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
Vector decay											
Mamba2	0.42	2.720	29.8	46.8	38.3	61.2	67.1	39.5	49.6	60.1	28.3
GLA	0.42	2.743	31.0	56.5	43.8	58.5	67.4	39.5	50.9	60.0	27.3
Hgrn2	0.42	2.736	30.4	45.3	37.9	58.4	66.3	39.2	50.8	58.9	28.1
LightNet	0.42	2.784	31.2	55.6	43.4	60.8	66.9	38.0	50.8	58.9	27.0
Mamba ablation											
Mamba2 w/o $\mathbf{A}$	0.42	2.717	29.5	45.4	37.5	61.1	66.5	39.7	53.9	61.5	28.8
Mamba2 w/o Δ	0.42	2.793	33.6	65.4	49.5	61.5	66.3	37.4	49.5	57.8	27.0
Mamba2 w/o $\mathbf{A},\Delta$	0.42	2.738	30.7	51.8	41.3	57.8	66.5	39.4	51.9	58.5	27.4
Parameter share											
Mamba2	0.41	2.719	29.5	45.7	37.6	61.7	67.0	39.6	50.4	60.9	28.9
GLA	0.41	2.805	34.3	79.0	56.7	59.1	65.4	37.6	52.7	60.4	28.0
Hgrn2	0.41	2.738	30.5	45.1	37.8	61.5	66.8	38.9	51.9	60.7	28.1
LightNet	0.41	2.859	34.5	126.8	80.6	61.0	63.9	35.2	52.0	55.9	25.4
Scalar decay											
Mamba2	0.42	2.733	30.3	50.8	40.6	61.9	67.0	39.4	50.2	59.1	28.0
GLA	0.42	2.769	31.4	56.4	43.9	56.7	67.3	38.5	53.0	59.0	27.8
Hgrn2	0.42	2.753	32.3	70.1	51.2	60.7	67.0	38.5	51.7	58.7	27.0
LightNet	0.42	2.787	32.2	61.0	46.6	61.4	65.9	37.0	50.4	59.2	26.0
TNL	0.41	2.759	31.9	50.6	41.3	58.9	66.2	38.3	51.6	60.3	28.0
TNL-L	0.41	2.747	30.8	54.9	42.8	61.6	67.7	38.4	52.2	59.6	27.7
Rope											
Mamba2	0.42	2.732	30.4	50.0	40.2	61.4	67.1	39.4	50.2	60.0	28.2
GLA	0.42	2.764	32.4	67.5	50.0	60.7	66.2	38.5	49.6	59.4	27.8
Hgrn2	0.42	2.751	31.6	53.4	42.5	57.9	66.8	38.6	50.5	59.5	28.8
LightNet	0.42	2.757	31.3	53.8	42.6	61.0	66.3	38.5	51.2	59.3	27.8
TNL	0.41	2.749	31.4	53.7	42.6	61.5	66.4	38.8	51.4	59.9	28.1
TNL-L	0.41	2.740	30.7	54.5	42.6	60.7	66.8	39.0	50.8	60.0	29.0
Tpe											
Mamba2	0.42	2.712	29.5	47.2	38.4	60.3	67.6	39.8	50.6	60.7	29.1
GLA	0.42	2.751	31.4	70.1	50.7	60.0	65.8	39.2	50.9	60.0	26.4
Hgrn2	0.42	2.730	41.5	98.8	70.1	60.7	63.9	33.0	52.4	53.4	25.9
LightNet	0.42	2.754	31.1	59.4	45.2	61.6	67.1	38.6	49.4	59.6	28.1
TNL	0.42	2.740	30.9	50.7	40.8	61.0	67.7	39.5	51.2	61.5	29.1
TNL-L	0.42	2.725	30.0	47.3	38.6	60.8	67.0	39.4	49.6	61.4	29.0
Baseline											
LLaMA	0.41	2.720	28.5	46.7	37.6	60.7	66.7	38.9	51.6	58.6	28.3