Cognitive Simulation: combining simulation and experiment with artificial intelligence

October 12, 2020

University of California, Santa Cruz

Brian K. Spears

spears9@llnl.gov

@bkspears9

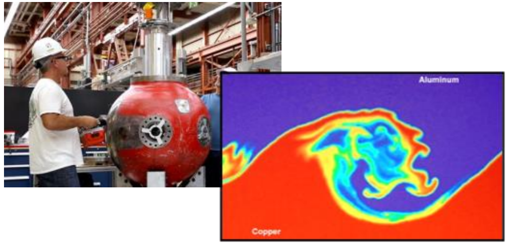
LLNL-PRES-757656

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

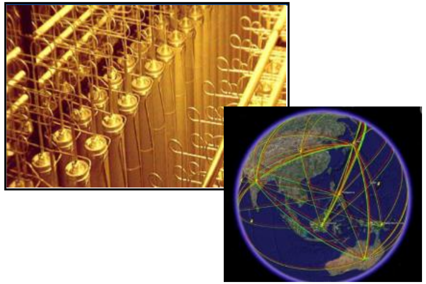
Lawrence Livermore National Laboratory

We’re developing solutions for complex data problems across LLNL’s national security missions

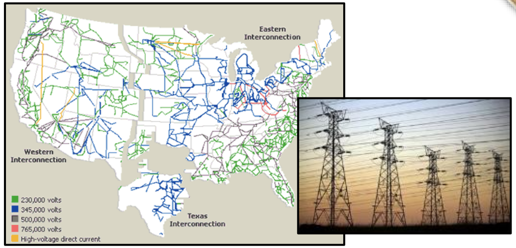
Ensuring nuclear security through stockpile stewardship



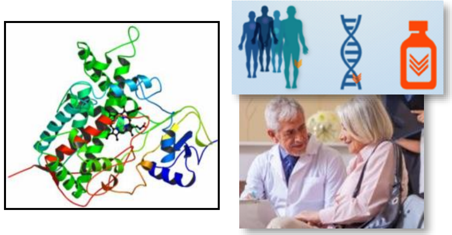
Detecting and preventing nuclear proliferation



Protecting our national critical infrastructure

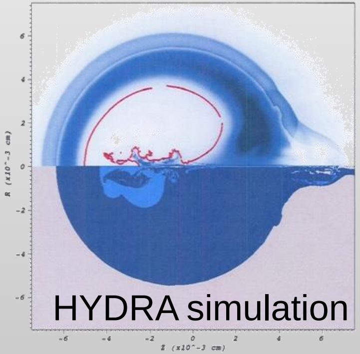


Biodefense and health security



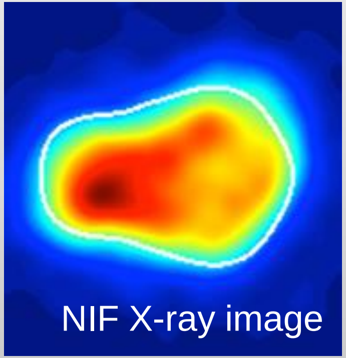
We advance our understanding by challenging our simulations with experimental observation

Traditional pillar
High-performance computing



HYDRA simulation

Traditional pillar
Large-scale experiments

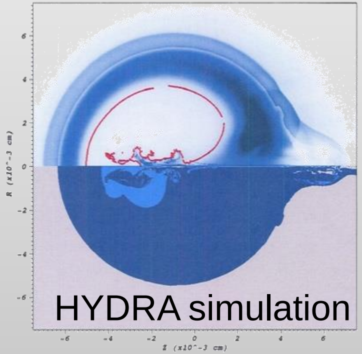


NIF X-ray image

Our successes have yielded data of overwhelming size and complexity

Traditional analysis techniques are ill-suited for modern research environments

Traditional pillar
High-performance computing



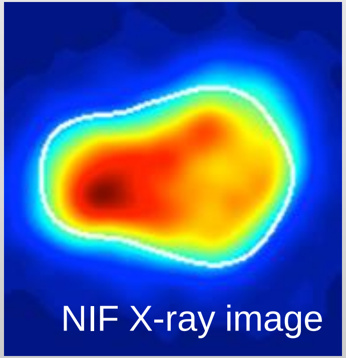
We compare a few key scalars – leaving our models less constrained

Y_{sim}
$T_{ion,sim}$
$P_{2,sim}$

$\longleftrightarrow$

Y_{expt}
$T_{ion,expt}$
$P_{2,expt}$

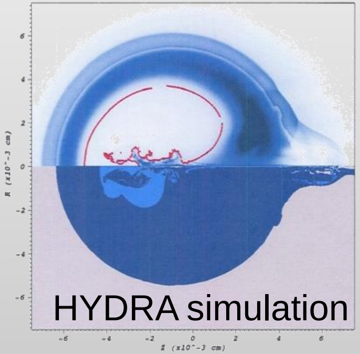
Traditional pillar
Large-scale experiments



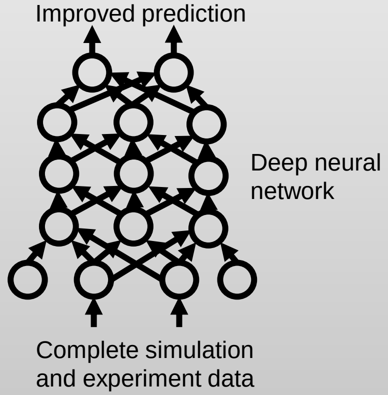
We need new techniques that improve our predictions in the presence of experimental evidence

Machine learning allows us to improve predictive modeling across applications

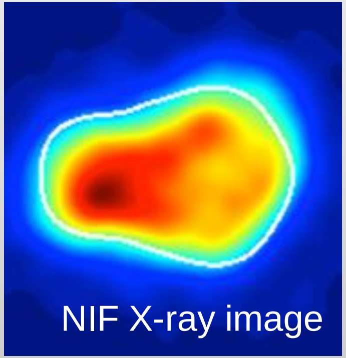
Traditional pillar
high-performance computing



<u>New pillar</u>
Machine learning to compare simulation and experiment



Traditional pillar
Large-scale experiments



Machine learning will allow us to use our full data sets to make our models more predictive

What would it take to revolutionize the way that HED science data is captured and consumed?

Self-driving laser selects a new, optimal experiment at 3 Hz

Laser control loop

Experiment performance loop

Modeling and simulation loop

1. Maximize safe laser delivery
2. Steer experiments to maximal performance
3. Select experiments to minimize simulation error

- Essential capabilities
  - Ability to use all the data
  - Quantify uncertainty in predictions
  - Detect and remove bias between simulation and experiment
  - Compute on time scales commensurate with experiment
  - Optimization strategies to seek out desired performance

CogSim brings this within reach for a new class of self-driving experimental facilities

We need three technological advances to transform our approach to predictive science

- We need improved models
  - That fully utilize available data sets
  - That estimate uncertainty
  - That improve by exposure to experimental data
- We need software tools to develop and guide those models
- We need computational platforms that support the advancement of these predictive tools

Simulation, experiment, and deep learning

Computational workflows and big data

Heterogeneous exascale computers

These advances can improve the modeling chain across programs and missions

Deep learning allows us to learn structure in data

- We use deep neural net models to map inputs to outputs
- Deep neural networks better capture rich data structure
  - Hidden layers build representation of data
  - Called a latent (or feature) space spanned by latent variables
  - Learn by minimizing the loss function (prediction matches truth)

y = f(x)

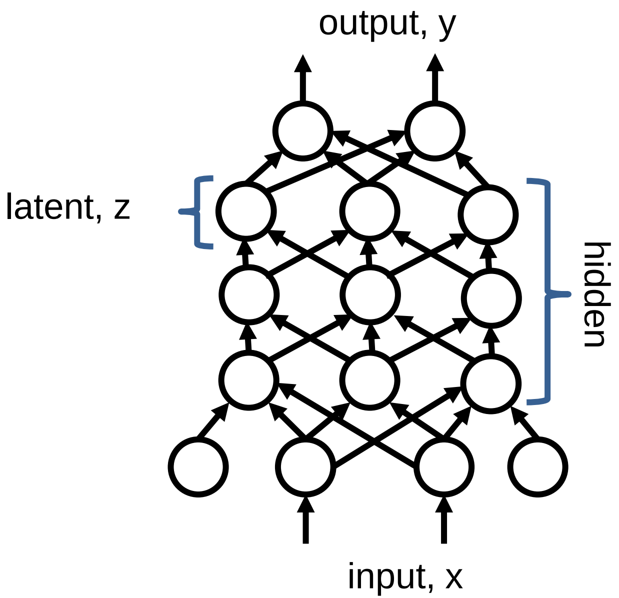
input, x
Capsule radius
Laser brightness

latent, z

output, y
Neutron yield
Ion temperature

z = h(x)

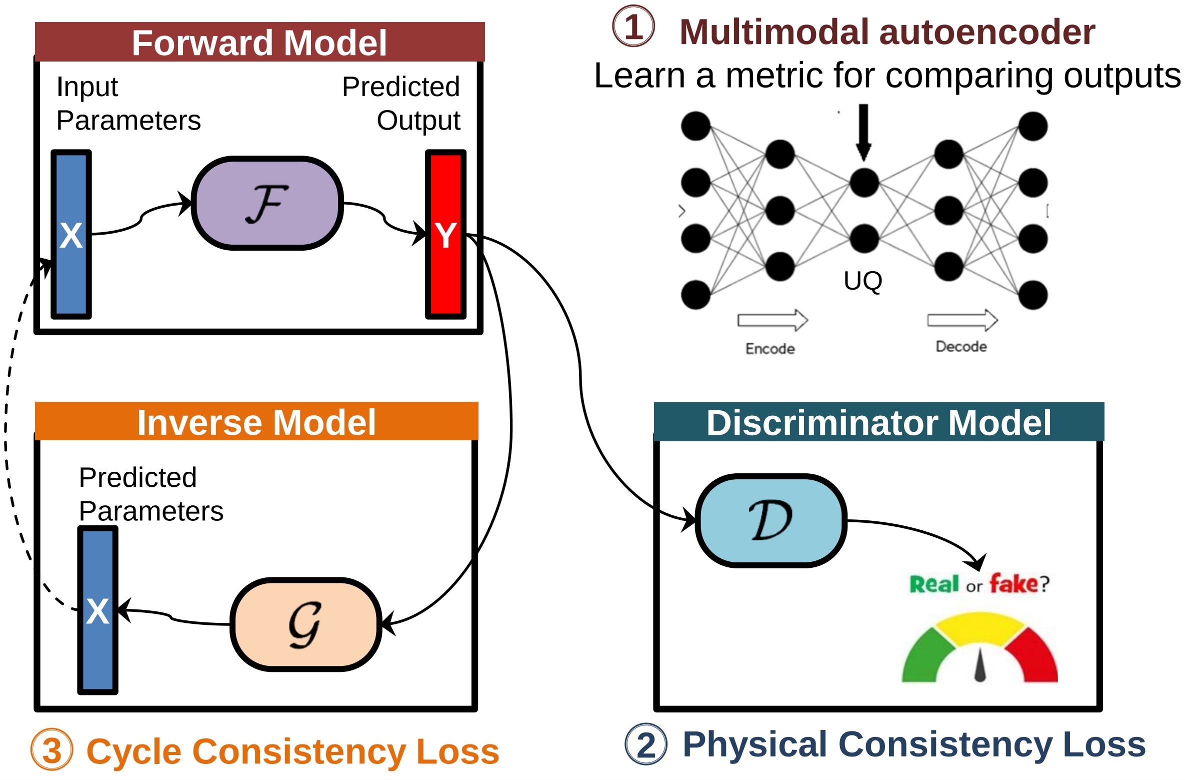
y = g(z)



Engineering and exploiting the latent space is one of our key strategies

Contemporary deep learning: a guide for practitioners in the physical sciences arXiv:1712.08523

We developed a cyclic system of sub-networks to engineer required performance features

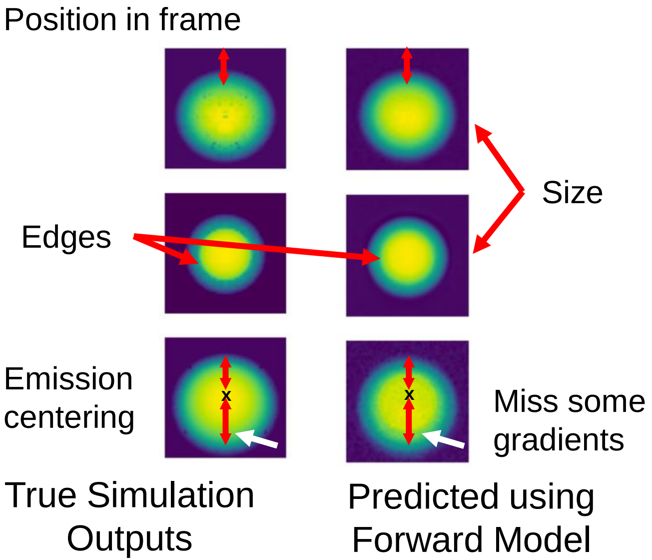


Performance features

1. **Uses all the data**
   engineers the latent space
2. **Enforces physical consistency**
   predictions look like training examples
3. **Enforces self consistency**
   regularizes ill-posed inverse

The learning system reproduces and recovers key physics information

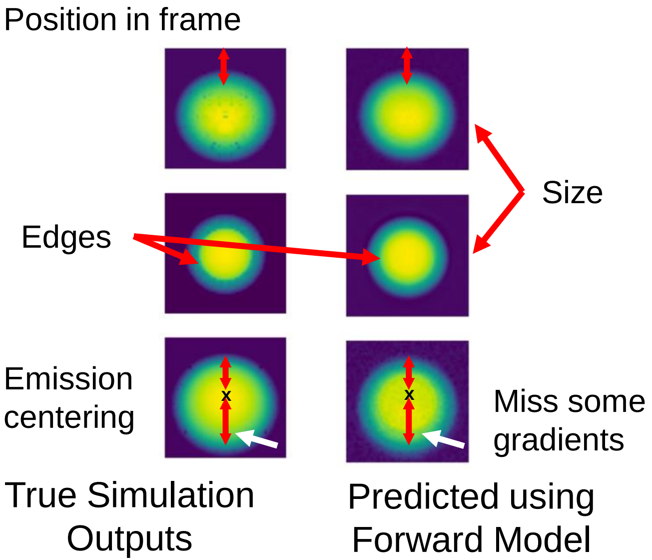
Integral features are predicted without explicit programming



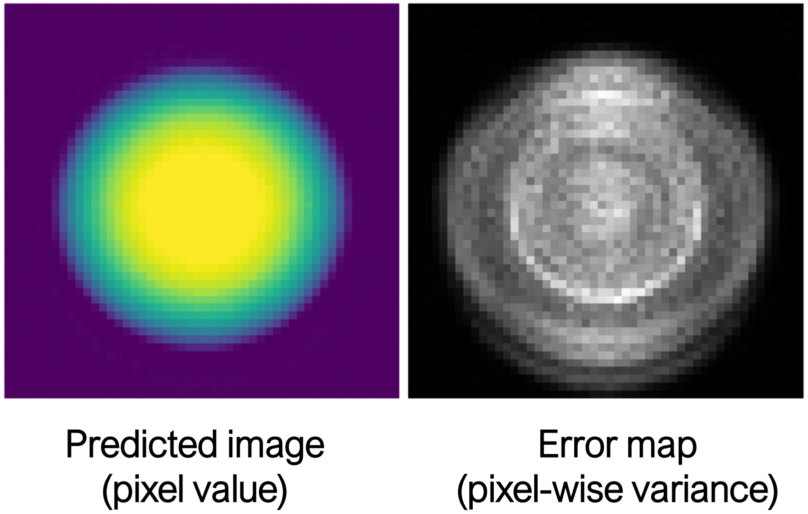
The network performance requirements have led to successful prediction

Variational extensions have equipped all output quantities with uncertainty measures

Integral features are predicted without explicit programming



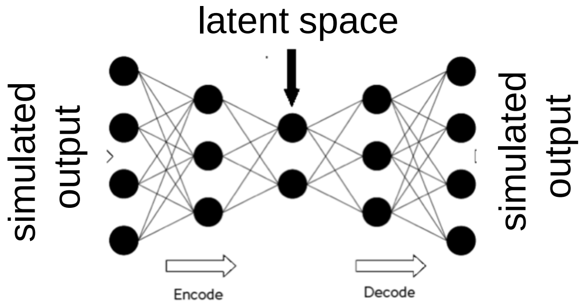
Predicted image and its error map



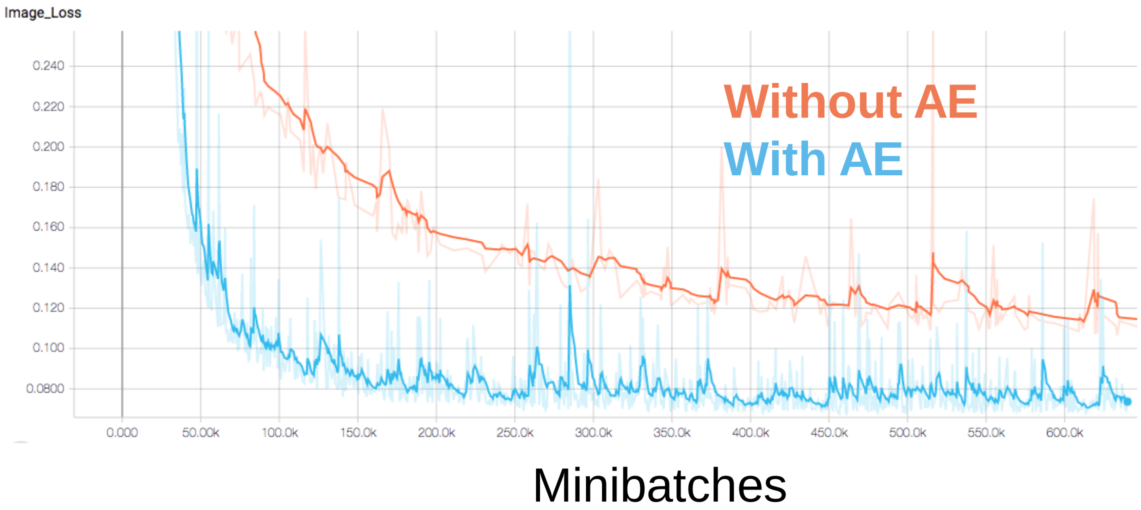
Our predictive tools are prepared for statistical comparison with experiment

Combining data through the autoencoder leads to faster training and more accurate models

unsupervised methods
inject useful correlations
from multimodal
training data



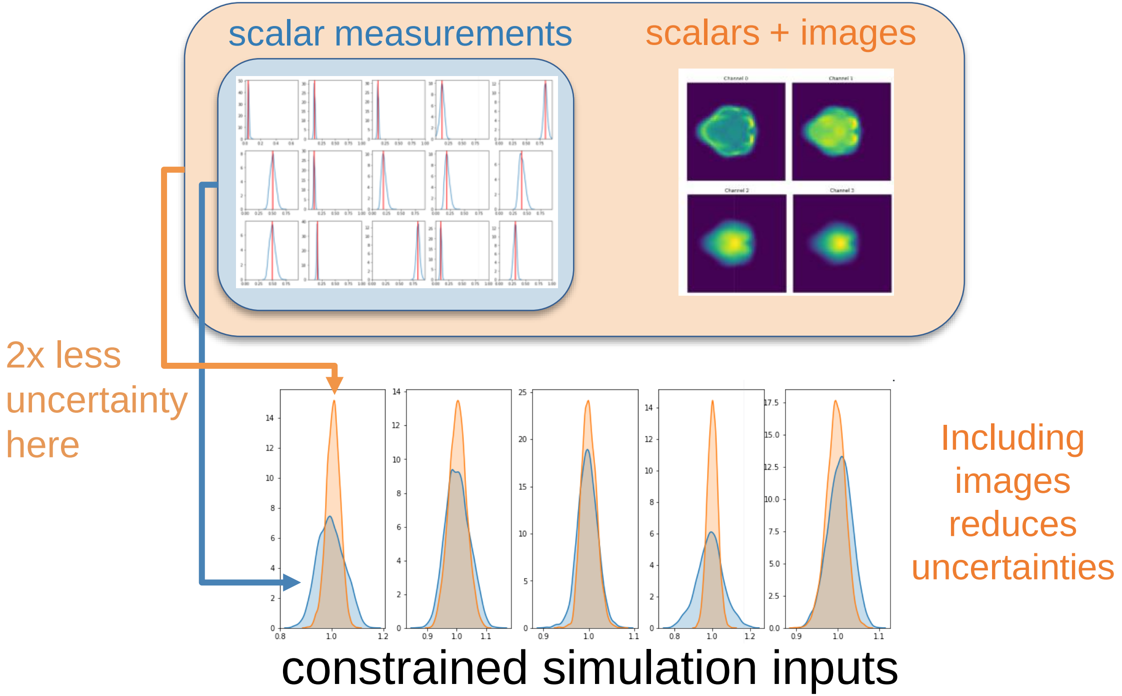
{input; latent; output}



We've confirmed our hypothesis that capitalizing on correlations in observables improves models

New Cognitive Simulation techniques allow us to use experimental data more effectively

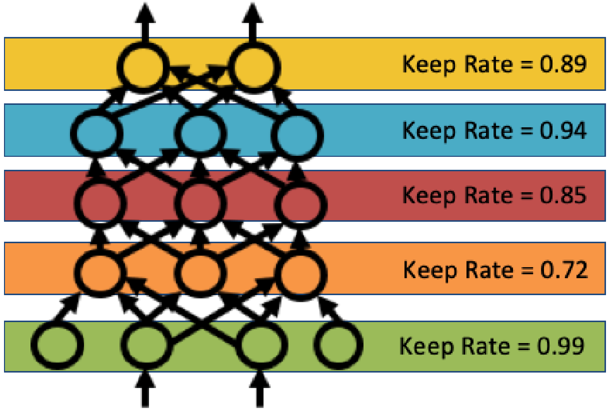
- Deep learning combines scalars and <u>complete</u> images
- Reduces uncertainties in key parameters
- Quantifies the value of new data
  - more images
  - more experiments



We're applying these techniques, developed in ICF, to other security missions

UQ for high-consequence applications requires new capabilities and scrutiny of existing ones

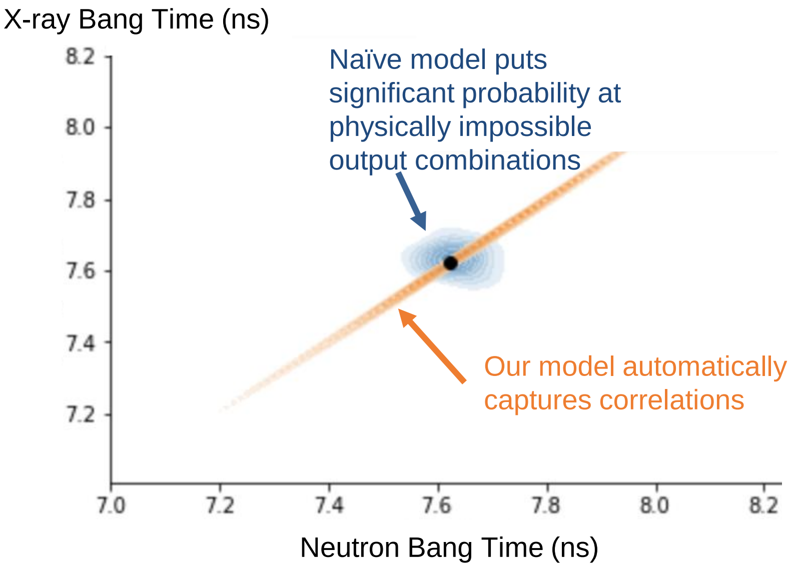
- Existing uncertainty analyses are uncalibrated!
- UQ models require validation against test data
- We need more parameters to tailor all confidence intervals
- This is complex and compute intensive
  - Search for the right combination of parameters
  - LBANN for optimal parameter search
  - Sierra for training during the search
  - Sierra or an accelerator for high-speed testing



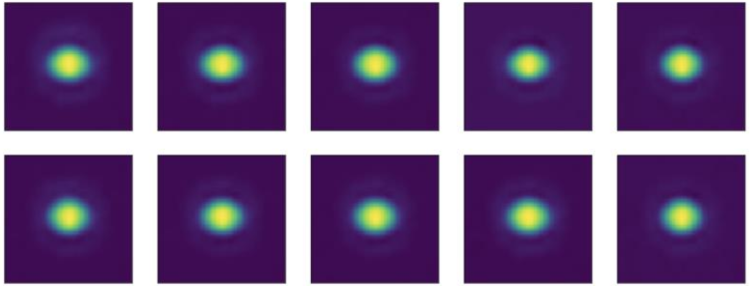
You can't even think about this kind of high-precision UQ without our flagship resources

Our CogSim UQ framework is both *calibrated* and *physically realistic*

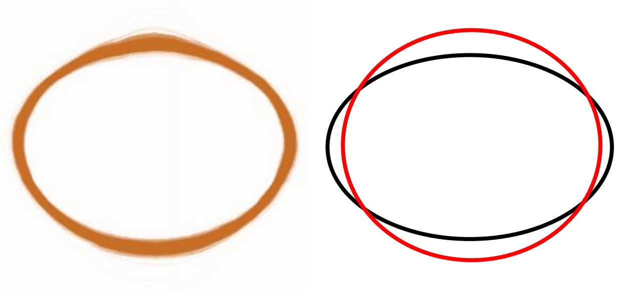
CogSim UQ allows only observations that are physically consistent with underlying simulations



Physically constent pixel-to-pixel correlations preserve important features



CogSim UQ methods capture shape variation for NIF shot N180128



These UQ methods are widely applicable across missions

Depending on the situation, networks can avoid or inherit human bias

- Learned models know what they're taught, and only what they're taught
- Humans (even scientists and engineers) can be distracted by context

Audience interaction

Find the toothbrush in 1 second!

From Heather Murphy Oct. 6, 2017 NYTimes

University of California, Santa Barbara

Audience interaction

Is there a parking meter present?


ICA

Audience interaction

Trained neural nets recognize large targets.

Humans often miss giant targets*.

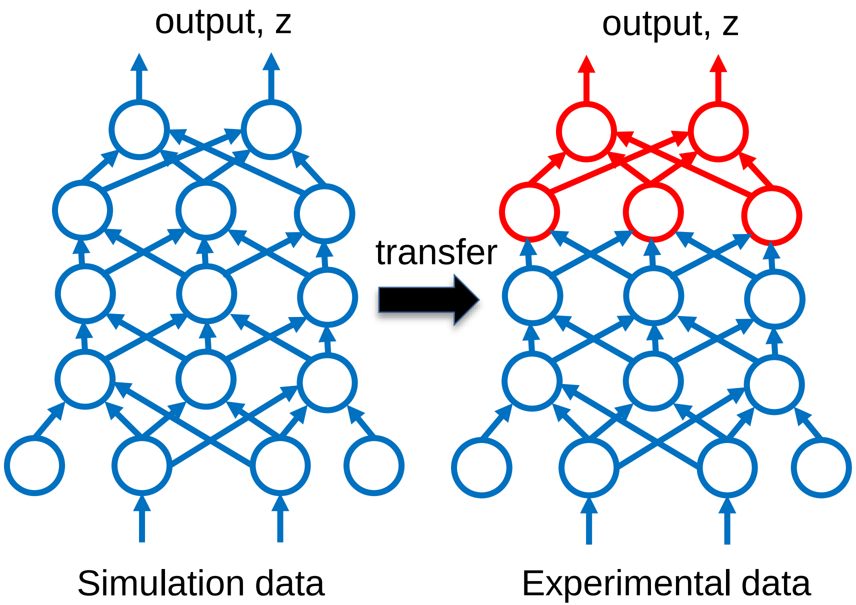
Expectations (e.g., about scale) sometimes prevent us from finding obvious patterns.

But, what if we've used our simulations to build in bias?

* "Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes" Miguel P. Eckstein, Kathryn Koehler, Lauren E. Welbourne,Emre Akbas

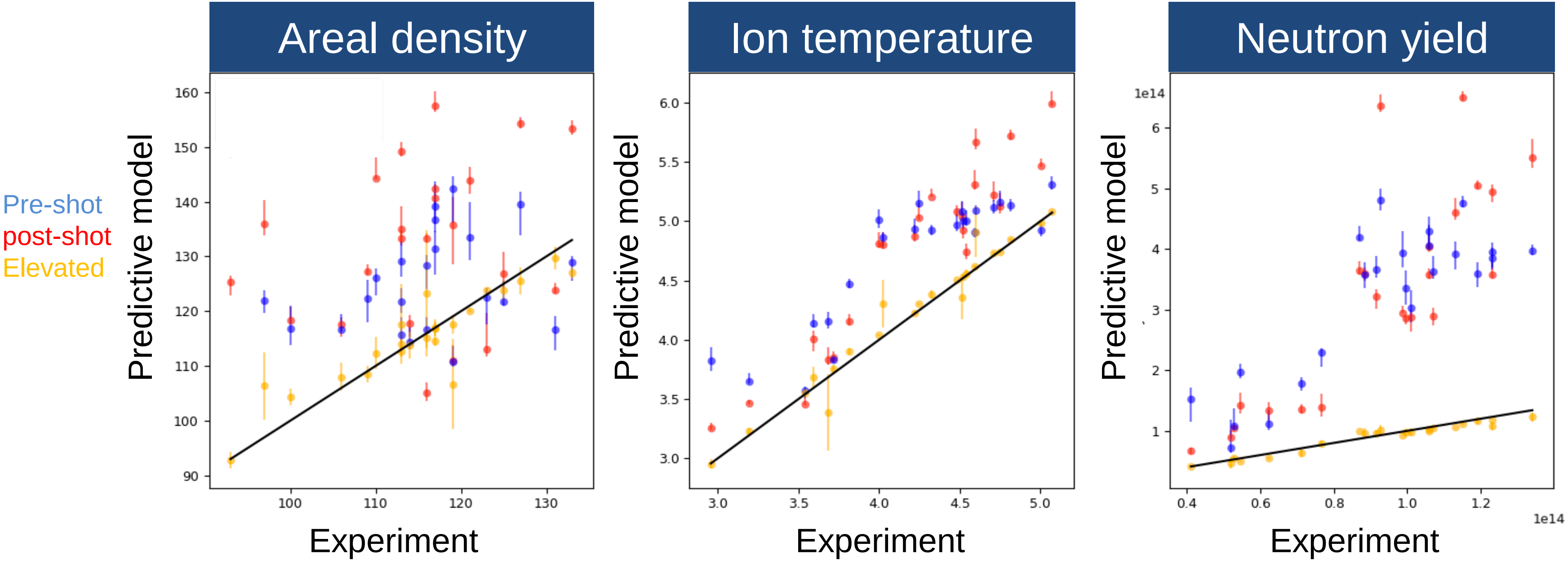
Next, we turn to transfer learning to remove simulation bias and better match experimental data

- Train the network on simulated data
- Re-train networks to predict experimental data
- Well-suited to ICF data
  - Improves prediction accuracy
  - Requires **much less data** than initial training
  - Measures discrepancy as a <u>*function*</u> of input parameters



Transfer learning produces elevated models that incorporate simulation and experiment

We can adapt our learned models to experimental data to enhance their predictive capability



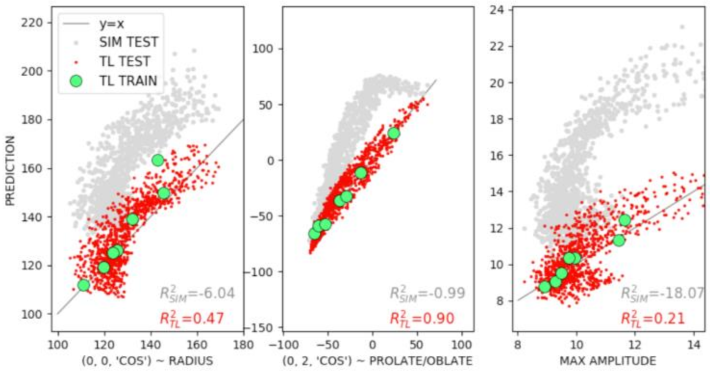
Experimental data from LLE 1D campaign

A single model holds across all shots and all observables

Recent CogSim advances can predict a broad range of scalars and images for more challenging NIF data

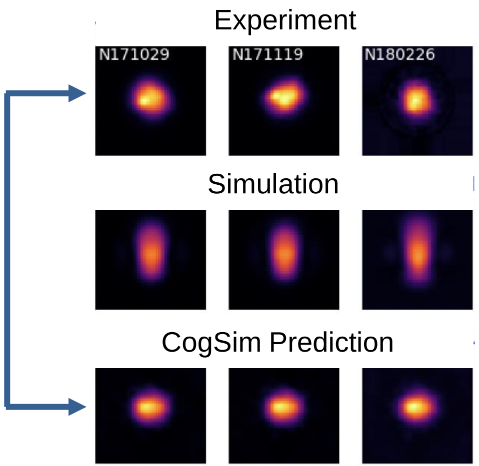
- Better model predictions with fewer NIF experiments – reduced experiment demand
- Predicts more measurement types with challenging discrepancies

Nearly unlimited scalars with strong nonlinearities



P_0 P_2 P_0

Extending to full X-ray image



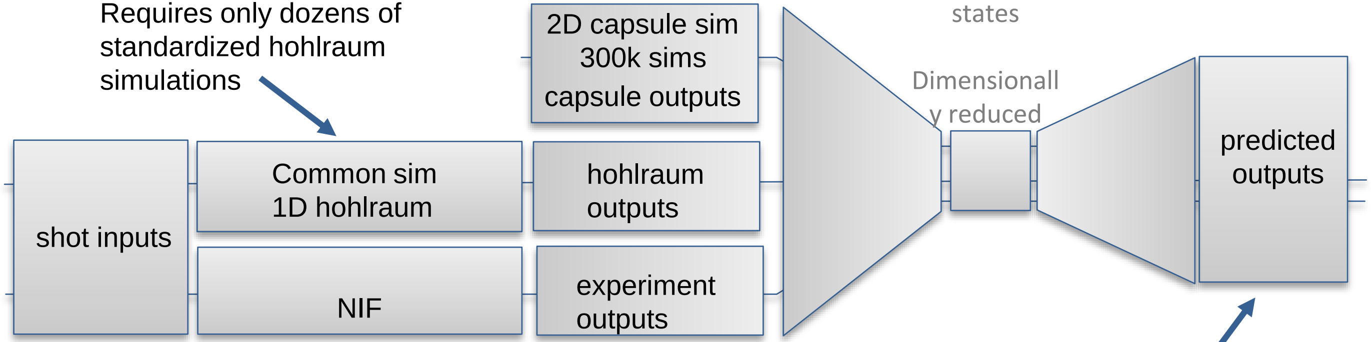
We’ve adapted our S&T tools to deliver new ICF program capabilities

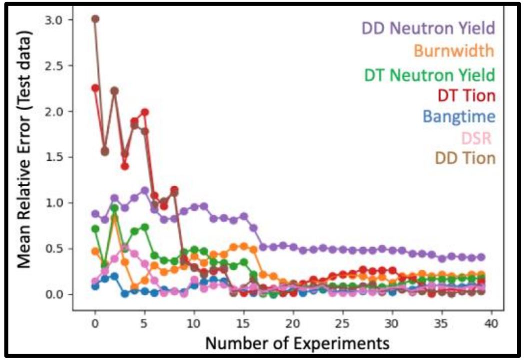
CogSim model pipeline corrects a common simulation model using data from ALL existing NIF ignition campaigns

Exploits years worth of extensive capsule simulation ensembles

Requires only dozens of standardized hohlraum simulations

Archive of imploded states





Uses all NIF layered implosions – all campaigns

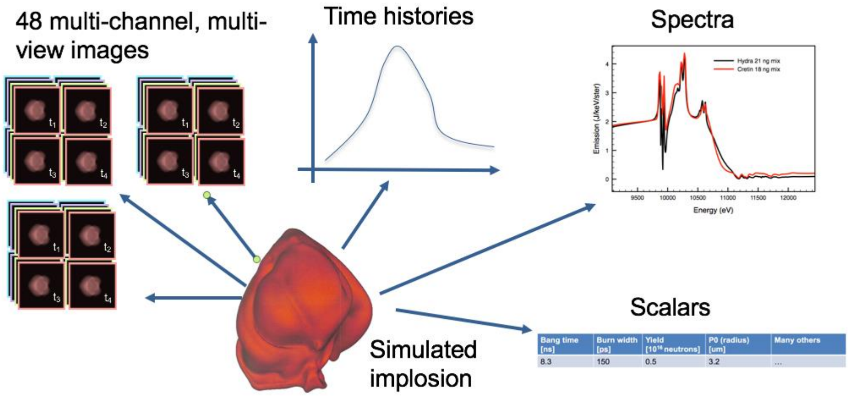
Makes better predictions that improve with additional experiments

1 decade of data

Provides a framework for tracking predictive modeling progress for both traditional simulations and CogSim models

Our newest, largest computers are enabling machine learning at an unprecedented scale

- Generated 100 million ICF implosion simulations
  - 1.5 billion scalar outputs
  - 4.8 billion images
- Built a state-of-the-art machine learning solution
- Hosting a shareable data set for scientific machine learning
- Sharing challenging and meaningful problems unique to the scientific ML community

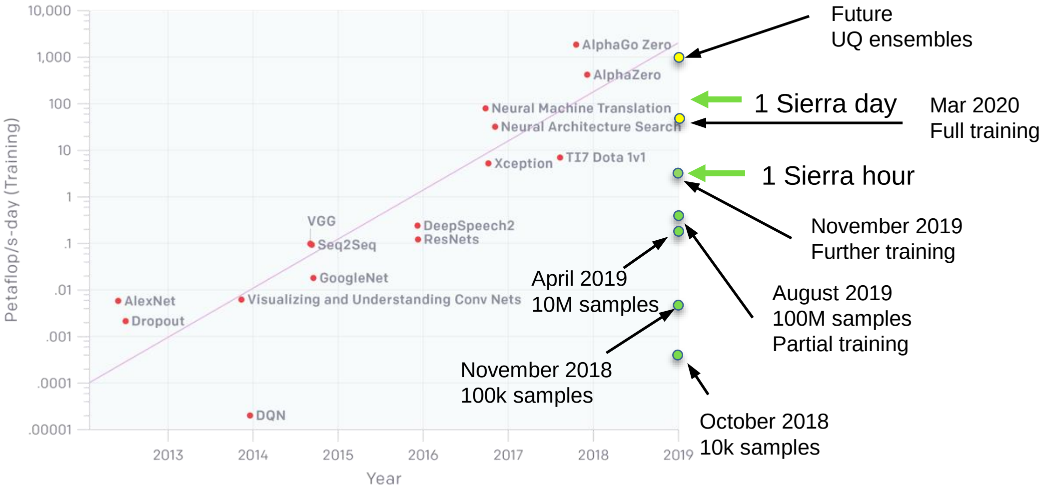


We have released this data for sharing:
https://data-science.llnl.gov/open-data-initiative

Lab technologies are operating at singular scales for applied scientific AI

- Computing needs are exploding in machine learning – doubling every 3.5 months
- Merlin, LBANN, and Sierra provide a unique capability
  - 100M simulations
  - 1.2B images and 1.5B scalars
  - Largest multi-modal network ever trained
  - Total compute rising to state-of-the-art

Models trained on Sierra have put LLNL at the state of the art



We've demonstrated AI training on all of Sierra ~ 17000 GPUs

What commerce wants from a next-generation computer may not match what science wants

Tens of billions of dollars

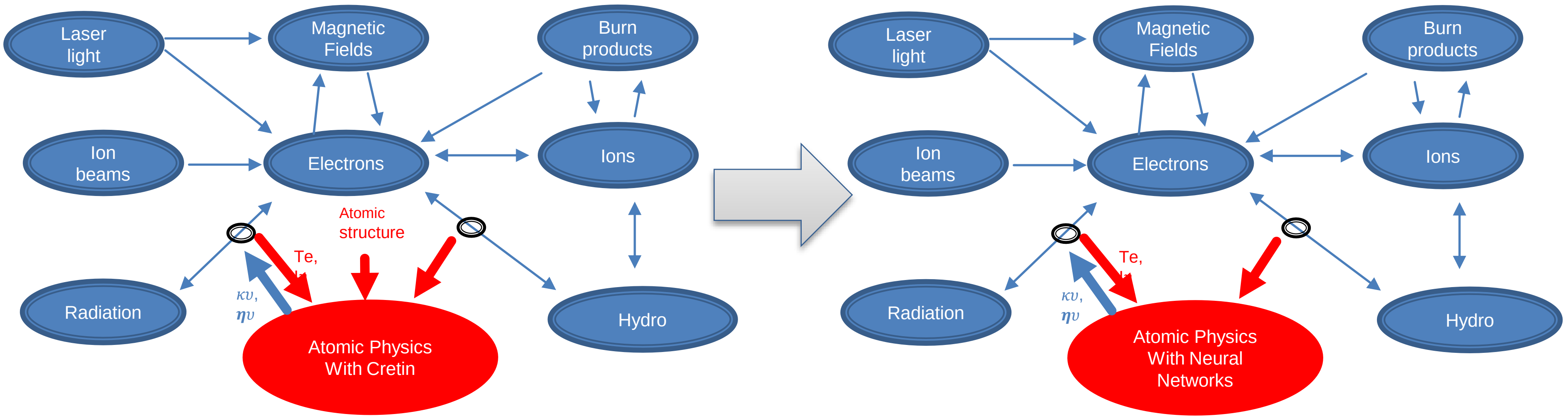
Infrequent, high-cost training
Frequent, low-cost evaluation

Frequent, low-cost training
Less-frequent evaluation?

We are using our shareable data sets to engage in co-design with vendor partners to develop machines appropriate for science

New AI-driven computing methods may change the computing architectures we're used to

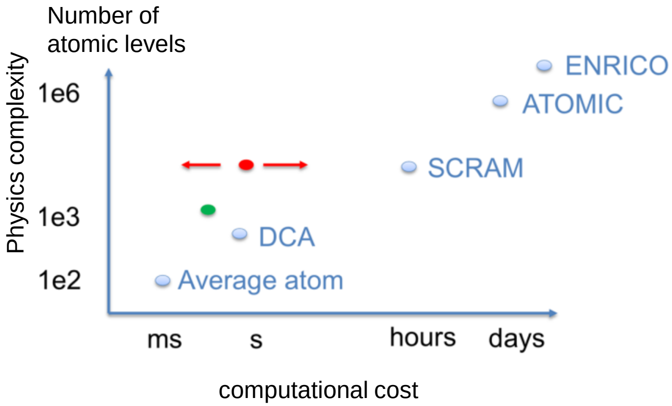
Multiscale, multiphysics simulations are expensive



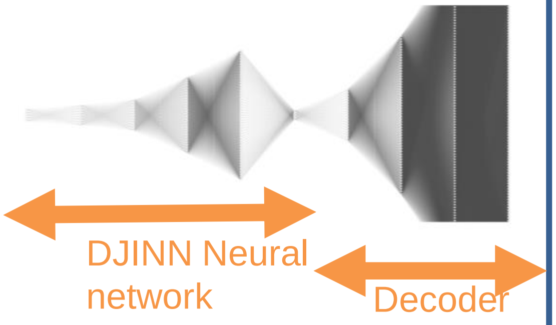
Replace expensive finite-difference physics calculation with fast AI surrogate

AI can accelerate our computing and improve our physics predictions at the same time

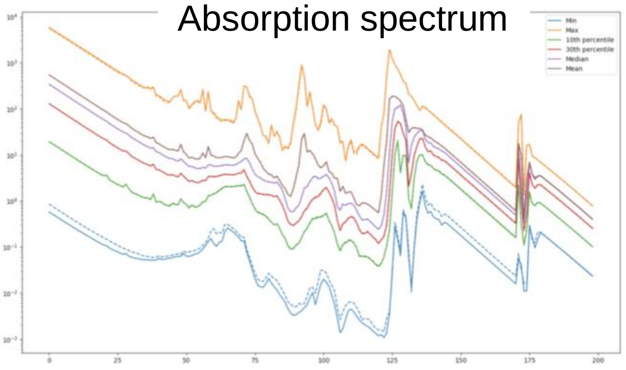
Atomic physics models are expensive



Deep neural networks learn the model



Radhydro code calls the fast DNN

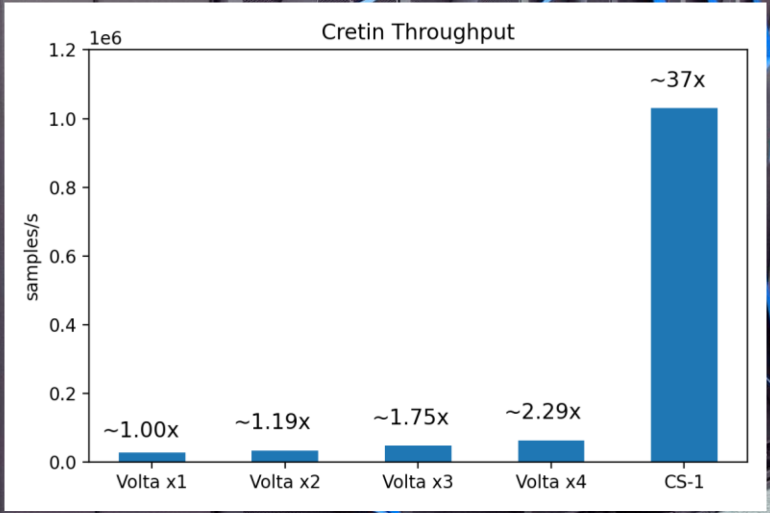


HYDRA test hohlraum simulation: **6.5x speed up**

Novel processor architectures could revolutionize the way we train and deploy this kind of model

Integrating the Cerebras CS-1 with Lassen will give the NNSA ASC Program one of the world's leading cognitive systems.

CS-1 returns CogSim physics data much faster than Sierra GPU hardware

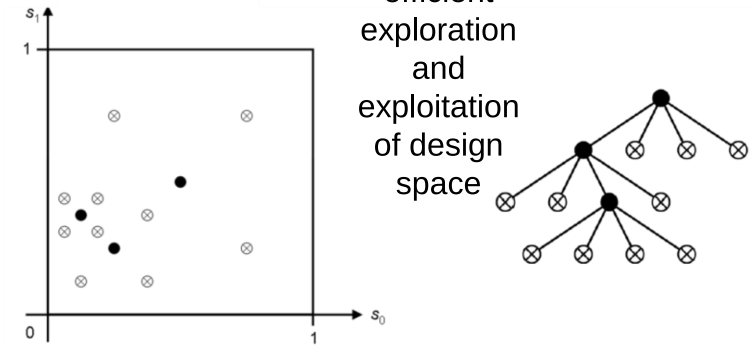


cerebras
ASC
Lawrence Livermore National Laboratory

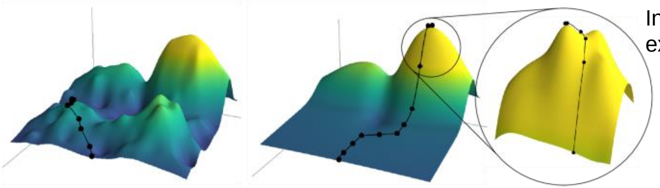
CogSim design optimization strategies will enable faster design in rich design spaces that humans can’t navigate

New methods to optimize complex designs in higher dimensions

efficient exploration and exploitation of design space



New primed CogSim models that support advanced design optimization



Naïve CogSim model

Primed CogSim models for exploration/exploitation

Design optimization benefits numerous projects and long-range plans

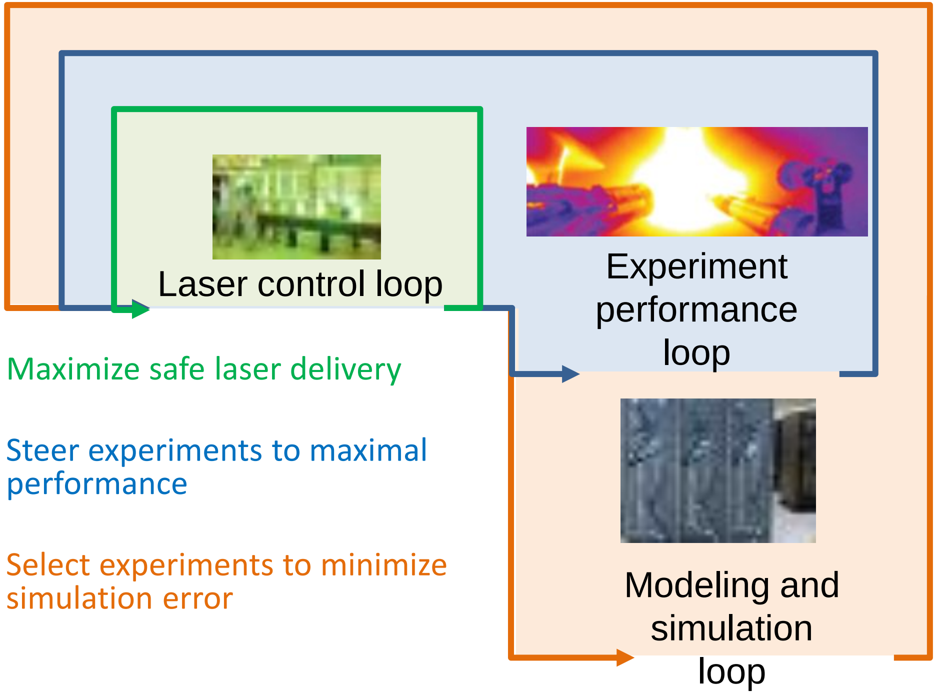
HRR lasers, ICF, stockpile projects, therapeutics design , and more

Our CogSim leadership helped capture funding for an ambitious project in high-repetition-rate lasers – *project snowball*

- Essential capabilities
  - Use all the data
  - Quantify uncertainty in predictions
  - Detect and remove bias between simulation and experiment
  - Compute on time scales commensurate with experiment
  - Optimization strategies to seek out desired performance

Now the hard part:
bolting this together to do science

Self-driving laser selects a new, optimal experiment at 3 Hz



1. Maximize safe laser delivery
2. Steer experiments to maximal performance
3. Select experiments to minimize simulation error

Lawrence Livermore National Laboratory

The LLNL Data Science Institute focuses on growing and strengthening LLNL's Data Science workforce.

- Data Science Summer Institute
- Data Science Institute endorsed training programs and courses
- Targeted recruiting and university collaboration
- Community outreach through seminar series, workshops, competitions, and web presence

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

ICME
Stanford | Institute for Computational & Mathematical Engineering

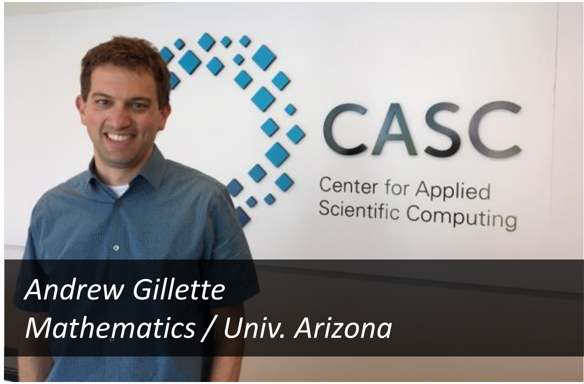
UC San Diego
JACOBS SCHOOL OF ENGINEERING

https://data-science.llnl.gov

datascience@llnl.gov

New pilot Faculty Mini-Sabbatical Program

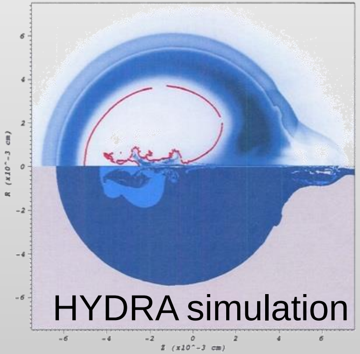
- Designed to increase the number of faculty–staff research partnerships and strengthen our S&T by bringing in top academic talent
  - Faculty hired 1–3 months
  - Hosted by staff scientist and approved by committee
  - Paid a monthly salary and travel costs
  - Faculty learns new research capabilities and gains greater knowledge set
- LLNL has an existing sabbatical program for staff
  - Salary paid for up to 1 year to visit universities



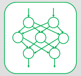
For more information visit https://st.llnl.gov/about-us/university-relations/faculty-sabbatical-program

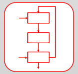
We are advancing the way we develop predictive models using large-scale scientific machine learning

Traditional pillar
high-performance computing



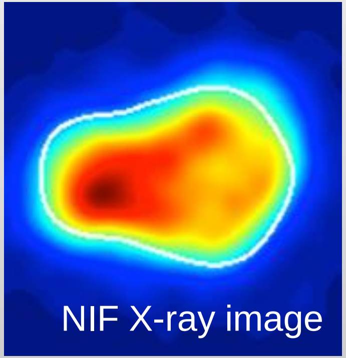
<u>New pillar</u>
Machine learning to improve predictive science


Deep learning to improve prediction


Advanced workflows to support learning


New architectures for modern computation

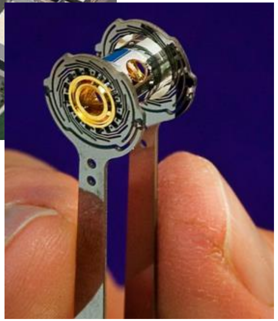
Traditional pillar
Large-scale experiments

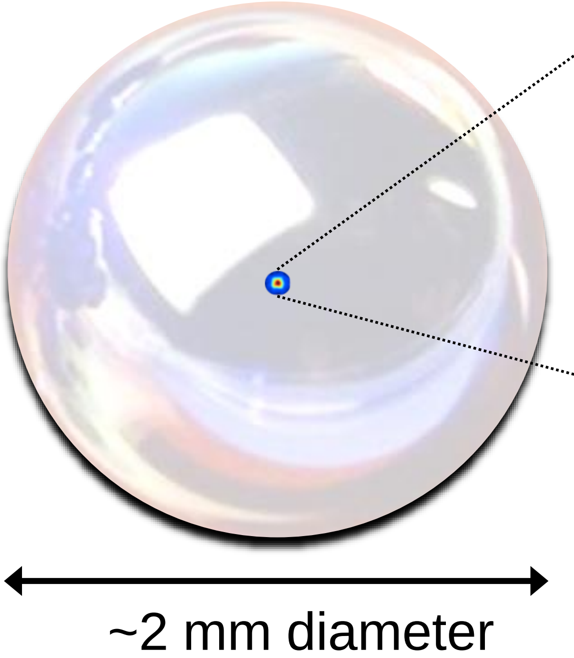


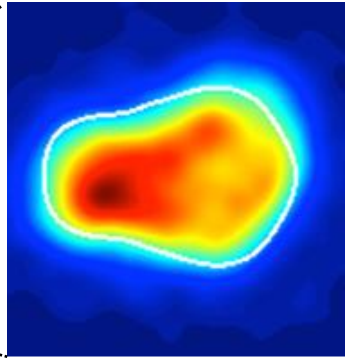
We're developing these techniques for a range of critical missions, and we need more of the best and the brightest

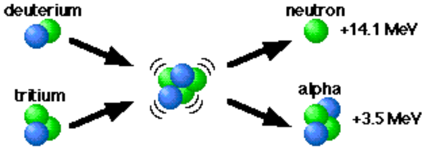
spears9@llnl.gov

Inertial confinement fusion (ICF) is a perfect testbed for our AI development









We use incredibly sophisticated simulations and experiments to understand laser-driven fusion

I am proud to present the work of a wonderful team

Machine-learned Predictive Models
PI, Brian Spears

Machine Learning Element
Timo Bremer

Architectures
Jay Thiagarajan
Rushil Anirudh
Shusen Liu

Elevation and UQ
Jim Gaffney
Bogdan Kustowski
Gemma Anderson
Francisco Beltran
Michael Kruse

Large-scale Learning
Sam Ade Jacobs
Brian Van Essen
David Hysom
Jae-Sung Yeom

Workflow Element
Luc Peterson

Workflow Tools
Peter Robinson
Jessica Semler
Luc Peterson
Ben Bay
Scott Brandon

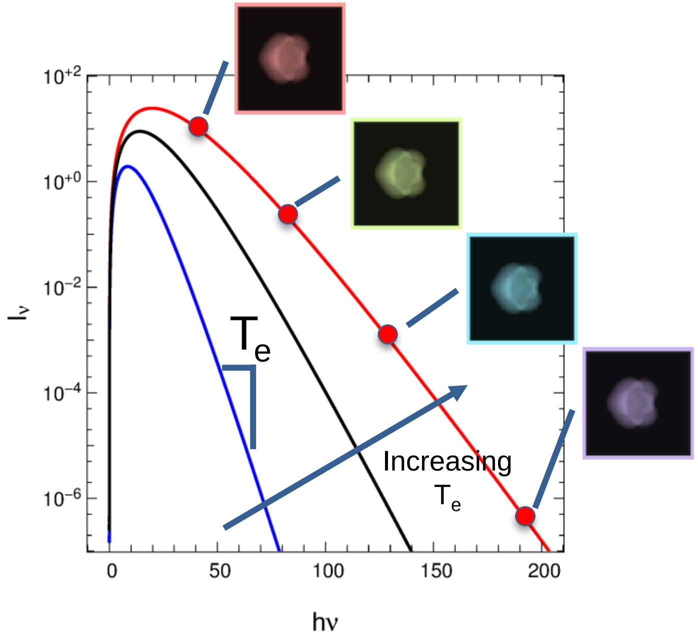
Intelligent Sampling
Vic Castillo
Bogdan Kustowski
Kelli Humbird
David Domyancic
Richard Klein

In-situ tools
John Field
Steve Langer
Joe Koning

Data Harvesting
Michael Kruse
Dave Munro
Robert Hatarik

What does it mean for an AI prediction to be "physical"?

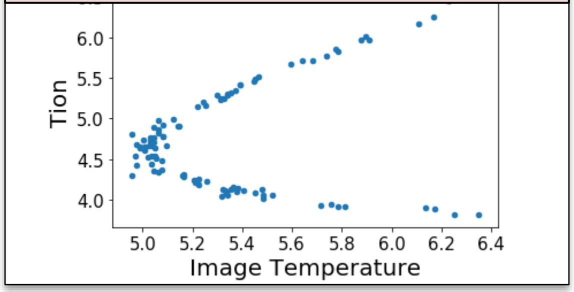
- The prediction should
  - Get the right answer
  - Respect physical laws
- An example
  - Predictions match simulations
    - Predicted images look like simulated images
    - Predicted Tion is close to simulated Tion
  - Predictions are physical
    - Temperature inferred from *predicted* images matches *predicted* Tion



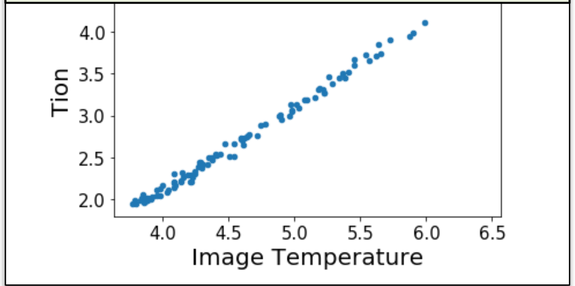
$T_e \sim T_i$

Physical relationships can guide performance improvement

Ad hoc regularization



Optimized constraint matching



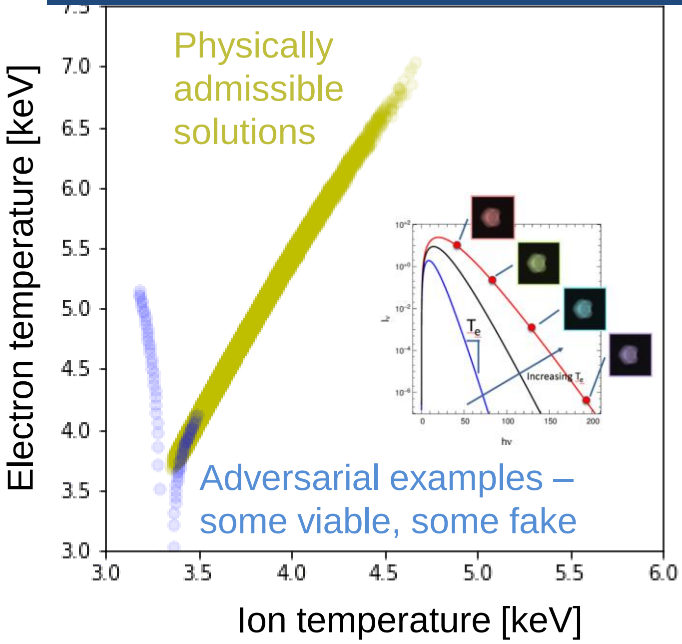
Loss = reconstruction + λ_1*adversary + λ_2*cycle

- Can we formulate physical constraints that we demand to be respected?
- Can we force models to respect physical constraints exactly?
- Should we force models to respect these constraints?

We can both generate and detect physics "deepfakes"

- We're interrogating exceptionally complicated neural networks to make them interpretable for physics
- Some model states are accessible by simulation, some aren't
- We aim to place constraints ***inside*** the model

Artificial predictions must satisfy physics



Deep learning needs big data, so we'd better be able to produce it

Merlin

Hierarchical Ensembles of HPC Simulations

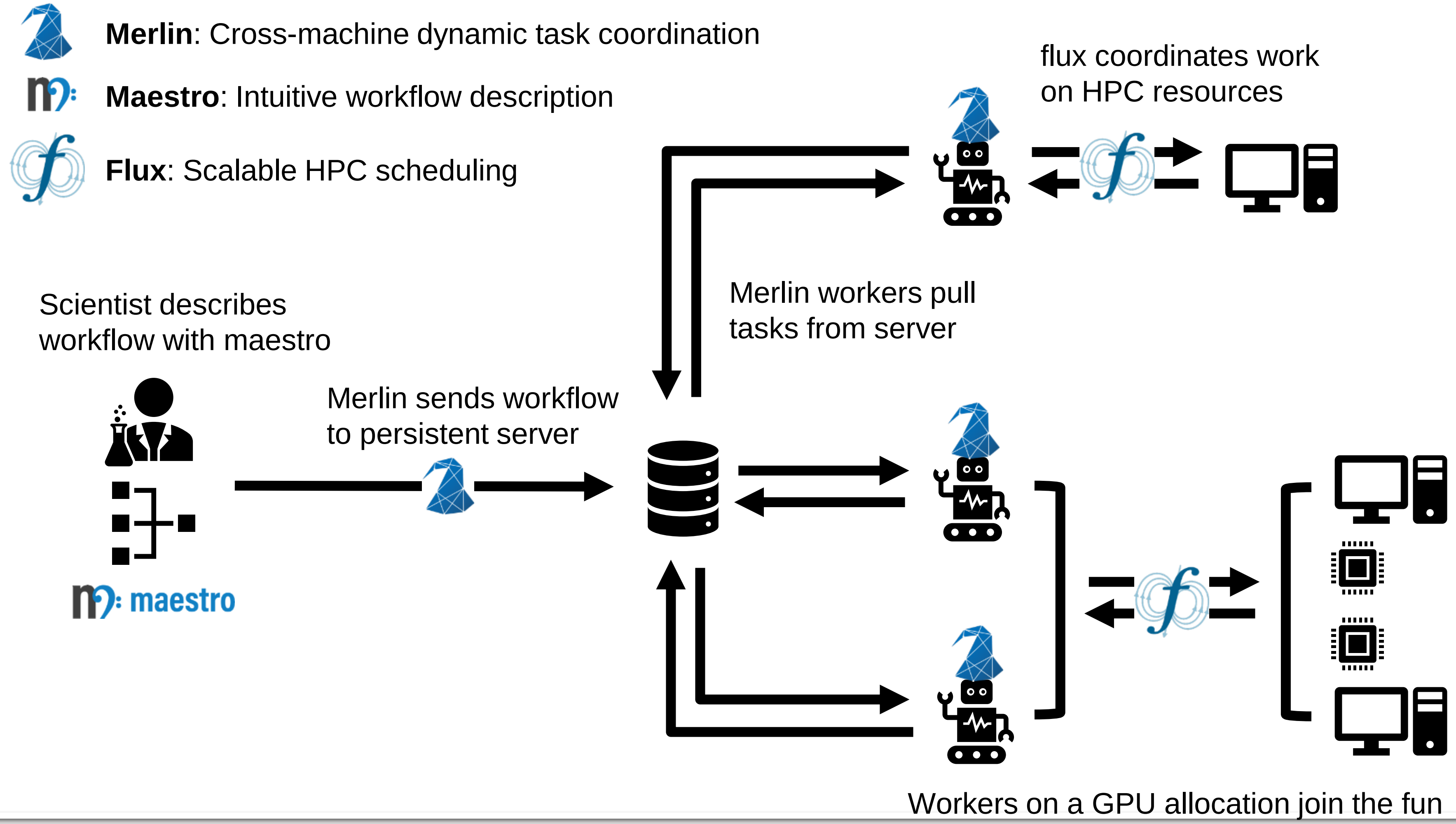
In-situ postprocessing

Deep Learning at Scale

Intelligent Sampling

Scalable Data Exploration

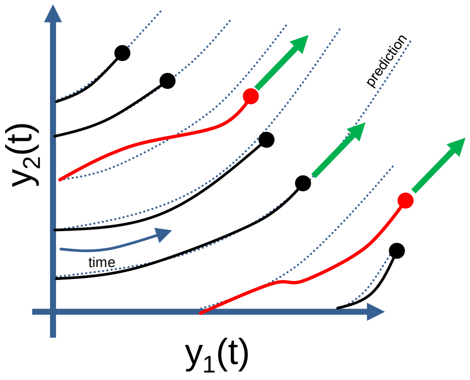
Merlin is a custom workflow tool for driving large-scale simulation and machine learning
https://github.com/LLNL/merlin


Merlin: Cross-machine dynamic task coordination
Maestro: Intuitive workflow description
Flux: Scalable HPC scheduling
flux coordinates work on HPC resources
Scientist describes workflow with maestro
Merlin workers pull tasks from server
Merlin sends workflow to persistent server
maestro
Workers on a GPU allocation join the fun

Even at very large scale, we must choose carefully which simulations to execute

Speculative sampling

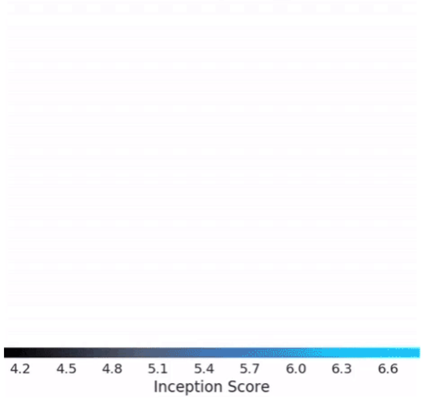
- During the run, is the simulation evolving as predicted?
  - Yes? No new information. Terminate. Invest in a new simulation.
  - No? Unpredicted behavior! Continue.
- Speculate on many more simulations than we can finish.
- More completely probe parameter space for further cost reduction.



Speculative sampling may require far less data than random sampling

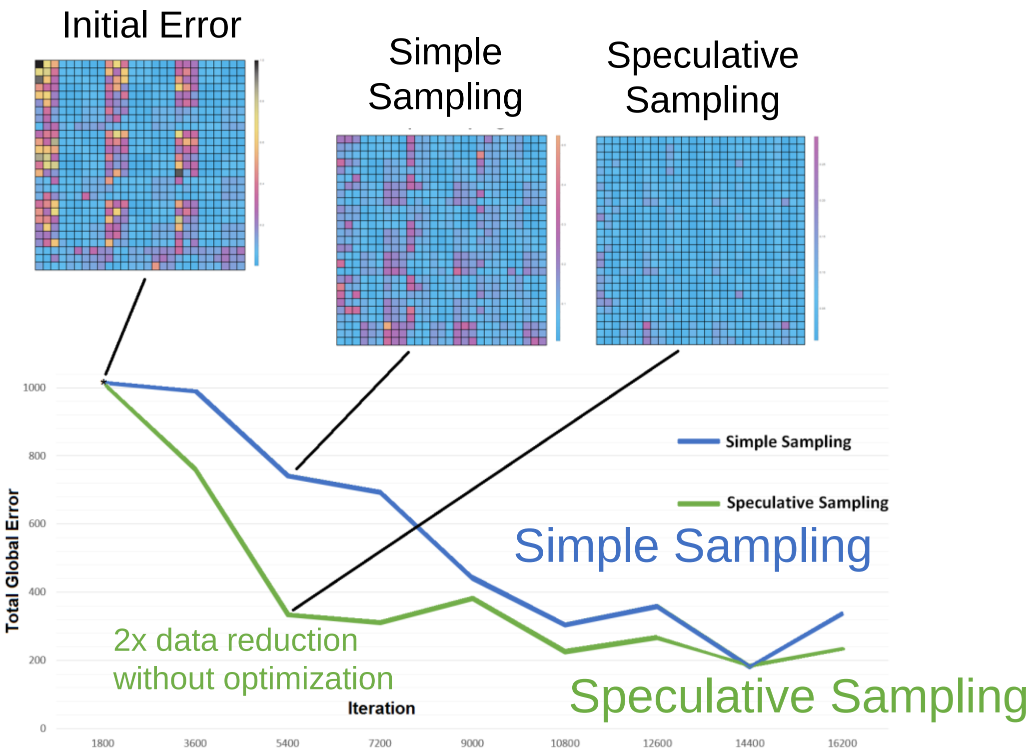
Initial speculative sampling experiments delivered better learned models for much less data

Agent-based exploration
Deep Mind, November, 2017



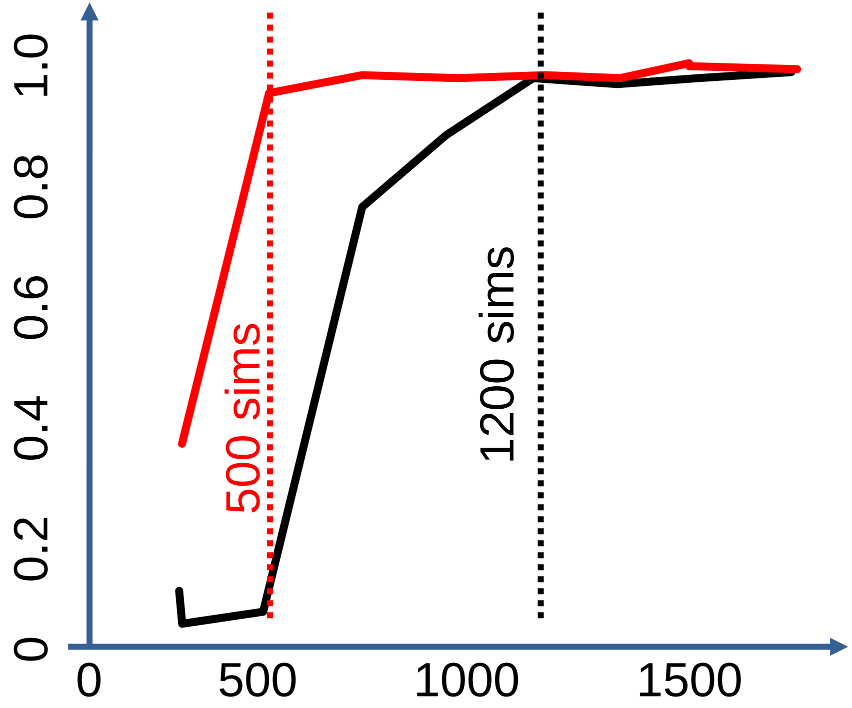
Model error in local regions

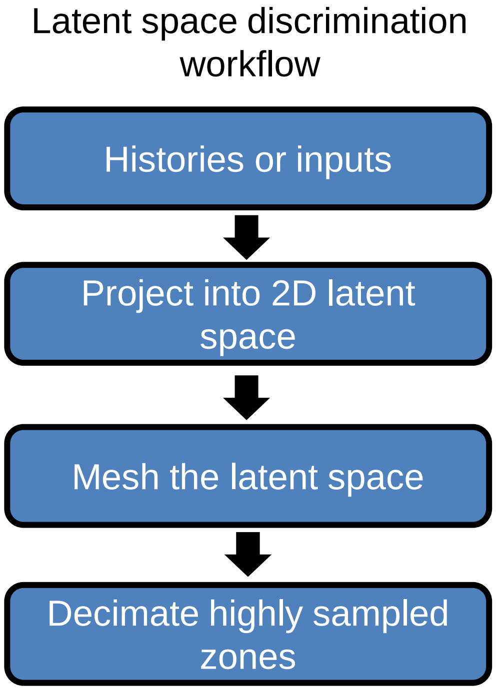
Total Global Model Error vs Iteration



Castillo *et al.*

We've applied speculation to in-flight radiation hydrodynamics simulations



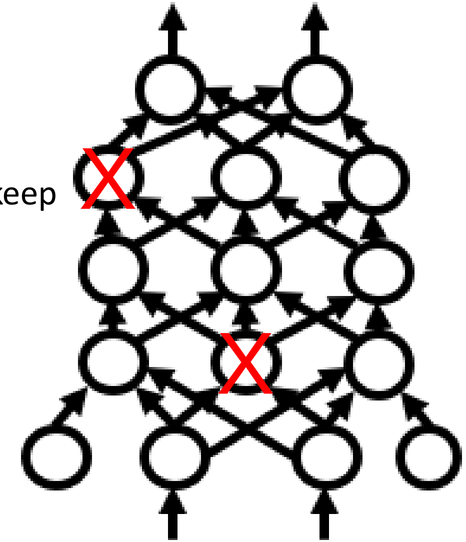


Speculative sampling requires 60% fewer radhydro simulations

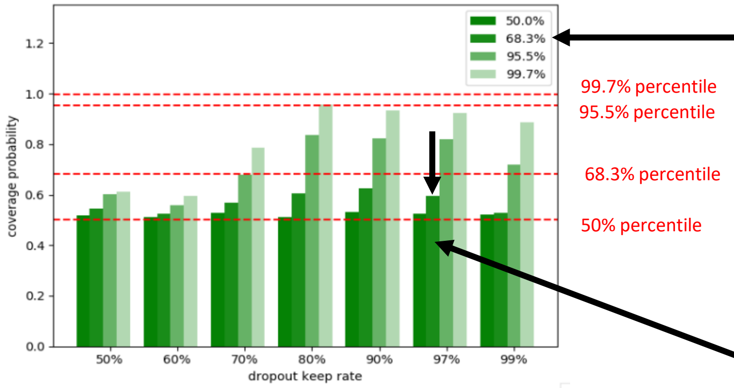
UQ for science requires new capabilities and scrutiny of existing ones

- Existing uncertainty analyses are uncalibrated!
- UQ models require validation against test data

Dropout keep rate = 0.9



Predicted confidence intervals must be tuned to be correct



Predictive model believes that 68.3% of the time values will fall within a specified range

In testing, values fall within the 68.3% confidence interval only 52% of the time.

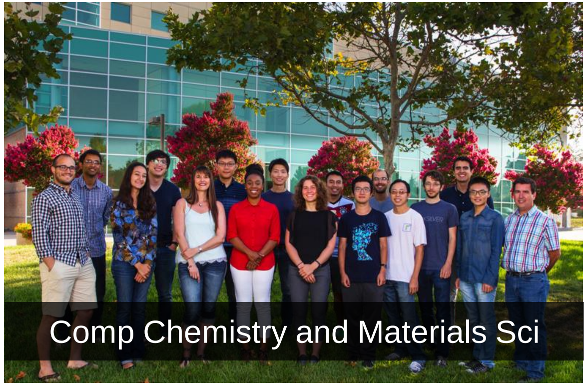
In 2019, more than 1,150 students engaged in research at LLNL that focused on our core mission areas

- Nuclear Forensics Summer Program
- Data Science Summer Institute
- Computational Chemistry and Materials Science Summer School
- Computation Scholar Program
- HED Science and WCI Summer Programs
- DHS Global Security Summer Program
- DOE Science Undergraduate Laboratory Internship (SULI)
- Science undergraduate lab interns


Seaborg Institute Nuclear Forensics


Cyber Defenders


Radiochemistry


Comp Chemistry and Materials Sci

For more information email kersting1@llnl.gov or visit https://st.llnl.gov/opportunities/student-opportunities

LLNL postdoc program

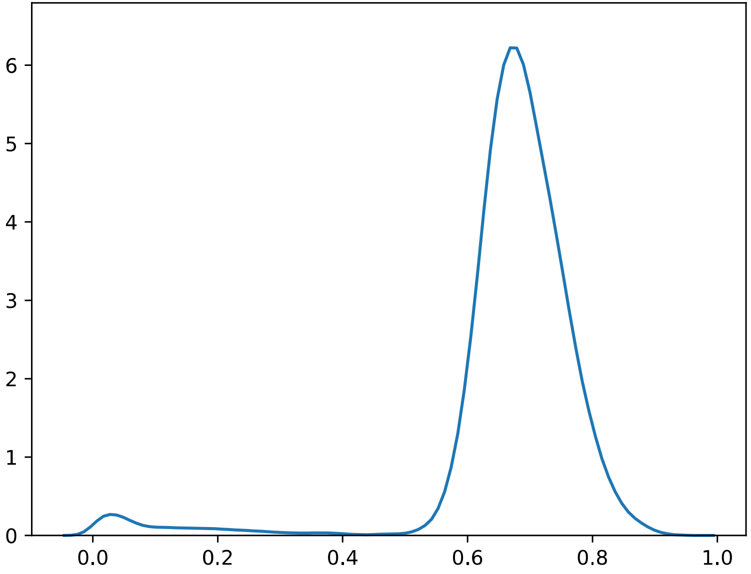
- Professional development
  - Research that is complementary to funded project
  - Maintain university collaborations
  - Travel and professional training activities
- LLNL culture
  - Networking and team building
  - Postdocs allowed to PI grants
  - Publishing is a priority
- Emphasis on mentoring
  - One-on-one meetings to help postdocs succeed


2017 Diving Dog Event

For more information email kulp1@llnl.gov or visit https://st.llnl.gov/opportunities/postdocs

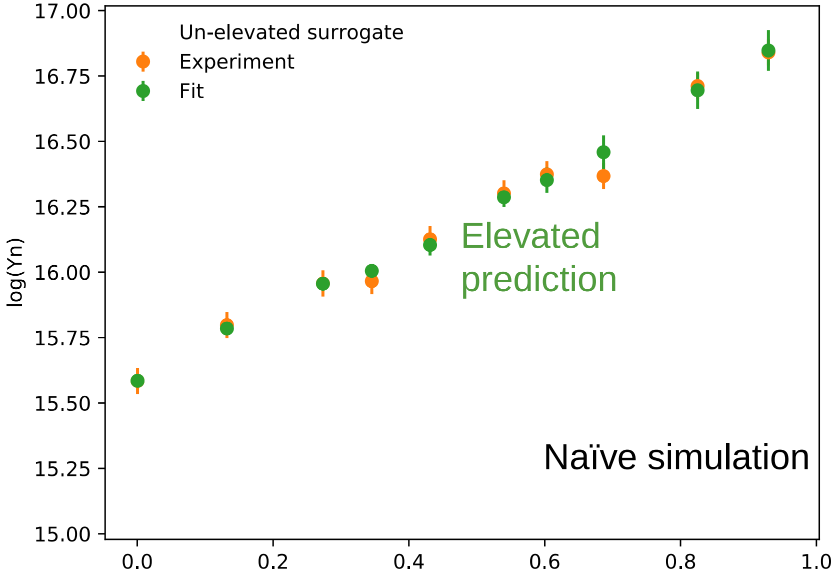
Coupling model elevation and calibrated UQ represents a capstone achievement for cognitive simulation

Inference and UQ for uncertain physics parameters



scaled compressibility

Elevated AI model that matches data AND is consistent with inferred physics



scaled implosion velocity