2025 ”From Modeling to Learning with HPC”

AI, Machine Learning, & HPC Enabled Scientific Discovery

William M. Tang (Princeton University, Princeton Plasma Physics Lab)

Jesse Rodriguez (Princeton U. CSML, Oregon State U.
Juan-Felipe Gomez (Harvard University)
Kyle Felker (Argonne National Lab)

BROWN UNIVERSITY
PROVIDENCE, RI

SEPTEMBER 13 , 2025

ABSTRACT: (for 45 minutes invited talk)

We focus on the scientific and engineering advances being driven by HPC together with advanced statistical methods featuring artificial intelligence/deep learning/machine learning (AI/DL/ML). Accelerated progress in delivering scientific discovery via predictions with these methods guided by Validation, Verification & Uncertainty Quantification (VVUQ) are enabling data-driven discoveries essential for realizing the grand challenge potential of fusion energy. The best validated fusion devices are tokamaks, which are vulnerable to large scale plasma instabilities called “disruptions” that can stop the reaction and damage the device.

Reliably predicting and avoiding these damaging events is essential, and recent advances feature successful capabilities involving the deployment of recurrent and convolutional neural networks in Princeton's Deep Learning Code – FRNN – for carrying out efficient transfer learning across very different tokamak devices. Essential validation vs. a huge data base is documented and highlights the applicability of FRNN to both the large JET and smaller DIIID. It is further shown that this AI/DL capability can provide not only the “disruption score,” as an indicator of the probability of an imminent disruption but also a “sensitivity score” in real-time to indicate the underlying reasons for the predicted event. Consequently, this deep learning capability provides the detailed information for a plasma control system (PCS) that can help improve disruption avoidance in real-time and to thereby optimize plasma performance. Application of such AI/DL methods for real-time prediction and control has been further advanced with the innovative introduction of a surrogate model simulator (SGTC).

These capabilities are now leading to exciting avenues for moving from passive prediction to active control and ultimately, to the optimization of the design for a first-of-a-kind fusion pilot plant.

• This **Talk highlights** our cross-disciplinary project that **builds and validates a data-driven, real-time predictive control system capable of realistically anticipating dangerous instabilities in magnetic fusion energy plasmas** ***guided by Validation, Verification and Uncertainty Quantification.***

• The aim of our ***associated Fusion Foundation Model ("FFM")*** is to transform fusion-energy research , enable human-AI co-discovery, and accelerate commercialization this decade.

• Large amounts of DOE HPC cycles have been secured via INCITE and an Early Science AURORA project ➔ *available for immediate use.*

•. Accelerated progress has been achieved on our *"Fusion GPT" -- an "AI Killer AP" within a foundation model framework with connections to NLP and Image Recognition* that was initially presented at international "Trillion Paramter Consortium (TPC)" 2024 meetings in Barcelona and SC'24 in Atlanta.

• ***Exciting new results have arrived via unique access to the CSCS ALPS supercomputer in Switzerland with it's 5K NVIDIA Grace-Hopper "Superchips." !***

TALK Outline

• Key scientific challenges for Magnetic Fusion Energy (MFE) Plasmas:
➔ *Disruptions; • Plasma Modeling and Real-Time Control; • Validation, Verification, & Uncertainty Quantification*

• Fusion ripe for deep learning engagement with advanced methodologies featuring HPC:
➔ *Key role of HPC in current & future Exascale computing [(Frontier, Aurora) in US; ALPS in Switzerland; JUPITER in Germany]*

• Deep Learning (DL) in MFE research:
➔ *High dimensional data (spatial/temporal) with rapid development facilitated by cross-cutting connections to natural language processing (NLP) & image recognition*.

• Fusion can help benchmark AI progress as *a societal and scientifically meaningful challenge with nonlinear data pushing progress in ML.*

• AI/ML/DL are “evolutionary” modern “tools in theoretical science" *for achieving diverse and complex plasma control capabilities needed for development of realistic next-generation power plant designs to accelerate delivery of real-world fusion reactors.*

Fusion: an Attractive Energy Source with Formidable Scientific Goal !

- Abundant fuel, available to all nations
  - Deuterium and lithium easily available for millions of years
- Environmental advantages
  - No carbon emissions, short-lived radioactivity
- Cannot "blow up or melt down," resistant to terrorist attack
  - Less than a minute's worth of fuel in the chamber
- Low risk of nuclear materials proliferation
  - No fissile materials required
- Compact relative to solar, wind and biomass
  - Modest land usage
- Not subject to daily, seasonal or regional weather variation;
  no requirement for local CO_2 sequestration
  - Not limited in its application by need for large-scale energy storage nor for long-distance energy transmission
- Fusion is complementary to other attractive energy sources

SCIENTIFIC GOAL →

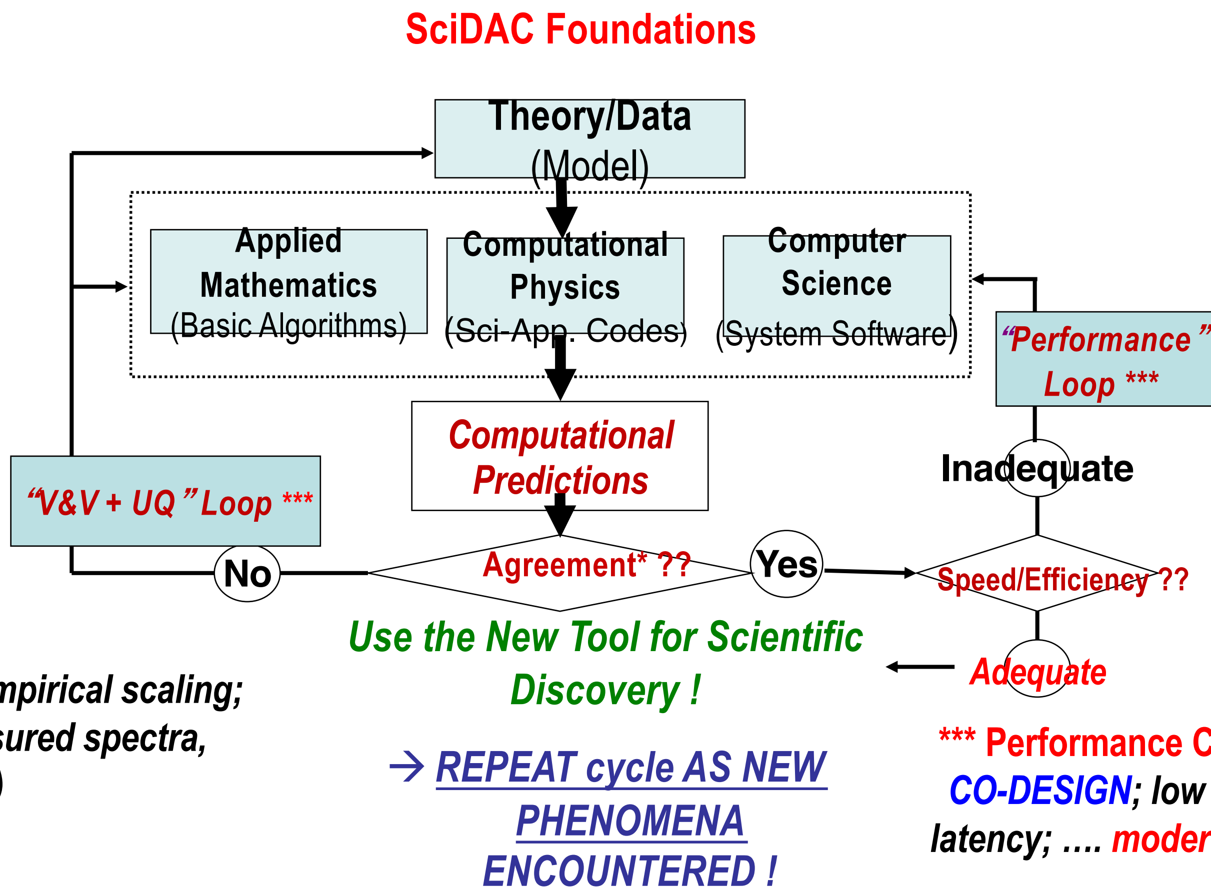
achieving "triple product of density/temperature/CONFINEMENT TIME conditions" IS HUGE GRAND CHALLENGE !

["Lawson Criterion"] for burning plasmas leads to a viable fusion power plant and eventual working reactor !
(was posed in the 1950's with D-T fuel having the most viable cross-section).

Other Considerations: Compact High-Field Tokamaks with Superconducting Materials (common interest with CERN/HEP !)

Validated Advanced Scientific HPC Codes:

"a measure of the state of understanding of natural & engineered systems"



*** VVUQ Challenges: *empirical scaling; sensitivity studies; measured spectra, correlation functions, ...)*

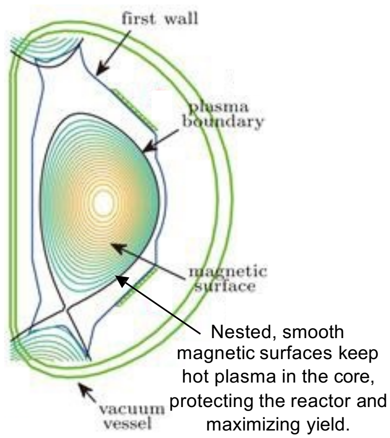
*** Performance Challenges: *CO-DESIGN; low memory/core; locality; latency; .... modern AI/DL/ML, ...)*

Prediction & Mitigation of Abrupt Disruptions for Fusion Device Operations

Equilibrium Tokamak Plasma

Unstable Tokamak Plasma

Stable Operation



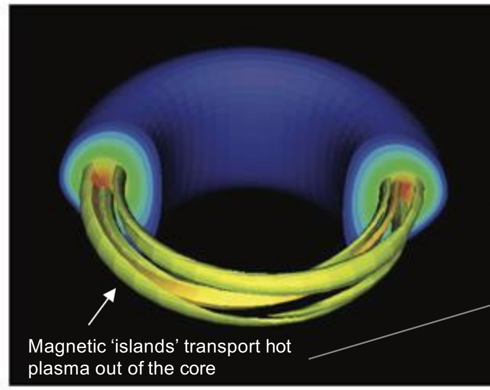
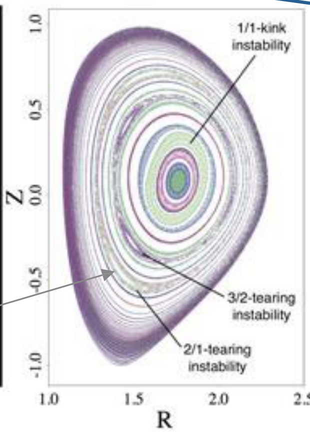


Figure by D. P. Brennan



Disruption Active !

Necessary for ignition with preservation of plasma-facing components

Power generation made impossible with damages to expensive ccmponents !

- Disruptions are caused by chaotic/turbulent phenomena that <u>*require the ability to predict in real-time.*</u>
  - On future ITER-scale experiment, a single disruption <u>*could cause major damage and halt operations !*</u>

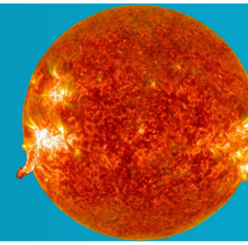
MISSION: To dramatically improve the stability of tokamaks by developing and training machine-learning

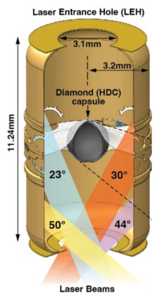
Goal: ***To develop "VVUQ enabled" software @ HPC exascale & beyond for prediction & mitigation of disruption events !***

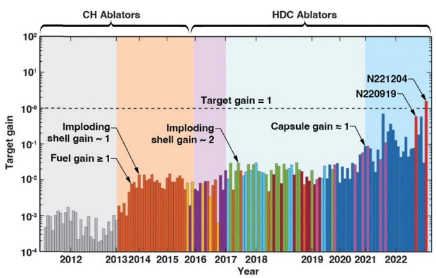
Exciting Time for Fusion as the "Final Frontier of Sustainable Clean Energy"

Nuclear fusion is moving closer to delivering on grand challenge of harnessing process by which Sun produces its energy !

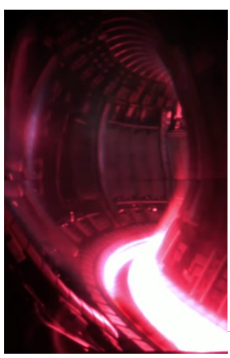
BUT "POWER ON THE GRID" REMAINS HUGE CHALLENGE FOR BOTH MFE & IFE APPROACHES

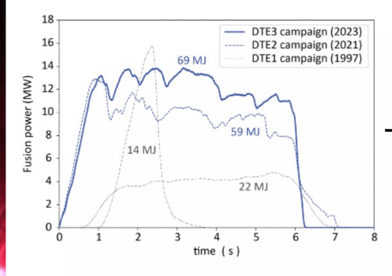






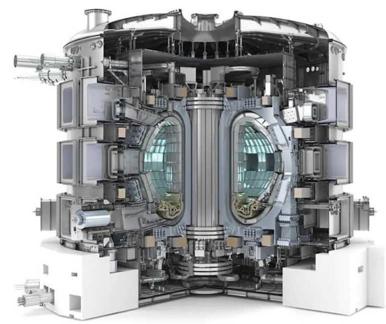
→ ***LLNL-NIF achieved fusion break-even (ignition) in late 2022 !***

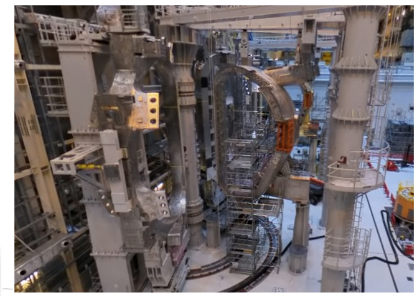




→ ***Eurofusion JET achieved ~ 70 MJ of fusion yield in 2023 !***

<u>(but huge BREXIT political obstacle resulting in "premature termination")</u>





- **ITER (under construction) targets a "burning plasma" ~2035 towards goal of fusion gain ~ 10.**

- **Meanwhile, billions of dollars flooding in from *private sector*.**

(CAUTIONARY: This is a "2-edged sword" with major unmet "VVUQ" requirements !!)

AI/DL/ML is essential for enabling ITER Goals and a "First of a Kind" Fusion Pilot Plant

Machine Learning Workflow

Identify Signals
• Classifiers

Preprocessing and feature extraction

Normalization

Train model, Hyper parameter tuning

Use model for prediction

AI/DL/ML predictions advanced to *multi-D time trace signals*

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 (e.g.) 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 ML/DL software on ***new data***

Francois Chollet (Google)
"Deep Learning with Python"
powerful Keras Library
"understanding via intuitive explanations and practical examples"

Artificial Intelligence/Deep Learning brings new technology to accelerate progress

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

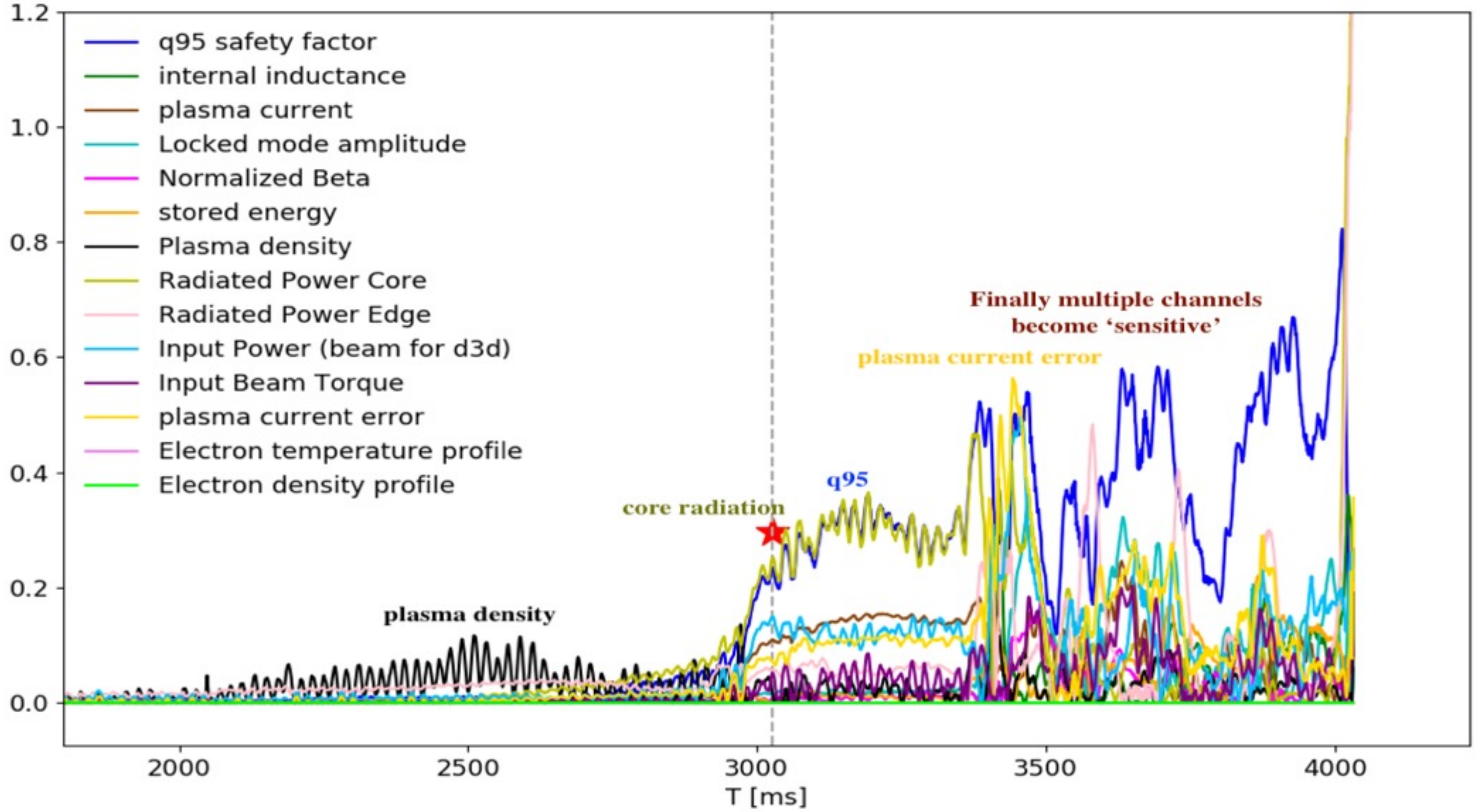
NATURE: (April, 2019 – DOI: 10.1038/s41586-019-1116-4) > 300 Google Citations

- ***Julian-Kates Harbeck, Alexey Svyatkovskiy, & William Tang***

"FRNN" software *(using "CNN & RNN" with supercomputers integrated huge amounts of space/time* ***data*** *to predict plasma disruptions with unprecedented accuracy & speed,* ***PLUS*** *essential demonstration of* ***"transfer learning" !***

- ***Ge Dong**, **et al.***, **"Deep Learning-based Surrogate Model for First-principles Global Simulations of Fusion Plasmas,"** *NUCLEAR FUSION 61 126061 (2021) addresses* ***CRUCIAL ISSUE OF "REALISTIC SURROGATES*** for ***FUTURE FUSION SYSTEMS "WHERE NO 'GROUND-TRUTH' MEASUREMEN EXIST – e.g., for ITER AND FUSION PILOT PLANTS !!***

Relation of Physics-based Signals to ML Predicted Disruption Score



Evolution of the sensitivity score of DIII-D shot #164582

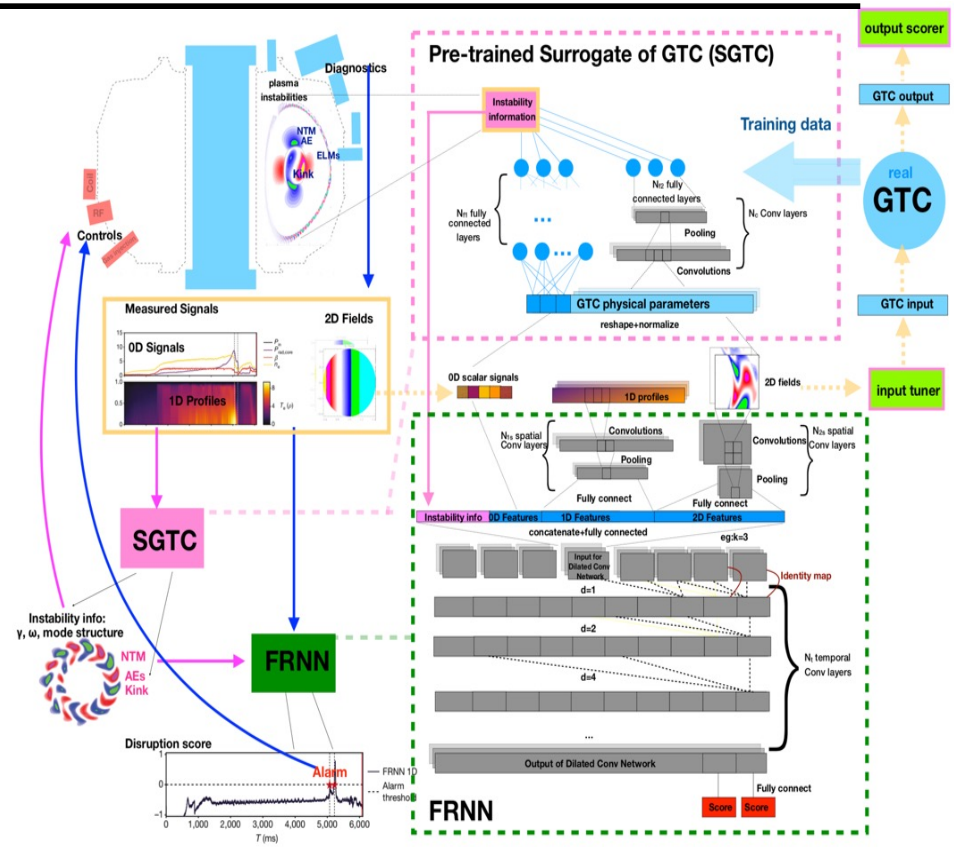
SGTC "Surrogate RT Simulator": *Instability Predictive Model Workflow*

- SGTC established with workflow of ***neural-network-based surrogates of the experimentally-validated first-principles EM gyrokinetic toroidal code (GTC)***

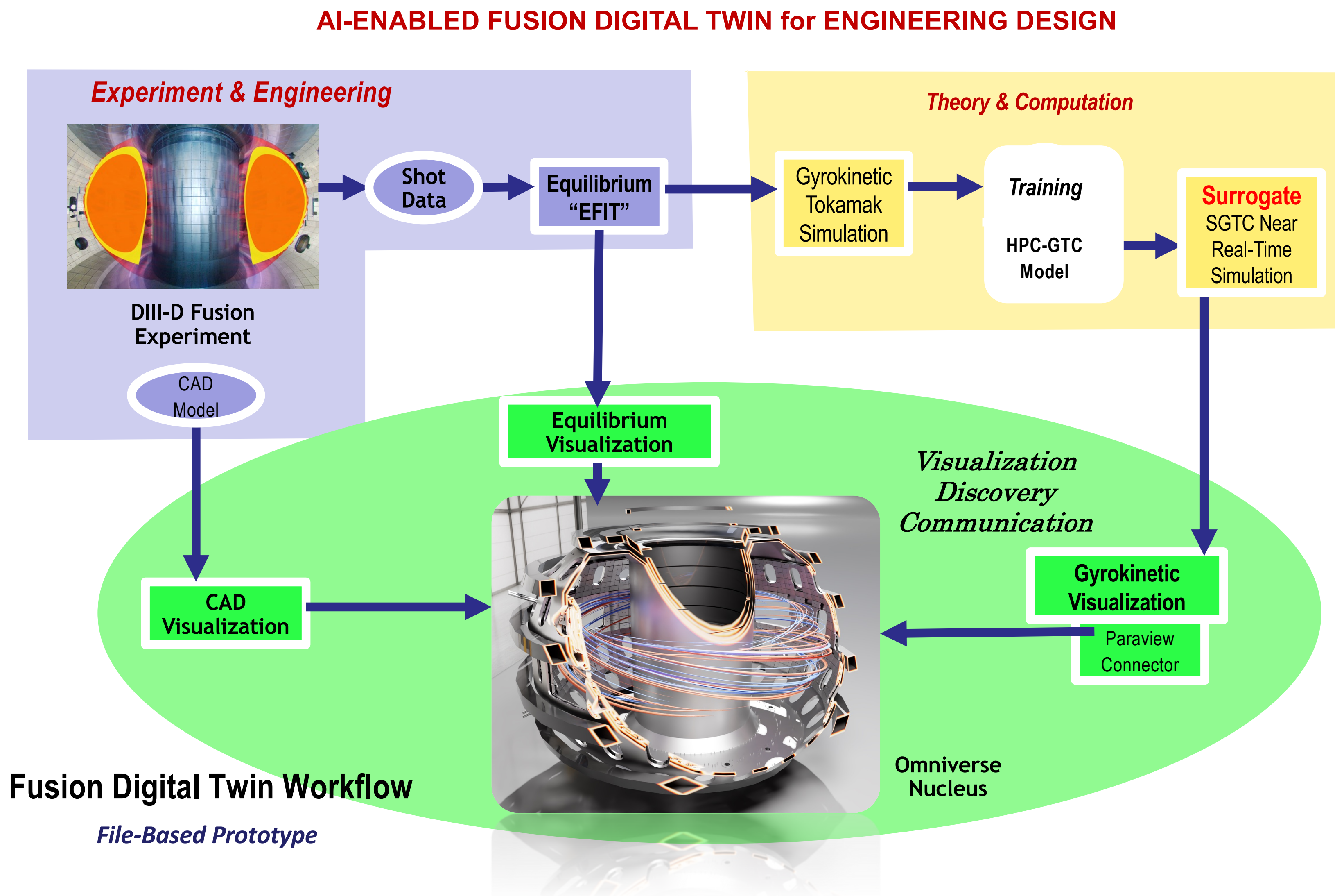
- Trained with extensive simulated data from GTC enabling realistic inference (prediction) of plasma instability properties

→**SGTC** models satisfy compatibility requirements of a plasma control system (PCS) and deliver ***inference times on order of milliseconds (ms)***

→ **SGTC** deliver results ***<u>5 order of magnitude faster</u> than GTC runs on Exaflop supercomuters !!***



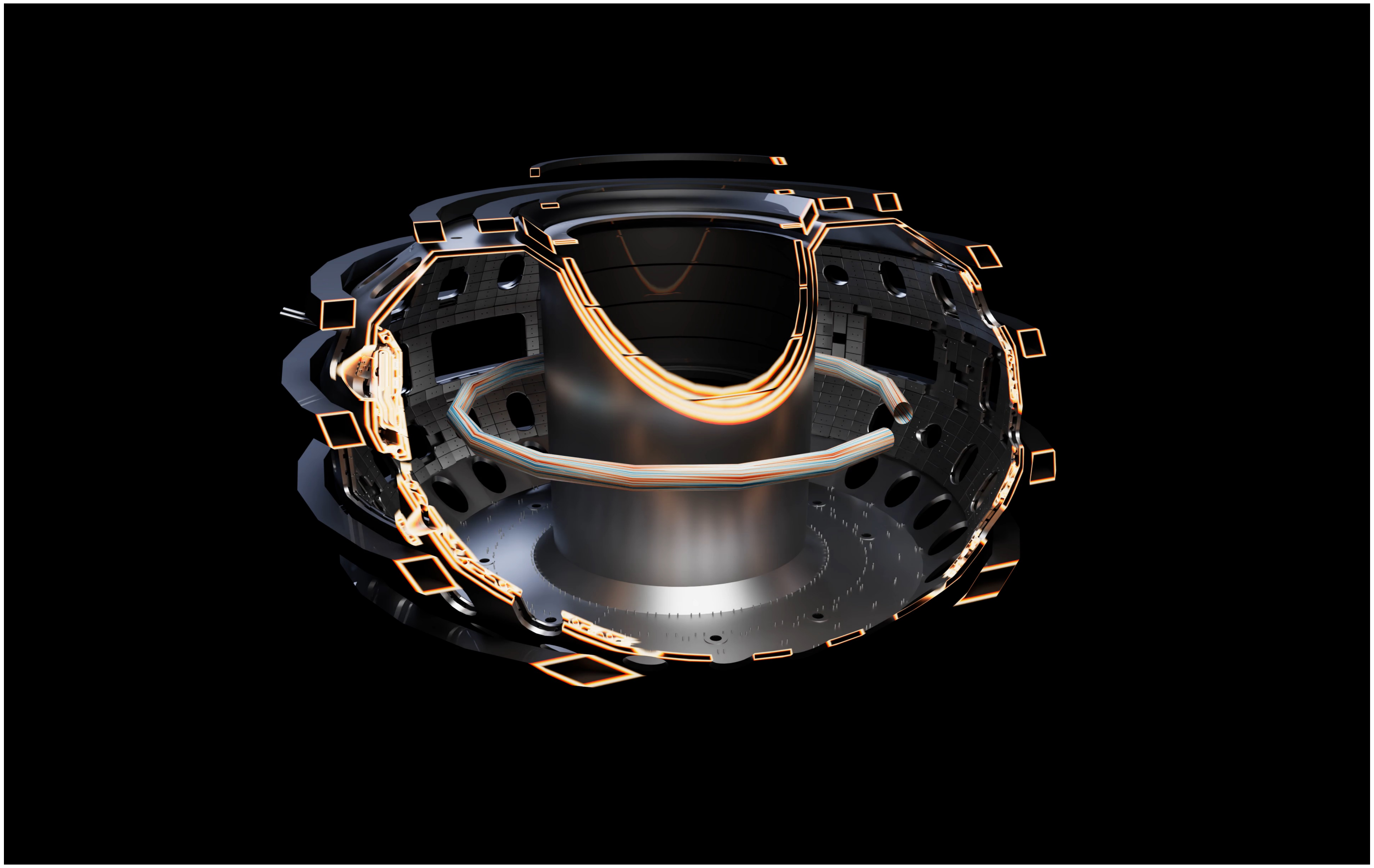
G. Dong, X. Wei, J. Bao, G. Brochard, Z. Lin and W. Tang (Nuclear Fusion, 2021)



AI-ENABLED FUSION DIGITAL TWIN for ENGINEERING DESIGN
Experiment & Engineering
Shot Data
Equilibrium "EFIT"
DIII-D Fusion Experiment
CAD Model
Theory & Computation
Gyrokinetic Tokamak Simulation
Training
HPC-GTC Model
Surrogate
SGTC Near Real-Time Simulation
Equilibrium Visualization
Visualization
Discovery
Communication
CAD Visualization
Gyrokinetic Visualization
Paraview Connector
Omniverse Nucleus
Fusion Digital Twin Workflow
File-Based Prototype

AI-Machine Learning-Enabled Tokamak Digital Twin

William Tang, et al, Proceedings of the 23rd IAEA FEC Conference, Paper TH-C, London, UK
Submitted for Publication, Nucl. Fusion (2023)



"TRILLION PARAMETER CONSORTIUM" FOUNDATION MODELS

. Validated scientific and engineering advances driven by Exascale Computing together with modern statistical methods featuring artificial intelligence/deep learning/machine learning (AI/DL/ML) ➔ ***must properly embrace verification, validation, and uncertainty quantification (VVUQ) to truly establish credibilit**y**.

- **Foundation Model efforts have especially promising potential for impact:**

  e.g., general framework can come from rapidly evolving LLM's & Image Recognition Models

→ Training of multi-billion parameter models on mix of experimental and simulation data – with technology leading to fine-tuning of huge models multiple times into smaller distilled models for "multitask learning for complex & diverse control needs"

"FUSION GPT" – CURRENT 1.5B Parameter Foundation Model

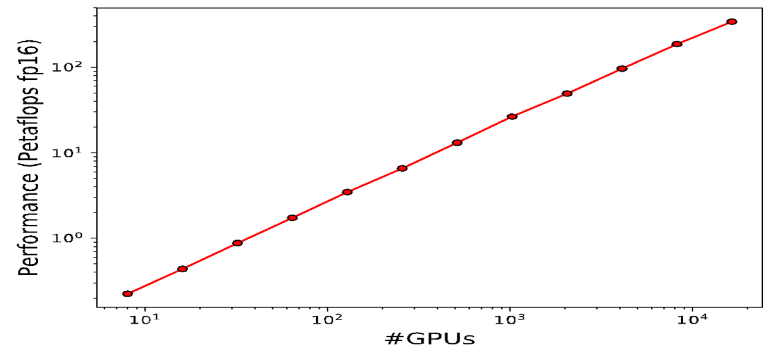
- Feasibility of the **1.5B parameter foundation model (FusionGPT)** using auto-regressive pre-training on experimental ECEi **data** + fine-tuning for downstream task of disruption prediction

  • WaveNet-type model trained on **higher temporal resolution** for 500 kHz ECEi (Electron-Cyclotron Emission Imaging)

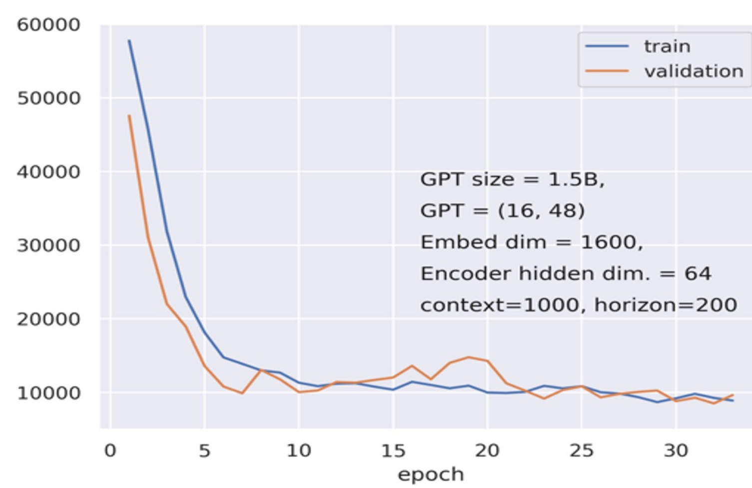
- **Data set created for optimized data loader --** efficiently handles 2D ECEi data along with 0-D and 1-D diagnostic data for input training samples

- Source Exascale GTC Simulation software currently running successfully on **Frontier** with initial performance optimization completed.

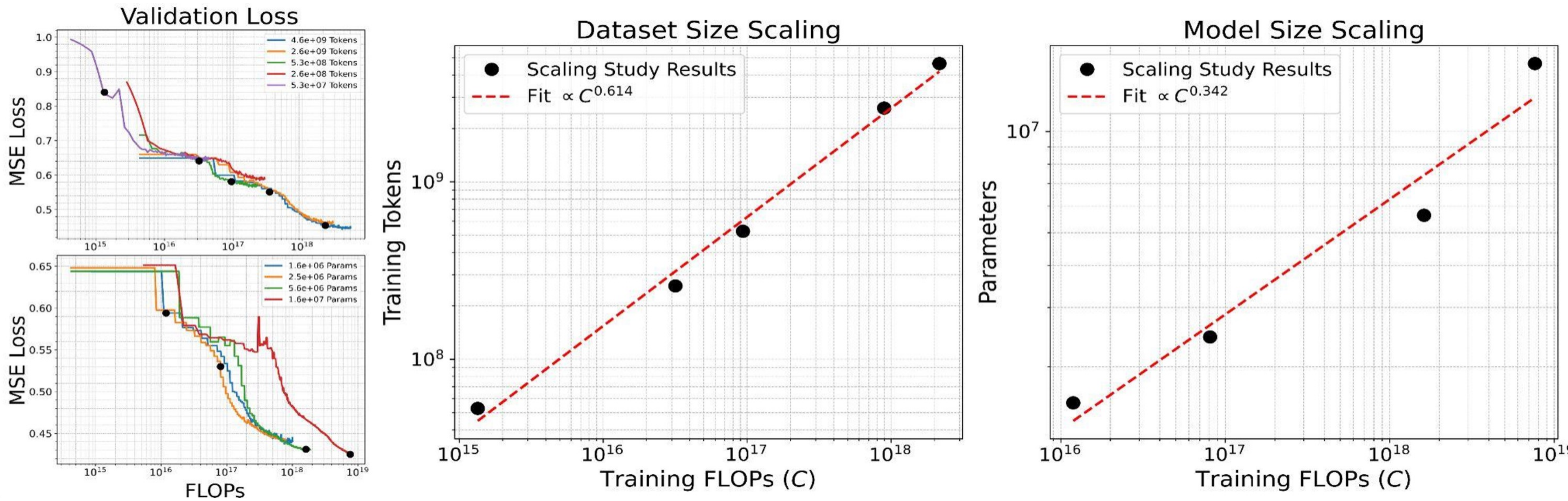
- DIII-D inference capability now prepared for using on-site GPU's.



Scaling of fusion model training with DeepSpeed to 2048 nodes and 16,384 GCDs.
→ Achieved 0.4 fp16 exaflops.



Achievements leading to Future Work: “Do standard practices for LLMs transfer to Science?”

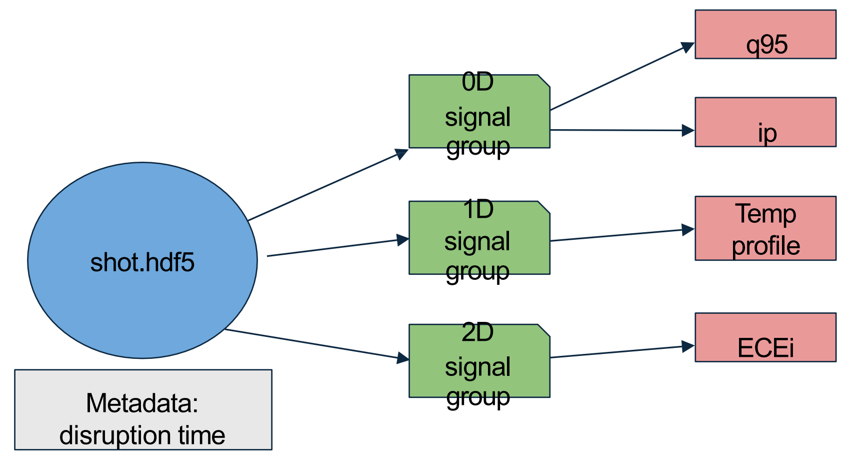


‘Chinchilla’-style scaling study illustrates **differences between the typical ‘infinite data’ environment of LLMs and the data-limited environment of fusion energy science**
[Jesse Rodriguez, Princeton U/CSML; Oregon State U]

- Demonstrates standard results and practices for LLMs ***DO NOT*** necessarily transfer to models trained on physical data → helps inform our compute budget and model size going forward.

More Data Soon !

Signal description	Numerical scale
Plasma current	3.8 e-1 MA
Plasma current target	3.9 e-1 MA
Plasma current error	3.1 e-2 MA
Plasma current direction	1.0
Internal Inductance	2.02 e-1
Plasma density	1.19 e19 m^{-3}
Input power (beam for DIII-D)	1.85 e6 W
Radiated power core	4.58 e2 W
Radiated power edge	4.94 e2 W
Stored energy	2.79 e5 J
Locked mode amplitude	1.14 e-6 T
Safety factor q95	1.0
Normalized plasma pressure	6.91 e-3
Input beam torque	1.47 Nm
Electron temperature profile	9.53 e-1 keV
Electron density profile	1.47 e19 m-3



Model pretraining with more data & associated compute will continue with ALPS engagement !

FUTURE OBJECTIVES: Clean Energy Fusion & Climate as 21st Century Grand Challenges Accelerated by Multi-disciplinary Research & the Exciting New HPC Capability

→ **Validation** vs. large "ground-truth" observational/experimental database to enable improved prediction and real-time control

→ **Real-time Simulators featuring Surrogates** backed by validated 1st-principles-based exascale-class HPC from

e.g.:
- **FRONTIER (ORNL) & AURORA (ANL) Exascale Systems in US**
- **ALPS 5K Grace-Hopper "superchips at CSCS In Switzerland**
- **Upcoming JUPITER (JSC) Exascale System in Germany**

→ **Industry Engagement** including:
- ***Microsoft's Transformers*** to accelerate rapid transfer of large multi-D complex temporal data
- ***Nvidia's "Superchips" – powered HPC Systems & Omniverse Viz*** to deliver near-real-time Digital Twins

→ **Essential Role of AI** advances emerging from NLP & Image Recognition Foundation Models

-- e.g., "Fusion GPT"

COMMENT: For major grand challenges, the "human in the loop" aspect will likely continue to be a significant necessity -- as long as true VV&UQ practices are properly followed !

Perhaps"realistically-validated" Quantum Information Science technology will change this scenario ??

- ***Future progress demands Attracting & Training the best and brightest talent to Grand Challenge growth areas !***

Delivery of Promise from "AI for Science" will require multi-disciplinary R&D, industry partnerships, & strategic new international collaborations featuring the ALPS at CSCS in Switzerland and JUPITER at JSC in Germany !

RICHARD P. FEYNMANN

"It doesn't matter how

beautiful our theory is,

It doesn't matter how

smart you are. If it doesn't

agree with experiment,

It's wrong. "

BILL @ IHPC-2015 IN TIANJING, CHINA !