

Table 1: Performance comparison of different model methods under various configurations. AVG represents average perplexity (lower is better) or average correct score rate.

Method	Pa	Loss	PPL ↓			Accuracy ↑					
			Wiki	LMB	AVG	BOQA	PIQA	Hella	Wino	ARC-e	ARC
<i>Vector decay</i>											
Mamba2	0.16	2.947	40.1	92.9	66.5	60.4	63.6	33.3	51.4	54.5	24.9
GLA	0.16	2.975	42.2	131.3	86.7	60.0	63.6	33.3	48.9	53.1	26.8
Hgrn2	0.16	2.966	41.5	107.2	74.4	60.6	64.3	33.0	50.3	52.5	24.9
LightNet	0.16	3.027	51.7	173.3	112.5	61.1	62.1	30.4	50.4	51.7	24.9
<i>Mamba ablation</i>											
Mamba2 w/o A	0.16	2.946	39.7	98.8	69.2	58.0	64.1	33.3	51.6	53.8	25.9
Mamba2 w/o Δ	0.16	3.019	44.9	132.1	88.5	61.6	62.8	31.8	49.0	53.4	24.9
Mamba2 w/o A , Δ	0.16	2.973	42.3	116.4	79.3	59.6	63.7	32.4	49.8	52.6	25.9
<i>Parameter share</i>											
Mamba2	0.16	2.947	39.9	91.5	65.7	60.7	63.9	33.3	49.3	54.6	27.9
GLA	0.16	3.048	46.9	180.7	113.8	61.0	64.0	32.2	49.3	53.2	25.9
Hgrn2	0.16	2.966	40.9	94.2	67.6	58.1	63.6	33.2	51.9	54.2	25.9
LightNet	0.16	3.104	48.1	312.6	180.3	61.1	62.5	30.5	50.2	49.7	24.9
<i>Scalar decay</i>											
Mamba2	0.16	2.960	41.0	104.7	72.9	58.6	63.2	33.1	50.1	54.1	25.9
GLA	0.16	3.008	43.8	128.0	85.9	57.2	63.7	32.4	51.4	53.8	25.9
Hgrn2	0.16	2.987	44.6	181.5	113.0	59.1	63.2	32.4	51.4	52.6	25.9
LightNet	0.16	3.032	44.6	149.4	97.0	59.0	62.7	31.4	52.6	53.2	25.9
TNL	0.16	2.985	42.4	117.5	79.9	61.9	63.2	32.4	49.3	54.8	26.9
TNL-L	0.16	2.970	41.3	118.7	80.0	61.2	64.0	32.6	50.8	54.5	23.9
<i>Rope</i>											
Mamba2	0.16	2.959	40.9	110.7	75.8	59.4	63.9	32.9	50.4	54.3	26.9
GLA	0.16	3.010	44.9	185.3	115.1	57.7	63.7	32.6	51.1	52.9	24.9
Hgrn2	0.16	2.990	43.1	128.3	85.7	56.9	63.1	32.6	50.5	51.8	25.9
LightNet	0.16	3.002	42.5	128.7	85.6	60.9	63.4	31.8	50.1	53.5	25.9
TNL	0.16	2.975	41.7	109.0	75.3	62.0	63.1	32.3	49.5	53.9	24.9
TNL-L	0.16	2.972	41.4	111.0	76.2	57.0	64.5	32.5	51.5	53.9	25.9
<i>Tpe</i>											
Mamba2	0.16	2.931	38.8	95.6	67.2	53.5	63.9	33.8	51.5	53.5	25.9
GLA	0.16	2.986	43.0	155.3	99.2	61.2	63.8	32.2	50.7	53.1	26.9
Hgrn2	0.16	2.969	41.5	98.8	70.1	60.7	63.9	33.0	52.4	53.4	25.9
LightNet	0.16	2.988	41.6	114.6	78.1	56.1	63.8	32.1	50.6	53.2	25.9
TNL-L	0.16	2.948	40.0	108.0	74.0	60.0	63.9	33.3	51.5	54.3	26.9
<i>Baseline</i>											
LLaMA	0.16	2.921	37.0	87.5	62.2	60.6	64.1	32.8	48.5	53.7	25.9