Deep Learning/AI of Accelerated Advances in Fusion Energy Science Disruption Predictions with Implications for Plasma Control

William M. Tang

Princeton University/Princeton Plasma Physics Laboratory (PPPL)

ISC High Performance: International Supercomputing Conference
Frankfurt, Germany

Invited Presentation: ISC-Deep Learning for Science Workshop

June 20, 2019

Princeton University / PPPL Fusion AI/DL/ Team:

Julian Kates-Harbeck (Harvard U/PPPL), Alexey Svyatkovskiy (Princeton /Microsoft)

Eliot Feibush (PPPL/Princeton), Kyle Felker (Princeton/ANL), Ge Dong (Princeton/PPPL), Dan Boyer (PPPL) & Keith Erickson (PPPL)

Artificial Intelligence (AI)

Context: F. Chollet (Google)

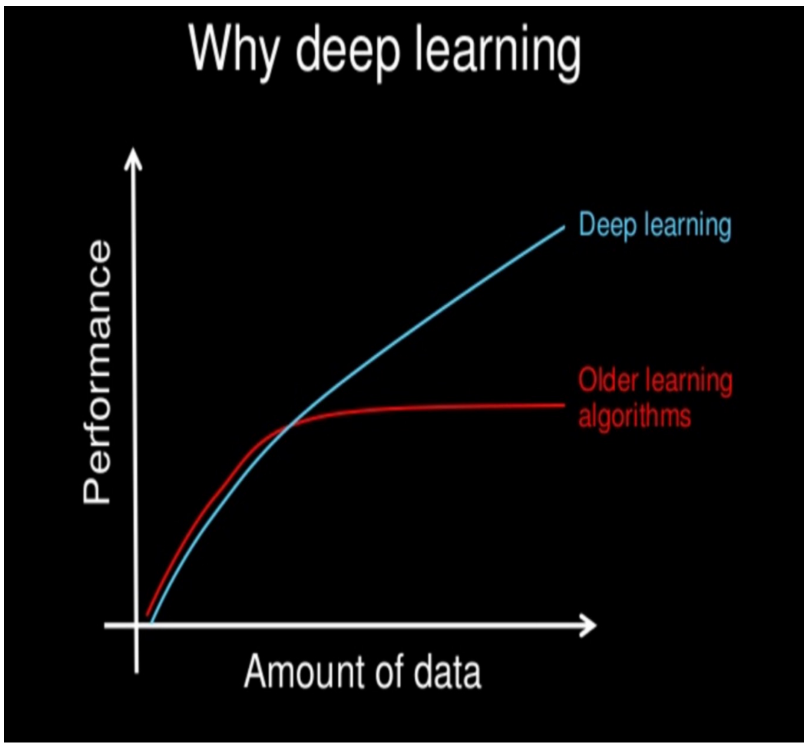
"*Deep Learning with Python,*" *Nov. 2018*

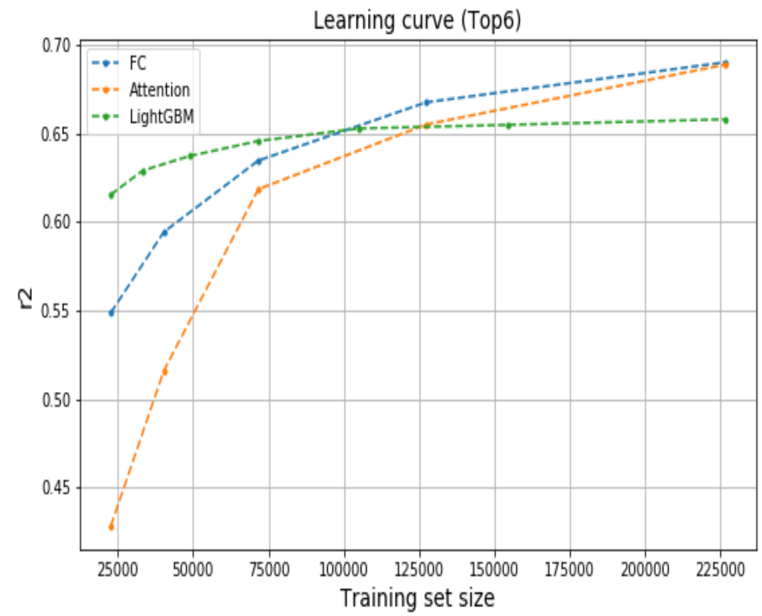
"Automation of intellectual tasks normally performed by humans"

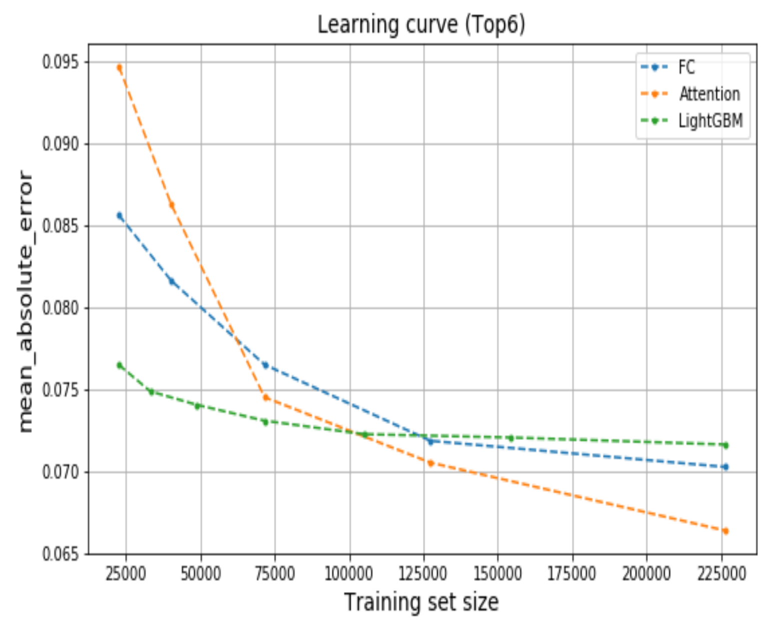
- *general area including Machine and Deep Learning (ML/DL)*
- *Machine Learning (ML) – focus on training rather than explicit programming*
- *Deep Learning (DL): Focus on complex data sets with temporal images including multi-pixels*

➔ *Requires deployment of stacks of modern Convolutional & Recurrent Neural Networks*

➔ *Automated-search ("Hyperparameter Tuning") usually required for best representations*







Reference: Rick Stevens (ANL/U.Chicago)
2019 International Symposium on Simulation, Big-Data, & AI, Kobe, Japan

APPLICATION FOCUS FOR AI/DL IN FES

Most Critical Problem for Fusion Energy →

Accurately predict, mitigate, & ideally avoid large-scale major disruptions in magnetically-confined thermonuclear plasmas such as the ITER –the $25B international burning plasma "tokamak"

•Most Effective Approach: Use of big-data-driven statistical/machine-learning predictions guided by observations for the occurrence of disruptions in world-leading facilities such as EUROFUSION "Joint European Torus (JET)" in UK, DIII-D (US), and other tokamaks worldwide such as KSTAR, EAST, JT60-SA (Asia)

•Recent Status: ~10 years of R&D results (led by JET) using Machine Learning (via Support Vector Machines) on *zero-D (scalar)* time trace data executed on CPU clusters yielding *success rates in mid-80 to 90% range for JET 30 ms before disruptions,*

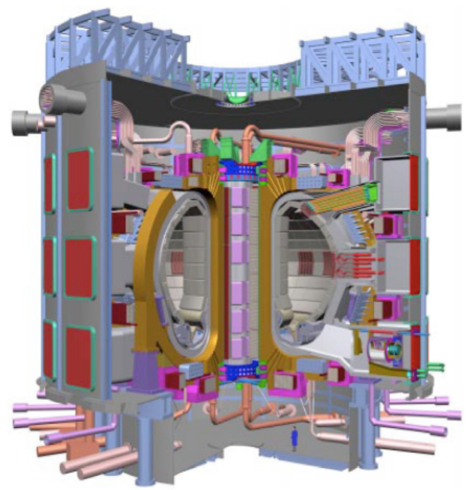
BUT > 95% accuracy with false alarm rate < 5% at least 30 milliseconds before actually needed for ITER ! **Reference** – P. DeVries, et al. (2015)

Success of ITER Requires Sufficiently Low Disruption Rate

Reference: Dave Humphries, GA/DIII-D

- Mid-pulse disruptions eliminate planned discharge time following disruptive event → *greatly reduces physics productivity*

- Disruptions can require *long recovery time bad for overall shot frequency*



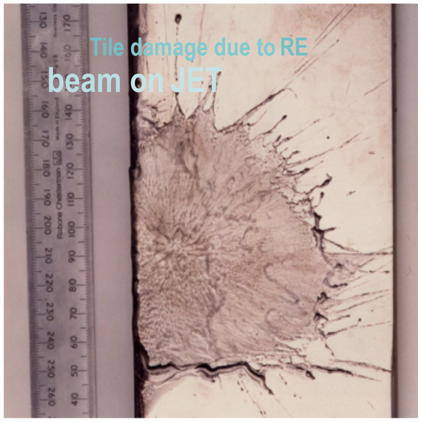
Availability > 80% (during operation periods)

Design target <10% disruptivity

- Disruption heat fluxes can *reduce component lifetime* (e.g. divertor target ablation)

- Damage to in-vessel components *can require shutdown for repair*

Tile broken by disruption forces in DIII-D



Tile damage due to RE beam on JET

AI/DL/Machine Learning Workflow

Identify Signals
• Classifiers

Preprocessing and feature extraction

Normalization

Train model, Hyper parameter tuning

Use model for prediction

Princeton U/PPPL AI/DL predictions now works with *multi-D time trace signals* (*beyond zero-D*)

Measured sequential data arranged in patches of equal length for training

All data placed on appropriate numerical scale ~ O(1)
e.g., Data-based with all signals divided by their standard deviation

• All available data analyzed;
• Train LSTM (Long Short Term Memory Network) iteratively;
• Evaluate using ROC (Receiver Operating Characteristics) and cross-validation loss for every epoch (equivalent of entire data set for each iteration)

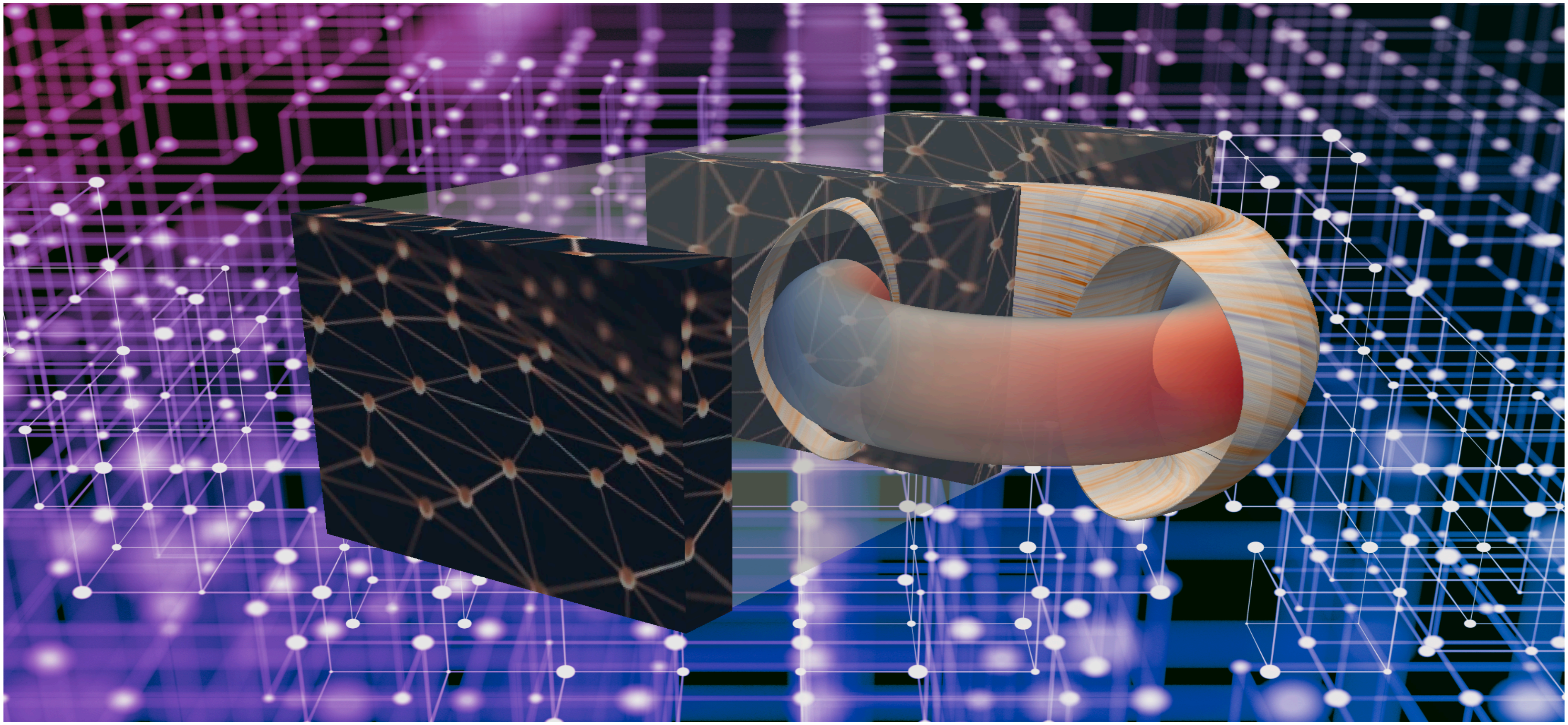
Apply AI/DL/ML software on plasma state signals

Artificial Intelligence/Deep Learning brings new technology to accelerate progress

"Predicting Disruptive Instabilities in Controlled Fusion Plasmas through Deep Learning"

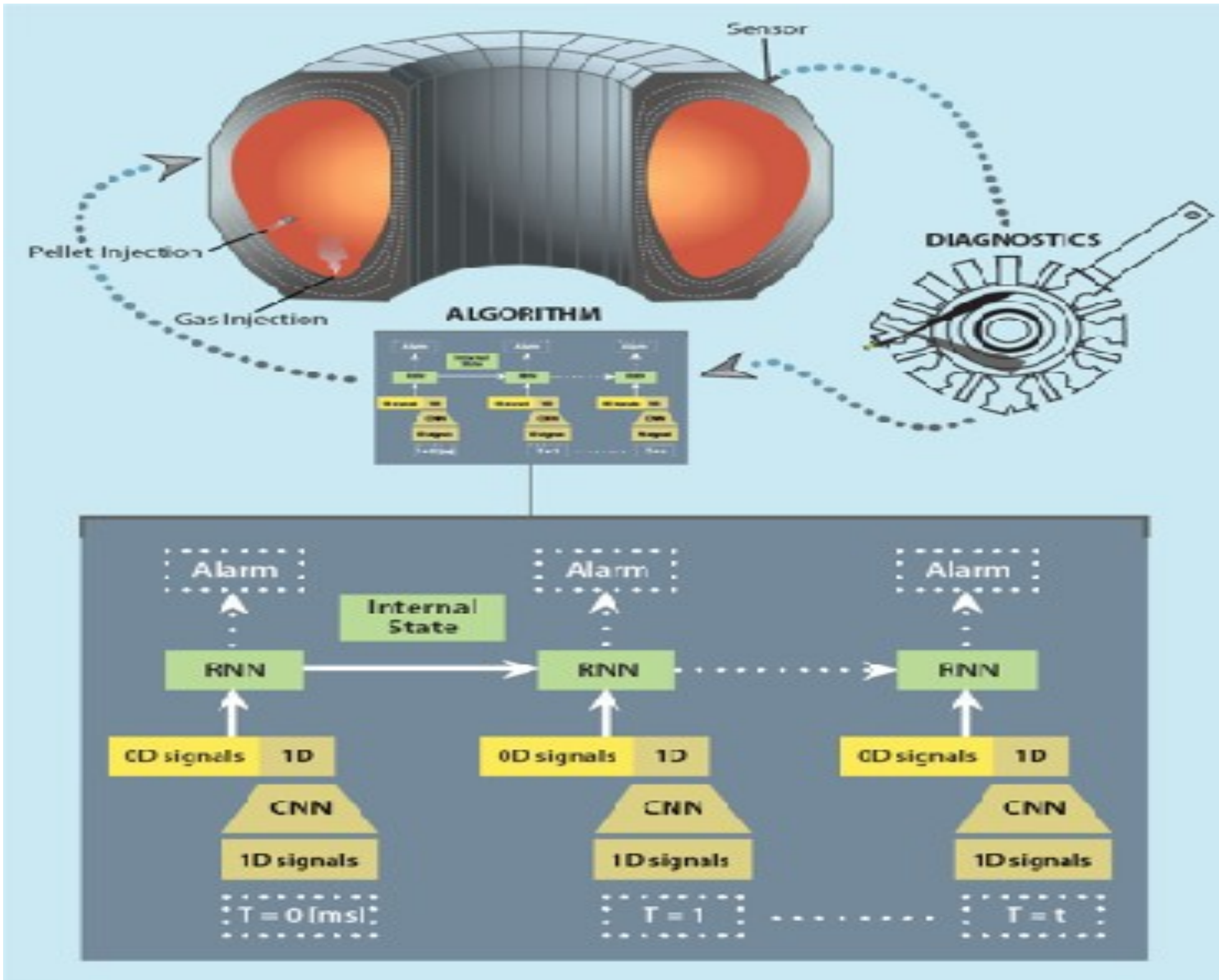
NATURE: (accepted for publication, Jan. 2019, published, April 17, 2019 – DOI: 10.1038/s41586-019-1116-4)

Princeton's Fusion Recurrent Neural Network code (FRNN) uses convolutional & recurrent neural network components to integrate both spatial and temporal information for predicting disruptions in tokamak plasmas with unprecedented accuracy and speed on top supercomputers



Data flow and summary of the AI/DL FRNN algorithm

→ highlights key challenge of associated plasma control



Highlights of KEY ACHIEVEMENTS featured in NATURE PAPER (2019)

→ ***Implementation of modern AI/Deep Learning advances enabled key achievements for Fusion Energy Science including:***

(1) Establishing ability to deal with one-dimensional "vector"physics signals for the first time (*overcoming "curse of dimensionality"*) – a significant improvement over previous Machine Learning R&D with focus on scalar-only "zero-D" signals.

(2) First demonstration of crucially-needed ability for predictive software trained on one experimental device (e.g., DIII-D tokamak) to make accurate predictions on another (e.g., the much larger, more powerful JET system) –> a key requirement for ITER – (*enabling cross-tokamak facility training*)

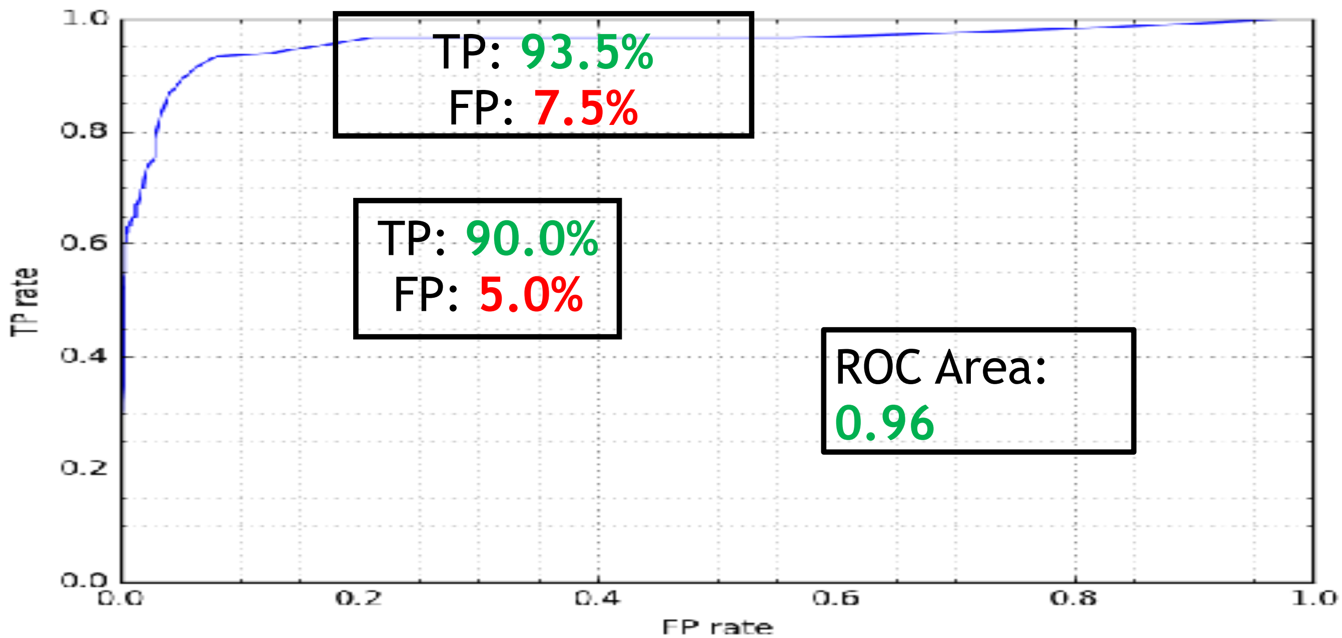
(3) Unique demonstration of AI/DL software capability to efficiently utilize leadership class supercomputers for *aggressive hyperparameter tuning scans enabling efficient training on big databases* – carried out, e.g. on *Titan, Summit in US; Tsubame-3, ABCI in Japan* –> and exascale systems in near future including *Aurora-21 and Frontier* in US; *Fugaku (formerly Post-K)* in Japan, and other emerging systems worldwide.

FRNN Code PERFORMANCE: ROC CURVES

JET ITER-like Wall Cases @30ms before Disruption

EUROfusion

Performance Tradeoff: Tune True Positives (good: correctly caught disruption) **vs. False Positives** (bad: safe shot incorrectly labeled disruptive).



JET Data (~50 GB), 0D signals:

- Training: on **4100 shots** from **JET C-Wall** campaigns
- Testing **1200 shots** from **Jet ILW** campaigns
- **All shots used**, no signal filtering or removal of shots

JET Data courtesy of J. Vega and A. Murari

CNNs & RNNs with HPC Innovations Engaged

GPU training

- Neural networks use dense tensor manipulations, efficient use of GPU FLOPS
- Over 10x speedup better than multicore node training (CPU's)

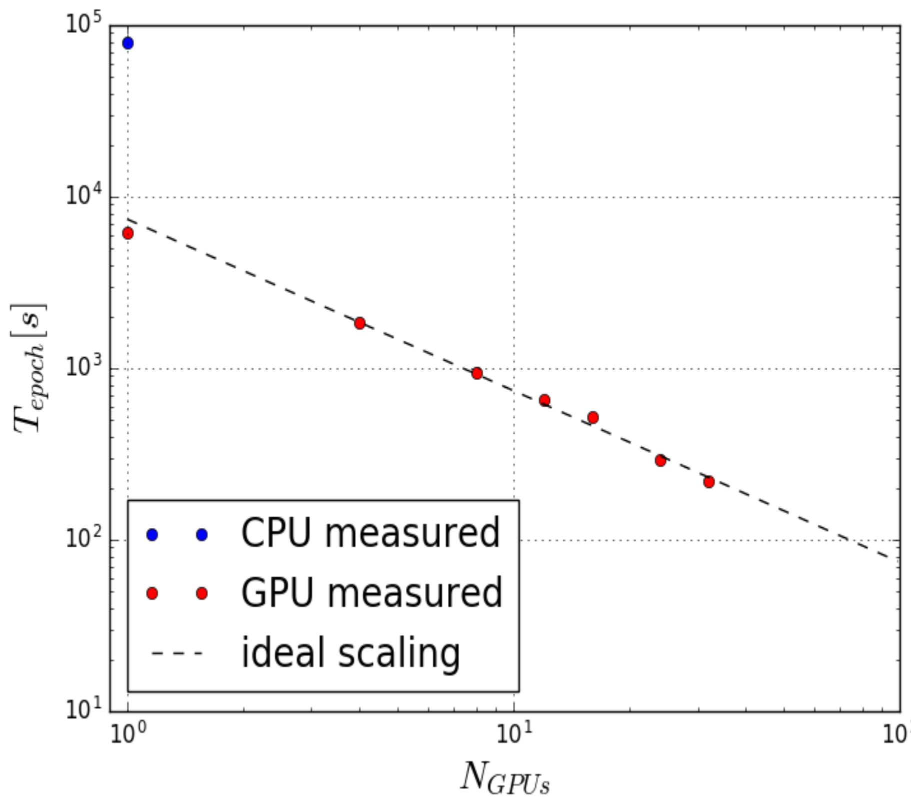
Distributed Training via MPI

Linear scaling:

- Key benchmark of "time to accuracy": we can train a **model that achieves the same results nearly N times faster with N GPUs**

Scalable

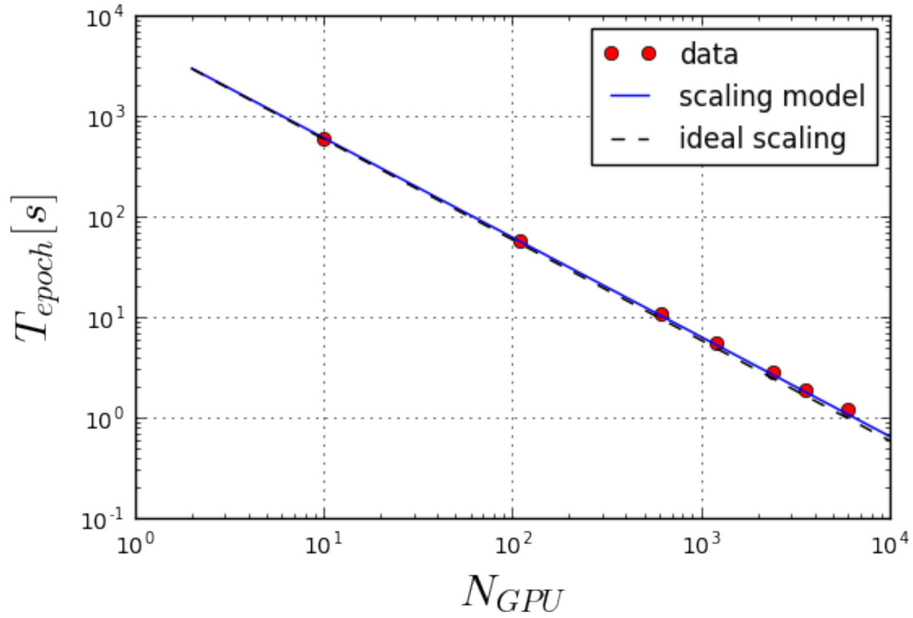
- **to 100s or >1000's of GPU's on Leadership Class Facilities**
- **TB's of data and more**
- Example: Training time on representative full dataset (~40GB, 4500 shots) of 0D signals
  - **SVM** (JET) **> 24hrs** (on CPU cluster)
  - **FRNN** (Princeton U – on 20 GPU's) **~40min**



FRNN Scaling Results on GPU's

- Tests on OLCF Titan CRAY supercomputer
  - ***OLCF DD AWARD***: *Enabled Scaling Studies on Titan currently up to 6000 GPU's*
  - Total ~ 18.7K Tesla K20X Kepler GPUs

Tensorflow+MPI



***** FRNN DL/AI software reliably scales to 1K P-100 GPU's on TSUBAME 3.0 "Grand Challenge Runs" (Tokyo Institute of Technology),Japan)**

→ associated production runs contribute strongly to Hyper-pameter-Tuning-enabled physics advances !

Hyper-parameter Tuning enabled by HPC

- **Example →** random grid of 100 iterations with 100 GPUs per each trial

  -- Trials run asynchronously to convergence

  -- Distributed training performed with ***data-parallel synchronous "Stochastic Gradient Descent (SGD)*** – standard approach in DL applications

  –Master loop determines the best set of parameters based on the validation level

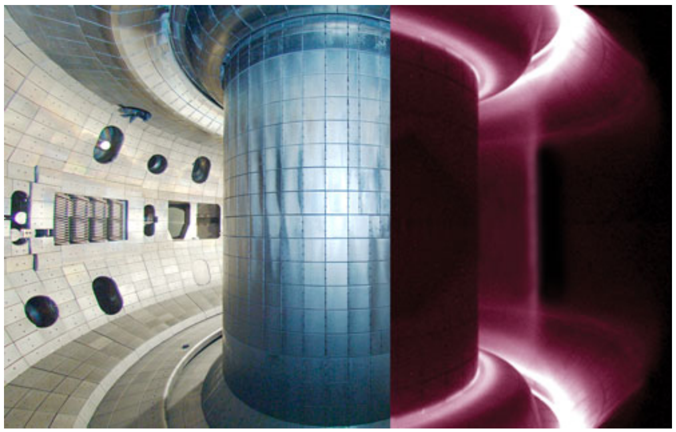
- **Exciting New Trends Emerging →** aggressive large-scale hyper-parameter tuning trials carried out on the "Titan" ***exhibit very promising potential for shifing the minimum warning time before disruptions to 50 ms and now up to 100 ms and above.***

  → *Strongly motivates new HPC-enabled studies enabled by deployment of **new half- precision version FRNN*** using **NVIDIA Volta GPU's on SUMMIT** at ORNL*

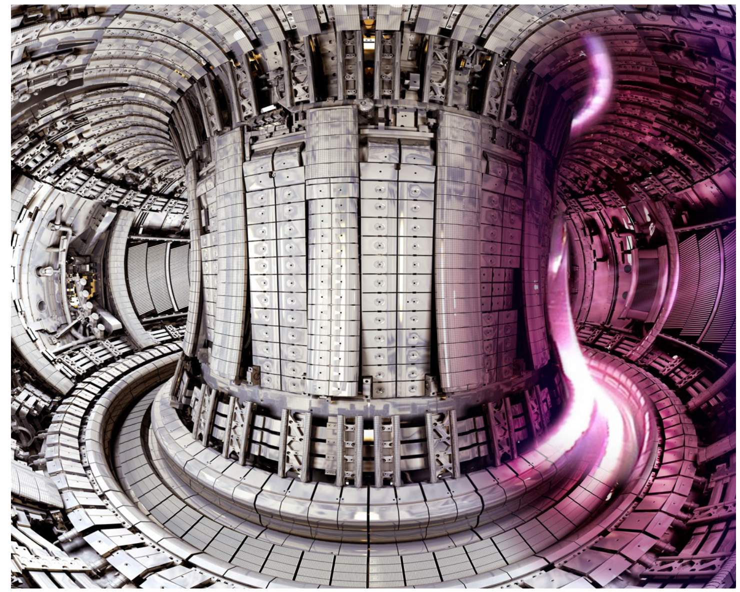
** **Significance**: ***Key to enabling future risk mitigation for ITER via achieving increased pre-disruption warning time***

Cross Machine Disruption Prediction (DIII-D to JET)

First demonstration of predictive DL software trained on one experimental device (DIII-D) to make accurate predictions on another (JET) – *critical for ITER*



Train (DIII-D)



Test (JET)

FRNN 1D	**0.836**
FRNN 0D	0.817
XGBoost	0.616

Integration of HPC (using GTC Exascale Code) with Deep Learning Workflows (using FRNN DL Code)

- *"Knowledge & experience" now in place for carrying out path-to-exascale HPC simulations of ITER-relevant burning plasmas with powerful*

GTC code

→*ESP selection for SUMMIT and 2019 INCITE awardee of 740K SUMMIT Node Hours – 151% above our request !*

Example: Neoclassical tearing modes (NTM's) already experimentally observed in JET, but NO realistic models yet developed as improved pre-disruption classifiers in Machine Learning workflows → ***because of inability to include measured higher-D profiles (only scalars)***

- ***CNN & RNN allow including realistic 1D & higher-D measurements of profiles to enable first-principles-based reduced models of NTM's (supported by exascale GTC code) to be used in FRNN workflows.***

Example of "integration of HPC with DL" !

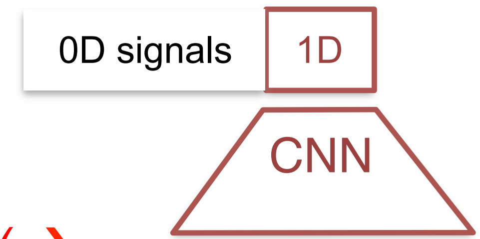
Cross-Disciplinary R&D Opportunities
(e.g., for AI/DL Applications & Applied Math)

Example: Improving Efficiency of Dense Matrix Operations in Keras Methodology used in DL/AI FRNN Code

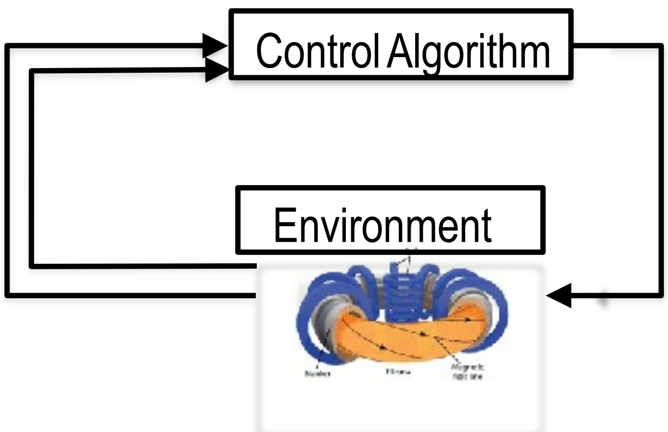
- Hierarchical Matrix Operations on GPUs: Matrix-Vector Multiplication and Compression
- Explore use of KAUST Basic Linear Solver (KBLAS) Packages

DL/AI Vision Summary in Moving from Prediction to Control

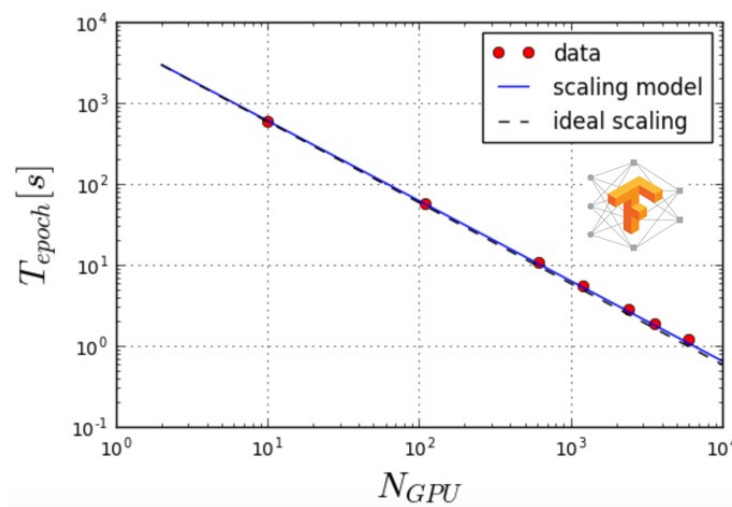
ZERO-D to HIGHER-D SIGNALS via CONVOLUTIONAL NEURAL NETS (CNN)



- *Enables immediate learning of generalizable features (→ helps enable **cross-tokamak portability of DL/AI software**)*



- ***Reinforcement learning enables transition from PREDICTION to CONTROL !***

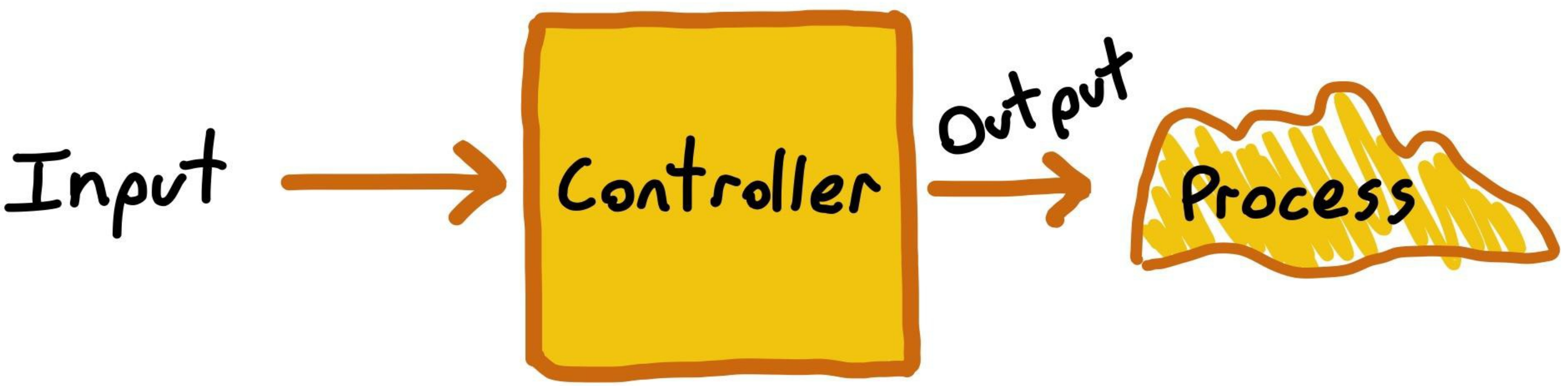


- *Takes advantage of increasingly powerful **world class HPC (supercomputing) facilities !***

Control Methods with Containers

Ref: Vallery Lancey, Lead DevOps Engineer, "Checkfront"

- **Managing a system using human and internal controls**
- **Inputs dictate what the controller should do (setpoint)**
- **Outputs dictate what the controlled process should do**
- **Closed Loop Container: (i) Contains feedback from the process to the controller; (ii) Controller able to self-correct to achieve desired outcome**



Control System Management

Traditional: ”Sysadmin" examines the system, makes a judgement, and performs an action

Automatic: System tracks its own state & translates the state to some internal action

POSSIBLE FRNN DEPLOYMENT INTO PCS of Tokamak Facilities e.g., DIII-D, JET, KSTAR, ...

A. Svyatkovskiy, Princeton U/PPPL/Microsoft

Approach: Deploy AI/DL/ML FRNN disruption predictor (described in NATURE) as a ”web-like service *within Tokamak facilities* using modified versions of Microsoft’s “Azureml”/Azure Container Service

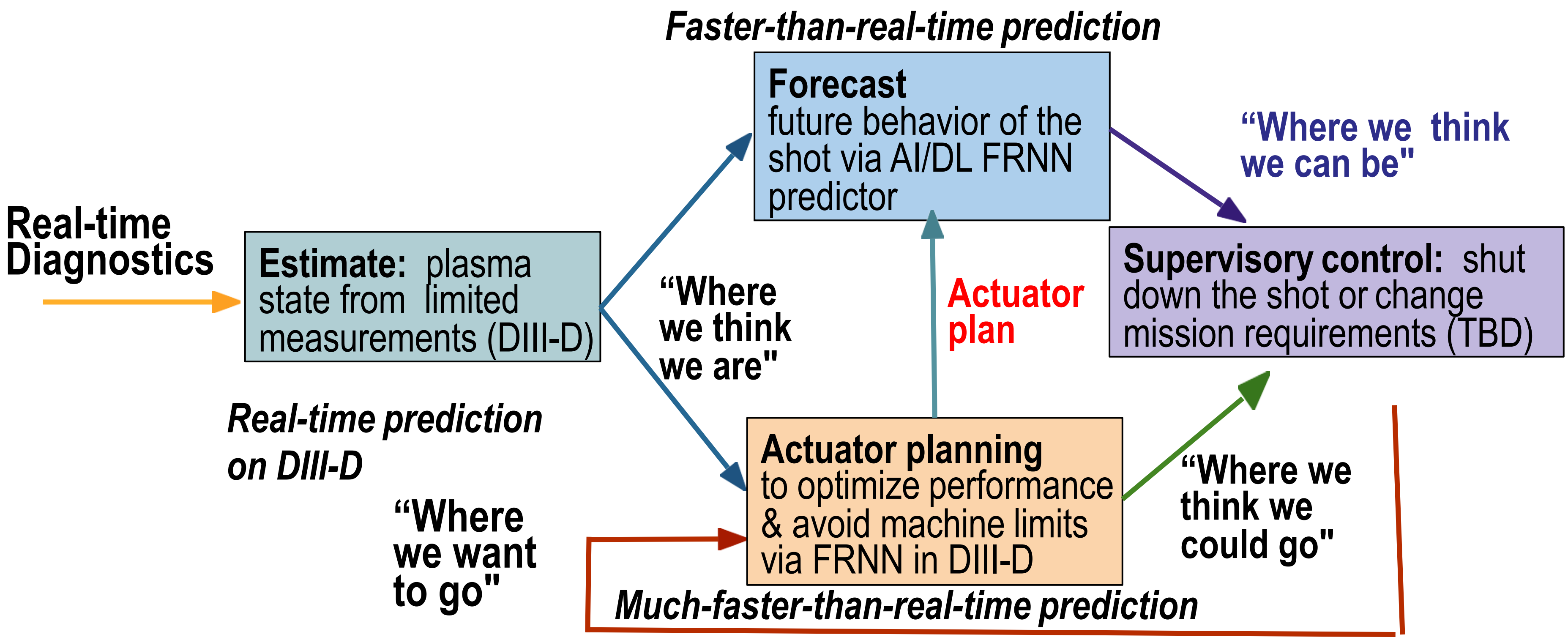
1) Can either use current version of FRNN or choose to train new pre-disruption classifiers with more realistic “reduced HPC-enabled classifiers for – e.g., for NTM’s, ITER-relevant alpha-driven instabilities, etc.

2) Prepare a "helper code" to deploy the model & interact via “RESTful API”–(details under development with Microsoft)

3) This approach has potential to carry out predictions on the order of a few 10’s of micro-seconds including network latency *

* ***examples available from other cases in Microsoft applications deployment portfolio***

"Computing at the Edge": Real-Time Experimental Planning



- Can we make our AI/DL FRNN Predictor fast & accurate enough?

--- e.g., via reinforcement learning/inference/ applied math ……

- Can we make our actuator models suffiently fast & realistic enough?

--- e.g., via focused actuator planning with experimental partners

KEY UPCOMING AI/DL PROJECT FOCUS:

➔ Moving from AI/DL-based Tokamak Prediction to real-time Plasma Control:

-- first need to ***strongly complement AI/DL prediction results (NATURE paper) with dedicated new runs enabled by experimental proposals submitted to DIII-D and JET – plus new ones on long-pulse KSTAR, EAST, and JT-60SA***

-- need to begin ***experimental control studies involving deployment of DL/AI predictors within actual Plasma Control System (PCS) at DIII-D, JET, KSTAR, EAST, & JT60SA***

→ ***involves reinforcement learning, inference, etc. + deployment of novel actuators developed with strong engagement by diagnostics experts for PCS deployment of AI/DL predictors to initiate control studies.***

******* News: US Executive Order signed for huge upcoming investment in ARTIFICIAL INTELLIGENCE/DEEP LEARNING !***

(Feb.11, 2019)