

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
<i>1.45B models</i>													
M2	-	1.45	2.514	22.8	25.2	24.0	61.7	70.0	47.7	52.8	67.1	30.7	37.6
SD	0.8	1.45	2.516	22.9	24.5	23.7	61.7	70.4	47.8	53.0	66.1	32.3	36.8
SD	0.9	1.45	2.512	22.7	25.6	24.2	60.7	70.6	47.5	53.7	65.6	31.6	36.2
SD	0.95	1.45	2.511	22.7	23.9	<b>23.3</b>	62.1	70.2	48.1	51.1	66.0	32.4	35.4
SD	0.99	1.45	2.511	22.6	24.3	23.5	58.4	70.1	47.7	55.9	66.7	33.4	36.4