Table 1: Performance comparison of Mamba2(M2) and Simple Decay (SD) with different initializations p. AVG represents average perplexity (lower is better) or average correct score rate.

Me	p	Pa	Loss	PPL ↓			Accuracy ↑						
				Wiki	LMB	AVG	BOQA	PIQA	Hella	Wino	ARC-e	ARC-c	OBQA
160M models													
M2	-	0.16	2.947	40.1	92.9	66.5	60.4	63.6	33.3	51.4	54.5	24.9	31.2
SD	0.8	0.16	2.954	41.0	117.6	79.3	61.0	63.6	33.4	50.1	54.8	25.6	30.8
SD	0.9	0.16	2.949	40.6	105.6	73.1	62.0	64.0	33.1	50.8	53.7	26.5	31.0
SD	0.95	0.16	2.939	39.7	97.5	68.6	59.5	64.2	33.5	49.2	54.4	26.4	31.6
SD	0.99	0.16	2.940	39.4	96.7	68.0	61.3	63.9	33.4	48.8	53.8	24.7	32.0
410M models													
M2	-	0.42	2.720	29.8	46.8	38.3	61.2	67.1	39.5	49.6	60.1	28.3	32.2
SD	0.8	0.42	2.727	30.2	45.3	37.8	59.8	67.7	40.0	51.1	59.3	29.3	34.6
SD	0.9	0.42	2.722	29.8	45.6	37.7	61.0	68.1	40.1	51.2	59.3	27.2	30.6
SD	0.95	0.42	2.716	29.5	48.9	39.2	60.5	67.0	39.7	52.8	60.1	27.4	34.0
SD	0.99	0.42	2.711	29.4	46.7	38.1	61.0	66.8	40.0	50.7	60.9	28.9	33.6