

Accurate & Confident Prediction of Electron Beam Longitudinal Properties using Spectral Virtual Diagnostics

Adi Hanuka

SLAC National Laboratory
Stanford University

June 13th 2022

hanukaa@eikontx.com

 [@AccAdkoo](https://twitter.com/AccAdkoo)



Thanks to the wonderful team!



O. Convery
UCSB



C. Emma
SLAC



Z. Huang
SLAC



M. Hogan
SLAC



A. Fisher
SLAC



T. Maxwell
SLAC



B. Jacobson
SLAC

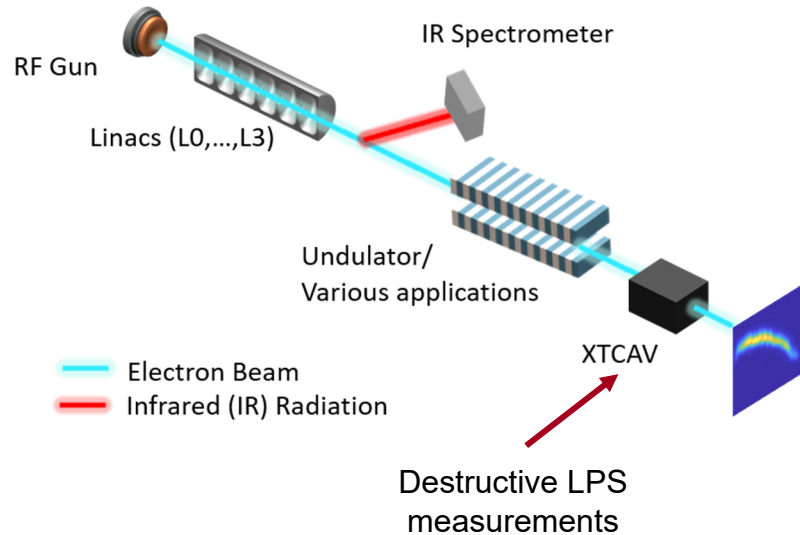


Y. Gal
Oxford

Outline

1. ML-based virtual diagnostics (VD) – Motivation & Background.
2. Spectral virtual diagnostics – 3 case studies:
 - Improved accuracy over scalar VD (LCLS)
 - Shot-to-shot prediction of fine features (LCLS-II)
 - Going beyond current diagnostic resolution (FACET-II)
3. Incorporating uncertainties – Methods & Results.
 - OOD robustness
 - Latent space interpretability
4. Summary

Motivation

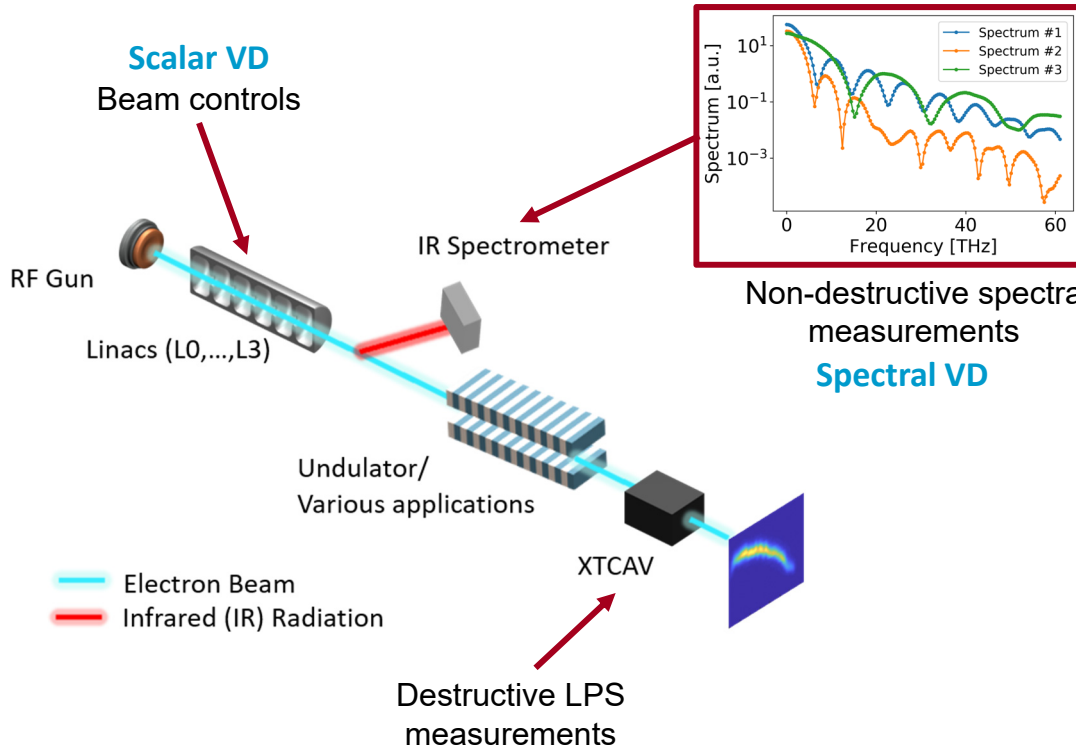


Current diagnostic methods for measuring Longitudinal Phase-space (LPS) are **destructive**.

Machine learning based diagnostics can predict the beam properties on **shot-to-shot** basis **non-destructively** during transport and delivery to experiments.

ML-based Virtual Diagnostics (VD)

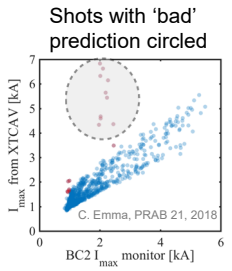
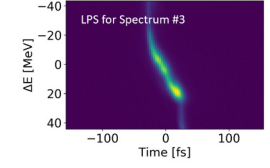
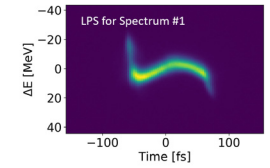
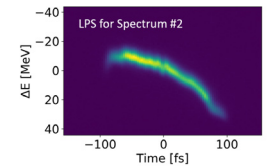
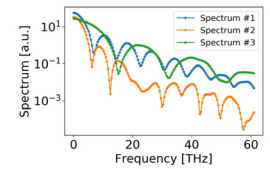
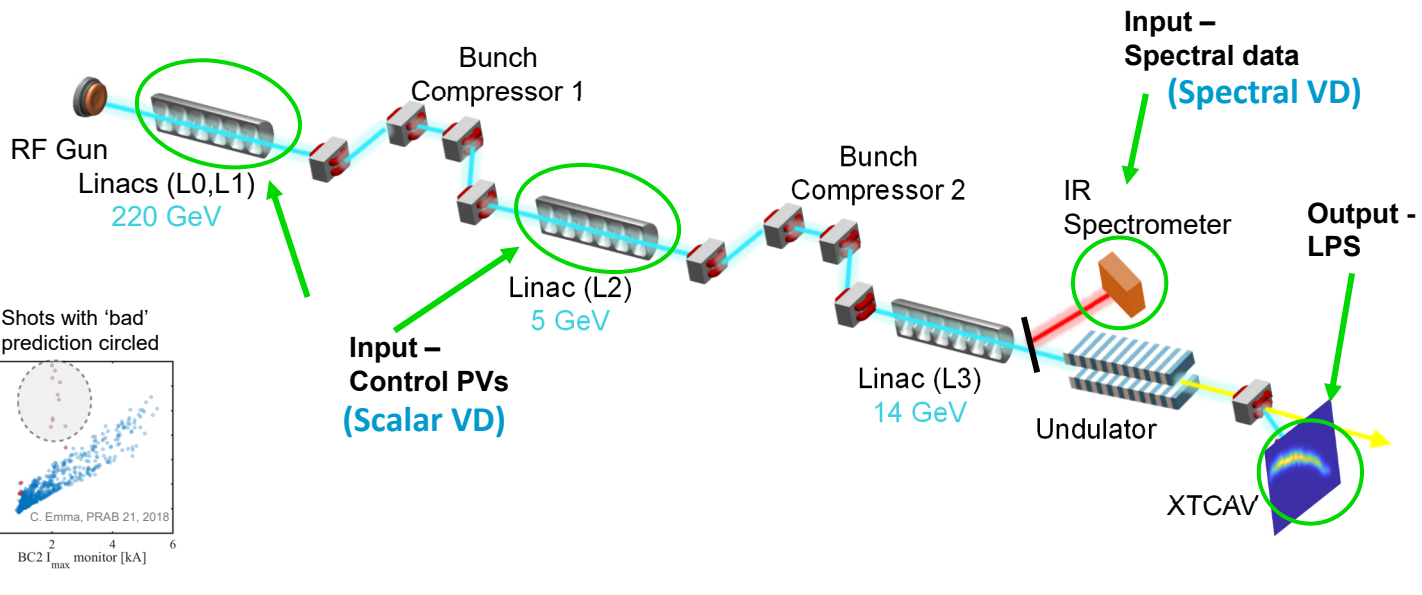
Goal: Get otherwise unavailable (single-shot) information about the beam non-destructively to improve machine characterization, optimization, and data analysis.



- **Supervised**: pairs of input-output
- Once trained, **fast** to execute!
- Train on measured data and/or (slow) high fidelity simulations.

Spectral VD Circumvent the Limitations of Scalar VD

1. Readback scalars are wrong → **GIGO!**
2. Readback scalars are integrated signals → **can not predict** shot-to-shot fluctuation effects like microbunching.
3. Bad predictions can result from large **discrepancy** between diagnostic input (e.g. Bunch Compressor 2 current) and XTCAV current.

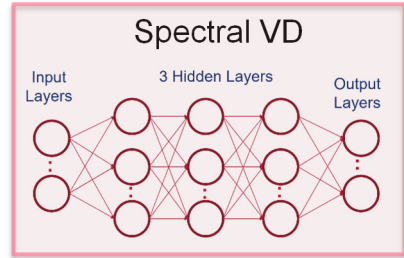
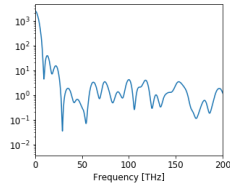


*May be exacerbated in more complicated accelerator operation modes.

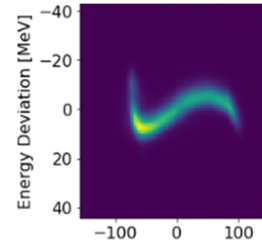
Spectral Virtual Diagnostic (VD)

Neural Network – mapping millions of inputs to similarly numerous outputs.

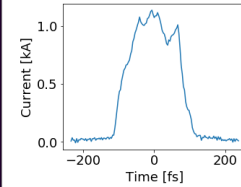
Spectral measurement
(non-destructive)



LPS 2D images



1D Current profile

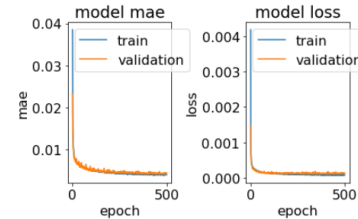


Only train **once!**
Fast prediction of beam

VD Class in Python is easy to use

```
from VD_class import VD
vd = VD(spectrum, Iz)
Iz_predict = vd.vd_trainer(batch_size=64, epochs=500, mc=False, mbi=False)

get_model
fit_model
predict_model
```



loss: 7.725056511245074e-05
Validation loss: 0.0001323133176889696
Test loss: 9.489419417711619e-05
Test accuracy: 0.004323772620409727

Outline

1. ML-based virtual diagnostics (VD) – Motivation & Background.
2. Spectral virtual diagnostics – 3 case studies:
 - Improved accuracy over scalar VD (LCLS)
 - Shot-to-shot prediction of fine features (LCLS-II)
 - Going beyond current diagnostic resolution (FACET-II)
3. Incorporating uncertainties – Methods & Results.
 - OOD robustness
 - Latent space interpretability
4. Summary

Accurate & Confident Predictions - Case Studies

Accuracy would come from designing the neural network architecture & its training.

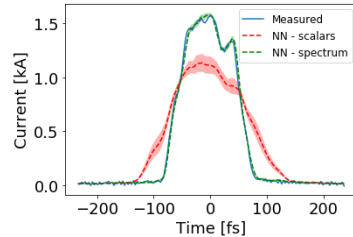
Confidence would come from various methods depending on the case.

Accuracy

Confidence

LCLS

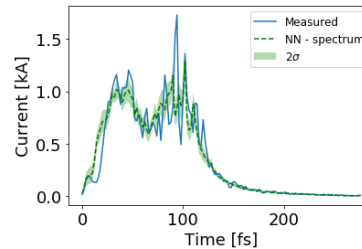
- Experimental
- 1D/2D outputs



Comparing Scalar VD vs Spectral VD

LCLS-II

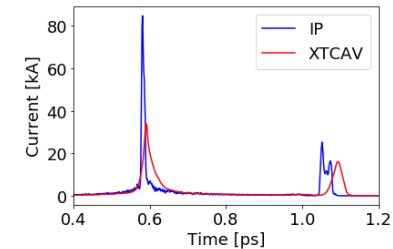
- Microbunching
- Elegant SC SXR simulation



Prediction uncertainty from ensemble

FACET-II

- 2-bunch mode
- Lucretia simulation



Correlating prediction with spectral intensity

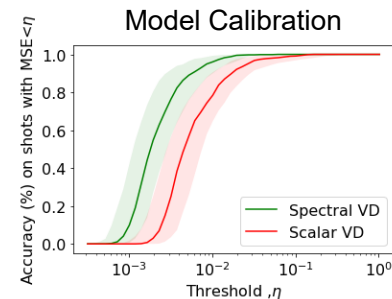
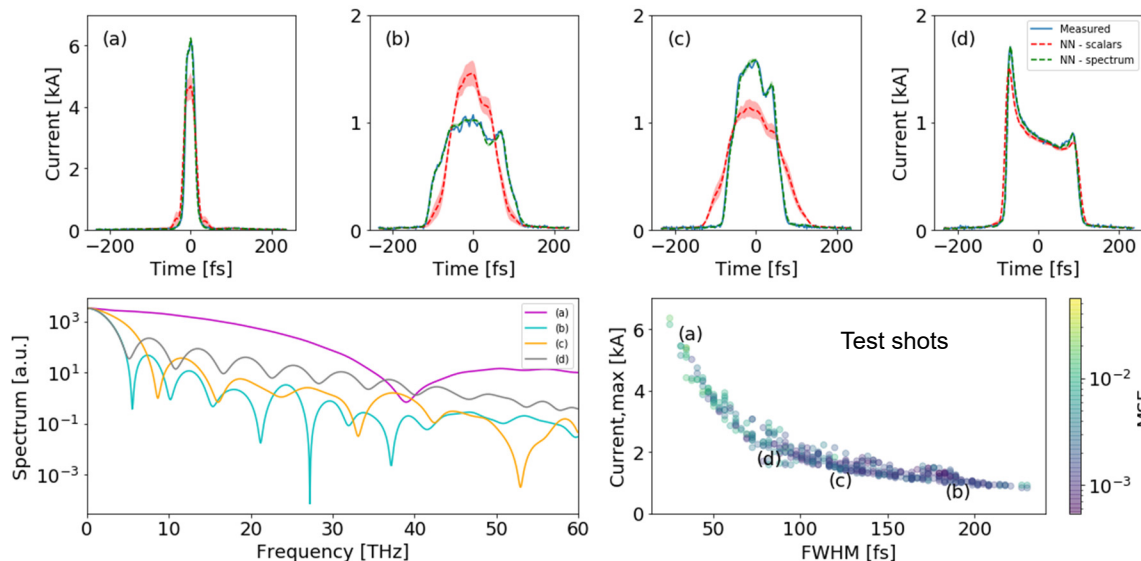
Improved accuracy over scalar VD (LCLS)

- Train ~4000 examples, Test ~600 examples.
- Scalar VD: optimized NN architecture consistently improved the results by 15%.
- Spectral VD has lower MSE than scalar VD.

LCLS Experiment:

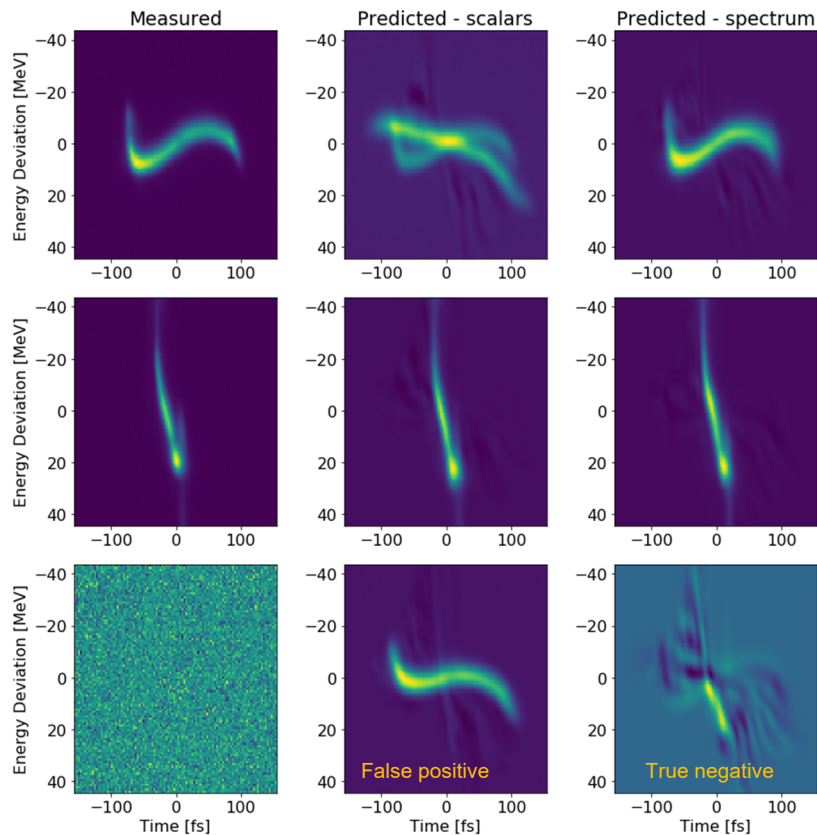
Machine parameters scanned:
L1s phase from -21 to -27.8 deg
BC2 peak current from 1 to 7 kA

Inputs to Scalar VD:
L1x voltage & phase,
L1x voltage, BC1 and BC2 current



Spectral VD better predicts LPS images (LCLS)

- Improved accuracy of the spectral VD*.
- Increased confidence from multiple diagnostic predictions.

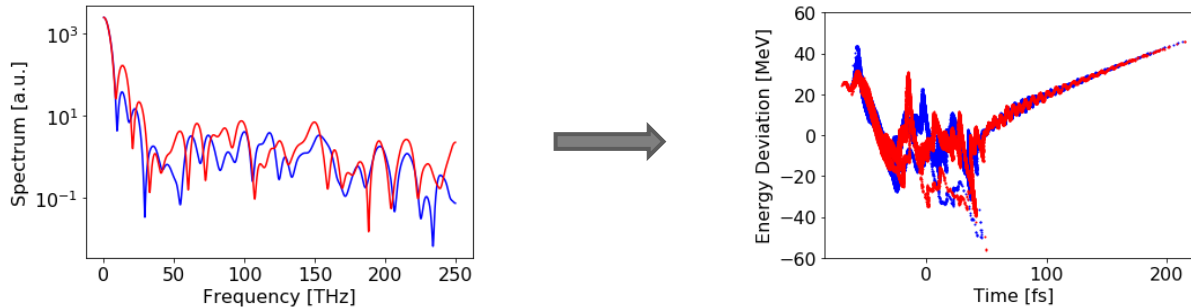


* $MSE=0.054, 0.079$; $SSIM=0.97, 0.96$ for *spectral*, *scalars*

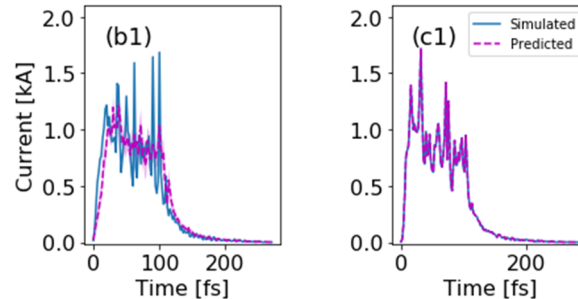
SSIM=structural similarity index measure [0,1]

Shot-to-shot prediction of fine features (LCLS-II)

- LCLS-II has a 1km bypass line between the linac and the undulator → MBI is especially pronounced → seed the growth of unwanted radiation modes in the FEL.
- Two Simulations of LCLS-II SC SXR – same input, different output → Only Spectral VD can predict!



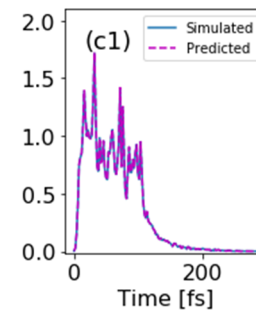
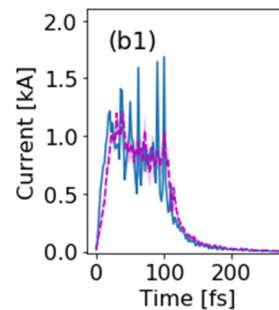
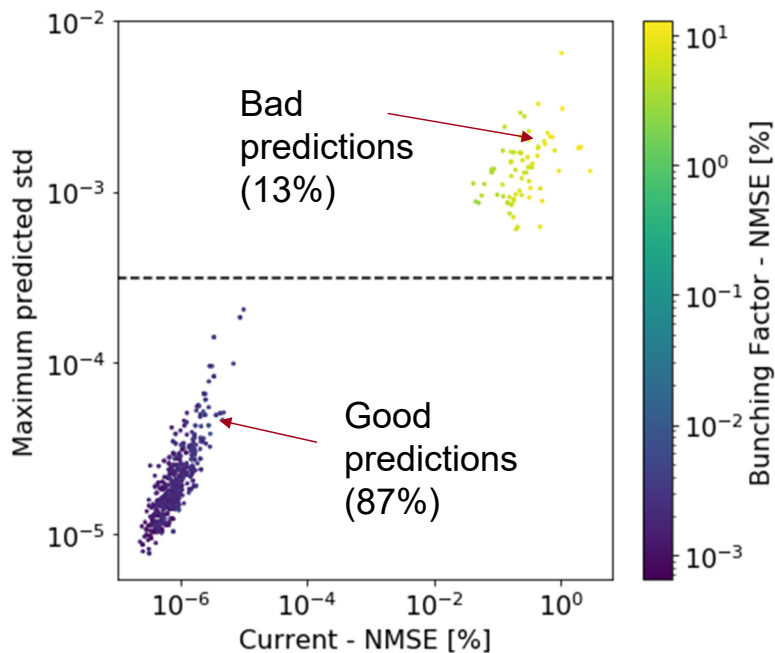
- Ensemble of NN trained on 4000 Elegant simulations → Predict current profile (NMSE=1.1%)



$$\text{NMSE}(y, \hat{y}) = \text{MSE} / \sum_{i=0}^{N-1} y_i^2$$

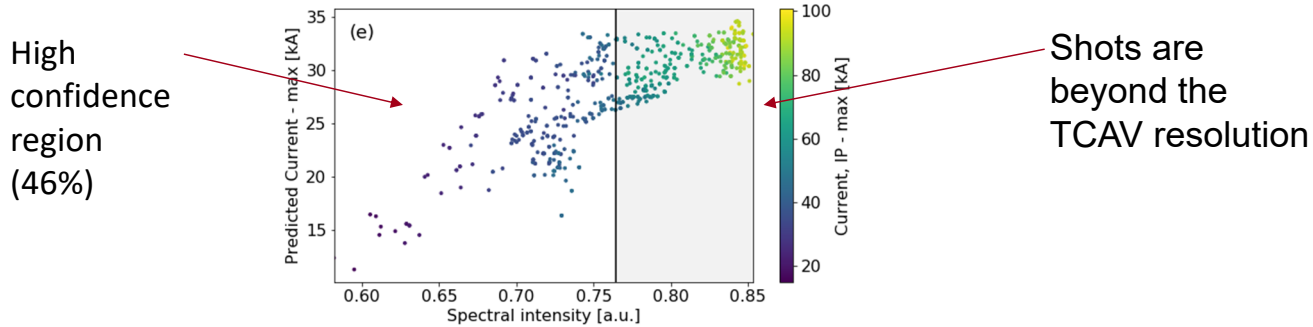
Increasing prediction's confidence (LCLS-II)

- Correlate max std and mean MSE from ensemble of NN with random initializations.
- Once deployed on the machine, the ground truth will not be available for calculating MSE → we can flag bad shots by setting a threshold (dashed line)



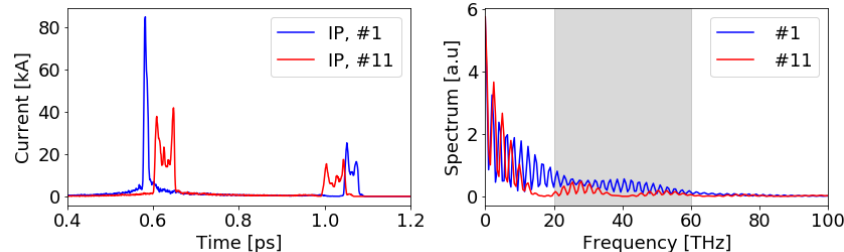
Going beyond current diagnostic resolution (*FACET-II*)

- Spectral VD resolves features that are beyond the TCAV limited resolution.



- Optimize the frequency band to distinguish between high peak current (>35 kA) shots to lower ones.

3000 LUCRETIA simulations

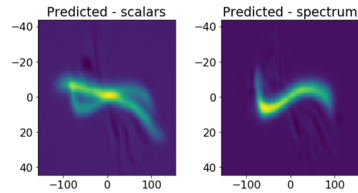


Accurate & Confident Predictions - Summary

LCLS

- Experimental
- 1D/2D outputs

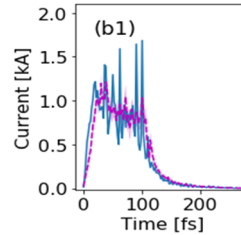
Customized NN architecture.



LCLS-II

- Microbunching
- Elegant SC SXR simulation

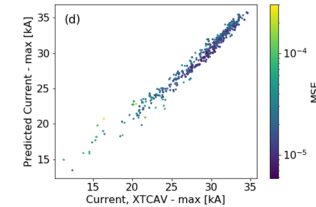
Wider NN trained longer



FACET-II

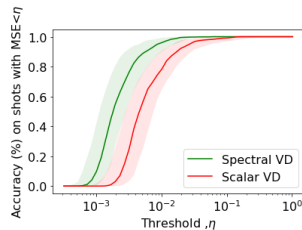
- 2-bunch mode
- Lucretia simulation

Customized NN architecture.

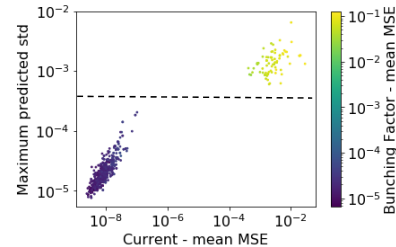


Accuracy

Comparing Scalar VD vs Spectral VD

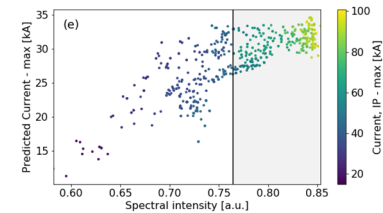


Prediction std from ensemble



Confidence

Correlating prediction with spectral intensity



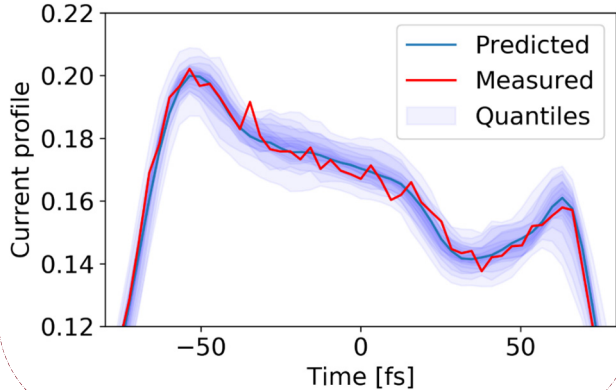
Outline

1. ML-based virtual diagnostics (VD) – Motivation & Background.
2. Spectral virtual diagnostics – 3 case studies:
 - Improved accuracy over scalar VD (LCLS)
 - Shot-to-shot prediction of fine features (LCLS-II)
 - Going beyond current diagnostic resolution (FACET-II)
3. Incorporating uncertainties – Methods & Results.
 - OOD robustness
 - Latent space interpretability
4. Summary

Uncertainty in particle accelerators

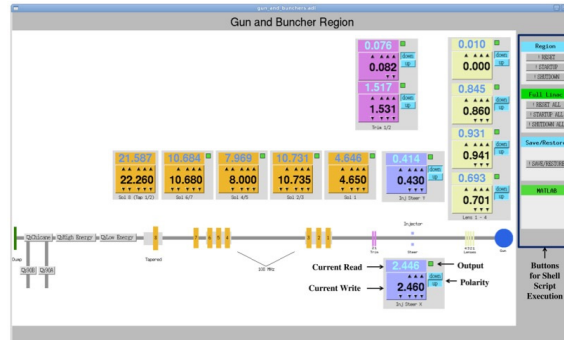
Analysis

- Provide accurate beam information
- UQ captures all possible outputs the machine can produce in a given state



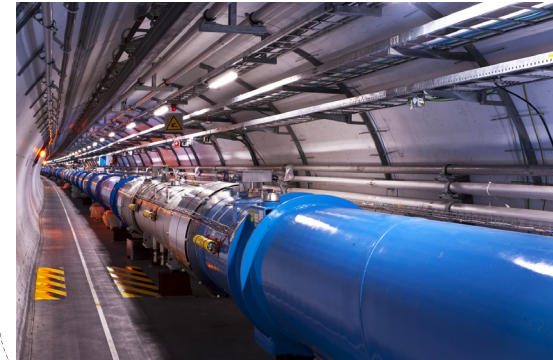
Control

- Tuning loops are often controlled by VD
- Avoiding uncertain state systems can save valuable tuning time

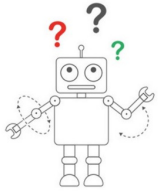


Safety

Avoiding unsafe control actions can lead to costly mistakes (Damage to walls, vacuum state, components)

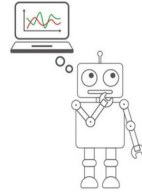


Incorporating Uncertainties – know what we don't know

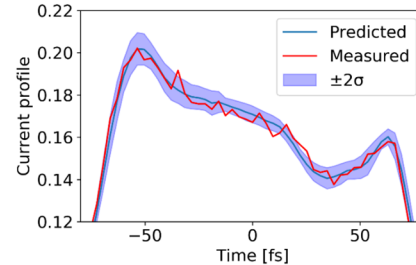


- Neural network is not aware of what it does not know!

- New shots might be out of trained distribution → prediction is unreliable.



Need estimates of std along with prediction mean.



Epistemic uncertainty:

- What the model doesn't know.
- Can be reduced by adding more data.

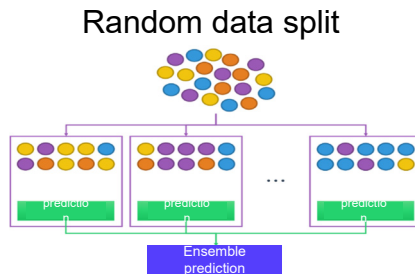
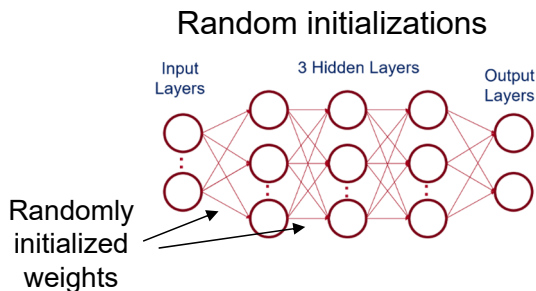
Aleatoric uncertainty:

- What we can't understand from the data.
- Arises from experimental error or inherent measurement noise.

Incorporating Uncertainties – Methods

Goal: Quantify how reliable the mean prediction is.

1. **Ensemble methods**: a collection of neural networks



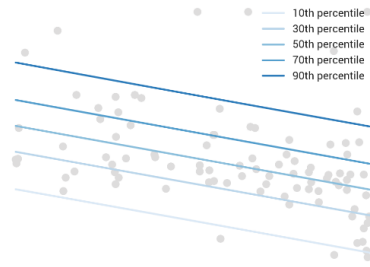
Epistemic uncertainty

2. **Quantile regression**: separate NN to predict each quantile

$$\xi_i = y_i - f(\mathbf{x}_i)$$

$$\mathcal{L}(\xi_i|\alpha) = \begin{cases} \alpha \xi_i & \text{if } \xi_i \geq 0, \\ (\alpha - 1)\xi_i & \text{if } \xi_i < 0. \end{cases}$$

$$\mathcal{L}(\mathbf{y}, \mathbf{f}|\alpha) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i - f(\mathbf{x}_i)|\alpha)$$



Aleatoric uncertainty

Incorporating Uncertainties – Results

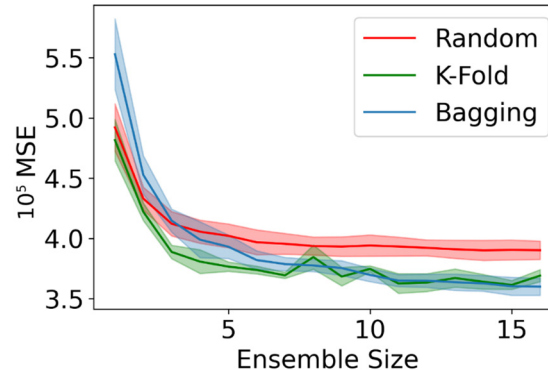
VD:
200,100,50
RELU
Batch 32
500 epochs

- Before exploring the quantified uncertainty, we found which ensemble method yielded the best MSE on the mean prediction.

Random initializations

K-fold = random subsets of data (w/o replacement)

Bagging = random initialization + random “bags” of data (w/ replacement)



Mean Ensemble Prediction:

$$\vec{L}_{\text{predicted}} = M^{-1} \sum_{m=1}^M \vec{l}_{\text{predicted},m}$$

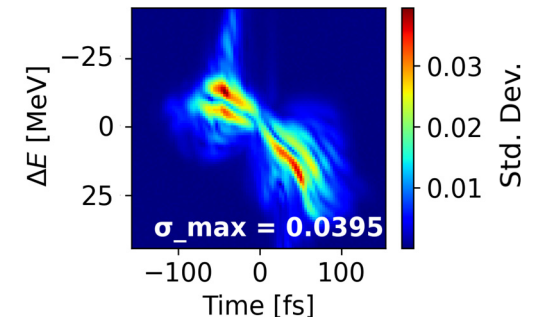
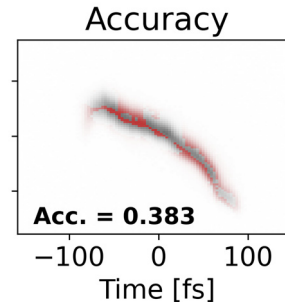
Predictive Standard Deviation:

$$\sqrt{M^{-1} \sum_{m=1}^M (\vec{l}_{\text{predicted},m} - \vec{L}_{\text{predicted}})^2}$$

- Accuracy evaluates how much of the measured value falls within a $\pm 2\sigma$ bounds

$$\text{Accuracy} = \frac{\sum_{t,e=1}^{T,E} \alpha_{t,e} \cdot L_{\text{measured},t,e}^2}{\sum_{t,e=1}^{T,E} L_{\text{measured},t,e}^2}$$

$$\alpha_t = 1 \text{ if } I_{\text{lower},t} < I_{\text{measured},t} < I_{\text{upper},t}$$

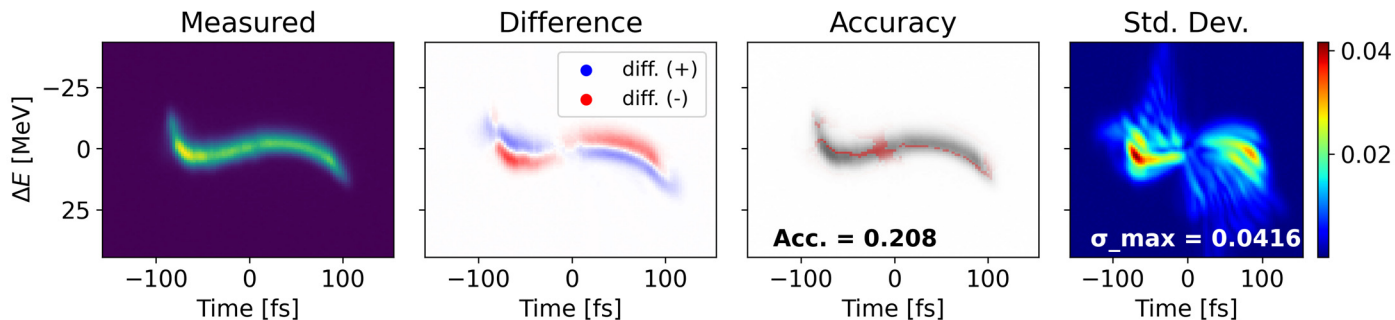


Common prediction errors with LPS images

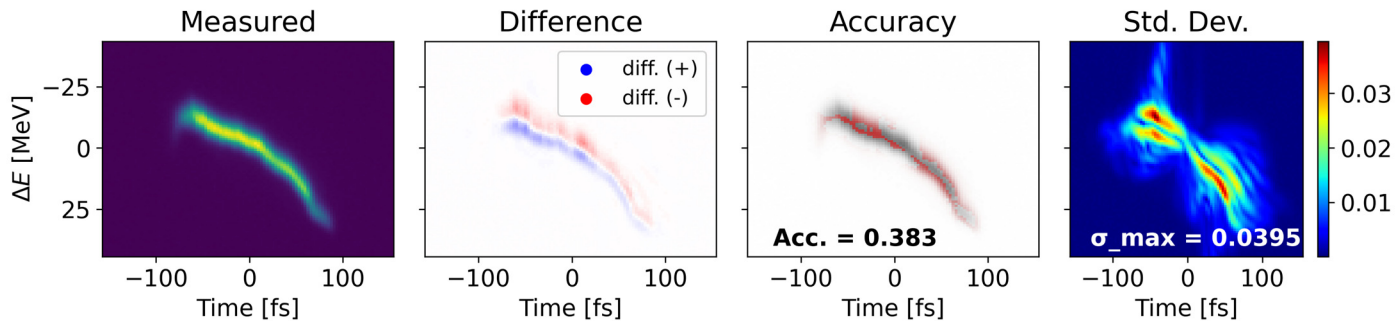
$$\text{Accuracy} = \frac{\sum_{t,e=1}^{T,E} \alpha_{t,e} \cdot L^2_{\text{measured},t,e}}{\sum_{t,e=1}^{T,E} L^2_{\text{measured},t,e}}$$

$\alpha_t = 1$ if $I_{\text{lower},t} < I_{\text{measured},t} < I_{\text{upper},t}$

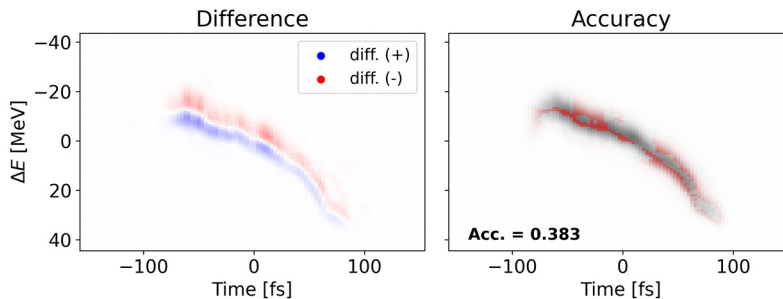
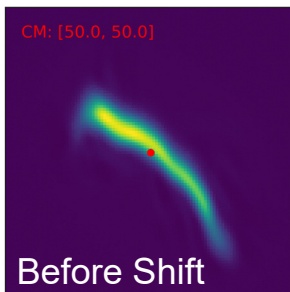
Shape error: the prediction is of the wrong shape



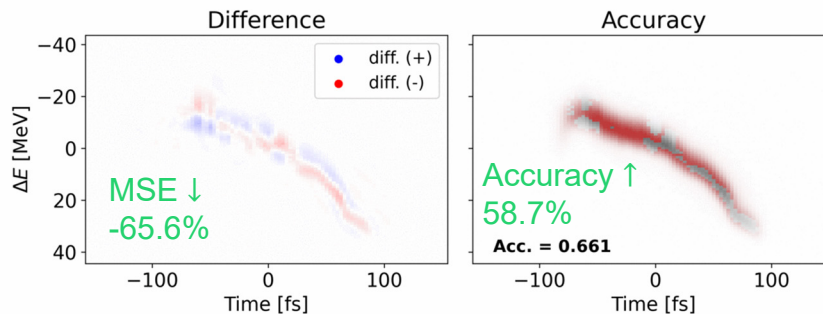
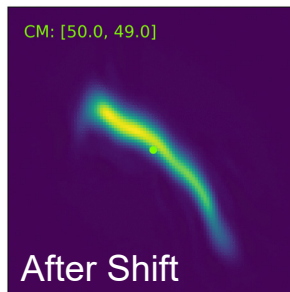
Translational error: the prediction is in the wrong place



Alleviating Translational Error



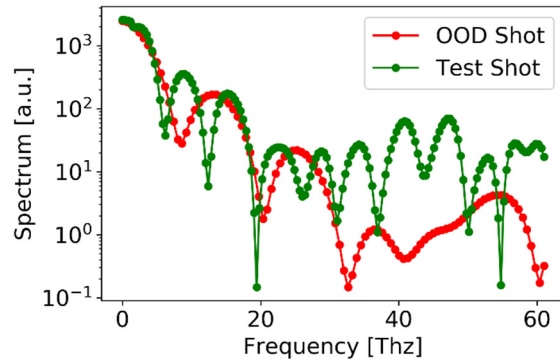
Center of Mass Correction (Prediction \rightarrow Truth) during training



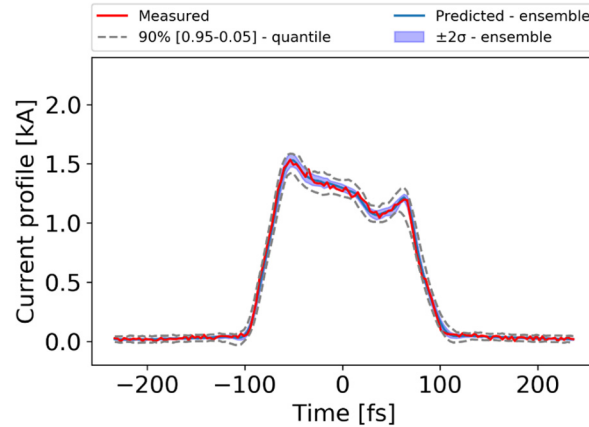
Out-of-distribution (OOD) Robustness

- Evaluate the robustness of the model by predicting an out-of-distribution (OOD) example, e.g. a change in the machine setting.
- Out-of-Distribution \rightarrow Higher Uncertainty

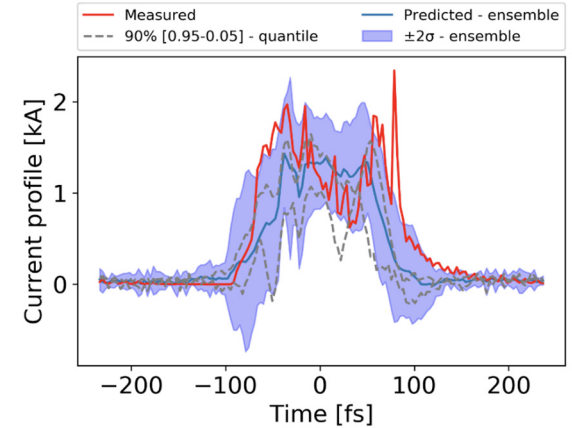
Spectral data



Test shot within the trained distribution

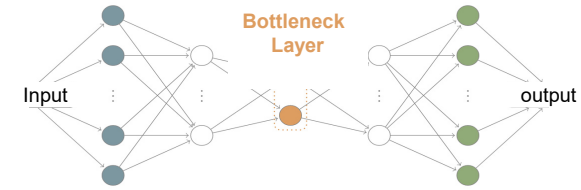


Out-of-distribution



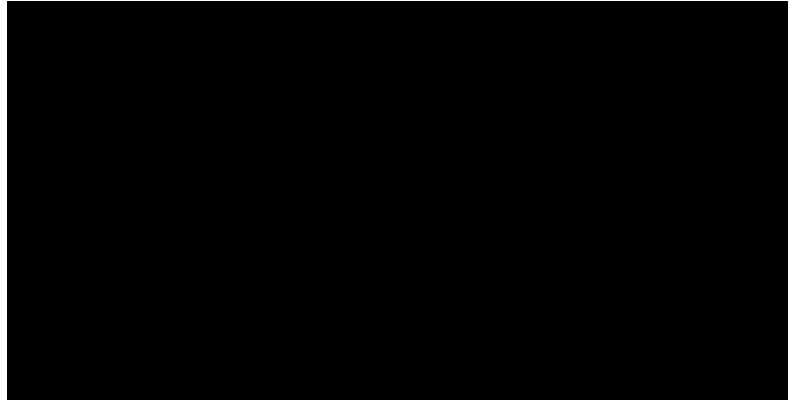
Latent space interpretability

Bottleneck Layer NN architecture followed by t-SNE Clustering



Latent space of size 64 for each of the 810 test shots is reduced to 2D using t-SNE.

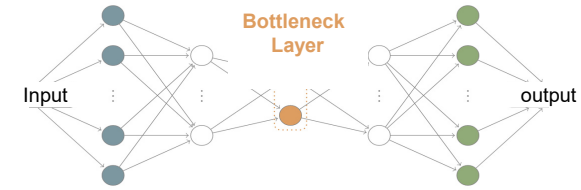
Shots are grouped by shape and maximum current in the latent space.



T-SNE:

- perplexity = 30
- learning rate = 200
- 1000 iterations
- embedding vector length 64

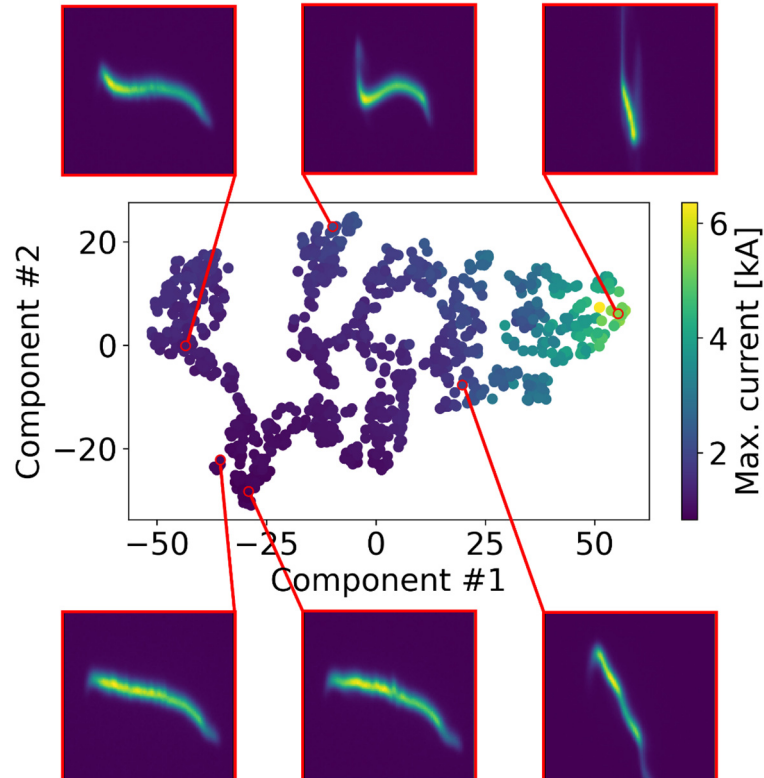
Latent space interpretability



Bottleneck Layer NN architecture followed by t-SNE Clustering

Latent space of size 64 for each of the 810 test shots is reduced to 2D using t-SNE.

Shots are grouped by shape and maximum current in the latent space.

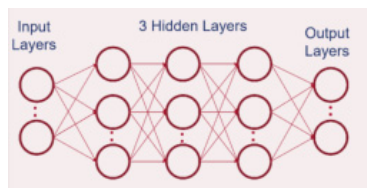


T-SNE:

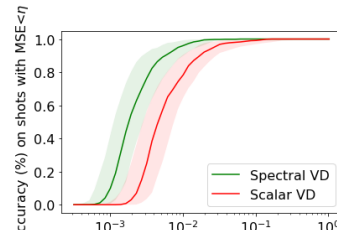
- perplexity = 30
- learning rate = 200
- 1000 iterations
- embedding vector length 64

Summary

Non-destructive, shot-to-shot of bunch diagnostic

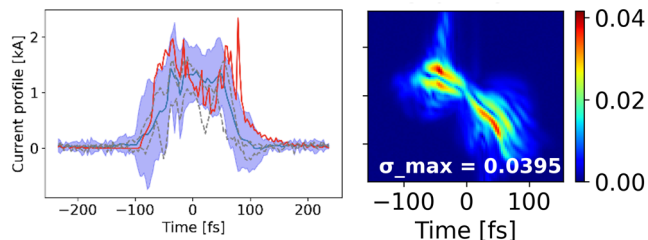


Spectral VD increases confidence & improve accuracy

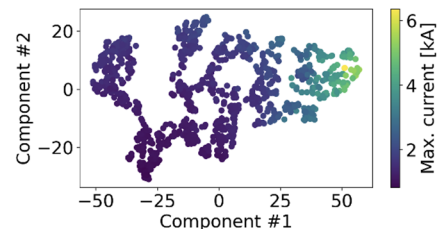


ML-based virtual diagnostic with quantified uncertainties for single shot prediction will provide additional information for analysis, tuning and control.

Quantify prediction robustness (OOD)



Bottleneck architecture to learn an interpretable latent space

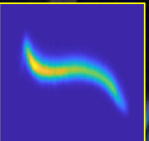


Thank You!

This work was supported by the Department of Energy,
Laboratory Directed Research and Development program
at SLAC National Accelerator Laboratory, under contract
DE-AC02-76SF00515.

<https://journals.aps.org/prab/abstract/10.1103/PhysRevAccelBeams.24.074602>

<https://www.nature.com/articles/s41598-021-82473-0>



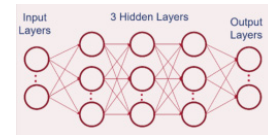
deepart.io



Summary of Virtual Diagnostics

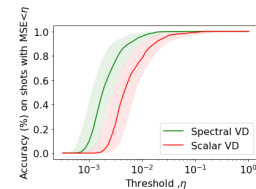
Non-destructive, shot-to-shot of bunch diagnostic during transport and delivery to experiments.

- **Fast & online** – doesn't require convoluted data processing.
- Fill in **missing** information – high peak current, repetition rate, etc.
- Understand **exotic** configs – by combining ML model with simulation.
- **Reverse** engineering of machine settings for a pre-defined current profile.



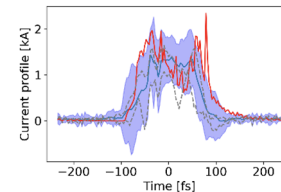
Spectral VD:

- **Increase** confidence – flag bad shots by cross check with scalars VD.
- **Improved** accuracy over scalars VD.
- In some cases is the **only** option! (e.g. microbunching)



Quantify prediction sensitivity:

- Flag **bad** predictions.
- Flag a **change** in the machine – out-of-distribution prediction.



ML-based virtual diagnostic for single shot prediction will provide additional information for users, and a signal for LPS feedback, tuning and control.

Bottleneck architecture to learn important features

Bottleneck architectures can reduce over-fitting by decreasing the NN complexity.

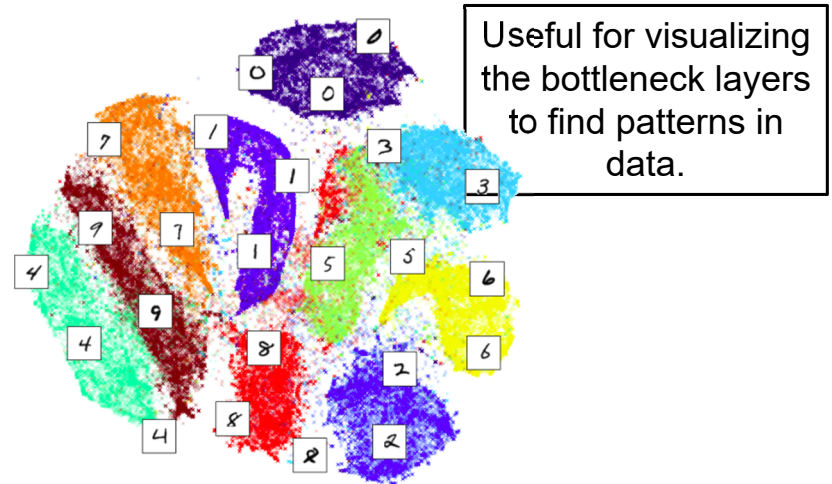
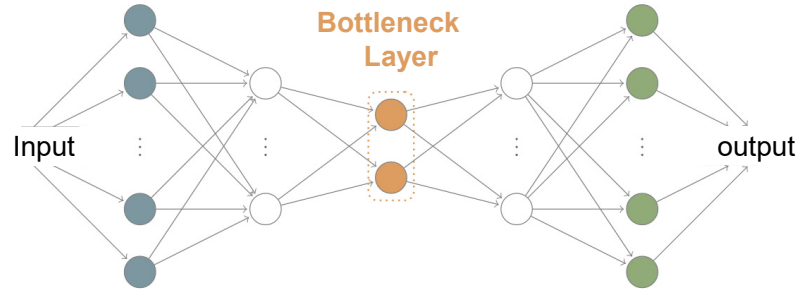
Condense information about the I/O relation.

t-SNE for dimensionality reduction

T-distributed Stochastic Neighbor Embedding (t-SNE) to visualize high-dimensional space in a 2D-3D space.

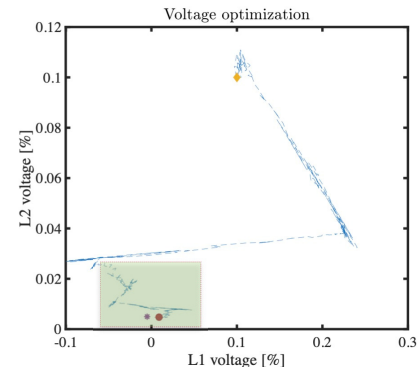
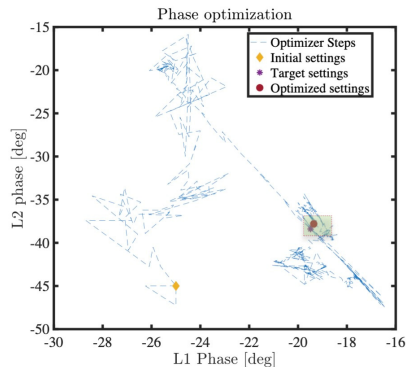
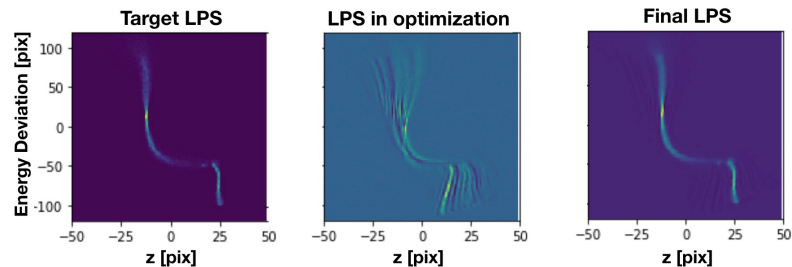
MNIST images visualized in two dimensions.
Colors indicate the digit of each image.

<https://bigsnarf.wordpress.com/2016/11/17/t-sne-attack-data/>

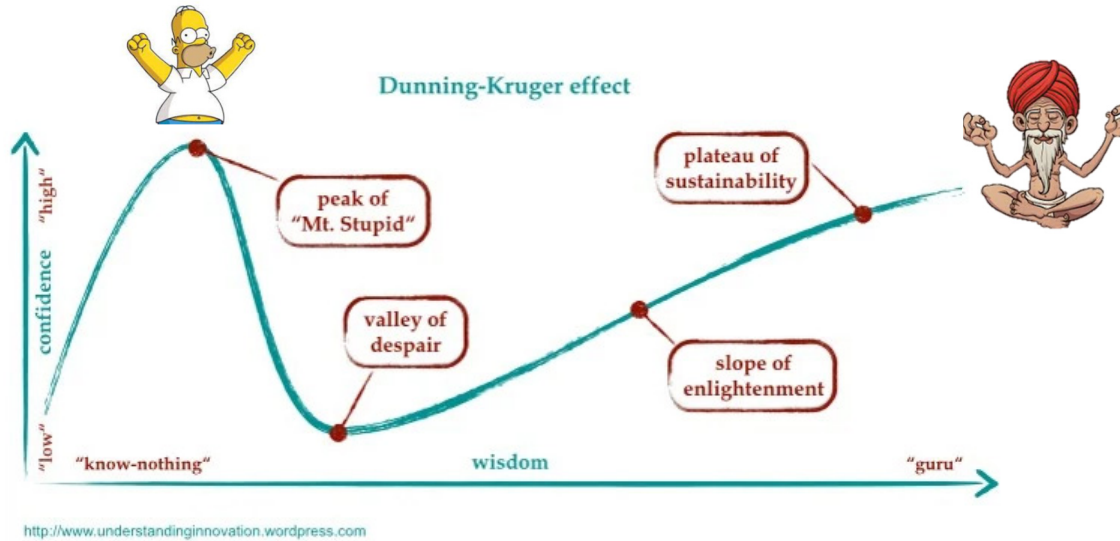


Optimization using LPS virtual diagnostics for two-bunch at FACET-II

- ML prediction of LPS used with conventional optimizer to tune L1-2 phases/voltages for desired LPS.
- Initial settings outside training set of ML model.
- Model shows ability to interpolate within training data.



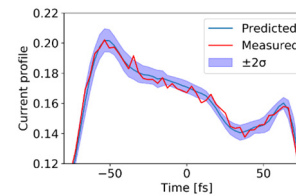
Incorporating Uncertainties – know what we don't know



- Neural network is not aware of what it does not know!
- New shots might be out of trained distribution → prediction is unreliable.

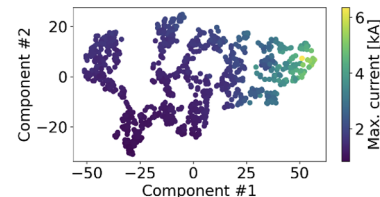
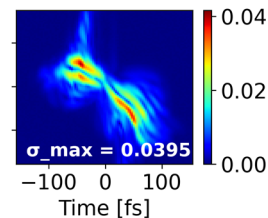
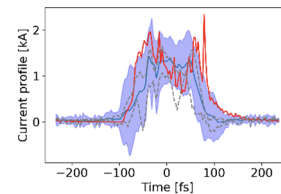
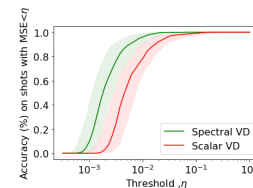
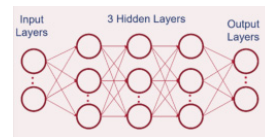


Need estimates of std along with prediction mean.



Summary

- Non-destructive, shot-to-shot of bunch diagnostic
- Spectral VD increases confidence & improve accuracy
- Quantify prediction sensitivity (OOD)
- Bottleneck Architecture to learn latent space



ML-based virtual diagnostic for single shot prediction will provide additional information for analysis, tuning and control.