seq512-hd64-12h-8b | Q=512, K=512, head_dim=64, heads=12, batch=8

Kernel	Avg Forward Time (ms)	Relative to Triton Flash
scaled_dot_prod_attention	9.28	5.08x slower
vectorized_torch	7.10	3.89x slower
vectorized_torch_compiled	0.80	2.27x faster
flash_attention_triton	**1.83**	**1.00x**