

Running on CPU

--- (**x1 + x2 + x3**) Final Form Benchmark (Forward + Backward Time) ---

D\_MODEL=64, BATCH\_SIZE=8

| N\_SEQ | ResNet-BK (MoE) (ms) | Attention (Autograd) (ms) |

[illegible]

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

