xLSTM-Mixer: Multivariate Time Series Forecasting by Mixing via Scalar Memories

Maurice Kraus*, Felix Divo*, Devendra Singh Dhami, Kristian Kersting

NeurIPS 2025 @ San Diego

TECHNISCHE UNIVERSITÄT DARMSTADT

dfki ai

hessian.AI

AIML Lab

TU/e

KompAKI

With funding from the:

Federal Ministry of Research, Technology and Space

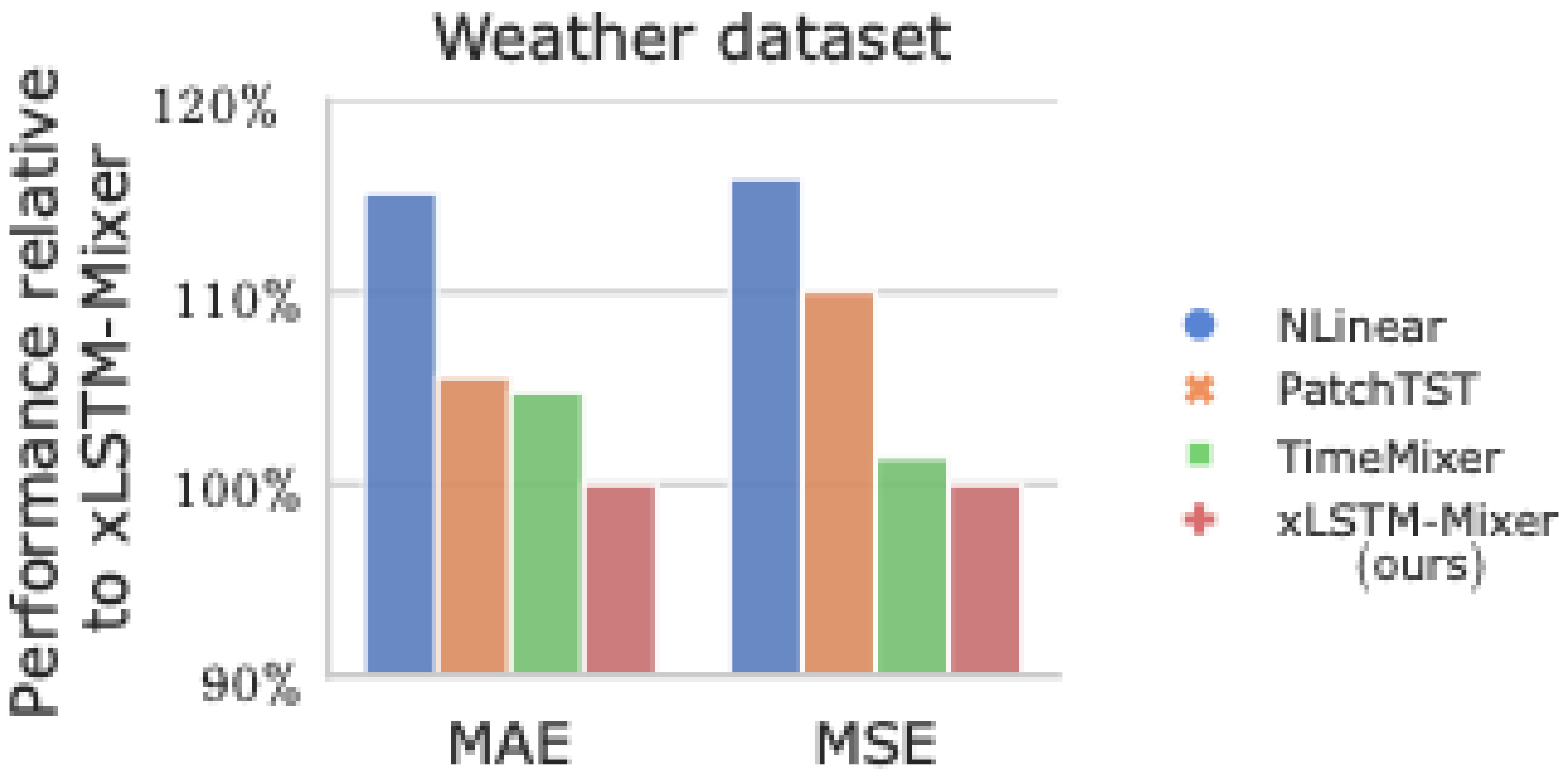
NEURAL INFORMATION PROCESSING SYSTEMS

Why Another Time Series Model?

Why Forecasting is hard

Recurrent models to the rescue?

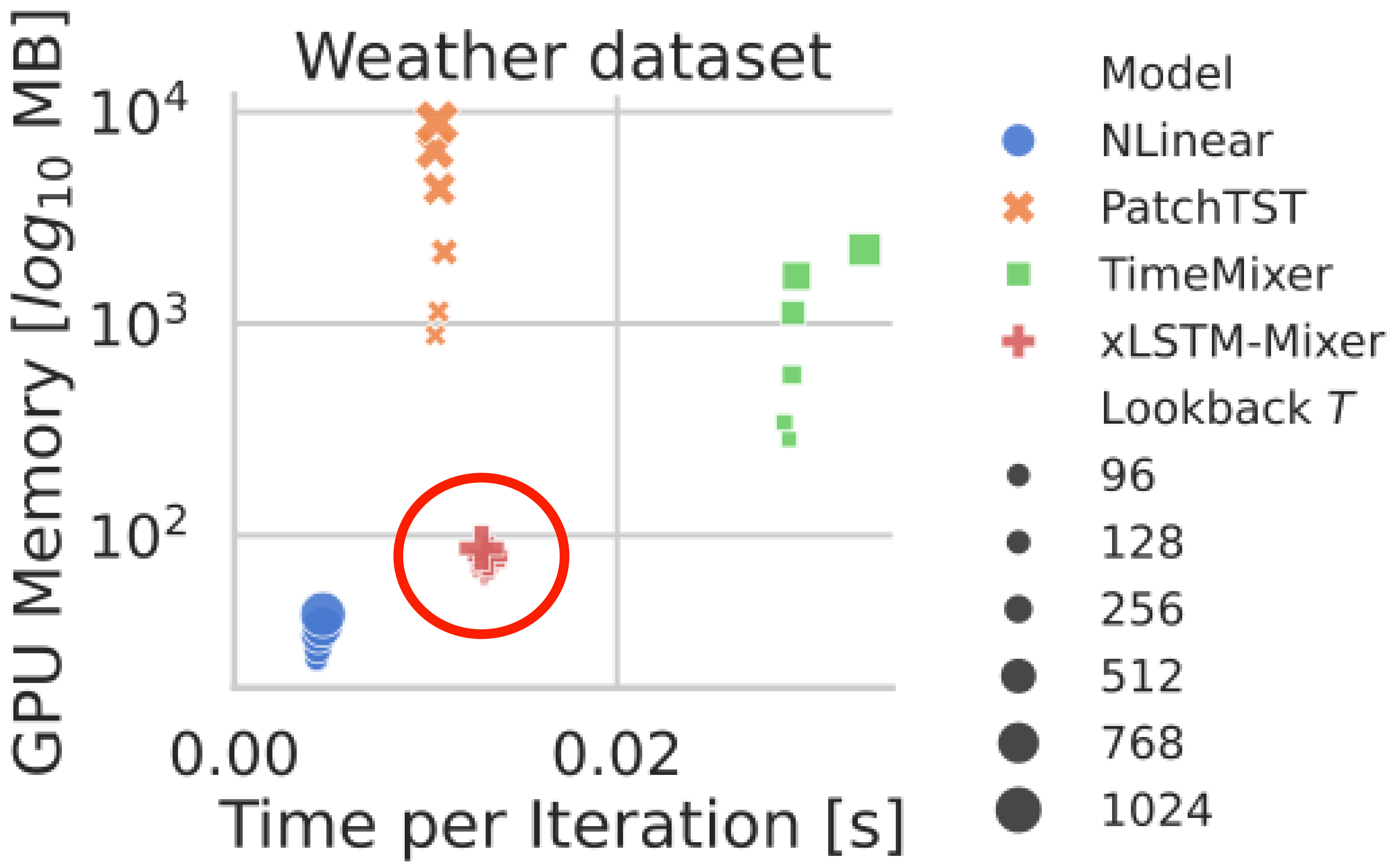
- Long-horizon, multivariate forecasting is still hard.
- Transformers are accurate but memory-heavy.
- Also MLPs don't scale well.
- We need SOTA models that run under tight compute.



Efficiency

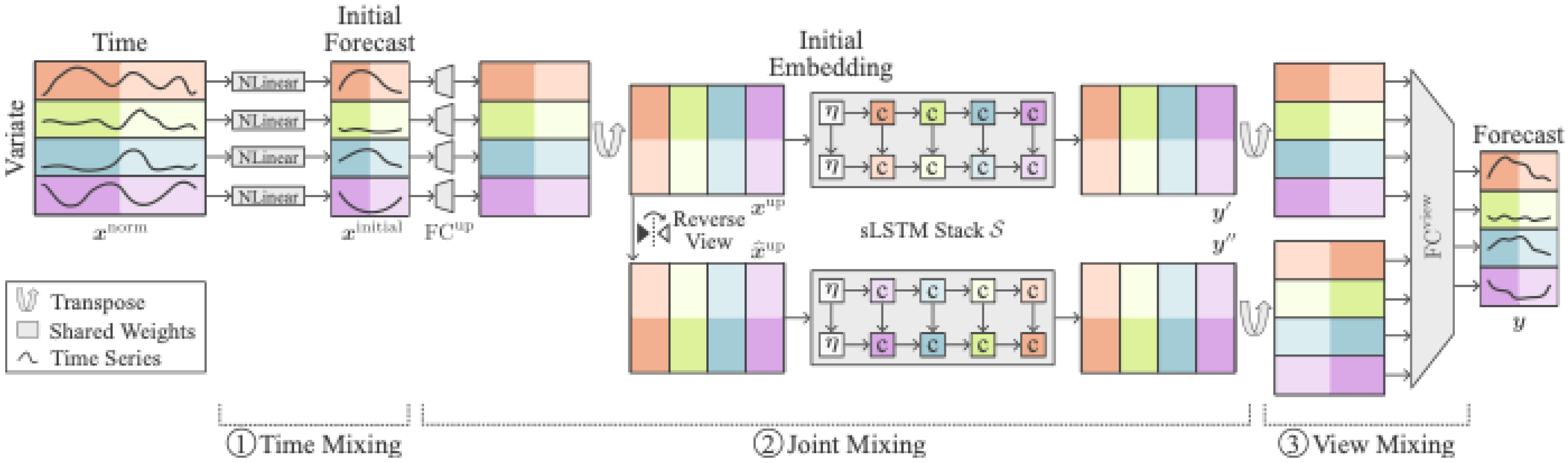
Why xLSTMs here?

- xLSTMs [1] use **scalar memories** and gating → strong sequence modeling without quadratic attention.
- **Very low GPU memory** and **competitive iteration time**.
- Fits edge / constrained deployments.



[1] Beck, Maximilian, Korbinian Pöppel, Markus Spanring, et al. “Xlstm: Extended Long Short-Term Memory.” NeurIPS, 2024.

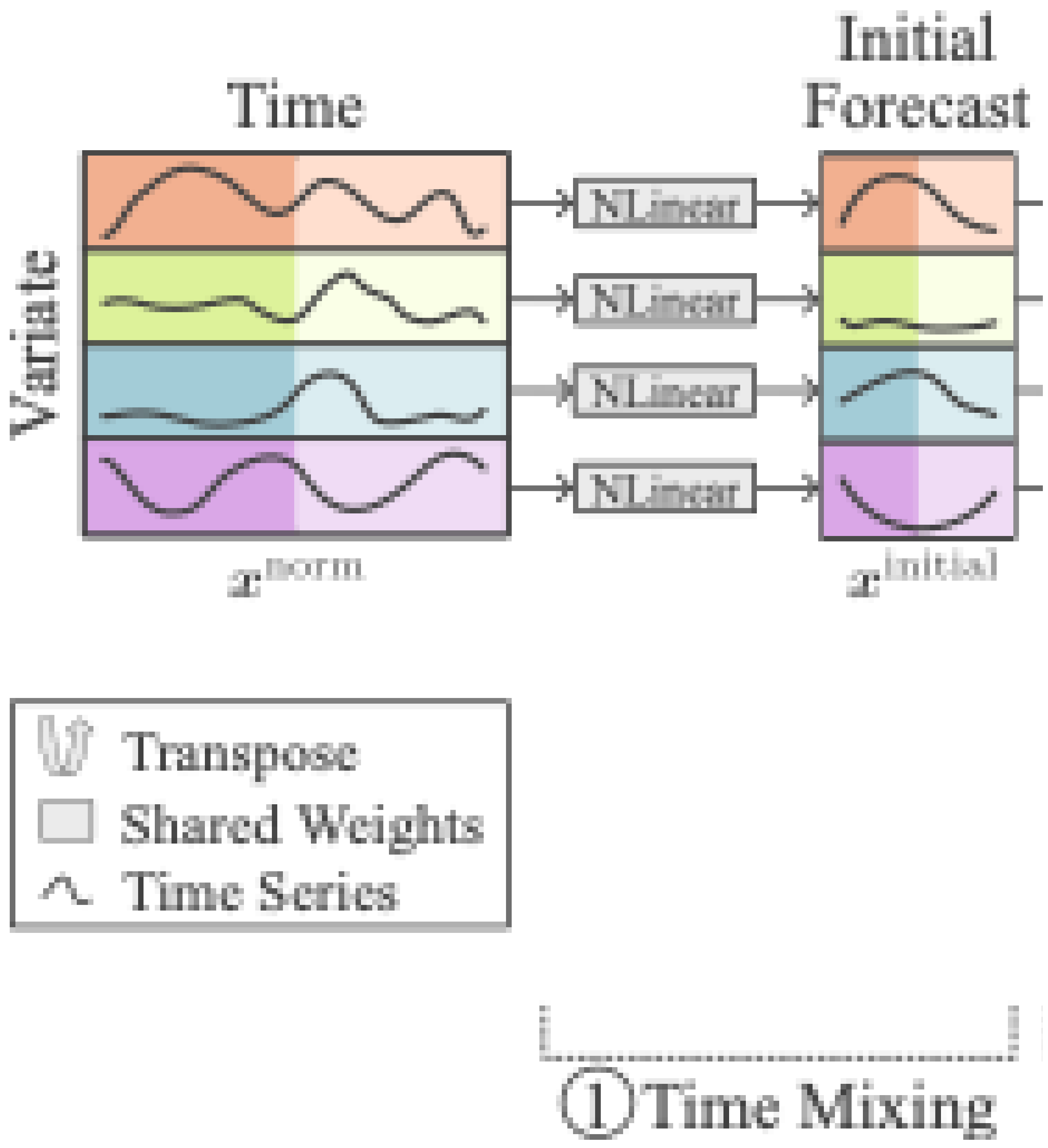
The Mixing Process



The Mixing Process

Time Mixing

- Start with a **shared linear forecast** [2] (cheap, channel-independent).



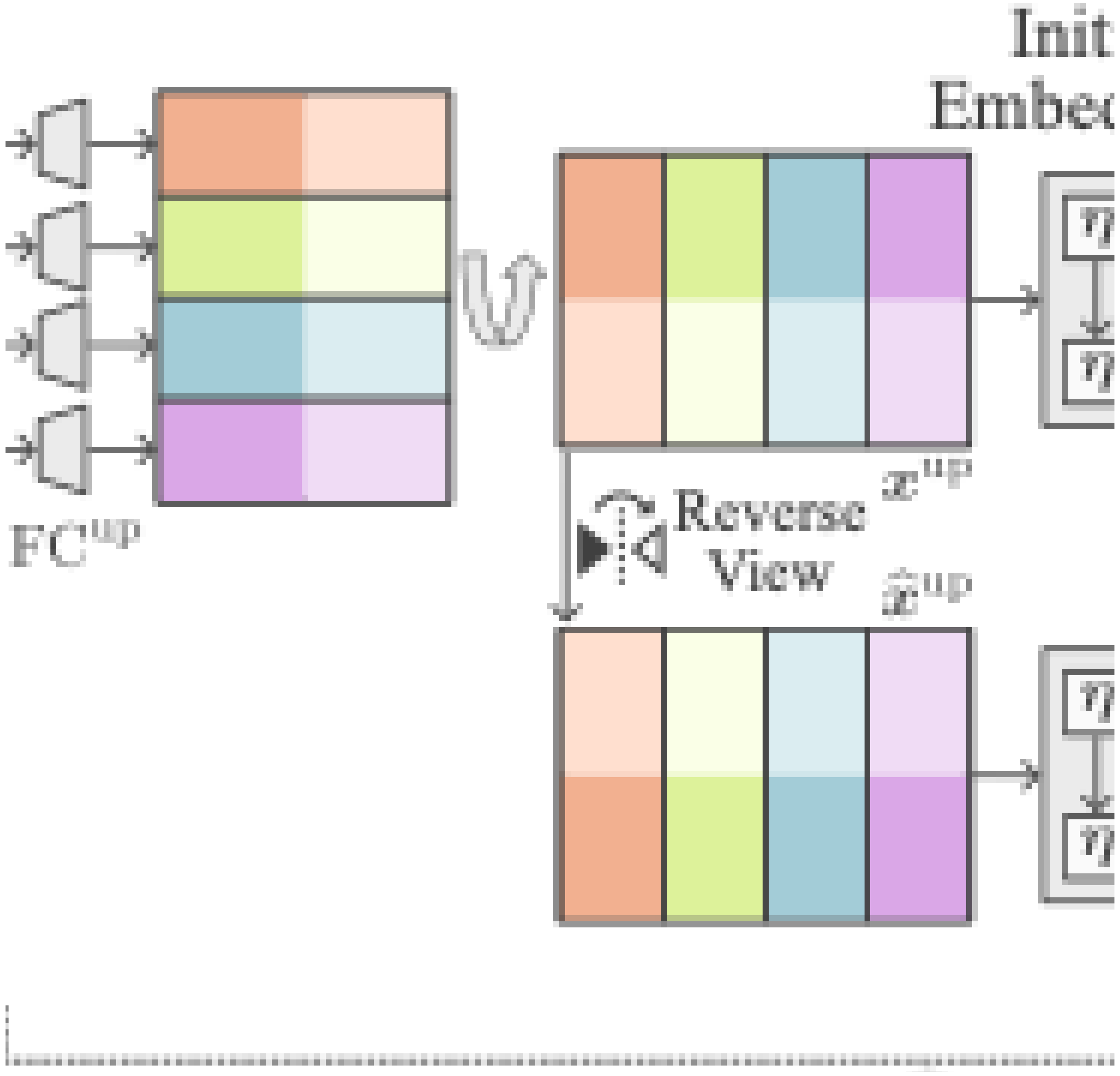
[2] Zeng, Ailing, Muxi Chen, Lei Zhang, and Qiang Xu. "Are Transformers Effective for Time Series Forecasting?" *AAAI*, 2023.

The Mixing Process

Joint Mixing

- Start with a **shared linear forecast** (cheap, channel-independent).

- **Two views** (forward + reversed) → **view mixing**.

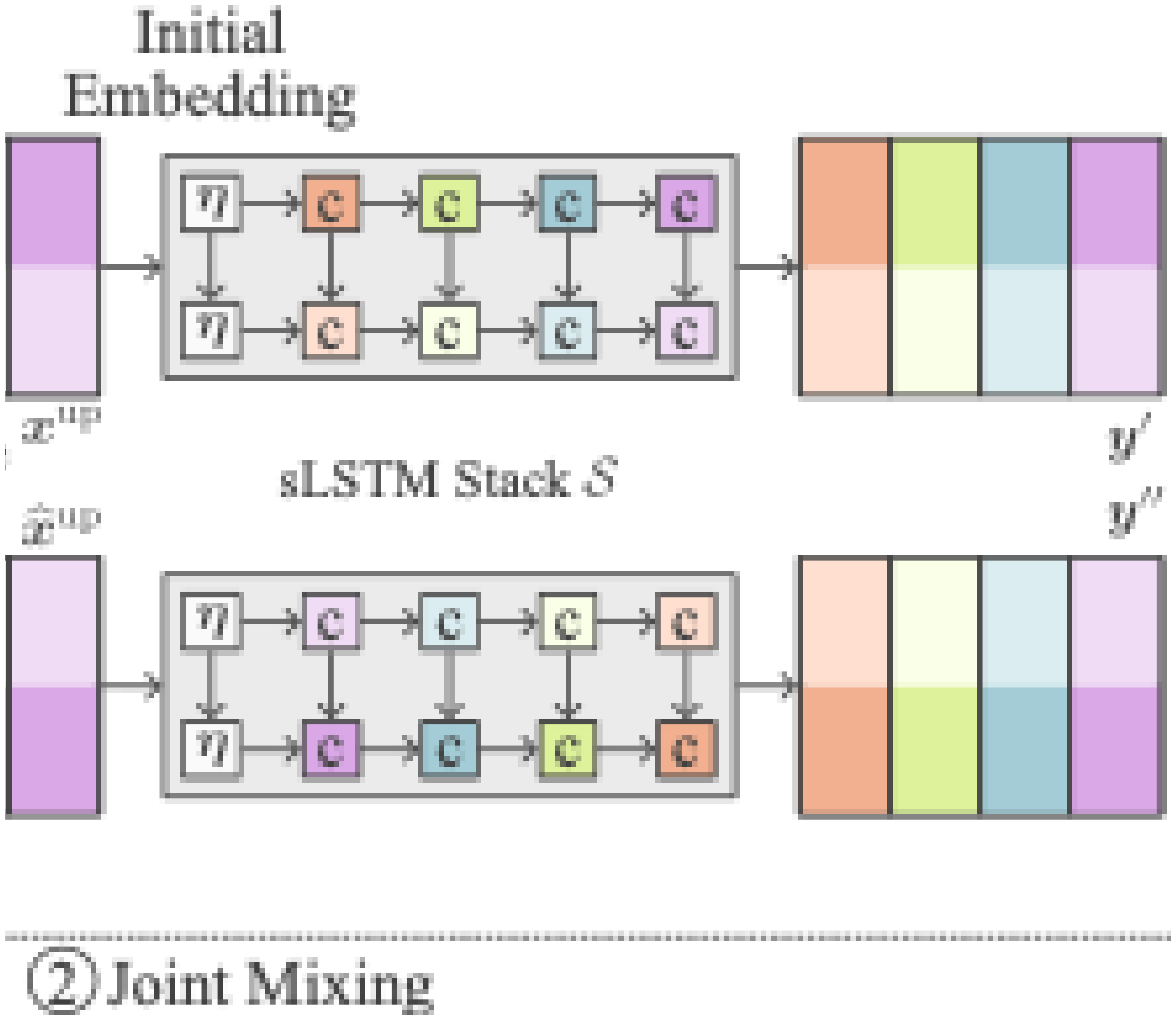


[3] Liu, Yong, Tengge Hu, Haoran Zhang, et al. "Itransformer: Inverted Transformers Are Effective for Time Series Forecasting." *ICLR*, 2023.

The Mixing Process

Joint Mixing

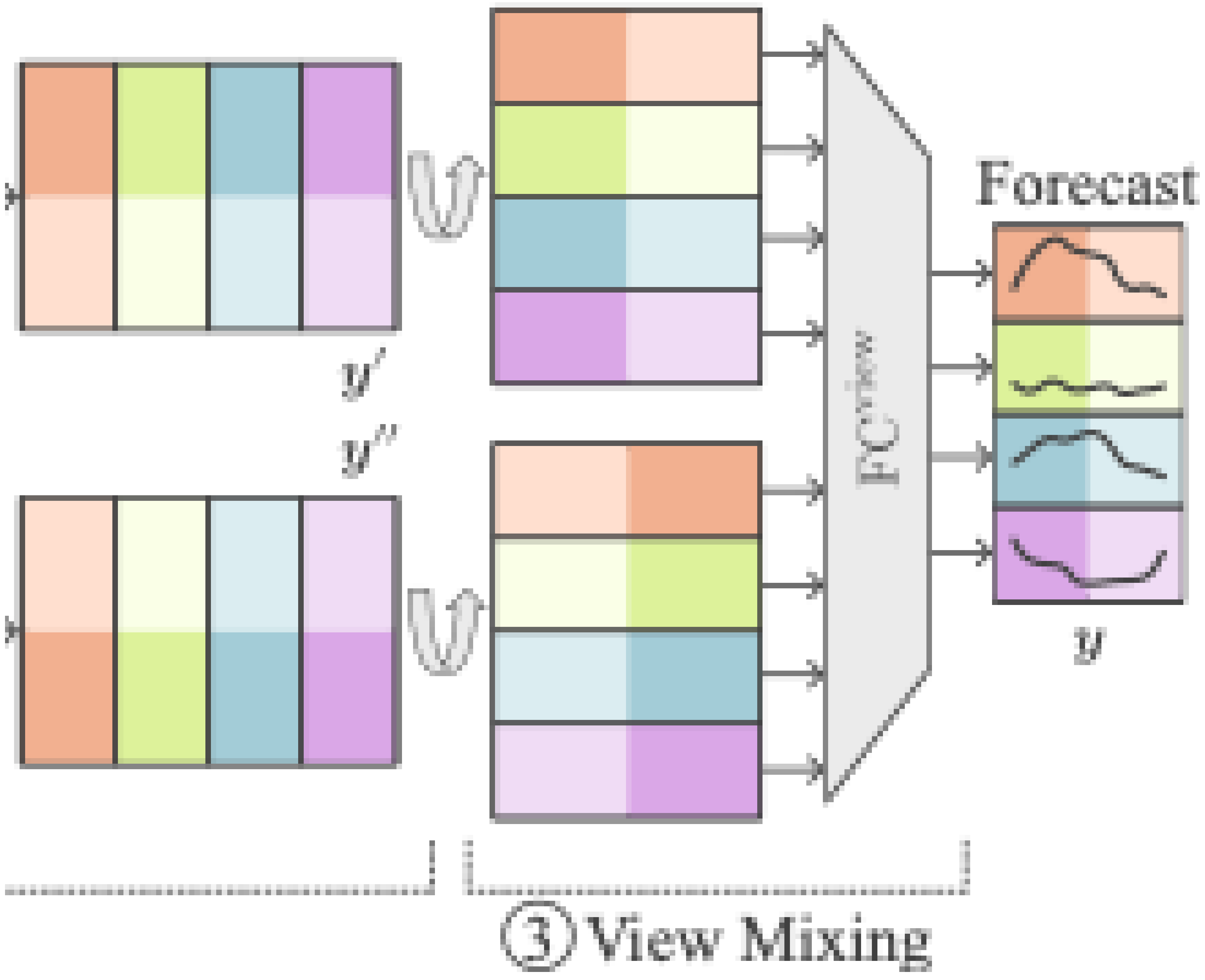
- Start with a **shared linear forecast** (cheap, channel-independent).
- **Two views** (forward + reversed) → **view mixing** → final forecast.
- **Refine** it with xLSTM block(s) that mix time + variates.



The Mixing Process

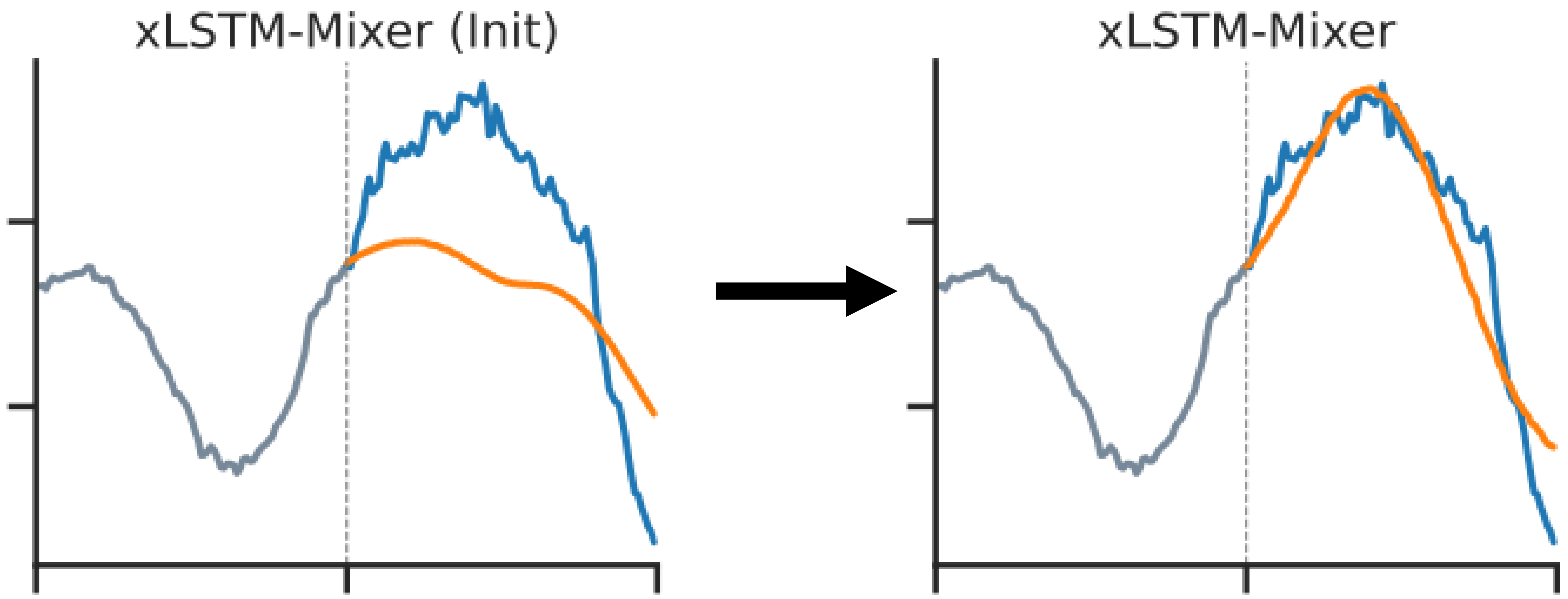
View Mixing

- Start with a **shared linear forecast** (cheap, channel-independent).
- **Two views** (forward + reversed) → **view mixing** → final forecast.
- **Refine** it with xLSTM block(s) that mix time + variates.
- **Result**: As expressive as big models yet parameter-frugal like RNNs.



Iterative refinement

- Think: **‘rough guess → smarter correction’**.
- Early stage handles what’s easy - xLSTM stages focus capacity on what’s hard.
- Multi-view mixing regularizes and reduces parameters via shared weights.



Benchmark Performance

- SOTA on standard multivariate benchmarks.
- Strong probabilistic forecasts on GIFT-Eval.
- Also works as an embedding model.

Models	Recurrent						Mixer						MLP					
	xLSTM-Mixer		xLSTMTime 2024		LSTM 1997†		TimeMix.++ 2025a		TimeMix. 2024a		TSMixer 2023c		CycleNet 2024		DLinear 2023		TiDE 2023	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	**0.219**	**0.250**	0.222	0.255	0.444	0.454	0.226	0.262	0.222	0.262	0.225	0.264	0.223	0.264	0.246	0.300	0.236	0.282
Electricity	**0.153**	**0.245**	0.157	0.250	0.559	0.549	0.165	0.253	0.156	0.246	0.160	0.256	0.156	0.251	0.166	0.264	0.159	0.257
Traffic	0.392	**0.253**	0.391	0.261	1.011	0.541	0.416	0.264	0.387	0.262	0.408	0.284	0.403	0.282	0.434	0.295	**0.356**	0.261
ETTh1	**0.397**	0.420	0.408	0.428	1.198	0.821	0.419	0.432	0.411	0.423	0.412	0.428	0.435	0.440	0.423	0.437	0.419	0.430
ETTh2	0.340	0.382	0.346	0.386	3.095	1.352	0.339	0.380	**0.316**	0.384	0.355	0.401	0.367	0.405	0.431	0.447	0.345	0.394
ETTm1	**0.339**	**0.366**	0.347	0.372	1.142	0.782	0.369	0.378	0.348	0.375	0.347	0.375	0.360	0.388	0.357	0.379	0.355	0.378
ETTm2	**0.248**	**0.307**	0.254	0.310	2.395	1.177	0.269	0.320	0.256	0.315	0.267	0.322	0.263	0.324	0.267	0.332	0.249	0.312
Wins	**11**	**16**	0	2	0	0	0	2	2	2	0	0	1	0	0	0	5	1

Model	**MASE ↓**	**CRPS ↓**	**Rank (CRPS) ↓**
TiRex	0.724	0.498	1
xLSTM-Mixer (ours)	0.780	0.510	2
TEMPO_ensemble	0.862	0.514	3
Toto_Open_Base_1.0	0.750	0.517	4
TabPFN-TS	0.771	0.544	5
YingLong_300m	0.798	0.548	6
timesfm_2_0_500m	0.758	0.550	7
YingLong_110m	0.809	0.557	8
sundial_base_128m	0.750	0.559	9
YingLong_50m	0.822	0.567	10

Take-home

- New: **3 Step-mixing** for multivariate forecasting.
- Cheap base forecast, then xLSTM refines.
- Two views regularize + reduce params.
- Delivers **SOTA accuracy with tiny memory footprint**.