

# PhaseFormer: From Patches to Phases for Efficient And Effective Time Series Forecasting



Yiming Niu<sup>\*,1</sup>, Jinliang Deng<sup>\*,1,2</sup>, Yongxin Tong<sup>#,1</sup>

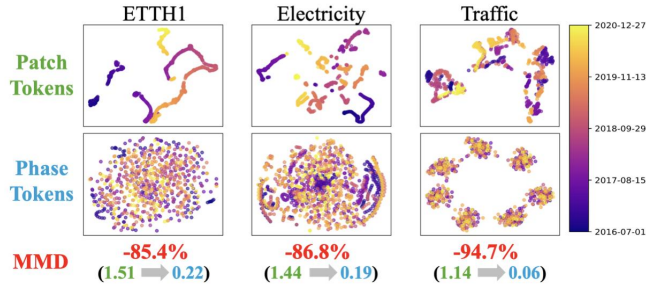
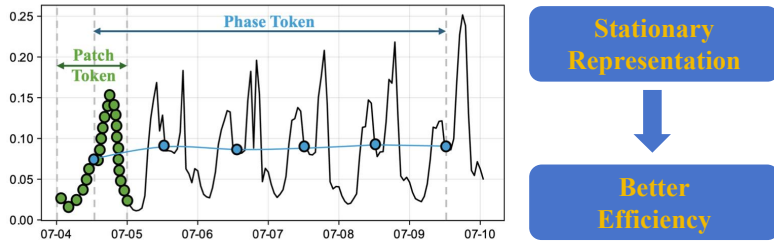
<sup>1</sup>Beihang University, <sup>2</sup>Southern University of Science and Technology

\*Equal Contribution, #Corresponding Author

## Motivation

### Key Discovery 1: Stationarity of Phase Tokens

Phase Tokens are globally more stationary compared to Patch Tokens

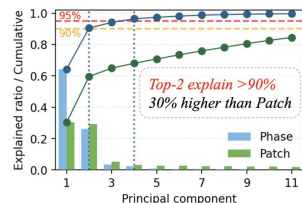


Significant MMD reductions for using Phase Tokens vs. Patch Tokens

### Key Discovery 2: Low Dimensionality of Phase Tokens

Phase Tokens are intrinsically low-dimensional, enabling concise and efficient representations

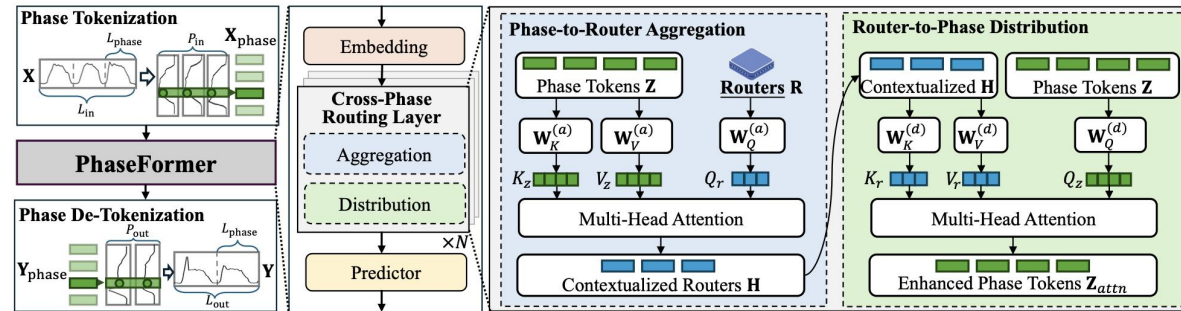
- Our PCA result shows that Phase Tokens have ~90% variance explained by **only 2 components**, vs. 11+ for Patch Tokens
- Most variance of Phase can be explained by only 1 components



## Methodology

### Overview: The PhaseFormer Architecture

A Lightweight Transformer-based Predictor with Phase Tokenization



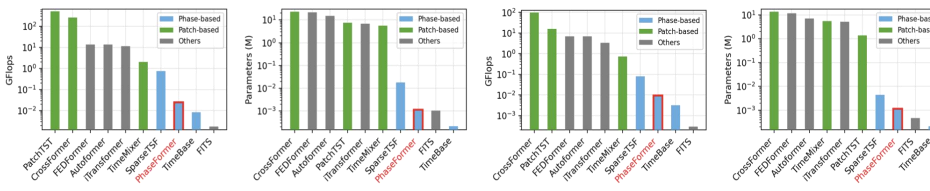
Efficient SOTA: >99.9% FLOPS & Params Reduction

## Results & Conclusions

### Main Results: High Accuracy with High Efficiency

PhaseFormer achieves SOTA accuracy on most datasets with high efficiency.

Dataset	PhaseFormer		PatchTST		iTransformer		Crossformer		FEDformer		TimeBase		SparseTSF		FITS		TimeMixer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	<b>0.403</b>	<b>0.415</b>	0.420	0.439	0.453	0.467	0.517	0.512	0.523	0.523	<b>0.404</b>	<b>0.416</b>	0.406	0.418	0.419	0.435	0.452	0.474
ETTh2	0.346	0.388	0.344	0.390	0.392	0.422	1.468	0.867	0.428	0.469	<b>0.347</b>	<b>0.397</b>	<b>0.345</b>	<b>0.383</b>	<b>0.334</b>	<b>0.382</b>	0.386	0.425
ETThm1	<b>0.346</b>	<b>0.374</b>	<b>0.354</b>	0.383	0.370	0.401	0.390	0.417	0.438	0.465	0.356	<b>0.380</b>	0.362	0.383	0.359	0.382	0.383	0.413
ETThm2	<b>0.250</b>	<b>0.313</b>	0.251	0.319	0.278	0.337	0.392	0.426	0.401	0.452	<b>0.250</b>	<b>0.314</b>	0.252	0.316	0.285	0.336	0.314	0.367
Electricity	<b>0.160</b>	<b>0.250</b>	0.169	0.265	<b>0.165</b>	0.263	0.180	0.273	0.235	0.348	0.167	<b>0.258</b>	0.168	0.263	0.172	0.270	0.171	0.273
Traffic	<b>0.386</b>	<b>0.249</b>	<b>0.394</b>	<b>0.266</b>	0.406	0.290	0.545	0.282	0.638	0.400	0.418	0.278	0.413	0.280	0.410	0.290	0.421	0.298
Weather	<b>0.223</b>	<b>0.260</b>	<b>0.223</b>	0.264	0.233	0.273	0.255	0.304	0.354	0.393	0.227	<b>0.262</b>	0.243	0.285	0.241	0.283	0.237	0.281



### Conclusions and Future Work

- Achieves state-of-the-art accuracy with only ~1k parameters, establishing a superior efficiency-effectiveness trade-off.
- Delivers over 99.9% reduction in both parameter count and computational cost (FLOPs) compared to PatchTST and Crossformer.
- We need modeling non-stationarity and complex drifts to develop more resilient phase representations.