6 Appendix

In this appendix, we examine the extrapolation ability of **HGRN** and provide the training and inference speed comparison of **HGRN** and existing efficient sequence modeling methods. We also illustrate the forget rates of each layer on a trained language model of **HGRN**.

6.1 Extrapolation test

In this section, we tested **HGRN** 's extrapolation ability by directly inferring the model with a variety of sequence lengths. As shown in Table 9, our method has the ability to train short and test long.

6.2 Speed comparison

In this section, we benchmark the speed of our method on the LRA benchmark. Our method achieves state-of-the-art training and inference speed.

6.3 Visualization

In this section, we visualize the forget rates (decay rates) of each layer on a model trained on language modeling tasks.

Figure 4: Visualization forget rates (decay rates) in each layer.

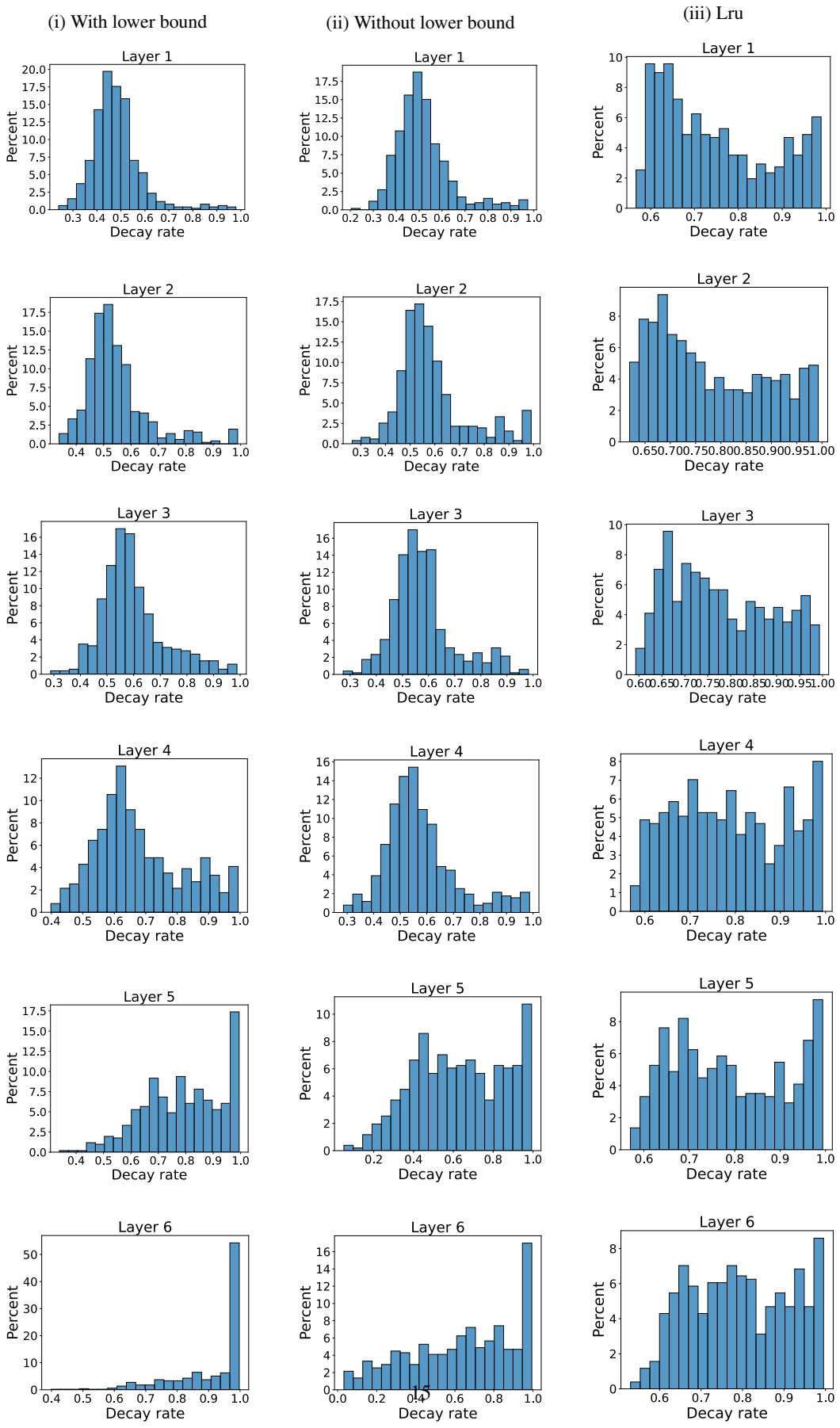

Table 9: The extrapolation performance of competing methods. The best result is highlighted in **bold** and the second in underline.↓ means *lower is better*.

Seqlen	Transformer PPL↓	LS PPL↓	FLASH PPL↓	1+elu PPL↓	Performer PPL↓	cosFormer PPL↓	gMLP PPL↓	S4 PPL↓	DSS PPL↓	GSS PPL↓	ALiBi PPL↓	TNN PPL↓	LRU PPL↓	**HGRU** PPL↓
512	24.78	24.05	24.69	28.05	63.16	27.06	29.13	30.74	41.07	39.66	24.15	24.67	31.12	24.85
768	41.36	23.49	16950.45	47.35	159.74	32.90	1.34E+9	30.41	40.50	39.76	23.38	24.25	30.72	24.4
1024	62.35	23.21	174165.47	70.47	504.30	55.28	8.93E+12	30.24	40.22	39.91	22.98	24.05	30.5	24.16
1280	82.52	23.07	346502.88	91.88	1020.28	102.88	1.58E+15	30.15	40.03	40.82	22.74	23.91	30.38	24.03
1536	100.17	22.97	647788.12	111.56	1568.83	175.26	4.96E+16	30.08	39.94	41.04	22.57	23.83	30.3	23.94
1792	118.42	22.97	1719873.5	129.92	2138.50	267.65	5.67E+17	30.04	39.85	41.08	22.52	23.79	30.24	23.88
2048	133.44	22.99	6.25E+6	147.09	2693.89	368.02	3.59E+18	30.00	39.79	41.53	22.43	23.73	30.19	23.82
3072	188.95	23.25	4.17E+10	206.88	4945.82	820.77	2.19E+20	29.91	39.64	44.08	22.24	23.63	30.09	23.71
4096	246.06	23.83	2.67E+13	267.87	7170.91	1335.51	1.61E+21	29.88	39.59	48.27	22.17	23.58	30.04	23.66
5120	270.93	24.56	1.26E+15	299.31	8443.15	1735.50	5.08E+21	29.85	39.54	53.32	22.11	23.54	30.01	23.62
6144	311.65	25.45	1.58E+16	352.62	10234.07	2146.19	1.16E+22	29.83	39.51	57.73	22.08	23.53	29.99	23.6
7168	346.58	26.42	8.11E+16	389.02	11420.56	2494.79	1.98E+22	29.82	39.49	60.25	22.07	23.51	29.97	23.58
8192	372.18	27.11	3.40E+17	411.50	12557.09	2902.24	2.78E+22	29.82	39.49	63.36	22.05	23.51	29.97	23.58
9216	387.29	28.78	1.22E+18	453.27	14847.66	3028.72	3.93E+22	29.80	39.46	74.92	22.03	23.49	29.96	23.56
10240	395.94	30.13	4.03E+18	457.06	13623.83	3247.83	4.93E+22	29.79	39.45	81.87	22.02	23.48	29.94	23.55
11264	426.54	31.14	1.07E+19	504.19	14661.77	3341.91	5.70E+22	29.79	39.46	87.67	22.00	23.48	29.94	23.55
12288	463.50	33.21	2.52E+19	555.38	17959.85	3644.81	7.18E+22	29.79	39.44	92.11	22.00	23.48	29.94	23.55
13312	506.35	34.72	4.96E+19	584.01	20026.35	3851.70	8.04E+22	29.78	39.43	96.00	22.00	23.47	29.93	23.54
14336	486.86	36.05	1.28E+20	589.83	20971.31	3951.26	9.41E+22	29.78	39.43	101.47	21.99	23.46	29.92	23.53
Avg	261.36	26.71	1.16E+19	299.86	8684.79	1764.75	2.41E+22	29.97	39.75	60.26	**22.40**	23.70	30.17	23.80

Table 10: Speed comparison on LRA benchmark. The 1K,...,5K represent the input sequence length. We mark it with - if a method is out of memory. The higher the better for all metrics.

	Train Speed(steps per second)↑					Inference Speed(steps per second)↑				
Method	1K	2K	3K	4K	5K	1K	2K	3K	4K	5K
Transformer [64]	13.58	4.84	-	-	-	23.67	8.22	-	-	-
Performer [27]	18.40	10.77	7.66	6.30	5.64	30.04	17.36	12.80	10.55	9.52
LS [71]	20.29	11.24	8.05	6.51	5.89	39.05	21.11	15.02	12.6	11.66
Fnet [30]	25.19	15.62	11.24	9.41	8.18	48.81	27.89	19.52	16.27	14.46
cosFormer [49]	22.00	12.80	9.47	7.93	7.13	39.05	22.31	16.62	13.95	12.60
S4 [17]	13.13	7.33	4.91	3.84	3.04	30.04	16.27	10.85	8.58	6.79
FLASH [23]	17.36	9.03	6.54	5.19	4.68	30.04	15.94	11.32	9.19	8.40
TNN [47]	17.55	9.89	6.79	5.68	4.54	33.96	17.75	12.40	10.28	8.22
HGRU	22.31	13.58	9.52	7.40	7.44	43.39	25.19	16.62	14.20	13.95