seq2048-hd64-12h-8b | Q=2048, K=2048, head_dim=64, heads=12, batch=8

Kernel	Avg Forward Time (ms)	Relative to Triton Flash
scaled_dot_prod_attention	58.56	5.23x slower
vectorized_torch	65.13	5.82x slower
vectorized_torch_compiled	21.46	1.92x slower
flash_attention_triton	**11.19**	**1.00x**