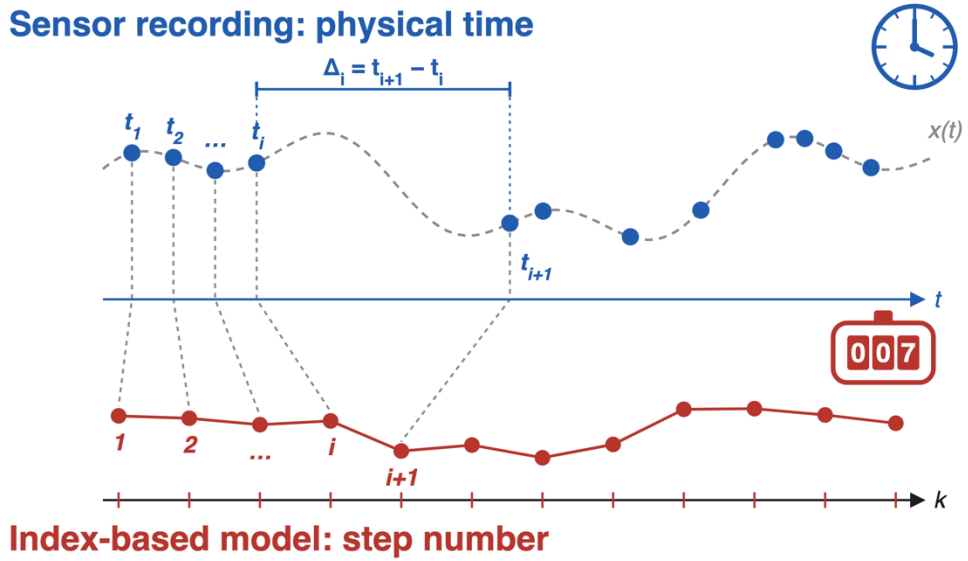
# TIDES: Implicit Time-Awareness in Selective State Space Models

**Taylan Soydan, Miguel A. Bessa, Dirk Mohr,**

**Rui Barreira***

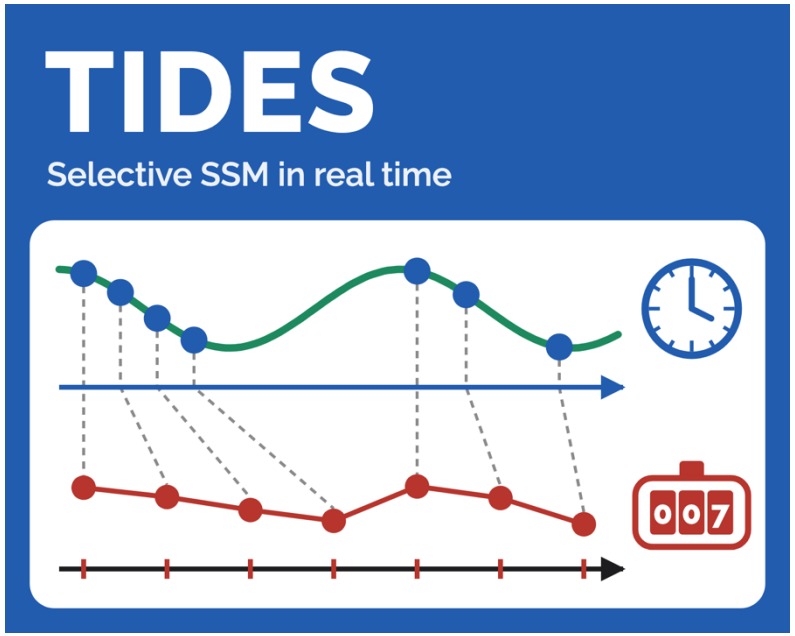



*rbarreira@ethz.ch

arXiv:2605.09742

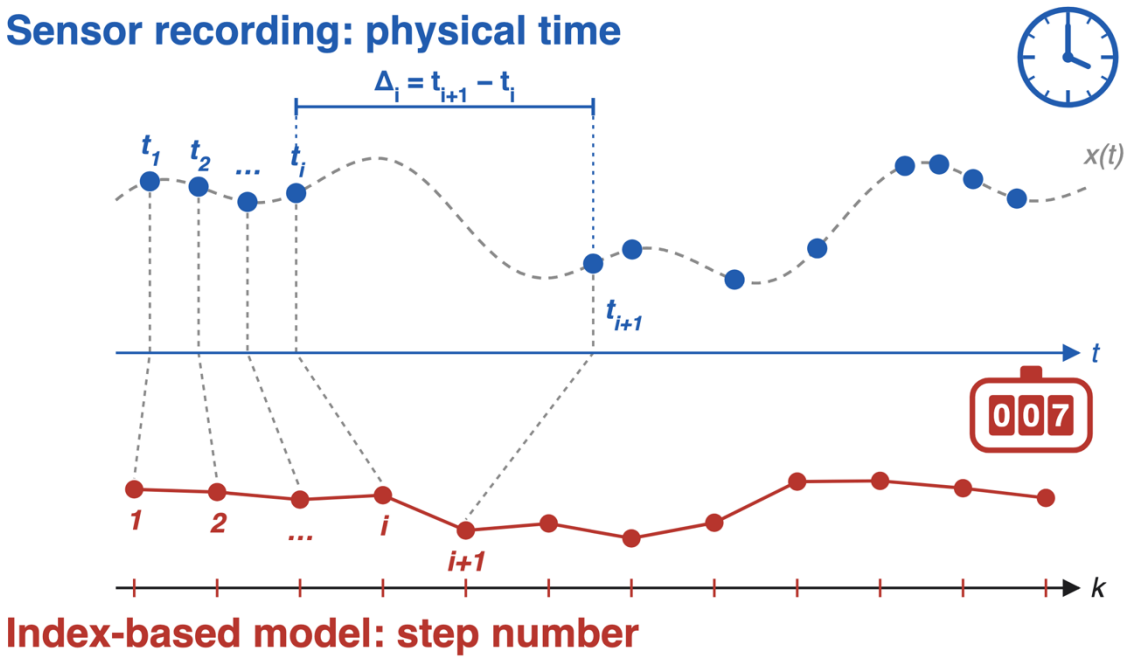
# Outline

1. Motivation – real-world data is irregular
2. Background – two families of state space models
3. Method – selectivity & physical time
4. Fading Flash – two failure modes
5. Benchmarks
6. Ablation – random drop on EigenWorms
7. Summary



# Motivation – real-world data is irregular

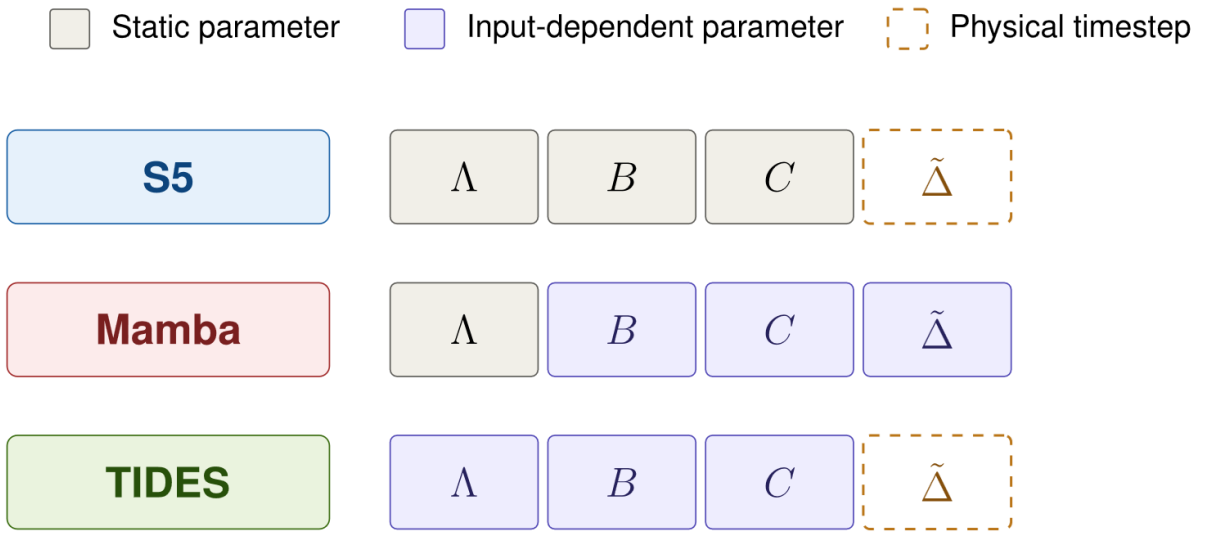
- Clinical measurements, astronomical surveys, and event sensors are rarely sampled on a regular grid;
- The elapsed time between observations, $\Delta$, carries information about the underlying continuous process;
- An index-based model only counts steps: it places observations on a uniform grid and loses the gaps $\Delta_i$.



# Background – two families of SSMs

- **Continuous-time SSMs (e.g., S5)** discretize $\dot{x} = Ax + Bu, y = Cx$ with the physical time interval, $\tilde{\boldsymbol{\Delta}} = \Delta$.
  - Irregular sampling comes naturally, but the model cannot choose what to remember or forget.

- **Selective SSMs (e.g., Mamba)** make $\tilde{\boldsymbol{\Delta}}$ , $\boldsymbol{B}$, $\boldsymbol{C}$ functions of the input.
  - $\tilde{\boldsymbol{\Delta}} \neq \Delta$: a real timestamp must be appended as an extra input feature;
  - Expressive; however, the models do not extrapolate.

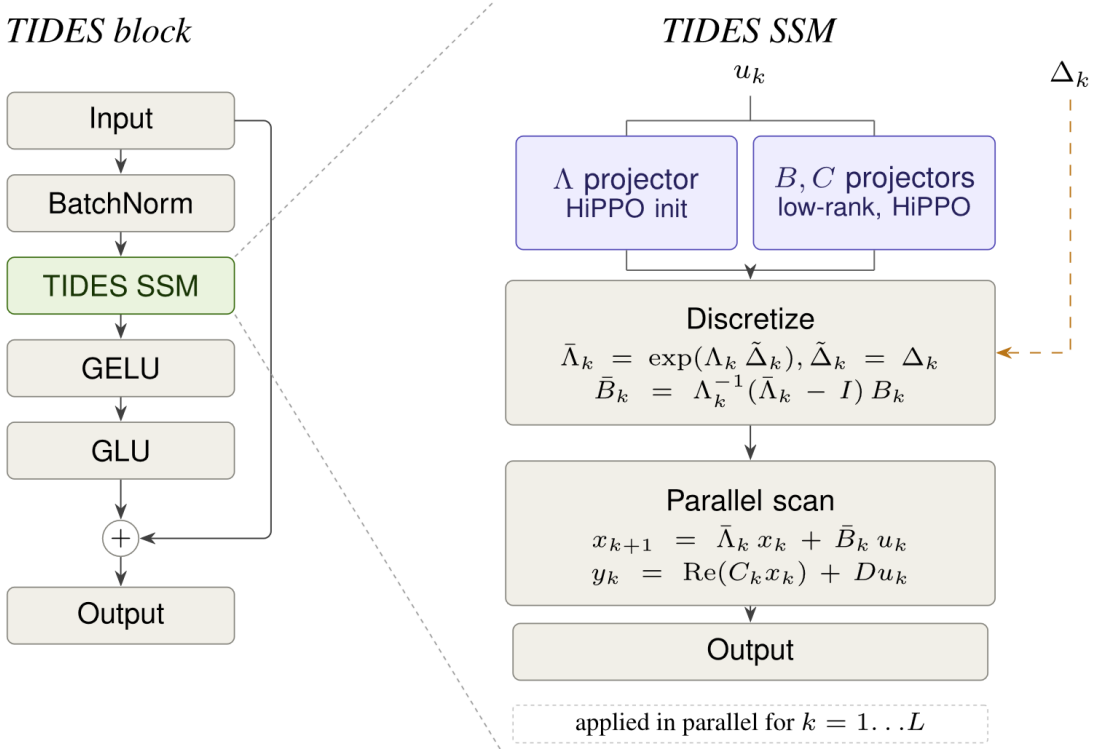


# Method – selectivity & physical time

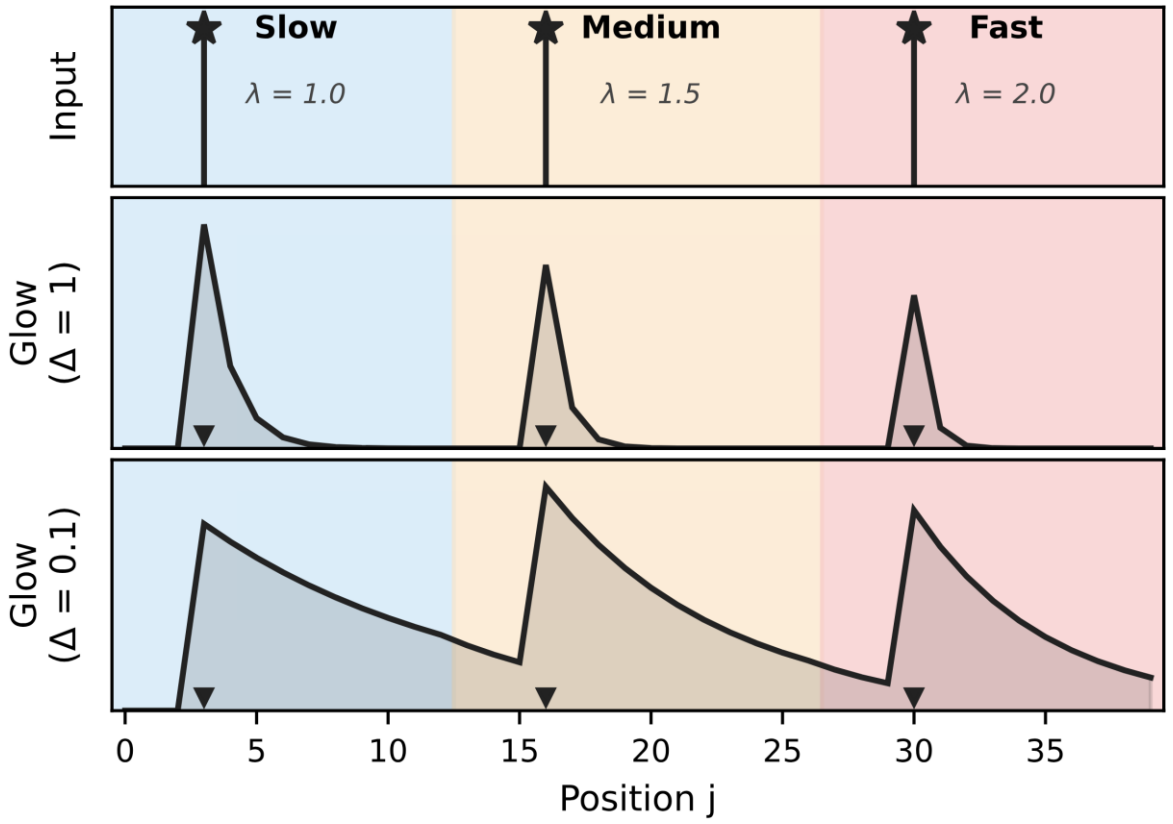
- Diagonal SSM recurrence under zero-order hold:

$$x_{k+1} = \exp(\Lambda_k \tilde{\Delta}_k)\, x_k + \bar{B}_k u_k, \qquad y_k = \mathrm{Re}(C_k x_k) + D u_k$$

- TIDES's design choices:
  - $\tilde{\Delta} = \Delta$ **stays physical**: observation timestamp;
  - $\mathbf{Re}(\Lambda)$ **input-dependent**: a mode can "forget faster" at an event boundary or "hold longer";
  - $\mathbf{Im}(\Lambda)$ **static**: per-token frequency modulation has no coherent continuous-time meaning;
  - $\boldsymbol{B}, \boldsymbol{C}$ **input-dependent**: extra expressivity.



# *Fading Flash* – task setup




- 40 detectors are hit by sparse flashes; each glows and fades exponentially;
- Three zones have different fade rates;
- A global clock Δ stretches or compresses the trajectory;
- Tests **expressivity** (three decays need input-dependence) and **extrapolation** (unseen Δ);
- **S5**, **Mamba,** and **TIDES** matched in parameters.

# *Fading Flash* – two failure modes

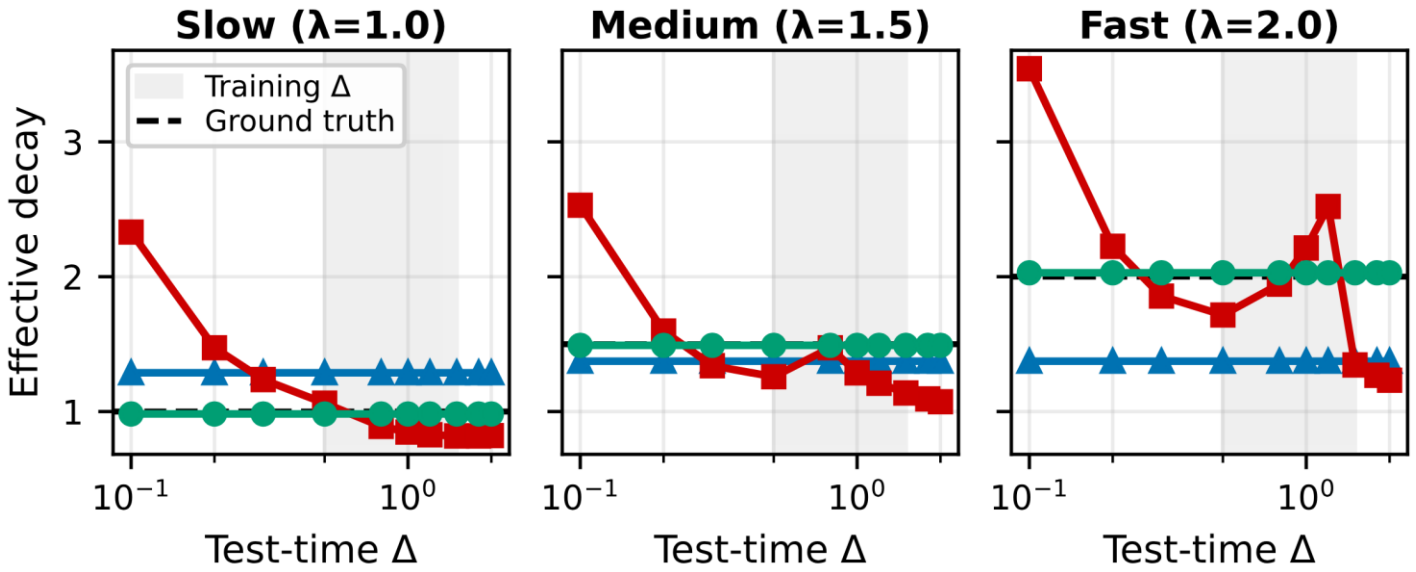



*Fig: Learned decay per zone. Shaded: training range.*

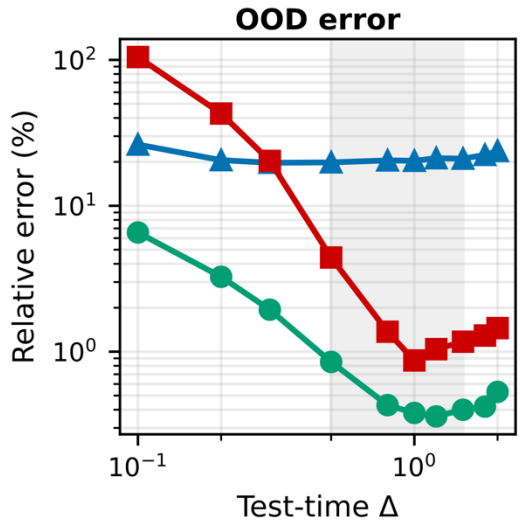



*Fig: relative error vs. test Δ.*

## Findings

**S5 lacks expressivity:** robust across Δ, but learns poorly

**Mamba lacks extrapolation:** error grows outside training Δ

**TIDES overcomes both:** learns target decays and stays robust across Δ

# Benchmarks – TIDES is state-of-the-art!

### UEA classification (6 datasets)

| Model | Acc. | Avg. Rank ↓ |
|---|---|---|
| **TIDES** | **66.1** | **3.00** |
| Mamba-3 | 64.7 | 4.17 |
| LogNCDE | 64.3 | 4.25 |
| RFormer | 64.5 | 5.17 |
| S5 | 61.8 | 7.08 |

### Physiome-ODE (50 datasets)

| Model | Wins | Avg. Rank ↓ |
|---|---|---|
| **TIDES** | **16** | **2.40** |
| LinODEnet | 16 | 2.62 |
| GraFITi | 5 | 2.82 |
| GraFITi-C | 10 | 3.54 |
| CRU | 3 | 3.68 |

### Natively irregular benchmarks vs. reference baseline

| Dataset | Metric | TIDES | Base | Baseline |
|---|---|---|---|---|
| MACHO | Acc. ↑ | 92.8 | 92.7 | iTCN |
| ASAS-SN | Acc. ↑ | 94.2 | 94.5 | iTCN |
| TimeSen2Crop | OA ↑ | 89.80 | 88.70 | MLP+LTAE |
| EuroCropsML | Acc. ↑ | 80.26 | 79.14 | DirPA |
| SHD | Acc. ↑ | 96.1 | 95.9 | Event-SSM |
| SSC | Acc. ↑ | 88.6 | 88.4 | Event-SSM |
| DVS128 | Acc. ↑ | 98.9 | 99.2 | Event-SSM |
| USHCN | MSE ↓ | 0.147 | 0.159 | APN |
| USHCN | MAE ↓ | 0.211 | 0.261 | APN |

*Astronomy (MACHO, ASAS-SN), agriculture (TimeSen2Crop, EuroCropsML), neuromorphic (SHD, SSC, DVS128), climate (USHCN).*

# Ablation – randomly drop training timesteps

- Train with a random 50% of steps dropped; test at drop rates $r$ = 0.1 to 0.9, timestamps kept during training;
- **TIDES**: best mean accuracy (0.74); maintains 0.71 at r = 0.9;
- **Mamba**: 0.77 → 0.42 at r = 0.9 (poor extrapolation);
- **S5**: 0.42 at every rate (poor expressivity);
- **RFormer**: 0.77 → 0.56 at r = 0.9.

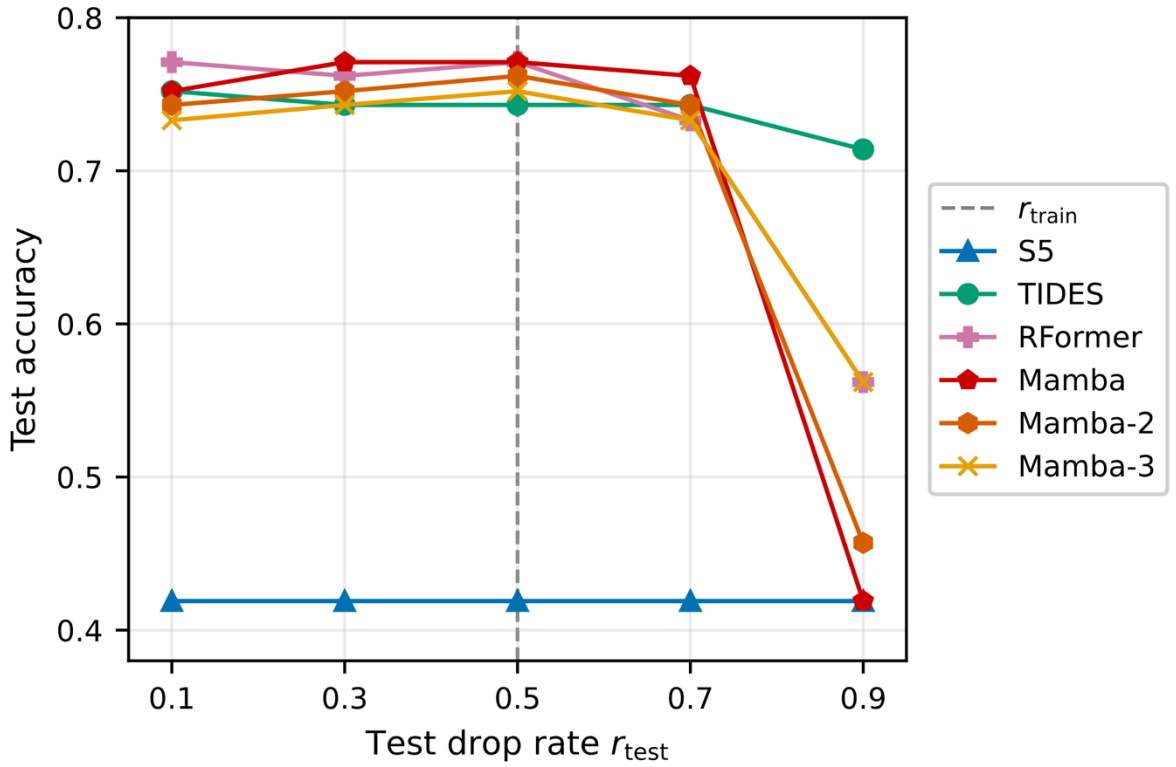
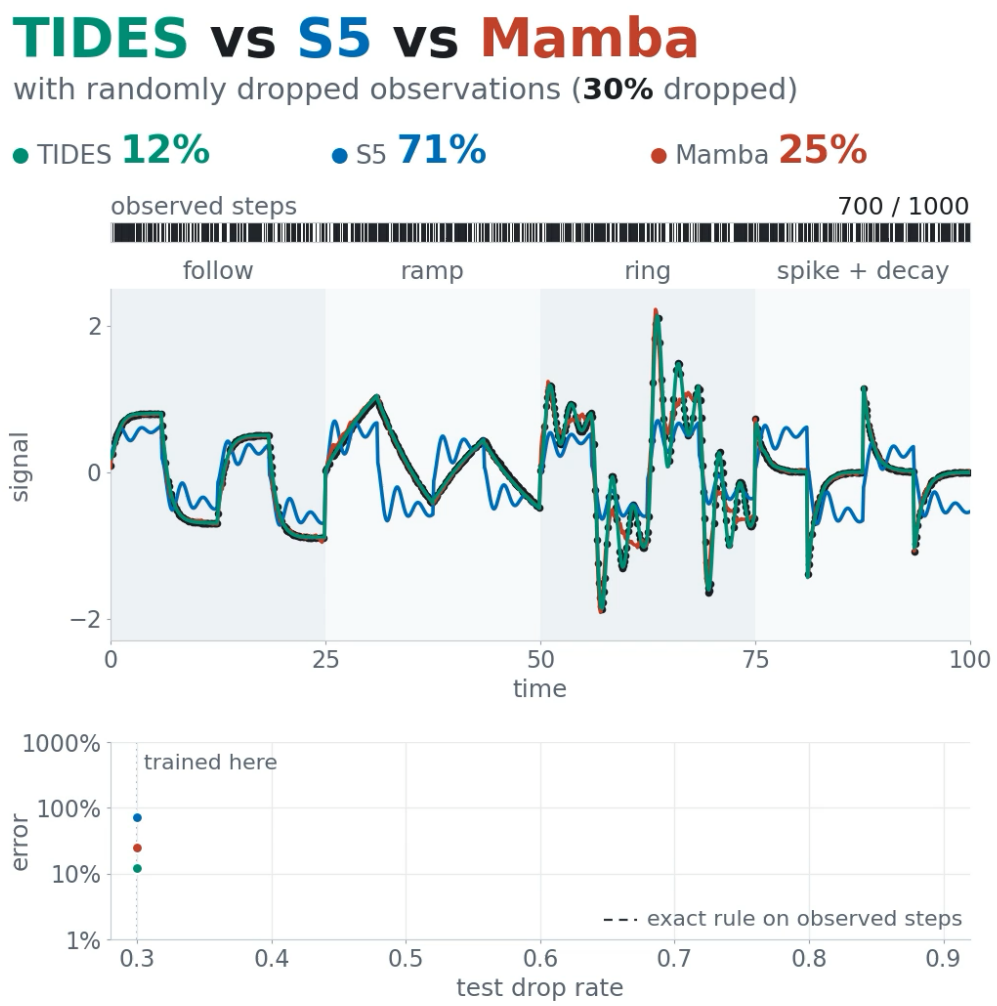



*Fig: Test accuracy vs. drop rate; trained at r = 0.5, mean of 3 seeds.*

# Summary




## Problem

**Mamba** learns its own step size, so the real-time gap is lost;
**S5** keeps it, but cannot select.

## Fix

**TIDES** makes decay rates $\mathrm{Re}(\Lambda)$ input-dependent and steps by the physical gap $\Delta$

## Evidence

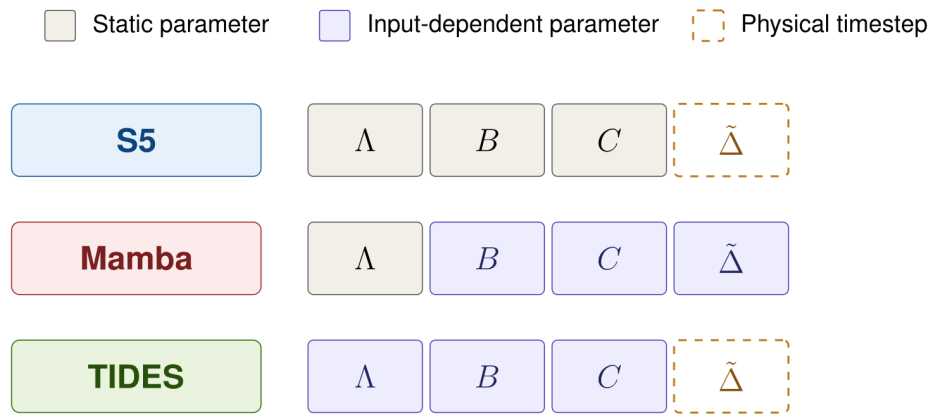
**TIDES** learns true decays in *Fading Flash.* State-of-the-art accuracy on the UEA and Physiome-ODE datasets (among others).

Code


*rbarreira@ethz.ch
arXiv:2605.09742

# Appendix

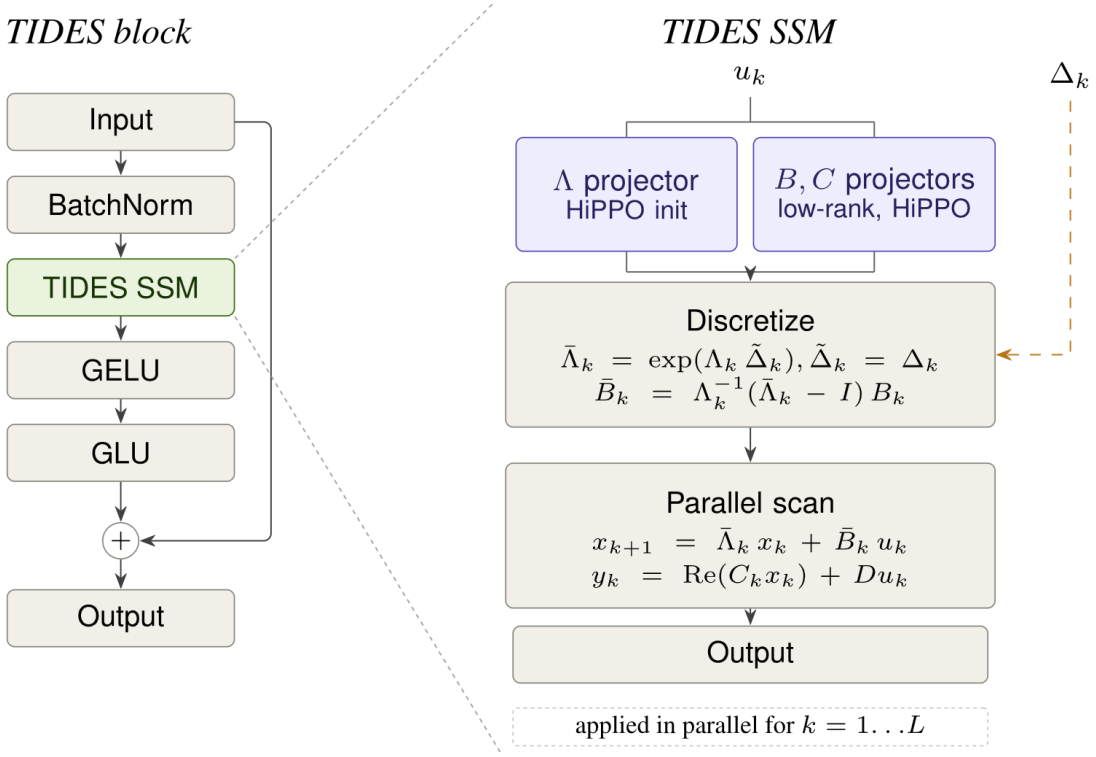
- A1 – Selectivity heads
- A2 – Ablation variants
- A3 – Failure analysis



# A1 – Selectivity heads

- **Initialization:** biases start at S5's HiPPO values, weights at zero – training begins from a well-conditioned recurrence;
- **Low-rank $W_B$ ,$W_C$:** cost $O(PH^2) \rightarrow O(rPH)$;
- **RMSNorm** on projected $Re(\Lambda_k)$, $B_k$, $C_k$;
- Stable reparameterization of $Re(\Lambda)$.

$$Re(\Lambda_k) = W_\Lambda u_k + Re(\Lambda_0)$$
$$B_k = W_B u_k + B_0, \qquad C_k = W_C u_k + C_0$$



# A2 – Ablation variants and full results

*Columns 2–5: which component is **input-dependent (ID)** or **static/linear time invariant (LTI)**. Columns 6–10: test accuracy at drop rate r.*

| Model | Re(Λ) | Im(Λ) | B, C | Δ̃ | r=0.1 | 0.3 | 0.5 | 0.7 | 0.9 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|
| S5 | LTI | LTI | LTI | LTI | 0.419 | 0.419 | 0.419 | 0.419 | 0.419 | 0.419 |
| Mamba (Δ input) | LTI | LTI | ID | ID | 0.705 | 0.705 | 0.705 | 0.686 | 0.400 | 0.640 |
| **TIDES** | **ID** | **LTI** | **ID** | **LTI** | **0.752** | **0.743** | **0.743** | **0.743** | **0.714** | **0.739** |
| TIDES-Λ | ID | LTI | LTI | LTI | 0.457 | 0.457 | 0.457 | 0.467 | 0.467 | 0.461 |
| TIDES-BC | LTI | LTI | ID | LTI | 0.695 | 0.695 | 0.695 | 0.695 | 0.686 | 0.693 |
| TIDES-full | ID | ID | ID | LTI | 0.676 | 0.667 | 0.676 | 0.648 | 0.648 | 0.663 |
| RFormer | – | – | – | – | 0.771 | 0.762 | 0.771 | 0.733 | 0.562 | 0.720 |

- Trained on EigenWorms UEA dataset, at a drop rate of r = 0.5, mean of 3 seeds;
- SSM variants matched at ≈30k parameters (L = 1, P = 16, bidirectional, ZOH);
- Static **B**, **C** caps expressivity (S5, TIDES-Λ); input-dependent $\widetilde{\mathbf{\Delta}}$ breaks extrapolation (Mamba); selective **Im**(**Λ**) slightly hurts (TIDES-full).

# A3 – Failure analysis

## Mamba (Δ as input)

The learned step is computed from the input and the gap.

$$\tilde{\Delta}_k = \mathrm{softplus}(W_\Delta\,[u_k,\,\Delta_k])$$

It is fit only on the training range of Δ; outside thatrange,e the learned step is unconstrained

(0.705 → 0.400 at r = 0.9)

## Rformer (non-SSM baseline)

10 fixed path-signature windows, baked into positional embeddings, result in ≈900 observations per window at r = 0.5, but only ≈180 at r = 0.9.

The features are out of distribution

(drops from 0.771 → 0.562 at r = 0.9)

## TIDES

Δ enters only through the exponential:

$$e^{\Lambda_k \Delta_k}.$$

A new gap changes the decay exactly as continuous time prescribes, with no retraining.

It extrapolates well across resolutions
(0.752 → 0.714 at r = 0.9).