seq1024-hd128-16h-32b | Q=1024, K=1024, head_dim=128, heads=16, batch=32

Kernel	Avg Forward Time (ms)	Relative to Triton Flash
scaled_dot_prod_attention	97.68	3.10x slower
vectorized_torch	108.24	3.44x slower
vectorized_torch_compiled	31.83	1.01x slower
flash_attention_triton	**31.47**	**1.00x**