Coordinate Check Results (Step 100)

Layers (in activation order): 1. layers.0, 2. layers.1, 3. logits

$\mu P - SSM$, μP, SP

$\frac{\|x\|_2}{\sqrt{N_x}}$, $\frac{\|\Delta x\|_2}{\sqrt{N_x}}$ versus $\log_2$ width multiplier