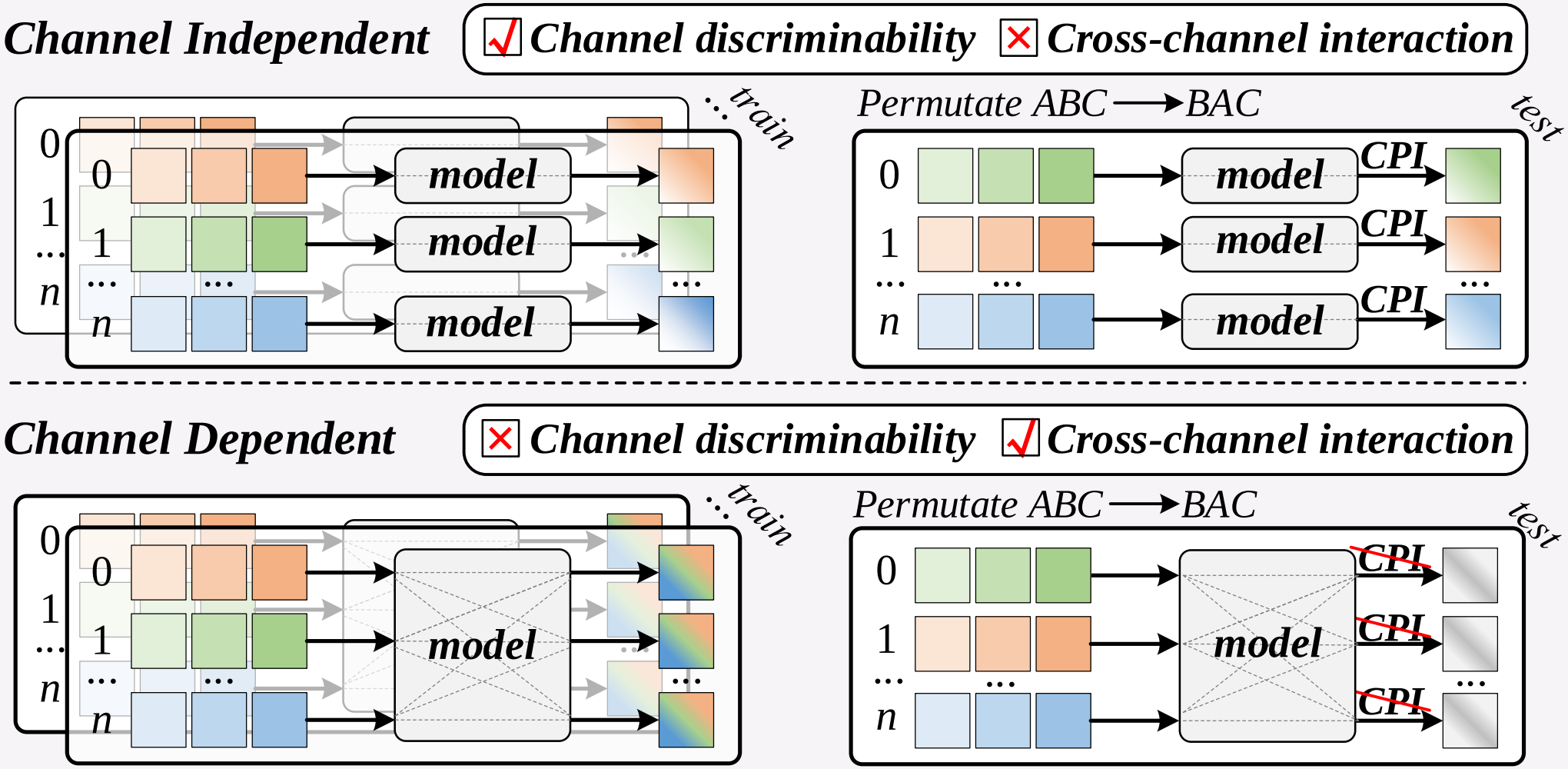
CPiRi:
Channel Permutation-Invariant Relational Interaction
for Multivariate Time Series Forecasting

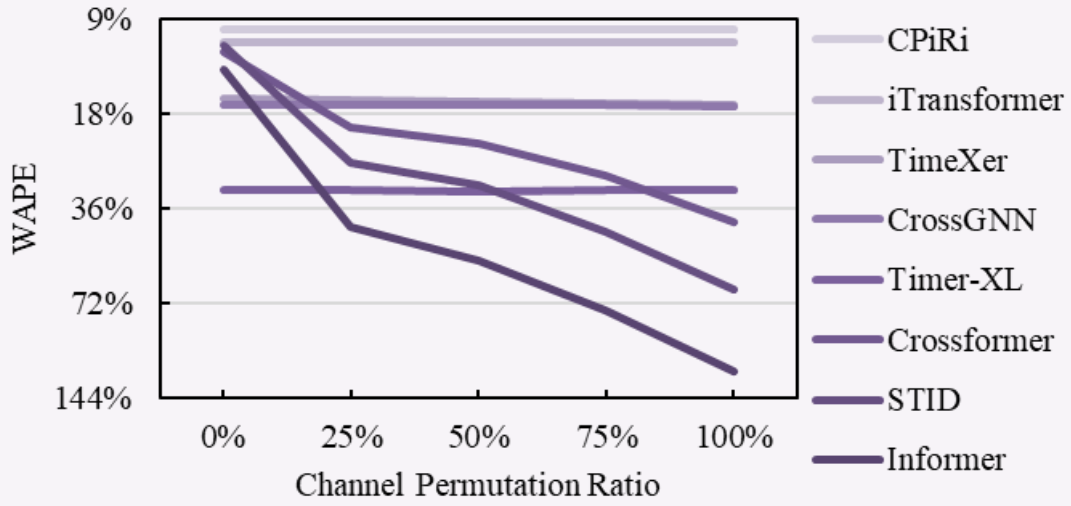
Jiyuan Xu
Zhejiang University of Finance and Economic

straka1@zufe.edu.cn

Background: The Core Dilemma in Multivariate Time Series Forecasting



Background: The Fragility of CD Models



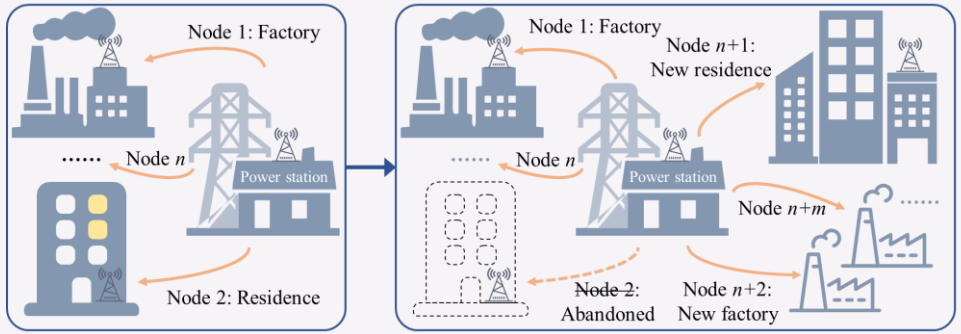


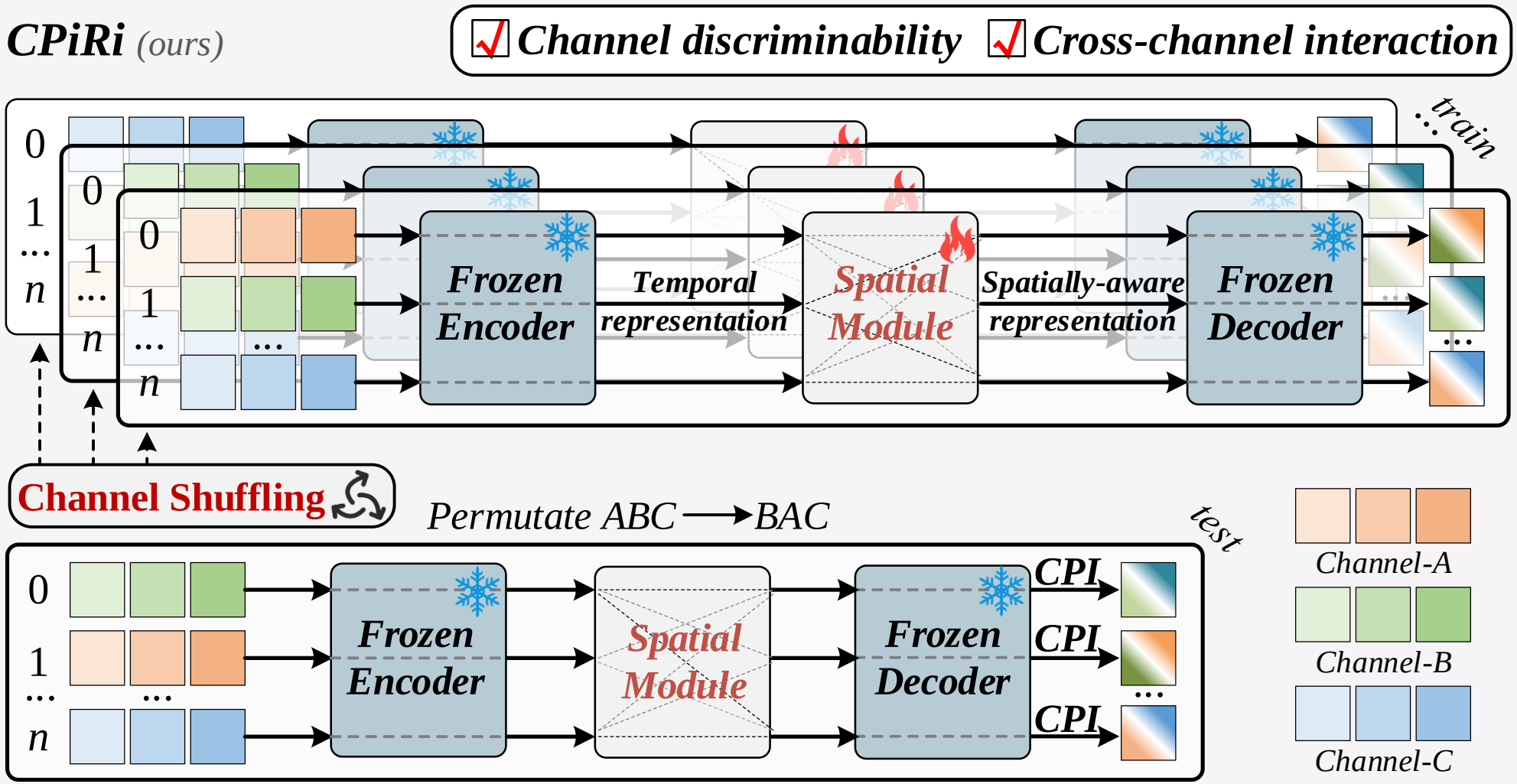
Table 2: Channel shuffling robustness analysis. For each model, we show: (1) performance when trained normally but tested with shuffled channels (Test Shuffle), and (2) performance when trained with channel shuffling (Train Shuffle). The severe degradation of most CD models under test-time shuffling reveals their dependence on fixed channel order, while CPiRi maintains stable performance in all conditions.

Model	Shuffle	METR-LA		PEMS-BAY		PEMS-04		PEMS-08		SD	
		WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓
Informer	Test	20.19%	10.38	9.99%	6.03	83.53%	150.02	118.19%	145.58	19.55%	43.10
	Train	16.48%	8.11	7.30%	4.38	49.39%	90.51	74.00%	81.12	17.38%	38.15
CrossGNN	Test	12.85%	6.53	5.20%	3.15	17.95%	37.23	17.01%	33.89	19.51%	44.00
	Train	12.78%	6.48	5.19%	3.15	17.92%	37.18	16.79%	33.78	19.49%	43.93
TimeXer	Test	13.79%	7.08	5.92%	3.57	17.22%	35.80	16.74%	33.27	18.46%	40.27
	Train	11.84%	6.09	4.86%	2.95	16.72%	34.94	15.96%	31.88	14.77%	32.38
iTransformer	Test	11.27%	5.70	4.21%	2.55	12.99%	26.79	10.70%	20.17	12.45%	27.28
	Train	11.50%	5.86	4.21%	2.55	13.02%	26.84	10.58%	19.98	12.40%	27.21
STID	Test	18.07%	9.23	7.20%	4.35	52.31%	86.25	65.18%	69.20	12.51%	26.64
	Train	10.11%	5.15	4.30%	2.60	13.75%	28.17	11.82%	21.93	12.98%	28.18
Crossformer	Test	18.06%	9.12	6.66%	4.03	43.83%	78.36	39.85%	54.72	12.50%	27.21
	Train	9.87%	4.90	4.47%	2.69	14.75%	30.38	12.82%	23.57	12.85%	27.65
CPiRi (ours)	Test	9.23%	4.67	4.02%	2.45	11.93%	24.57	10.08%	18.20	13.46%	29.21
	Train	**9.14%**	**4.62**	**3.90%**	**2.36**	**11.67%**	**23.96**	**9.43%**	**17.46**	**12.25%**	**26.85**

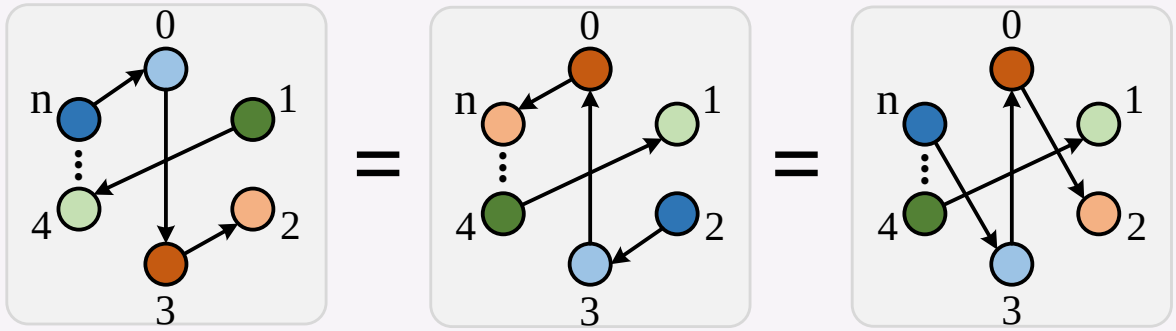
Table 3: Performance degradation under partial channel shuffling on the PEMS-08 dataset. The performance of non-invariant CD models collapses progressively as the percentage of permuted channels increases at test time, while CPiRi remains perfectly robust across all conditions. The best results are highlighted in **bold**, while the second-best results are underlined.

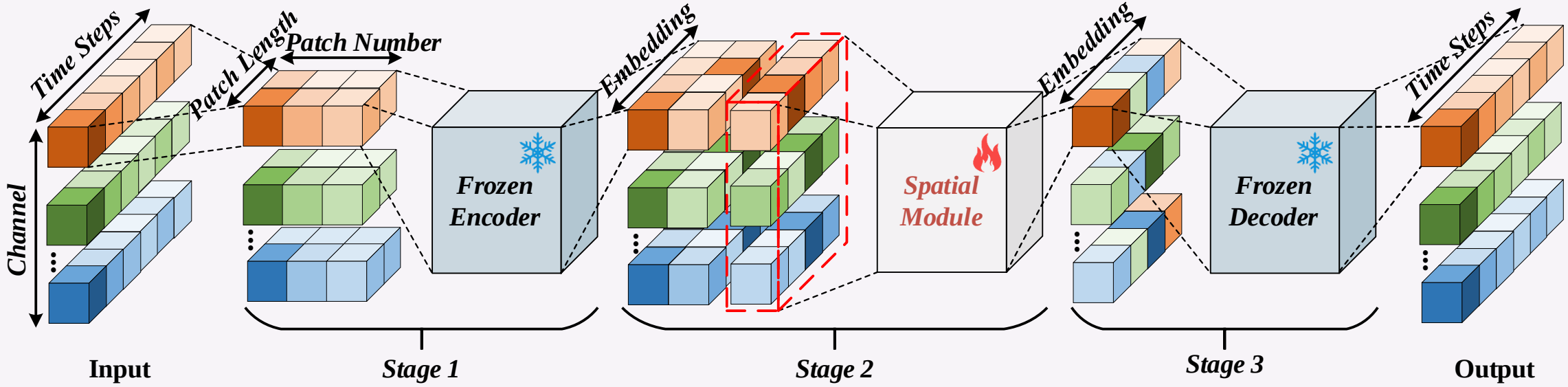
Model	100% shuffle		75% shuffle		50% shuffle		25% shuffle		0% shuffle	
	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓
Informer	118.19%	145.58	76.21%	113.26	52.73%	83.18	41.25%	59.07	13.02%	27.36
STID	65.18%	69.20	42.66%	56.13	30.17%	41.31	25.61%	32.19	10.90%	20.60
Crossformer	39.85%	54.72	28.33%	46.86	22.29%	38.41	19.90%	33.06	11.43%	22.03
Timer-XL	31.52%	68.11	31.52%	68.07	31.54%	68.11	31.53%	68.11	31.52%	68.07
CrossGNN	17.01%	33.89	16.91%	33.86	16.86%	33.84	16.89%	33.85	16.83%	33.81
TimeXer	16.74%	33.27	16.57%	32.97	16.40%	32.63	16.19%	32.06	16.02%	31.66
iTransformer	10.70%	20.17	10.70%	20.17	10.70%	20.17	10.70%	20.17	10.70%	20.17
CPiRi (ours)	**9.43%**	**17.46**	**9.43%**	**17.46**	**9.43%**	**17.46**	**9.43%**	**17.46**	**9.43%**	**17.46**

Our New Framework: CPiRi



Spatio-Temporal Decoupling Architecture and Permutation-Invariant Regularization





Key Results: Accuracy and Robustness

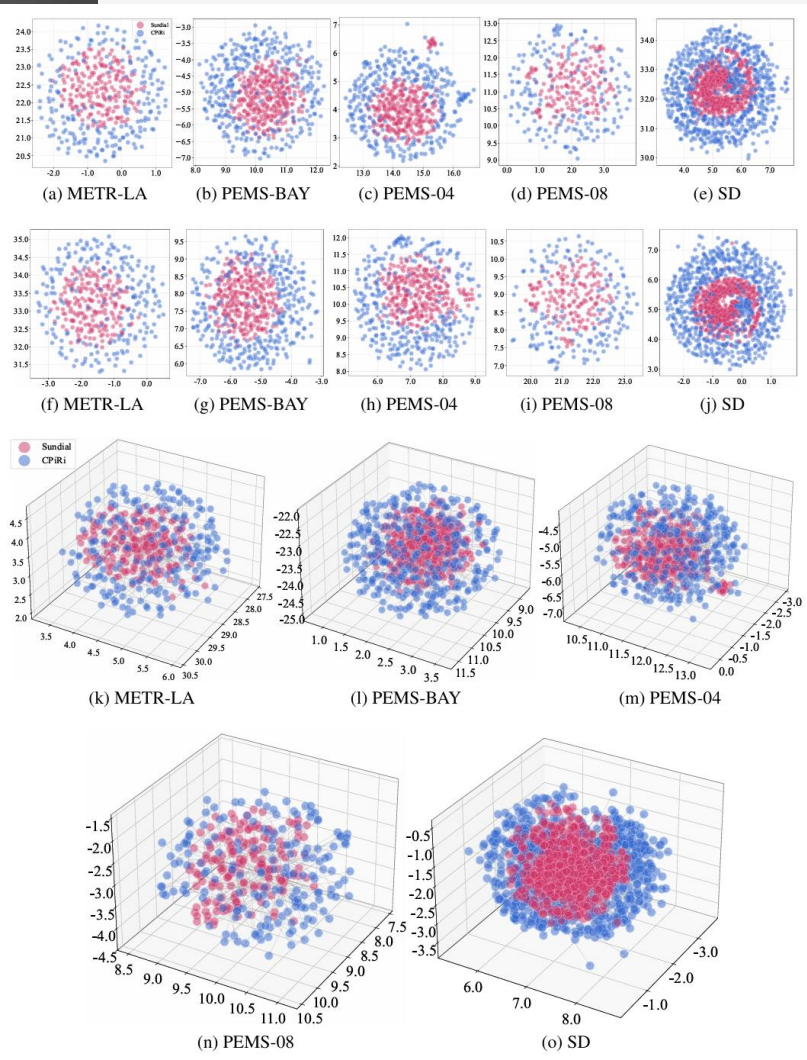
Paradigm	Model	METR-LA		PEMS-BAY		PEMS-04		PEMS-08		SD	
		WAPE	STD	WAPE	STD	WAPE	STD	WAPE	STD	WAPE	STD
CI	Chronos-Bolt*	24.19%	–	8.52%	–	34.48%	–	32.83%	–	19.71%	–
	Sundial*	16.51%	–	5.79%	–	18.77%	–	22.69%	–	24.40%	–
	Dlinear	14.93%	0.11	5.22%	0.07	17.77%	0.13	16.50%	0.12	19.46%	0.16
	PatchTST	10.51%	0.09	4.87%	0.06	15.54%	0.11	12.37%	0.09	13.41%	0.12
CD trained with **fixed** channel order	Timer-XL*	23.64%	–	8.22%	–	36.33%	–	31.52%	–	46.07%	–
	Informer	10.36%	0.09	4.47%	0.06	13.57%	0.11	13.02%	0.10	19.55%	0.17
	CrossGNN	12.82%	0.11	5.19%	0.07	17.90%	0.14	16.83%	0.13	19.51%	0.17
	TimeXer	11.18%	0.10	4.61%	0.06	16.43%	0.13	16.02%	0.12	14.43%	0.13
	iTransformer	11.28%	0.10	4.21%	0.05	12.99%	0.10	10.70%	0.08	12.45%	0.11
	STID	**8.48%**	0.07	3.91%	0.05	12.43%	0.10	10.90%	0.09	12.51%	0.11
	Crossformer	8.84%	0.08	4.07%	0.05	13.28%	0.11	11.43%	0.09	12.50%	0.11
CD **trained** with channel **shuffling**	Informer	16.48%	0.25	7.30%	0.15	49.39%	0.45	74.00%	0.60	17.38%	0.28
	CrossGNN	12.78%	0.18	5.19%	0.10	17.92%	0.25	16.79%	0.24	19.49%	0.28
	TimeXer	11.84%	0.17	4.86%	0.09	16.72%	0.23	15.96%	0.22	14.77%	0.21
	iTransformer	11.50%	0.16	4.21%	0.08	13.02%	0.18	10.58%	0.15	12.40%	0.17
	STID	10.11%	0.15	4.30%	0.08	13.75%	0.19	11.82%	0.17	12.98%	0.18
	Crossformer	9.87%	0.14	4.47%	0.08	14.75%	0.20	12.82%	0.18	12.85%	0.18
CD **tested** with channel **shuffling**	Informer	20.19%	0.30	9.99%	0.20	83.53%	0.70	118.19%	0.90	19.55%	0.35
	CrossGNN	12.85%	0.20	5.20%	0.12	17.95%	0.28	17.01%	0.26	19.51%	0.30
	TimeXer	13.79%	0.22	5.92%	0.14	17.22%	0.27	16.74%	0.25	18.46%	0.29
	iTransformer	11.27%	0.18	4.21%	0.10	12.99%	0.20	10.70%	0.17	12.45%	0.19
	STID	18.07%	0.28	7.20%	0.16	52.31%	0.50	65.18%	0.55	12.51%	0.20
	Crossformer	18.06%	0.28	6.66%	0.15	43.83%	0.45	39.85%	0.40	12.50%	0.20
CI+CD	**CPiRi** (ours)	9.14%	0.08	**3.90%**	0.05	**11.67%**	0.09	**9.43%**	0.07	**12.25%**	0.11

Key Results: Inductive Generalization and Efficiency

Variant	Channels	METR-LA		PEMS-BAY		PEMS-04		PEMS-08		SD	
		WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓	WAPE↓	MAE↓
w/ strategy	w/o 75%	9.35%	4.73	4.10%	2.48	12.07%	24.72	10.72%	19.57	13.40%	28.96
	w/o 50%	9.29%	4.71	3.99%	2.42	12.06%	24.70	10.70%	18.93	13.01%	28.01
	w/o 25%	9.20%	4.66	3.96%	2.40	12.01%	24.62	10.12%	18.19	12.96%	28.00
	–	9.14%	4.62	3.90%	2.36	11.67%	23.96	9.43%	17.46	12.25%	26.85
w/o strategy	w/o 75%	10.27%	5.20	4.52%	2.73	15.27%	30.54	14.22%	25.86	16.91%	35.37
	w/o 50%	9.55%	4.82	4.38%	2.65	13.12%	26.65	11.76%	20.64	15.97%	34.16
	w/o 25%	9.33%	4.70	4.16%	2.51	12.59%	25.60	10.91%	19.51	14.32%	30.95
	–	9.23%	4.67	4.02%	2.45	11.93%	24.57	10.08%	18.20	13.46%	29.21

Paradigm	Model	Maximum batch size	Inference time (s)		Avg. time (s)		GPU memory (GB)		Avg. memory (GB)		Complexity
			Base	Compiled	Base	Compiled	Base	Compiled	Base	Compiled	
CI	Sundial	4	2.55	1.61	0.64	0.40	54.44	20.68	13.61	5.17	$O(T^2)$
CI+CD	**CPiRi** (ours)	4	2.66	1.62	0.67	0.41	54.56	32.00	13.64	8.00	$O(T^2 + C^2)$
CD	Timer-XL	1	OOM	1.07	–	1.07	>80	75.68	–	75.68	$O((T \times C)^2)$
CD	iTransformer	2	0.52	0.40	0.26	0.20	45.18	35.37	22.59	17.69	$O((T \times C)^2)$

Key Results: Inductive Generalization and Efficiency



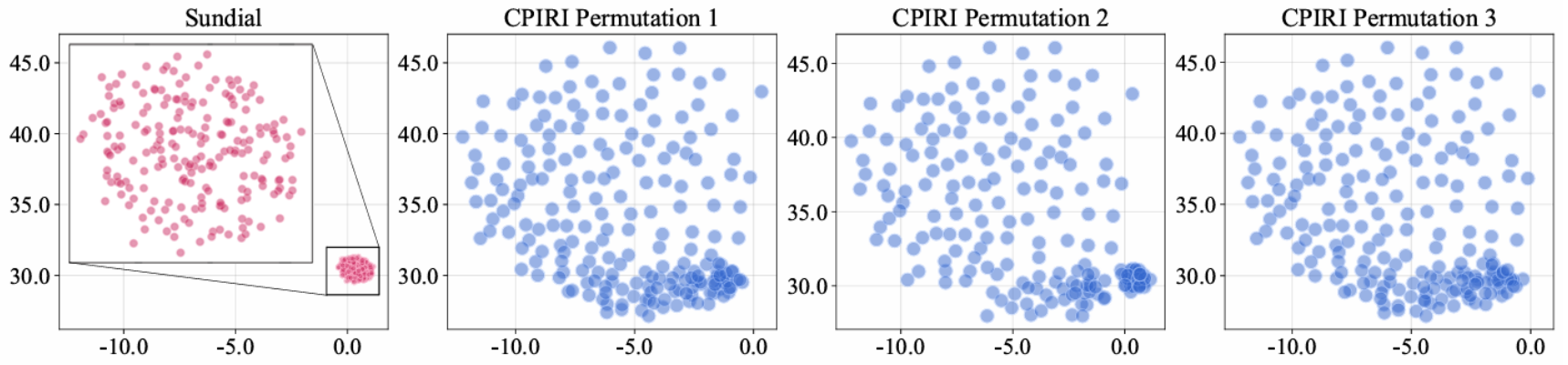


Figure 5: UMAP visualization of channel representations on METR-LA. Left: Sundial. Right: CPiRi under three random channel permutations, near-identical geometry with clearer separation.

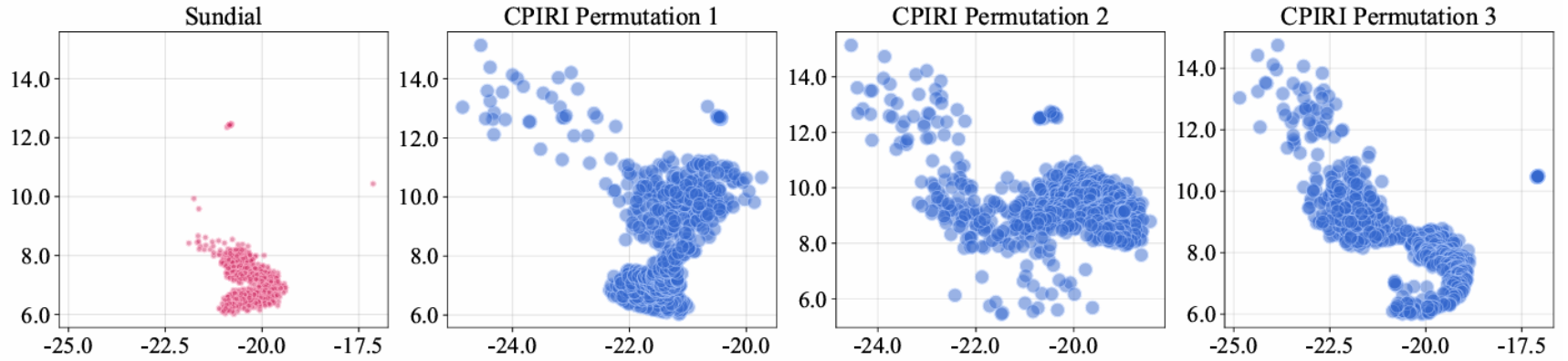


Figure 6: UMAP visualization of channel representations on METR-LA after fine-tuning CPiRi, embeddings become more compact and less separable.

Thanks

https://openreview.net/pdf?id=tgnXCCjKE3

https://github.com/JasonStraka/CPiRi

Jiyuan Xu Zhejiang University of Finance and Economic