GyroSwin: 5D Surrogates for Gyrokinetic Plasma Turbulence Simulations

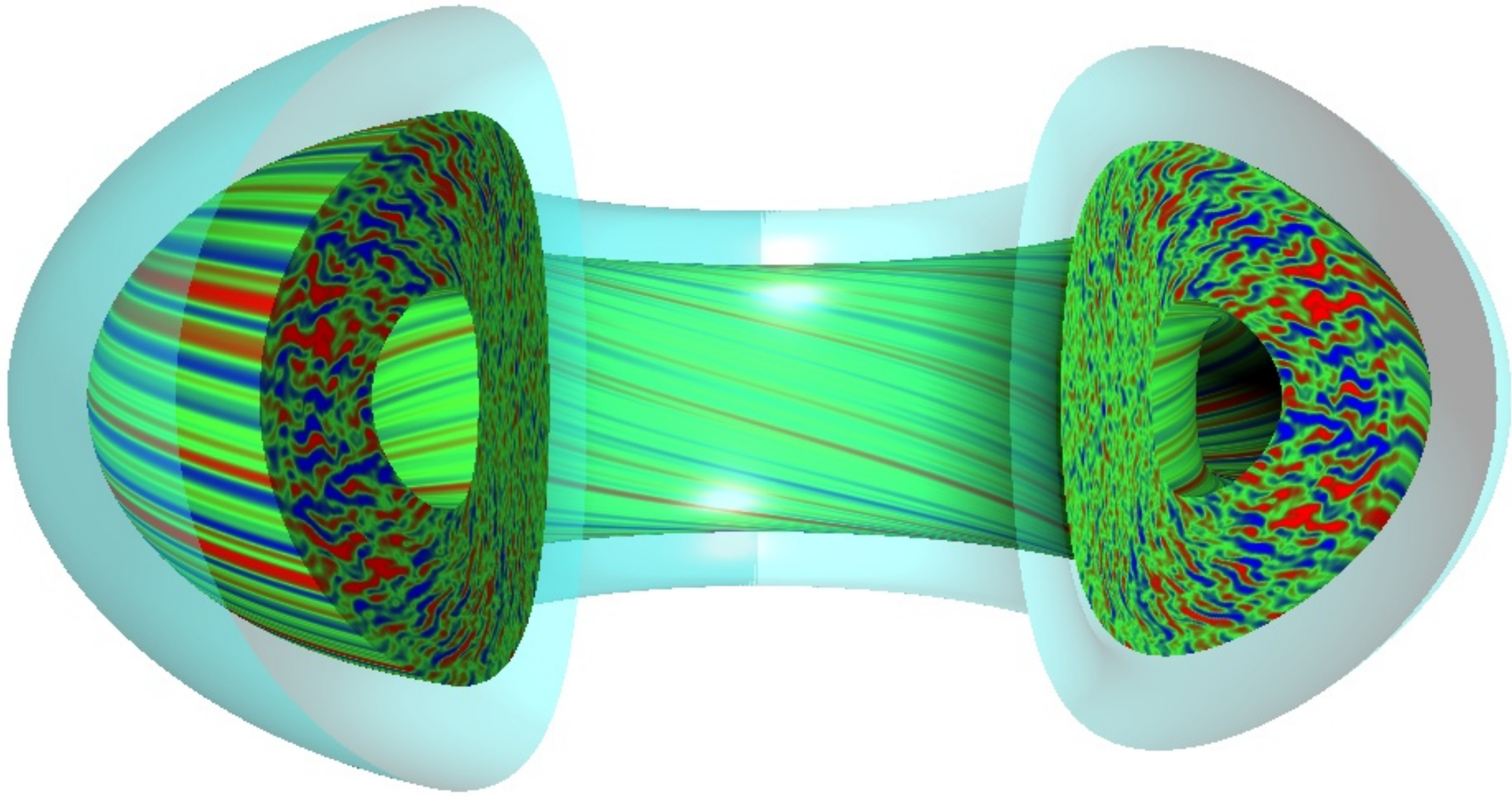
Fabian Paischer, Gianluca Galletti, William Hornsby, Paul Setinek, Lorenzo Zanisi, Naomi Carey, Stanislas Pamela, Johannes Brandstetter

JOHANNES KEPLER UNIVERSITÄT LINZ

UK Atomic Energy Authority

Plasma Turbulence

- Turbulence arises due to micro-instabilities in plasma (Francis F. Chen, 2016)
  - Leads to particle and energy transport (Romanelli et al. 2014)
  - Causes plasma to leak out of its magnetic cage => needs to be controlled!
  - To be able to control it, we first need to be able to simulate it!



Source: https://w3.pppl.gov/~hammett/viz/viz.html

Modeling Plasma Turbulence

- Described via nonlinear Gyrokinetic equations (Peeters et al. 2009)
  - Evolve a 5D (2x velocity, 3x spatial) distribution function over time
  - Numerically solving those equations can take months!
- Practitioners rely on reduced numerical solutions without nonlinear term (Bourdelle et al. 2007)

$$\underbrace{\frac{\partial f}{\partial t} + (v_{\parallel}\boldsymbol{b} + \boldsymbol{v}_D)\cdot\nabla f - \frac{\mu B}{m}\frac{\boldsymbol{B}\cdot\nabla B}{B^2}\frac{\partial f}{\partial v_{\parallel}}}_{\text{Linear}} + \underbrace{\boldsymbol{v}_{\chi}\cdot\nabla f}_{\text{Nonlinear}} = S$$

Motivation

- Quasilinear approximations (Bourdelle et al. 2007, Staebler et al. 2007)
  - Are based on linear simulations (3D) + saturation rules => not general (Giacomin et al. 2024)
  - Neglect nonlinear phenomena (zonal flows) => severely affect emerging turbulence

- Can we build a **scalable surrogate model** for 5D plasma turbulence modelling?

GyroSwin

- 5D local window attention (Liu et al. 2021)
- Multitask learning
    - 5D distribution function
    - 3D electrostatic potential fields
    - Scalar heat flux time trace
- Inductive Biases
    - Channelwise mode separation
    - Latent integral modules

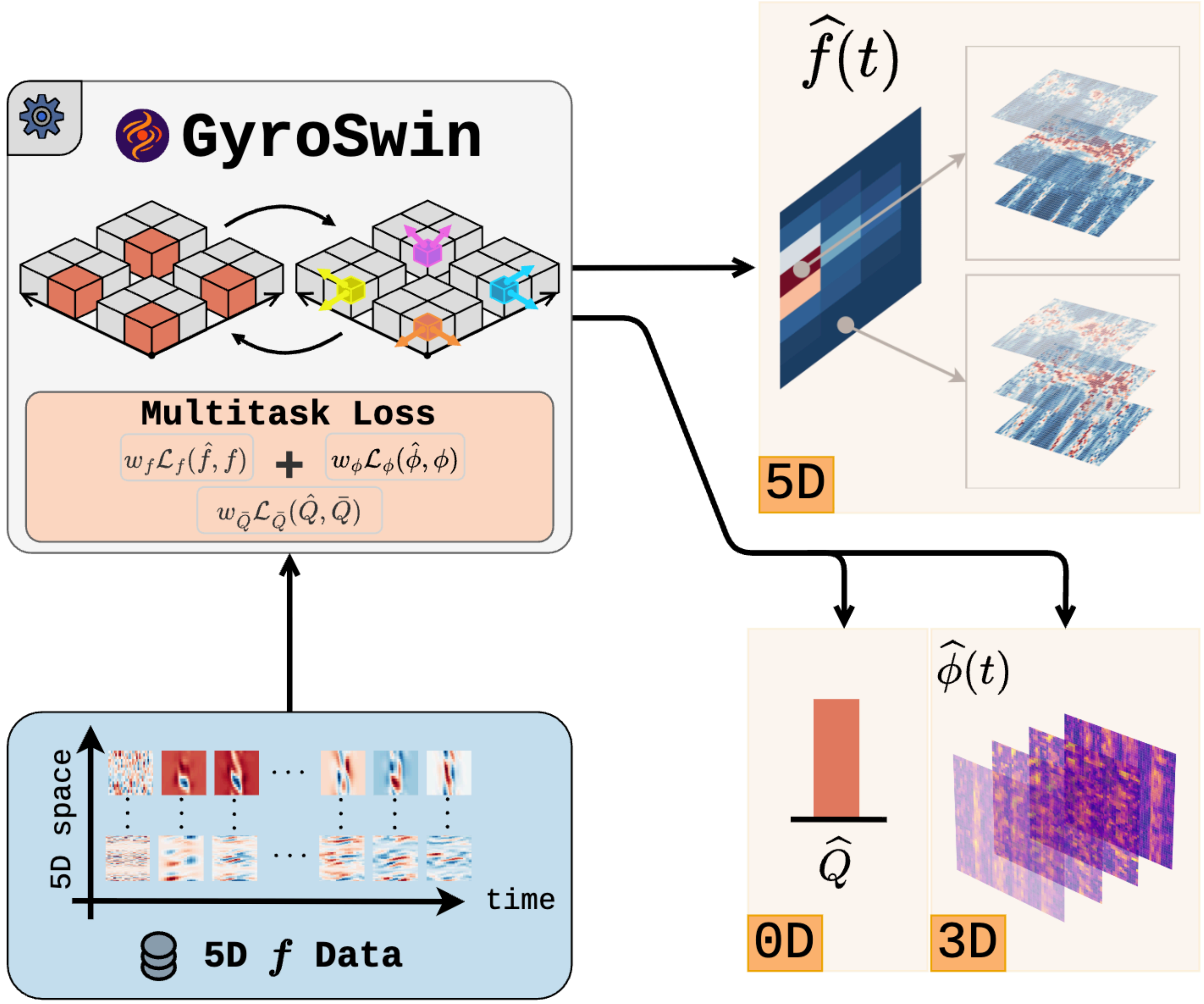


Figure: GyroSwin models the 5D distribution function of nonlinear gyrokinetics and incorporates integration blocks to predict 3D electrostatic potential fields and scalar heat flux.

Experiment design

- Data Generation using GKW [1]
  - Consider adiabatic electron approximation (resolution=32x8x16x85x32)
  - Generate 255 simulations (~6TB) with 4 operating parameters
  - Small training set (48 sims) for comparisons, large set (241 sims) for scaling experiments
  - Evaluate on 6 in-distribution (ID) and 5 out-of-distribution (OOD)

- Different surrogate approaches offer different capabilities:

Method	**Average Flux**	**Diagnostics**	**Zonal Flows**	**Turbulence**
Tabular Regressors, e.g., GPR, MLP	**1D→0D**	✗	✗	✗
SOTA Reduced Numerical modelling, e.g., QL	**3D→0D**	**3D→1D**	✗	✗
Neural Surrogates, e.g. GyroSwin (Ours)	**5D→0D**	**5D→1D**	**5D→1D**	**5D→5D**

Table: Comparison of different surrogate approaches by capabilities.

Turbulence modelling & average flux

Method	Input	f ID (↑)	f OOD (↑)	$\bar{Q}$ ID (↓)	$\bar{Q}$ OOD (↓)
SOTA Reduced Numerical modelling					
QL (Bourdelle et al., 2007)	3D	n/a	n/a	89.53 ± 11.76	95.22 ± 21.57
Classical Regression Techniques					
GPR (Hornsby et al., 2024)	0D	n/a	n/a	43.82 ± 10.84	59.28 ± 17.55
MLP	0D	n/a	n/a	50.50 ± 10.79	61.98 ± 18.41
Neural Surrogate Models (48 simulations)					
FNO (Li et al., 2021)	3D	9.33 ± 0.56	9.20 ± 0.58	119.88 ± 13.15	124.96 ± 23.27
PointNet (Qi et al., 2016)	5D	7.33 ± 0.21	7.40 ± 0.24	119.93 ± 13.15	125.05 ± 23.29
Transolver (Wu et al., 2024)	5D	9.83 ± 1.40	10.80 ± 1.46	119.93 ± 13.15	125.05 ± 23.28
ViT (Dosovitskiy et al., 2021)	5D	16.83 ± 1.49	19.20 ± 1.36	119.63 ± 13.13	125.13 ± 23.29
GyroSwin (Ours)	5D	26.50 ± 3.55	28.60 ± 8.82	67.68 ± 10.28	70.48 ± 17.21
Scaling GyroSwin to 241 simulations					
$\text{GyroSwin}_{\text{Small}}$ (Ours)	5D	98.00 ± 27.53	76.40 ± 17.60	23.72 ± 4.05	53.54 ± 18.10
$\text{GyroSwin}_{\text{Medium}}$ (Ours)	5D	94.17 ± 21.96	91.20 ± 18.61	37.24 ± 9.60	44.17 ± 17.68
$\text{GyroSwin}_{\text{Large}}$ (Ours)	5D	$\mathbf{110.33} \pm 19.74$	$\mathbf{111.80} \pm 23.86$	$\mathbf{18.35} \pm 1.56$	$\mathbf{26.43} \pm 9.49$

Table: Evaluation for 5D turbulence modelling and nonlinear heat flux prediction. We evaluate all methods across six ID and five OOD simulations. For Q we report RMSE of time-averaged predictions after an autoregressive rollout. For f we report correlation time for autoregressive rollouts with threshold τ = 0.1. Higher correlation time is better.

Conclusions

- GyroSwin is the first 5D neural surrogate for nonlinear Gyrokinetics

1. **Scalable**: Based on local window attention
2. **Efficiency**: Three orders of magnitude faster than nonlinear numerical code GKW
3. **Stable rollouts**: Trained on MSE on next time step. During inference model is rolled out for over 100 time steps without diverging.

neural-gyrokinetics

Paper

Blogpost

JOHANNES KEPLER UNIVERSITÄT LINZ

UK Atomic Energy Authority