

---

# 5D NEURAL SURROGATES FOR NONLINEAR GYROKINETIC SIMULATIONS OF PLASMA TURBULENCE

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

<https://arxiv.org/abs/2502.07469>

ICLR 2025 Workshops

---

April 28 / 29th 2025

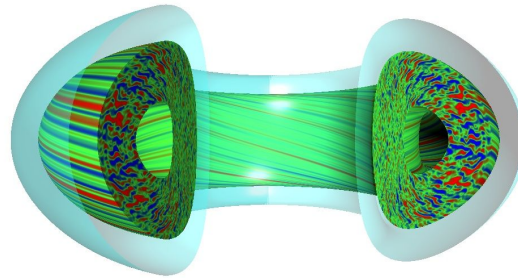


# Nuclear fusion

Produce energy by fusing hydrogen isotopes

We focus on **Tokamaks**, the most established fusion reactor design

- requires magnetic **plasma confinement**
- turbulence arises due to instabilities in plasma [1]
  - ↪ leads to particle, energy, and momentum transport [2]
  - ↪ **needs to be modeled** for **power plant design** and **control!**



*Plasma turbulence visualization [3]*

[1] Introduction to Plasma Physics and Controlled Fusion, Francis F. Chen, 2016

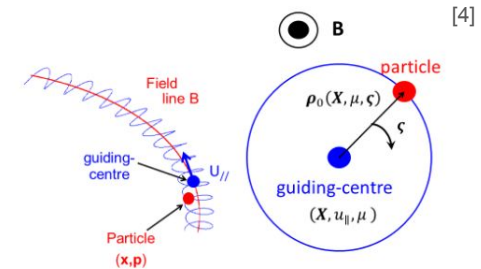
[2] JINTRAC: A system of codes for integrated simulation of tokamak scenarios, Romanelli et al., Plasma Fusion Research, 2014

[3] <https://w3.pppl.gov/~hammett/viz/viz.html>

# Plasma turbulence modelling

## **Turbulence is a key driver of plasma confinement degradation**

- Described via nonlinear **Gyrokinetics** equations [1]
  - evolve **5D distribution function** over time
  - solving these equations numerically is **very expensive**  
↪ hours to days for a single simulation on HPC!
- Development of fusion control requires knowledge of transport in “saturated state”
  - saturation is caused by nonlinear phenomena, e.g. **“zonal flows”** [2]
  - **Current reduced-order methods** (quasilinear) **don’t capture zonal flow** [3]



# Motivation

## Quasilinear approximations [1,2]

- rely on saturation rules derived from linear simulations  
↪ to date there is **no general quasilinear model** [3]
- **neglect nonlinear phenomena**

## What about machine learning?

- Current neural surrogates [4] mimic quasilinear models → **no zonal flow**
- Can we learn GK surrogates that **directly evolve the 5D distribution function?**

# Contributions

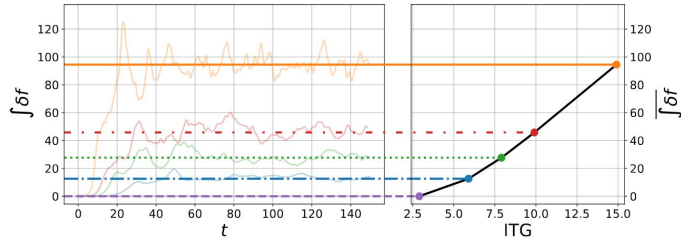
- **Data generation and inspection**
- **5D neural surrogate model**
  - Based on hierarchical vision transformers
- **Evaluation of physical quantities**
  - Electrostatic potentials
  - Heat flux

# Contributions

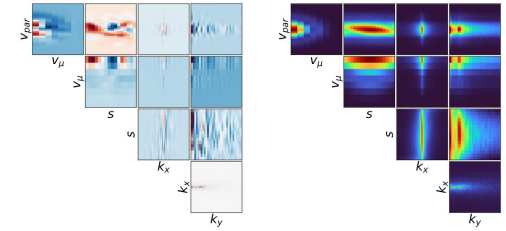
- Data generation and inspection

Numerical code **GKW** [1]

Sampled ion temperature gradient (ITG)

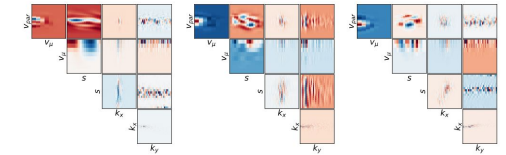


Heat fluxes for our dataset



(a) Mean

(b) Standard Deviation

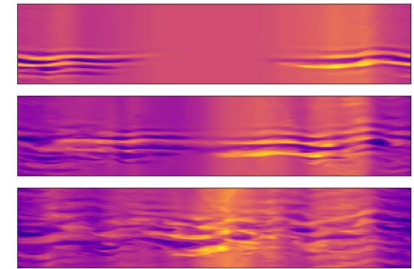


(c) Start slice

(d) Middle slice

(e) End slice

5D density function



3D electrostatic potential

# Contributions

- Data generation and inspection
- **5D neural surrogate model**
  - Based on hierarchical vision transformers

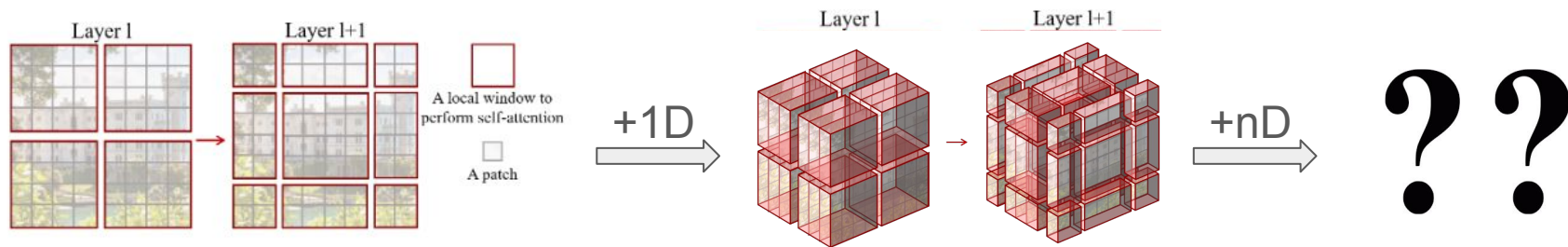


**Problem: How can we build neural surrogates that can process 5D data?**

# Contributions - neural surrogate

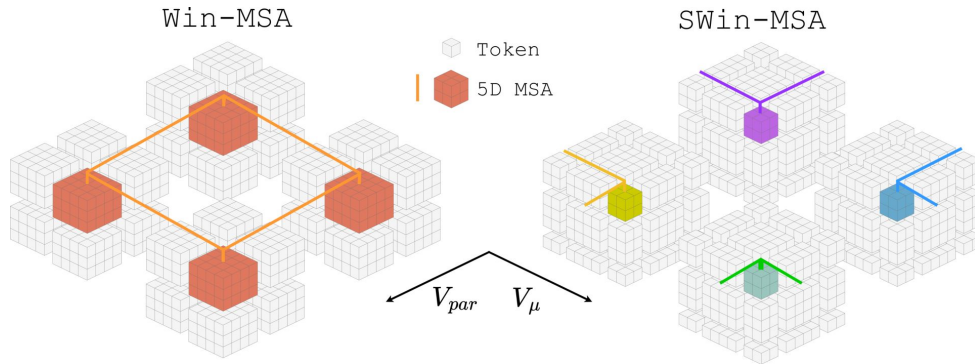
How can we build neural surrogates that can process 5D data?

- CNNs [1] ?  
↳ Impractical due to runtime and memory constraints
- (Vision) transformers [2]?  
↳ 5D results in extremely long sequences, self-attention scales quadratically...  
↳ **Swin** [3,4]: **performs attention in parallel in local windows**



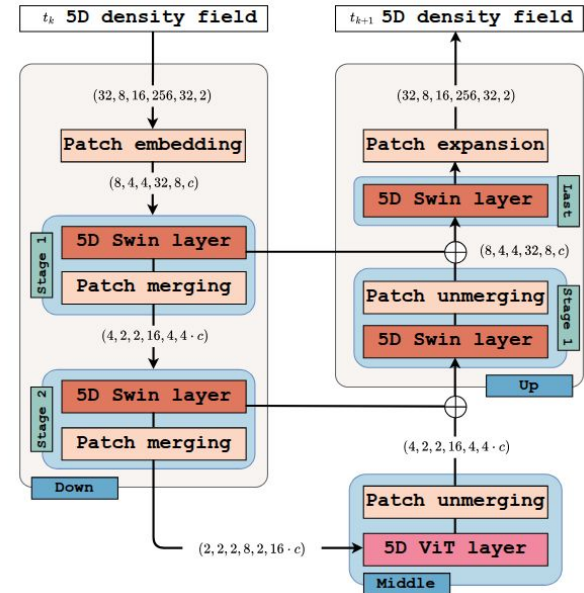
# Contributions - 5D Hierarchical Swin Transformer

**5D swin attention** → generalized nD swin



Hierarchical **UNet** structure [1]

- Downsampling (Patch merging)
- Upsampling (Patch expansion)

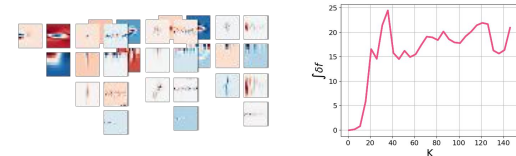


# Contributions

- Data generation and inspection
- 5D neural surrogate model
  - Based on hierarchical vision transformers
- **Evaluation of physical quantities**
  - Electrostatic potentials
  - **Heat flux**

Physical quantities:  
**electrostatic potential, heat flux**

Visual inspection

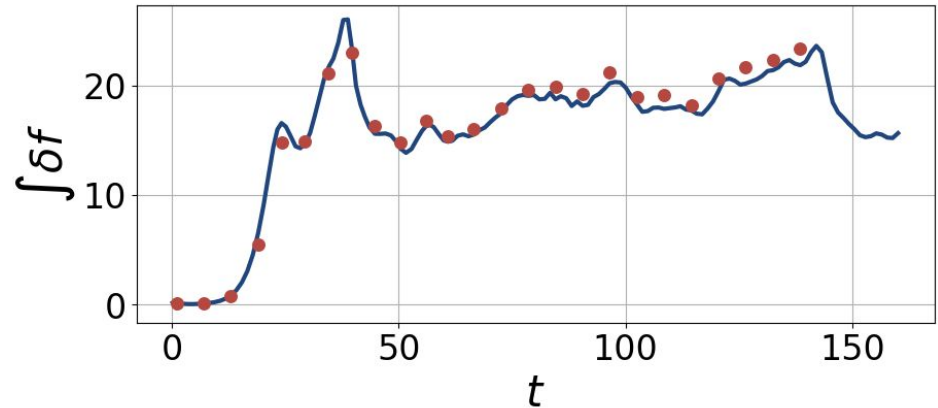
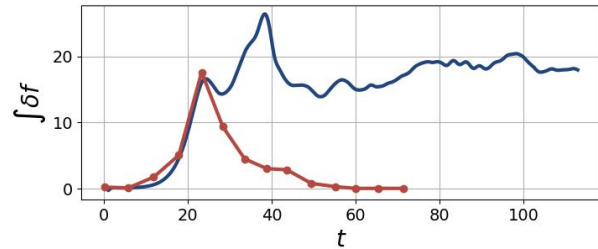


# Evaluation - heat flux time trace

- Heat flux time trace for single-step prediction is decent ✓
- (At least) **two orders of magnitude speedup vs GWK** ✓
- Generalization to unseen ion temperature gradient ✓

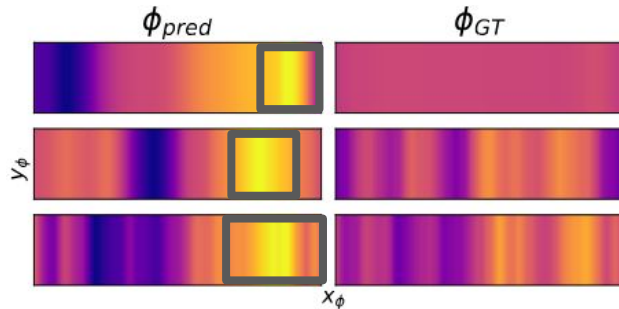
... BUT

↳ **autoregressive** ✗

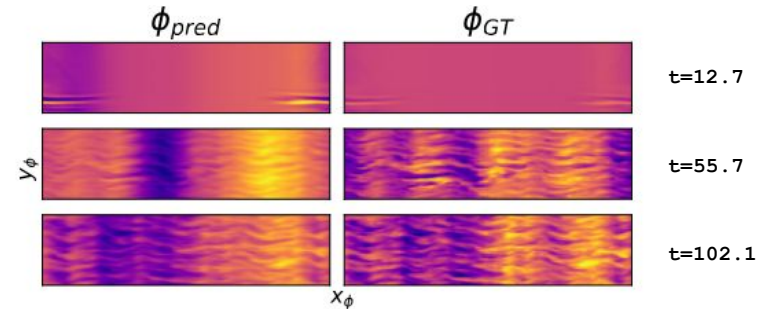


# Evaluation - electrostatic potentials

- Y wave vector is well reproduced ✓  
↳ explains good alignment of predicted heat flux
- **Zonal flow is overestimated** ⚠  
↳ dampens turbulence during autoregressive rollout

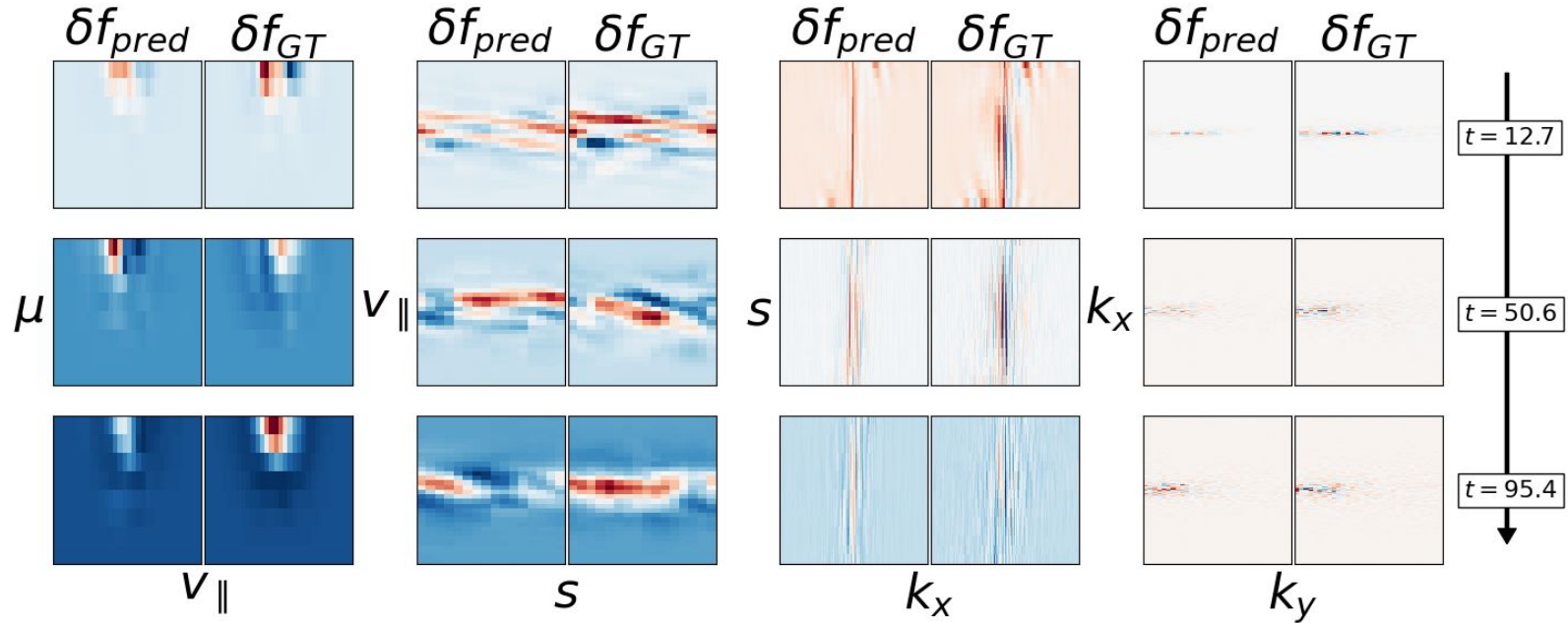


(a) Zonal mode of the electrostatic potential.



(b) Electrostatic potential with zonal mode.

# Evaluation - visual inspection



# Conclusions and future work

## First 5D neural surrogate for nonlinear gyrokinetics

- 5D density function rollouts, one-step heat flux
- ~two orders of magnitude faster than GKW
  - ↪ increases for higher fidelity simulations

## Open challenges

- data storage ⚠
- physical grounding
- verification

## Future directions

- Improve zonal flow modeling → remedy heat flux decay (autoregressive)
- Move from low fidelity to high fidelity

# Thank you for listening!



Gianluca

Fabian

Paul

Johannes



William

Naomi

Lorenzo

Stan

X 2502.07469

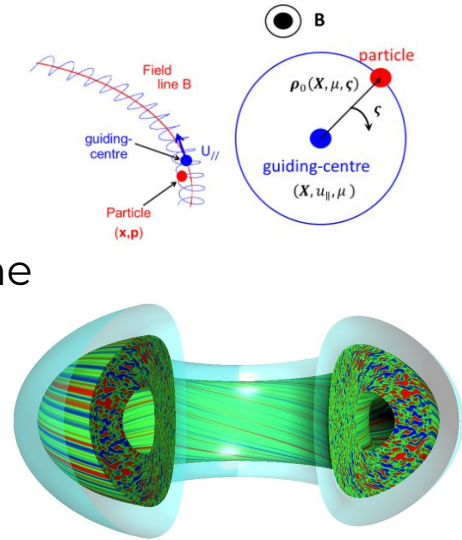


# Supplementary - Experimental design

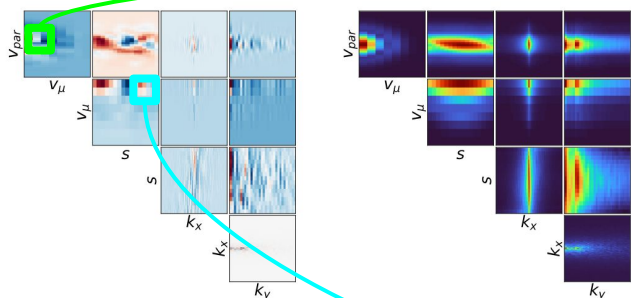
- Train on ITG = [4.9, 5.9, 7.9, 8.9] and evaluate on holdout simulation (ITG = 6.9)
  - ~**500GB** of data (only 5 trajectories, **still very limited**)
  - Enough to answer: “*can the surrogate generalize across configurations?*” **YES!**
- Objective for the surrogate is to predict the next time slice of the simulation
  - training on (1-step) normalized MSE loss  $\mathcal{L}(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{\|x_i - y_i, 1\|_2^s}{\|y_i\|_2^s}$
  - training is done in **real space**, raw data in frequency space
  - autoregressive at inference
- Data-parallel across multiple GPUs with mixed precision
  - allows for larger batch sizes and more efficient training

# Supplementary - Gyrokinetics

- In a confined plasma, the magnetic field is very strong ( $\sim 11$  tesla in ITER)
- Particle gyratory motion is extremely fast with respect to the turbulent modes that perturb its orbit
- This leads to the ability to “average out” the particle gyro motion, and thus reduce 6D phase space to 5D
- Turbulence is highly elongated along the field line (See Fig.) but has small scales perpendicular to the confining field
- Turbulence can generate large scale “zonal flows” which regulate the turbulence. Steady state is a delicate balance between the turbulence drive and the damping effects of flows
  - Turbulence is one of the dominant mechanisms at to how heat “leaks” out of the confined plasma



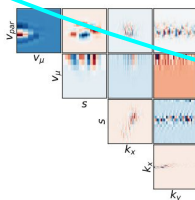
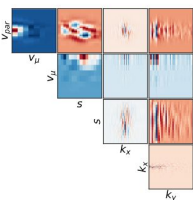
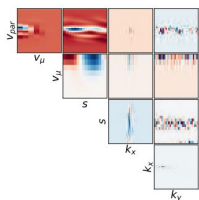
# Supplementary - how to visualize 5D?



(a) Mean

(b) Standard Deviation

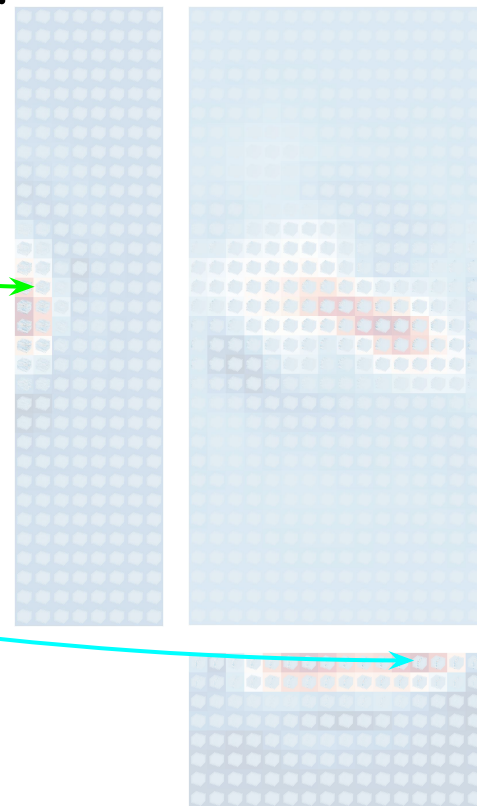
...



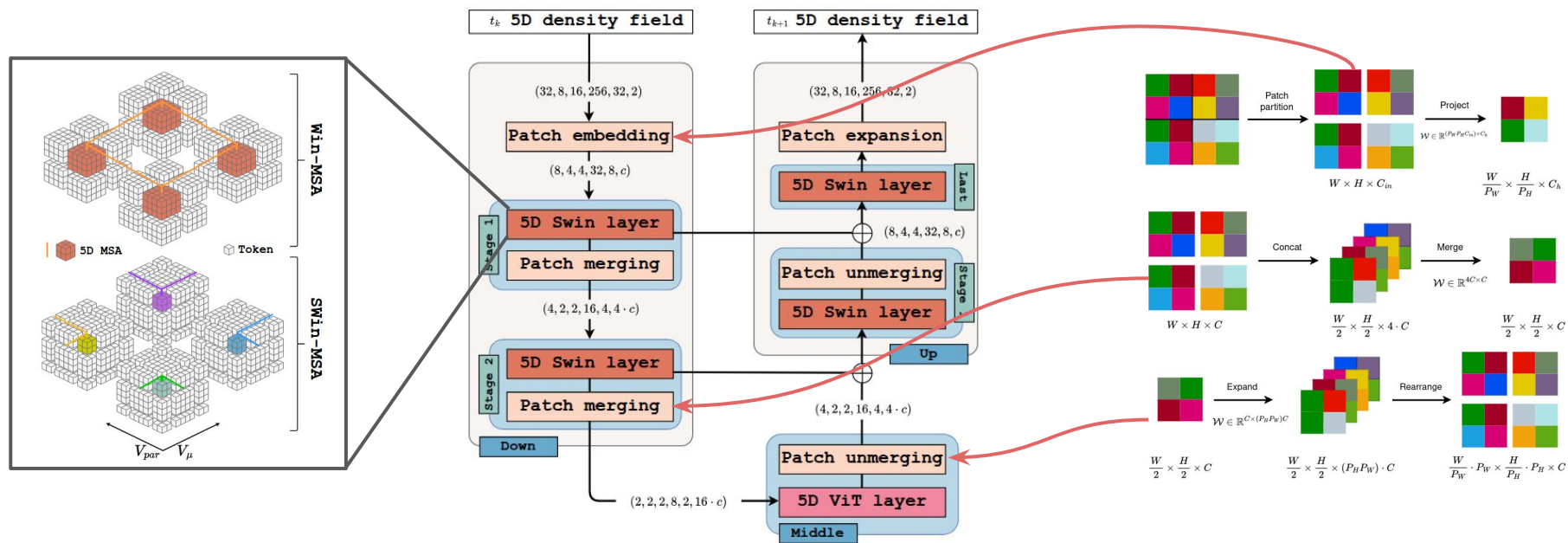
(c) Start slice

(d) Middle slice

(e) End slice



# Supplementary - 5D Hierarchical Swin Transformer



# Supplementary - potential over time

