Machine Learning for an X-ray Laser

Oct. 21, 2018

D. Ratner, E. Cropp, J. Duris, A. Edelen, C. Emma, K. Kabra, D. Kennedy, T. J. Lane, S. Li, T. Maxwell, P. Musumeci, X. Ren, J. Wu, X. Zhang

SLAC National Accelerator Laboratory

U.S. DEPARTMENT OF ENERGY
Office of Science

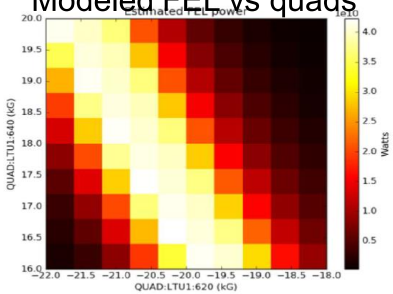
SLAC NATIONAL ACCELERATOR LABORATORY

Recent Machine Learning Highlights at LCLS

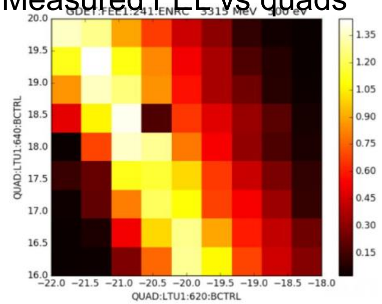
Optimization

- Model independent (simplex, gradient, RCDS, ES, etc.)
- Reinforcement learning
- Bayesian optimization

Modeled FEL vs quads

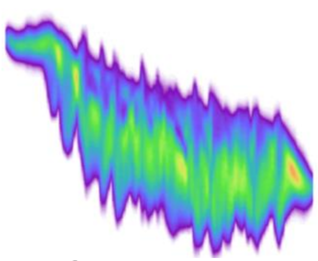


Measured FEL vs quads

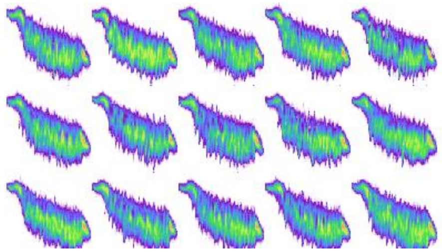


Surrogate models

- Generative adversarial nets



Genesis

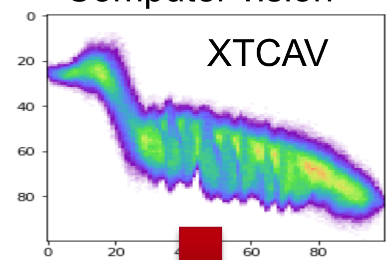


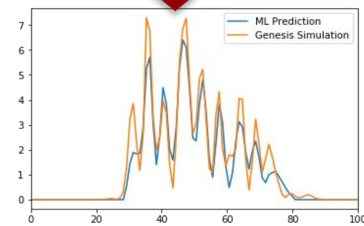
GAN (neural net)

Data analysis

- Computer vision (conv-nets)
- Compressed sensing (convex opt.)

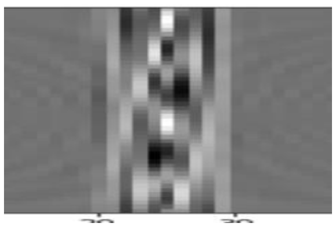
Computer vision



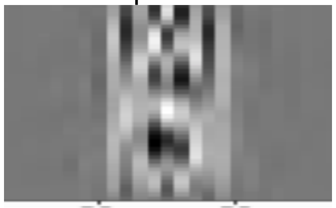


Inverse problems

True Wigner Matrix

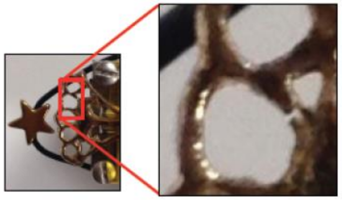


NN prediction

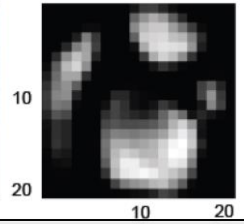


Statistical analysis

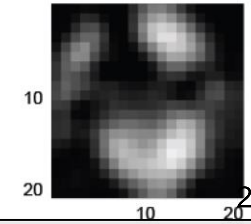
target



ground truth



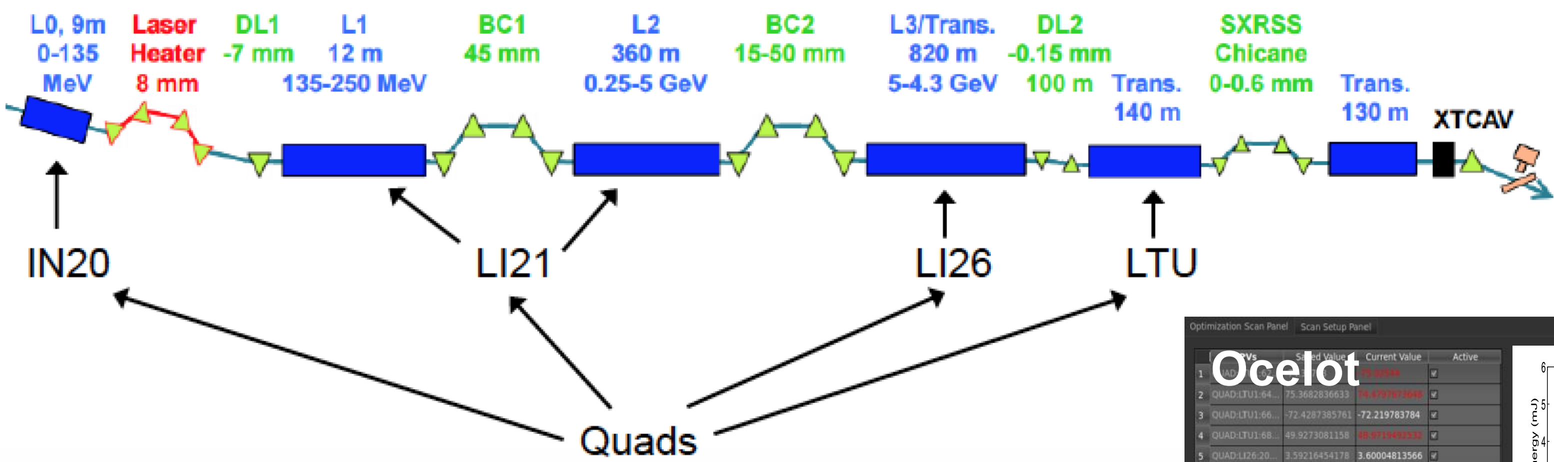
MP, MSE=0.022

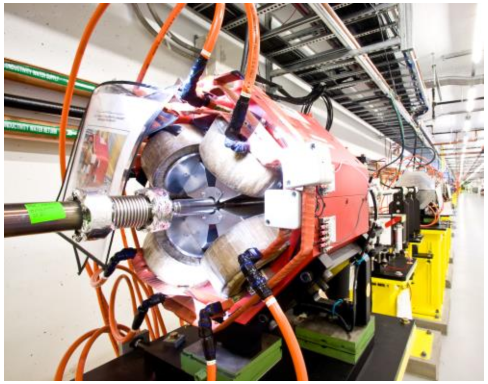


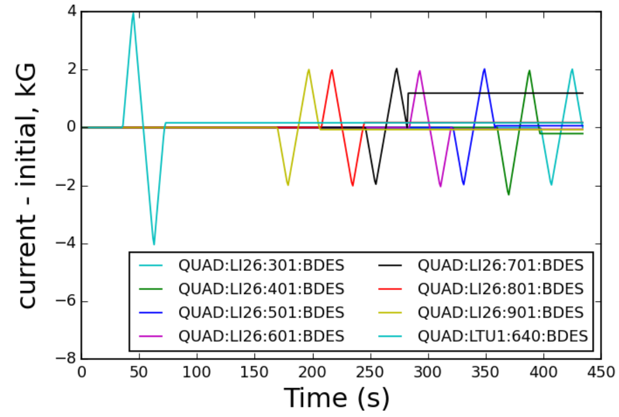
Optimization

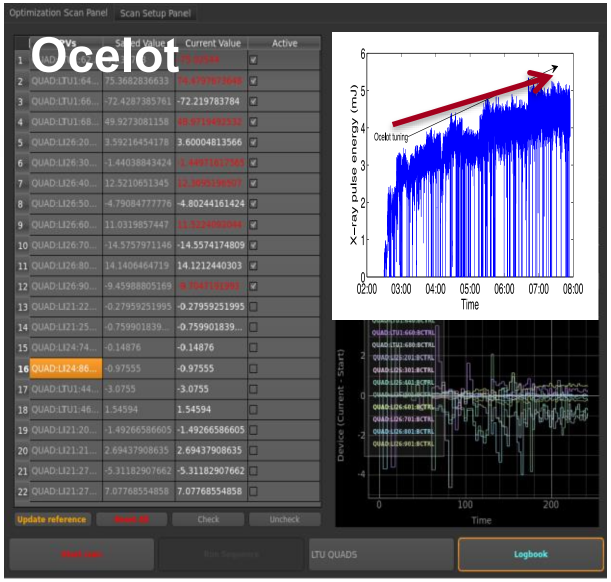
Online tuning:

- Twice daily, ~500 of hours/year
- A single task, quadrupole tuning, required 1 hour/day





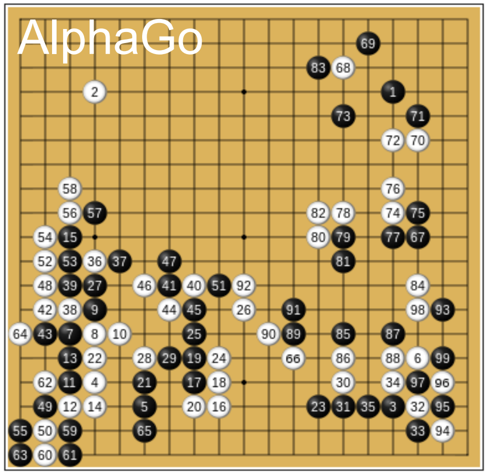


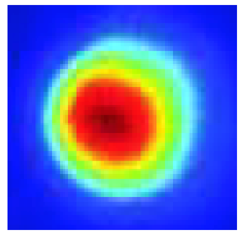


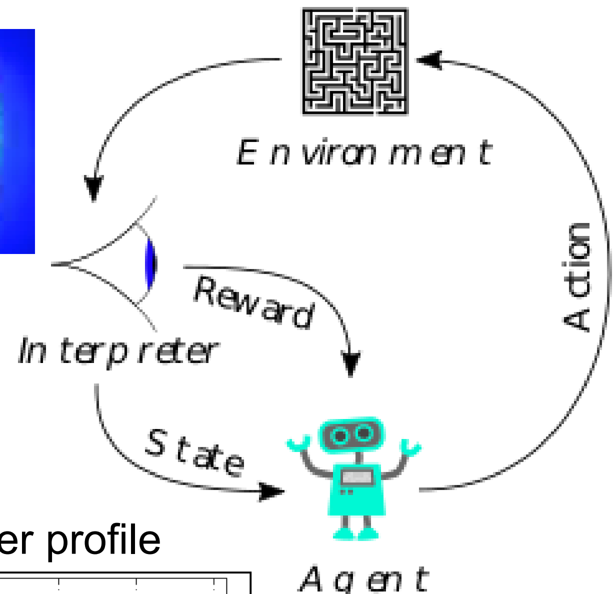
Optimization: Reinforcement learning

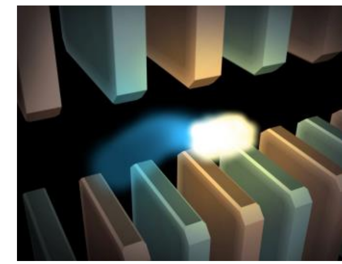
SLAC

Treat optimization like a game: FEL power is the score

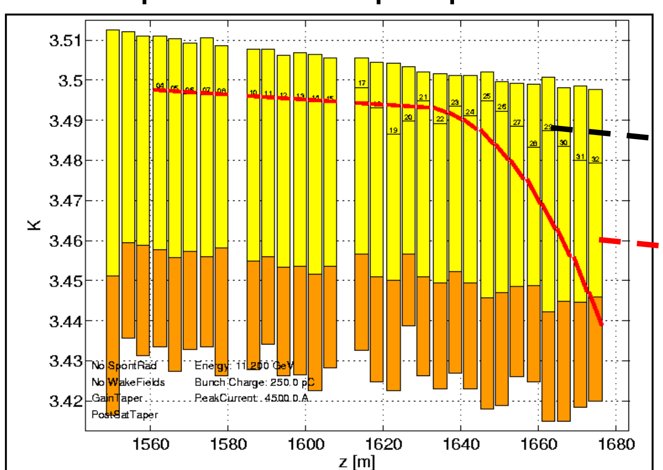




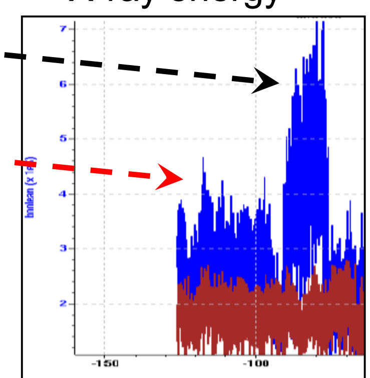




Experiment: Taper profile



X-ray energy



LCLS Experiment: 5.5 KeV Self-seeding FEL, **Zig-zag** doubles power from **continuous profile**

Juhao Wu

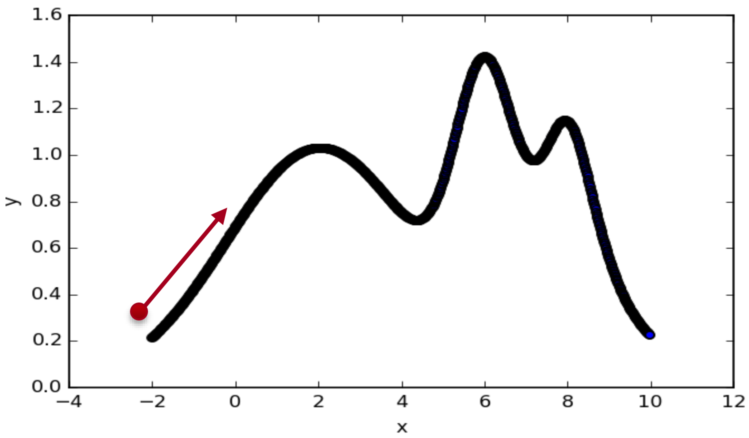
Optimization: Bayesian optimization

Model-based optimization

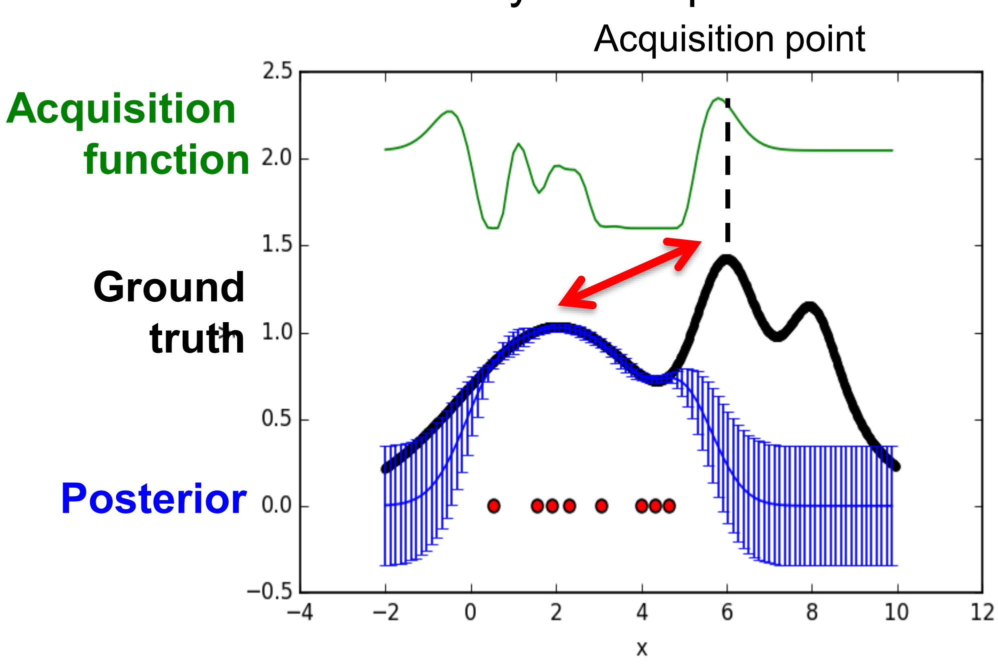
Advantage 1: Balance “exploitation vs. exploration”

➔ Find global maximum

Gradient optimizer



Bayesian optimizer



Joe Duris

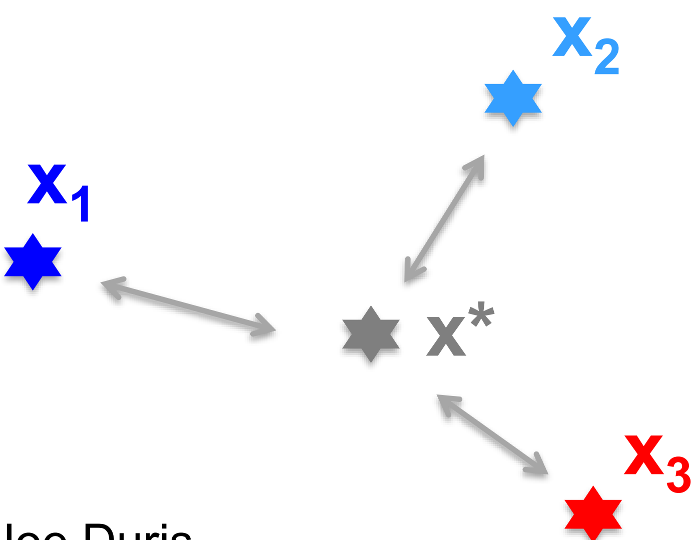
Optimization: Bayesian optimization

Model-based optimization

Advantage 2: Existence of model enables use of physics, archived data

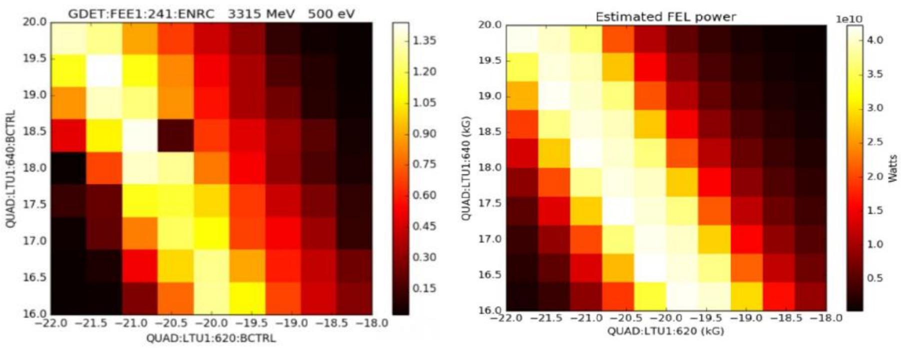
Gaussian process: instance based learning method

Kernel function: $$k(x_1, x_2) = \theta e^{-(x_1 - x_2)^T \Lambda (x_1 - x_2)}$$



FEL vs quads

Measured Modeled

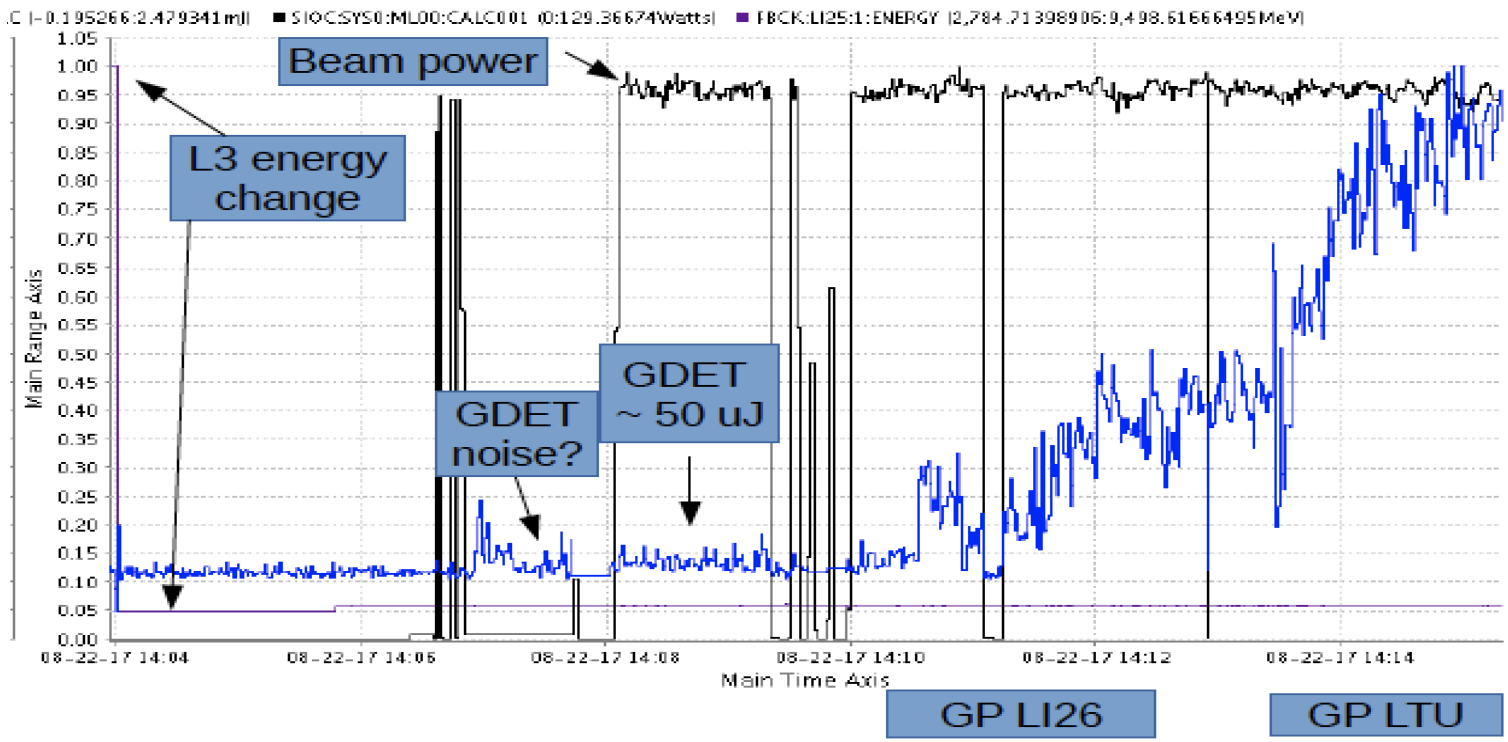


Joe Duris

Optimization: Bayesian optimization

Gaussian process: instance based learning method

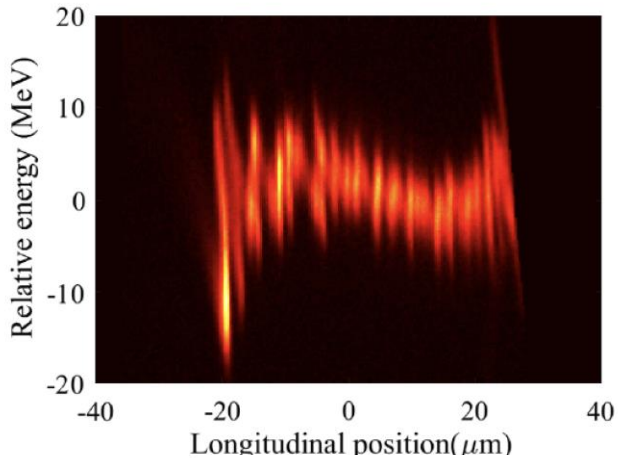
Example: tuning quadrupoles from pure noise



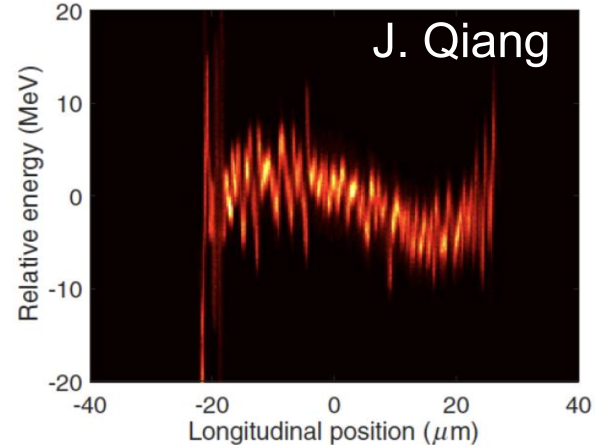
Surrogate Models: Accelerator models

High fidelity physics simulations are remarkable:

LCLS microbunching instability



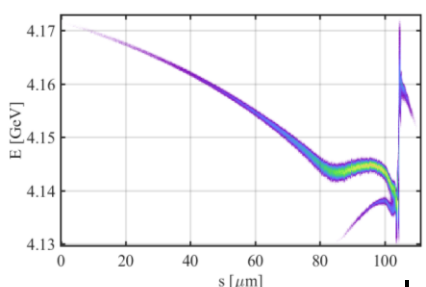
measurement

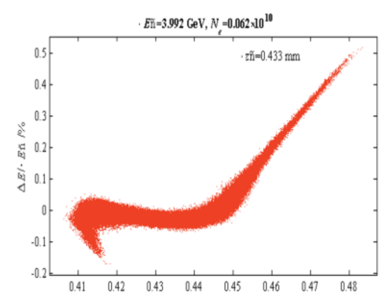


simulation

…but also slow. (e.g. hours on NERSC)

How can we best support design of a new machine?

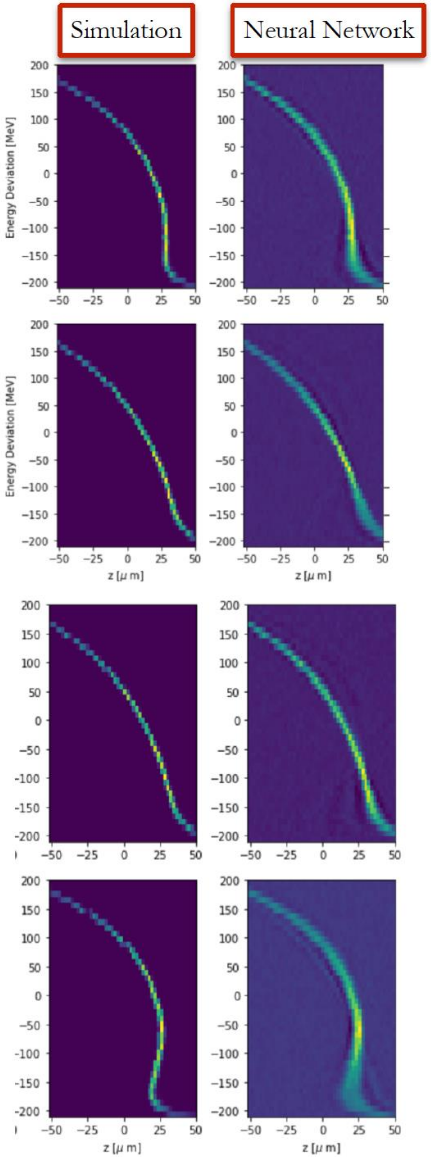




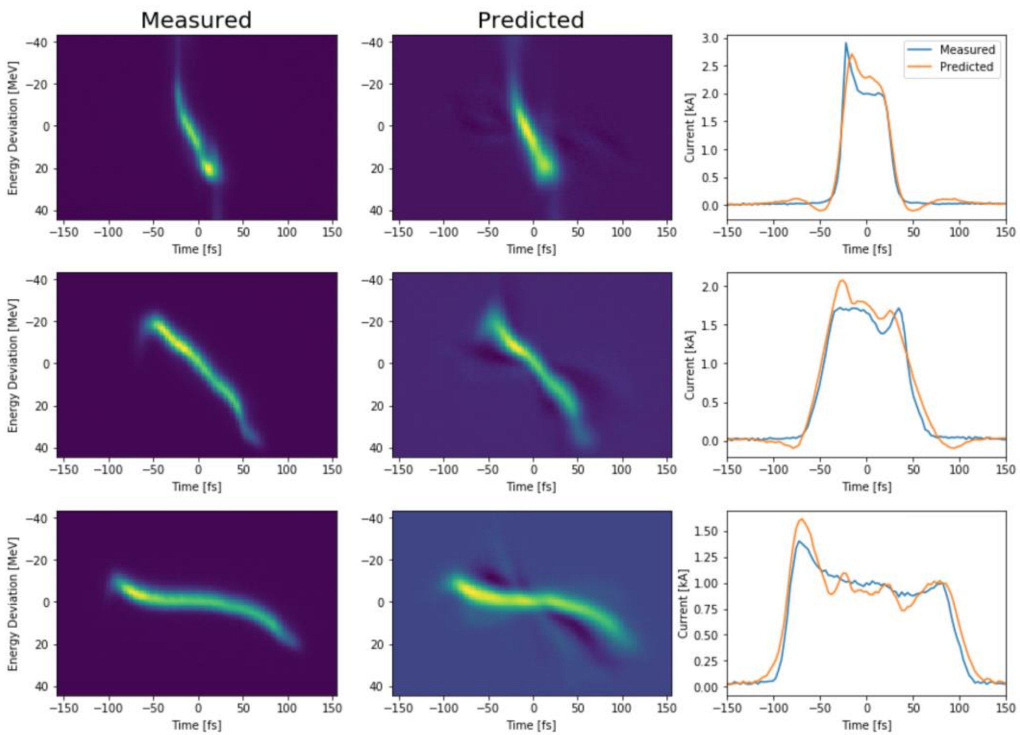
LCLS-II simulations, Y. Ding

Surrogate Models: Accelerator models

SLAC



- Predict XTCAV image from other diagnostic output or upstream machine settings to create a non-destructive **virtual diagnostic**
- Simulation + neural network results match well for FACET-II *(see left)*
- Small study with LCLS machine data and XTCAV images *(scan of L1S phase and BC2 peak current at 13.4 GeV)*



Emma, Edelen, et al. in preparation

Surrogate Models: FEL simulations

SLAC

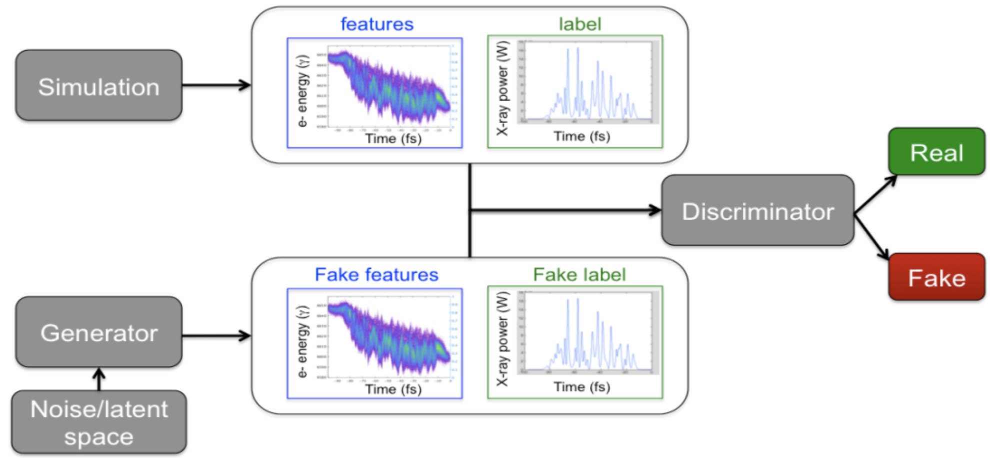
What happens if you turn around your trained network?

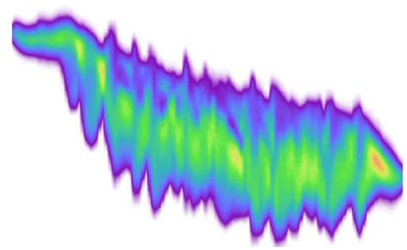
Deep dreaming of dogs

Style transfer

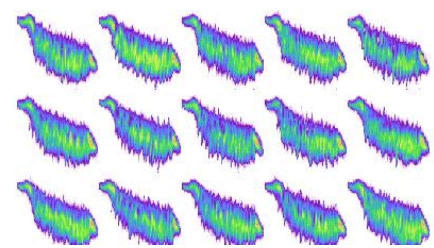
Gatys, et al.

Generative adversarial network (GAN)





Genesis:
~1000 cpu-sec



GAN (neural net):
~0.001 gpu-sec

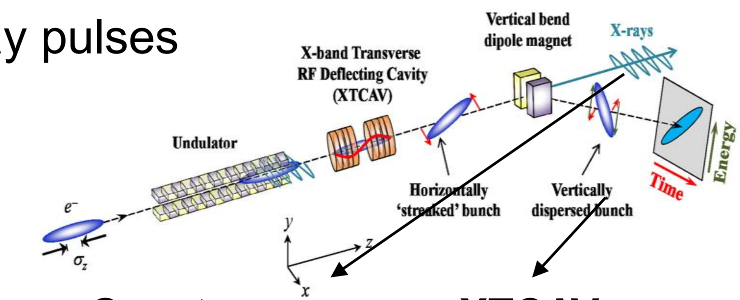
X. Ren

Data Analysis: Pulse reconstruction

Best measurements of X-ray beam come from electrons
Users require:

1. High resolution X-ray power profile shot-by-shot
2. Full (phase and amplitude) reconstruction of X-ray pulses

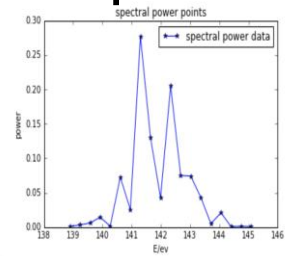
Maxwell, Timothy J., et al. International Society for Optics and Photonics, 2014.



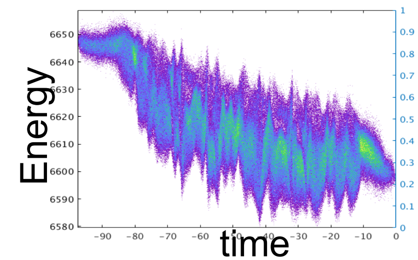
Current analysis algorithm has ~5 fs resolution
➔ Use computer vision to improve algorithm, speed up reconstruction

Combine spectra and high resolution power to reconstruct full FEL pulse
➔ Neural network speeds up to beam rate for users

Spectrum

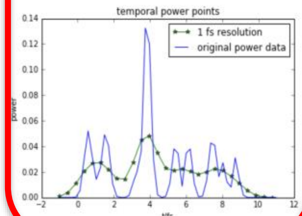


XTCAV



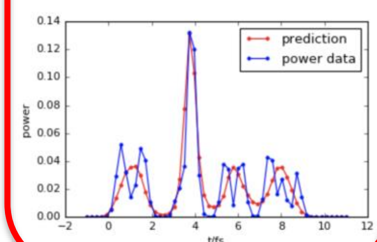
Neural network

Power



Neural network

Full pulse reconstruction



A. Edelen, X. Zhang, X. Ren

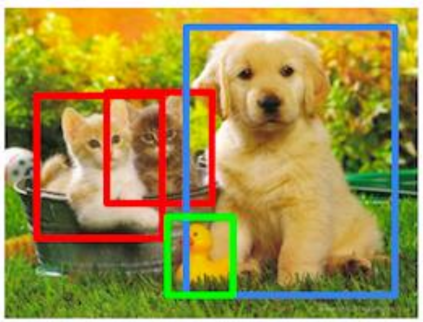
Data Analysis: Pulse reconstruction

Computer vision

Classification + Localization

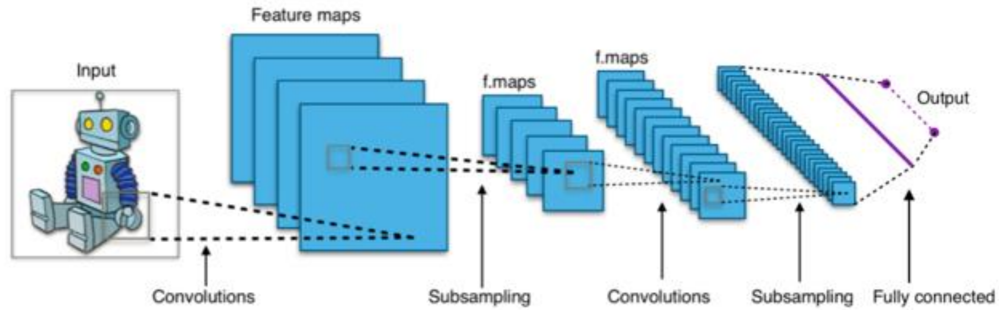
CAT

Object Detection

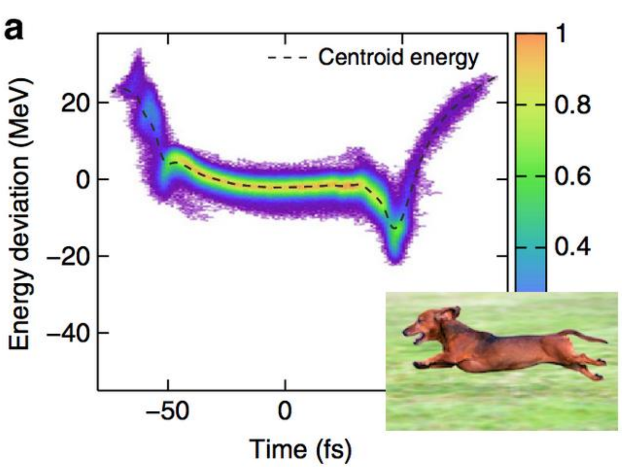


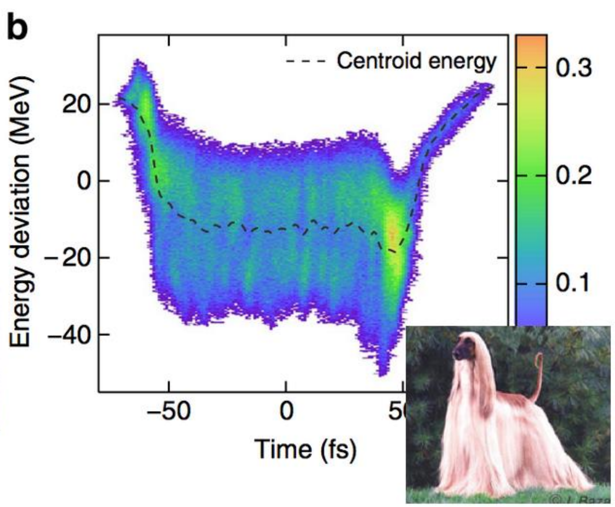
CAT, DOG, DUCK

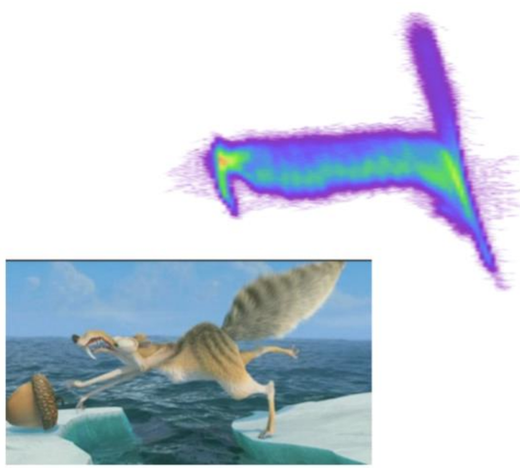
Stanford CS231n



Aphex34 https://commons.wikimedia.org/w/index.php?curid=45679374



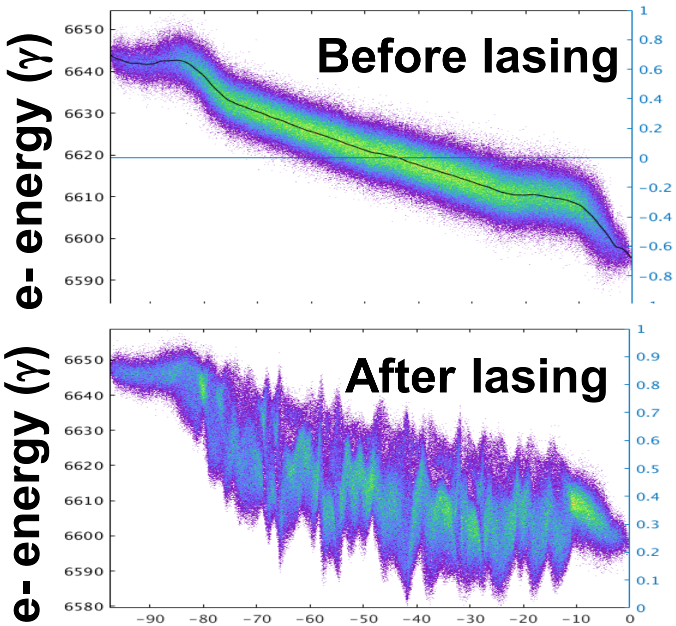


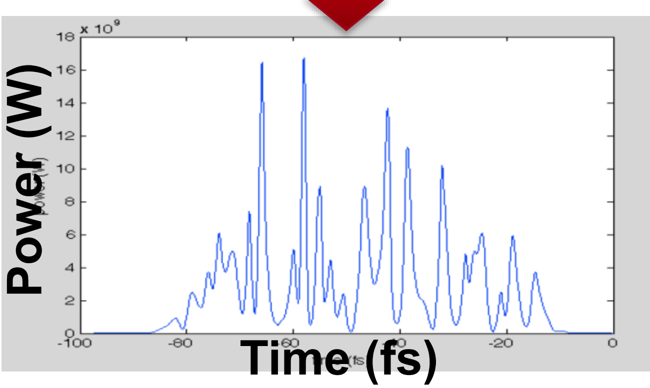


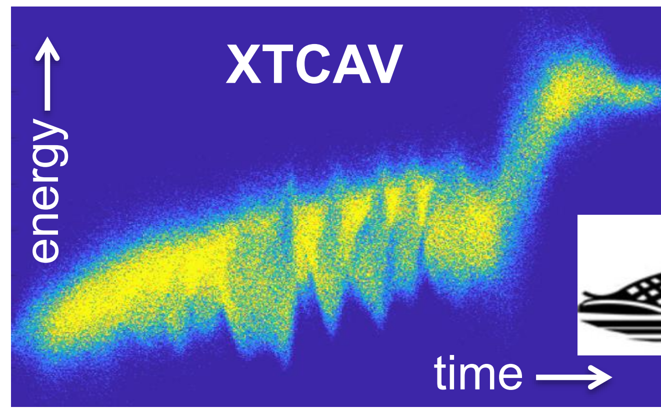
D. Mendez

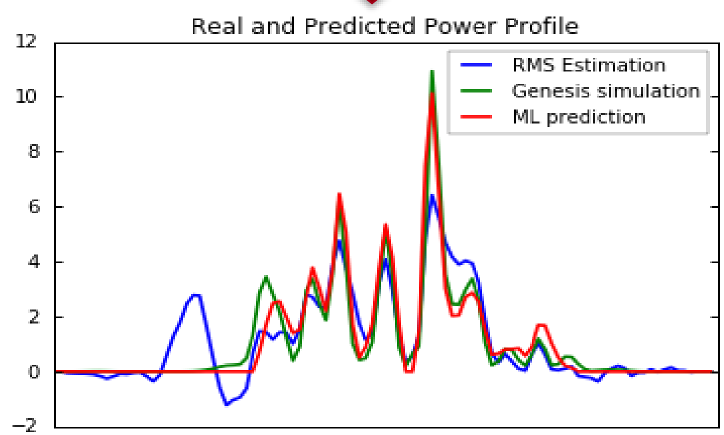
Data Analysis: Pulse reconstruction

XTCAV Analysis







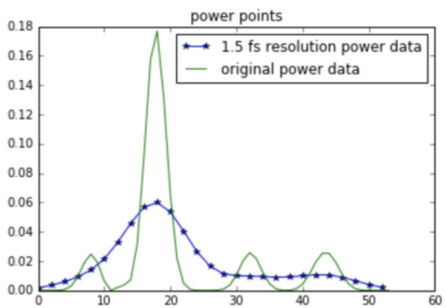


CNN prediction

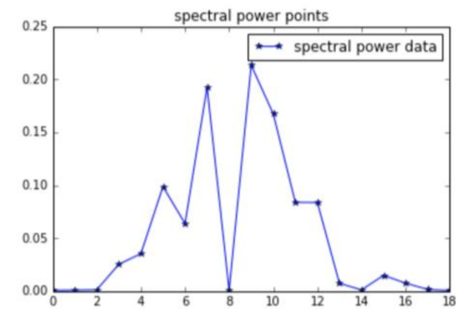
Xinyu Ren

Full beam reconstruction

- Combine time and frequency-domain measurements to retrieve field
- Iterative optimization methods used historically, but slow (minutes to hours)
- Neural network approach: do the optimization once (hours to days) then each example is fast (minutes)



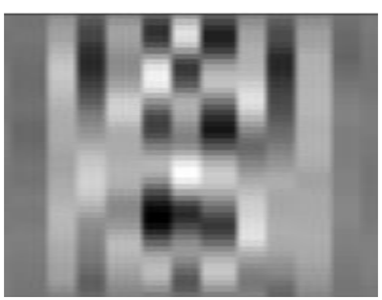
Power



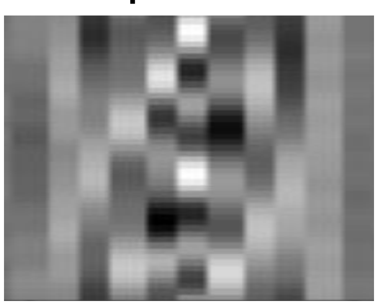
Spectrum

X-ray Field

True Wigner Matrix



NN prediction

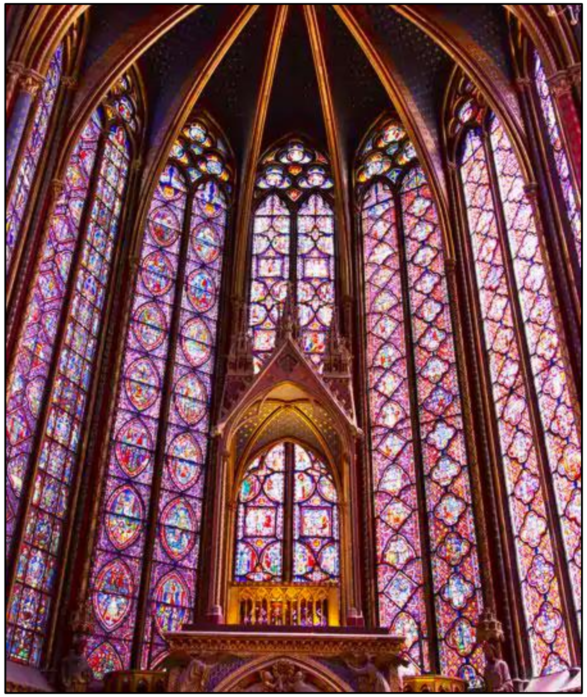


Xiao Zhang

Ghost Imaging / Single Pixel Camera

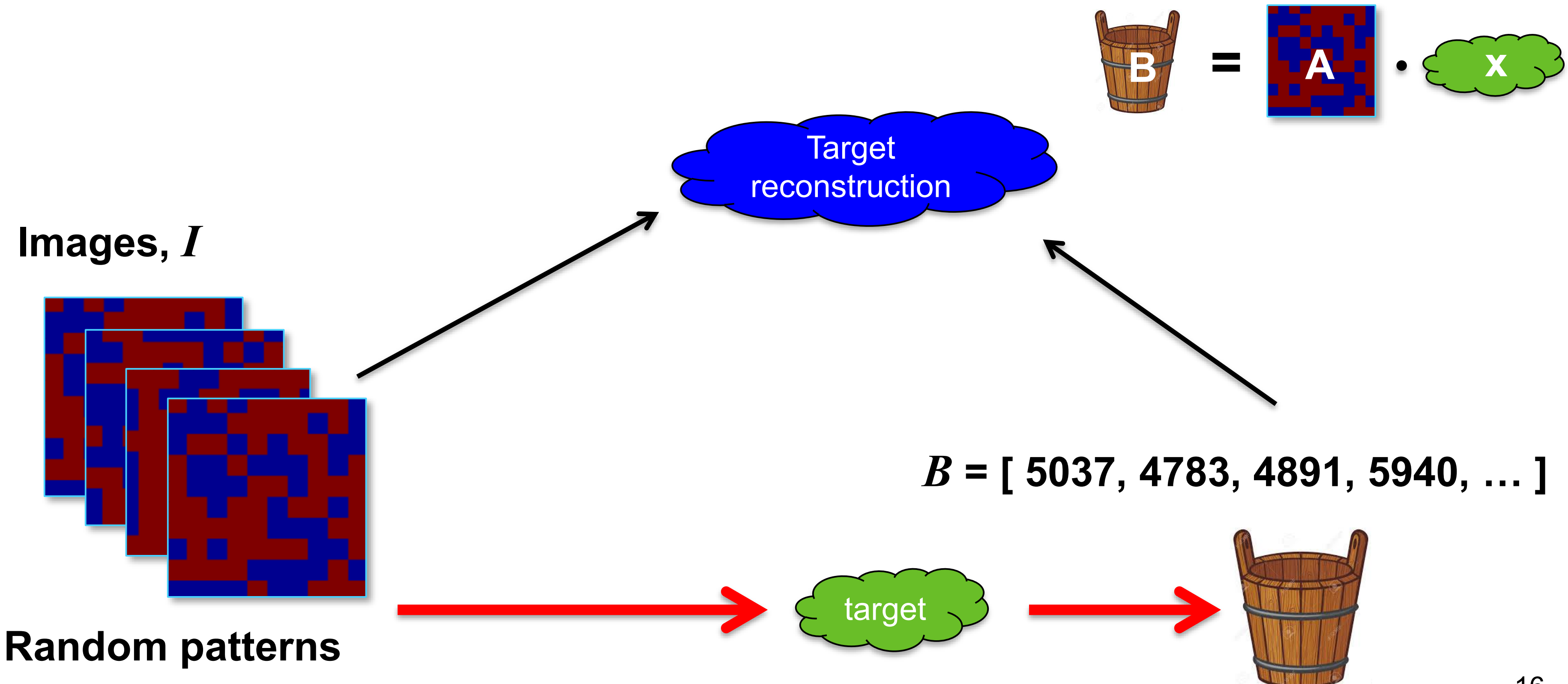
Riddle: How can I take a picture with a spectrometer?

Answer: Have a friend with a flashlight



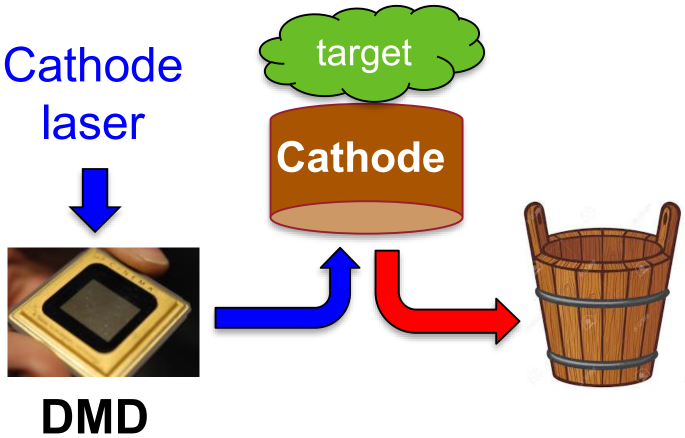
Ghost Imaging / Single Pixel Camera

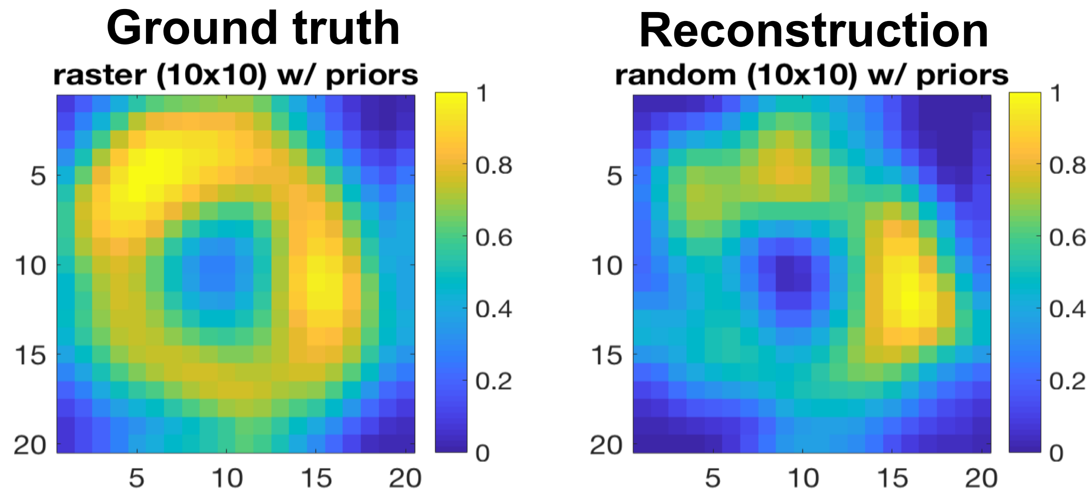
$$\mathbf{x}^{\star} = \mathrm{argmin}_{\mathbf{x}} \left(||\mathbf{Ax}\text{-}\mathbf{B}||^2 + \lambda_2 ||\mathbf{x}||^2 + \lambda_1 \sum_j |x_j| \right) \text{ subject to } x_j \geq 0$$

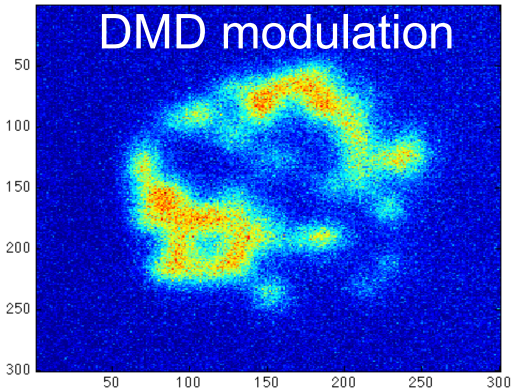


Data Analysis:
Statistical methods for data analysis

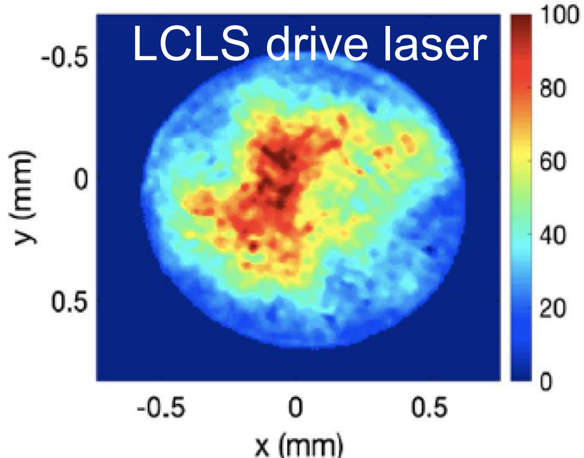
Example application: photocathode quantum efficiency







Don’t need DMD:
exploit natural variation,
jitter of drive laser



Summary:

X-ray FELs are complex, challenging machines. We need all the computational help we can get!

Applications include:

1. **Online tuning:** transverse matching, longitudinal phase space, X-ray spectrum, emittance minimization, etc.)
2. **Surrogate modeling:** efficient machine design, user support, predictive control
3. **Data analysis:** X-ray pulse reconstructions, electron parameters, user experiments

Conclusion

Thanks for your attention!

And thanks to the people who did the work:

E. Cropp, J. Duris, A. Edelen, K. Kabra, D. Kennedy, T. J. Lane, S. Li, T. Maxwell, P. Musumeci, X. Ren, J. Wu, X. Zhang

Gaussian Process Optimizer

Gaussian process: instance based learning method

Covariance function: $k(x_1, x_2) = \theta e^{-(x_1 - x_2)^T \Lambda (x_1 - x_2)}$

observations

new point to predict

prior mean

new point

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix} \right)$$

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

$$K_* = \begin{bmatrix} k(x_*, x_1) & \cdots & k(x_*, x_n) \end{bmatrix}$$

$$K_{**} = k(x_*, x_*)$$

taken from M. Ebner, GP for Regression

Gaussian Process Optimizer

SLAC

Gaussian process: instance based learning method

Covariance function: $k(x_1, x_2) = \theta e^{-(x_1 - x_2)^T \Lambda (x_1 - x_2)}$

observations

new point to predict

prior mean

new point

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix}\right)$$

Prediction of new point: $\overline{y}_* = K_* K^{-1} \mathbf{y}$

Variance of new point: $\mathrm{var}(y_*) = K_{**} - K_* K^{-1} K_*^{\mathrm{T}}$

taken from M. Ebner, GP for Regression

Gaussian Process Optimizer

Gaussian process: instance based learning method

Covariance function: $k(x_1, x_2) = \theta e^{-(x_1 - x_2)^T \Lambda (x_1 - x_2)}$

prior mean

observations

new point to predict

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix}\right)$$

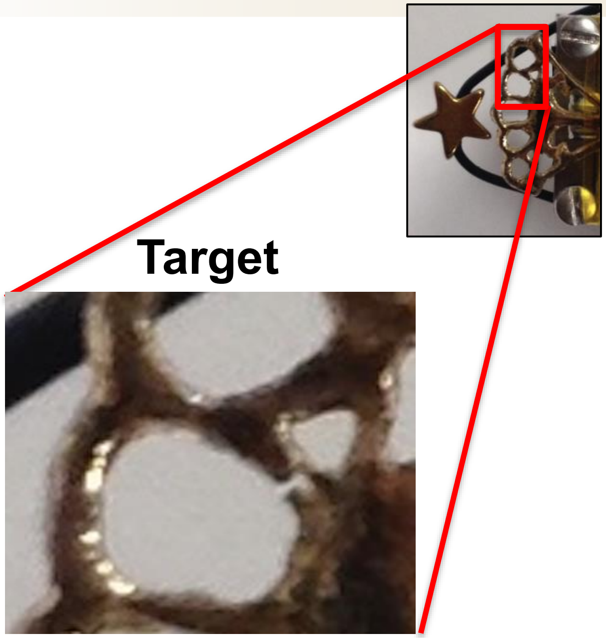
new point

Acquisition function:

$$UCB(x^*) = \mu(x^*) + \sqrt{(\nu \tau_t)}\sigma(x^*)$$

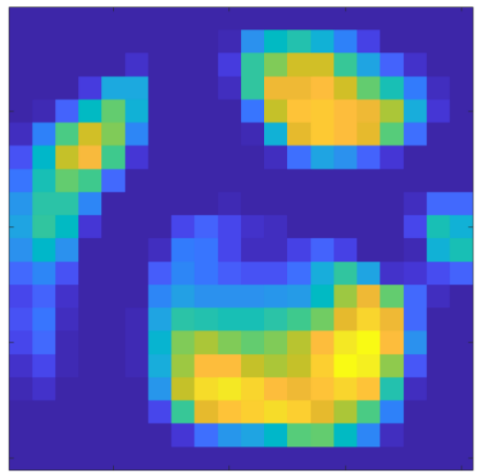
$$\tau(t) = 2\log(t^{d/2+2}\pi^2/3\delta),\ 0 < \delta < 1,\ 0 < \nu$$

Data Analysis:
Statistical methods for data analysis



Experimental Results

Ground truth



ADMM reconstruction