N-RING: REDEFINING MULTIVARIATE TIME-SERIES FORECASTING WITH MULTIPLE DIRECTED RINGS OF LENGTH N

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time-series forecasting is widely used in energy management, traffic monitoring, finance, and environmental sensing, yet many high-performing methods rely on deep architectures, complex correction modules, or temporal downsampling, increasing computational cost or risking information loss. We propose N-ring, a compact shallow model that decomposes long-horizon forecasting into structured local components without reducing the original sampling resolution. The model represents temporal dynamics with reusable directional rings and jointly learns their length, direction, phase, and local waveform, transforming monolithic value regression into coordinated dynamical-state composition. Across 32 input-output configurations and seven mainstream baselines, including PatchTST, DLinear, and MixLinear, the 3,249–7,244-parameter N-ring model achieves first place in 20/32 MAE and 6/32 MSE comparisons, and second place in 5/32 MAE and 6/32 MSE comparisons; relative to the two best lightweight baselines, MixLinear and SparseTSF, it reduces MAE by 3.02%–11.97% and MSE by 2.46%–9.34% across multiple datasets, with average reductions of 6.40% and 5.74%, respectively. These results establish structured dynamical-state decomposition and coordinated local specialization as an effective paradigm for compact and efficient long-horizon sequence modeling.

Keywords: Multivariate Time-Series Forecasting; Structured Temporal Decomposition; Directed Temporal Dynamics; Conditional Expert Routing; Lightweight Long-Horizon Forecasting

1 INTRODUCTION

Multivariate forecasting supports energy, transportation, weather, and financial decisions, where accuracy must be balanced against deployment cost (Hyndman & Athanasopoulos, 2021; Box et al., 2015; Trindade, 2015). A central difficulty is that one predictor is often responsible for local shape, turning points, periodicity, and extrapolation simultaneously.

Deep forecasters distribute these responsibilities across attention, decomposition, frequency, channel, recurrent, probabilistic, and state-space mechanisms (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022). PatchTST, TimesNet, and iTransformer reshape the temporal or channel axes (Nie et al., 2023; Wu et al., 2023; Liu et al., 2024); gated and state-space models compress history into learned states (Hochreiter & Schmidhuber, 1997; Lai et al., 2018; Gu & Dao, 2024). Distributional and LLM-based variants add flexibility but retain substantial backbone complexity (Salinas et al., 2020; Lim et al., 2021; Jin et al., 2024).

Lightweight models replace deep encoders with linear or structured mixing (Zeng et al., 2023; Ekambaram et al., 2023; Wang et al., 2024). SparseTSF (TPAMI 2026) and MixLinear (ICLR 2026) obtain extremely small parameter counts through sparse or reduced-resolution representations (Lin et al., 2026; Ma et al., 2026), although temporal reduction can suppress short extrema or phase changes before their relevance is evaluated.

We instead compress the task. N-ringrepresents valid trajectories as directed monotonic rings, assigns local fitting and structural continuation to separate compact predictors, and retains native-resolution samples. A shallow periodic/state path supplies calendar and boundary context.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Datasets, chronological splits, and training-prefix ring statistics. ETT uses calendar-month partitions; other datasets use percentages. "Retained" follows the missing/constant filter. $N > 18$ is the tail beyond the compact bank and is processed in direction-preserving blocks.

Dataset	Channels	Sample interval	Split (train/val./test)	Train rows	Retained	Excluded const.	$N > 18$
ETTh1	7	1 h	12/4/4 months	8,640	99.13%	77	2.79%
ETTh2	7	1 h	12/4/4 months	8,640	96.99%	245	2.51%
ETTm1	7	15 min	12/4/4 months	34,560	99.55%	109	14.63%
ETTm2	7	15 min	12/4/4 months	34,560	94.29%	1,346	7.67%
Electricity	321	1 h	70/10/20%	18,412	99.50%	3,784	1.73%
Traffic	862	1 h	70/10/20%	12,280	99.77%	2,787	3.78%
Weather	21	10 min	70/10/20%	36,887	94.30%	2,108	21.92%
Exchange	8	1 day	70/10/20%	5,311	99.79%	3	47.84%

Sources: ETT and Weather are distributed with the long-term forecasting benchmark introduced by Zhou et al. (2021); Electricity is the UCI ElectricityLoadDiagrams20112014 collection (Trindade, 2015); Traffic and Exchange follow the benchmark protocol of Lai et al. (2018).

Our contributions are: 1) an original-resolution directed-ring representation with shared endpoints, exact reconstruction, and direction-preserving subdivision of long runs; **2)** a compact architecture that replaces a monolithic deep backbone with specialized low-rank ring experts, structural routing, and shallow periodic/state calibration; and **3)** a matched eight-dataset evaluation of accuracy, convergence, parameters, and runtime against Transformer, recurrent, linear, and sparse baselines.

Section 2 defines the method and protocol; Section 3 reports accuracy, convergence, ablations, and cost; Sections 4 and 5 state the empirical scope and conclusions.

2 METHOD

Figure 1 summarizes the route: directed-ring extraction, compact local experts and structural routing, periodic/state calibration, and the training protocol.

2.1 ZONE 1: DATA, RING STATISTICS, AND THE PRIMARY PATH

Datasets and preprocessing. Table 1 gives the eight datasets and chronological splits. Normalization and ring statistics use the training prefix only. Non-finite intervals and constant runs of at least eight samples are excluded; Figure 2 reports the retained ring-length distribution.

The $N > 18$ mass is small for Electricity and Traffic (1.73% and 3.78%) but substantial for Weather and Exchange (21.92% and 47.84%), motivating both the compact $N \leq 18$ bank and the direction-preserving continuation rule.

Let $\mathbf{x}_{1:L} \in \mathbb{R}^{L\times C}$ and $\mathbf{y}_{1:H}$ denote input and target. For each channel, let $\Delta_t = x_{t+1} - x_t$ and $0 = \tau_0 < \cdots < \tau_K = L - 1$ be turning points and boundaries. Ring j has direction $d_j \in \{-1, +1\}$ with $d_j\Delta_t \geq 0$ and inclusive length

$$N_j = \tau_{j+1} - \tau_j + 1, \qquad L = 1 + \sum_{j=0}^{K-1} (N_j - 1). \tag{1}$$

Neighboring rings share the turning point, hence the shortest valid ring has $N = 2$.

For a non-constant ring, define $a_j = x_{\tau_j}$, $A_j = |x_{\tau_{j+1}} - x_{\tau_j}|$, and the direction-normalized phase shape

$$s_{j,k} = \frac{d_j \left(x_{\tau_j+k} - a_j\right)}{A_j}, \qquad k = 0, \ldots, N_j - 1. \tag{2}$$

Here $s_{j,0} = 0$, $s_{j,N_j-1} = 1$, and the native samples are recovered by

$$x_{\tau_j+k} = a_j + d_j A_j s_{j,k}. \tag{3}$$

Thus $r_j = (N_j, d_j, a_j, A_j, \mathbf{s}_j)$ is exact; the learned bank approximates only $\mathbf{s}_j$.

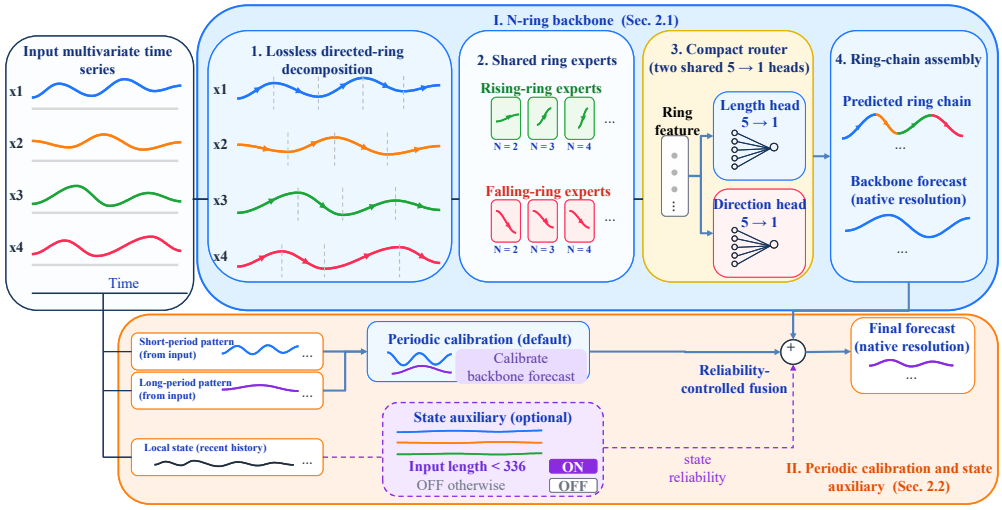


Figure 1: N-ring forecasting route. Zone 1 extracts directed rings; the primary path fits local shapes and routes their continuation; the periodic and state auxiliary paths provide repeated phase and causal boundary information; the final forecast is formed at the original resolution.

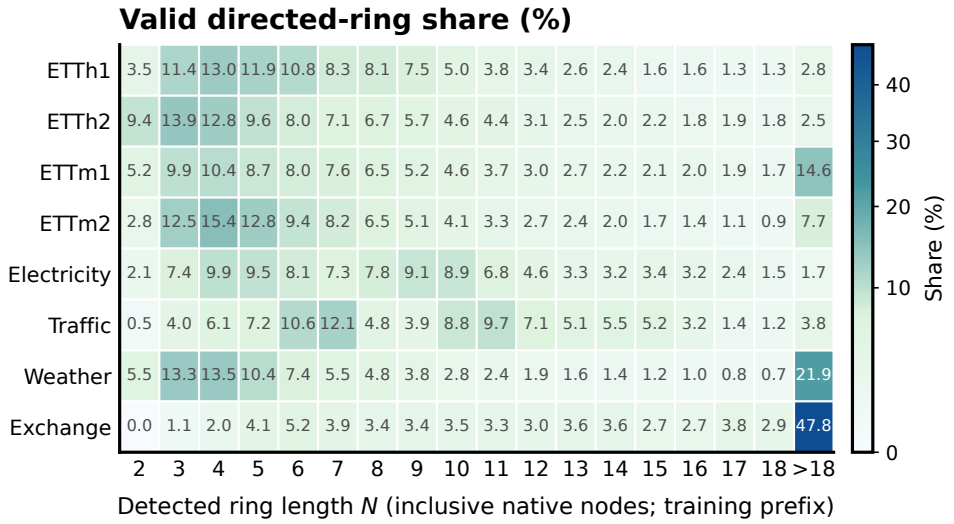


Figure 2: Per-dataset distribution of valid directed-ring lengths in the training prefixes. Consecutive rings share their turning-point endpoint, and N counts inclusive native nodes. Missing and constant intervals are excluded before normalization; the final category is the combined $N > 18$ tail.

The router uses $\mathcal{N} = \{2, \ldots, 18\}$. A physical ring with $N_{\text{phys}} > 18$ is partitioned into overlapping blocks of at most 18 nodes that share endpoints and retain direction, phase/age, and remaining length. This creates no artificial extremum and preserves the original samples; only the router category is clipped to $N = 18$. Appendix A.4 proves the block rule.

The two directions give 34 shared experts. For fixed (N, d), phase-aligned columns feed rank-R temporal and low-frequency factors:

$$\hat{\mathbf{y}}^{(N,d)} = \text{reshape}\left(\mathbf{U}^{t}_{N,d}\,\sigma(\mathbf{D}^{t}_{N,d}\mathbf{z}) + \mathbf{U}^{f}_{N,d}\,\sigma(\mathbf{D}^{f}_{N,d}\mathbf{z}_{\text{fft}})\right). \tag{4}$$

The factors are shared across channels and windows, with $R = 2$. The router factorizes structural uncertainty:

$$p(N, d \mid \mathbf{x}) = p(N \mid \mathbf{x})p(d \mid N, \mathbf{x}), \qquad \hat{\mathbf{y}} = \sum_{N\in\mathcal{N}} \sum_{d\in\{-1,+1\}} p(N, d \mid \mathbf{x})\hat{\mathbf{y}}^{(N,d)}. \tag{5}$$

Local values are fitted within short rings; continuation becomes a length, direction, and phase decision.

Let Z collect future ring structure and boundary amplitudes. The law of total variance gives

$$\mathrm{Var}(\mathbf{y} \mid \mathbf{x}) = \mathbb{E}[\mathrm{Var}(\mathbf{y} \mid Z, \mathbf{x}) \mid \mathbf{x}] + \mathrm{Var}(\mathbb{E}[\mathbf{y} \mid Z, \mathbf{x}] \mid \mathbf{x}) . \tag{6}$$

The experts model conditional local shape and the router estimates structural uncertainty through $p(Z \mid \mathbf{x})$; Appendix A gives the exact reconstruction and approximation bounds.

2.2 Periodic calibration and short-window state auxiliary

The ring bank lacks an explicit repeated calendar reference. A shallow branch receives its forecast and recent daily/weekly slices; for period P,

$$\mathbf{r}^{(P)}_{1:H} = \mathrm{repeat}\,(\mathbf{x}_{L-P+1:L})_{1:H} .$$

Two references, their forecast differences, a short trend, and phase features give

$$\hat{\mathbf{y}}_{\mathrm{cal}} = \hat{\mathbf{y}}_{\mathrm{ring}} + g_w(\mathbf{r}^{(P_w)} - \hat{\mathbf{y}}_{\mathrm{ring}}) + g_d(\mathbf{r}^{(P_d)} - \hat{\mathbf{y}}_{\mathrm{ring}}) + \delta, \tag{7}$$

where two small one-dimensional convolutions produce the gates and δ.

For $L \in \{96, 192\}$, state_aux supplements incomplete periodic context. Causal origin features encode current direction, age since reversal, dispersion, slope, and acceleration; separate training targets describe the next-boundary direction and remaining length. At inference their predictions supply a small correction, without a deep encoder or future inputs.

2.3 Parameter count and computational complexity

For ring length N, define $M_N = \lceil L/N \rceil$ and $Q_N = \lceil H/N \rceil$. The rank-R (N, d) expert contains temporal and spectral factors whose parameter count is

$$P_{N,d} = R(M_N + Q_N) + R(2F_N + Q_N) + 1, \tag{8}$$

where F_N is the number of retained real/imaginary low-frequency features. The shared backbone count is $P_{\mathrm{ring}} = \sum_{N \in \mathcal{N}, d} P_{N,d} + P_{\mathrm{router}}$. This explains why the compact model remains in the thousands of parameters even when L and H are increased: it uses a finite bank of short-ring factors rather than a full $L \times H$ dense map.

The corresponding forward cost is

$$\mathcal{O}\left(C \sum_{N \in \mathcal{N}, d} N\,[R(M_N + Q_N) + M_N \log M_N] + CLH_0 \right), \tag{9}$$

where N covers all phase positions and H_0 is the calibration width. Complexity depends on input and output through M_N, Q_N, is linear in channels, and gains $\mathcal{O}(CH)$ from calibration. Counts include every labeled component.

2.4 Experimental design

We report test MAE/MSE on chronological splits. Short runs use $L, H \in \{96, 192\}$, seeds 2025–2027, batch size 64, ten epochs, and patience 3. Long runs use the same seeds and patience; their budgets are identified in the tables. Baselines are matched local reruns unless explicitly labeled as published results, and test data never select structures or variants.

3 Results

We first test whether compact ring specialization is competitive under a matched ten-epoch budget, then ask whether it remains effective with long histories, continues to benefit from additional optimization, and justifies each calibration component and its deployment cost.

3.1 MATCHED 10-EPOCH MATRIX AND MODEL POSITIONING

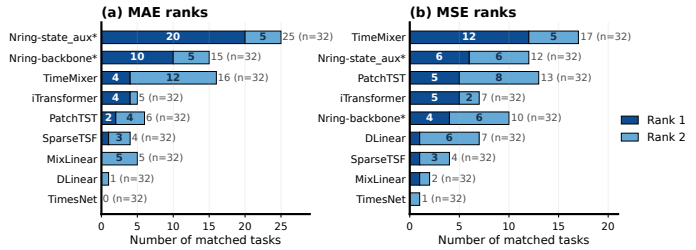


Figure 3: Rank-1 and rank-2 counts for MAE and MSE over the 32 matched tasks in Table 2. The starred backbone and state-calibrated results are two configurations of the same N-ring model family, so each is compared against the external baselines without using the other N-ring configuration as a competing method. For an external baseline, the better N-ring configuration occupies one family rank rather than two separate ranks. Dark and light segments denote Rank-1 and Rank-2, and n gives completed-cell coverage.

This experiment tests whether task decomposition alone can place a 3,249–7,244-parameter model among mainstream forecasters under the same limited optimization budget. Table 2 reports the 32 matched tasks from eight datasets and $L, H \in \{96, 192\}$; Figure 3 summarizes ranks. Nring-backbone is the local-shape model and Nring-state_aux the selected final configuration. Seven external baselines have complete matched coverage.

Nring-state_aux obtains 20/32 rank-1 and 5/32 rank-2 MAE outcomes, so it is top-two in 78.1% of the matrix; it is top-two in 12/32 MSE cells. It leads MAE throughout the four ETTm1 and four Exchange settings and is consistently strong on ETTh, whereas iTransformer, TimeMixer, and PatchTST retain clear MSE advantages on Electricity, Traffic, and Weather. The family-aware rule compares each N-ring variant with MixLinear (Ma et al., 2026), SparseTSF (Lin et al., 2026), iTransformer (Liu et al., 2024), TimeMixer (Wang et al., 2024), PatchTST (Nie et al., 2023), TimesNet (Wu et al., 2023), and DLinear (Zeng et al., 2023), without allowing the other N-ring variant to consume a rank. Thus the compact decomposition is an MAE-leading mainstream competitor, while its lower MSE rank identifies turning-point and large-error correction as the principal remaining weakness.

3.2 LONG-RANGE COMPARISON WITH COMPACT FORECASTERS

This comparison tests whether the ring formulation remains useful when a long input supplies substantially more temporal context. Table 3 compares N-ring with matched SparseTSF (Lin et al., 2026) and MixLinear (Ma et al., 2026) runs at $L = 720$ and $H \in \{96, 192, 336, 720\}$. DLinear uses the published reproduction (Song et al., 2026), with Exchange filled by our matched local rerun.

N-ring has the best MAE in 28/32 cells (87.5%) and the best MSE in 14/32. It wins all four MAE horizons on ETTm1, Electricity, Traffic, Weather, and Exchange; the four MAE losses occur only at the longest ETTh1 horizon, the two longer ETTh2 horizons, and ETTm2–720. MSE losses concentrate in ETT and Traffic, where a few displaced extrema dominate squared error. The long history therefore strengthens ring continuation without changing the model's central profile: broad absolute-error accuracy with less uniform extremum-sensitive performance.

3.3 COMPARISON WITH LIGHTWEIGHT DOWNSAMPLED BASELINES

This analysis tests whether native-resolution specialization can continue extracting information after the ten-epoch budget, rather than reaching an early ceiling like an aggressively compressed representation. Figure 4 measures post-epoch-10 validation gains for the complete three-seed $L = 720$ audit.

N-ring gains more on Electricity, Traffic, and Weather: MAE improves by 4.47%, 11.51%, and 1.18%, and MSE by 6.67%, 4.53%, and 0.71%. Gain correlates with channel count ($R^2 = 0.992/0.555$ for raw-count MAE/MSE and $0.868/0.844$ for $\log_{10} C$), a descriptive rather than causal trend. Across seeds and horizons, N-ring beats each lightweight baseline on 83/96 test-MAE runs; its MSE advantage is less uniform (67/96 versus MixLinear and 56/96 versus SparseTSF). These results support the proposed mechanism: retaining native samples leaves reusable cross-

Table 2: Dataset-level matched local ten-epoch results for every model used in Figures 3 and 5. Each entry is MAE/MSE averaged over three seeds; lower is better. The starred backbone and state-calibrated rows are two configurations of the same N-ring family; each is compared against all external baselines without using the other N-ring configuration as a competing method. N-back.* denotes Nring-backbone and N-state* denotes Nring-state_aux. Every listed model has complete coverage of the 32 dataset–configuration cells.

(a) Input $L = 96$, horizon $H = 96$

Dataset	N-back.*	N-state*	MixLinear	SparseTSF	iTrans.	TimeMixer	PatchTST	TimesNet	DLinear
ETTh1	0.400/0.393	**0.387/0.384**	0.407/0.405	0.400/0.395	0.409/0.395	0.401/0.397	0.407/0.396	0.415/0.401	0.403/0.390
ETTh2	**0.335/0.289**	**0.337/0.291**	0.349/0.310	0.347/0.305	0.349/0.302	0.346/0.297	0.341/0.293	0.360/0.317	0.365/0.308
ETTm1	0.359/0.341	**0.342**/0.326	0.372/0.353	0.377/0.360	0.369/0.336	0.356/**0.323**	0.360/0.326	0.371/0.333	0.362/0.336
ETTm2	0.258/0.183	0.257/0.181	0.272/0.189	0.267/0.185	0.268/0.188	**0.256/0.174**	0.262/0.177	0.264/0.186	0.273/0.182
Electricity	0.276/0.200	0.262/0.190	0.280/0.210	0.280/0.210	**0.241/0.149**	0.252/0.159	0.260/0.168	0.271/0.169	0.275/0.195
Traffic	0.382/0.641	0.348/0.625	0.396/0.664	0.395/0.663	**0.280/0.409**	0.303/0.490	0.287/0.450	0.315/0.603	0.386/0.648
Weather	0.221/0.194	0.214/0.187	0.241/0.199	0.240/0.200	0.218/0.177	**0.206/0.164**	0.211/0.172	0.214/0.166	0.249/0.198
Exchange	**0.197**/0.082	**0.195/0.078**	0.202/0.084	0.204/0.086	0.209/0.087	0.209/0.091	0.207/0.089	0.237/0.108	0.206/0.081

(b) Input $L = 96$, horizon $H = 192$

Dataset	N-back.*	N-state*	MixLinear	SparseTSF	iTrans.	TimeMixer	PatchTST	TimesNet	DLinear
ETTh1	0.430/0.451	**0.420**/0.443	0.424/0.440	0.420/0.436	0.440/0.448	0.428/0.444	0.433/**0.435**	0.441/0.441	0.433/0.439
ETTh2	**0.386**/0.371	**0.389**/0.376	0.396/0.389	0.397/0.389	0.405/0.391	0.390/0.368	0.391/**0.367**	0.409/0.402	0.429/0.410
ETTm1	0.382/0.387	**0.368**/0.374	0.399/0.401	0.392/0.395	0.395/0.384	0.382/**0.366**	0.387/0.372	0.392/0.376	0.383/0.377
ETTm2	0.299/0.246	0.305/0.249	0.312/0.254	0.304/0.248	0.314/0.259	**0.297/0.236**	0.307/0.247	0.306/0.254	0.325/0.252
Electricity	0.283/0.204	0.270/0.196	0.282/0.206	0.282/0.206	**0.254/0.162**	0.259/0.170	0.271/0.177	0.285/0.185	0.279/0.194
Traffic	0.373/0.611	0.340/0.608	0.369/0.612	0.370/0.611	**0.287/0.426**	0.304/0.498	0.291/0.459	0.327/0.638	0.362/0.599
Weather	0.259/0.238	0.252/0.229	0.278/0.245	0.273/0.242	0.258/0.223	**0.247/0.208**	0.254/0.221	0.262/0.224	0.283/0.237
Exchange	**0.292**/0.171	**0.285**/0.162	0.294/0.172	0.296/0.173	0.304/0.180	0.308/0.189	0.308/0.189	0.327/0.205	0.297/**0.161**

(c) Input $L = 192$, horizon $H = 96$

Dataset	N-back.*	N-state*	MixLinear	SparseTSF	iTrans.	TimeMixer	PatchTST	TimesNet	DLinear
ETTh1	0.398/0.387	**0.386**/0.384	0.390/**0.378**	0.391/0.379	0.412/0.395	0.408/0.398	0.413/0.395	0.430/0.419	0.399/0.382
ETTh2	**0.334/0.284**	**0.334/0.286**	0.347/0.301	0.347/0.302	0.346/0.295	0.348/0.294	0.358/0.313	0.370/0.323	0.351/0.292
ETTm1	0.348/0.310	**0.331**/0.302	0.356/0.321	0.353/0.318	0.356/0.312	0.343/**0.300**	0.355/0.310	0.363/0.314	0.344/0.306
ETTm2	**0.252**/0.173	**0.251**/0.172	0.270/0.184	0.258/0.174	0.263/0.178	0.252/**0.170**	0.263/0.178	0.268/0.186	0.265/0.173
Electricity	0.256/0.163	**0.228**/0.143	0.252/0.161	0.252/0.160	0.229/**0.135**	0.229/0.136	0.237/0.139	0.276/0.173	0.247/0.152
Traffic	0.317/0.469	0.261/0.469	0.298/0.456	0.297/0.456	0.276/0.391	0.276/0.415	**0.259/0.381**	0.315/0.609	0.301/0.455
Weather	0.207/0.175	**0.190**/0.158	0.237/0.192	0.233/0.189	0.217/0.172	0.200/**0.156**	0.204/0.161	0.211/0.164	0.234/0.184
Exchange	**0.197/0.082**	**0.195/0.079**	0.204/0.086	0.207/0.088	0.214/0.091	0.223/0.102	0.225/0.103	0.263/0.129	0.214/0.087

(d) Input $L = 192$, horizon $H = 192$

Dataset	N-back.*	N-state*	MixLinear	SparseTSF	iTrans.	TimeMixer	PatchTST	TimesNet	DLinear
ETTh1	0.429/0.439	0.415/0.433	0.419/0.426	**0.414/0.419**	0.444/0.449	0.424/0.431	0.444/0.449	0.462/0.476	0.423/0.423
ETTh2	**0.385**/0.362	0.393/0.377	0.391/0.369	0.392/0.375	0.403/0.382	0.388/0.358	0.392/**0.358**	0.420/0.398	0.413/0.384
ETTm1	0.370/0.351	**0.358**/0.348	0.372/0.352	0.370/0.351	0.380/0.355	0.366/**0.338**	0.386/0.360	0.398/0.387	0.365/0.342
ETTm2	**0.292/0.233**	**0.295**/0.235	0.307/0.243	0.298/0.235	0.306/0.246	0.304/0.241	0.303/0.237	0.309/0.249	0.311/0.237
Electricity	0.273/0.182	**0.242**/0.158	0.260/0.170	0.260/0.170	0.248/0.155	0.245/**0.154**	0.250/0.155	0.289/0.187	0.258/0.164
Traffic	0.342/0.504	0.268/0.487	0.300/0.463	0.299/0.463	0.284/0.411	0.286/0.432	**0.264/0.401**	0.327/0.630	0.305/0.462
Weather	0.246/0.217	**0.231**/0.200	0.273/0.236	0.269/0.232	0.255/0.217	0.238/**0.197**	0.242/0.204	0.262/0.220	0.270/0.222
Exchange	**0.292**/0.171	**0.285/0.161**	0.295/0.172	0.297/0.174	0.319/0.197	0.313/0.194	0.319/0.201	0.356/0.236	0.298/0.163

window structure for later optimization, especially in high-channel data, although the correlation does not establish channel count as its sole cause.

3.4 Ablations of periodic, state, amplitude, and routing components

This ablation identifies which shallow component supplies each gain and whether additional correction is consistently justified. Relative to the legacy calibration, prior-plus-phase reduces MAE/MSE from 0.262/0.190 to 0.248/0.162 on Electricity, from 0.350/0.624 to 0.291/0.565 on Traffic, and from 0.218/0.191 to 0.209/0.164 on Weather. It degrades ETTh2 and ETTm2, so periodic information is valuable chiefly where stable repeated structure is present. State_aux yields smaller improvements on ETTh1, ETTm1, and Exchange and complements phase conditioning on Electricity; explicit amplitude loss mainly improves Traffic MSE, while soft routing has no uniform benefit. The final design therefore retains the broadly stable state auxiliary but treats amplitude supervision and softened routing as dataset-specific options rather than default complexity.

3.5 Parameter count and deployment cost

This comparison asks whether the accuracy gains survive a deployment-level accounting of parameters and measured latency. Figure 5 and Table 5 show that Nring-state_aux reaches the best plotted

Table 3: Long-history comparison at input length $L = 720$. MixLinear (2026) (Ma et al., 2026) and SparseTSF (TPAMI 2026) (Lin et al., 2026) are local three-seed runs with batch size 64, at most 100 epochs, and patience 3. Nring is also averaged over three local seeds but uses its selected backbone checkpoint followed by a shorter periodic-calibration stage. DLinear (2023) (Zeng et al., 2023) values for seven datasets are the published single-seed reproduction under the same $L = 720$ protocol reported by Song et al. (2026); its Exchange entries are our matched local three-seed rerun with at most 100 epochs and patience 3. Entries are test MAE/MSE, and the best available MAE and MSE in each block are bolded independently.

Dataset	H	Nring	MixLinear	SparseTSF	DLinear	Dataset	H	Nring	MixLinear	SparseTSF	DLinear
ETTh1	96	**0.386**/0.373	0.402/0.369	0.389/**0.361**	0.398/0.371	Electricity	96	**0.218/0.132**	0.248/0.150	0.233/0.139	0.230/0.133
	192	**0.409**/0.410	0.414/**0.396**	0.411/0.401	0.433/0.417		192	**0.234**/0.149	0.251/0.156	0.245/0.151	0.244/**0.146**
	336	**0.424**/0.431	0.428/**0.421**	0.428/0.432	0.456/0.450		336	**0.251**/0.165	0.279/0.182	0.261/0.167	0.261/**0.162**
	720	0.469/0.519	**0.443/0.422**	0.447/0.427	0.512/0.497		720	**0.282**/0.200	0.299/0.210	0.294/0.205	0.294/**0.196**
ETTh2	96	**0.333/0.283**	0.341/0.284	0.344/0.287	0.351/0.285	Traffic	96	**0.237**/0.389	0.317/0.464	0.268/0.389	0.279/**0.389**
	192	**0.377**/0.348	0.378/0.342	0.378/**0.340**	0.453/0.440		192	**0.243**/0.412	0.325/0.478	0.270/0.399	0.280/**0.399**
	336	0.405/0.384	**0.396/0.359**	0.398/0.362	0.499/0.520		336	**0.250**/0.431	0.292/0.428	0.276/**0.411**	0.287/0.412
	720	0.425/0.398	**0.423**/0.383	0.424/**0.382**	0.648/0.834		720	**0.270**/0.470	0.308/0.460	0.297/**0.448**	0.310/0.452
ETTm1	96	**0.331/0.295**	0.377/0.338	0.357/0.316	0.350/0.307	Weather	96	**0.183/0.146**	0.236/0.178	0.225/0.170	0.226/0.168
	192	**0.356**/0.338	0.384/0.354	0.375/0.346	0.367/**0.336**		192	**0.224/0.188**	0.273/0.221	0.261/0.214	0.270/0.212
	336	**0.377/0.370**	0.409/0.393	0.391/0.373	0.393/0.372		336	**0.265/0.238**	0.307/0.267	0.297/0.260	0.308/0.257
	720	**0.410**/0.427	0.434/0.440	0.418/0.422	0.418/**0.416**		720	**0.319/0.306**	0.352/0.330	0.340/0.321	0.359/0.317
ETTm2	96	**0.244/0.160**	0.272/0.180	0.258/0.168	0.256/0.163	Exchange	96	**0.197/0.081**	0.205/0.085	0.214/0.091	0.224/0.094
	192	**0.285/0.220**	0.307/0.233	0.297/0.223	0.322/0.241		192	**0.288/0.167**	0.299/0.176	0.309/0.185	0.317/0.179
	336	**0.324**/0.277	0.340/0.283	0.328/**0.271**	0.353/0.291		336	**0.401/0.310**	0.411/0.327	0.417/0.330	0.459/0.358
	720	0.387/0.371	0.391/0.367	**0.380/0.352**	0.422/0.401		720	**0.679/0.816**	0.803/1.108	0.783/1.059	0.729/0.950

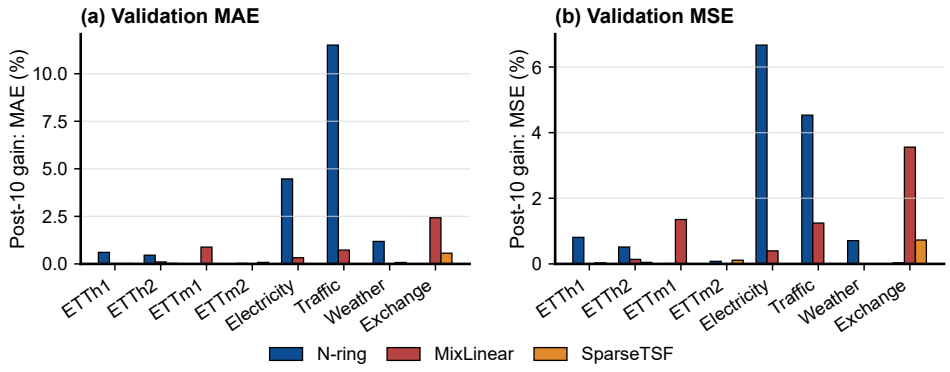


Figure 4: Dataset-level long-training benefit on the complete $L = 720$ matrix. Bars show the additional best validation-error reduction after epoch 10, averaged over four horizons and three seeds for each dataset: (a) MAE and (b) MSE. N-ring gains are largest on the high-channel Electricity, Traffic, and Weather datasets; MAE and MSE select their best epochs independently.

ETTm1 MAE and Exchange MSE with 7,244 parameters on Electricity; the backbone uses 3,249, periodic calibration 7,210, and state_aux adds only 34 parameters. Its 5,165-window test pass takes 2.09 ± 0.00 s, close to MixLinear/SparseTSF (1.91–1.94 s) and 9.6–22.7 times faster than the larger local baselines (20.13–47.44 s). Two-stage training takes 618.2 ± 7.8 s because the reported schedule trains the backbone and calibration separately, so the main deployment advantage is compact storage and inference rather than universally shortest training. The result confirms that the accuracy improvement is achieved without converting N-ring into a large or slow backbone.

Table 4: Complete three-seed 96-to-96 ablation matrix. Each entry is test MAE/MSE; the best MAE and MSE in each dataset row are bolded independently. All variants use the same frozen backbone, batch size 64, ten-epoch limit, and patience 3.

Dataset	Legacy	Prior + phase	State aux	Prior + phase + state aux	Amplitude loss	Soft route ($T = 2$)
ETTh1	0.388/0.383	0.391/**0.375**	**0.387**/0.384	0.398/0.379	0.401/0.399	0.395/0.378
ETTh2	**0.335/0.289**	0.382/0.357	0.337/0.291	0.383/0.357	0.385/0.366	0.383/0.359
ETTm1	0.343/0.327	0.345/**0.317**	**0.342**/0.326	0.346/0.319	0.346/0.320	0.348/0.319
ETTm2	**0.256**/0.181	0.287/0.214	0.257/**0.181**	0.291/0.216	0.291/0.219	0.284/0.210
Electricity	0.262/0.190	0.248/0.162	0.262/0.190	**0.247/0.160**	0.248/0.161	0.248/0.161
Traffic	0.350/0.624	0.291/0.565	0.348/0.625	0.287/0.564	**0.287/0.549**	0.288/0.564
Weather	0.218/0.191	0.209/0.164	0.214/0.187	0.207/0.162	0.208/0.165	**0.206/0.162**
Exchange	0.195/0.079	0.195/0.079	**0.195/0.078**	0.195/0.078	0.196/0.079	0.195/0.078

Table 5: Electricity 96-to-96 parameter count and locally measured runtime. Train sec. is the complete measured training time under each reported schedule; N-ring + state_aux includes 10 backbone epochs followed by 10 state-calibration epochs. Test sec. measures the complete 5,165-window test pass.

Model	Params	Train sec.	Test sec.	Model	Params	Train sec.	Test sec.
N-ring backbone	3,249	269.6	2.06	iTransformer (Liu et al., 2024)	4,833,888	175.8	20.81
N-ring + periodic	7,210	460.7	2.16	TimeMixer (Wang et al., 2024)	106,833	249.2	24.04
N-ring + state_aux	7,244	618.2	2.09	PatchTST (Nie et al., 2023)	6,903,904	656.1	37.44
MixLinear (Ma et al., 2026)	65	68.3	1.94	TimesNet (Wu et al., 2023)	150,304,769	1226.7	47.44
SparseTSF (Lin et al., 2026)	41	82.1	1.91	DLinear (Zeng et al., 2023)	18,624	138.8	20.13

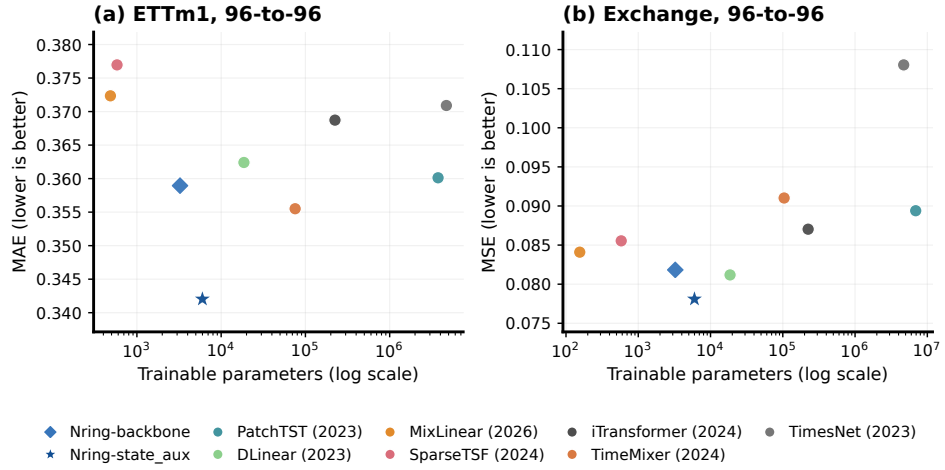


Figure 5: Accuracy–parameter trade-off for 96-to-96 forecasting: (a) ETTm1 MAE and (b) Exchange MSE. The star and diamond denote Nring-state_aux and Nring-backbone; all baselines are matched local reruns. Lower-left is better.

4 Discussion

The evidence supports task decomposition as compression: ring bookkeeping retains native samples, while learning compresses reusable shapes and structural decisions. The resulting lightweight model leads MAE and remains competitive on MSE.

Calibration remains dataset dependent: phase conditioning helps Electricity, Traffic, and Weather, while state_aux is smaller but more stable. Amplitude supervision and softer routing are not uniformly beneficial. Long-training gains concentrate on high-channel datasets, but the channel association is descriptive.

The weaker MSE ranking follows from limited extremum correction. Rank-2 experts and soft mixing can average nearby phase hypotheses: in Electricity 96-to-96, points within two steps of a turning point are 67.29% of samples but 75.74% of squared error (MSE 0.2141 versus 0.1411 elsewhere).

The first boundary is displaced by 2.59 steps on average, although 74.27% lie within two steps. N-ring omits deep residual correction and a dedicated extremum distribution; its length/direction router is two shared five-input linear heads (12 parameters) within a 3,249–7,244-parameter model. Rare, high-amplitude and phase-shifted events therefore remain difficult.

Weather and Exchange have substantial $N > 18$ tails: block reconstruction is exact, but the clipped expert category is approximate. Future work can add a compact hazard boundary model, phase-aligned extremum amplitude head, and squared-error-aware calibration. The DLinear comparison is not fully seed matched because seven datasets use a published single-seed reproduction and Exchange our local three-seed run.

5 Conclusion

N-ring decomposes original-resolution forecasting into directed-ring fitting, structural continuation, and shallow calibration. With 3,249–7,244 parameters, Nring-state_aux ranks first in 20/32 MAE and 6/32 MSE cells and second in 5/32 MAE and 6/32 MSE cells. At $L = 720$, it retains more post-ten-epoch gain than MixLinear and SparseTSF and lowers MAE in 83/96 matched runs against each. These results establish task decomposition and local specialization as an effective route to competitive lightweight forecasting; MSE-oriented extremum and long-ring modeling remain future work.

AI use statement

GPT-5.6 was used for literature retrieval and manuscript polishing.

Ethics statement

This study uses public aggregate time-series datasets and involves no human subjects, private records, or interventions. Original dataset licenses and evaluation splits are preserved.

Reproducibility statement

The anonymous artifact provides frozen N-ring inference graphs, retained public-baseline checkpoints where practical, result manifests, and one-command test verification. It covers the reported short-window variants and complete input-720 matrix over eight datasets and three seeds; larger baselines use an official-code snapshot, fixed rerun scripts, and per-seed records. The appendix documents splits, configurations, and provenance. N-ring training code is withheld during double-blind review; tables use machine-readable summaries without imputation.

References

George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time series analysis: Forecasting and control. *Wiley*, 2015.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/BF02288367.

Vijay Ekambaram, Arindam Jati, Chetan Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Tsmixer: An all-mlp architecture for time series forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 3 edition, 2021.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-llm: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations*, 2024.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *International Conference on Research and Development in Information Retrieval*, 2018.

Bryan Lim, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4): 1748–1764, 2021.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and C. L. Philip Chen. Sparsetsf: Lightweight and robust time series forecasting via sparse modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1):170–183, 2026. doi: 10.1109/TPAMI.2025.3602445.

Yong Liu, Tengge Hu, Haoran Zhang, Hang Wu, Shiyu Wang, L. Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

Aitian Ma, Dongsheng Luo, and Mo Sha. Mixlinear: Extreme low resource multivariate time series forecasting with 0.1k parameters. In *International Conference on Learning Representations*, 2026.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

Yuhao Song, Yiyuan Liu, Chang Wang, and Haiyan Li. Controlled local–global channel communication for multivariate long-term forecasting. *Mathematics*, 14(17):3132, 2026. doi: 10.3390/math14173132.

Artur Trindade. Electricityloaddiagrams20112014. *UCI Machine Learning Repository*, 2015.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *AAAI Conference on Artificial Intelligence*, 2023.

Haishi Zhou, Shangyang Zhang, Jieqi Peng, Shuenge Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI Conference on Artificial Intelligence*, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, 2022.

A Additional implementation details

A.1 Configuration-level convergence audit

Table A1 expands Figure 4 into all 32 dataset–horizon configurations. Each best epoch is the median over three seeds, while each gain is the mean relative improvement from the best validation value available by epoch 10 to the best value reached before early stopping. MAE and MSE are evaluated independently, so their selected epochs need not coincide.

Table A1: Configuration-level convergence audit for input length $L = 720$. Best epochs are medians across three seeds. Gains are mean relative reductions from the best validation value available by epoch 10 to the best value over the full run. MAE and MSE select their epochs independently.

		N-ring				MixLinear				SparseTSF			
Dataset	H	e^*_{MAE}	g_{MAE}	e^*_{MSE}	g_{MSE}	e^*_{MAE}	g_{MAE}	e^*_{MSE}	g_{MSE}	e^*_{MAE}	g_{MAE}	e^*_{MSE}	g_{MSE}
		e^*: best validation epoch; g: post-10-epoch gain (%)											
ETTh1	96	21.0	1.57	22.0	2.37	7.0	0.00	7.0	0.00	6.0	0.00	7.0	0.00
	192	14.0	0.62	17.0	0.72	10.0	0.08	10.0	0.00	7.0	0.01	8.0	0.00
	336	9.0	0.13	9.0	0.07	9.0	0.01	6.0	0.00	6.0	0.00	6.0	0.00
	720	11.0	0.12	9.0	0.08	7.0	0.00	7.0	0.01	8.0	0.08	7.0	0.13
ETTh2	96	16.0	1.07	16.0	1.19	9.0	0.02	9.0	0.09	6.0	0.12	6.0	0.19
	192	9.0	0.31	9.0	0.35	10.0	0.39	10.0	0.46	8.0	0.00	5.0	0.00
	336	9.0	0.10	9.0	0.05	8.0	0.00	8.0	0.00	8.0	0.00	8.0	0.00
	720	9.0	0.35	9.0	0.47	7.0	0.00	7.0	0.00	4.0	0.00	5.0	0.00
ETTm1	96	6.0	0.00	6.0	0.00	15.0	1.41	15.0	2.15	3.0	0.00	3.0	0.00
	192	10.0	0.03	10.0	0.05	14.0	0.78	11.0	1.19	5.0	0.00	2.0	0.00
	336	6.0	0.00	6.0	0.00	15.0	0.97	15.0	1.45	4.0	0.00	3.0	0.00
	720	9.0	0.00	8.0	0.00	9.0	0.38	9.0	0.62	5.0	0.00	4.0	0.00
ETTm2	96	10.0	0.06	9.0	0.15	6.0	0.00	5.0	0.00	5.0	0.00	5.0	0.05
	192	10.0	0.05	12.0	0.14	6.0	0.00	6.0	0.00	4.0	0.00	3.0	0.00
	336	8.0	0.00	8.0	0.00	10.0	0.00	10.0	0.00	10.0	0.04	10.0	0.09
	720	6.0	0.00	6.0	0.02	7.0	0.00	7.0	0.00	5.0	0.28	7.0	0.31
Electricity	96	40.0	3.67	40.0	4.96	12.0	0.63	10.0	0.69	9.0	0.00	7.0	0.00
	192	43.0	4.76	46.0	7.14	11.0	0.54	12.0	0.61	8.0	0.00	8.0	0.00
	336	42.0	4.88	41.0	7.60	13.0	0.09	11.0	0.28	8.0	0.00	7.0	0.00
	720	47.0	4.56	48.0	6.98	9.0	0.03	10.0	0.01	9.0	0.00	7.0	0.00
Traffic	96	75.0	11.99	73.0	4.08	8.0	0.00	13.0	1.26	7.0	0.00	8.0	0.00
	192	100.0	14.41	94.0	6.81	11.0	1.90	16.0	2.87	6.0	0.00	8.0	0.01
	336	76.0	9.86	77.0	3.86	13.0	0.35	13.0	0.41	6.0	0.00	7.0	0.00
	720	66.0	9.76	58.0	3.39	12.0	0.65	12.0	0.44	8.0	0.00	7.0	0.00
Weather	96	16.0	1.26	18.0	1.04	6.0	0.00	6.0	0.00	11.0	0.30	9.0	0.00
	192	16.0	0.83	16.0	0.70	7.0	0.00	6.0	0.00	5.0	0.00	4.0	0.00
	336	20.0	1.82	21.0	0.89	4.0	0.00	5.0	0.00	4.0	0.00	4.0	0.00
	720	19.0	0.82	12.0	0.21	7.0	0.00	7.0	0.00	8.0	0.00	8.0	0.00
Exchange	96	9.0	0.00	7.0	0.00	21.0	4.54	20.0	6.04	12.0	2.25	12.0	2.91
	192	10.0	0.03	11.0	0.13	13.0	2.37	16.0	3.85	9.0	0.00	7.0	0.00
	336	7.0	0.00	4.0	0.00	11.0	2.80	11.0	4.34	8.0	0.00	8.0	0.00
	720	4.0	0.00	6.0	0.00	3.0	0.00	3.0	0.00	1.0	0.00	1.0	0.00

A.2 Formal ring encoding and lossless reconstruction

Proposition 1 (lossless original-resolution encoding). Assume a finite channel $x_{0:T-1}$ has finite values and a fixed rule for assigning zero differences to neighboring monotone intervals. Let $\tau_0, \ldots, \tau_K$ be the resulting boundaries, let $N_j = \tau_{j+1} - \tau_j + 1$, and let each ring retain $(d_j, a_j, A_j, \mathbf{s}_j)$ from Eqs. 2–3. Then the encoder is injective on the set of sequences satisfying the detector rule, and the decoder reconstructs every native sample exactly.

Proof. The intervals $[\tau_j, \tau_{j+1}]$ cover the complete index set $\{0, \ldots, T-1\}$ and overlap only at adjacent endpoints. Equation 3 recovers every interior node of interval j. At a shared endpoint, the two reconstructions agree because

$$a_j + d_j A_j = x_{\tau_{j+1}} = a_{j+1}.$$

The number of newly introduced nodes is $N_j - 1$ for ring j, so

$$1 + \sum_{j=0}^{K-1} (N_j - 1) = T.$$

No node is omitted or counted twice. Applying the decoder to the encoded rings therefore returns x exactly. Conversely, the boundary indices are recovered from the cumulative edge counts, so two different valid sequences cannot have the same complete ring encoding. Hence the encoding is lossless and injective. □

This proposition concerns the full shape vectors $\mathbf{s}_j$. The finite expert bank, shared templates, and rank-R factors are learned approximations to this representation; they do not preserve every possible shape exactly.

A.3 Flat intervals, missing values, and detector smoothing

For a numerical tolerance ϵ, the implementation uses the sign map

$$\operatorname{sgn}_\epsilon(\Delta) = \begin{cases} +1, & \Delta > \epsilon, \\ -1, & \Delta < -\epsilon, \\ 0, & |\Delta| \le \epsilon. \end{cases}$$

Zero-sign runs are assigned to the neighboring interval according to the fixed plateau rule, or excluded from ring statistics when they are long constant runs. A constant interval has $A_j = 0$ and therefore has no direction-normalized shape; it must not be silently assigned an arbitrary direction. A missing or non-finite sample breaks the native-resolution identity and is removed from the candidate interval before statistics are computed. Consequently, the lossless proposition applies to each retained finite interval, not to an interval containing an unobserved value.

The detector may smooth first differences to avoid treating one-sample noise reversals as physical turning points. In that case, the theorem is exact for the detected partition and the observed values retained inside it; strict monotonicity of the raw samples is an empirical property. The reported monotonicity audit therefore uses the raw retained samples after the detector has chosen the boundaries rather than assuming that smoothing makes every interval strictly monotone.

A.4 Long-ring block lemma

Lemma 1 (direction-preserving finite-bank blocks). Let a physical monotone ring contain $N_{\text{phys}} > 18$ nodes and $E = N_{\text{phys}} - 1$ edges. Choose integers $q_0, \ldots, q_{M-1}$ such that $1 \le q_m \le 17$ and $\sum_m q_m = E$. Define $u_0 = \tau_j$, $u_{m+1} = u_m + q_m$, and blocks $B_m = [u_m, u_{m+1}]$. Then every block contains at most 18 nodes, adjacent blocks share one endpoint, and the union of the blocks is exactly the original physical ring.

Proof. The node count of B_m is $q_m + 1 \le 18$. Since u_{m+1} is both the final node of B_m and the first node of B_{m+1}, the blocks overlap at exactly one endpoint. Their edge intervals are disjoint and their edge counts sum to E, so their union contains the original first and last node and every

intermediate node exactly once after shared endpoints are merged. All blocks inherit the same direction because the physical ring is monotone throughout. Thus the block construction changes the expert bookkeeping but neither inserts an extremum nor discards an observed sample. □

The router's $N = 18$ category for this tail is therefore a finite approximation to a longer physical state. Exact continuation requires carrying the direction, global phase or age, and remaining length across block boundaries; treating the blocks as independent rings would invalidate the lemma's interpretation.

A.5 Low-rank shape approximation and error bound

Collect normalized phase curves of a fixed length and direction into a matrix S. Let S_R be its best rank-R approximation. The Eckart–Young–Mirsky result gives (Eckart & Young, 1936)

$$\|S - S_R\|_F^2 = \sum_{k>R} \sigma_k(S)^2,$$

where $\sigma_k(S)$ are the singular values in descending order. This identifies the empirical condition under which shared rank-factorized experts are appropriate: the tail singular-value energy must be small.

For a ring with amplitude A_j, if the normalized approximation satisfies

$$\|\mathbf{s}_j - \hat{\mathbf{s}}_j\|_\infty \le \varepsilon_j$$

and the two boundary values are exact, then

$$|x_{\tau_j+k} - \hat{x}_{\tau_j+k}| \le A_j \varepsilon_j.$$

Consequently,

$$\mathrm{MAE} \le \frac{\sum_j (N_j - 1) A_j \varepsilon_j}{\sum_j (N_j - 1)}, \tag{10}$$

$$\mathrm{MSE} \le \frac{\sum_j (N_j - 1) A_j^2 \varepsilon_j^2}{\sum_j (N_j - 1)}. \tag{11}$$

If the estimated start and amplitude have errors δa_j and δA_j, the pointwise bound gains the terms $|\delta a_j| + |\delta A_j| + |\delta A_j|\varepsilon_j$. Thus the local shape, boundary, and amplitude heads have separable contributions to the reconstruction error.

A.6 Empirical representation diagnostics

Table A2 reports the current complete oracle audit for the Electricity training prefix and its held-out chronological interval. The monotonicity values are computed from the detected raw intervals. The low-rank values are the cumulative singular-value energy of the fitted normalized-shape matrix, and the oracle errors are standardized-value MAE rather than end-to-end forecast MAE. These values are included as diagnostics of the representation; they are not used to claim that the same spectrum holds for every dataset.

A.7 Extremum and boundary error diagnostic

Table A3 isolates the peak/trough and boundary errors discussed in Section 4. The audit reuses the three completed Electricity 96-to-96 Nring-state_aux checkpoints and the held-out chronological test split; it performs no retraining or model selection.

The diagnostics show why the representation is useful but not exact after compression: direction-normalized rings are substantially easier to reconstruct than a linear baseline, while the singular-value tail remains nonzero. The finite rank-R expert is therefore a controlled approximation whose error can be measured with the bounds above.

Table A2: Empirical diagnostics of the original-resolution ring representation on Electricity.

Diagnostic	Definition	Value
Training monotonic success	retained training windows	0.7658
Held-out monotonic fraction	detected test rings	0.7619
Rank-1 shape energy	$\sum_{k\le 1}\sigma_k^2/\sum_k\sigma_k^2$	34.75%
Rank-2 shape energy	$\sum_{k\le 2}\sigma_k^2/\sum_k\sigma_k^2$	48.98%
Rank-4 shape energy	$\sum_{k\le 4}\sigma_k^2/\sum_k\sigma_k^2$	61.64%
Rank-8 shape energy	$\sum_{k\le 8}\sigma_k^2/\sum_k\sigma_k^2$	73.24%
Oracle linear reconstruction MAE	standardized test values	0.3065
Oracle shared (N, d) MAE	standardized test values	0.2495
Oracle channel-specific MAE	standardized test values	0.2045
Oracle channel low-rank-4 MAE	standardized test values	0.2174

Sources: `ring_train_split_analysis.json` and `nring_oracle_analysis.json`. The oracle audit uses the training prefix for template fitting and a later chronological interval for evaluation.

Table A3: Three-seed Electricity 96-to-96 error diagnostic for Nring-state_aux. Extrema neighborhoods contain samples within two steps of a true turning point. Values are mean $\pm$ standard deviation.

Diagnostic	Value
Global MAE / MSE	0.2616 ± 0.0002 / 0.19019 ± 0.00007
Extrema-neighborhood share of samples	67.29%
Extrema-neighborhood MAE / MSE	0.2814 ± 0.0003 / 0.2141 ± 0.0003
Non-extrema MAE / MSE	0.2209 ± 0.0001 / 0.1411 ± 0.0004
Share of total squared error near extrema	$75.74\% \pm 0.07\%$
Peak / trough amplitude MAE	0.4322 ± 0.0044 / 0.1703 ± 0.0023
First-boundary position MAE	2.59 ± 0.02 steps
First boundary within ± 2 steps	$74.27\% \pm 0.27\%$