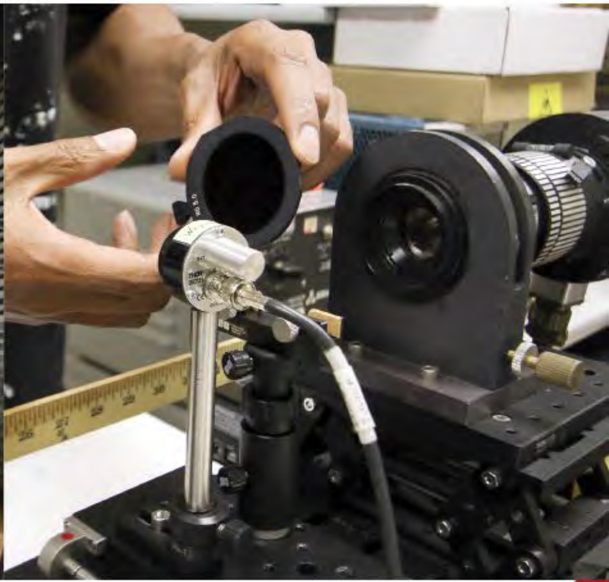
Machine Learning for Accelerator Applications



Bruce Dunham

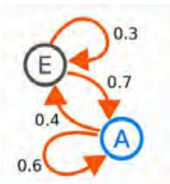
S&T Directorate

The Nevada National Security Site is managed and operated by MSTS under contract number DE-NA0003624.

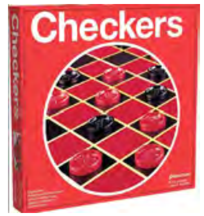
This work was done by Mission Support and Test Services, LLC, under Contract No. DE-NA0003624 with the U.S. Department of Energy, the Office of Defense Programs, and supported by the Site-Directed Research and Development Program. DOE/NV/03624--1695

Machine Learning (ML) Timeline

1812 – Bayes Theorem



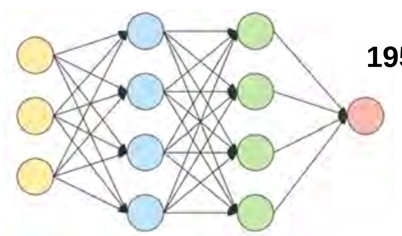
1913 – Markov Chains



1943 – Artificial Neuron

1950 – The Turing Test

1952 – Computer plays checkers



1957 – The Perceptron (Neural Networks)

1967 – The Nearest Neighbor Algorithm

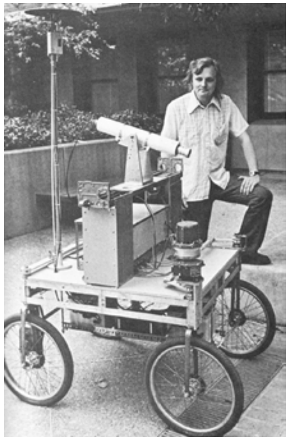
1979 – The Stanford Cart (first self driving ‘car’)

1989 – Reinforced Learning

1992 – Computer plays backgammon

1997 – IBM Deep Blue beats Kasparov

1999 – The GPU is invented



2006 – The Netflix Prize

NETFLIX

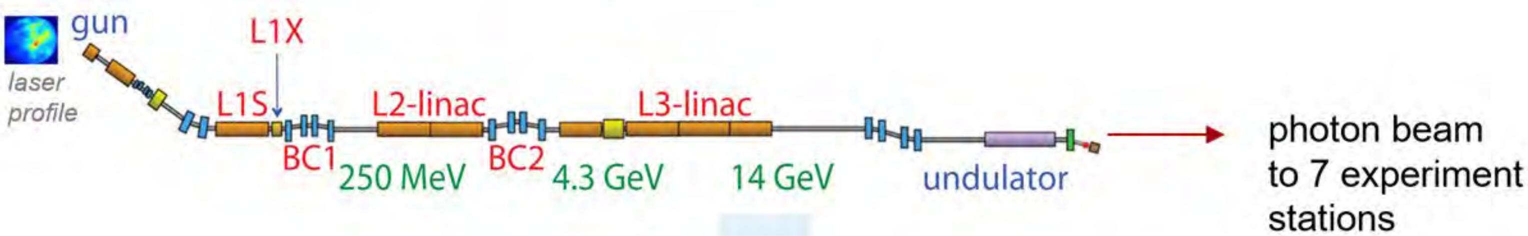
2009 – ImageNet is created

2012 – Recognizing cats on YouTube

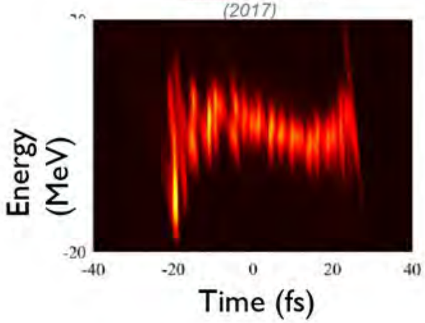
2021 – AlphaFold-2 protein structure prediction

2023 – ChatGPT

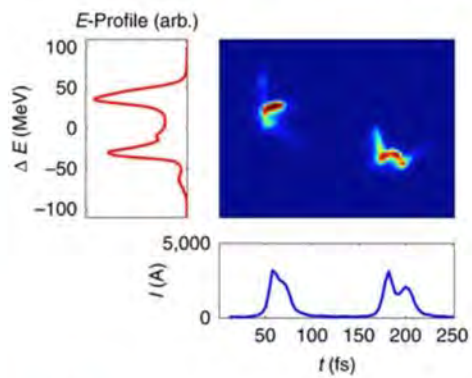
ML for Accelerators – Who Cares?



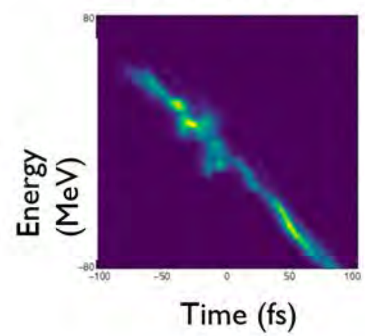
J. Qiang et al, PRAB (2017)



A. Marinelli, et al., Nat. Commun. 6, 6369 (2015)



A. Marinelli, IPAC'18

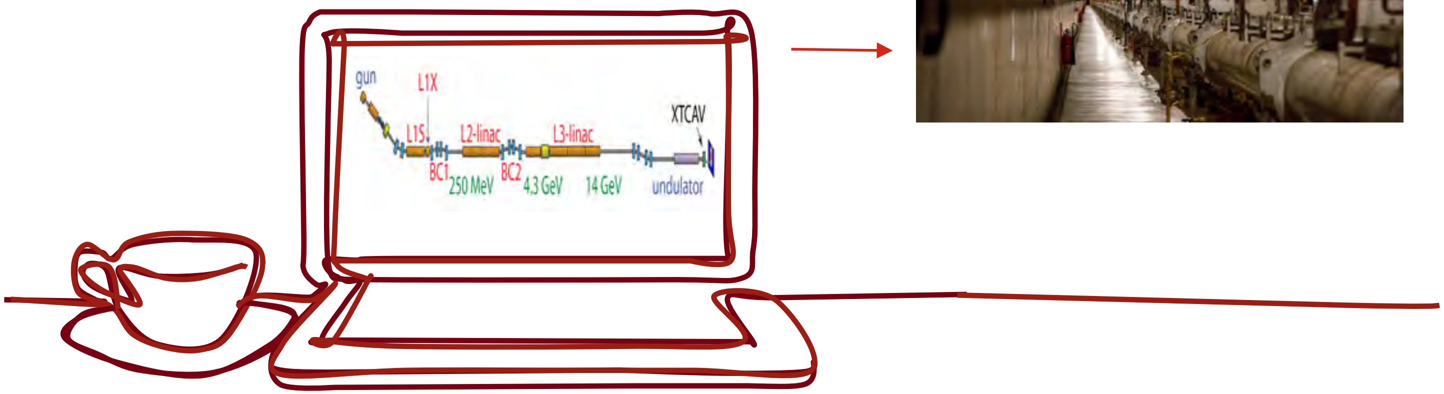


SLAC Annual Operating Budget: $145 million
Approximate hours of experiment delivery per year: 5000
About $30k per experiment-hour to run

Operators spend 400 hours hand-tuning the accelerator in a year → $12 million value
~10 additional experiments

Slide courtesy of Auralee Edelen

In a Perfect World…

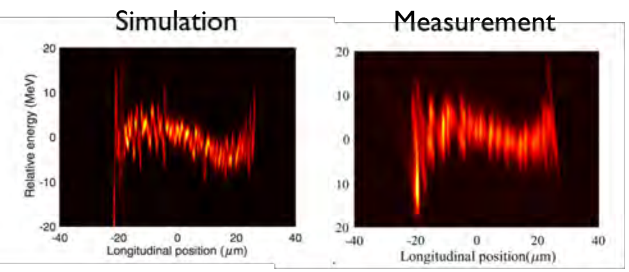


- Use a fast, accurate model …
- Experts find some knobs that give us the beam we want and apply those to the machine
- get info about unobserved parts of machine (online model / virtual diagnostic)
- do offline planning and control algorithm prototyping

Slide courtesy of Auralee Edelen

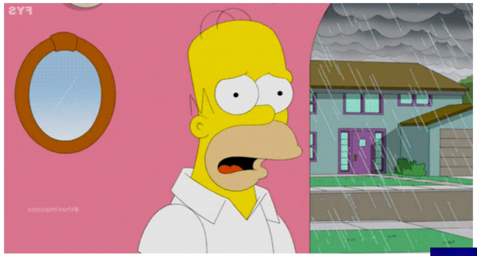
In Reality, Things are Much More Difficult…

computationally expensive simulations

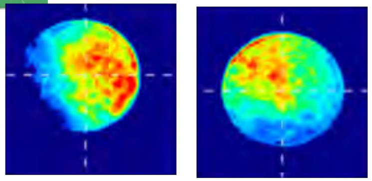


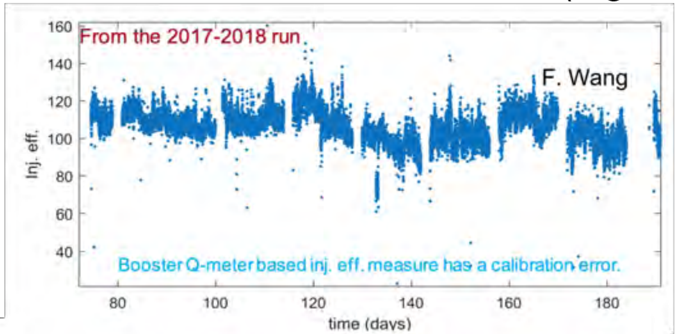
10 hours on thousands of cores at NERSC!

J. Qiang, et al., PRSTAB30, 054402, 2017

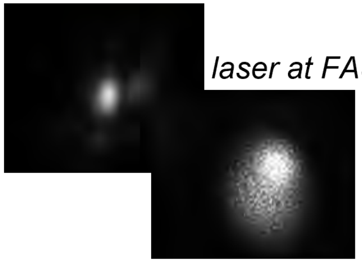


fluctuations/noise (e.g. laser spot)



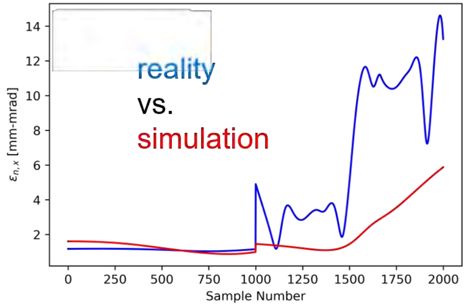


hidden variables / sensitivities

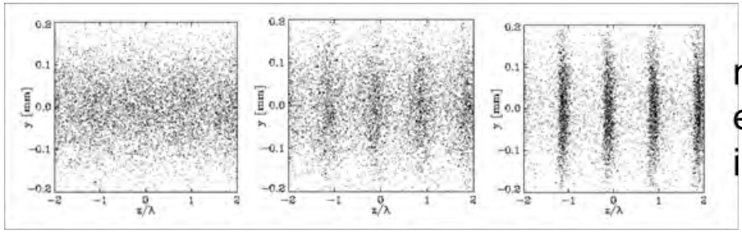


laser at FAST

drift over time



many small, compounding sources of uncertainty

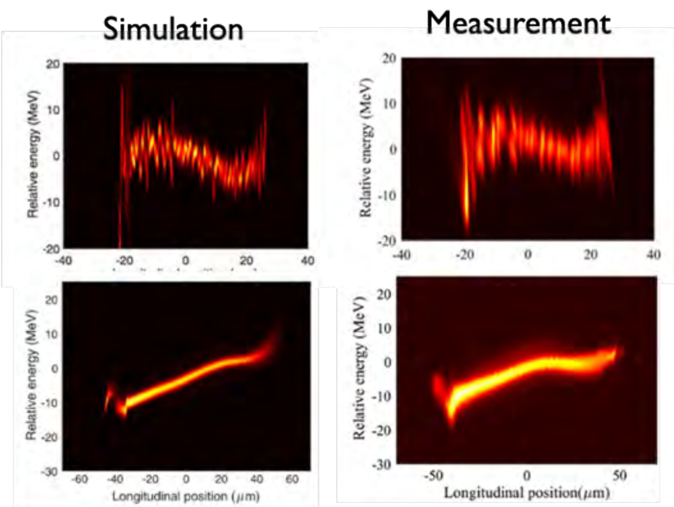


nonlinear effects / instabilities

Slide courtesy of Auralee Edelen

Digital Twins

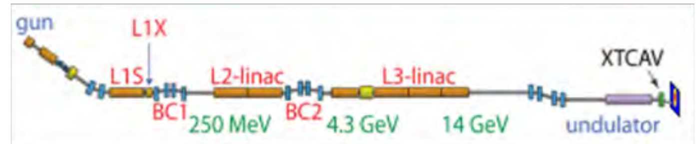
Accelerator simulations that include nonlinear and collective effects are powerful tools, but they can be computationally expensive



J. Qiang, et al., PRSTAB30, 054402, 2017

"10 hours on thousands of cores at the NERSC"

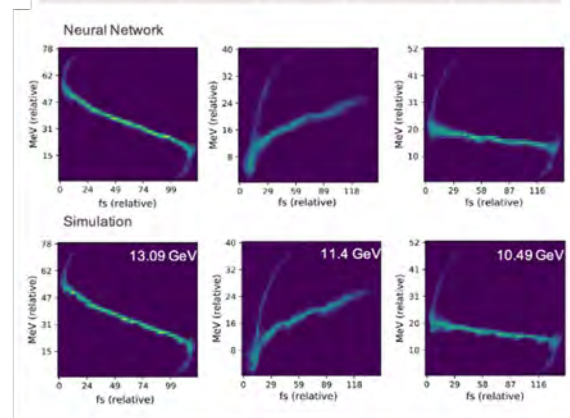
ML models can provide fast approximations to simulations



Linac sim in Bmad with collective beam effects

Scan of 6 settings in simulation

Variable	Min	Max	Nominal	Unit
L1 Phase	-40	-20	-25.1	deg
L2 Phase	-50	0	-41.4	deg
L3 Phase	-10	10	0	deg
L1 Voltage	50	110	100	percent
L2 Voltage	50	110	100	percent
L3 Voltage	50	110	100	percent



< ms execution speed

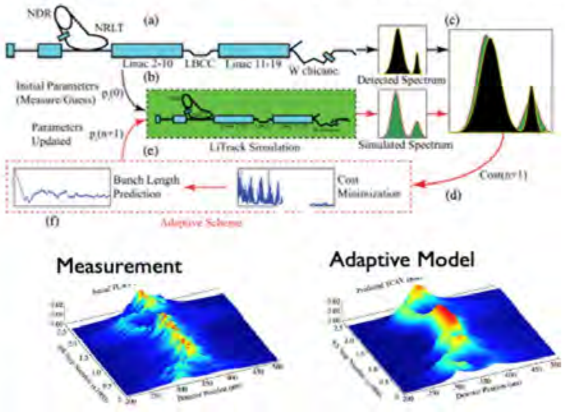
10^6 speedup

Slide courtesy of Auralee Edelen

Virtual Diagnostics

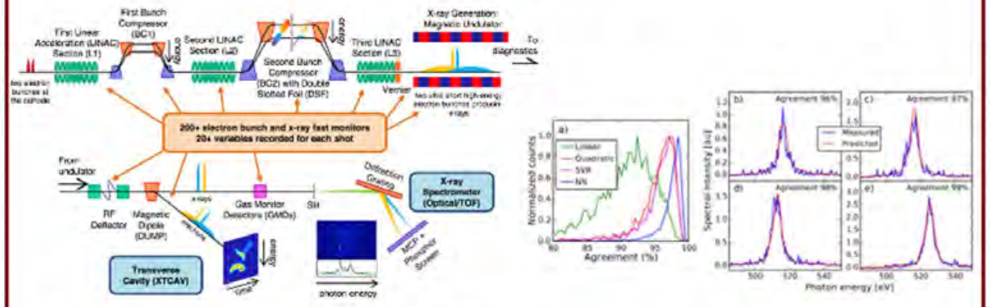
Provide information about parts of the system that are typically inaccessible (destructive, too slow, not directly measurable)

Adaptively tune a simple physics model



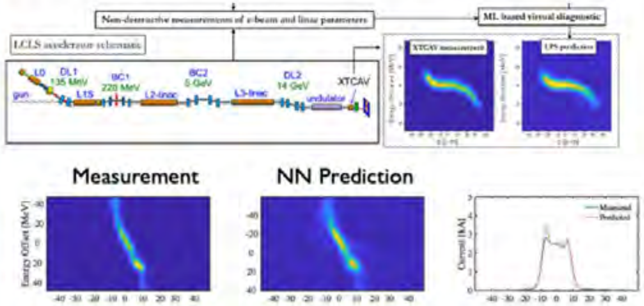
A. Scheinker, S.Gessner, PRAB 18, 102801 (2015)

Fill in shots: use archive data to learn correlation between fast and slow diagnostics



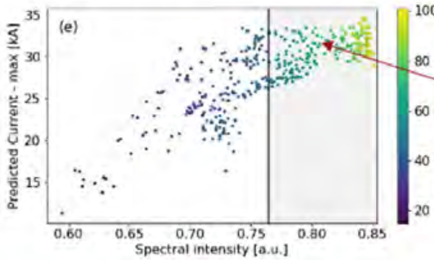
A. Sanchez-Gonzalez, et al., Nature Comms (2017)

Predict with a trained neural network



C. Emma, A. Edelen, et al., PRAB21, 112802 (2018)

Can use spectral information as input to predict beyond typical diagnostic resolution



Shots are beyond the TCAV resolution

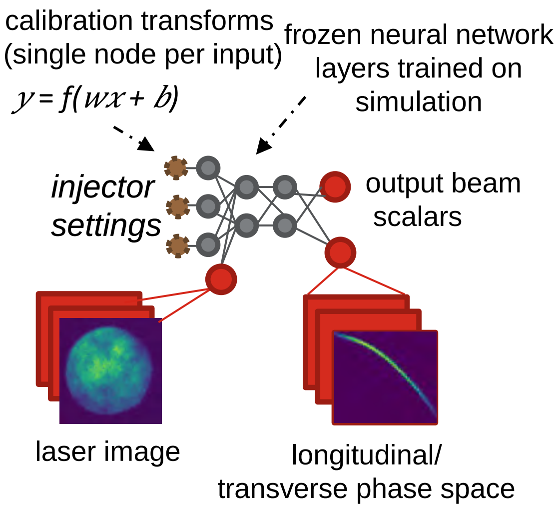
A. Hanuka, et al. 2009.12835 [accepted to Nature Scientific Reports]

Slide courtesy of Auralee Edelen

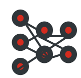
Finding Sources of Error between Simulations and Measurement (Uncertainty Quantification)

Many non-idealities not included in physics simulations:
- **static error sources** (e.g. magnetic field nonlinearities, physical offsets)
- **time-varying changes** (e.g. temperature-induced phase calibrations)

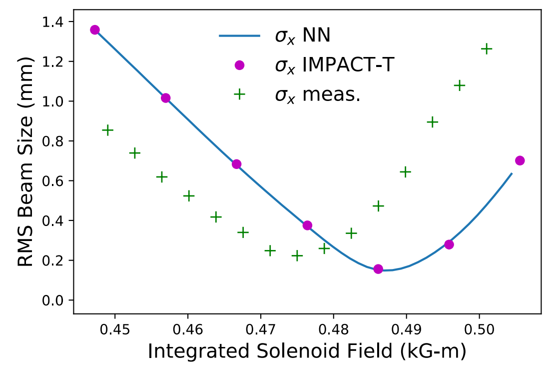
*Want to identify these to get **better understanding of machine** → fast-executing ML model allows fast/automatic exploration of possible error sources simultaneously*



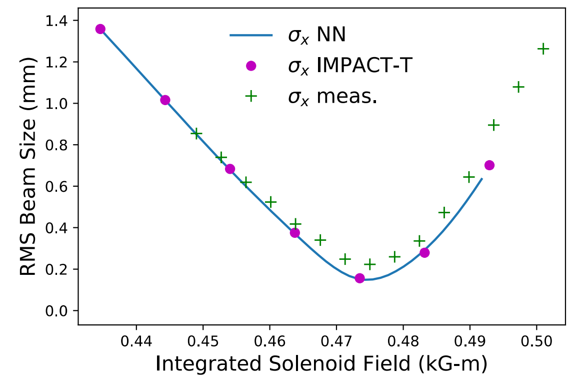
Inputs
Laser radius
Laser spot sizes
Pulse length
Charge
Solenoid
L0A phase
L0B phase
SQ quad
CQ quad
6 matching quads



Outputs
Beam size (x,y)
Emittance (x,y)
Bunch length



Without calibration



With calibration

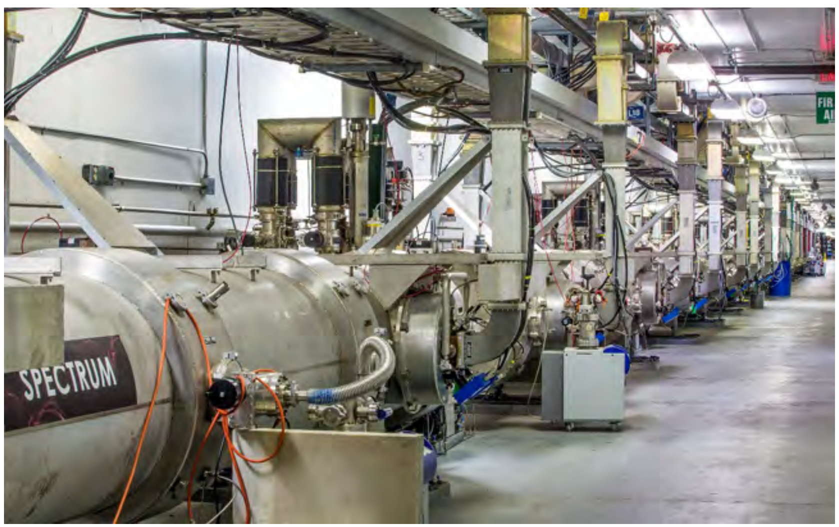
Calibration offset in solenoid strength found automatically with neural network model (trained in simulation, then calibrated to machine)
Example above is simulation-to-machine, but can adapt model over time as well

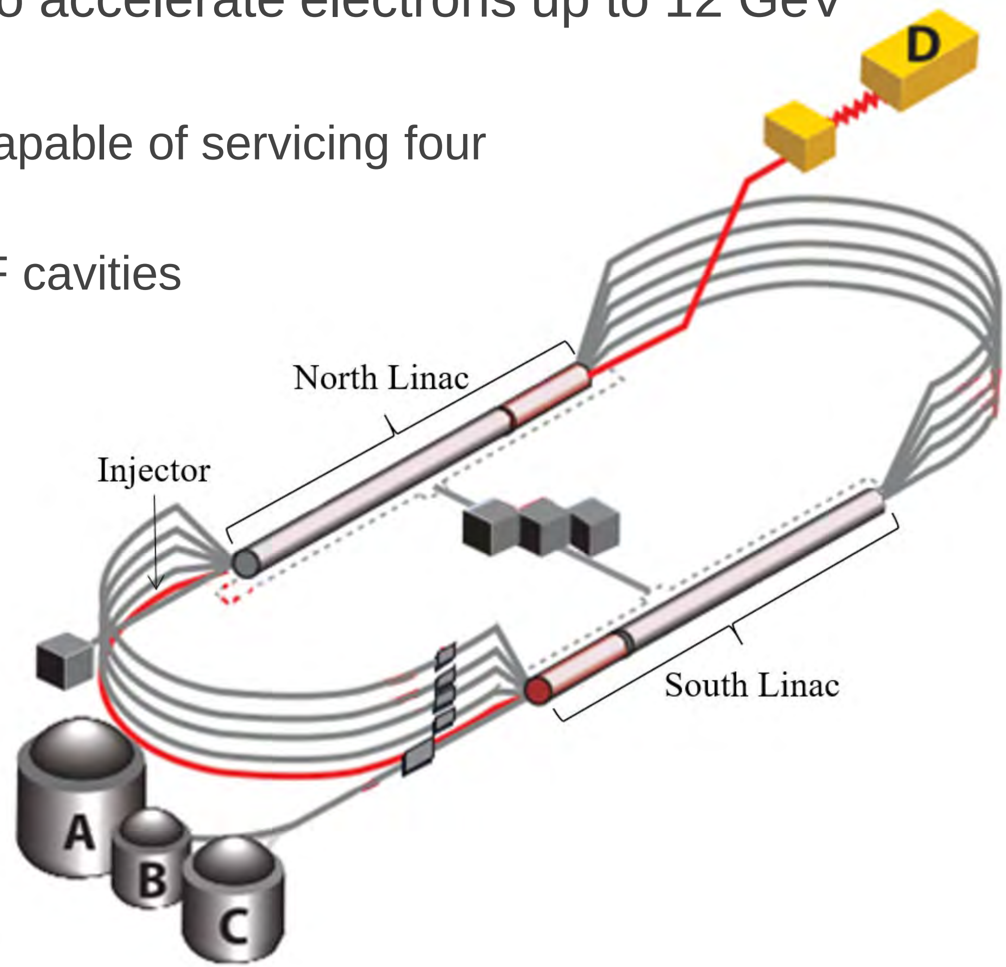
Slide courtesy of Auralee Edelen

Continuous Electron Beam Accelerator Facility

► The Continuous Electron Beam Accelerator Facility (CEBAF) is a continuous wave (CW) recirculating linac utilizing 418 superconducting radio frequency (SRF) cavities to accelerate electrons up to 12 GeV through five passes

- It is a nuclear physics user-facility capable of servicing four experimental halls simultaneously
- The heart of the machine is the SRF cavities





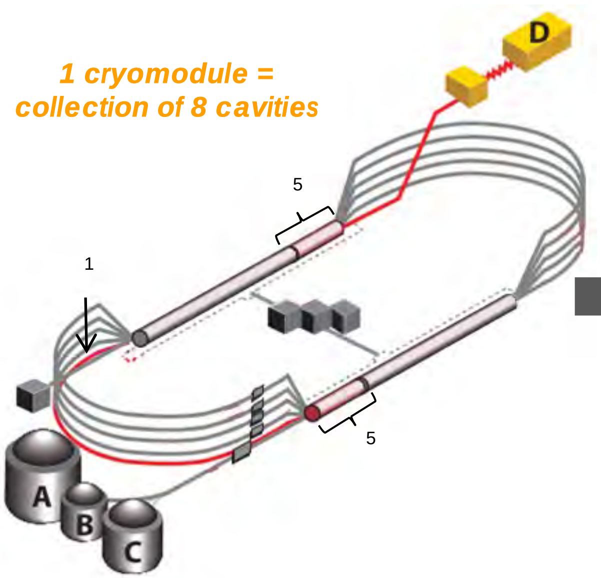
Slide courtesy of Chris Tennant

Predicting Failures

Fault Classification: Defining the Problem

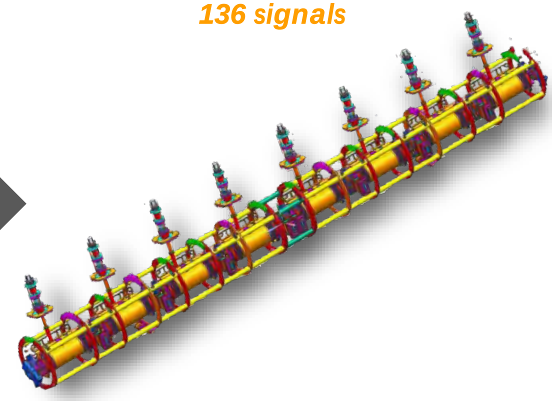
They record high-fidelity data from 12 cryomodules

1 cryomodule = collection of 8 cavities



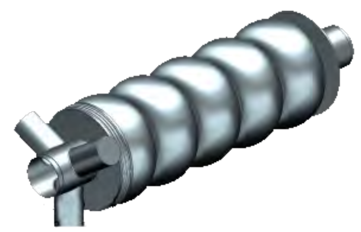
Question #1
Which of the 8 cavities faulted first?

17 signals/cavity × 8 cavities = 136 signals



Question #2
What kind of trip was it?

17 signals



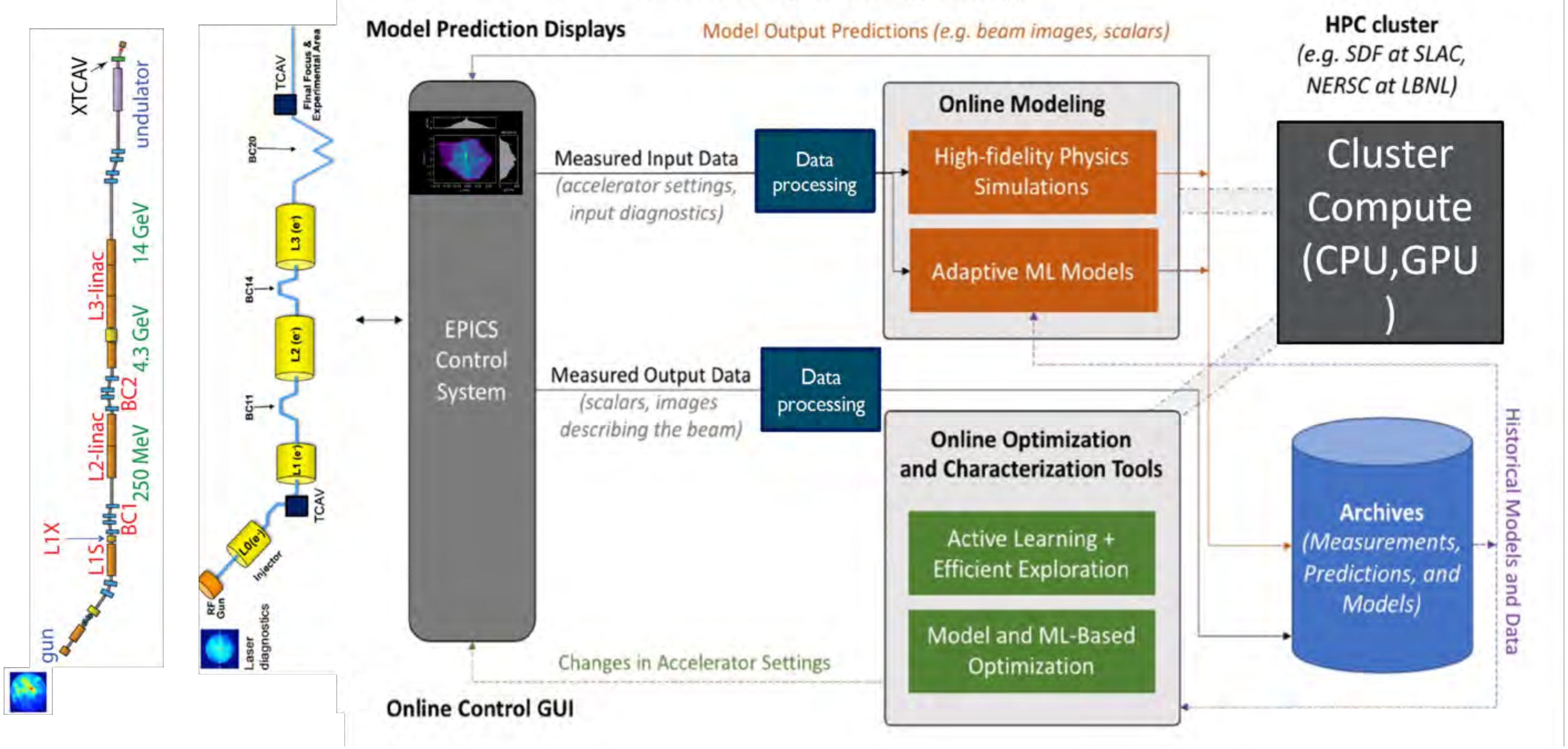
Train an ML algorithm to correctly classify the <u>cavity</u> and <u>type</u> of RF fault given waveform data. The results can be used to identify a maintenance action to take, for example

Slide courtesy of Chris Tennant

Full Integration of AI/ML Optimization, Data-Driven Modeling, and Physics Simulations is the Goal for Accelerator ML

*Want a **facility-agnostic*** ecosystem for online simulation, ML modeling, and AI/ML driven characterization/optimization

Will enable system-wide application to aid operations, and help drive AI/ML development *(e.g. higher dimensionality, robustness, combining algorithms efficiently)*



Slide courtesy of Auralee Edelen

Materials Discovery – Photocathodes for Accelerators

NEVADA NATIONAL NNSS SECURITY SITE

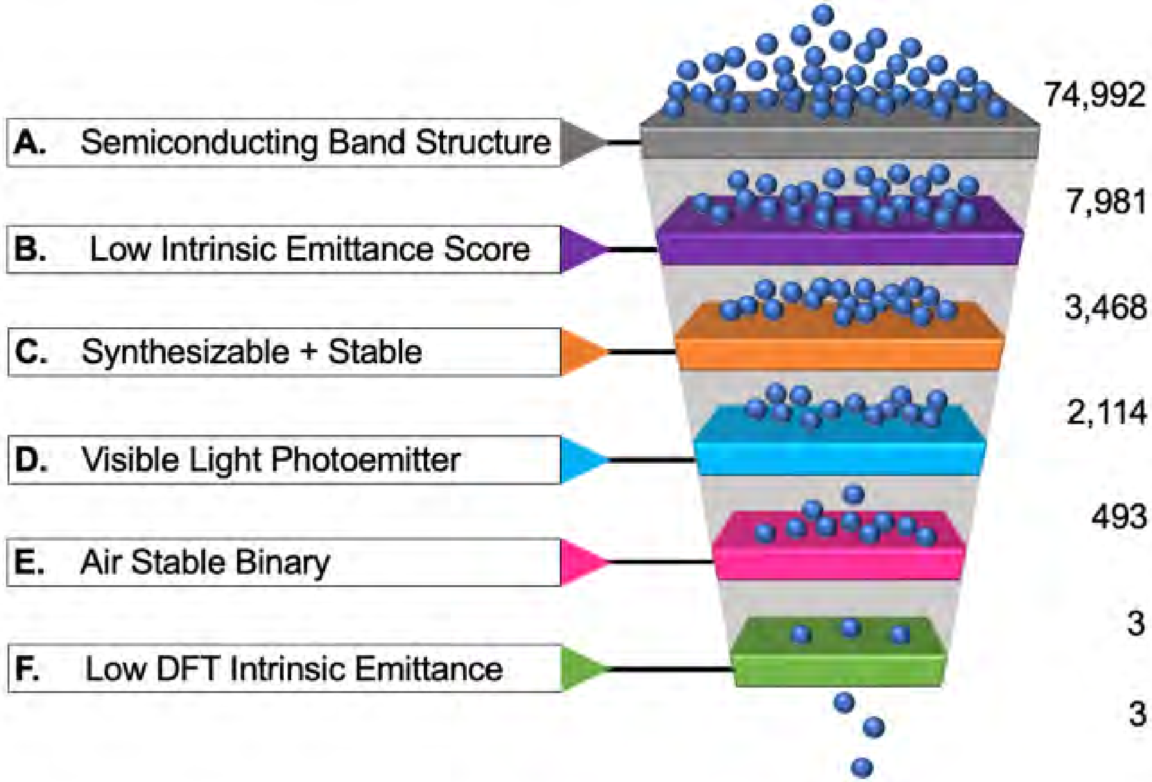
The goal of this project was to identify new photocathode materials for electron sources using ML and Big Data techniques.

We performed **multi-objective screening** to materials that are:

i) **Air-stable**
ii) **Visible light active**
iii) **Low emittance**

Air stability: Look for **oxide binaries**

Best candidate materials:
Na_2O, K_2O, Rb_2O

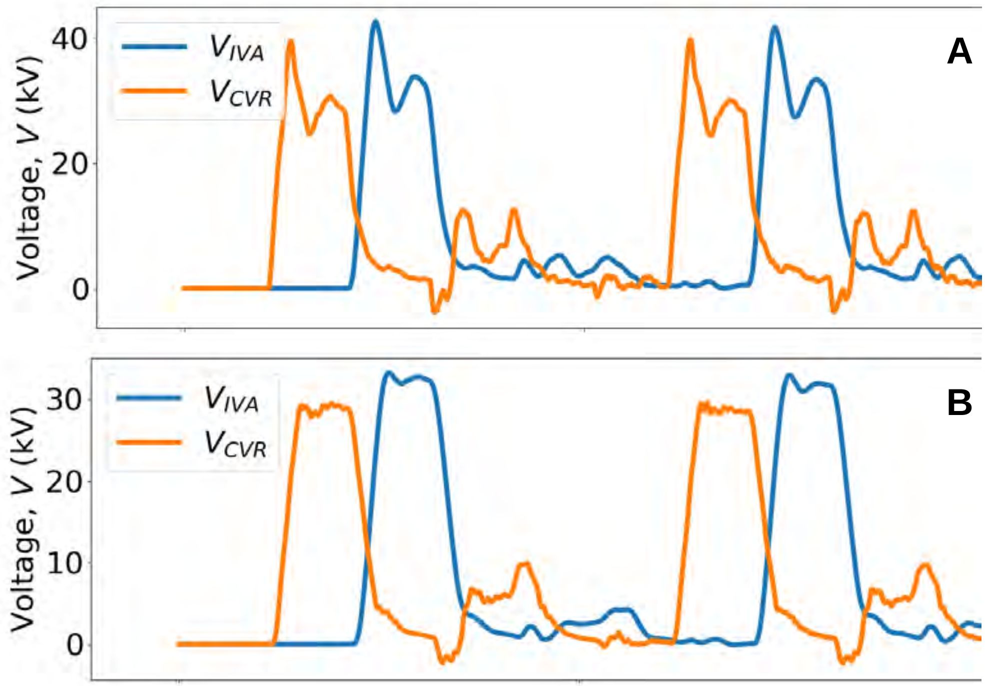


This technique can be applied to other applications of interest, such as searching for new scintillator materials for detectors

Antoniuk et al., Advanced Materials **(2021)**

ML Projects on Scorpius

- Exploring use of ML to speed optimizations on Scorpius
  - Solid-state pulsed power modulation for pulse shaping
- Collaboration with UNLV on interfacing ML model with controls
  - Important for field application of modulation



Demonstration of basic effects of an unmodulated pulse (A) and a staggered pulse (B) using CASTLE.

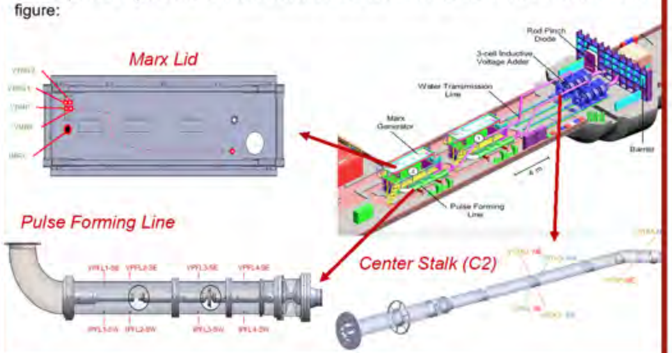
ASD Scorpius

Slide from Evan Scott

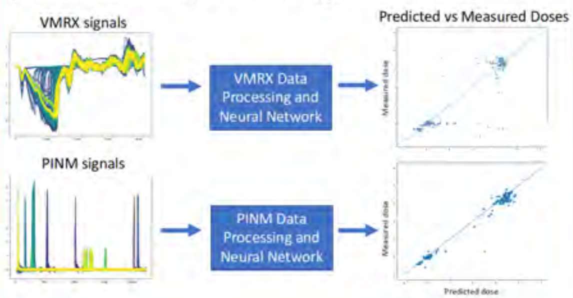
Health Assessment and Performance Monitoring of Large Machine Diagnostics

SDRD by Jesse Adams

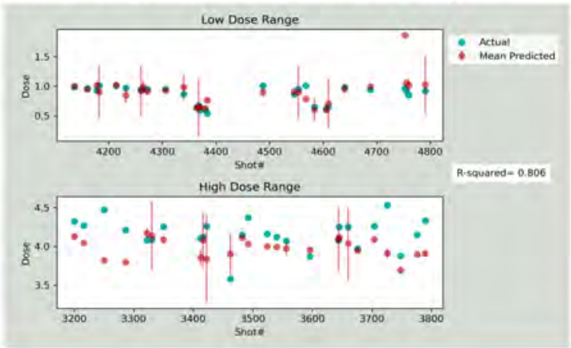
Diagnostic sensors on each Cygnus axis gather information on each experiment that can be potentially used to characterize the machine. Their locations are indicated in this figure:



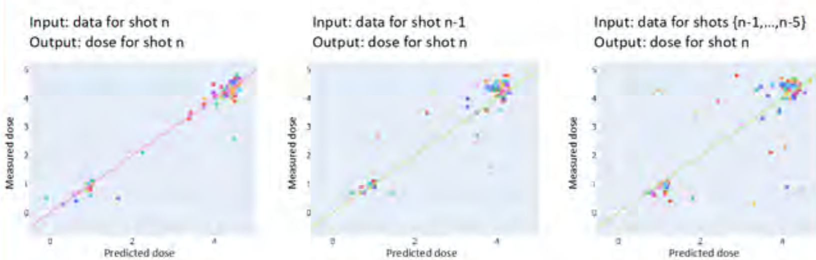
The raw trace data are processed, and used as input for a neural network. The experiment dose is the output. On the left, are two different signal inputs, and the right shows the relationship between the dose predictions and the truth (the nearer the diagonal line, the better).

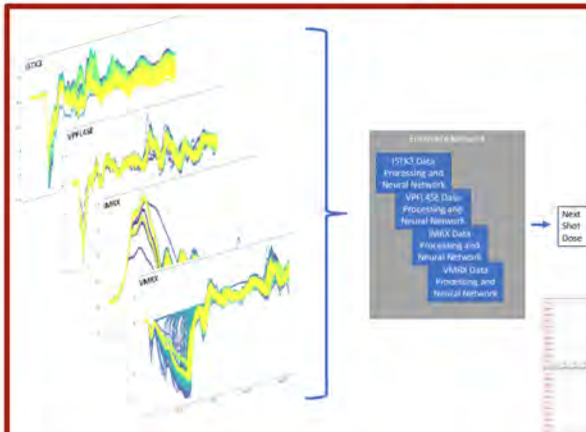


Monte Carlo dropout was used to produce an ensemble of deep neural networks in order to provide uncertainty quantification in the output

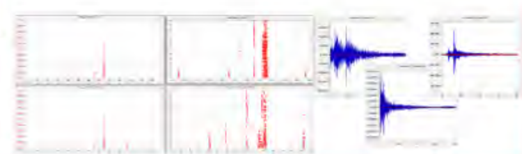


The ultimate goal is to predict machine failure. Networks were trained on previous shot inputs, in order to predict the next shot dose (again, nearer the diagonal line, the better the prediction).





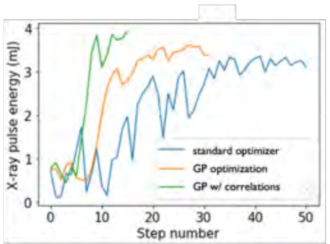
Our next steps are to create an ensemble network that incorporates multiple signals, and predicts the next shot dose (figure on the left). We have also deployed a new set of sensors, which can be introduced into the model once they have collected sufficient data (shown below).



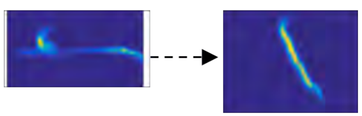
Summary - Broad Set of Areas for ML to Impact Accelerator Operations

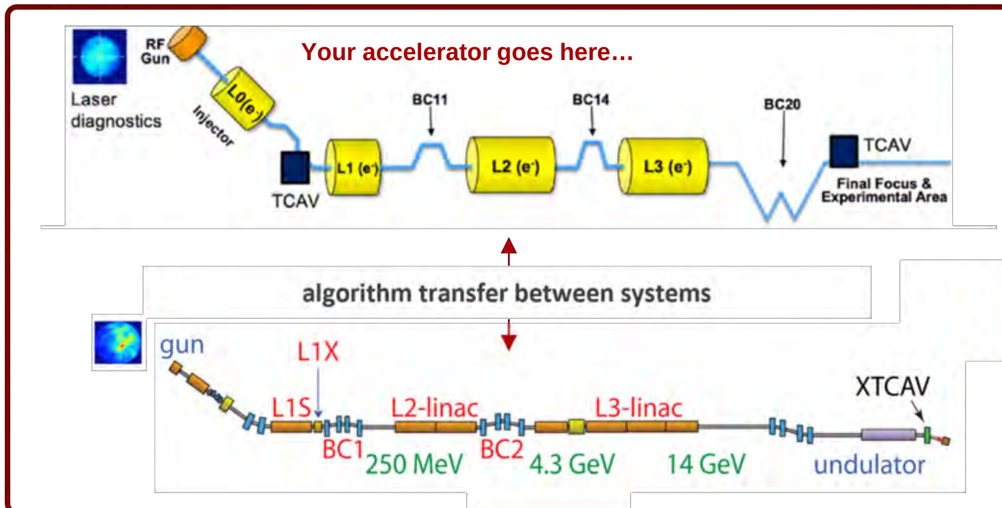
Data reduction/rejection *(kHz/MHz data streams)*
Event triggering

automated control + optimization

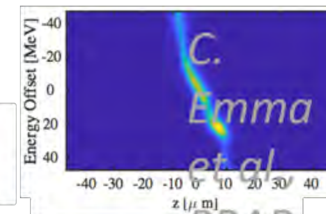


J. Duris et al., PRL, 2020



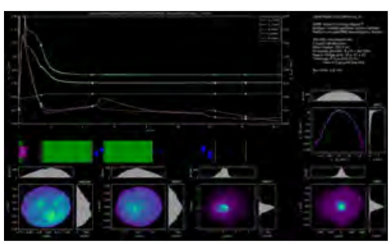


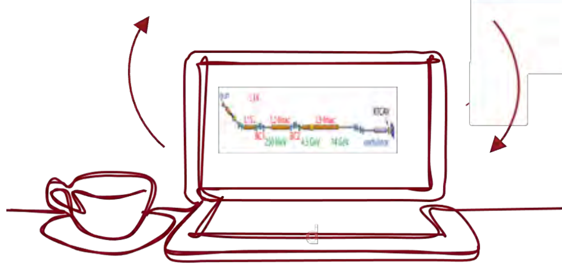
ML-enhanced diagnostics
(provide insight at faster rate, at higher resolution, non-invasively)



C. Emma et al. PRAB, 2018

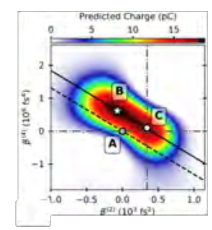
anomaly detection
failure prediction
(plan maintenance; alert to changes in machine; alert to interesting science)





digital twins + online modeling
(fast sims, differentiable sims, model calibration, model adaptation)

extract unknown relationships + correlations
(feed into future control / design)



R. Shaloo et al. arXiv:2007.14340

+ need uncertainty quantification for all
+ can incorporate physics information in all