PPPL HBPS WDMApp PRINCETON UNIVERSITY

Data Before Data in Fusion Science*

C.S. Chang[1] for SciDAC HBPS[2] and ECP-WDMApp[3] Teams

[1]Princeton Plasma Physics Laboratory

[2]SciDAC Center for High-fidelity Boundary Plasma Simulation

[3]ECP-App High-fidelity Whole Device Modeling Project (WDMApp)

*Funded by the DOE FES/ASCR SciDAC program and the ECP project

*Computing resources provided by OLCF and ALCF via INCITE and ALCC, and NERSC

SciDAC Scientific Discovery through Advanced Computing · Argonne NATIONAL LABORATORY · OLCF Oak Ridge Leadership Computing Facility · NERSC · ECP EXASCALE COMPUTING PROJECT

Outline

Extreme-scale/Exascale computers and AI/ML are in our playground

- We have been using after-data taken from storage in fusion science
  - Theory/computation
  - Experiment
- We need to use live-data (on-line)

  Emphases in the XGC group: examples to be presented
  - In-situ, on the fly
  - AI / Machine learning
  - Feature detection/discovery
  - Real-time control of experiment and simulation
  - Federated data management
  - Simulation acceleration
  - Extrapolation to future experiments

ADIOS

Theme of this talk

- Once the live data hit the storage,

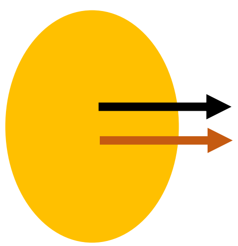
  - Beautiful “after-data” needs to be discovered and awakened by a Prince to be brought to life. Otherwise, they stay asleep for ever.

  - Many of the data in the storage are useless. They should have been thrown away.

- We need to let the data to enjoy their lives and do their work before falling into “after-data” storage. How?

Simulation data from LCF cannot be post-processed any more

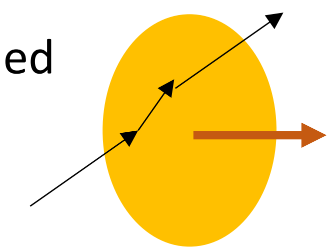
- **Data size from today's XGC simulation on 200PF Summit is too big for filesystem**
  - Trillions of particles with 10 scalars per particle x 1,000 time steps per day
    → Hundreds PB/day, cannot be allowed on the 250PB Summit GPFS
- **Exascale HPCs will produce ~EB/day data**
- **Why do we need the on-memory particle data?**
  - <u>3D particle motions</u> must be followed to calculate the <u>non-Maxwellian</u> edge plasma transport.



Simple fluid picture: plasma moves together with blobs.

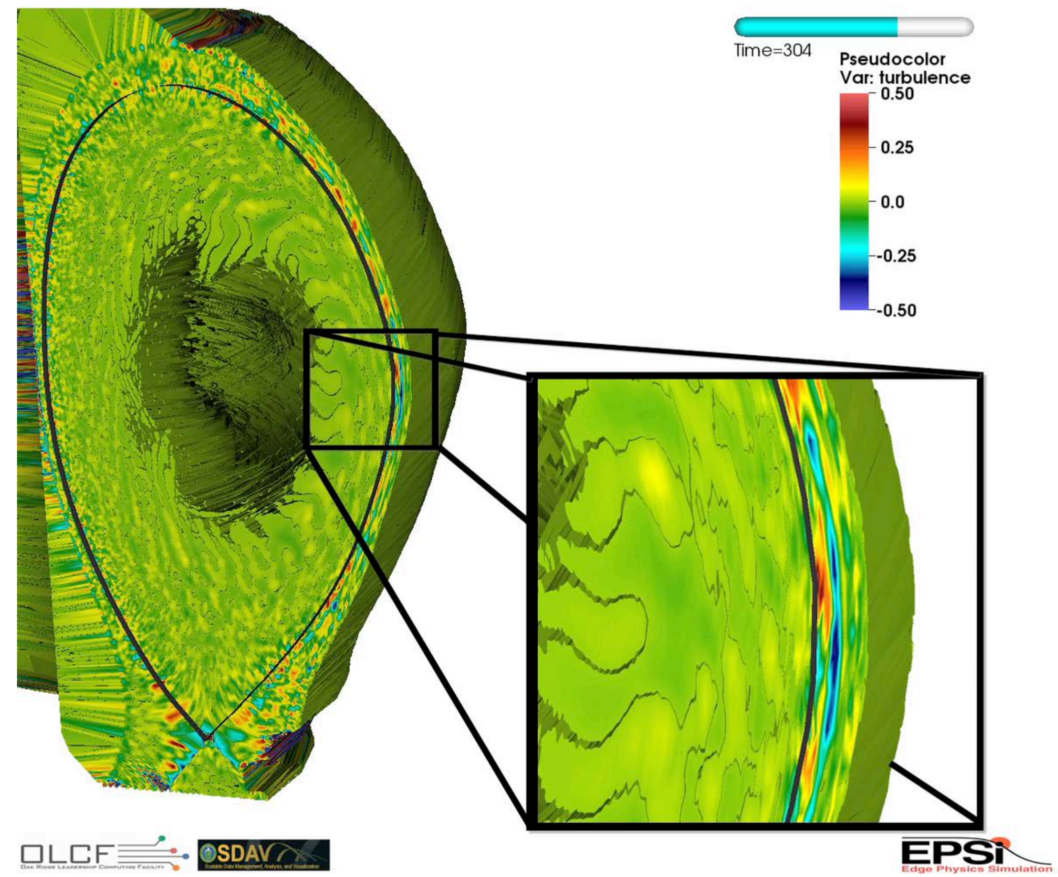
Very inaccurate in edge

AI/ML needed

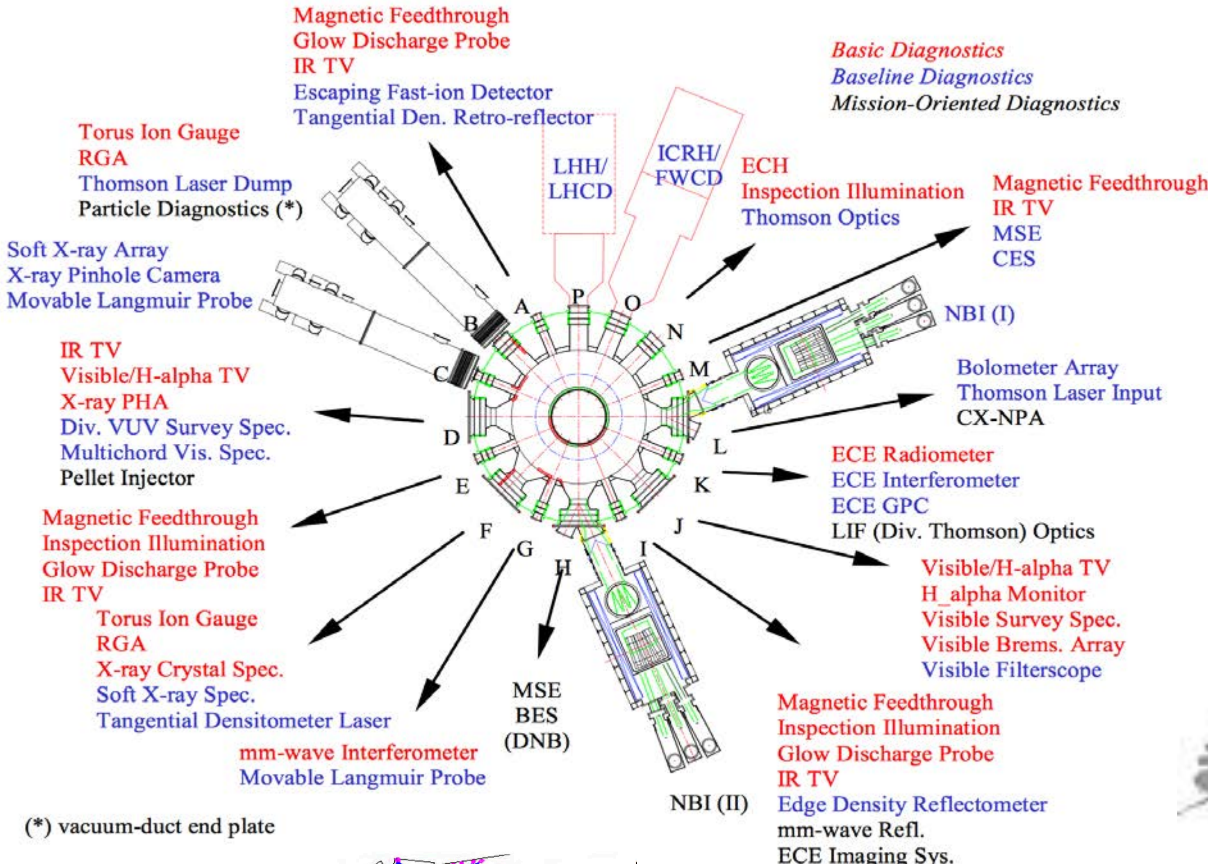


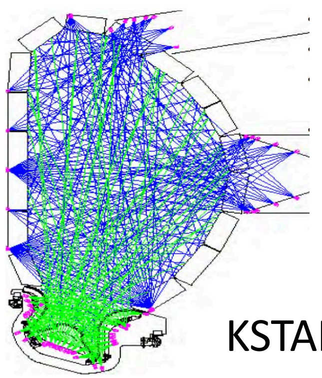
Particle picture: individual particle orbits are distorted by blobs while they go through them.

Vis by D. Pugmire

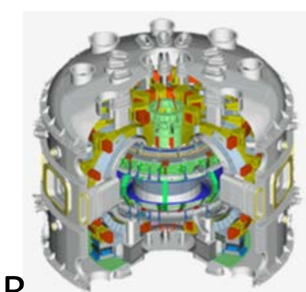


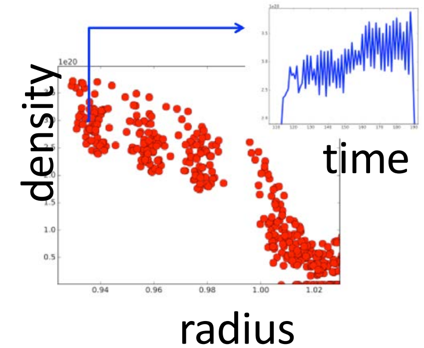
Fusion research can be greatly accelerated by federating thousands of sensors and distribted scientists/computers





KSTAR





Fusion data from ITER possess all the big-V properties

5 Vs of big ITER data

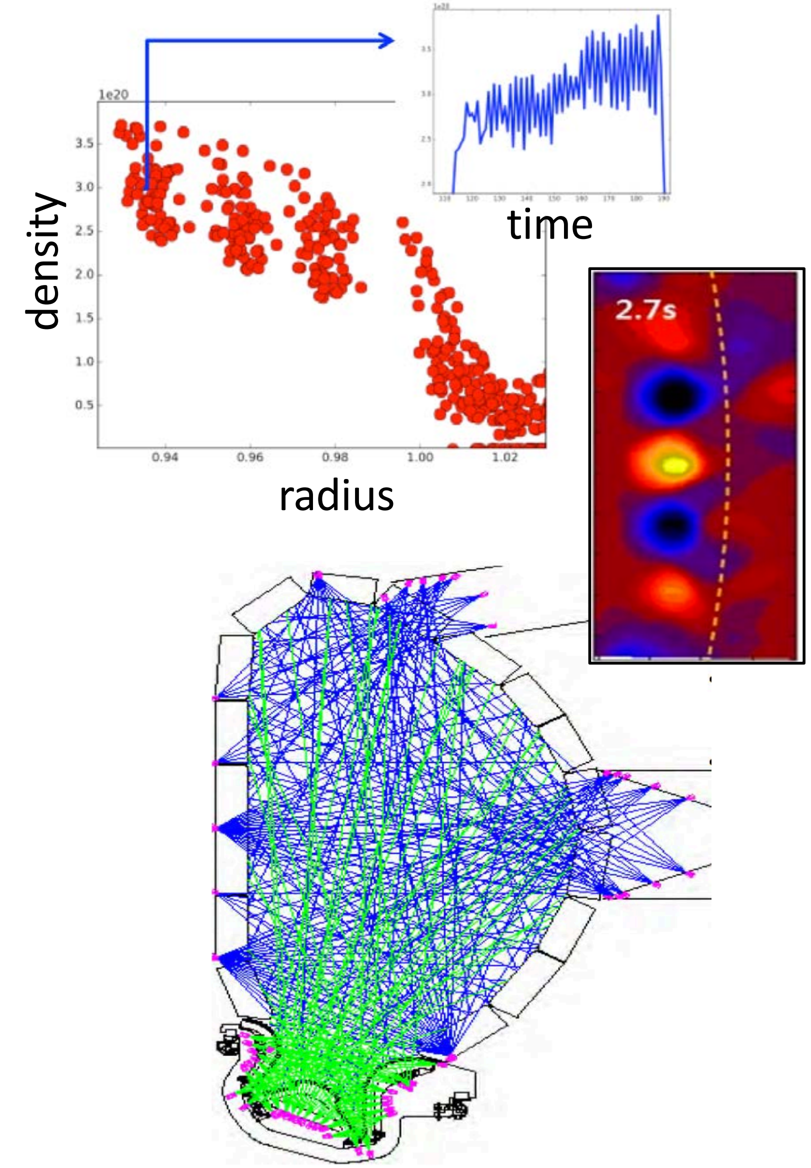
Volume
- ~1/2EB/year
- Subsequent simulations can greatly enhance data volume

Velocity
- >50GB/s for up to 1000s, each shot
- Hierarchy of analysis for real or near-real-time feedback

Value
- All the expensive sensors are there because they produce valuable data

Veracity
- All sensors have different quality: e.g., validity range, error, and noise

Variety
- Wide variety data from thousands of sensors
- Many-scale many-physics

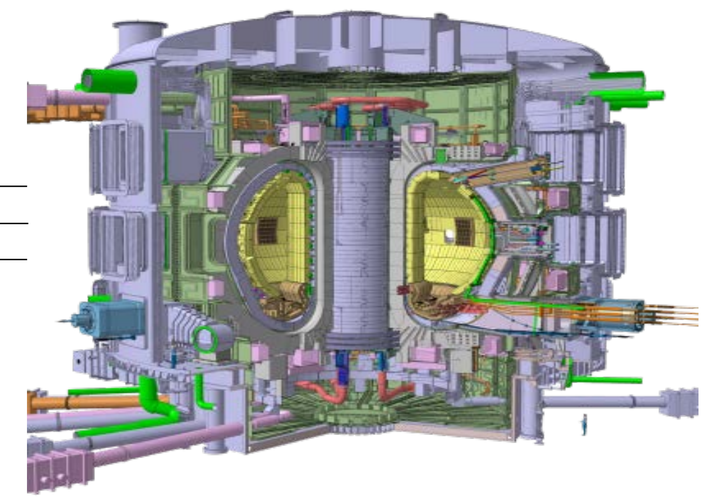


AI/ML in high demand

We have been using after-data → very slow scientific advance

ITER's progress plan is based upon this type experience.

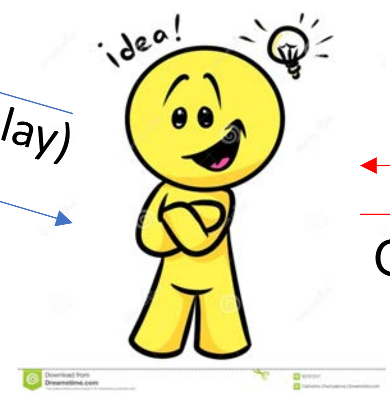
There is also a critical need for real-time control.



I observed something that seems to be important, and published it. But, ?

Months or years later, a theoretician heard/remembered about the experiment

(further delay)



Data back to life

Give me the data, PLEASE (human relationship, bureaucracy)

We understand it now.
Do "this" in the next experiment.
But, months or years have passed.

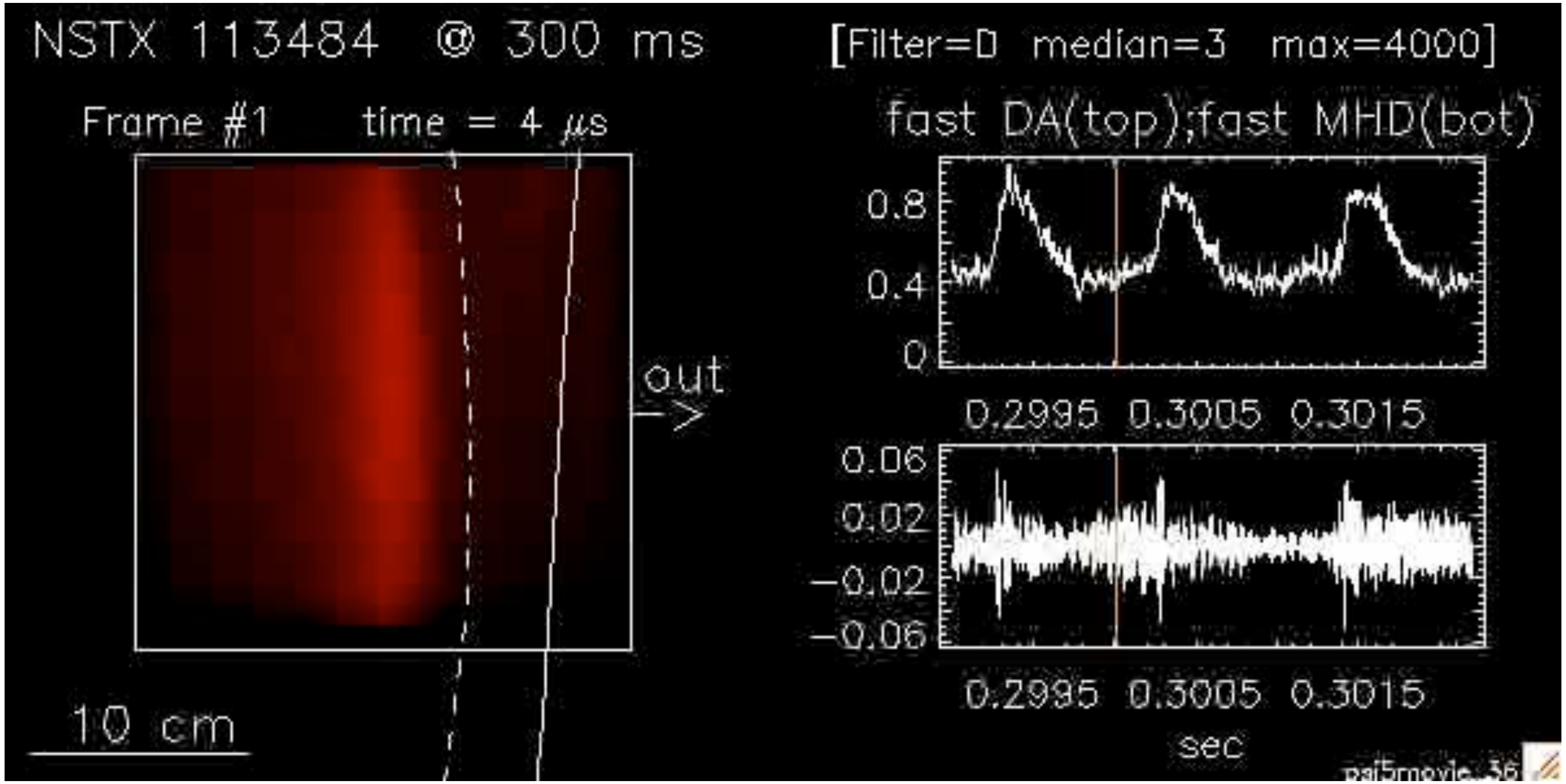
ITER requires real-time control

- Disruptions can damage the machine's structural integrity (~50ms response): AI/ML work in progress
- **Edge-localized-mode (ELM)** crashes will destroy ITER wall: AI/ML work at infancy
  Requires ~10ms response from physics precursor detection to actuator trigger

(Zweben)

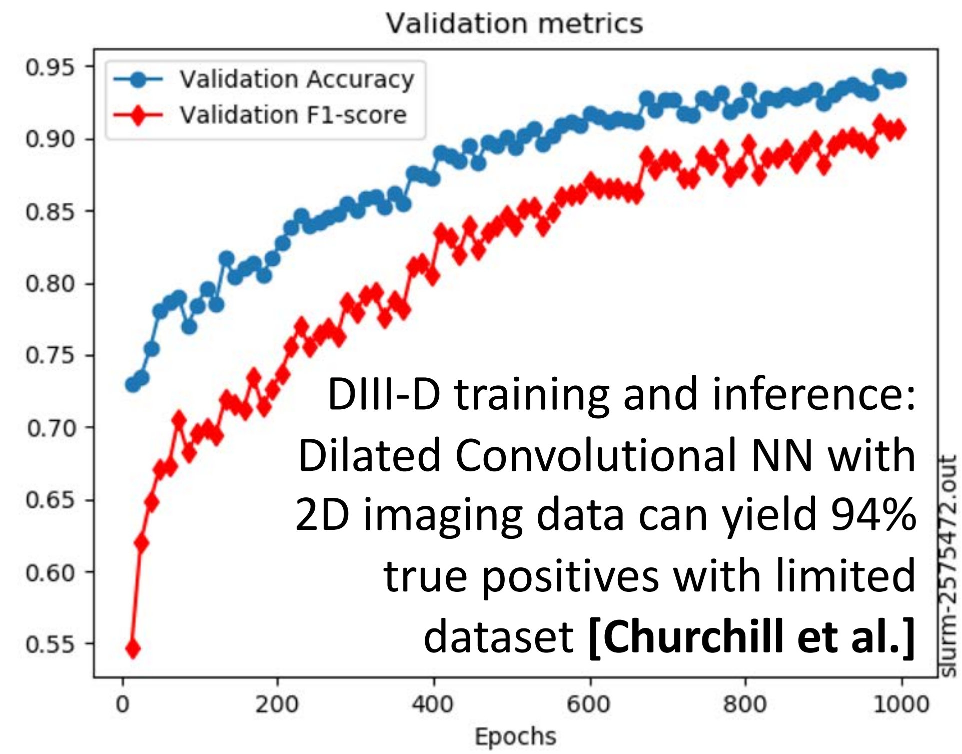
- Human intervention not possible
- Must be a highly trained AI/ML system
  - Detection
  - Actuator trigger
- Local inference near sensors, but remote training on HPCs

→ Federated data management system



ML prediction for disruption is making further progress

Classical ML →
DL + 0d →
DL + 1d profile (Kates-Harbeck et al.) →
Dilated Convolution NN + 2D instability dynamics
(R.M. Churchill et al., XGC group)



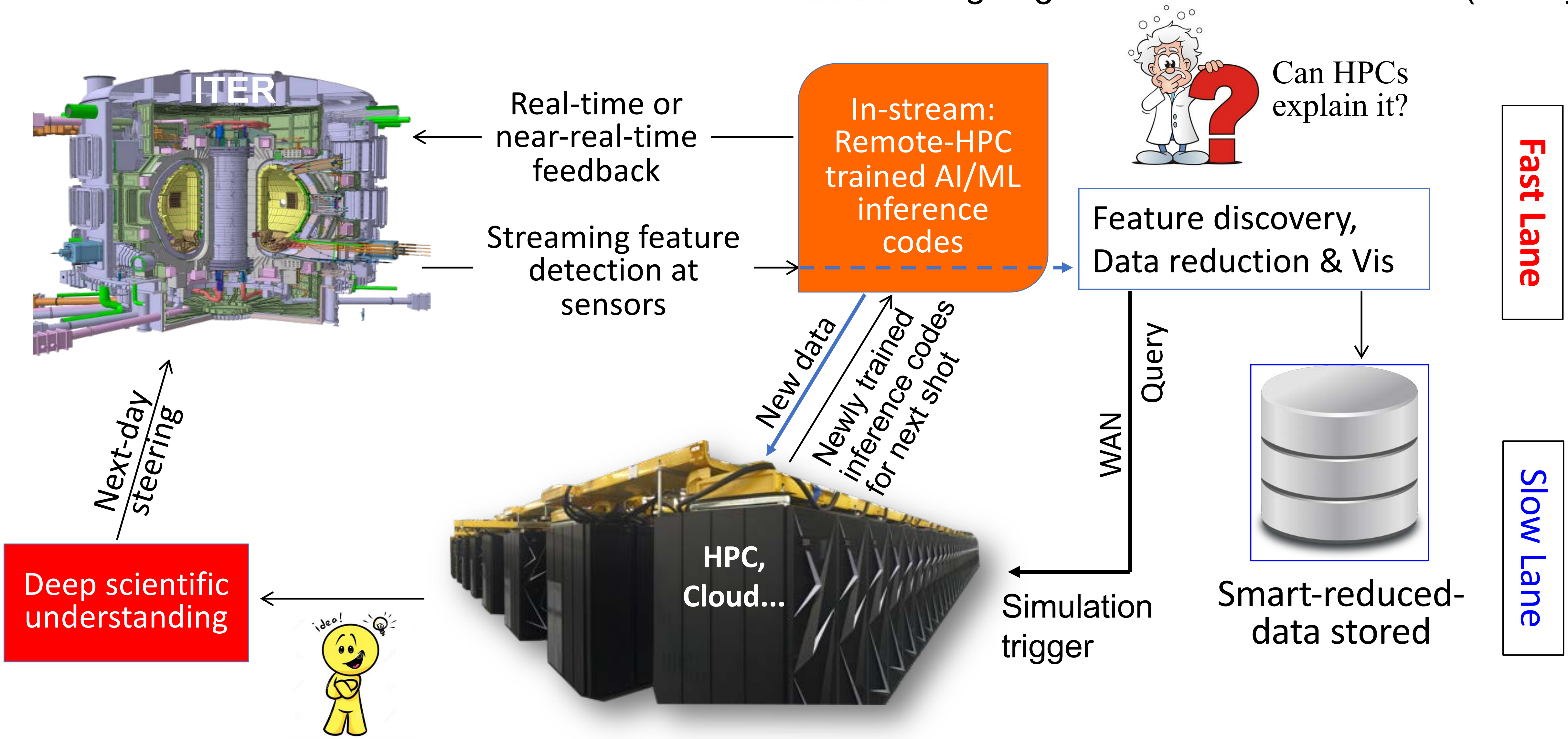
DIII-D training and inference: Dilated Convolutional NN with 2D imaging data can yield 94% true positives with limited dataset **[Churchill et al.]**

Model / AUC	AUC: JET training, JET inference	DIII-D training, JET inference	DIII-D training + 5 JET data, JET inference
DL + 1d profile		0.836	0.911 (Kate-Harbeck et al)
DL + 0d	0.952	0.817	0.879
Classical ML	0.893	0.616	0.851

[Kates-harbeck]

Federated Fusion Data System to Accelerate Fusion Research

ADIOS On-going collaboration with ORNL (Klasky)



Many scientists on multiple HPCs in different countries

Use ML for simulation acceleration: Fokker-Planck Eq.

(R.M. Churchill)

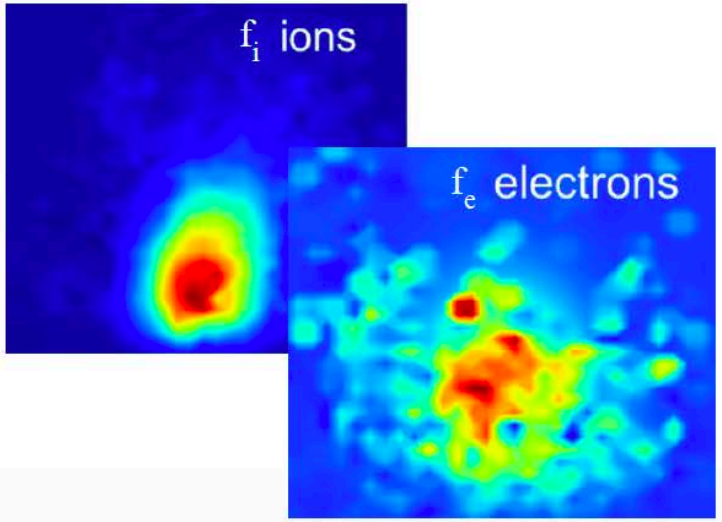
- **Collision operator equation maps e.g.** $\{f_a, f_b\} \rightarrow \{\delta f_a\}$

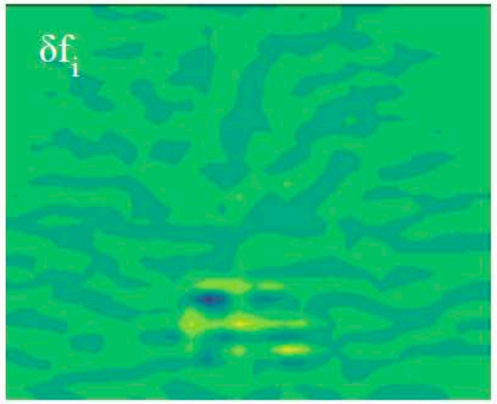
$$\left.\frac{d f_a}{dt}\right|_{col} = \sum_b^N C_{ab}(f_a, f_b')$$

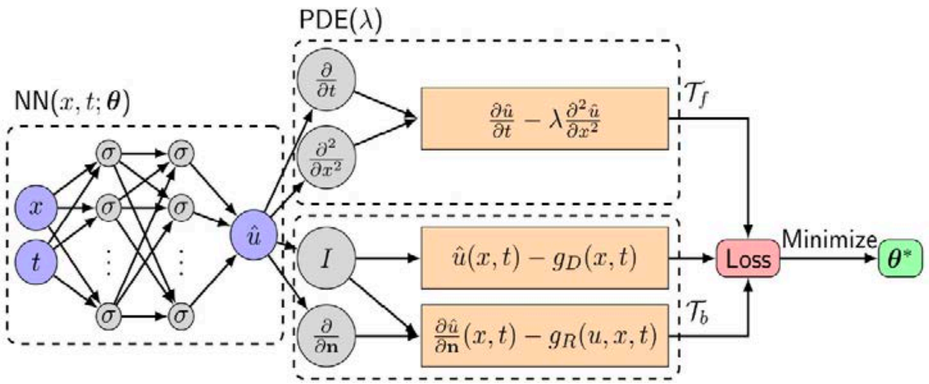
$$= -\sum_b^N \frac{e_a^2 e_b^2 \ln \Lambda_{ab}}{8\pi \epsilon_0^2 m_a} \nabla \cdot \left[\int d^3 v' \, \underline{\mathbf{U}} \cdot \left(\frac{f_a}{m_b} \nabla' f_b' - \frac{f_b'}{m_a} \nabla f_a \right) \right]$$

- **Task then is minimizing the per-velocity bin mean-square error (MSE) loss, averaged over all velocity bins.**

- Multispecies collision operation is a serious issue in kinetic ITER simulation: cost ∝ N^2
- We utilize Semantic Segmentation task: Hourglass Convolutional NN
- Include physics constraints in the loss after the main NN





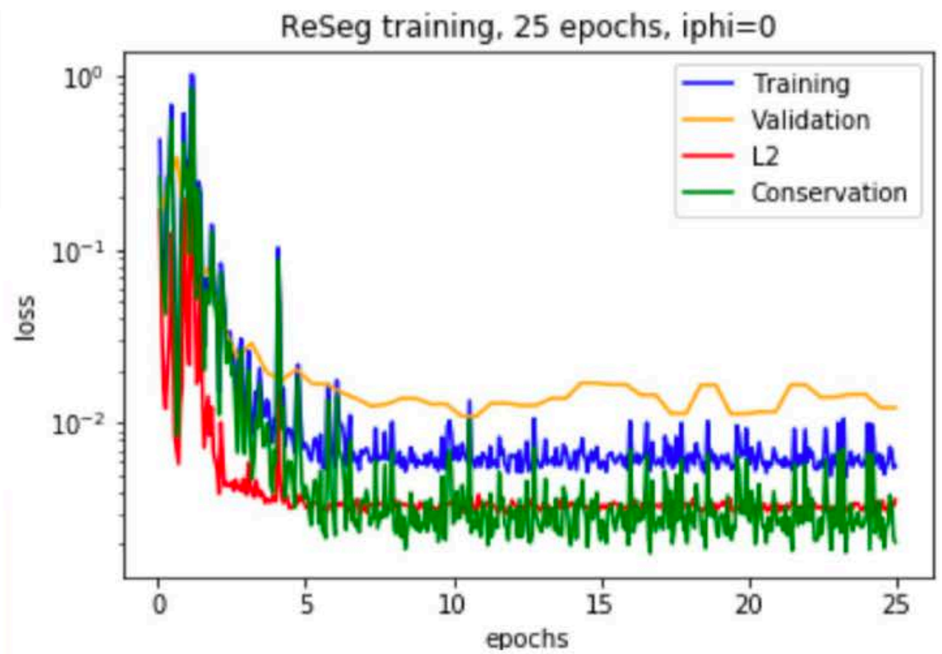


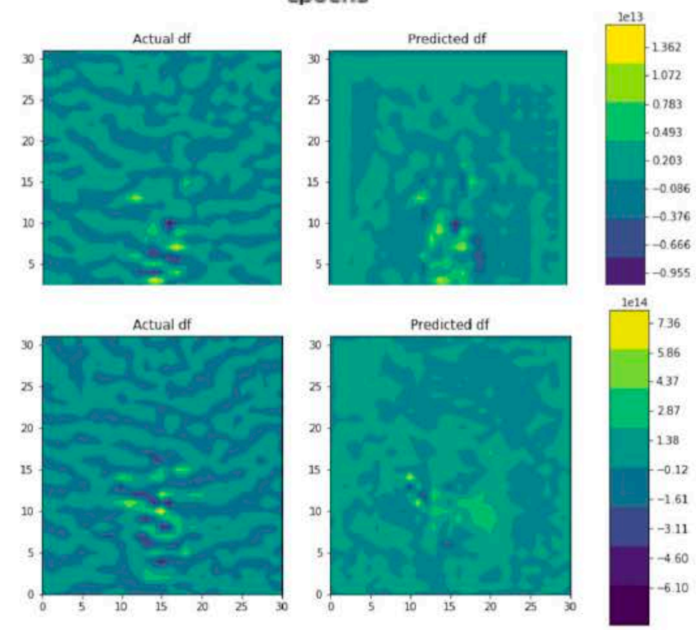
Deep NN trained Fokker-Planck operator: initial result

(R.M. Churchill)

- **Collision operator neural network trained on ReSeg architecture, using both L2 and conservation losses**
- **Single XGC1 simulation time point (JET discharge), ~17k "images"**
- **Mean conservation loss:**
    - Density: 1.61e-4
    - Momentum: 1.22e-3
    - Energy: 1.040e-5

  Target $\leq 10^{-5}$
- **Further work using more data and harder constraints is ongoing.**



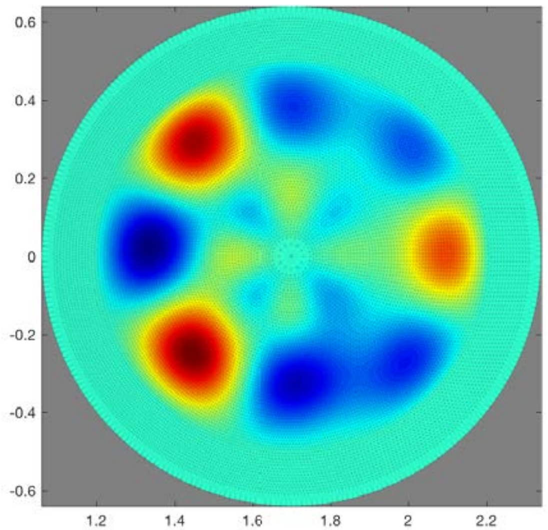


[Collaboration with ANL (T. Munson) has been initiated]

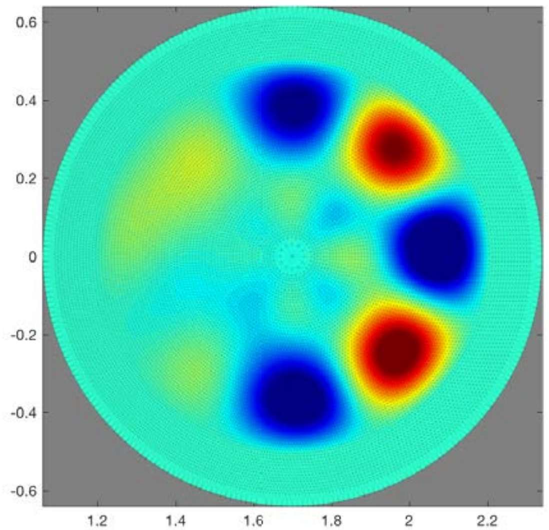
Toward Machine Learning in XGC's iterative EM solver preconditioner

R. Archibald (ORNL), Ben Sturdevant and C-S Chang (PPPL)

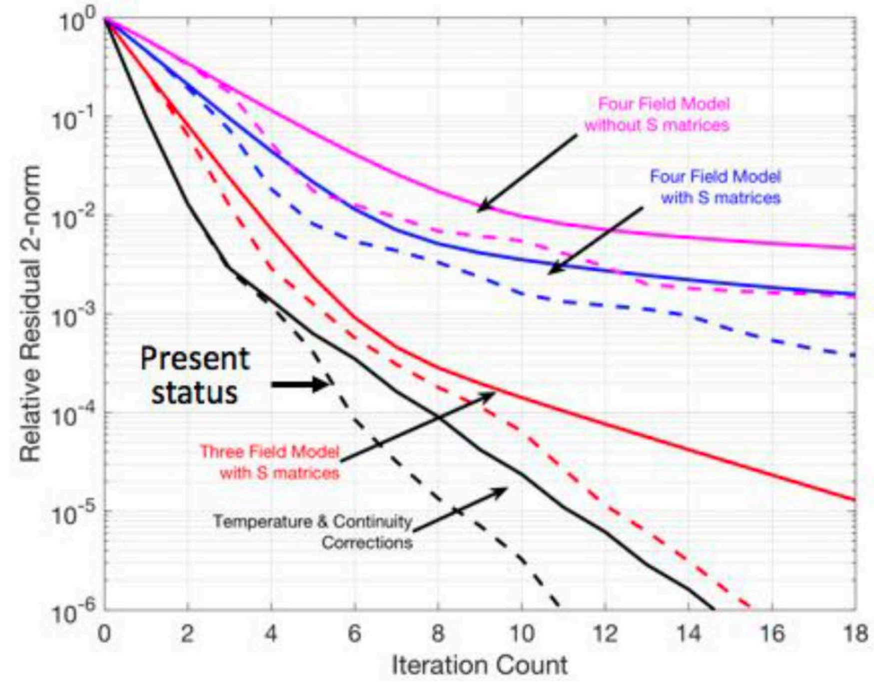
- **Goal: Use machine learning to act as effective preconditioner for fully implicit kinetic EM scheme in XGC**



Initial Condition of iterative solve



ML Prediction from Initial Condition



- **Result:** Use small XGC run, validated that machine learning (Convolutional DNN) that initial condition guess can be improved by over a factor of 4X in relative error.
- **Challenge:** Improve accuracy. Translate improvements in initial guess to reduced iterations of fully implicit XGC. Build lightweight adaptive training of Convolutional DNN.

For more information: Rick Archibald (ArchibaldRK@ornl.gov), Data Analytics lead

Anchored ML for prediction: Anchor the under-determined ML, using predictive data from first-principles-based simulations

(C.S. Chang)

Present data is usually an under-determined set: solution may not be extrapolated

Data from today's experiments → Interpolative AI/ML for today's experiment → Interpolation-based AI/ML prediction for future experiment **Validity?**

+

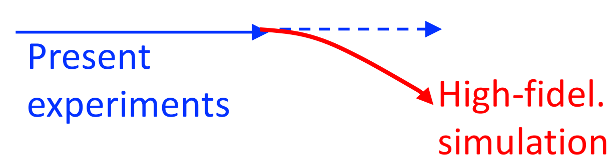
Data from predictive simulations

Must be validated on today's experiments

Interpolative AI/ML for today's experiment → Anchor-based extrapolative AI/ML for future experiment High-fidelity simulations

$\mathbf{D(x_i) \rightarrow y_k}$

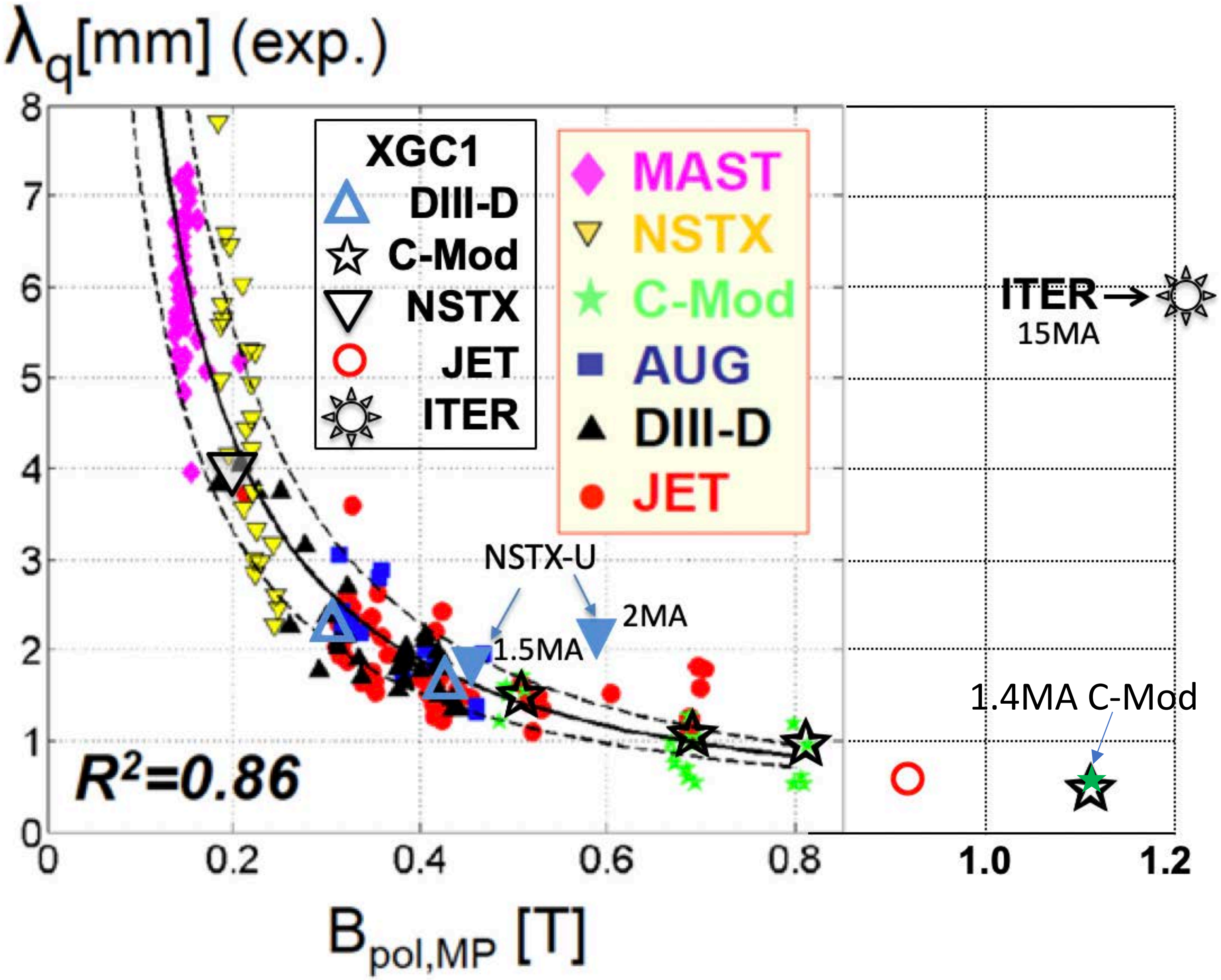
$\mathbf{D'(x_i + x_i') \rightarrow Y_k (\supset y_k)}$



This technique will get better as the accuracy and sampling of x_i' improves on more powerful computers with better algorithms (AI/ML algorithms included).

Predictions from gyrokinetic XGC have been validated on existing tokamaks

(Summit, Cori)



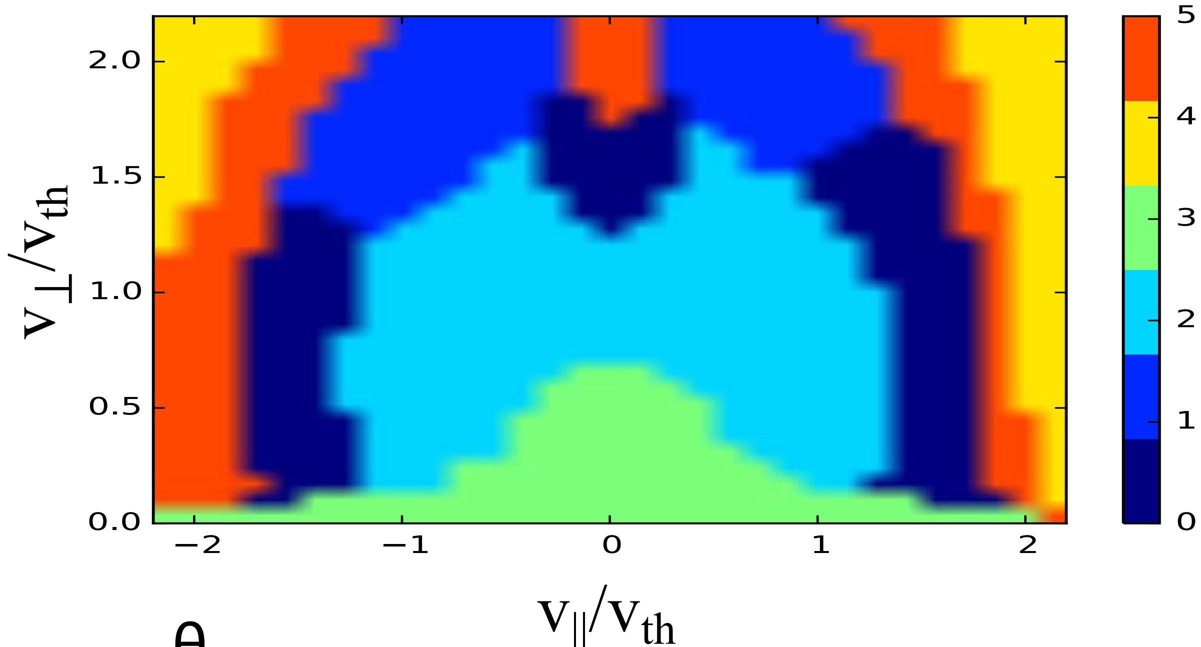
- But, the same code predicts λ_q(XGC) >6λ_q(Exp) for full-power ITER
- This result was challenged by the high-current C-Mod experiments.
  - ○ Double valued?
  - ○ Suggestive of hidden parameters
  - ○ Or XGC is wrong
- XGC on NSTX-U at 2MA also produced a wider λ_q
  - But, not at 1.5MA
  - Hidden parameters, again?

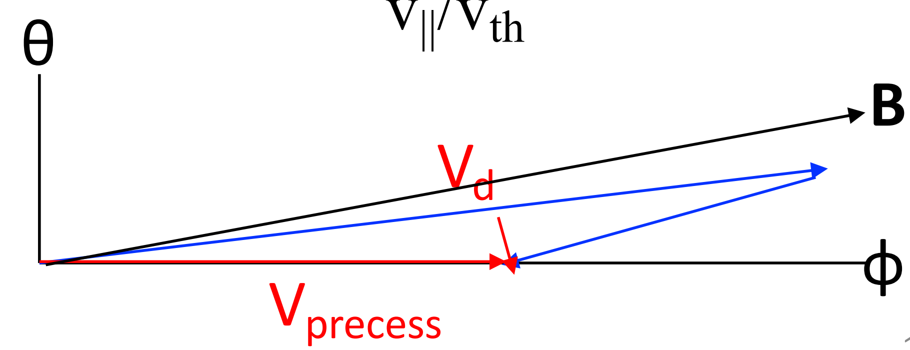
Machine learning reveals trapped electron interaction with turbulence in the 15MA ITER edge (R.M. Churchill)

(Summit data, NERSC)

- K-means clustering, with K=6
- At a higher energy band, trapped electrons show correlated response to turbulence
    - Another sign of TEM turbulence
- Because of the high $\omega_* \sim v(\rho/L)$ around the separatrix, q needs to be high for precession resonance by trapped electrons: $V_{precess} \sim v(\rho/R)(B/B_P)$

→ easier excitation of TEMs just inside the separatrix, ψ_N=0.98-1, where ∇T_e is high.

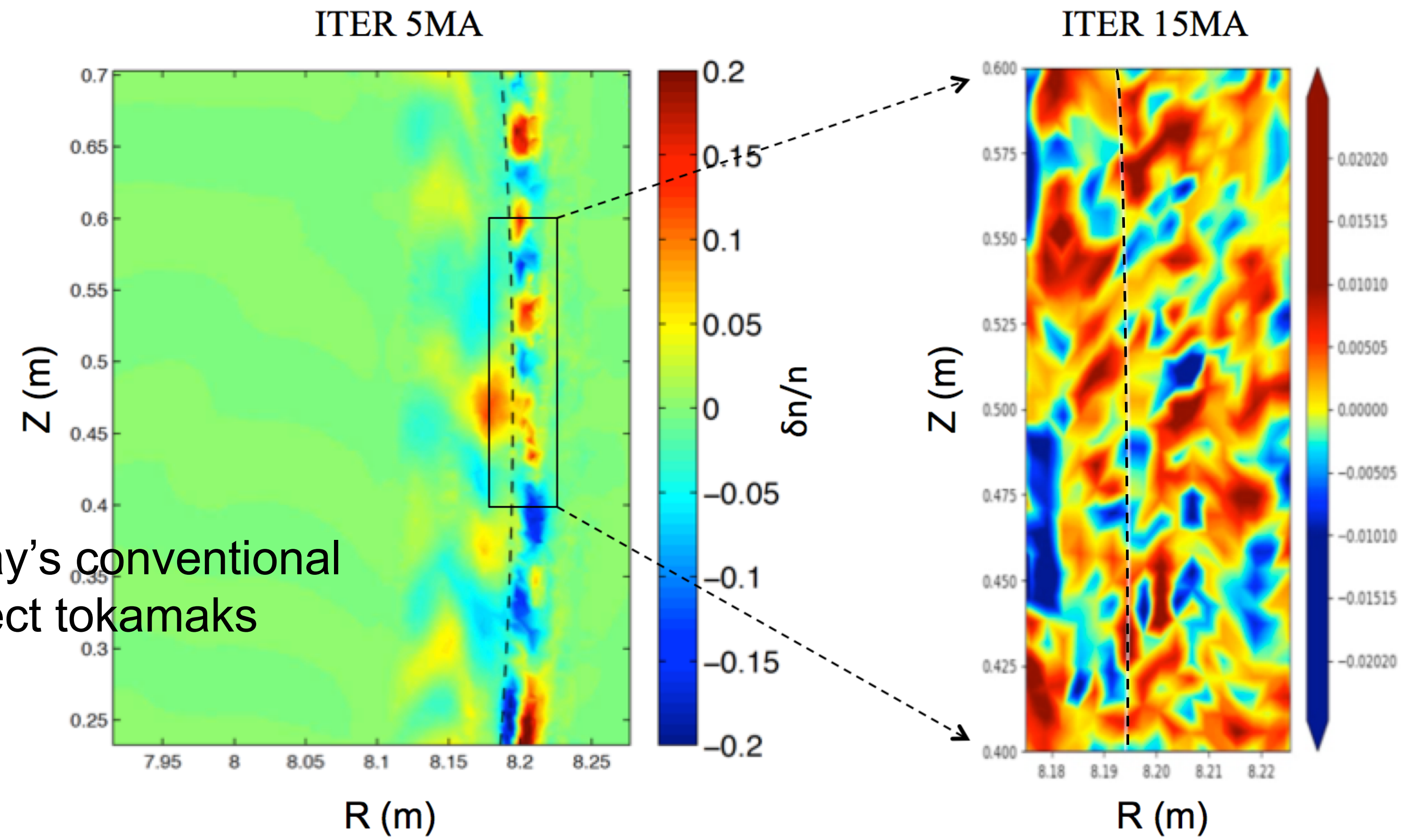




Electron heat-spread by strong trapped electron modes is the prime suspects

- **Fact: $\rho_{ip}/a \rightarrow 0$ in 15MA ITER yields little neoclassical ExB shearing,**
- **Fact: $(2a/R)^{1/2} \rightarrow 1$ in NSTX-U with warm T_e yields strong TEM drive**

(Summit)



Today's conventional aspect tokamaks

TEM streamers are the suspects. ITGs do not penetrated into SOL.

Full-current ITER

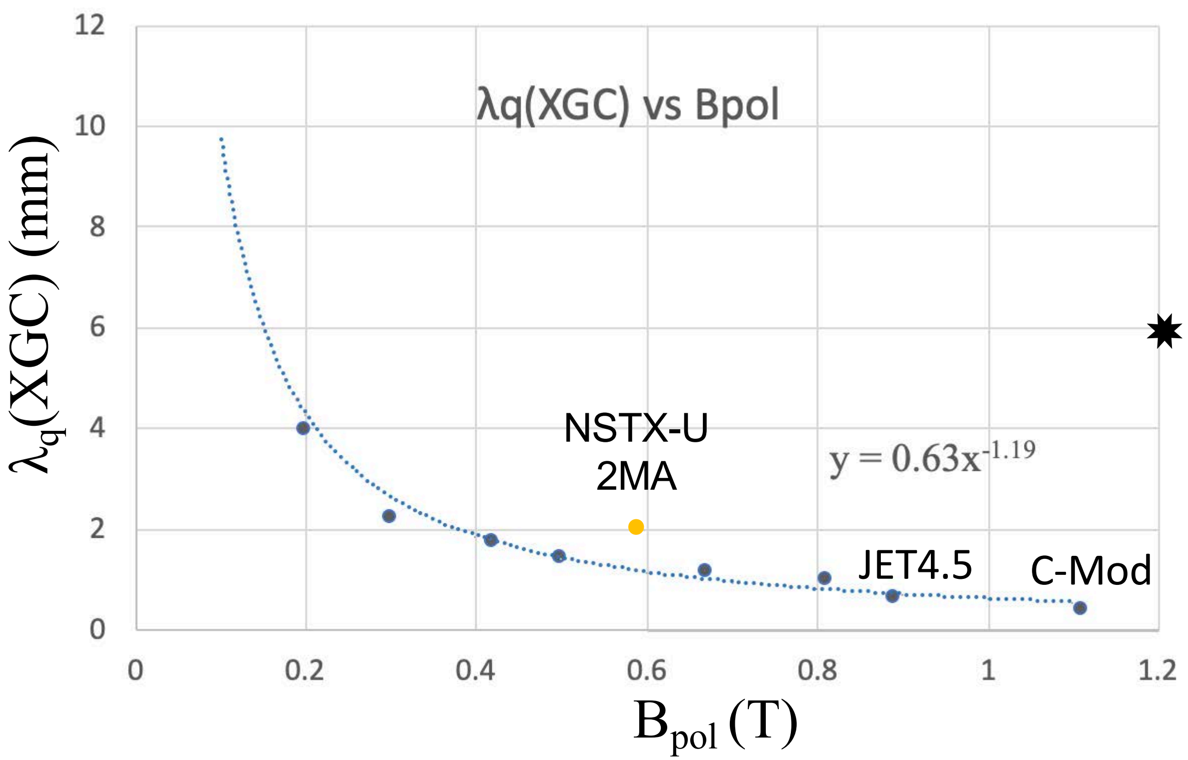
Isolated "**blobby**" turbulence (with **strong sheared-ExB** flow across separatrix)

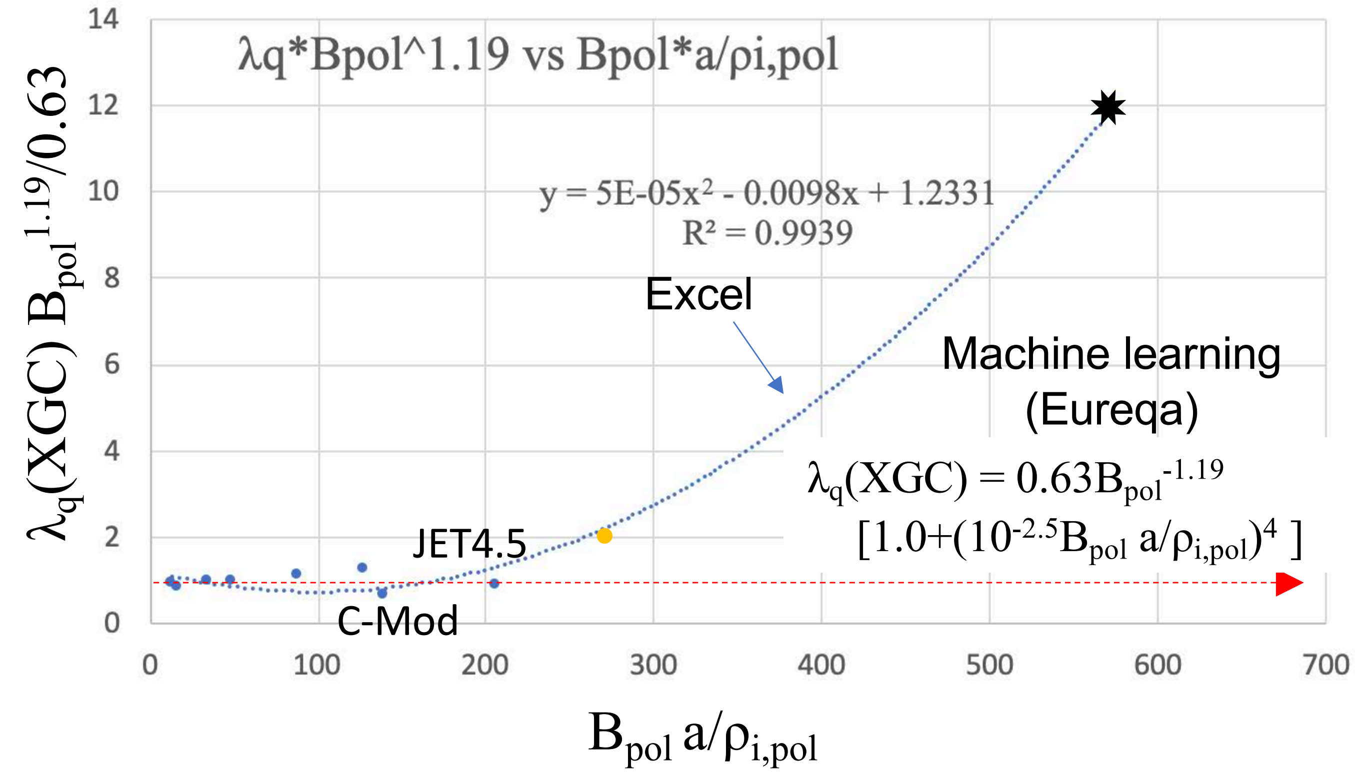
Connected "**streamer**"-type turbulence (with **weak sheared-ExB** flow across separatrix)

Looking for hidden parameters with feature engineering

- Large $a/\rho_{i,pol}$ weakens the neoclassical ExB shearing rate → stronger TEM

(C.S. Chang)

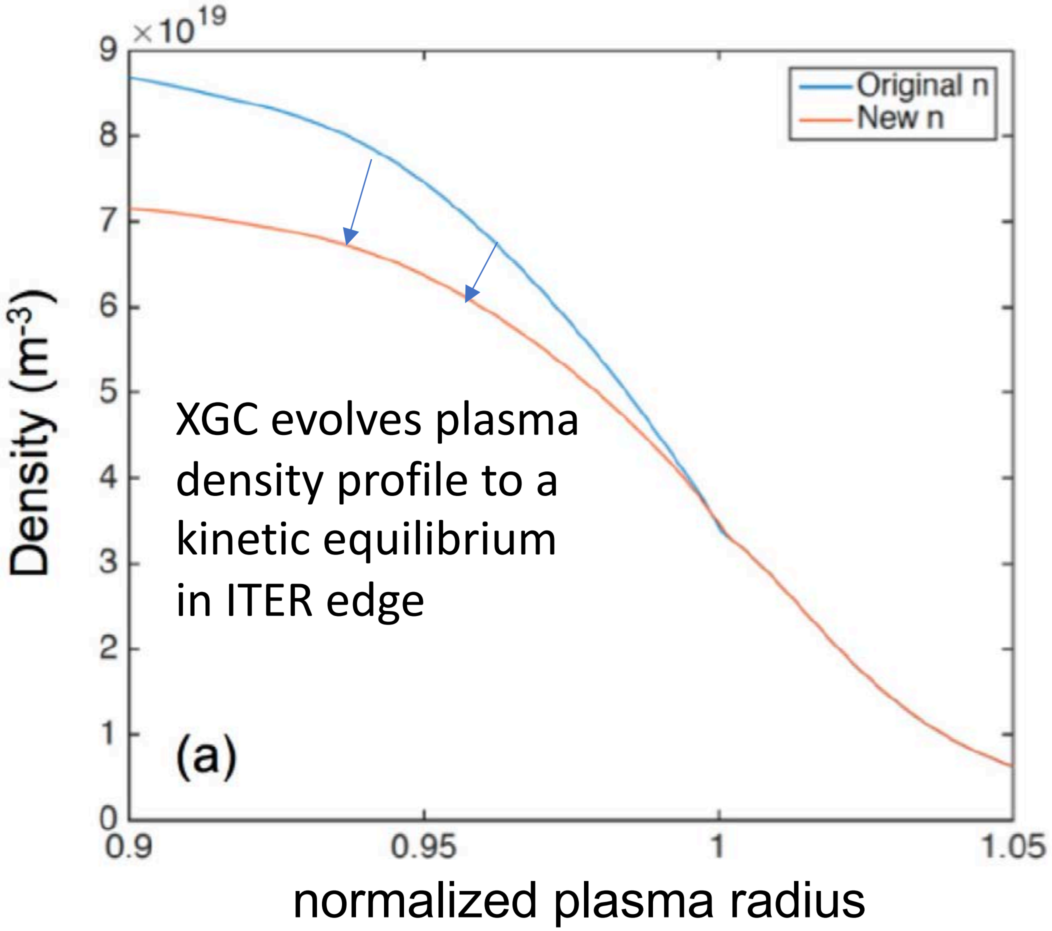




- In the present conventional aspect-ratio tokamaks, λ_q(XGC) follows λ_q(Exp).
- However, λ_q(XGC) shows double-value between high-Ip C-Mod and 15MA ITER.

- When we use $B_{pol}\ a/\rho_{i,pol}$ as the scaling variable,
  - λ_q(XGC) in the present tokamaks still follows λ_q(Exp)
  - and the double-valuedness disappears
  - NSTX-U 2MA data follows the ML anchored curve

There are several other AI/ML applications in dire needs*



- Kinetic evolution of background plasma profile is computationally expensive, but a critical problem for ITER.
  - When the initial plasma profile is far away from kinetic solution, XGC could spend 10x more exascale computing time
- Can we use AI/ML to telescope the background profile evolution?
  - Tried a Beysian algorithm, but with only a limited success

*Not all the collaborations are represented here; e.g.,
- Hanqui Guo: Deep learning feature discovery from blob-turbulence isocontours
- Jong Choi: this workshop

Federated Fusion Data System to Accelerate Fusion Research

ADIOS On-going collaboration with ORNL (Klasky)