Improvements in efficiency of gyrokinetic simulation runs with convolutional neural network models analyzing nonlinear saturation processes

E. Narita[1], M. Honda[2], S. Maeyama[3] and T.-H. Watanabe[3]

[1]National Institutes for Quantum Science and Technology (QST)

[2]Graduate School of Engineering, Kyoto University

[3]Department of Physics, Nagoya University

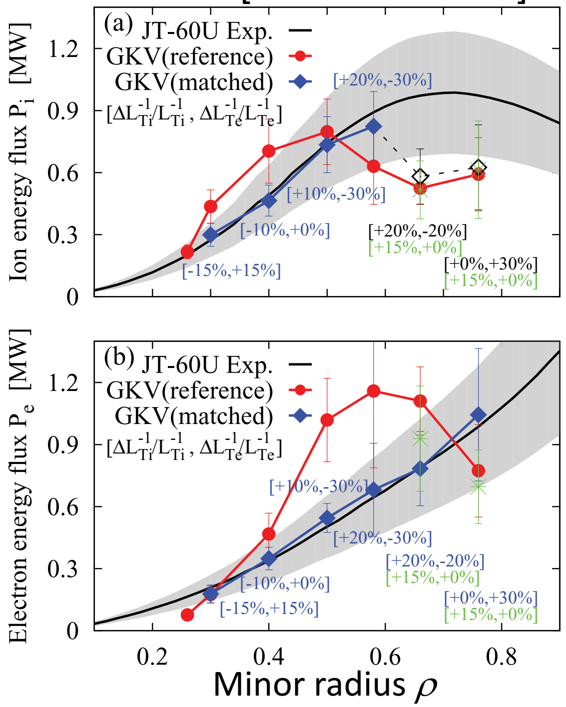
This work was supported MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Exploration of burning plasma confinement physics, JPMXP1020200103) and used computational resources of ITO provided by Kyushu University (Project IDs: hp200127 and hp210178). This work was also carried out using the JFRS-1 supercomputer system at Computational Simulation Centre of International Fusion Energy Research Centre (IFERC-CSC) in Rokkasho Fusion Institute of QST (Aomori, Japan).

Turbulent transport simulations with gyrokinetic codes

□ First-principle based gyrokinetic codes

- The time evolution of the perturbed distribution function is solved in the 5D phase-space.
  - → Predictions of turbulent fluxes
  - → Investigations into underlying turbulence physics
- Huge computational resources are required.
- An enormous amount of calculation data is generated.

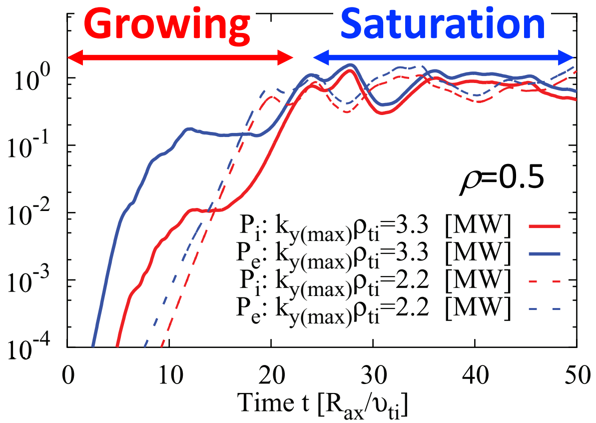
[Nakata NF2016]



Based on the data collected,

- ✓ evolving processes can be visualized as images and
- ✓ these images must contain much information on turbulence evolution.

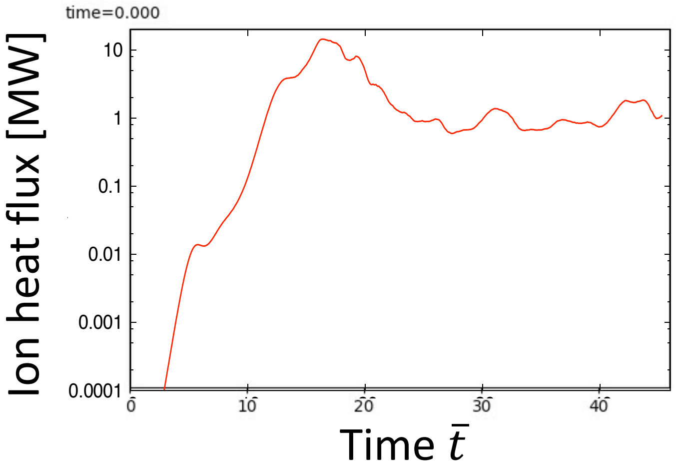
  ➔ A new tool to reduce the computational cost



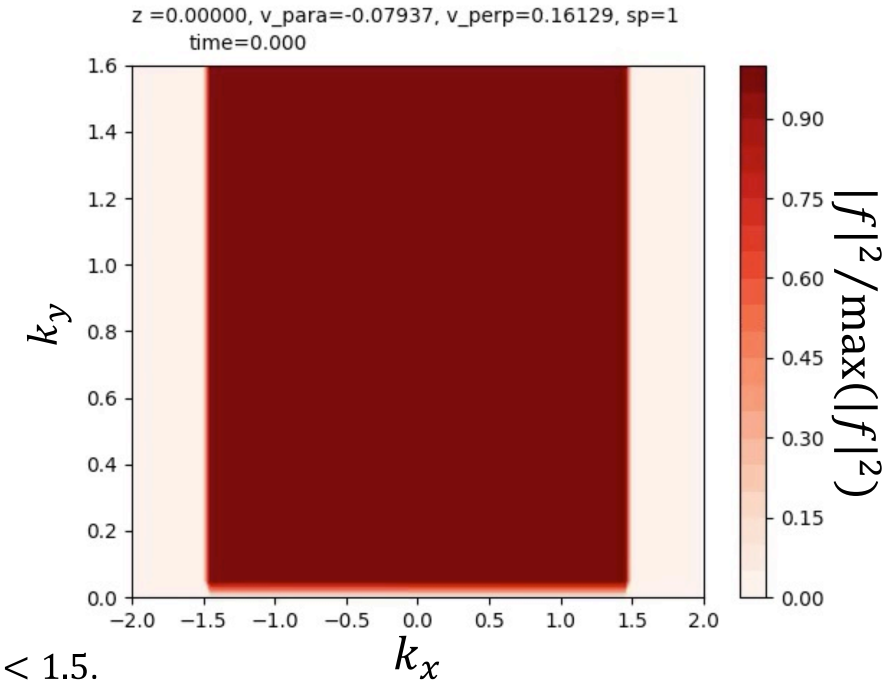
Patterns of the distribution function in the wavenumber space

- A nonlinear calculation with GKV for JT-60U plasma parameters
  - JT-60U #45072@ρ=0.76: ITG/TEM
  - 4,608 cores x 60 hours @ITO

Focusing on plots of the square of the perturbed distribution function, $|\tilde{f}|^2$, in the wavenumber space.



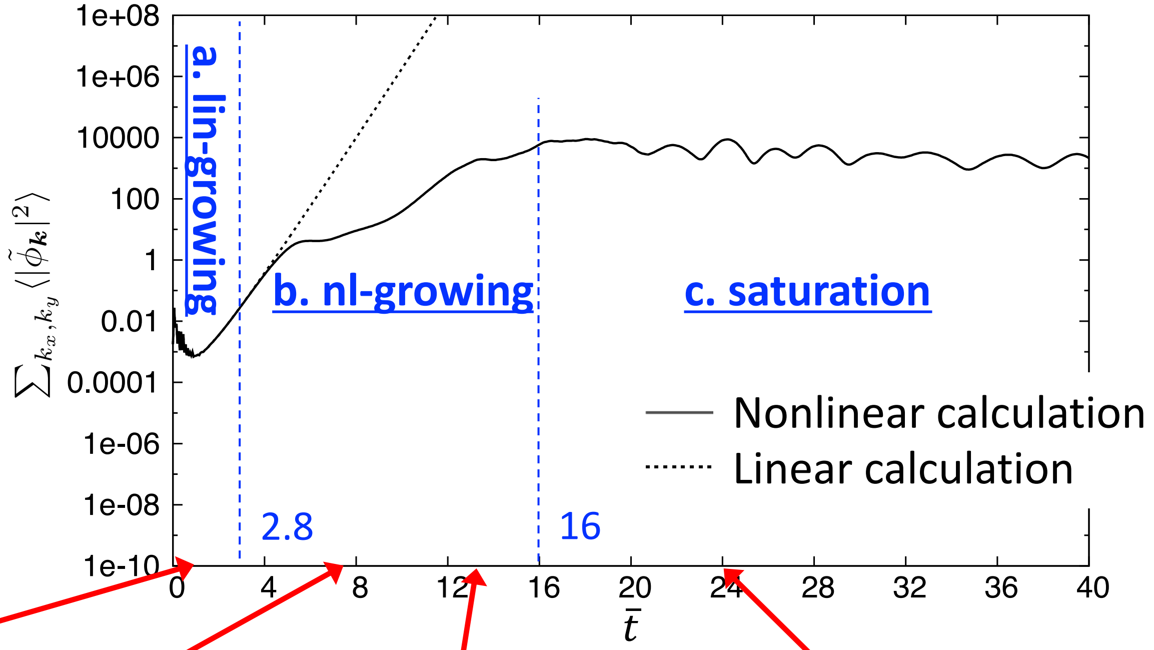
- ✓ $\bar{t} < 4$:
  High $|f|^2$ areas are found in $k_y > 1$
- ✓ $7 < \bar{t} < 9.5$:
  The fluctuation gets more intense at $k_y \sim 0.5$.
- ✓ $\bar{t} > 9.5$:
  The fluctuation spreads over $0 < k_y < 0.5$, and the transition to the saturation phase begins.



Initial fluctuations are given in $1.5 < k_x < 1.5$.

$\left|\tilde{f}\right|^2(k_x, k_y)$ images differ phase by phase

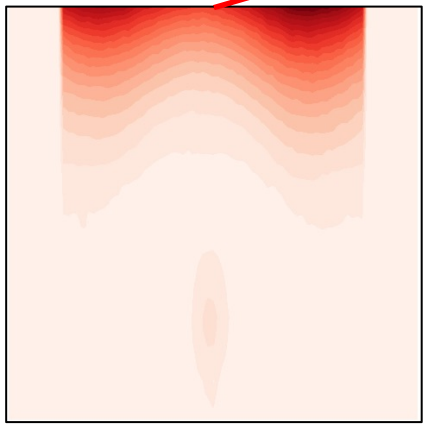
- Classification into 3 phases
  - a. linear-growing
  - b. nonlinear-growing
  - c. saturation



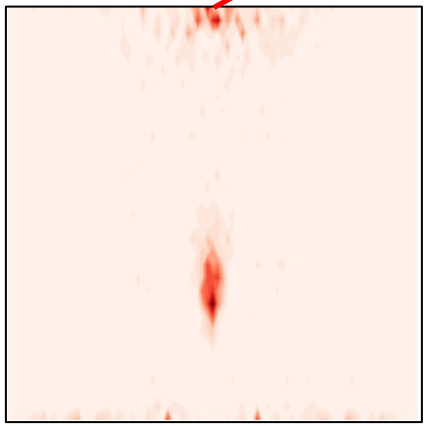
Characteristic patters observed in each phase

- Can a deep learning program detect the difference?
- Can a deep learning program predict the simulation time from the image?

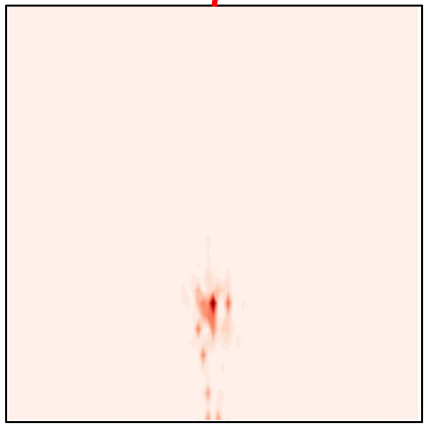
$\bar{t}$ =1.4



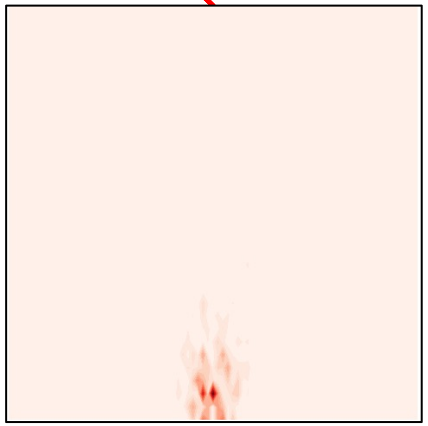
$\bar{t}$ =7.5



$\bar{t}$ =13



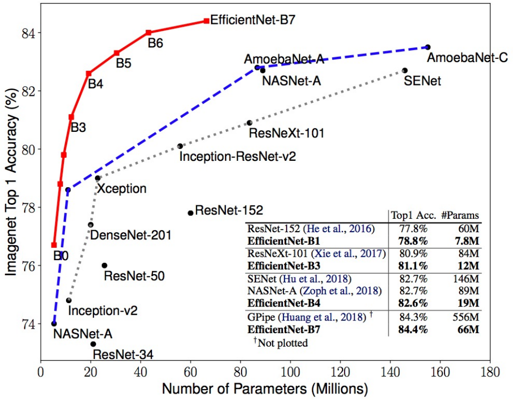
$\bar{t}$ =24



Convolutional Neural Network model: EfficientNet

EfficientNet (ENet) [Tan ICML19]

- A state-of-the-art convolutional neural network (CNN) model in 2019
- Pre-trained with ImageNet datasets
- Variants with different network depth: We use **EfficientNet-B4**.
- Very high transfer learning performance (important!) & Fine tuning



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

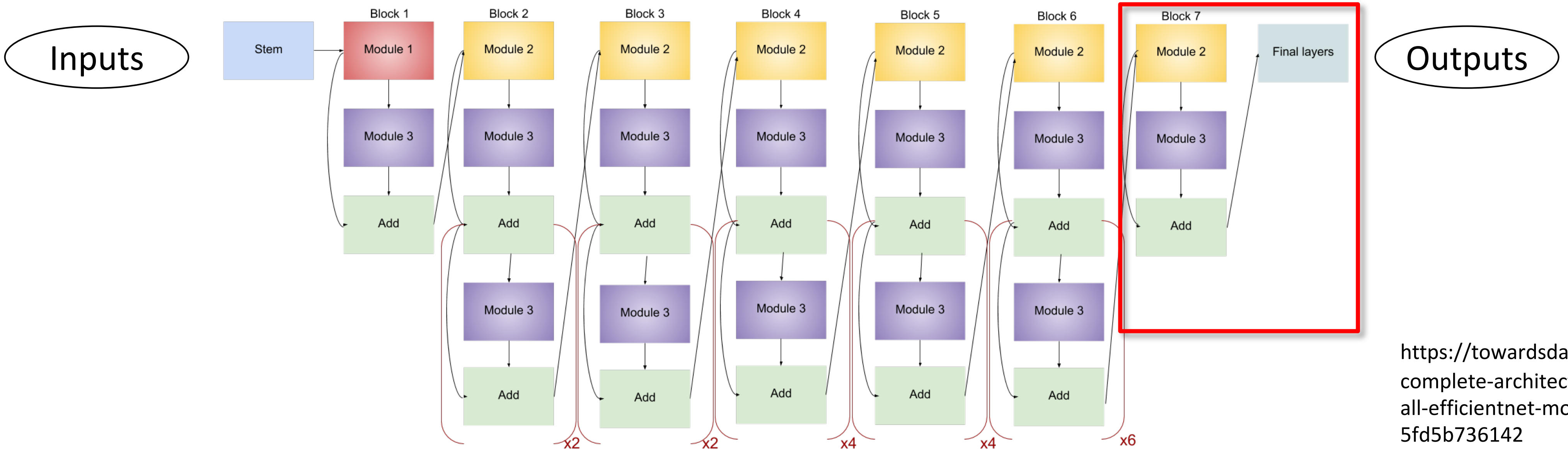
†Not plotted

https://arxiv.org/abs/1905.11946.

Feature learning

Fine tuning for specific problems

Classification / Regression



https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142

Freezing pre-trained weights

Modified for the $|\tilde{f}|^2$ patterns: unfrozen

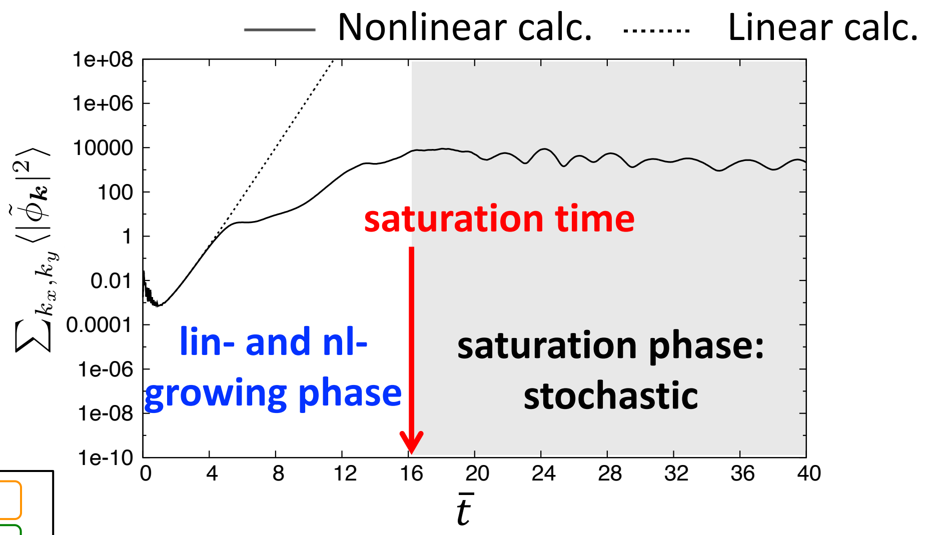
The simulation time is predictable from the image

Train ENet for regression (prediction)

- Feed images and their corresponding time in the linearly and nonlinearly growing phases into ENet for training
- In the saturation phase, the link b/w the image and the time is lost due to its stochastic nature.
- Number of data
  - Train: 5,403
  - Validation: 1,543
  - Test: 772

Prediction	5.005	10.734	2.755
True	5.100	11.000	3.100

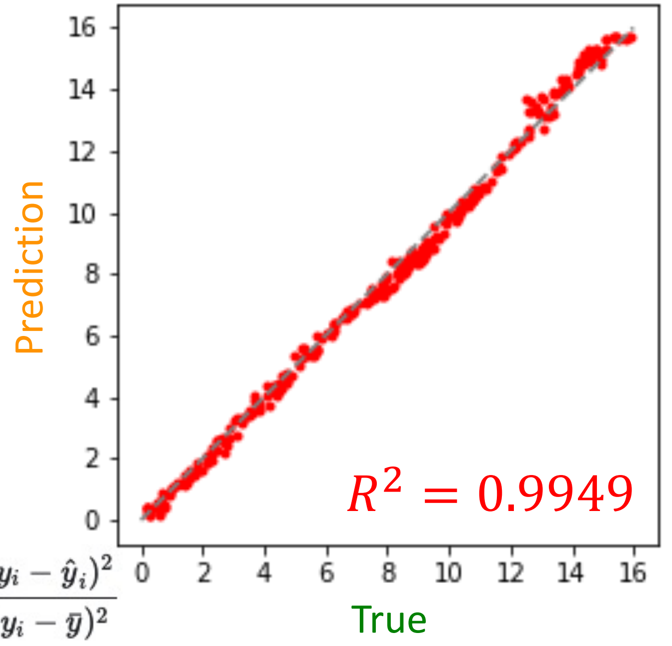
High predictability!



Extremely high determination coefficient R^2: **0.9949**

➔ Excellent prediction evenly over the entire time period

➔ Feasible to get ENet forecast the saturation time



$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

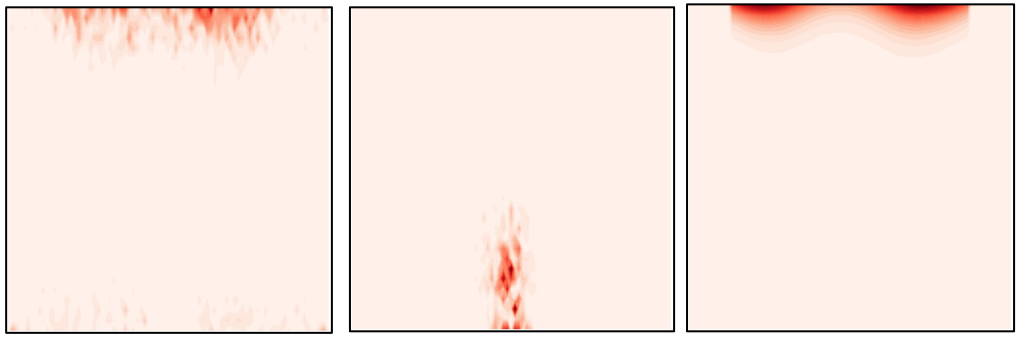
ENet as a predictor for efficient runs

Predictive capability makes it possible to choose the fastest case of all.

- Make use of ENet trained with “Base” case data (black line)
- Execute several GK runs with different initial amplitudes for a while and pick up the seemingly fastest case

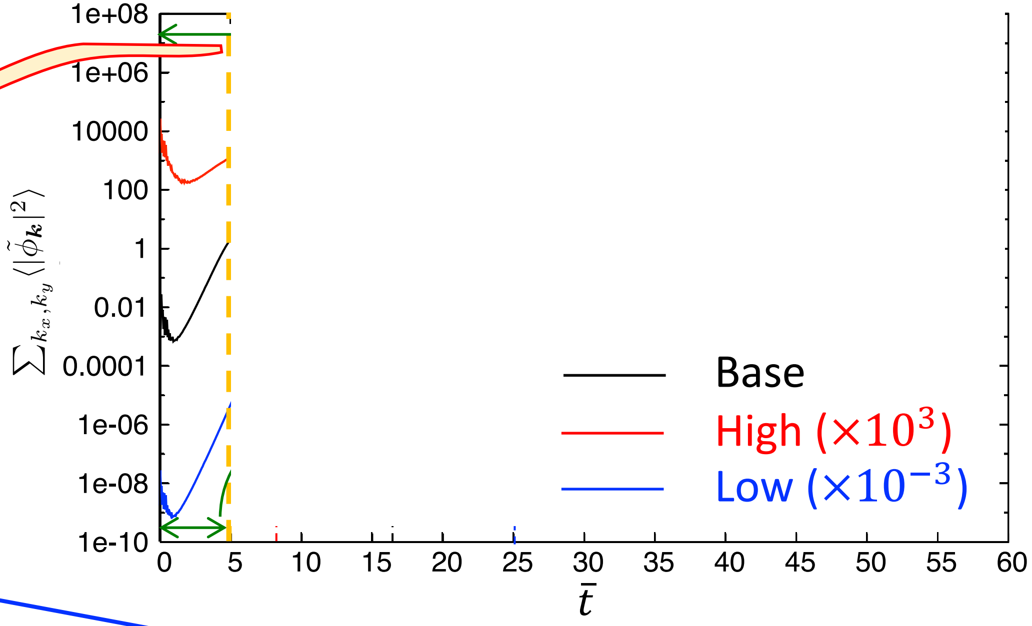
Images at $\bar{t}$ = 5

Base | High | Low



Estimated saturation time | $\bar{t}$ **= 6.726** | $\bar{t}$ **= 17.437**

Saturation time



It will be saturated soon

The simulation is still in the midst of the growing phase.

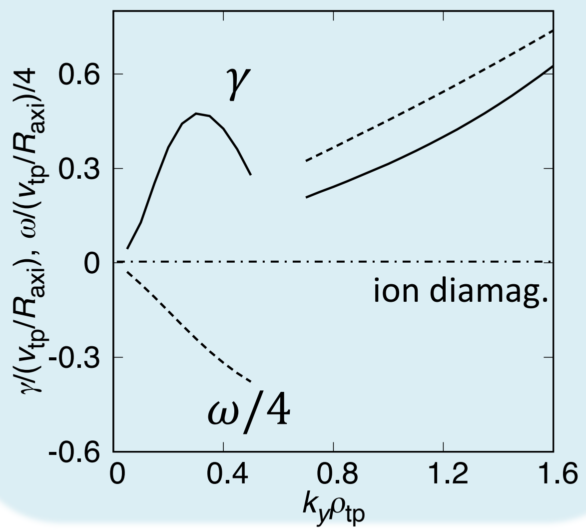
- ✓ The saturation time can be roughly forecasted at an early stage.
- ✓ Save numerical resources by keeping the fastest case and eliminating the rest.

Generalization of ENet-based predictor for different dominant instabilities

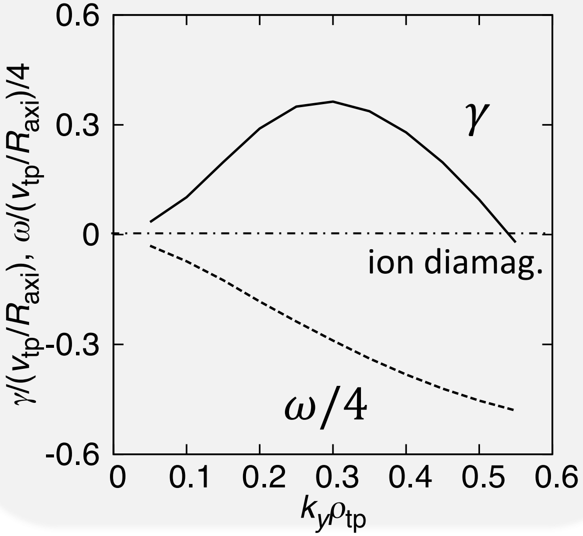
- Three cases prepared for different dominant instabilities based on the Cyclone base case (CBC), which is a *de facto* standard DIII-D parameter set for gyrokinetic simulation benchmarking tests.

	R/L_{T_i}	R/L_{T_e}	R/L_n	$T_\mathrm{e}/T_\mathrm{i}$
CBC	6.92	6.92	2.22	1
Pure ITG	6.92	0	2.22	1
Pure TEM	1	8	3	3

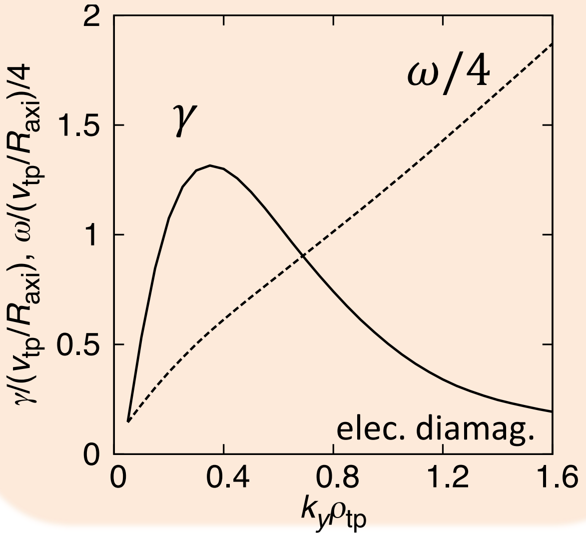
ENet for ITG/TEM (CBC original)



ENet for ITG



ENet for TEM

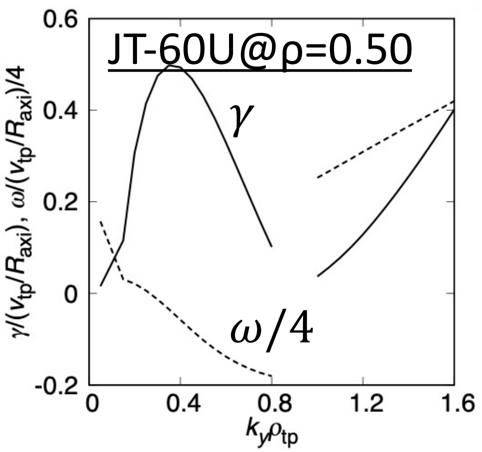


Linear calculations can be performed at low numerical cost.

➔ Choose the best model to better predict the time based on the linear results performed at initial.

Methodology for actual application

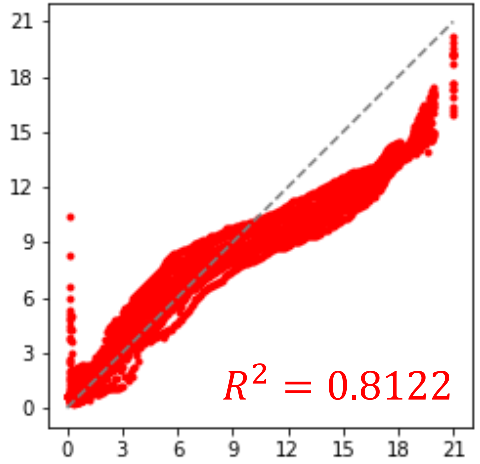
Linear calculation → **Choose the suitable model** → **Predict the time**

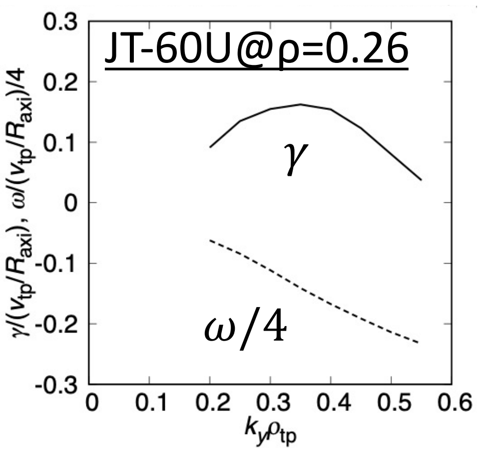


ITG/TEM

ITG

TEM

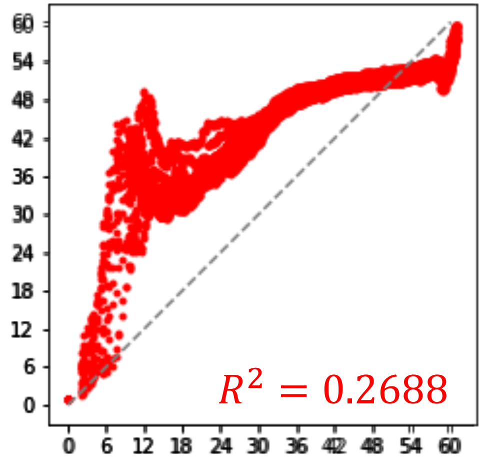




ITG/TEM

ITG

TEM



Pre-assessing the linear stability leads to higher predictability.

Conclusions and future work

- The powerful CNN model, **EfficientNet**, is able to distinguish the minuscule difference between images of fluctuations in the wavenumber space.
- ENet has an ability to select the simulation that finishes the fastest for obtaining the result as fast as possible and saving computational resources.
- By preparing multiple ENet models with different dominant instabilities, high prediction performance can be achieved for untrained cases.

The ENet models are helpful to study turbulent transport with gyrokinetic simulations effectively.

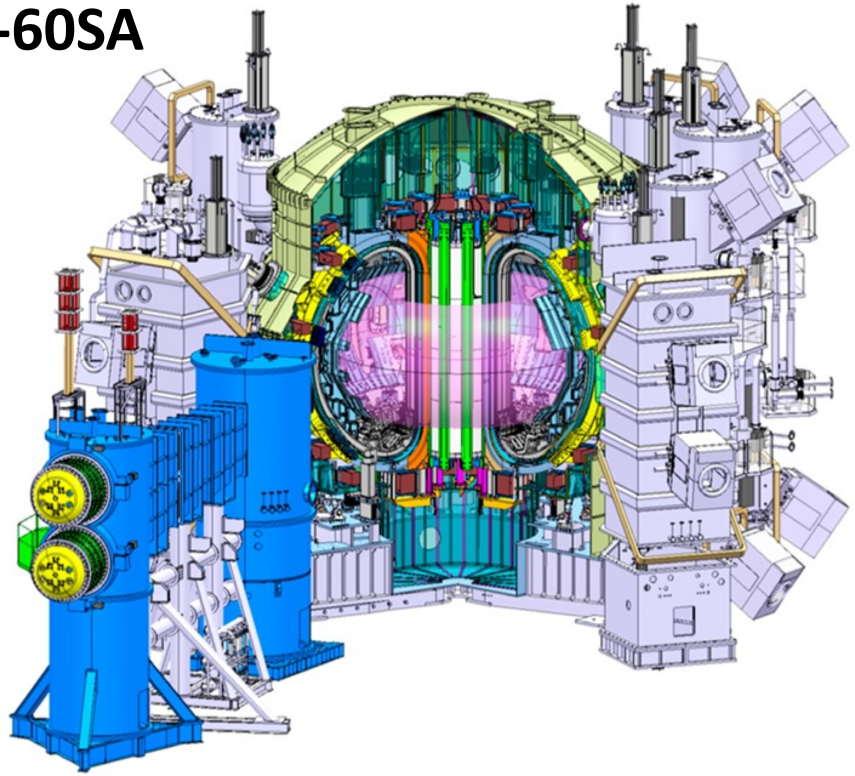
Future plan

- Multimodality for higher accuracy
- Predicting the turbulent saturation levels

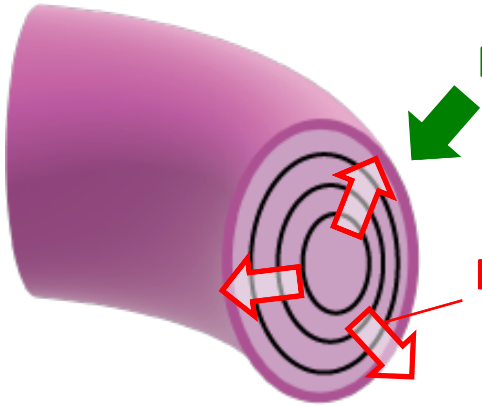
ADDITIONAL MATERIALS

Turbulence regulates plasma confinement

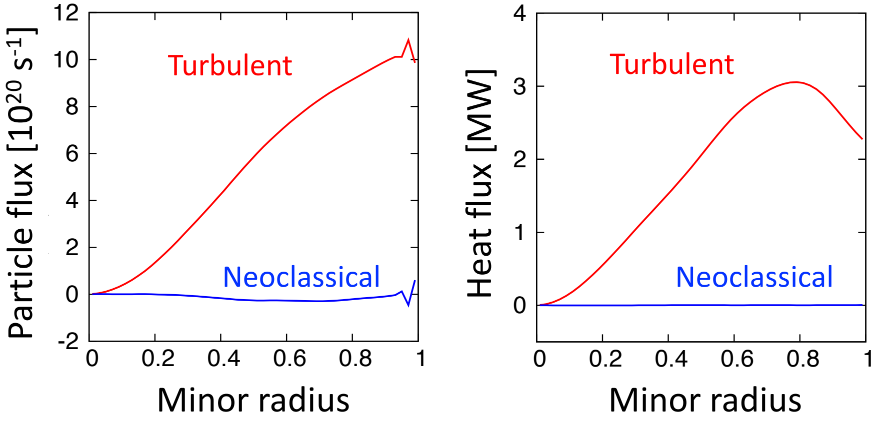
JT-60SA



- The balance between **transport** and **sources** determines density and temperature profiles.



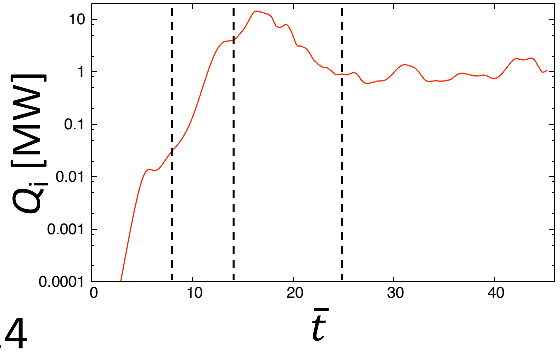
- Turbulence is dominant in tokamak plasmas.
- Massive computational cost is required to estimate turbulent fluxes.



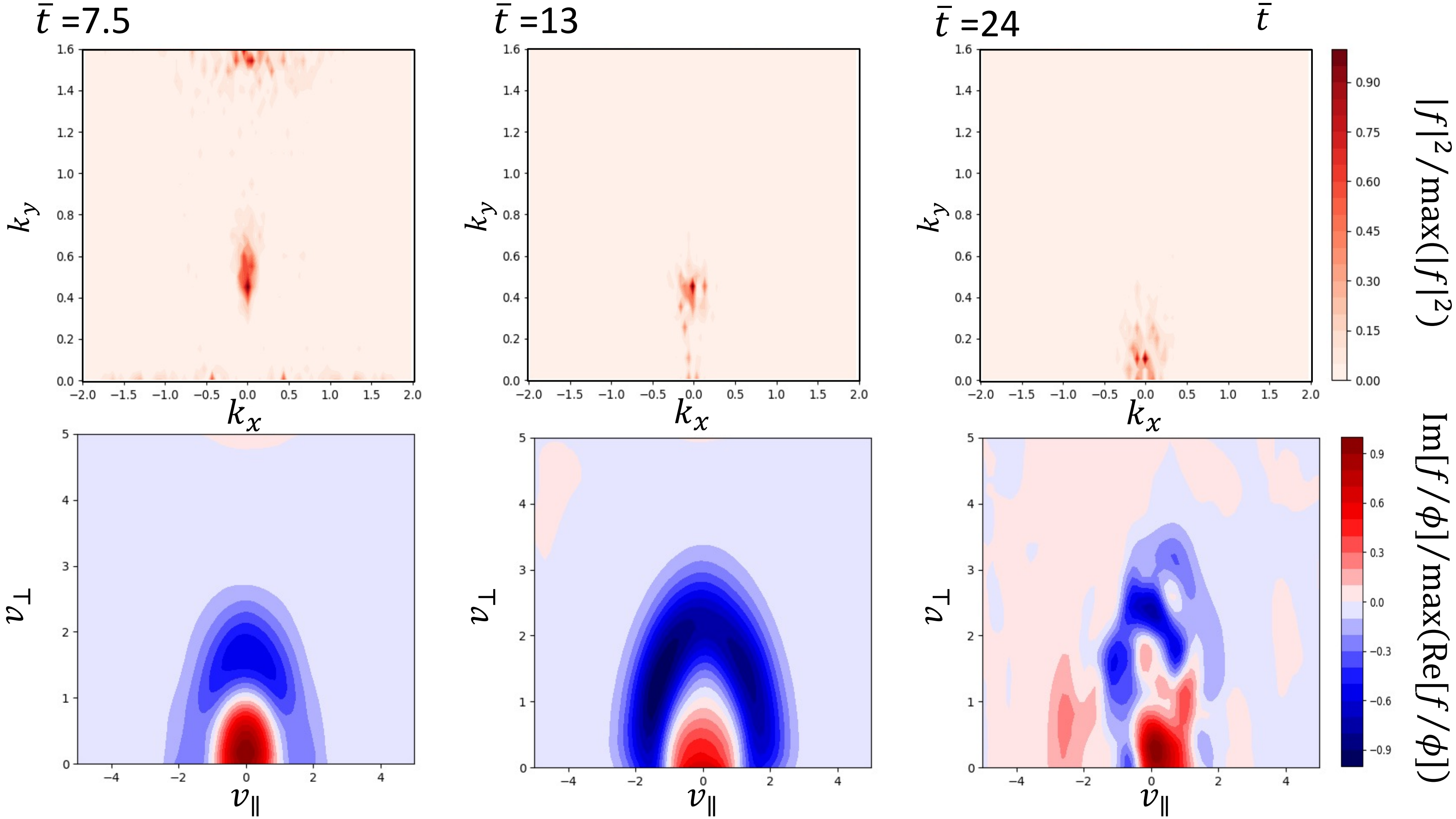
✓ **Predicting and understanding turbulent transport quickly and efficiently are crucial issues.**

➔ **Taking data-driven approaches, we have developed neural-network (NN) based models.**

High $|f|^2$ areas gradually move toward the low wavenumber region.



- ✓ The patterns in the velocity space have also been checked.
  - ➔ The structure changes in the growing phase, and it breaks in the saturation phase.

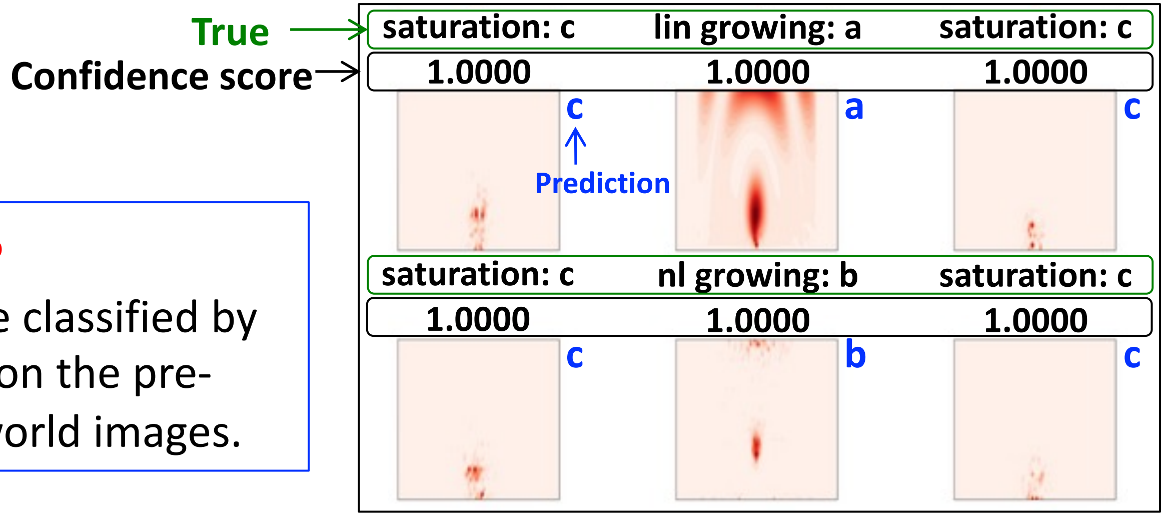


Classification of the $|f|^2$ patterns by the CNN model

- Number of data
  - Train: 5,403
  - Validation: 1,543
  - Test: 772

✓ Accuracy for test data: 99.9%

➔ The $|f|^2$ patterns can be classified by transfer learning based on the pre-trained CNN with real-world images.



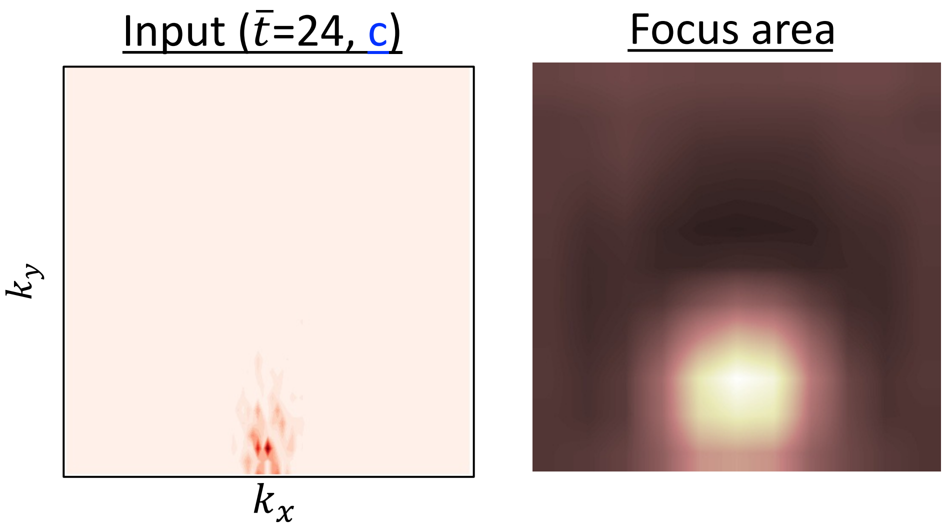
➢ Visualization of focus areas

In the saturation phase,

- high $|f|^2$ areas are located in the low k_y region.

✓ The model is focusing on the low k_y region.

➔ A potential tool to understand turbulence physics

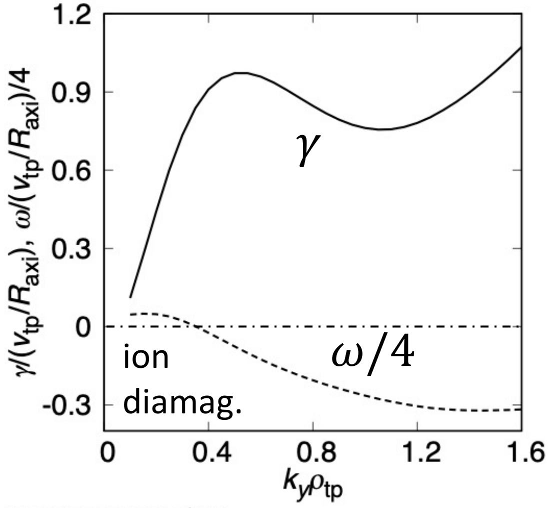


Generalization of ENet-based predictor

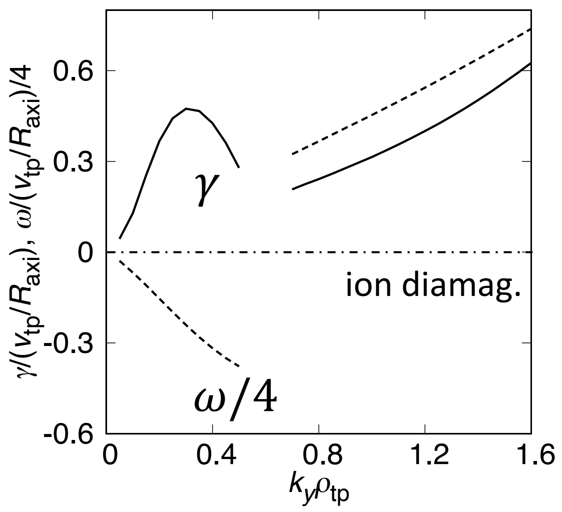
Caveats of the current ENet-based predictor, based on JT-60U data

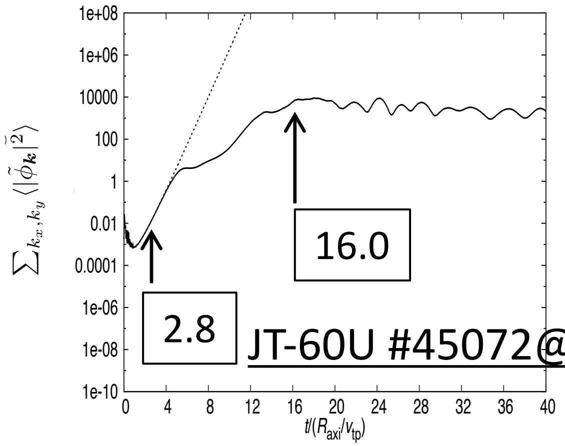
- It may not be applicable to the case away from that corresponding to the dataset used to train ENet.
- The cases with different linear dispersion relations will exhibit different fluctuation growth patterns.

JT-60U #45072@ρ=0.76

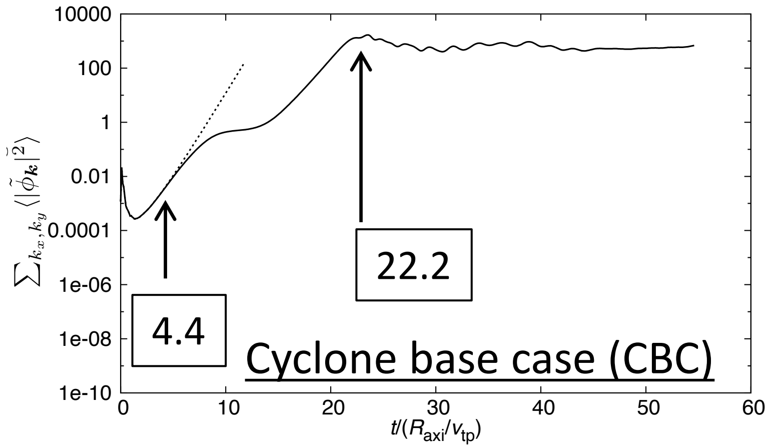


Cyclone base case (CBC)





JT-60U #45072@ρ=0.76



Cyclone base case (CBC)

> - Develop an ENet-based predictor trained on the **Cyclone base case (CBC)**, which is a *de facto* standard for gyrokinetic simulation benchmarking tests.
> - Test the performance when applied to the previous JT-60U case

ENet trained with Cyclone base case dataset

Test the predictive capability of ENet trained with CBC datasets for the CBC.

Extremely high R^2: 0.9945

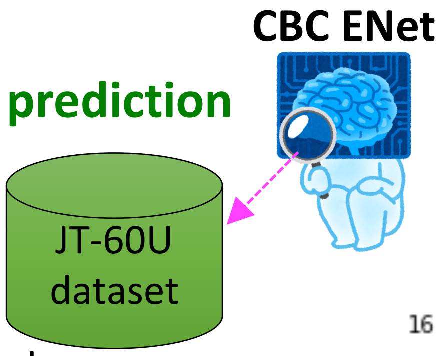
➔ The excellent performance was demonstrated for the CBC as well.

Apply the <u>CBC ENet</u> to JT-60U data (#45012@ρ=0.76) for prediction

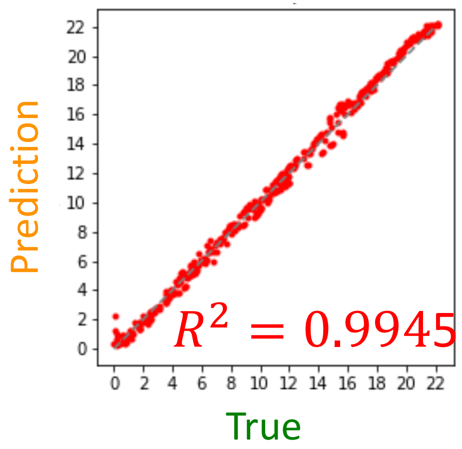
Fairly high R^2: 0.7634

➔ Overall trend has been captured well.

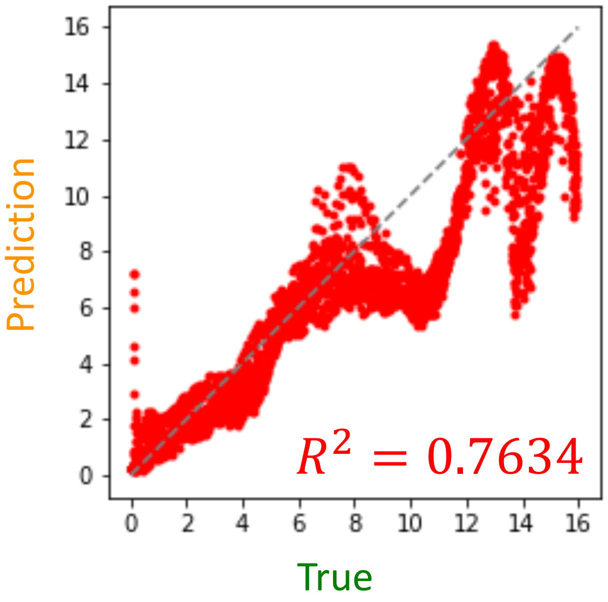
- The CBC ENet shows high predictability until some time after the transition to the nonlinearly growing phase ($\bar{t}\sim 8$).



<u>Cyclone base case (CBC)</u>



<u>JT-60U #45072@ρ=0.76</u>



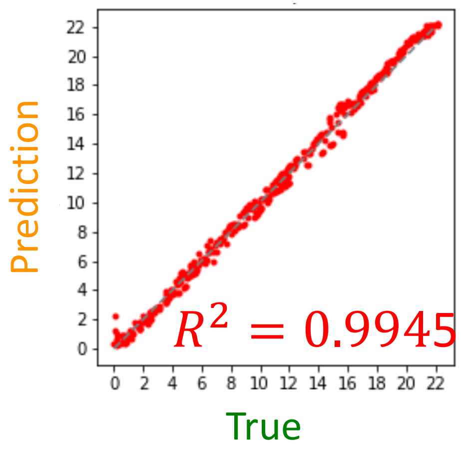
ENet trained with Cyclone base case dataset

Test the predictive capability of ENet trained with CBC datasets for the CBC.

Extremely high R^2: 0.9945

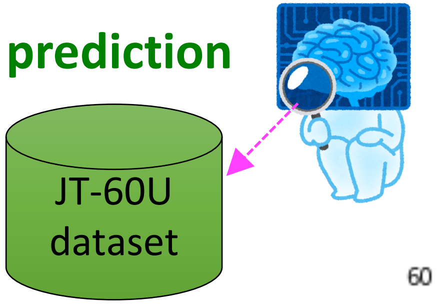
➔ The excellent performance was demonstrated for the CBC as well.

Cyclone base case (CBC)



Apply the CBC ENet to JT-60U data (#45012@ρ=0.26) for prediction

CBC ENet

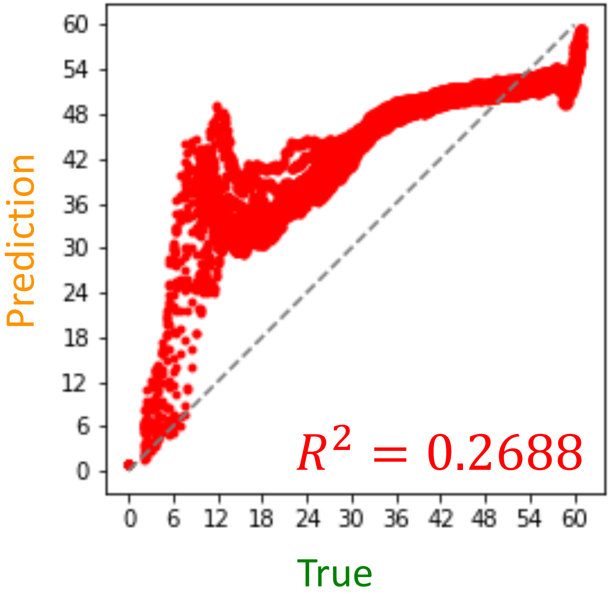


Poor R^2: 0.2688

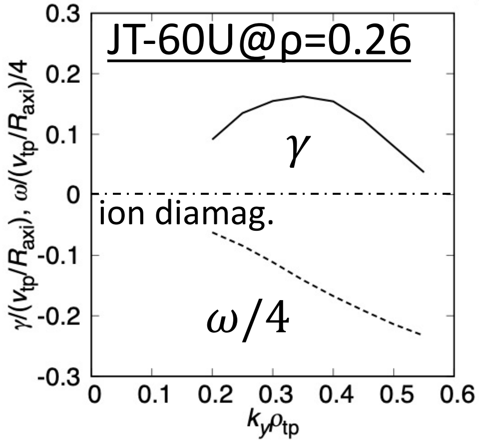
➔ Predicted a time farther into the future than it actually is

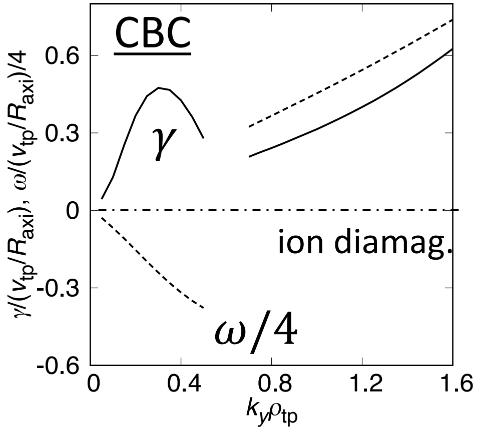
- Completely different linear dispersion relation b/w CBC and JT-60U #45072@ρ=0.26

JT-60U #45072@ρ=0.26



pure ITG





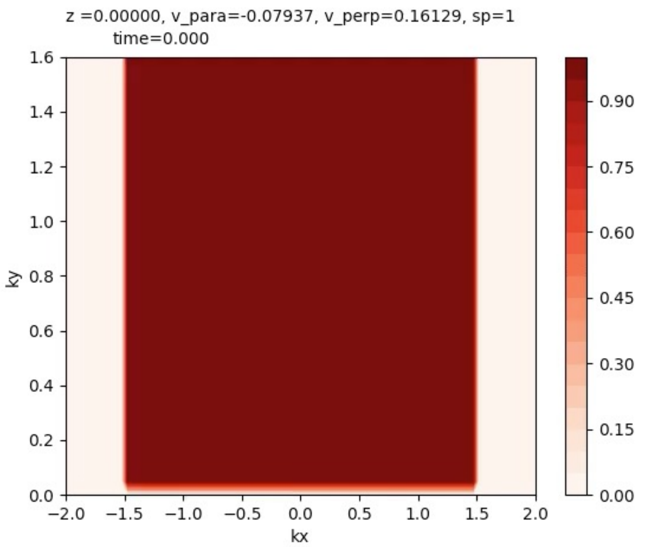
Multiple ENet models classified by dominant instabilities

Three cases prepared for different dominant instabilities

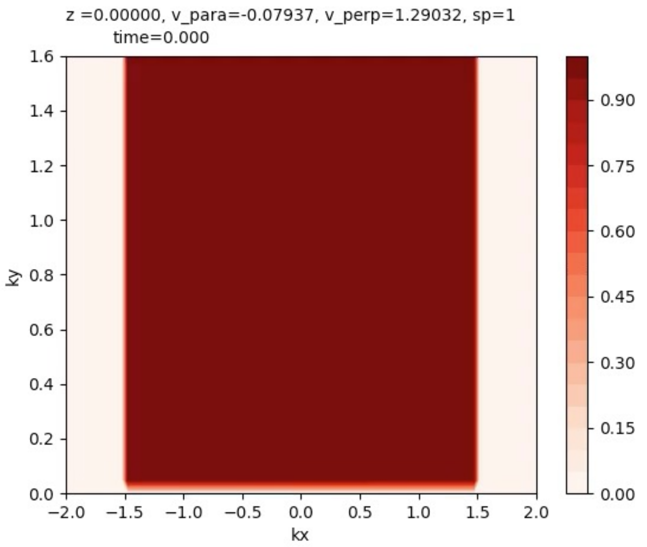
- CBC original for ITG/TEM
- CBC with flat T_e for pure ITG
- CBC with four parameters modified to develop pure TEM

	R/L_{T_i}	R/L_{T_e}	R/L_n	T_e/T_i
CBC	6.92	6.92	2.22	1
Pure ITG	6.92	0	2.22	1
Pure TEM	1	8	3	3

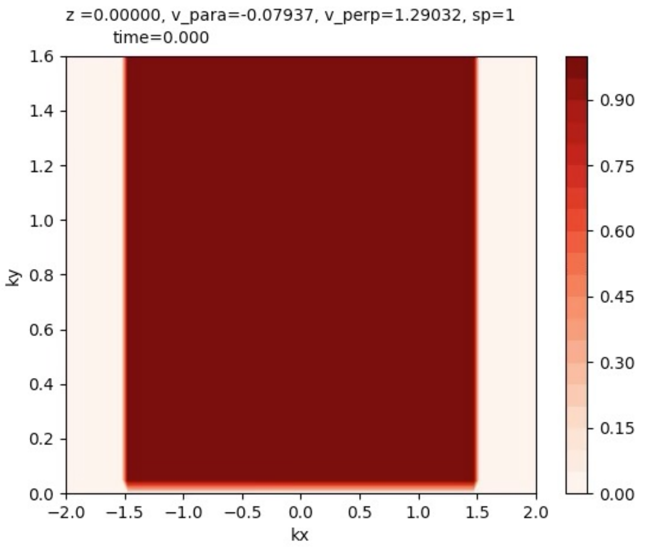
ENet for ITG/TEM



ENet for ITG



ENet for TEM



Linear calculations can be performed at low numerical cost.

➔ Choose the best model to better predict the time based on the linear results performed at initial.