

Running on CPU

--- ($\mathbf{X1} + \mathbf{X2} + \mathbf{X3}$) Final Form Benchmark (Forward + Backward Time) ---

D_MODEL=64, BATCH_SIZE=8

N_SEQ	ResNet-BK (MoE) (ms)	Attention (Autograd) (ms)
----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----		
64	13.353	4.231
256	42.650	27.721
1024	229.348	613.301
2048	377.409	2549.120

Benchmark graph saved to scaling_benchmark_final_moe.png

