Not All Frequencies Are Equal: Energy-Adaptive Diffusion for Time Series Forecasting

Speaker: Zining Qin
Supervisor: Dr. Chenhao Wang

ICML 2026 Accepted (Poster)

Introduction

- Context: Diffusion models have revolutionized generative.
- Current State: Growing application in time series forecasting.
- The Status Quo: Existing methods operate in the raw temporal domain.
- The Flaw: They apply uniform Gaussian noise equally across all time steps.

Motivation: The "Isotropic" Problem

SeriesContext: Diffusion model is strong but suffer in modeling Time Series

- Diffusion model have potential to act as a world model, but often fails in temporal modeling.
- Standard Diffusion adds Gaussian noise uniformly (Basically towards features).

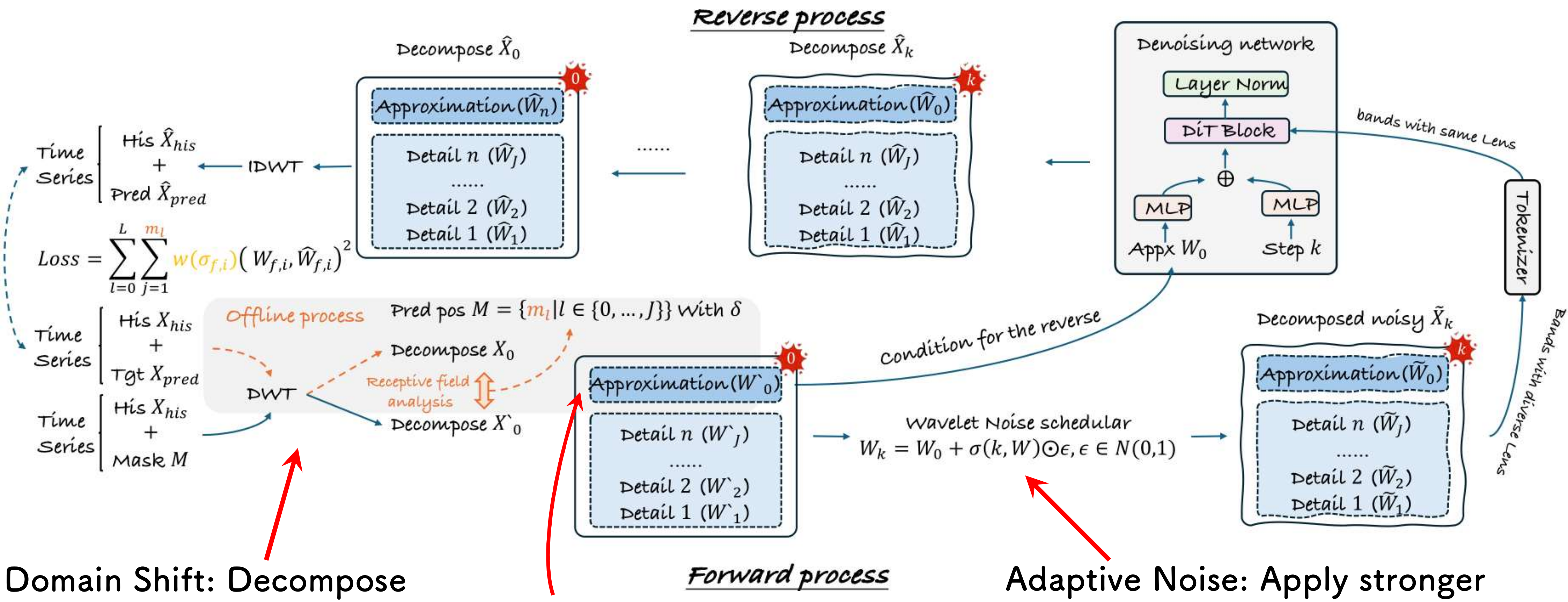
  Trend ($Amplitude \approx 100$) + Noise (1) → Signal is mostly preserved. Model learns lazily.

  Detail ($Amplitude \approx 1$) + Noise (1) → Signal is obliterated. Model learns nothing.

Our Goal

- need model diffusing "Hard features" (Trend) aggressively and
  "Fragile features" (Details) gently.

Solution: EADiff (Energy-Adaptive Diffusion)



Domain Shift: Decompose signals into frequency bands using Discrete Wavelet Transform (DWT).

Low-frequency trends act as generative conditions.

Adaptive Noise: Apply stronger perturbation to high-energy components and gentler corruption to low-energy details.

Algorithm Details: Adaptive Parameter γ

$$E_{f,\ell} = \log\left(\mathrm{Std}_t(\mathbf{W}_{f,:}^{(\ell)}) + \epsilon\right)$$

where $\mathbf{W}_{f,:}^{(\ell)} \in \mathbb{R}^{L_\ell}$ denotes the coefficients of band ℓ for feature f.

$$\tilde{E}_{f,\ell} = E_{f,\ell} - \frac{1}{J+1}\sum_{j=0}^{J} E_{f,j}$$

$$f_{f,i} = \exp\left(\gamma \cdot \tanh(\tilde{E}_{f,i}/C)\right)$$

$$\sigma_{\text{base}}(k) = \left(\sigma_{\min}^{1/\rho} + k \cdot (\sigma_{\max}^{1/\rho} - \sigma_{\min}^{1/\rho})\right)^{\rho}$$

$$\sigma_{f,i}(k, \mathbf{W}) = \sigma_{\text{base}}(k) \cdot f_{f,i}(\mathbf{W})$$

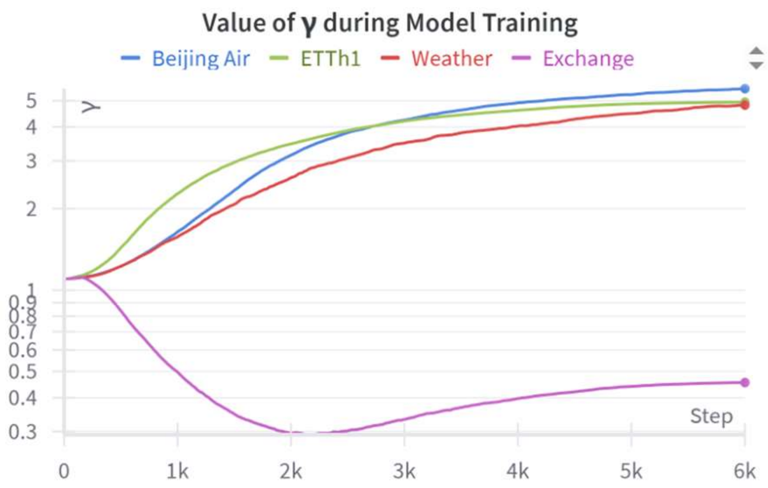


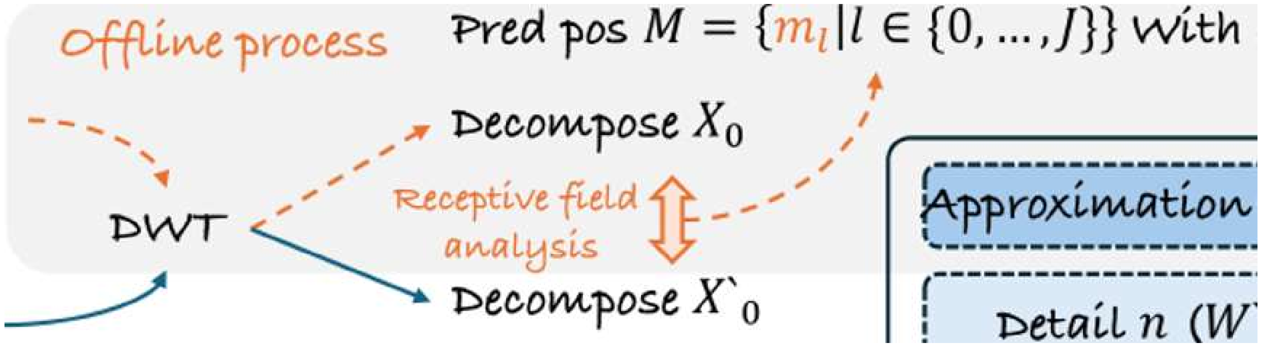
Figure 3. **Evolution of the Learnable Parameter γ during Training.** The parameter adapts differently depending on data characteristics: it increases for structured, periodic datasets (Beijing Air, ETTh1, Weather) to enhance spectral differentiation, but decreases for the highly stochastic Exchange dataset to prevent overfitting to spurious frequency energy.

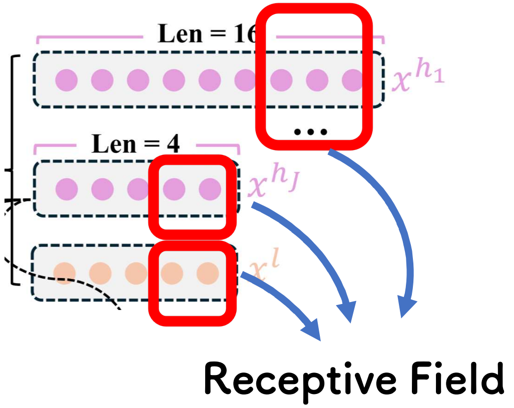
- Noise depends on both Bands and feature.
- Energy Estimation: Calculate temporal standard deviation of wavelet coefficients.
- Learnable Modulation (γ >0): adapts noise levels per-instance;
  - Amplifies noise for high-energy bands, attenuates for low-energy.
  - Unlinear scaler makes model more flexible on various datasets.
- Band-Specific Tokenization: Dense stride for trends, overlapping stride for details.

Algorithm Details: Learning Object

All Wavelet coeffs → Receptive Field only, and different noise for past and future.

- Old: model learning to reconstruct the whole Wavelet coeffs.
- New:





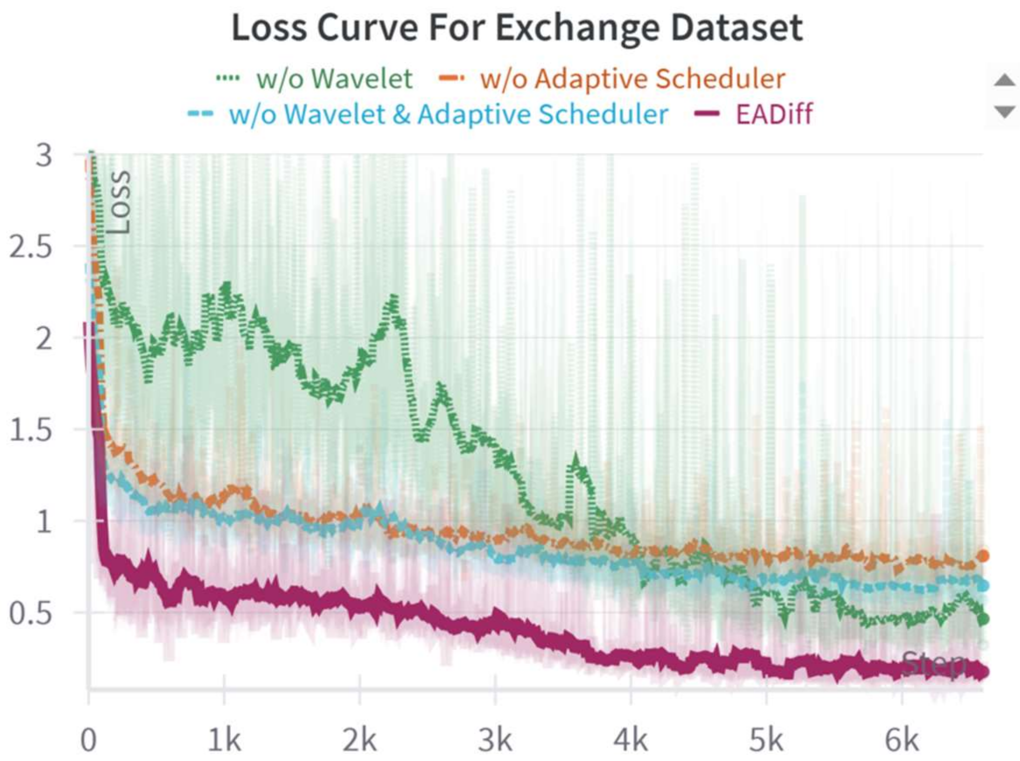
- Easier to learn as no need to reconstruct the past explicitly.
- Add noise to future: diffusion target.
- Add noise to past: Robust for autoregressive prediction. Simulate the error accumulation in AR.
- Add smaller noise to past: model can utilize such "trend" information.

Results

Table 1. **Performance on Short and Long-Term Forecasting.** We report **MAE** and **RMSE** (lower is better) under the unified autoregressive protocol. The best results are highlighted in **bold**, and the second best are underlined.

Dataset	Len	H	EADiff		TabDiff(2025)		TimeMixer(2024)		iTransformer(2024)		TabSyn(2024)		PatchTST(2023)		FEDformer(2022)		CSDI(2021)	
			MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Beijing Air	Short	48	0.2770	0.6087	0.4352	0.8132	0.4973	0.8178	0.4190	0.6566	0.7431	1.0435	0.3235	0.6385	0.4726	0.7562	**0.2712**	**0.6022**
		72	**0.2938**	**0.6255**	0.4378	0.7191	0.5396	0.8791	0.4527	0.6926	0.8803	1.1993	0.3465	0.6739	0.4820	0.7649	0.2967	0.6275
		96	**0.3207**	**0.6412**	0.4426	0.6577	0.5729	0.9179	0.4720	0.7155	1.1403	1.4456	0.3602	0.6952	0.4873	0.7699	0.3240	0.6449
	Long	192	**0.3512**	**0.6737**	0.4799	0.6959	0.6348	0.9790	0.5117	0.7627	1.7557	2.1148	0.3794	0.7262	0.4973	0.7797	0.3519	0.6837
		336	**0.3857**	0.7252	0.6132	1.0593	0.7470	1.1116	0.5361	0.7898	1.8438	2.2803	0.3913	0.7453	0.4149	**0.7152**	0.3863	0.7236
		720	0.4360	0.7986	0.6933	1.1457	1.0261	1.4395	0.5658	0.8286	1.9001	2.2515	0.4133	0.7860	**0.5081**	**0.8585**	0.4446	0.8037
Weather	Short	48	0.3561	**0.6299**	OOM	OOM	**0.3522**	0.6419	0.5087	0.8199	NAN	NAN	0.6352	0.9499	0.6110	0.9365	0.3682	0.7863
		72	**0.3979**	**0.6616**	OOM	OOM	0.3984	0.6941	0.5383	0.8561	NAN	NAN	0.7263	1.0599	0.6161	0.9414	0.3989	0.8134
		96	0.4224	**0.6843**	OOM	OOM	0.4384	0.7401	0.5546	0.8744	NAN	NAN	0.7872	1.1328	0.6187	0.9437	**0.4199**	0.8311
	Long	192	**0.4801**	**0.7266**	OOM	OOM	0.7230	1.2424	0.5890	0.9082	NAN	NAN	0.9674	1.3474	0.6224	0.9471	0.4848	0.8975
		336	**0.5374**	**0.8515**	OOM	OOM	1.4022	2.7241	0.6053	0.9231	NAN	NAN	1.0835	1.4676	0.6240	0.9484	0.5549	0.9883
		720	**0.6102**	0.9325	OOM	OOM	2.3338	6.2094	0.6167	**0.9248**	NAN	NAN	1.1788	1.5461	0.6242	0.9479	0.6553	1.1222
ETT-h1	Short	48	0.5447	**0.7537**	0.8040	0.9985	0.7185	0.9244	0.5806	0.7856	1.0308	1.2759	0.6796	0.8927	0.7635	0.9946	**0.5523**	0.8410
		72	0.5810	**0.7938**	0.7915	0.9862	0.7312	0.9418	0.6119	0.8203	0.9906	1.2252	0.7081	0.9282	0.7704	1.0017	**0.5695**	0.8587
		96	0.6183	**0.8367**	0.7735	0.9731	0.7307	0.9418	0.6372	0.8471	1.1319	1.3650	0.7311	0.9565	0.7782	1.0098	**0.5817**	0.8703
	Long	192	0.6675	0.9985	0.9497	1.1410	0.7427	0.9544	0.7127	0.9275	1.1491	1.3692	0.7885	1.0263	0.8116	1.0486	**0.6250**	**0.9125**
		336	0.7312	0.9922	0.9769	1.2056	0.7751	0.9821	0.7749	0.9963	1.2241	1.4884	0.8257	1.0704	0.8470	1.0974	**0.6738**	**0.9615**
		720	0.8375	1.0313	1.0902	1.3009	0.9517	1.2007	0.8540	1.0867	1.3107	1.5684	0.8594	1.1042	0.8709	1.1319	**0.7828**	**1.0723**
Exchange	Short	48	**0.3006**	**0.3710**	0.8135	0.8812	1.9388	2.2386	0.6362	0.7874	1.4457	1.5298	0.7570	0.9941	1.4935	1.7992	0.3444	0.5222
		72	**0.3572**	**0.4400**	0.9098	1.0009	2.1907	2.5150	0.7786	0.9547	1.4699	1.5736	0.8992	1.1982	1.4932	1.7921	0.4253	0.6317
		96	**0.4060**	**0.5004**	0.8734	0.9863	2.5175	2.9420	0.9169	1.1202	1.6232	1.7496	1.0181	1.3658	1.4792	1.7741	0.4873	0.7101
	Long	192	**0.5512**	**0.6823**	0.8796	0.9995	3.4653	4.1635	1.3058	1.5960	2.0518	2.1355	1.3499	1.7936	1.4071	1.6912	0.6205	0.8573
		336	**0.6851**	**0.8522**	1.0037	1.1285	5.2163	6.5516	1.3965	1.7334	2.2289	2.3272	1.6013	2.1144	1.2761	1.5599	0.7415	0.9972
		720	**0.8505**	**1.0534**	1.0192	1.0927	12.5190	17.7049	1.2978	1.6277	2.7411	2.7894	1.7953	2.4468	1.1122	1.3995	0.9134	1.1886

Evaluation: Ablation Study



Both **Wavelet transform** and **Adaptive Scheduler** are critical for convergence stability.

Table 2. **Ablation Results** ($H = 96$).

Method	Beijing Air MAE / RMSE	Exchange MAE / RMSE	ETTh1 MAE / RMSE	ETTh2 MAE / RMSE
EADiff (Ours)	**0.341 / 0.681**	**0.406 / 0.500**	**0.628 / 0.837**	**0.394 / 0.538**
w/o Adaptive	0.580 / 1.028	0.643 / 0.840	0.832 / 1.137	0.416 / 0.558
w/o Wavelet	0.357 / 0.682	1.056 / 1.301	0.913 / 2.742	0.497 / 0.663
w/o Wav & Adap	0.448 / 1.028	0.945 / 1.153	0.735 / 1.044	0.462 / 0.630

Evaluation: Adaptive Parameter γ

The Brain of EADiff (γ Evolution):

- Structured Data (Weather, Air quality,…):

  $\gamma \approx 5.0$, Enhances spectral differentiation.

- Stochastic Data (Exchange):

  $\gamma \approx 0.3$, Flattens to avoid overfitting to noise).

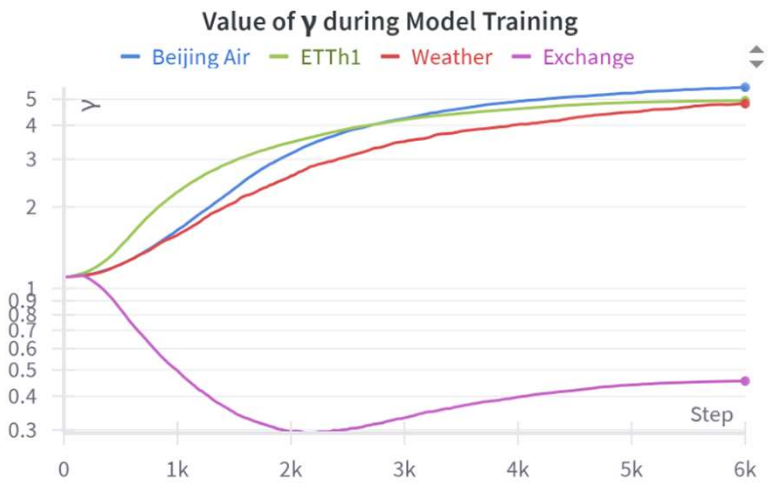


Figure 3. **Evolution of the Learnable Parameter γ during Training.** The parameter adapts differently depending on data characteristics: it increases for structured, periodic datasets (Beijing Air, ETTh1, Weather) to enhance spectral differentiation, but decreases for the highly stochastic Exchange dataset to prevent overfitting to spurious frequency energy.

Evaluation: Wavelet Hyperparameters the Paradox of Sparsity

Table 3. **Impact of Decomposition Levels** ($H = 96$). We fix the wavelet basis to `sym2` and vary the level from 1 to 5. A trade-off is observed: deep decomposition benefits non-stationary data (Exchange) by isolating trends, while shallow decomposition favors noisy data (Beijing Air, ETTh2) by preserving temporal resolution.

Dataset	Level 1		Level 2		Level 3		Level 4		Level 5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Beijing Air	**0.341**	**0.681**	0.465	0.759	0.580	0.907	0.583	0.915	0.748	1.106
ETTh1	**0.702**	**0.985**	0.825	1.119	0.749	1.105	0.919	1.366	0.800	1.031
ETTh2	**0.394**	**0.538**	0.441	0.583	0.599	0.931	0.798	1.001	0.656	0.876
Exchange	1.168	1.397	0.938	1.232	0.935	1.194	1.204	1.487	**0.406**	**0.500**

Table 4. **Impact of Wavelet Families** ($H = 96$). We compare different wavelet bases at deepest decomposition levels. `sym2` and `db1` generally outperform `bior3.1`, despite the latter having lower theoretical entropy. This highlights the importance of numerical stability and basis symmetry over pure sparsity.

Dataset	`sym2` (L5)		`db1` (L6)		`bior3.1` (L5)		`db4` (L3)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Beijing Air	0.748	1.106	0.900	1.172	4.195	6.866	**0.404**	**0.742**
ETTh1	0.800	1.031	**0.628**	**0.837**	2.896	5.732	0.875	1.136
ETTh2	0.656	0.876	**0.417**	**0.564**	1.120	1.633	0.508	0.663
Exchange	**0.406**	**0.500**	0.821	0.995	1.042	1.301	0.900	1.172

Noisier/more non-stationary the data is, higher level performs better.

How to "feel" the influence of wavelet basis?

Evaluation: Wavelet Hyperparameters the Paradox of Sparsity

Table 4. **Impact of Wavelet Families ($H = 96$).** We compare different wavelet bases at deepest decomposition levels. `sym2` and `db1` generally outperform `bior3.1`, despite the latter having lower theoretical entropy. This highlights the importance of numerical stability and basis symmetry over pure sparsity.

Dataset	`sym2` (L5)		`db1` (L6)		`bior3.1` (L5)		`db4` (L3)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Beijing Air	0.748	1.106	0.900	1.172	4.195	6.866	**0.404**	**0.742**
ETTh1	0.800	1.031	**0.628**	**0.837**	2.896	5.732	0.875	1.136
ETTh2	0.656	0.876	**0.417**	**0.564**	1.120	1.633	0.508	0.663
Exchange	**0.406**	**0.500**	0.821	0.995	1.042	1.301	0.900	1.172

Table 5. **Physical Properties vs. Model Performance ($H = 96$).** We compare `sym2` and `bior3.1` at Level 5. **Entropy**: Lower implies better theoretical compression. **Scale Ratio**: $\sigma_{max}/\sigma_{min}$, indicating energy imbalance (Lower is more numerically stable).

Dataset	**`sym2` (L5)**			**`bior3.1` (L5)**		
	Entropy ↓	Ratio ↓	RMSE ↓	Entropy ↓	Ratio ↑	RMSE ↓
ETTh1	5.33	28.9	0.800	5.15	72.3	2.896
Exchange	3.37	88.6	0.406	3.14	140.8	1.042
Beijing Air	5.29	121.5	0.748	5.28	175.8	4.195
ETTm1	5.90	36.5	0.573	5.76	72.4	0.692

- In general, more sparse → easier to learn
- However, the extreme sparsity may lead to the model loss many important details and,
- Consequently, sparse but balance in data metric could be better.

- Further hypothesis: if we arrange the points (wavelet parameters, RMSE) by Entropy in Desc., for each local minimum of RMSE, the corresponding Scale Ratio is also a local minimum of Ratios.

Conclusion

- Standard uniform noise creates a severe frequency-energy imbalance in TS diffusion.
- EADiff resolves this by operating in the wavelet domain with an energy-adaptive modulation mechanism.
- Autonomous adaptation to dataset characteristics (periodic vs. stochastic).
- Achieves robust SOTA performance without explicit model architecture design, especially in challenging long-horizon, non-stationary scenarios.

Thanks!