Second Visualizing Offline and Live Data with AI (VOLDA) Workshop - 2025

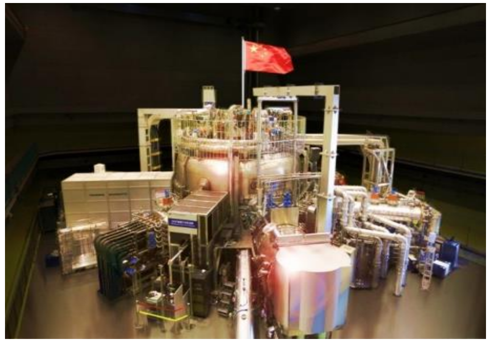
Intelligent processing of multimodal diagnostic data on EAST tokamak

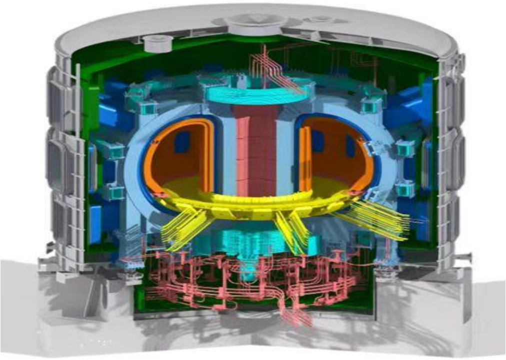
T. Lan[1], F. F. Long[2], S. X. Wang[1], L. F. Yang[1], J. S. Zhang[2], X. P. Xie[2], S. Z. Yuan[1], Y. F. Wang[1], H. B. Shen[1], Q. L. Ren[1], H. Q. Liu[1*], Y. T. Song[1], and EAST Team.

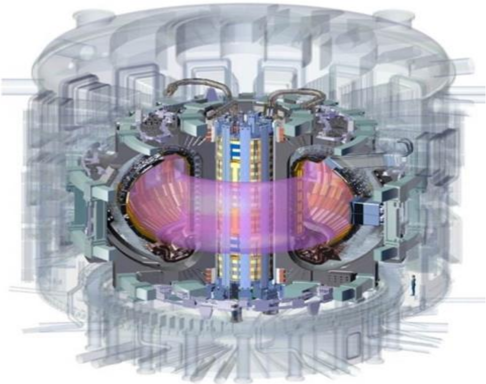
1. Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, Anhui 230031, China
2. School of Nuclear Science and Technology, University of Science and Technology of China, Hefei, Anhui 230026, China

ASIPP

AI for fusion







Introduce AI to serve the future burning plasma research

Simulation
Heavy calculation burden, difficulty to balance accuracy and speed

Theory
Multiscale and nonlinear physical processes

Experiment
Challenging engineering, and inconsistent data

Burning plasma:

More complex physics,

Greater computation burden,

Incomplete data,

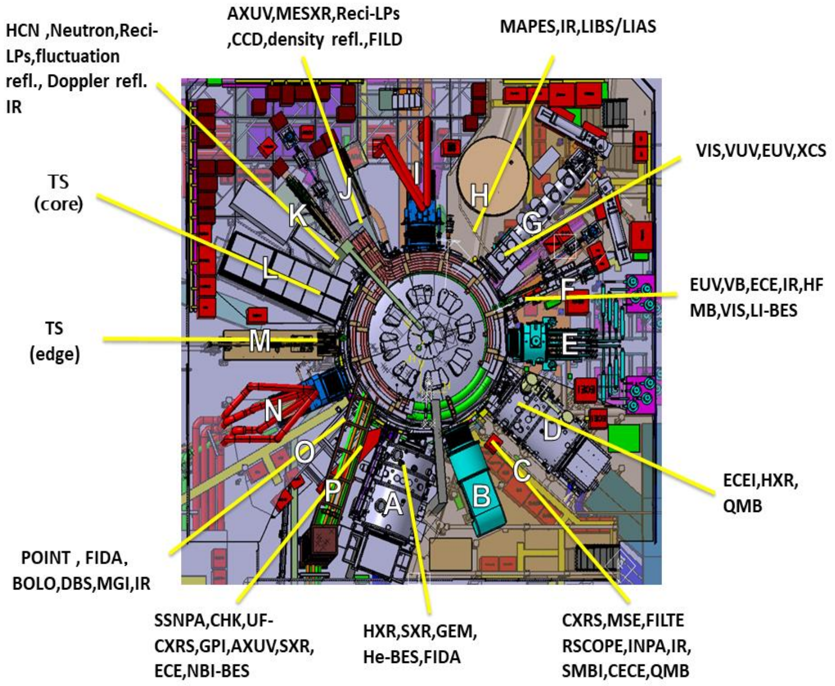
Unclear physical processes,

……

ASIPP

Multimodal diagnostic data on EAST tokamak

Layout of diagnostic systems on EAST tokamak



Types of fusion data

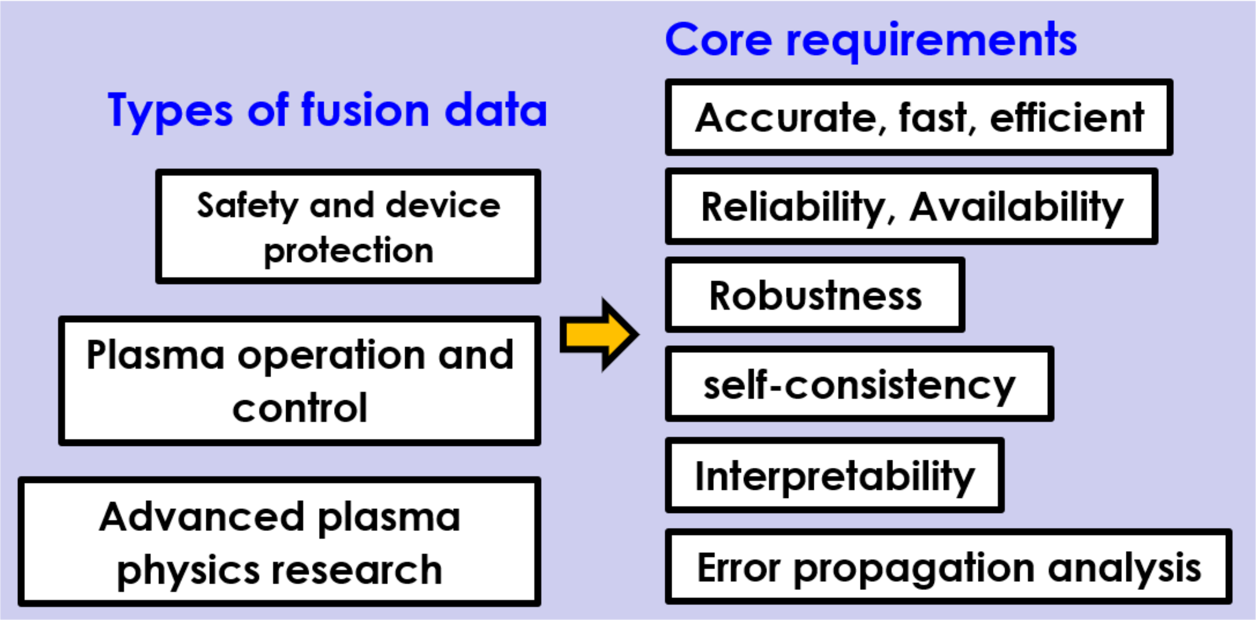
- Safety and device protection
- Plasma operation and control
- Advanced plasma physics research

➡

Core requirements

- Accurate, fast, efficient
- Reliability, Availability
- Robustness
- self-consistency
- Interpretability
- Error propagation analysis

Intelligent processing of multimodal diagnostic data on EAST tokamak

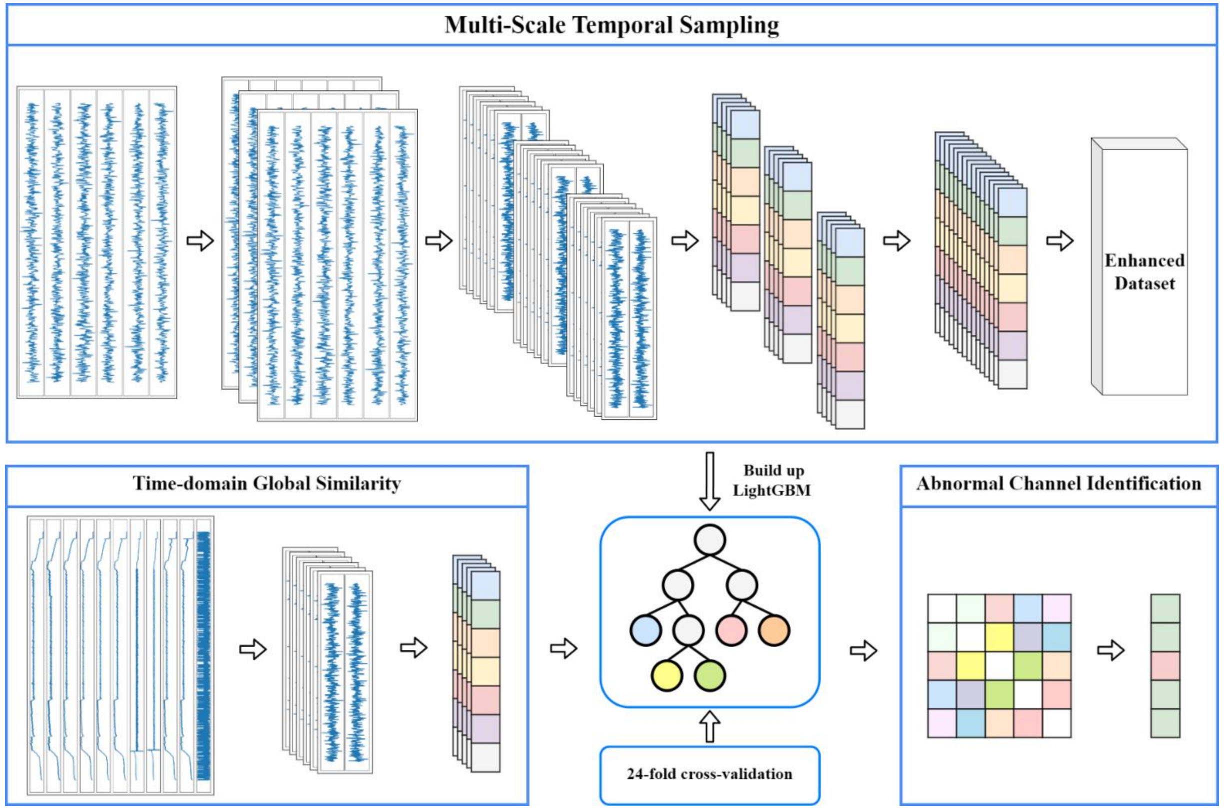


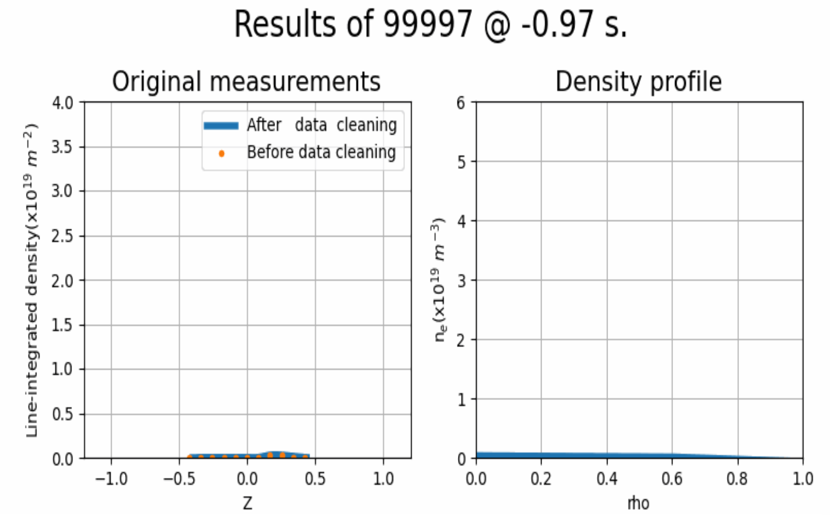
- **Improve reliability and availability of data:**
  - Real-time data cleaning
  - Phase jump correction
- **Improve speed and accuracy of data processing:**
  - Density profile inversion
  - CTS spectral analysis
- **Data consistency and uncertainty quantification:**
  - Multi-system data fusion
  - VVUQ

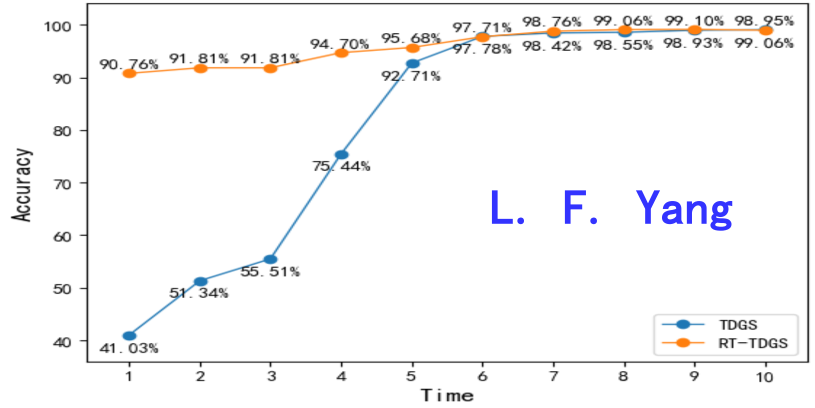
Real time data cleaning — Improve data reliability

- Data cleaning: accurately identify the useless data
- Apply RT-TDGS to EAST POINT system (accuracy: 0.9891±0.0038; time delay: 4.75 ms)

RT-TDGS（Real-Time Time-domain Global Similarity）



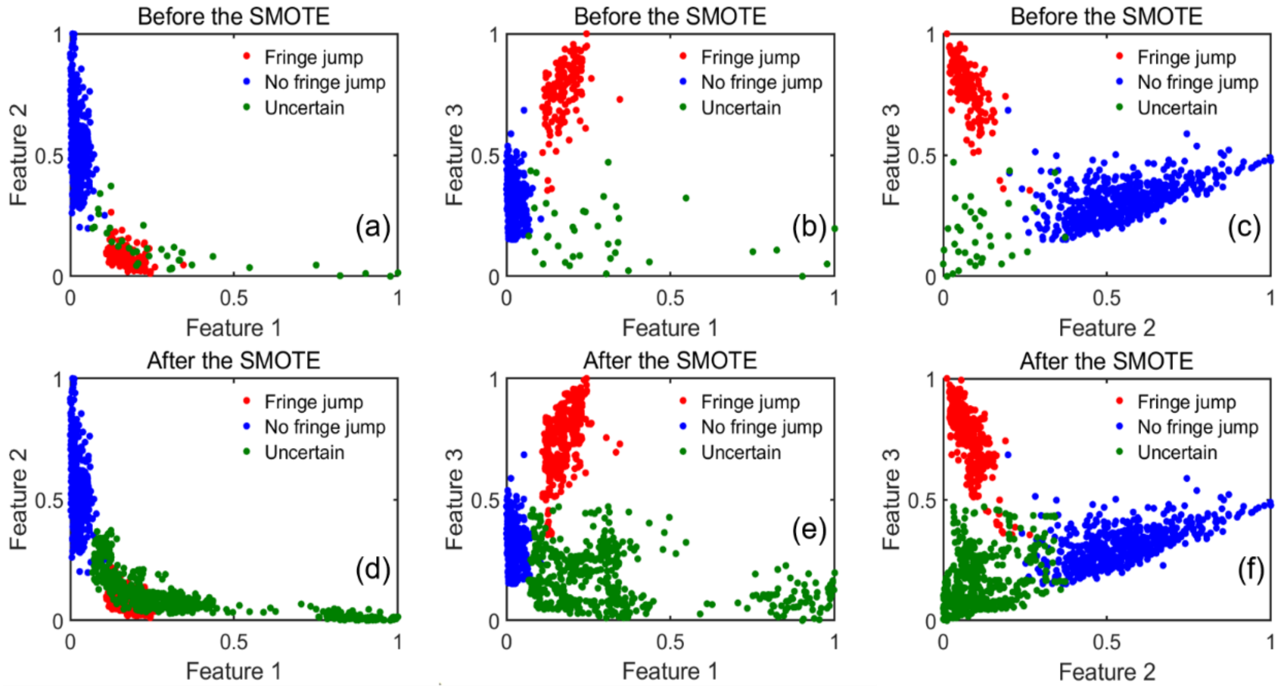




Yang L F., et al., ECPD, poster, 2025.

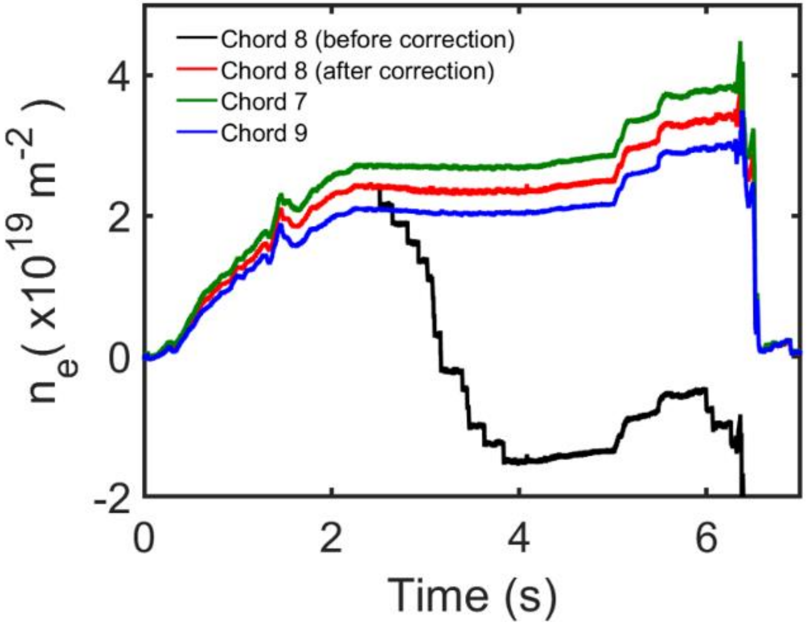
Automatic phase jump correction – Enhance availability

- A model based on supervised learning was established, which can detect and correct the phase jump in density measurement.
- This work provides a reference for data correction in other diagnostic systems.



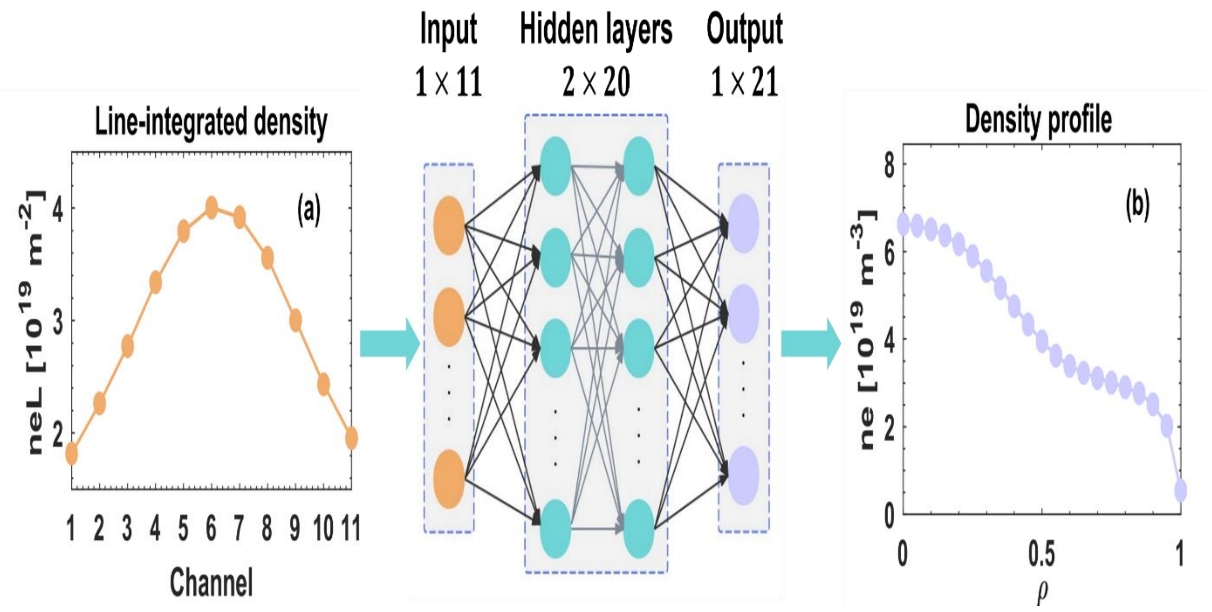
Wang Y., et al., PPCF, 2023, 65: 065001.

Y. F. Wang



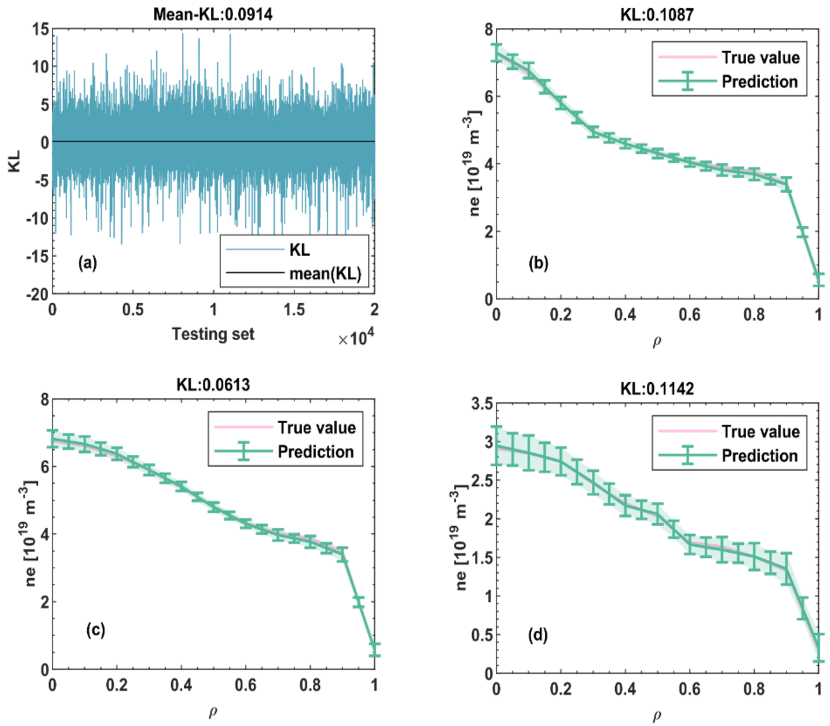
Density profile inversion - Improve processing speed

- BPNN model calculates density profile from the 11-channel integral measurement of POINT.
- Compared with park-matrix method, BPNN model is ten times faster.



Xie XP., et al., PPCF, 2025, 67: 045001.

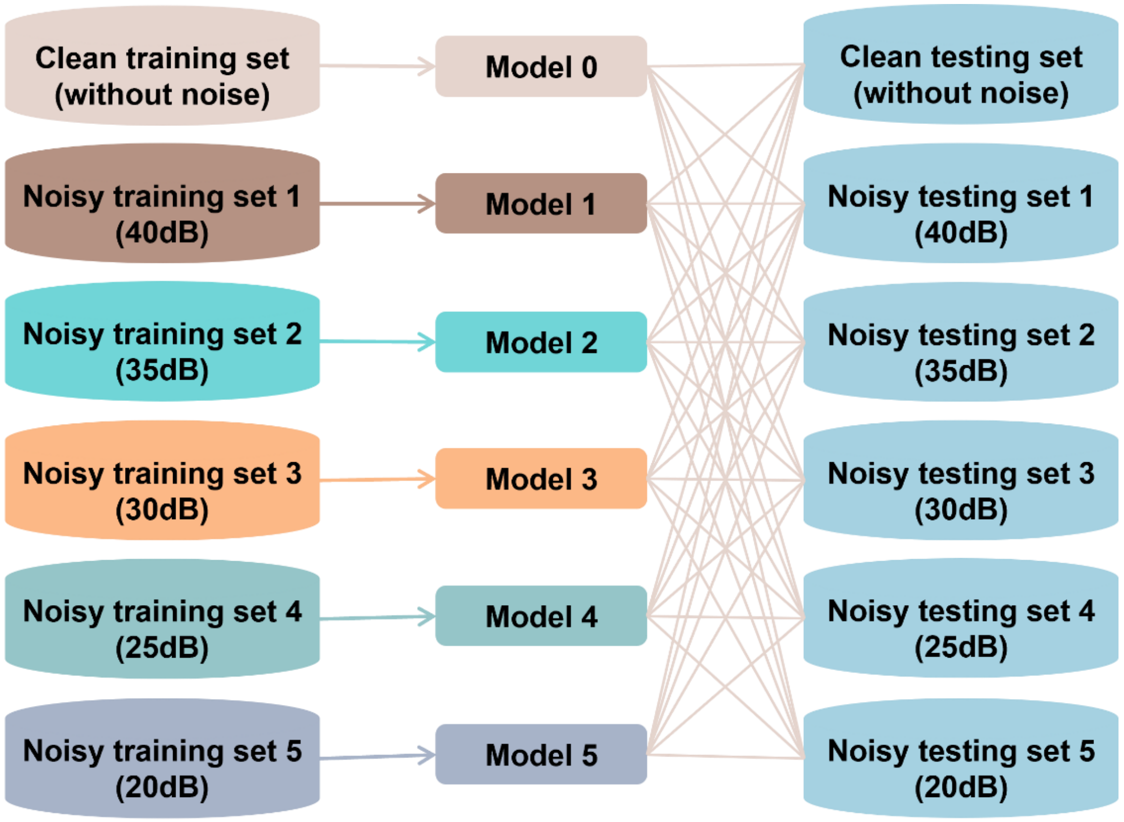
X. P. Xie



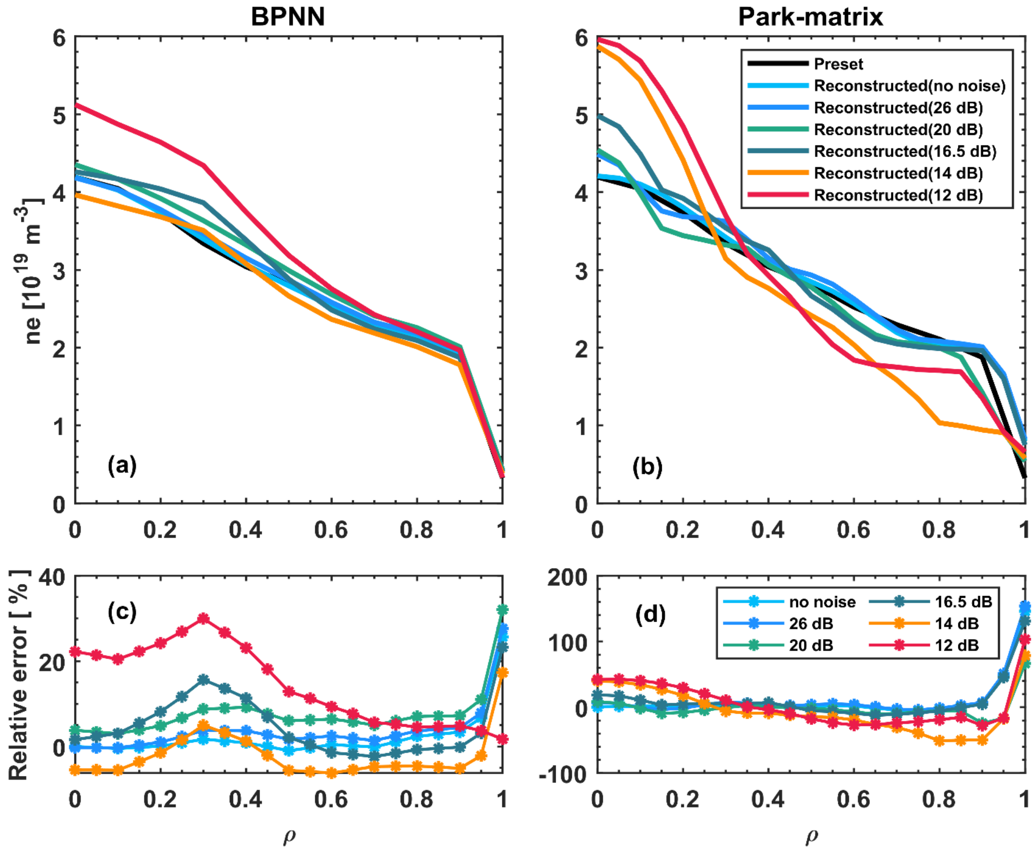
ASIPP

Robustness optimization of BPNN inversion model

Training with noisy data



BPNN is robust against noise!



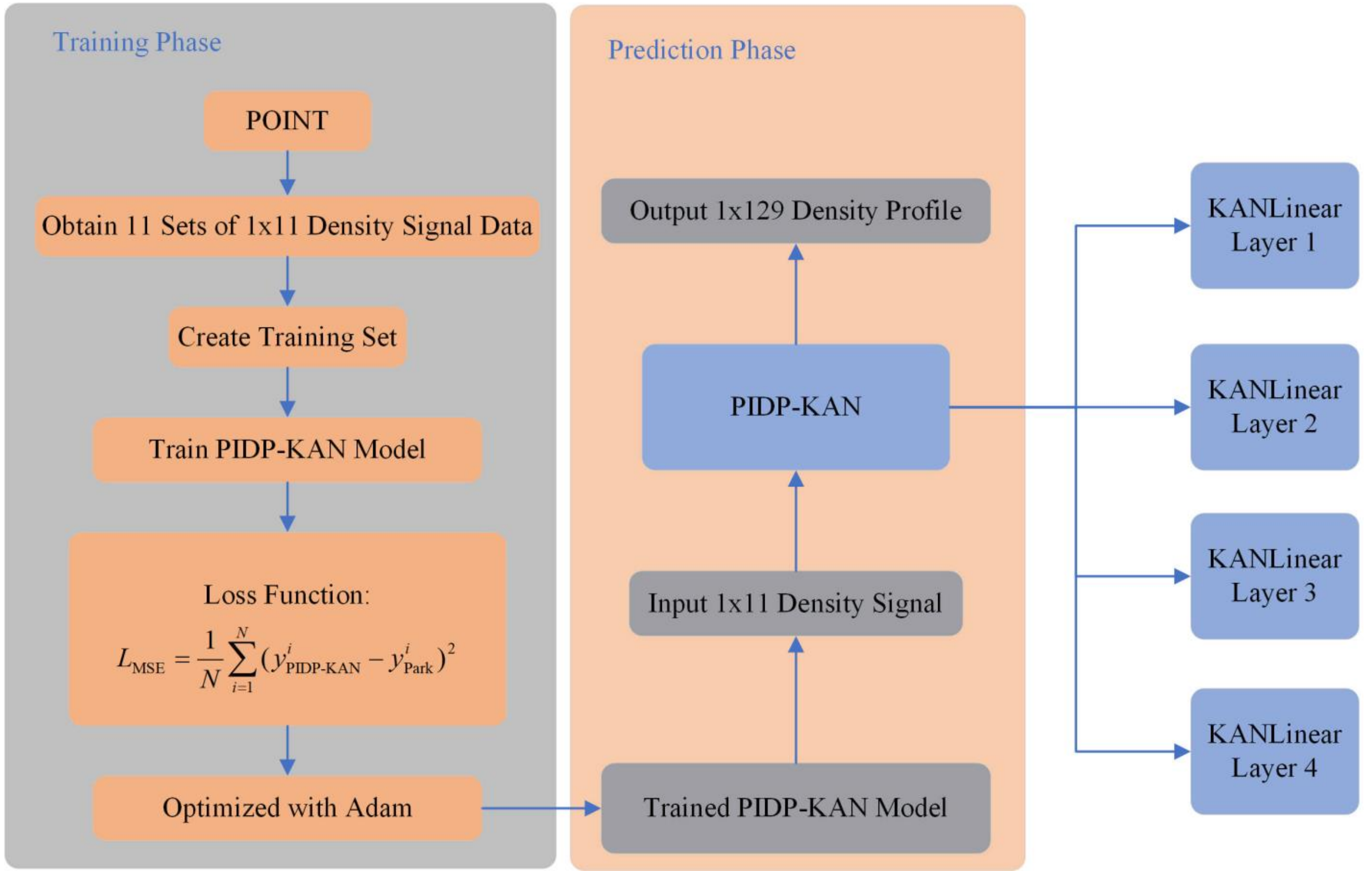
Density profile inversion using PIDP-KAN model

Kolmogorov-Arnold network (PIDP-KAN) model captures complex nonlinear relationships using spline functions.

PIDP-KAN structure:

- Consists of four KAN Linear layers parameterized by B-splines.
- Input: 11-channel POINT line integral density data.
- Output: Plasma density profile with 129 dimensions.

- By utilizing the adaptive mechanism based on spline functions, KAN method effectively fitted the complex nonlinear relationships.
- Compared with Park matrix method and MLP and BERT models, the proposed method demonstrates better performance in accuracy and speed.
- Calculation time is 2 milliseconds, meeting the feedback control requirements.



PIDP-KAN

H. B. Shen

Performance of PIDP-KAN model

Precision:

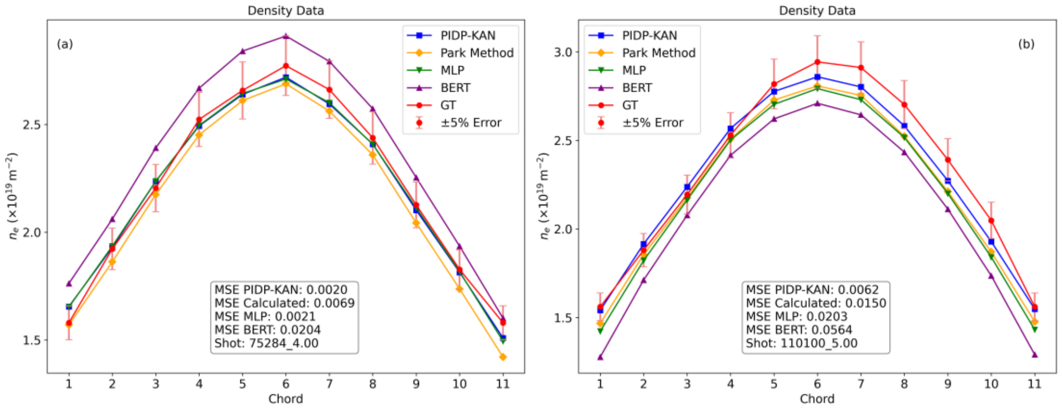
- MSE distribution of PIDP-KAN is highly concentrated, and outperforms the MLP and BERT models.
- All MSE are less than 0.035, indicating excellent global accuracy.

Robustness:

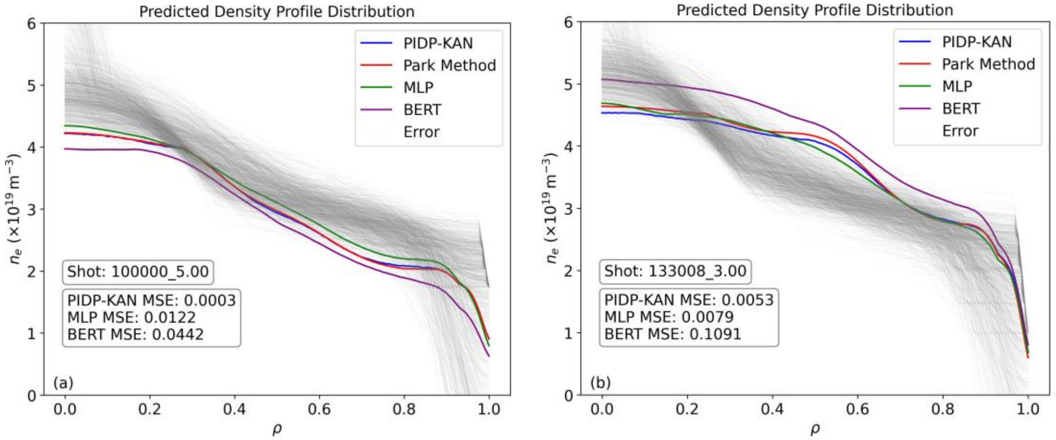
- Under 5% random noise, the prediction error still remains within 5%, demonstrating strong noise resistance.

Speed:

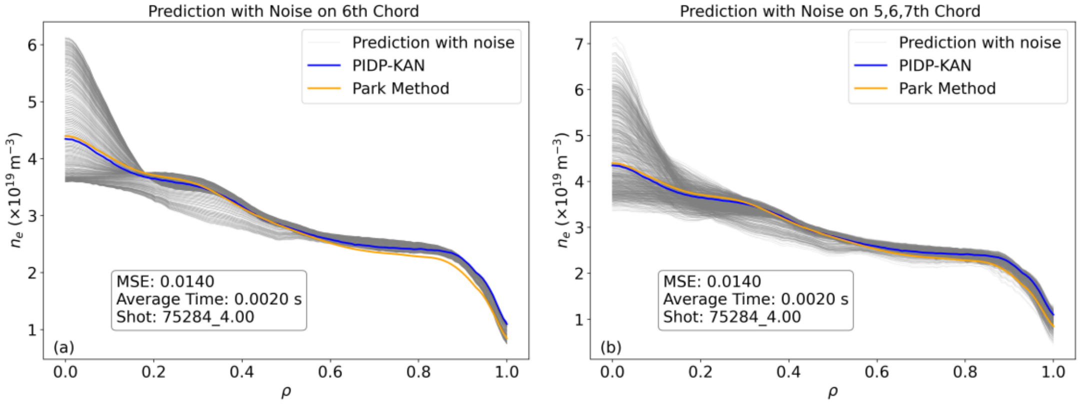
- Predict density profiles within 2 milliseconds, more efficient than traditional methods and has real-time computing capabilities.



Densities obtained by different methods, compared with original data of POINT.
(a) #75284@4s; (b) #110100@5s。



Density distributions obtained from different models.
(a) #100000@5s; (b) #133008@3s。



Density prediction profile with 5% random noise (#75284@4s)。

Spectrum analysis of Collective Thomson Scattering (CTS) – Improve accuracy

Fitness function of genetic algorithm

1. RMSE: $f = \exp\left(-\sqrt{\frac{(S_{mea}-S_{sim})^2}{N}}\right)$

2. Construct a new fitness function by the random weight method

Multi-Feature of Scattering Spectra Fitness Function (MFSS):

1) Dice Loss

$$f_1 = \exp\left(-\frac{1}{N}\sum_i \frac{2S_{mea}S_{sim}}{S_{mea}^2 + S_{sim}^2}\right)$$

2) Scattering spectrum slope

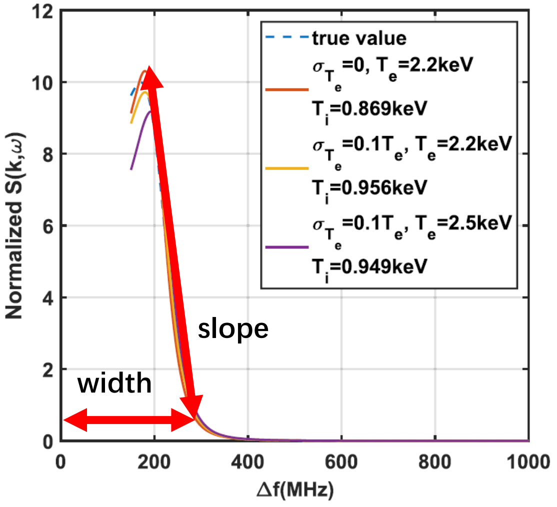
$$f_2 = \exp\left(-\left|\frac{\delta_{mea} - \delta_{sim}}{\delta_{mea}}\right|\right)$$

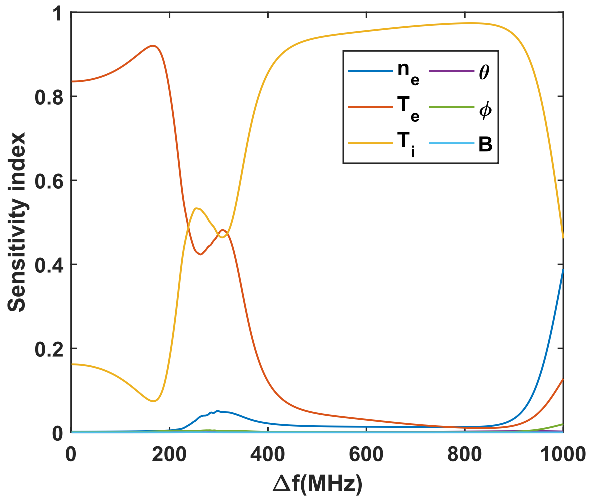
3) Scattering spectrum width

$$f_3 = \exp\left(-\left|\frac{\zeta_{mea} - \zeta_{sim}}{\zeta_{mea}}\right|\right)$$

$$f = \gamma_1 f_1 + \gamma_2 f_2 + \gamma_3 f_3$$

The weights of each individual's fitness function are randomly generated and satisfy $\sum_i \gamma_i = 1$





Fitness function	Maximum (keV)	Minimum (keV)	Average (keV)
RMSE	1.120	1.008	1.072
MFSS	1.090	0.995	1.036

Maximum error 12% → 9%

Average error 7.2% → 3.6%

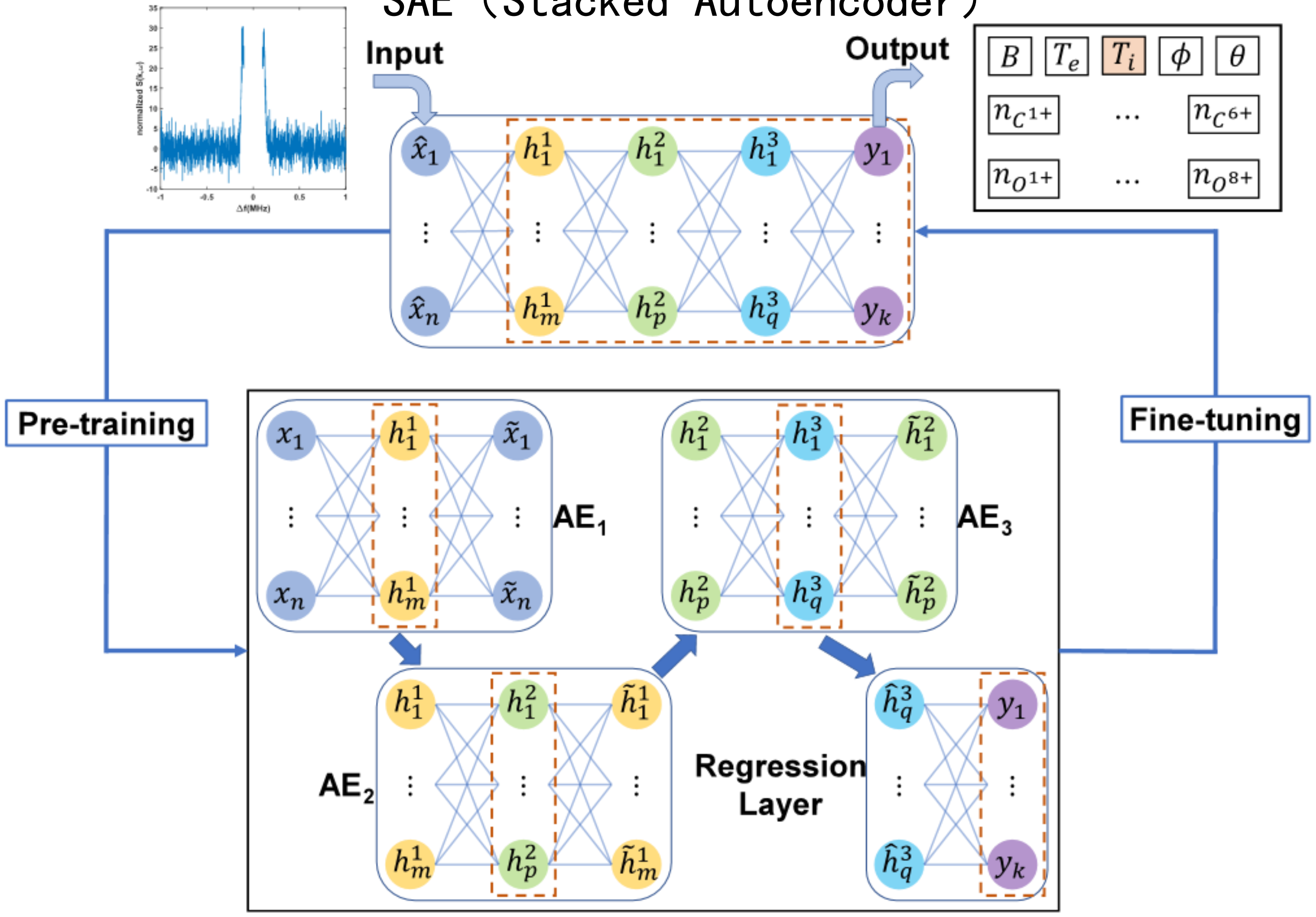
J. S. Zhang

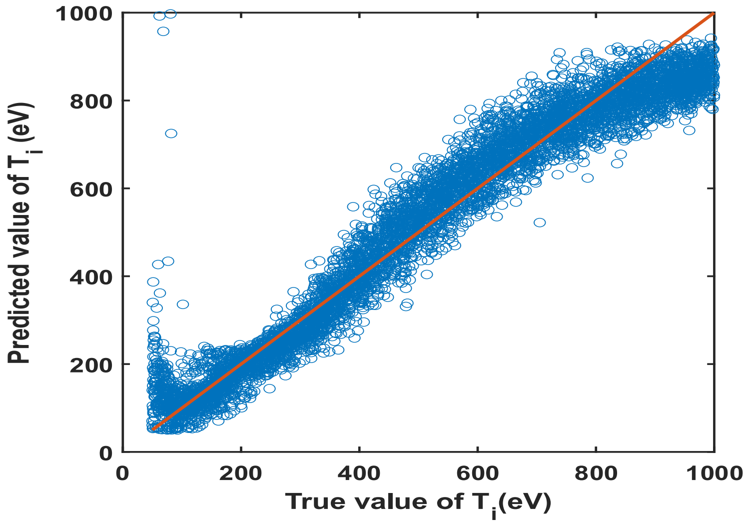
Zhang JS., et al., RSI, 2024, 95(9).

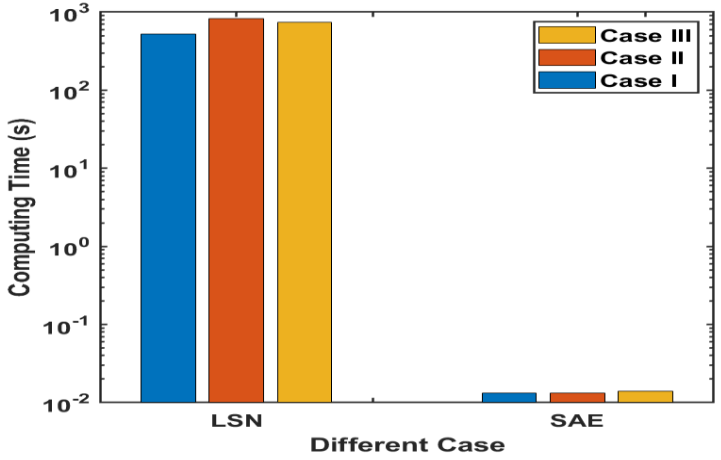
Spectrum analysis of CTS

- Real-time measurement of ion temperature

SAE（Stacked Autoencoder）







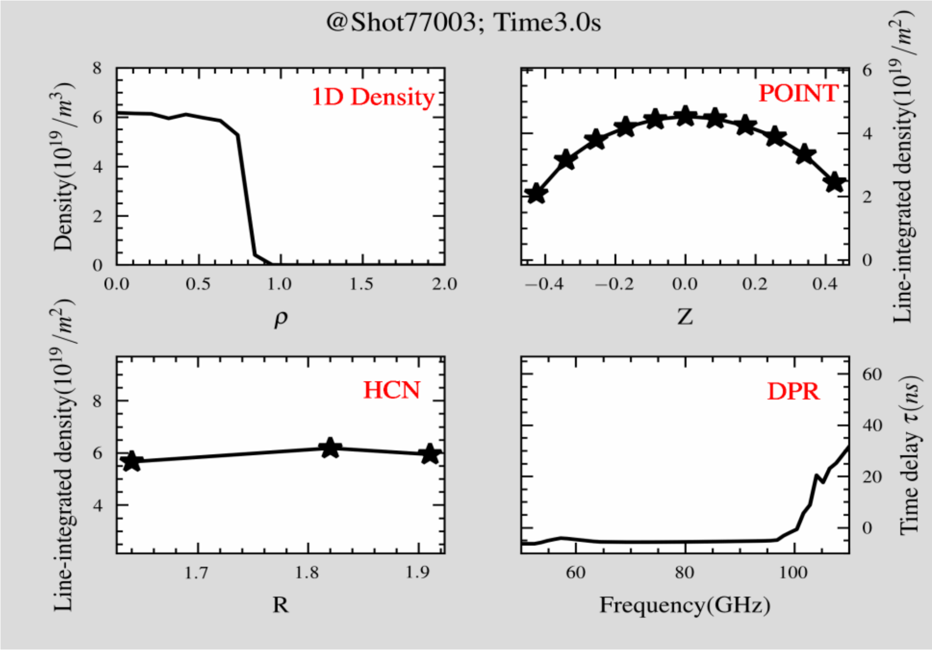
Bayesian data fusion – Enhance consistency

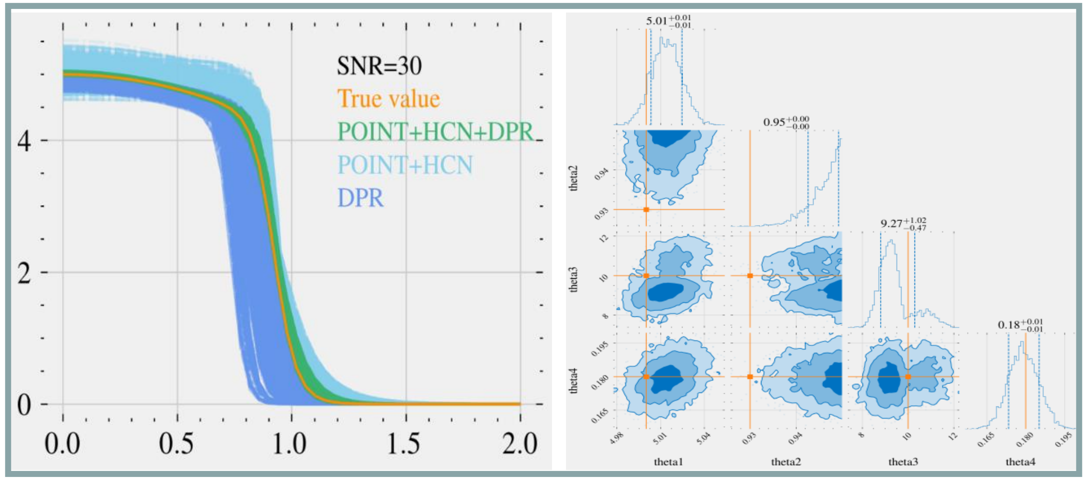
- Obtain distributions with higher accuracy and reliability than single diagnosis

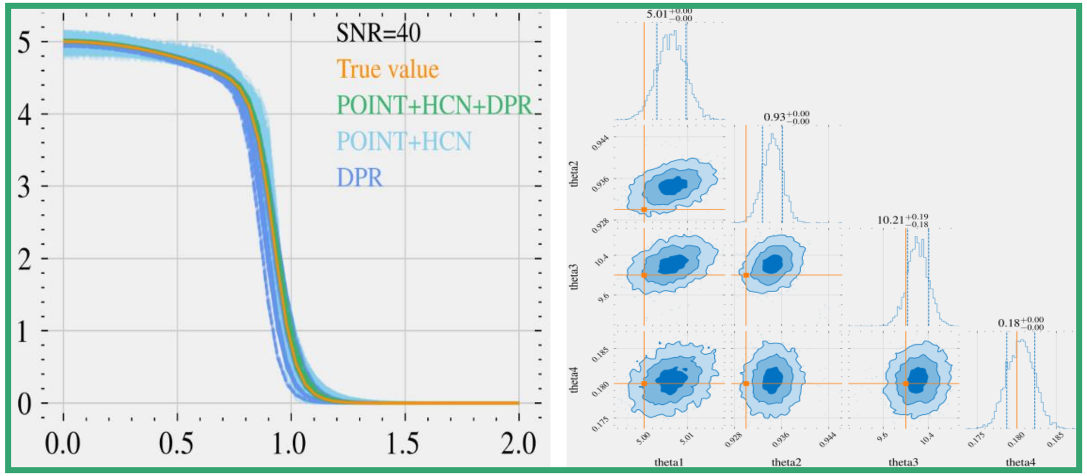
$$n_e(r_{eff}) = \theta_1 \cdot 10^{20} m^{-3} \cdot \left[\frac{1-\theta_4 \cdot (r_{eff}^2 / a^2)}{1+(r_{eff}^2/(\theta_2 \cdot a)^2)^{\theta_3}} \right]$$

Θ_1: maximum density　　Θ_3: steepness of boundary

Θ_2: gradient position　　Θ_4: degree of platform inclination

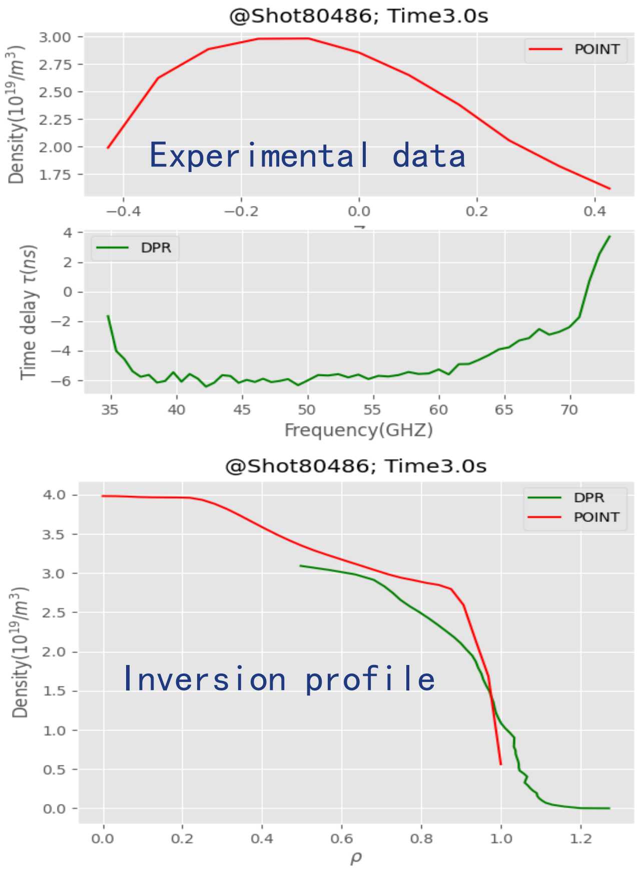


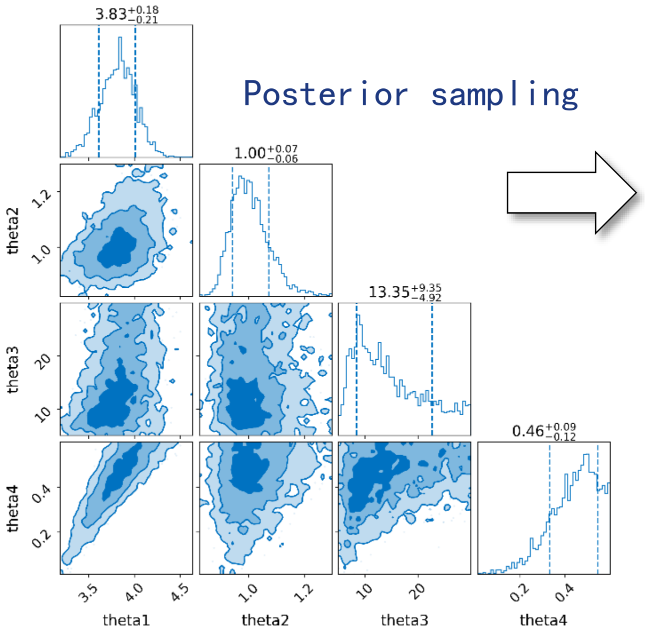


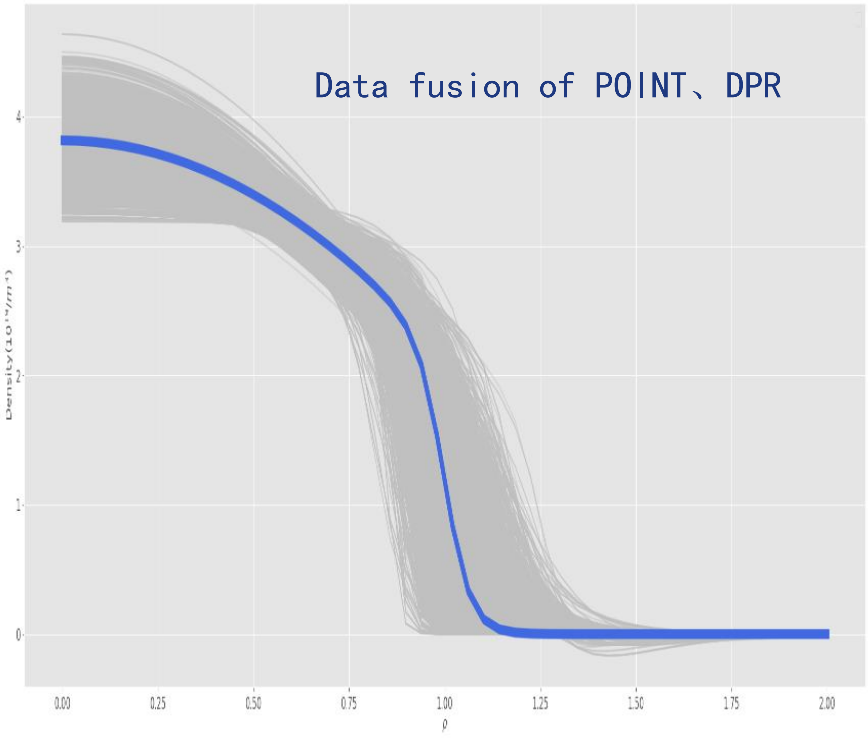


Bayesian-based data fusion of POINT and DPR data

- Quantify uncertainty and enhance the consistency between diagnostic systems

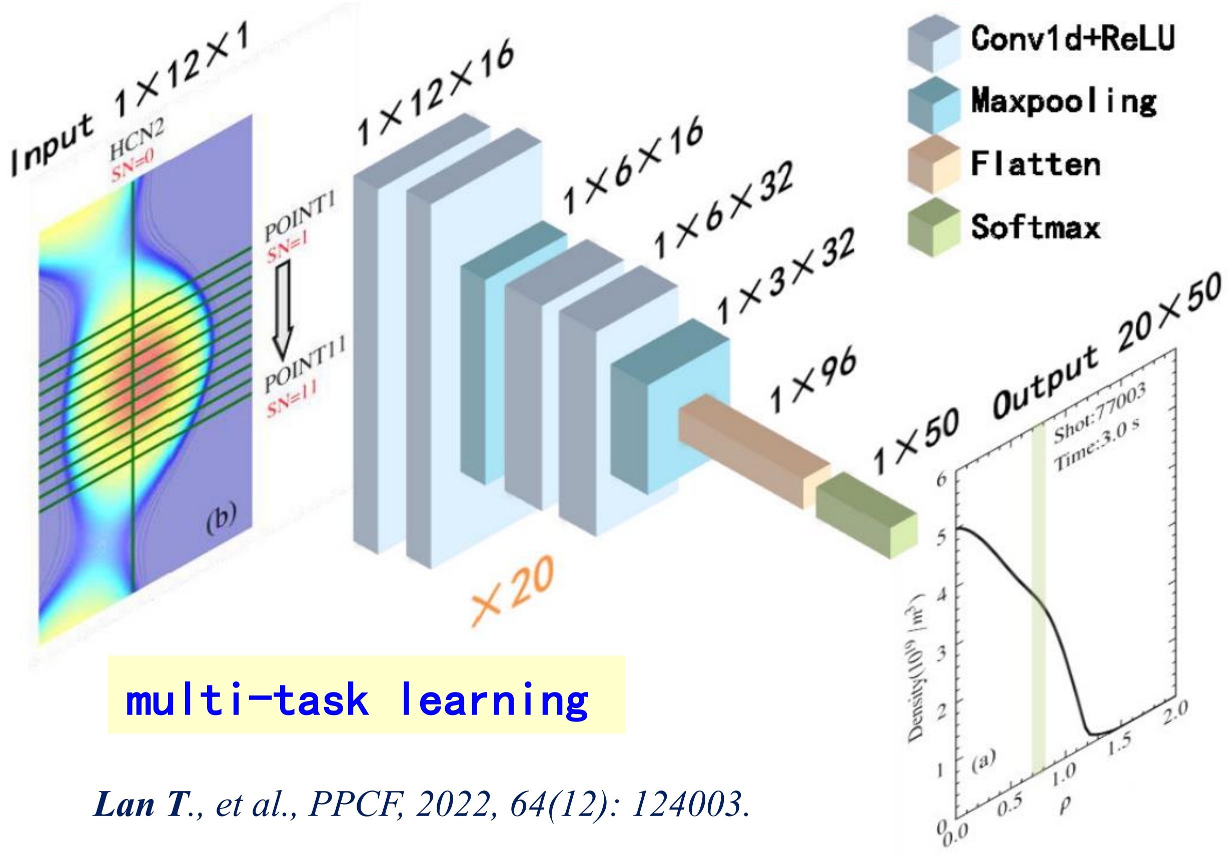






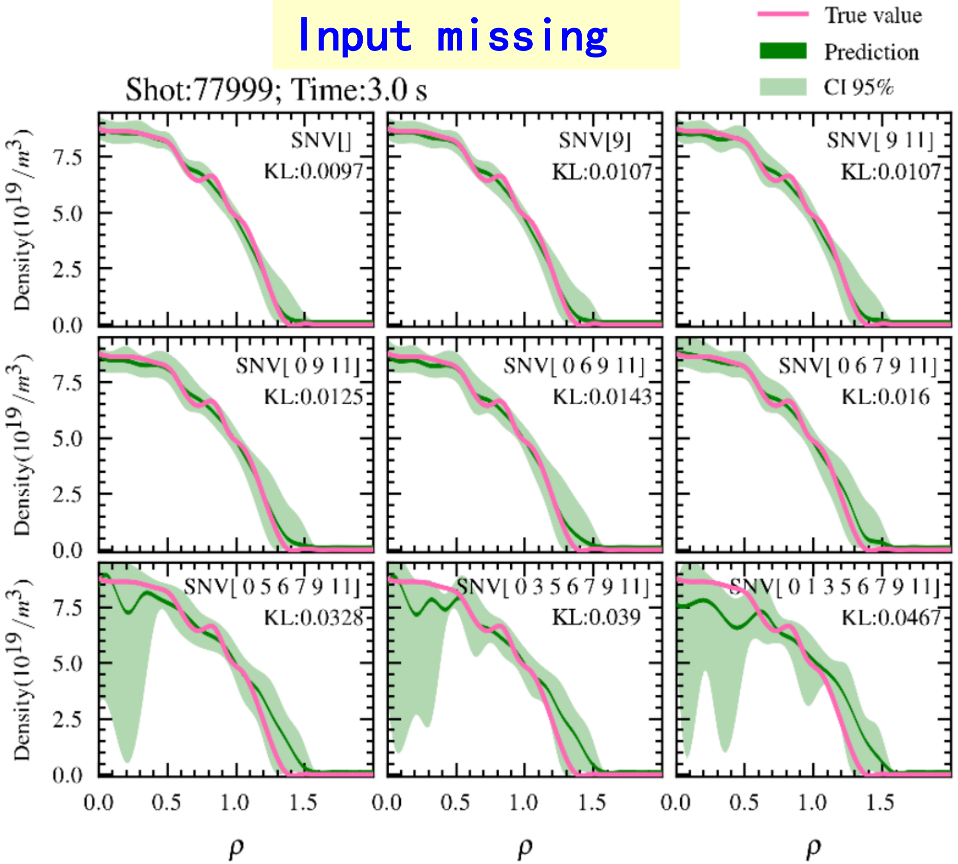
CNN-based data fusion of POINT and HCN data

- Not dependent on EFIT input, and robust against missing data.



multi-task learning

Lan T., et al., PPCF, 2022, 64(12): 124003.



ASIPP

Comparison of fusion methods

Bayesian inference

Advantage:

- Describes quantities and uncertainties in a probabilistic manner
- Provide an intuitive data fusion framework

Disadvantage:

- The priors significantly affect the posterior results
- Sampling process may take a relatively long time

VS

Deep learning

Advantage :

- Reduce method uncertainty brought by fusion model
- Fast, suitable for online data processing

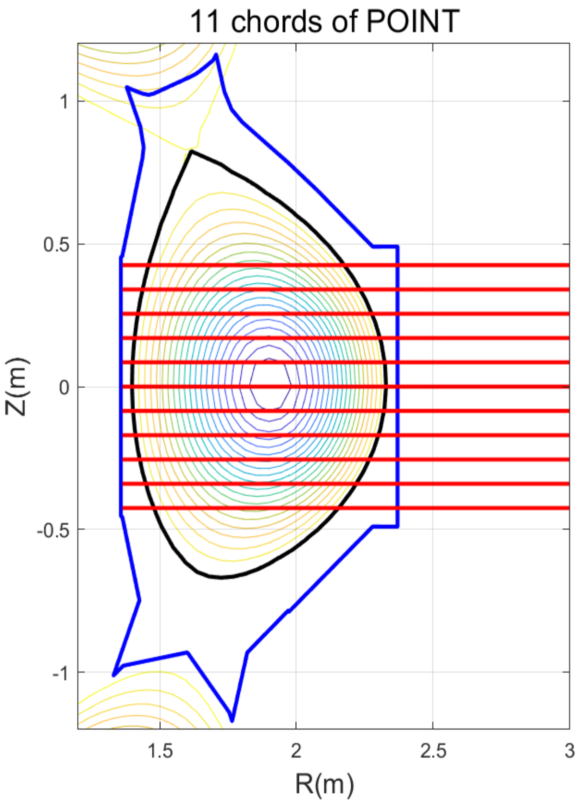
Disadvantage:

- Model training is time consuming
- Require fusion strategy design for data with varying principles, attributes, and features

VVUQ analysis – Quantify uncertainty

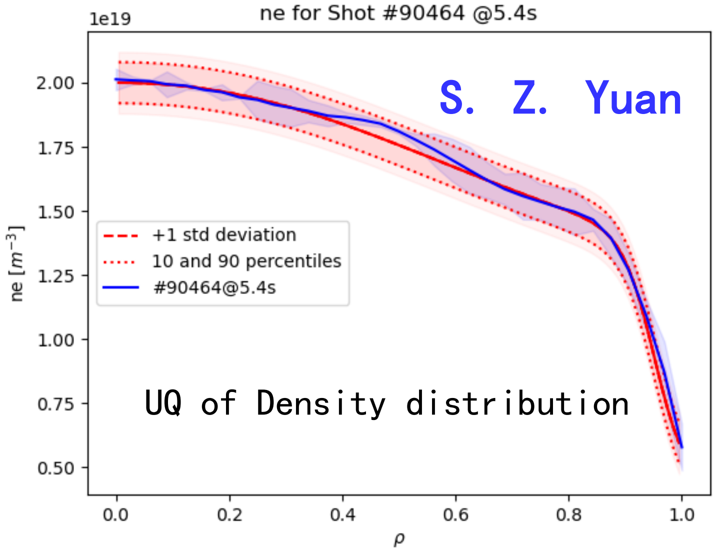
- VVUQ goal: quantify uncertainty, minimize errors, optimize diagnostic system and data processing

VVUQ of POINT



VVUQ analysis is performed by:

- **Verifying code quality and results.**
- **Validating the model against experimental data.**
- **Quantifying uncertainties from design, environment, and analysis.**



confidence coefficient: 99.5% (JS divergence calculation).

similarity: ≥97.5% (Core), 82.5% (boundary).

Confirmation process uncertainty: 3.12%, from simulation.

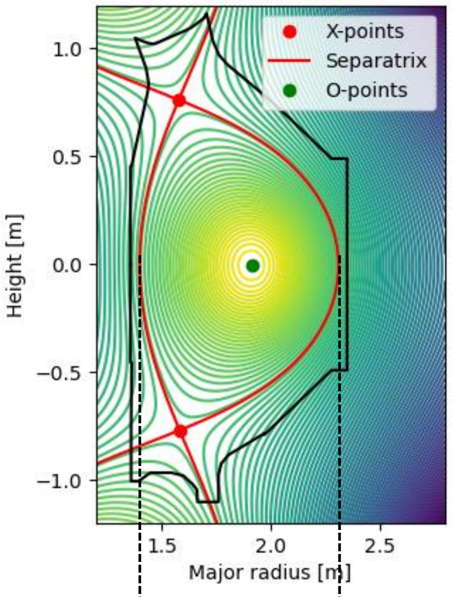
Uncertainty of POINT: 6.17%;

2.99%, from data processing,
3.18%, from engineering design.

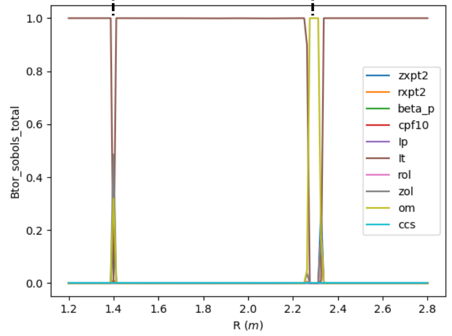
Uncertainty of POINT:

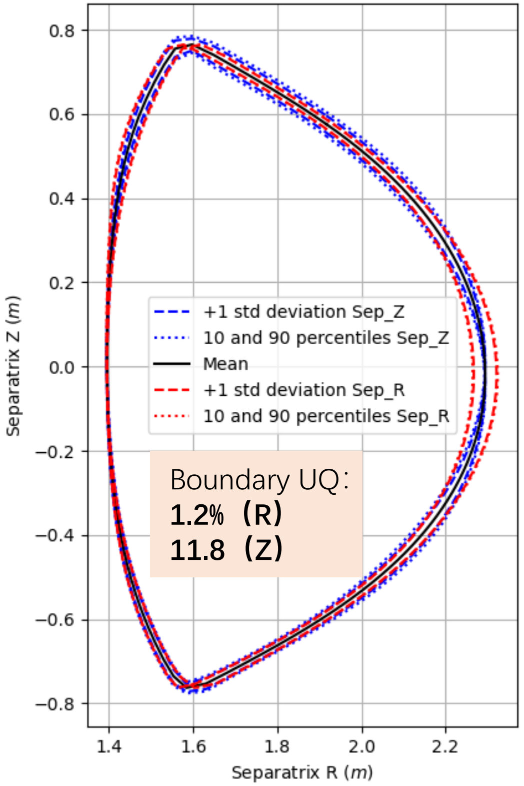
Source	Mechanical vibration	frequency variation	refraction effect	finite temperature effect	spline interpolation
uncertainty	1.0%	0.28%	0.4%	1.5%	2.99%

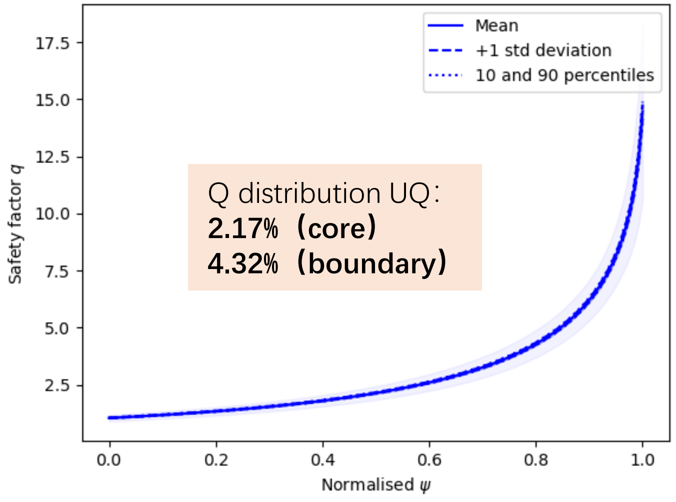
VVUQ of magnetic equilibrium

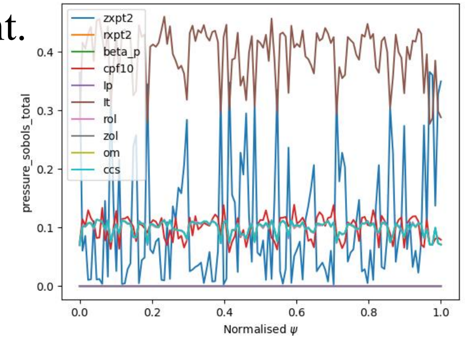


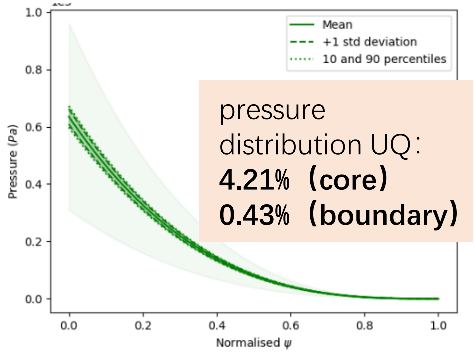
1. The influence of the internal plane position on the balance reconstruction accounts for 65%.
2. Among the poloidal field coils, the 10th coil (1.13679, -1.94092) has the greatest impact on q profile.
3. The influence ranking on q profile: toroidal current, plasma current (core part); R coordinate of X point, inner/outer mid-plane position (boundary)...
4. The influence ranking on plasma shape (boundary line position): inner/outer mid-plane position, R and X coordinates of striking point...
5. The uncertainty of toroidal field mainly occurs at plasma boundary, and influenced by positions of inner and outer mid-surfaces as well as the magnitude of toroidal current.
6. The distribution of central plane pressure is significantly influenced by toroidal current and Z-coordinate of X point.

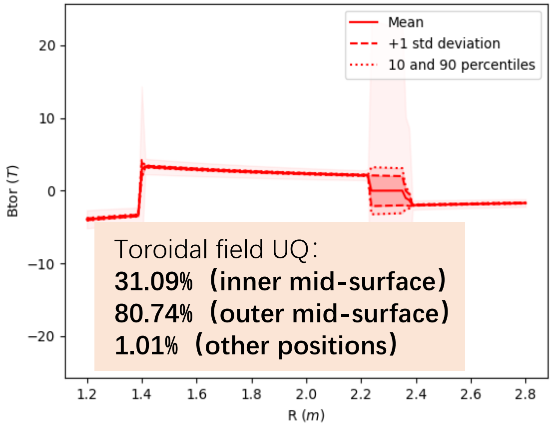




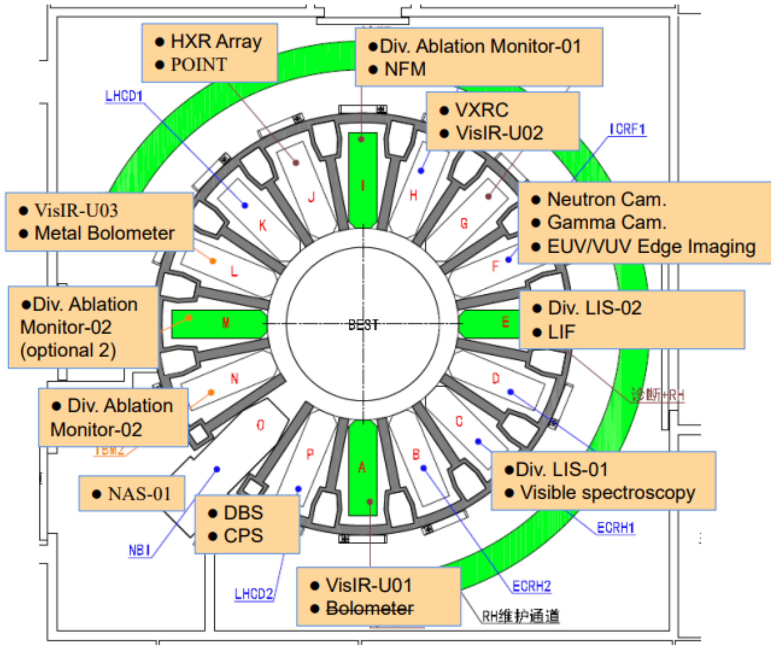








Intelligent processing of multimodal diagnostic data on EAST tokamak

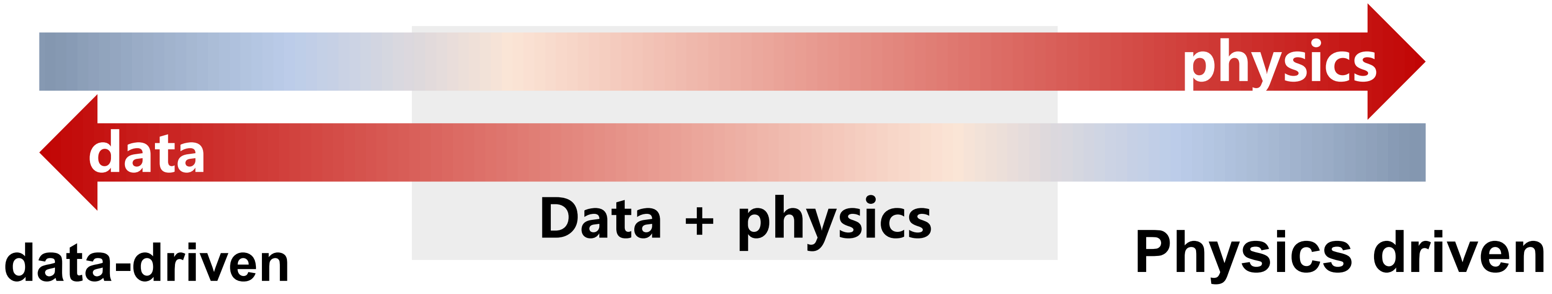


Types of fusion data

- Safety and device protection
- Plasma operation and control
- Advanced plasma physics research

Core requirements

- Accurate, fast, efficient
- Reliability, Availability
- Robustness
- self-consistency
- Interpretability
- Error propagation analysis



- Combine physics with AI can enhance the generalization ability of AI models and improve the understanding of physical laws.

ASIPP

Thank you for your attention!