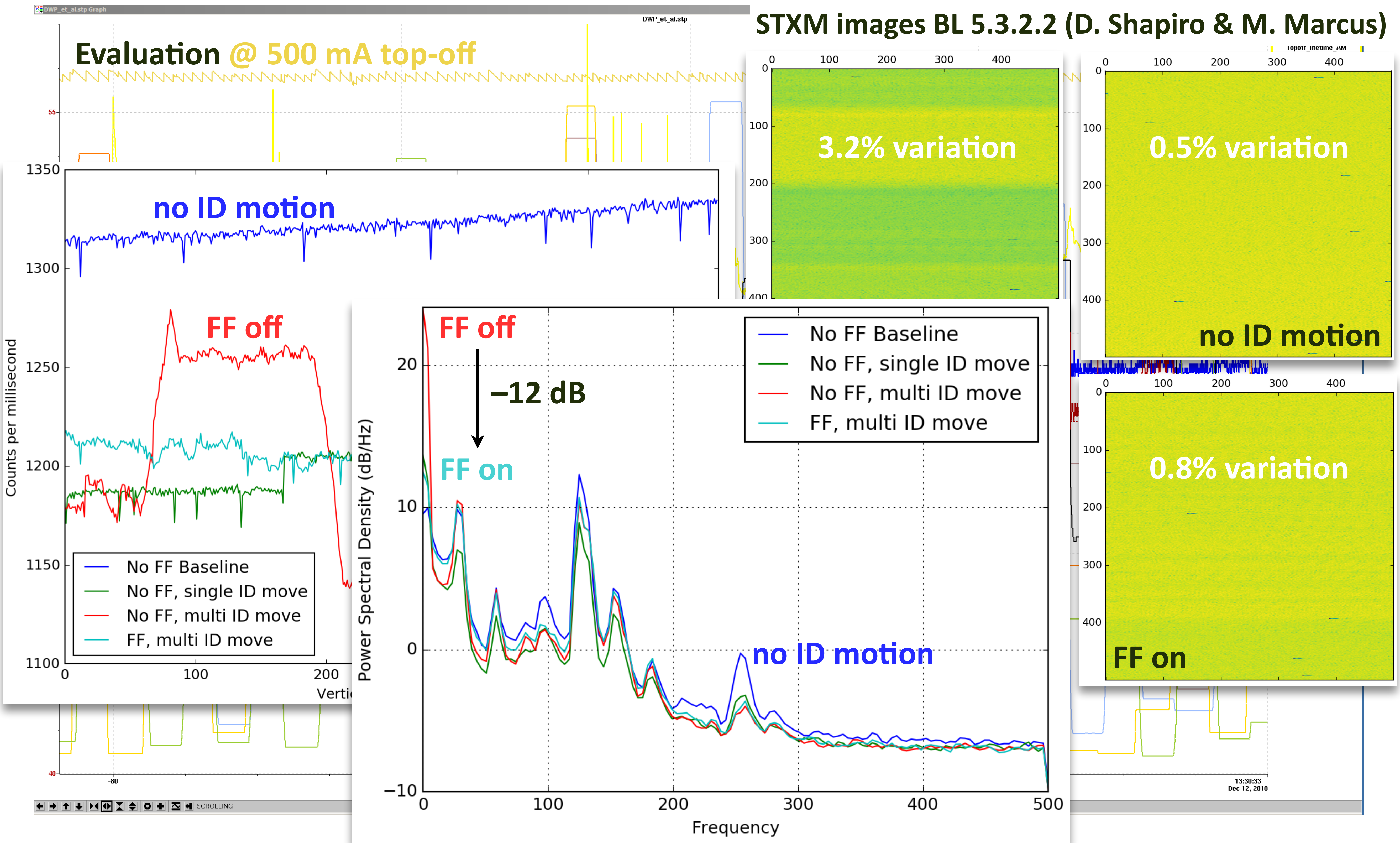
Physics Shift: Running NN-based ID Feed-Forward



Physics Shift: Running NN-based ID Feed-Forward



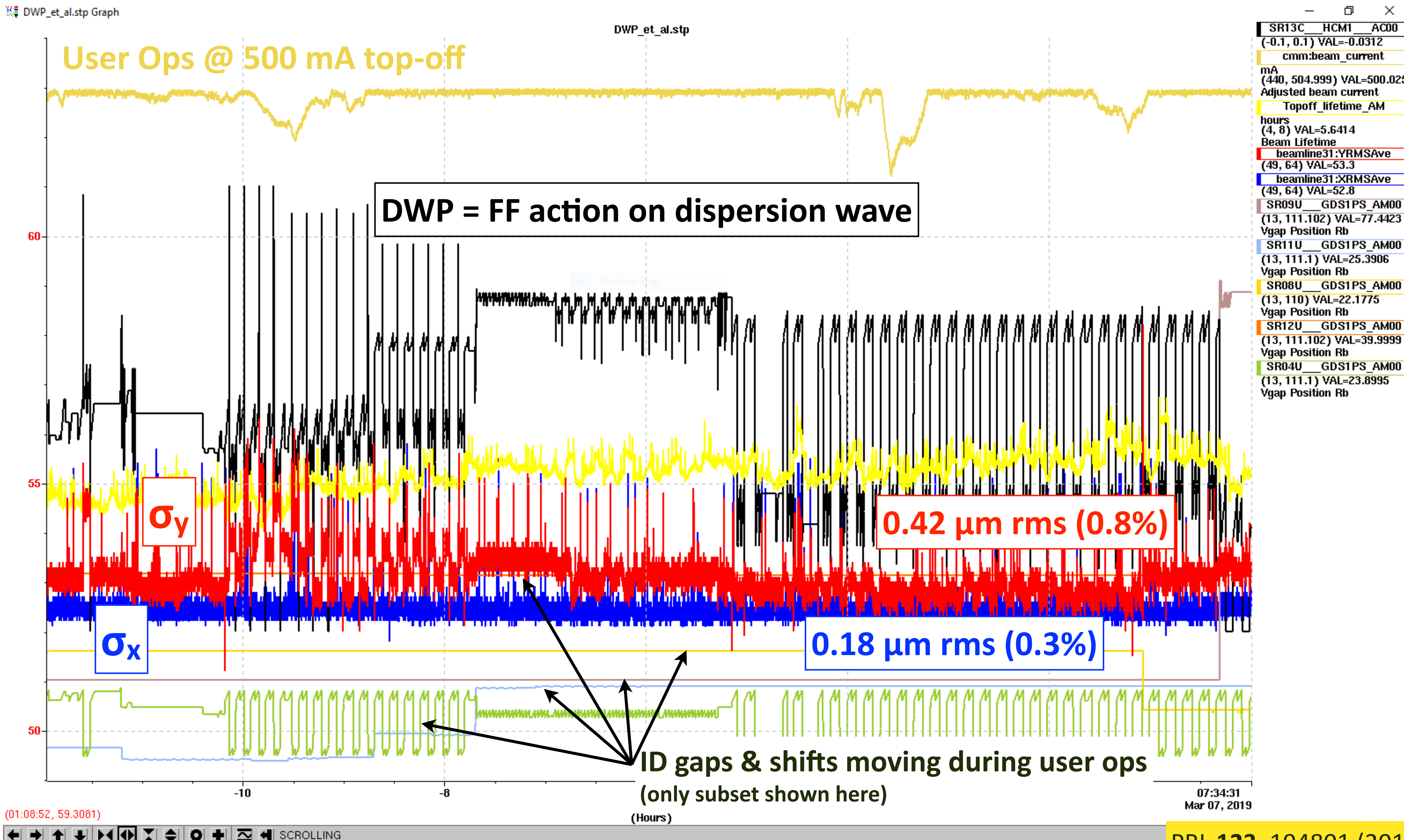
Physics Shift: Running NN-based ID Feed-Forward



Towards First Experiments During User Ops

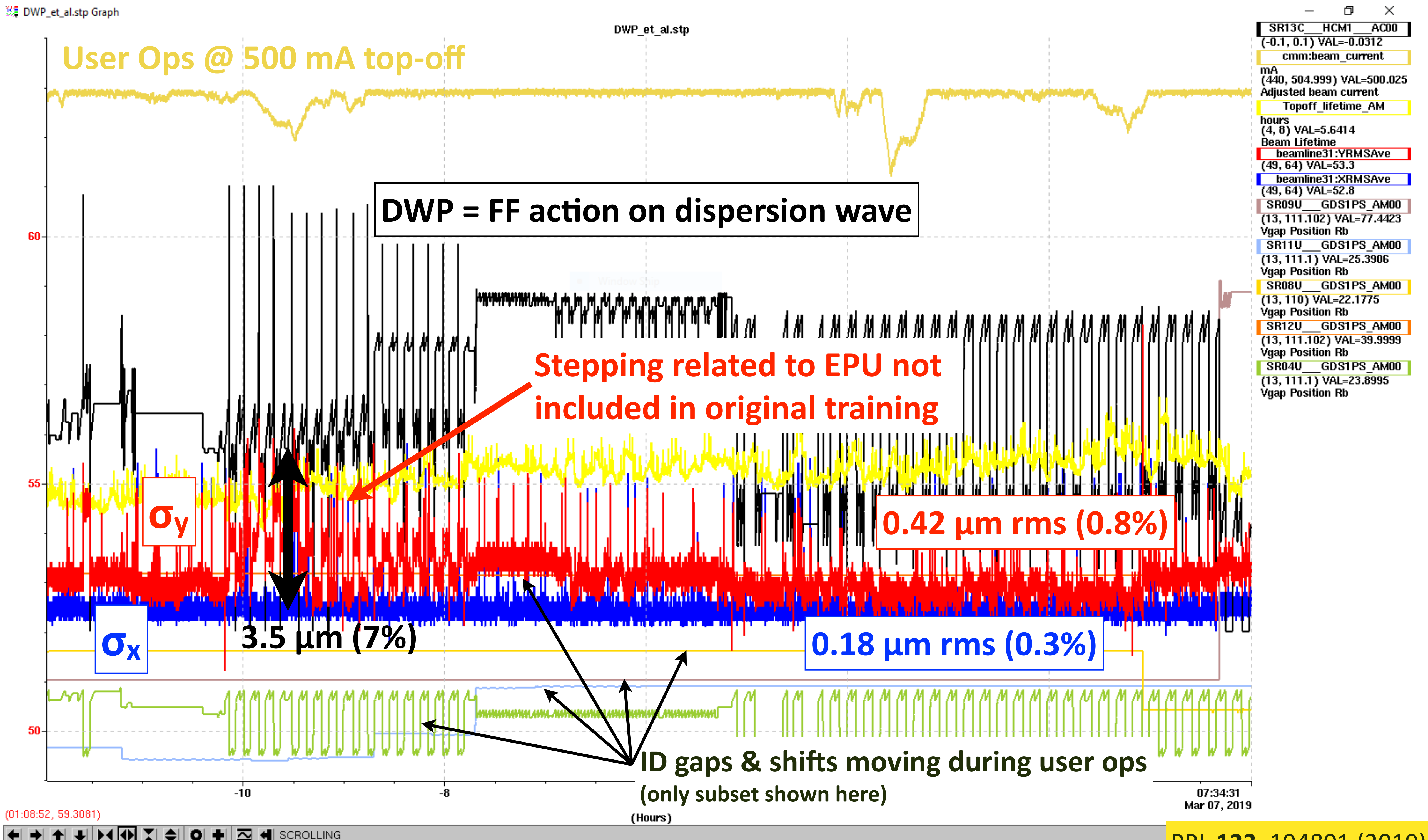
- Use machine shift to acquire training data by scanning operational IDs in a quasi-randomized fashion (favoring operational gap range) → train NN
- Put this NN into FF operation during user ops and evaluate

Stabilization Confirmed During First User Ops Trial



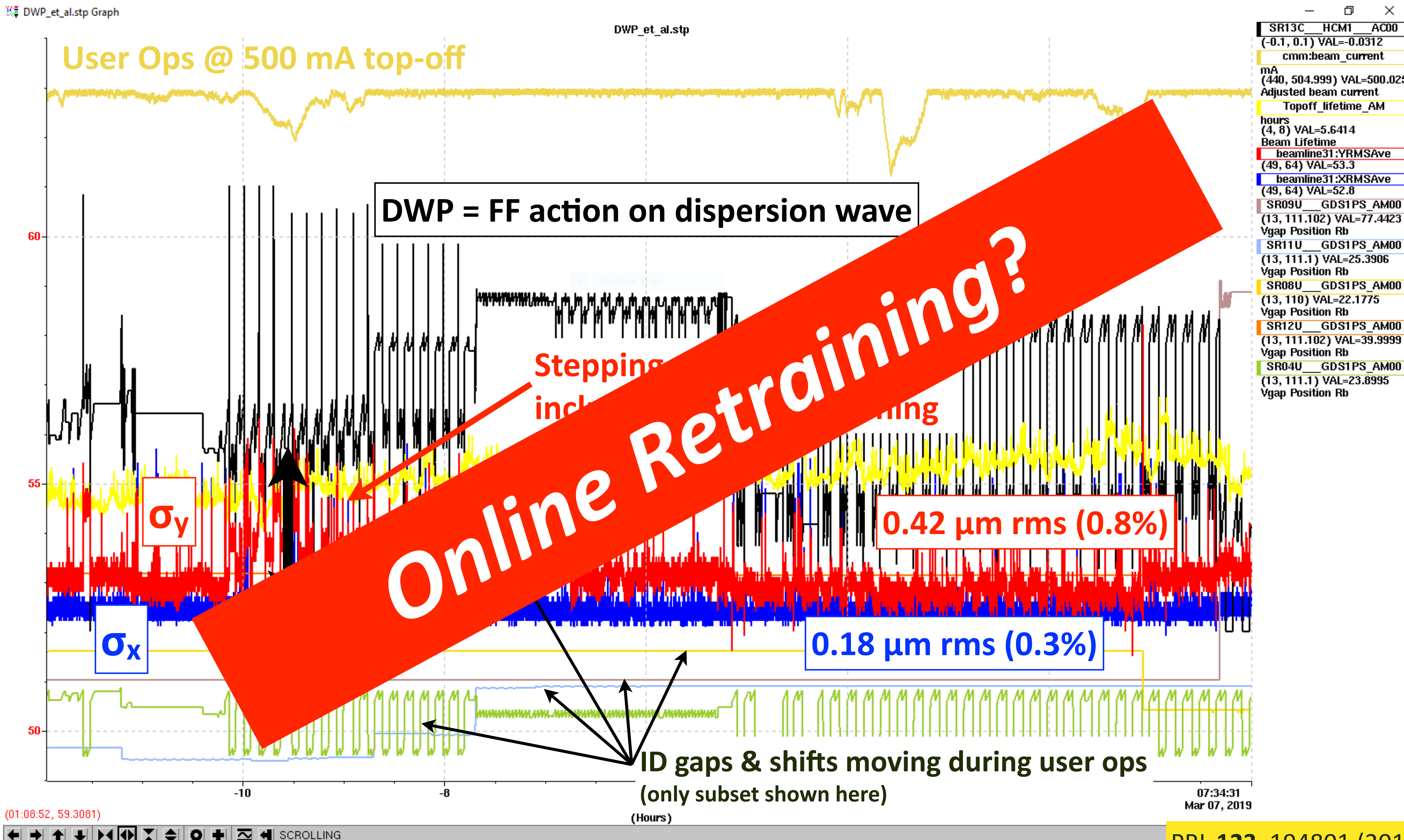
PRL **123**, 194801 (2019)

Stabilization Confirmed During First User Ops Trial



PRL **123**, 194801 (2019)

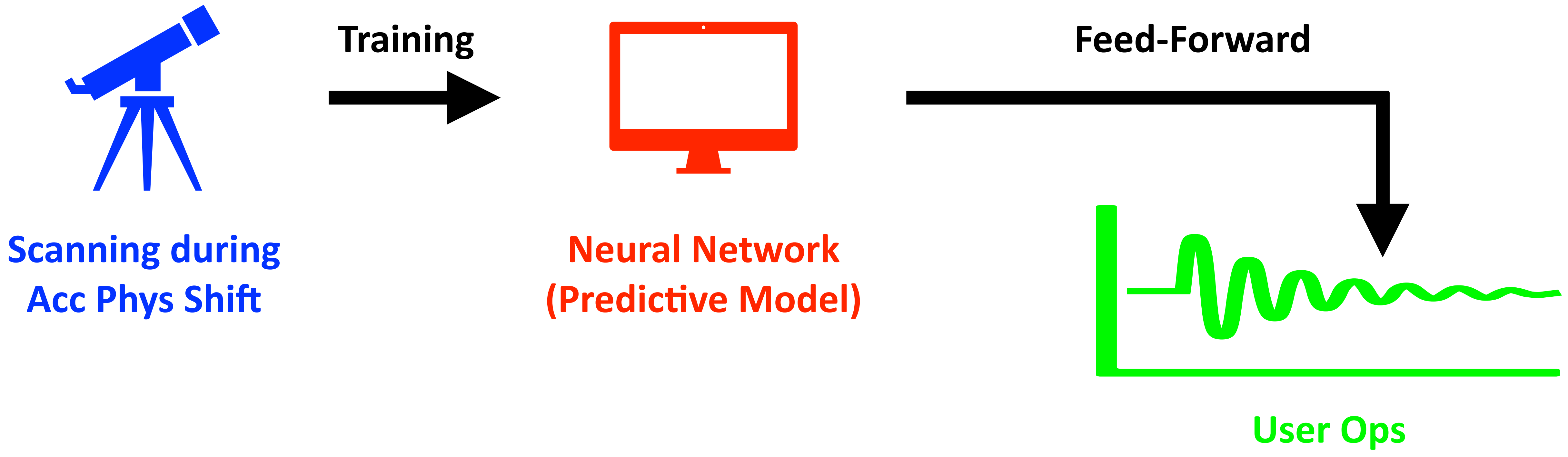
Stabilization Confirmed During First User Ops Trial



PRL **123**, 194801 (2019)

Online Retraining: Improve NN with User Ops Data

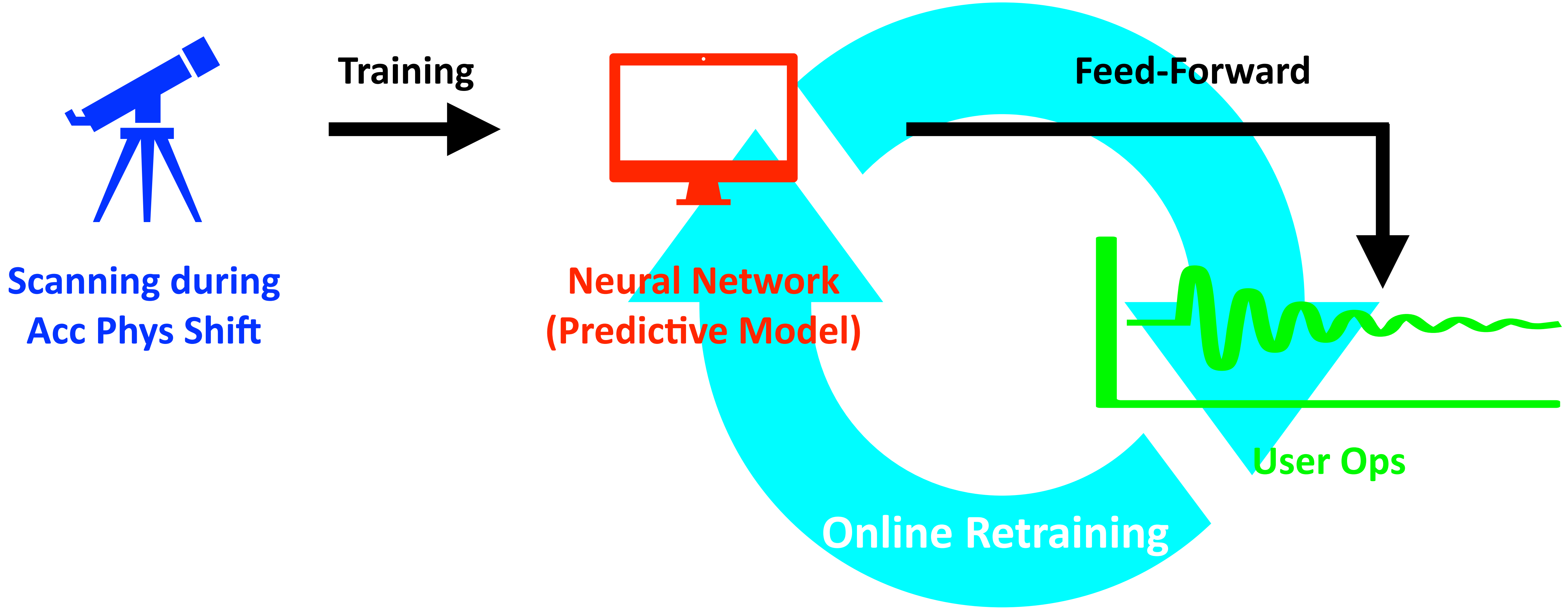
So far: "Conventional" Machine Learning



PRL **123**, 194801 (2019)

Online Retraining: Improve NN with User Ops Data

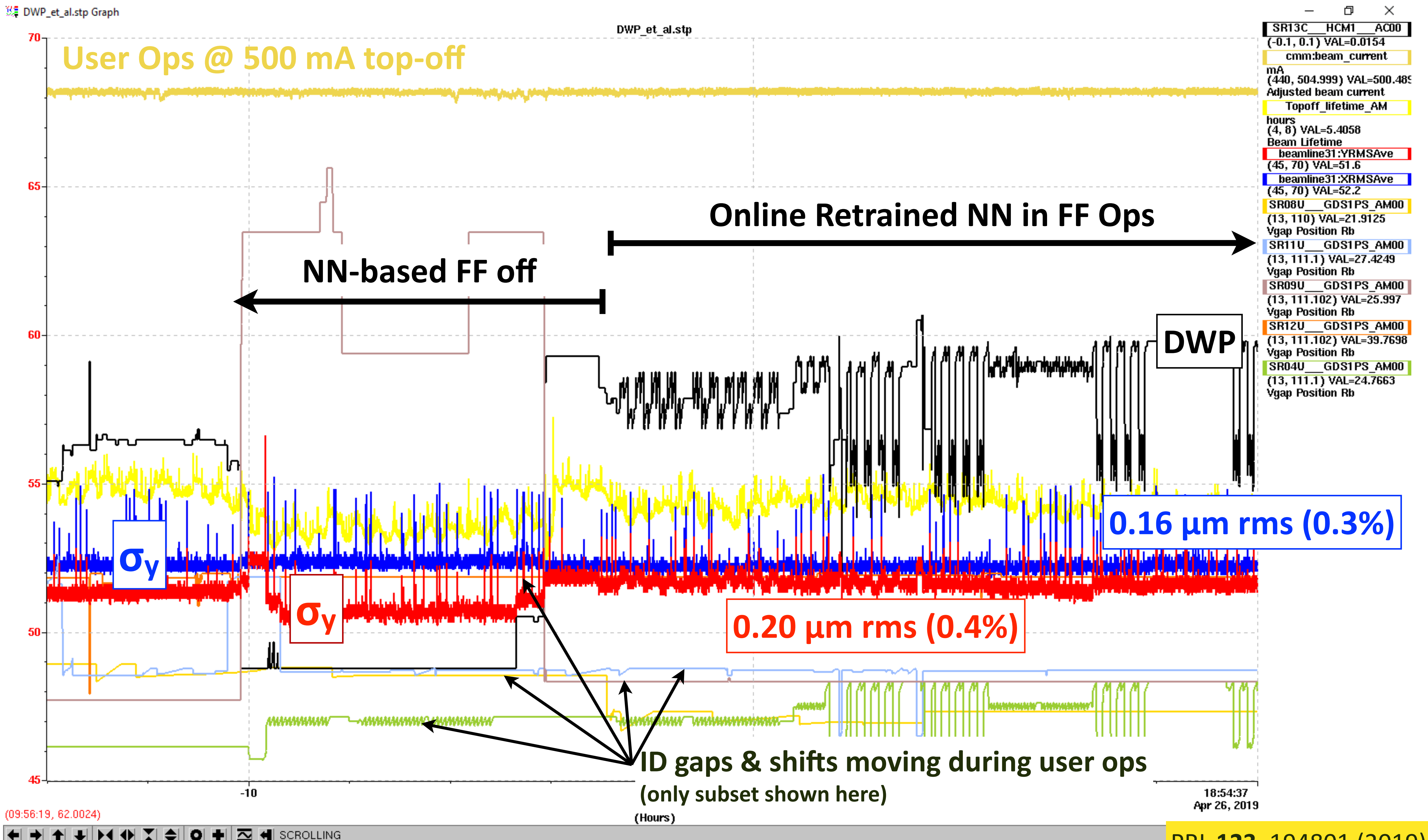
Online Retraining: apply user ops data to improve NN → swap NN used for ID FF on the fly



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

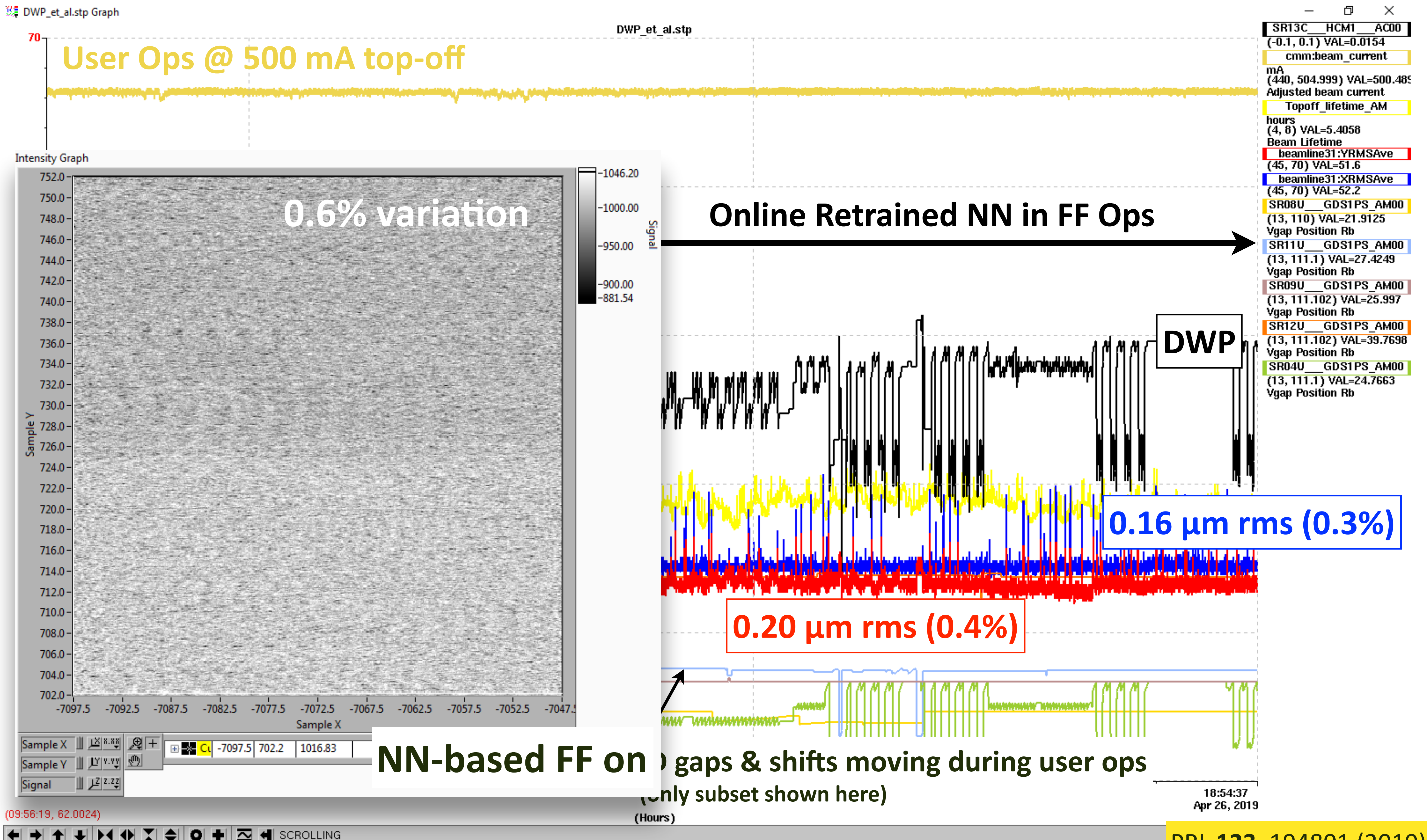
PRL **123**, 194801 (2019)

Substantial Improvement After Online Retraining



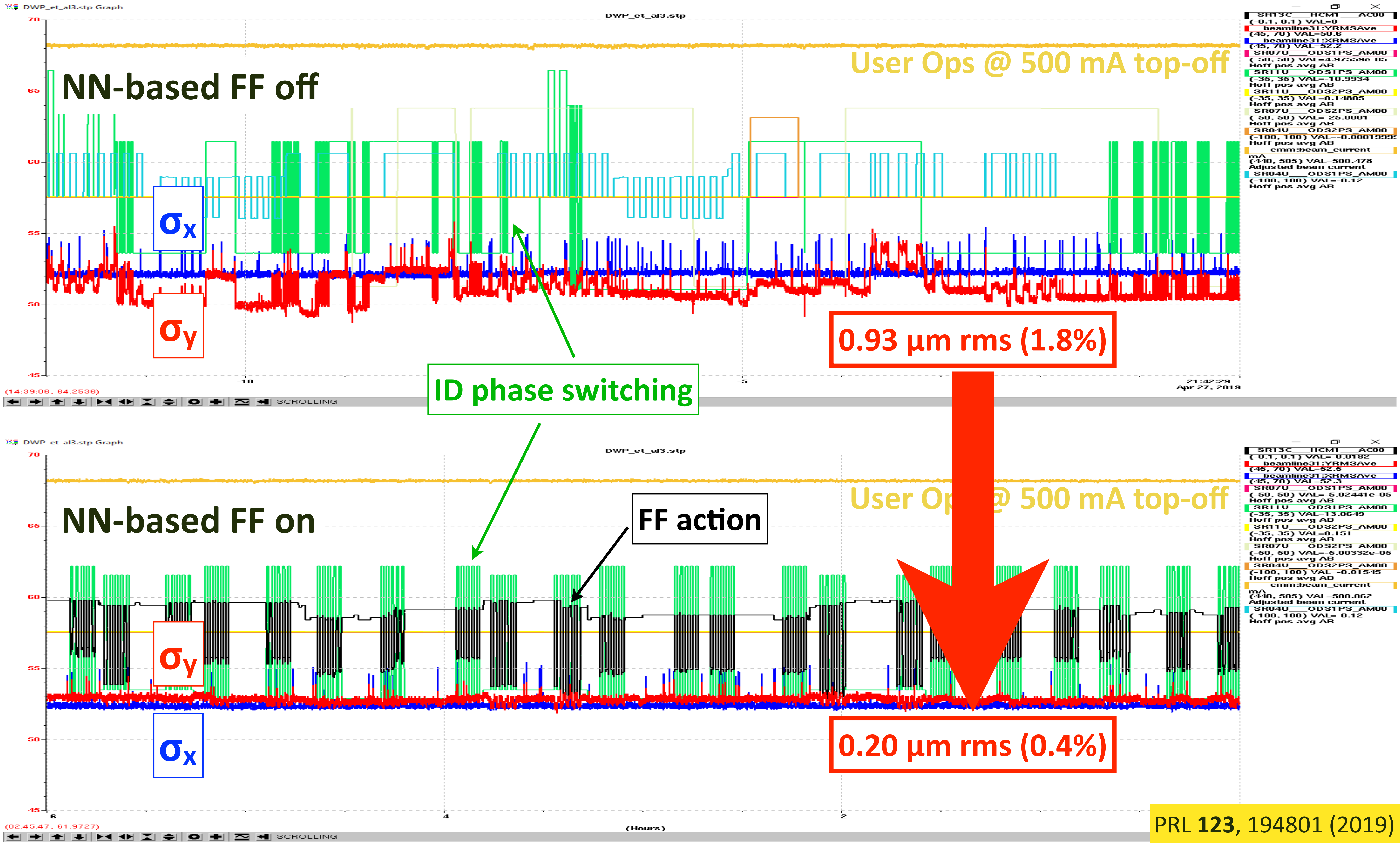
PRL **123**, 194801 (2019)

Substantial Improvement After Online Retraining



PRL **123**, 194801 (2019)

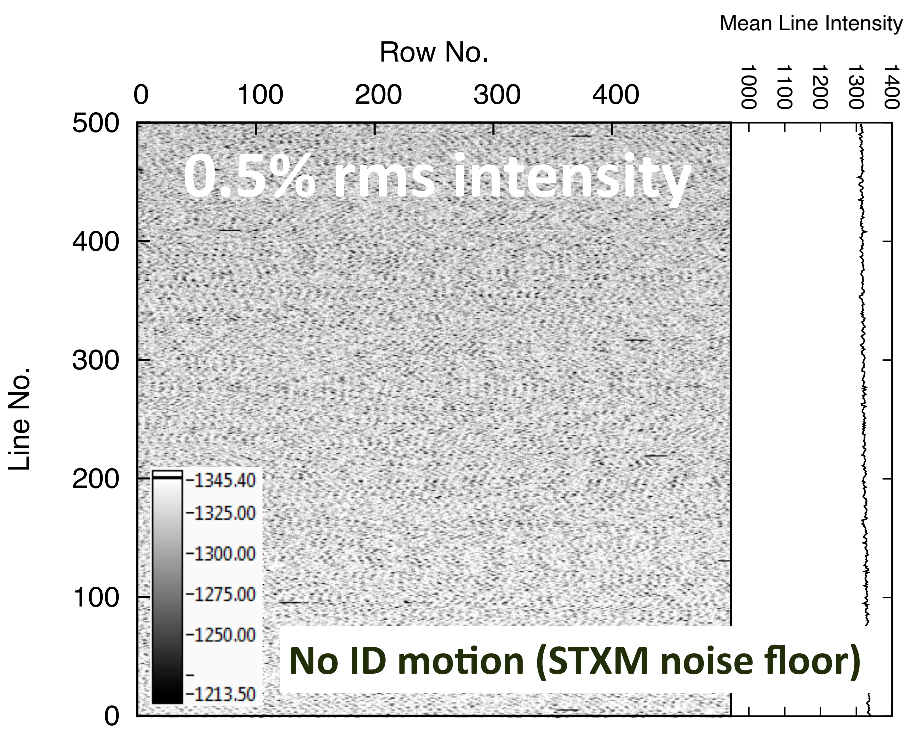
Results: NN-based FF Off vs. On During User Ops



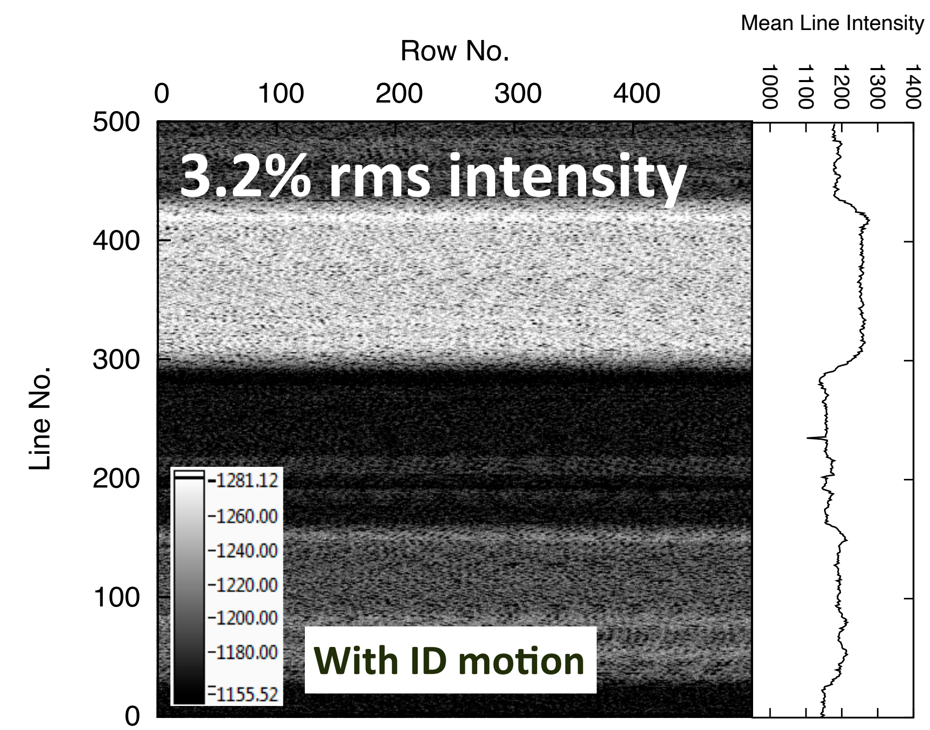
PRL **123**, 194801 (2019)

Stabilization Confirmed at Experiment

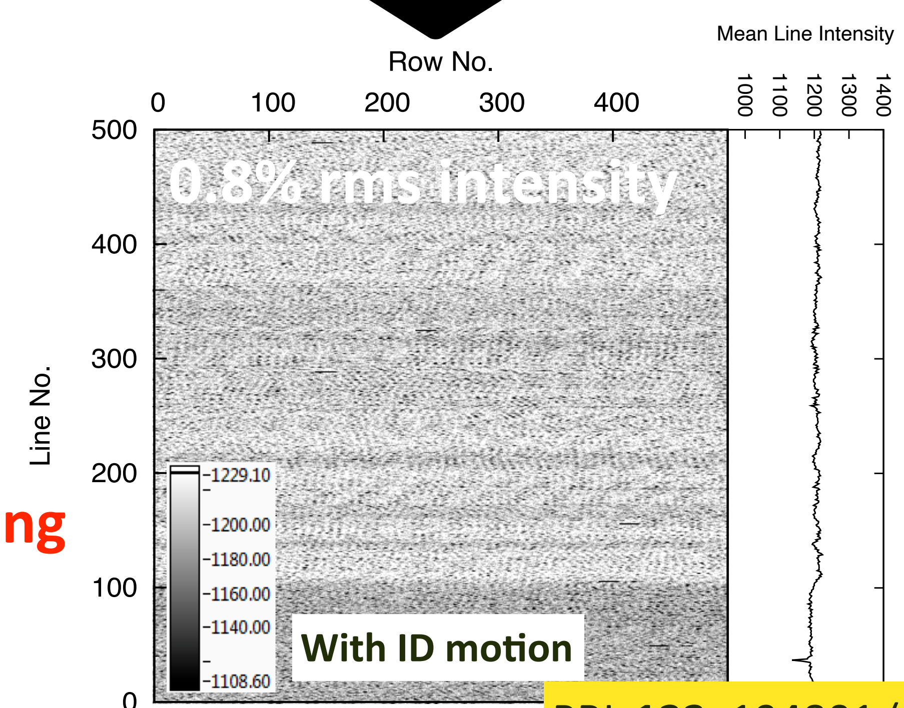
ALS Beamline 5.3.2.2



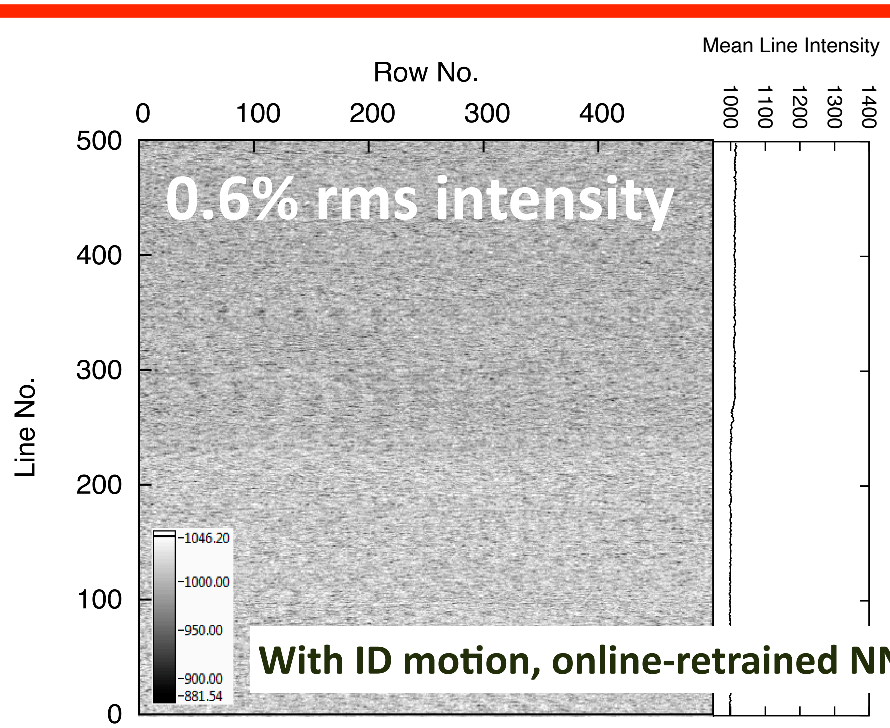
ID Motion



NN-based FF on



Online Retraining



Noise reduced to almost floor level

PRL **123**, 194801 (2019)

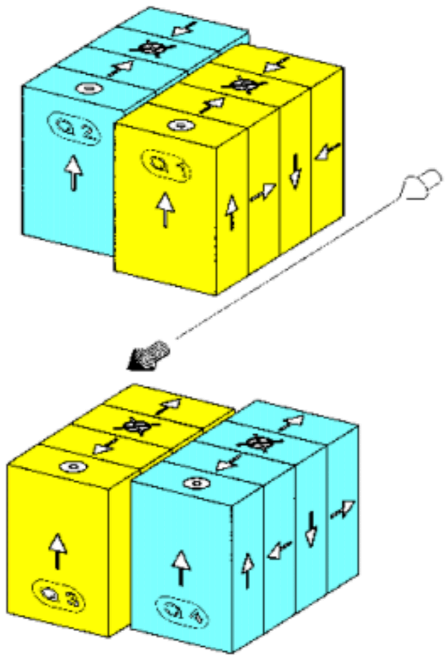
Thank You!

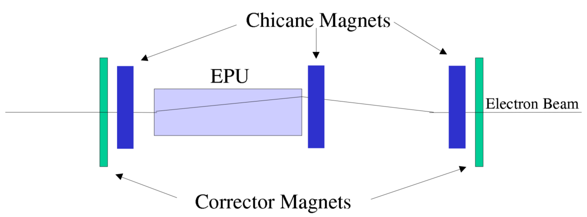
Acknowledgments:
Shuai Liu, Hiroshi Nishimura, Matthew A. Marcus, David Shapiro, Changchun Sun, Nathan Melton, Alex Hexemer, Dani Ushizima, Mike Ehrlichman, Gregg Penn, Thorsten Hellert, Yuping Lu, Erik Wallen, Warren Byrne, Fernando Sannibale, Marco Venturini, Andreas Scholl, Xiaobiao Huang (SSRL)

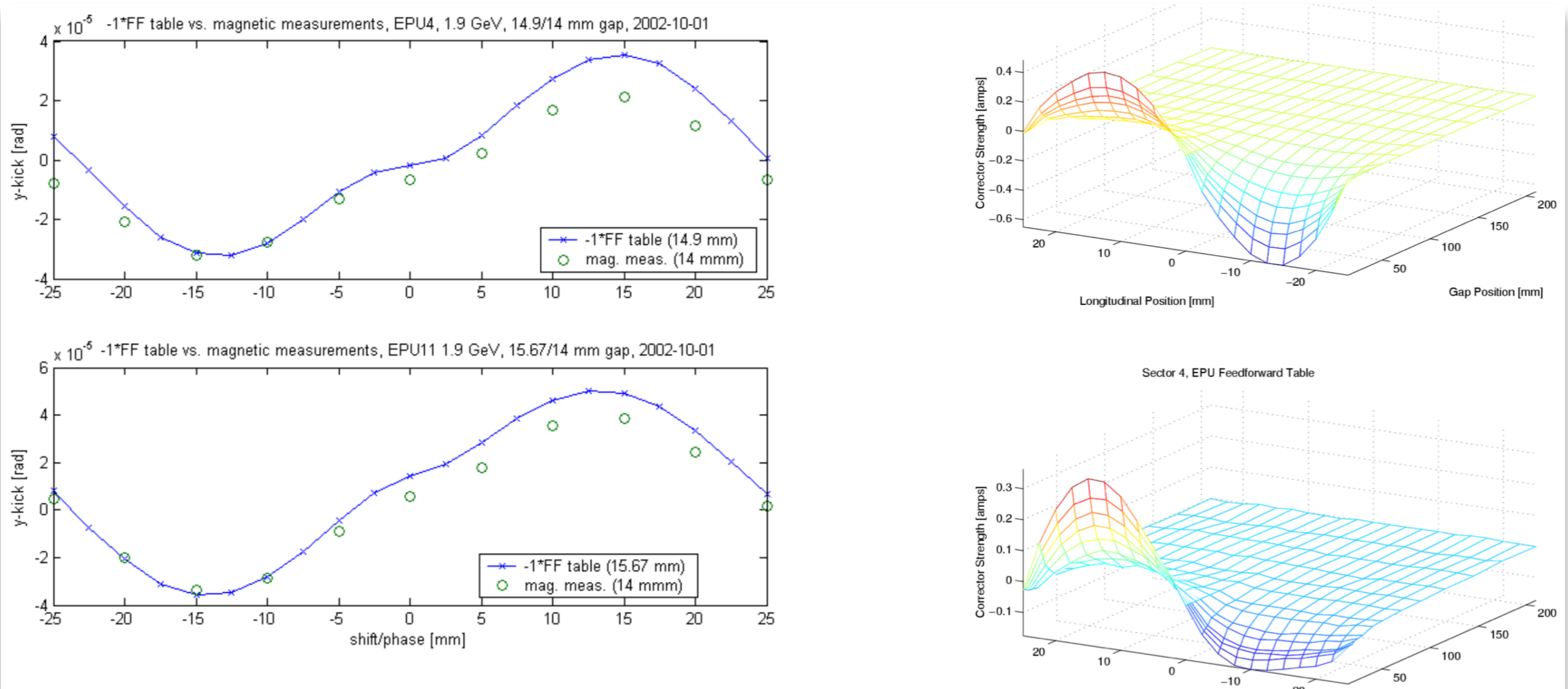
Backup Slides

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)

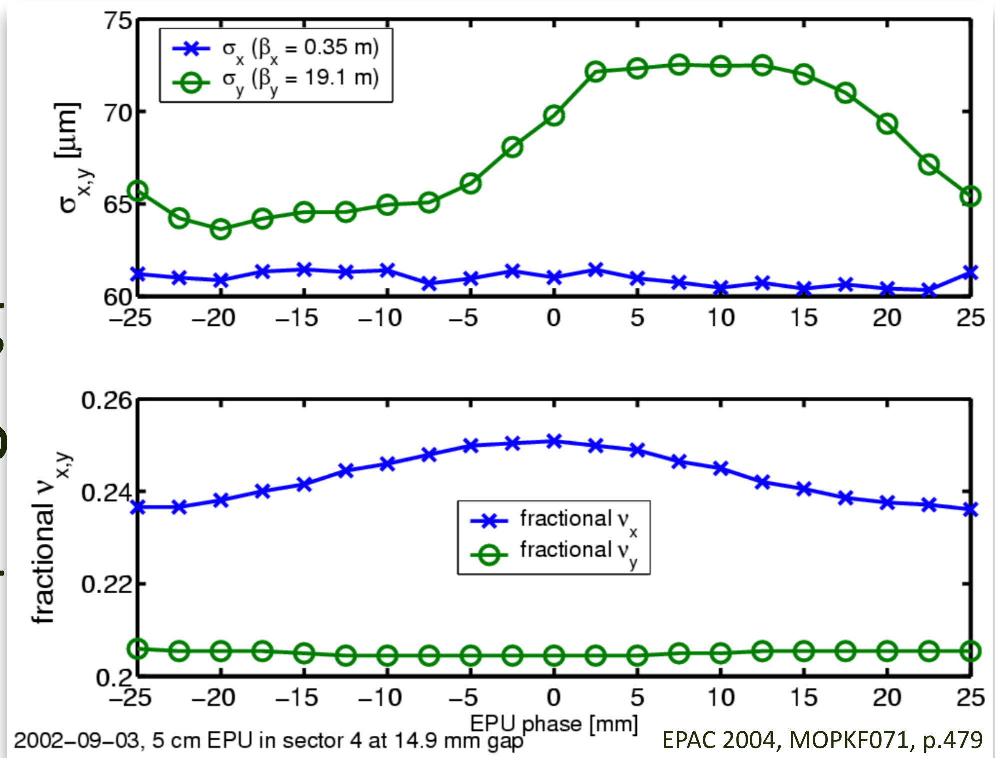






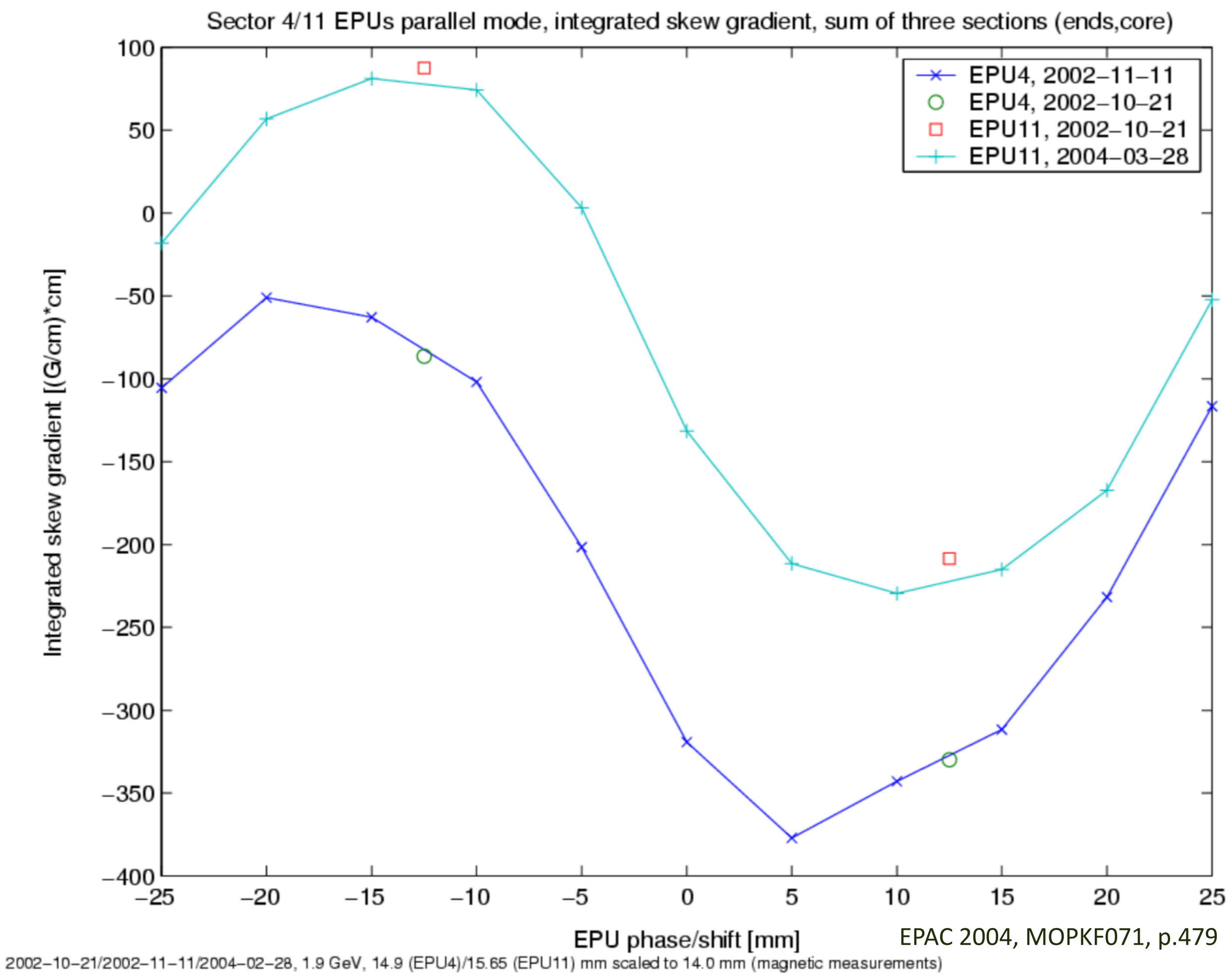
EPAC 2004, MOPKF071, p.479

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



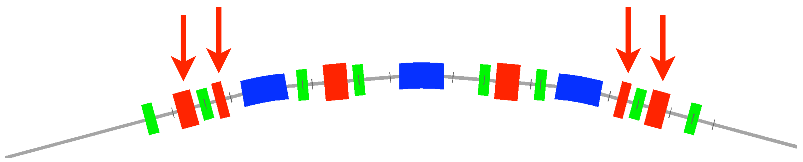
- Orbit distortions
  - caused by on-axis variation of fie
    - corrected by shims (magic fingers)
    - corrected by ring corrector magnet
- Beam size (primarily vertical)
  - caused by variation of ID focusing
    - corrected by local quad trims and g
  - caused by variation of ID-induced coupling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

Errors Caused by IDs & Required Corrections

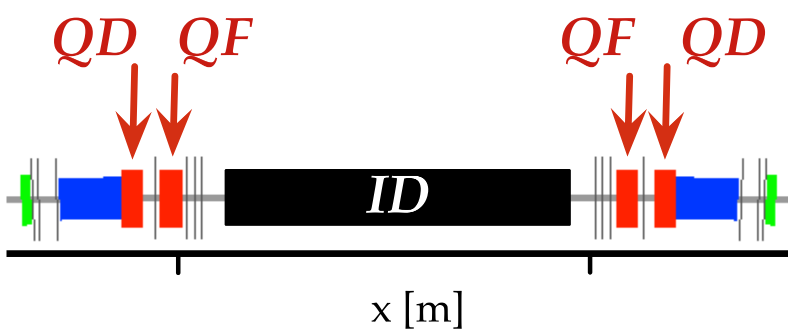
- Orbit distortions
  - caused by on-axis variation of field integrals (with gap or EPU phase)
    - corrected by shims (magic fingers) & local orbit correctors (FF, 200 Hz)
    - corrected by ring corrector magnets (FB, ≈1 Hz SOFB & 1.1 kHz FOFB)
- Beam size (primarily vertical)
  - caused by variation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



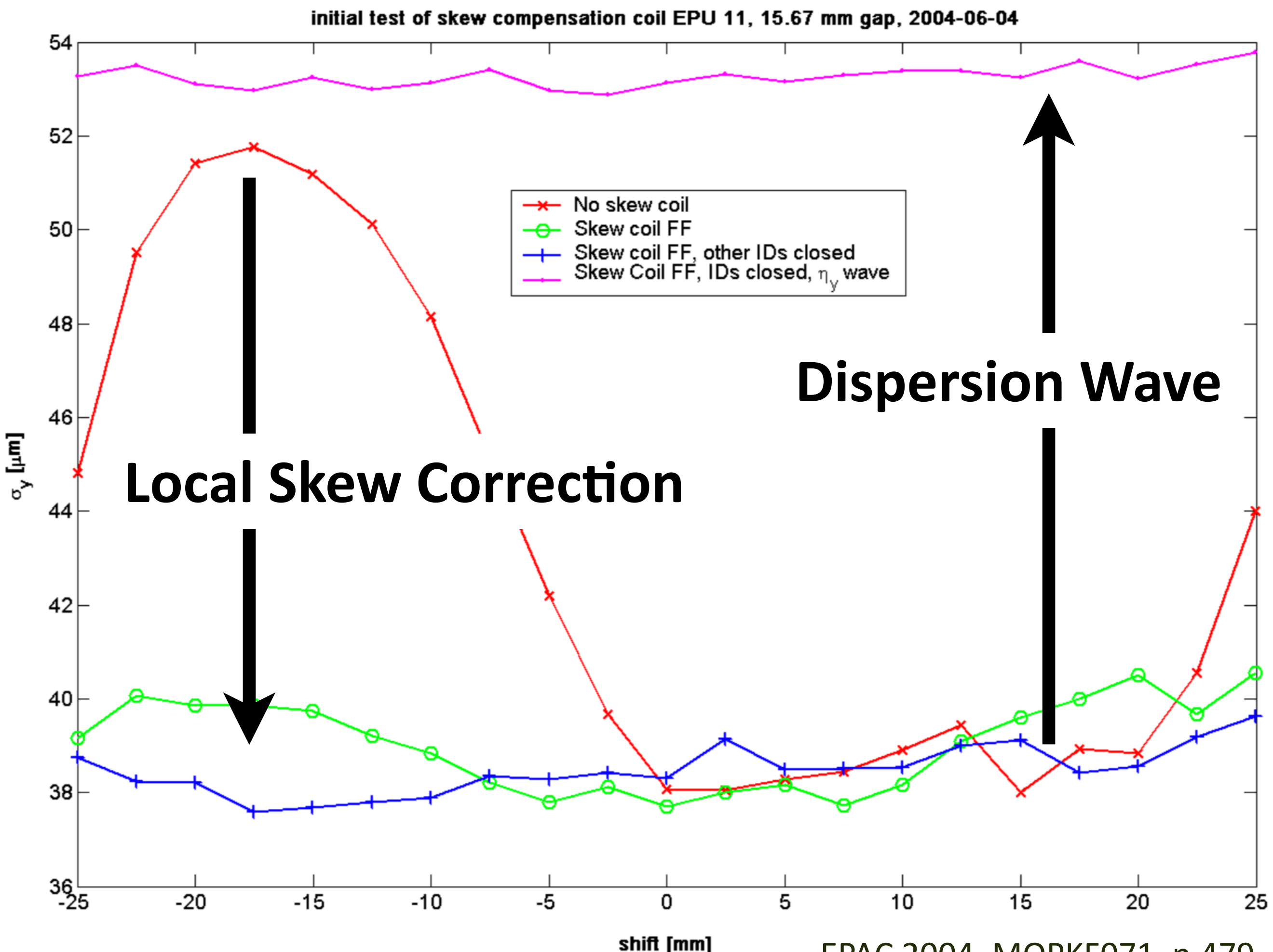
- Local Corrections for EPUs only
  - β_x beat (using 2 QF & 2 QD locally)
    ➔ locally also corrects $\Delta\nu_x$ since $\beta_x \approx 21$ m

Who is 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

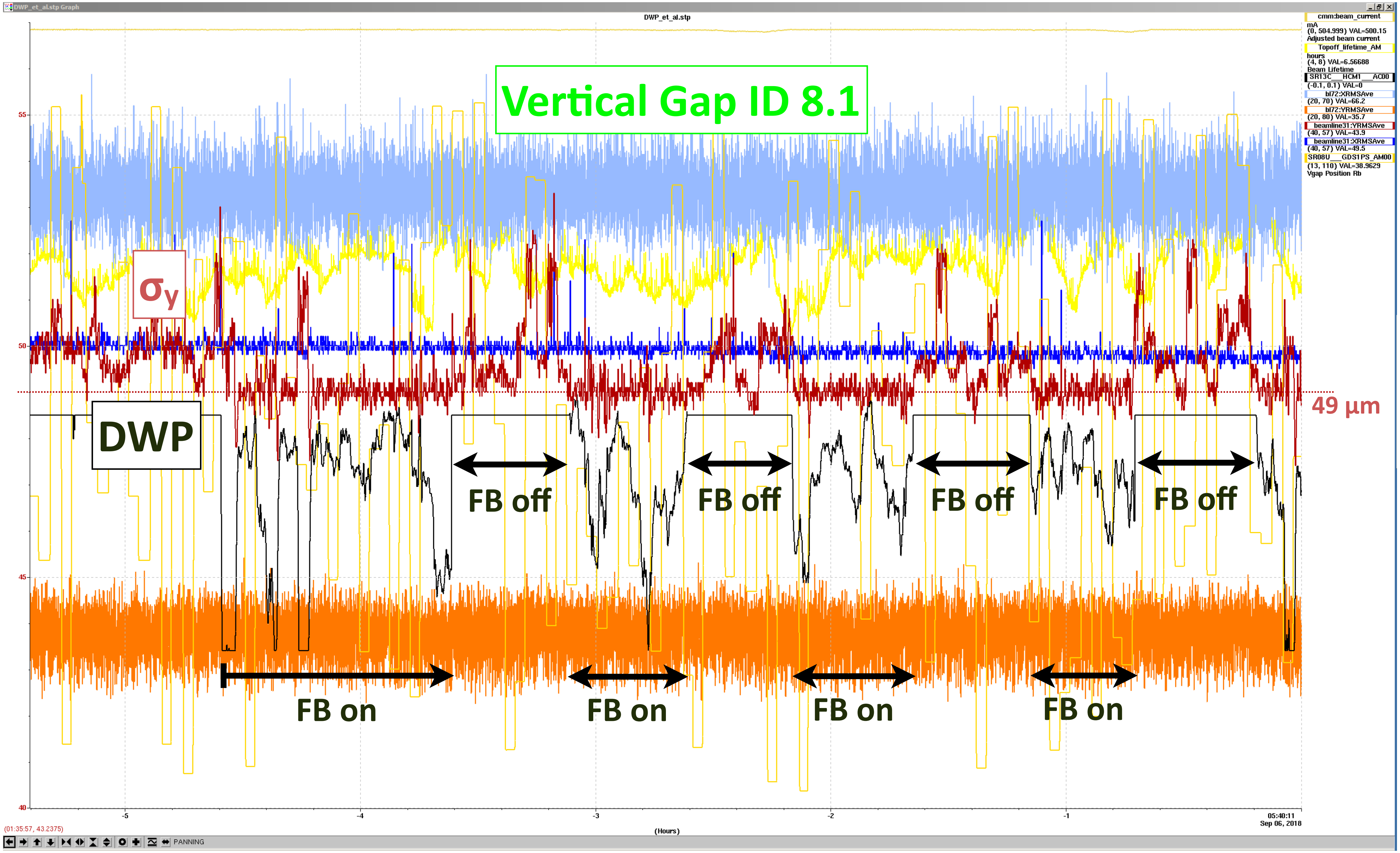
Vertical Dispersion Wave Determines Effective ε_y

- 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

PID FB Loop Adjusting DWP as Function of σ_y



PID FB Loop Adjusting DWP as Function of σ_y (cont.)