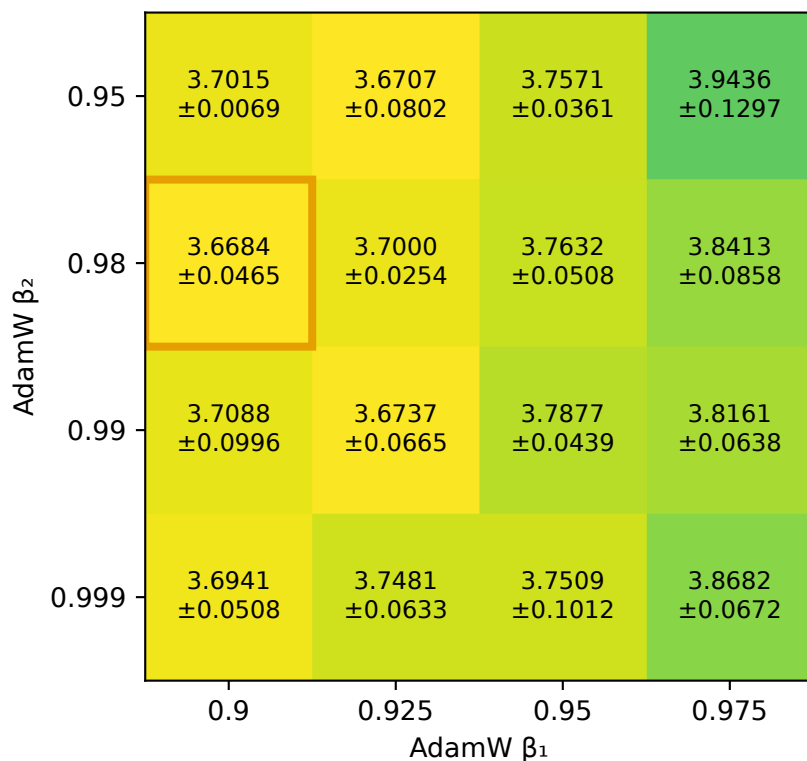
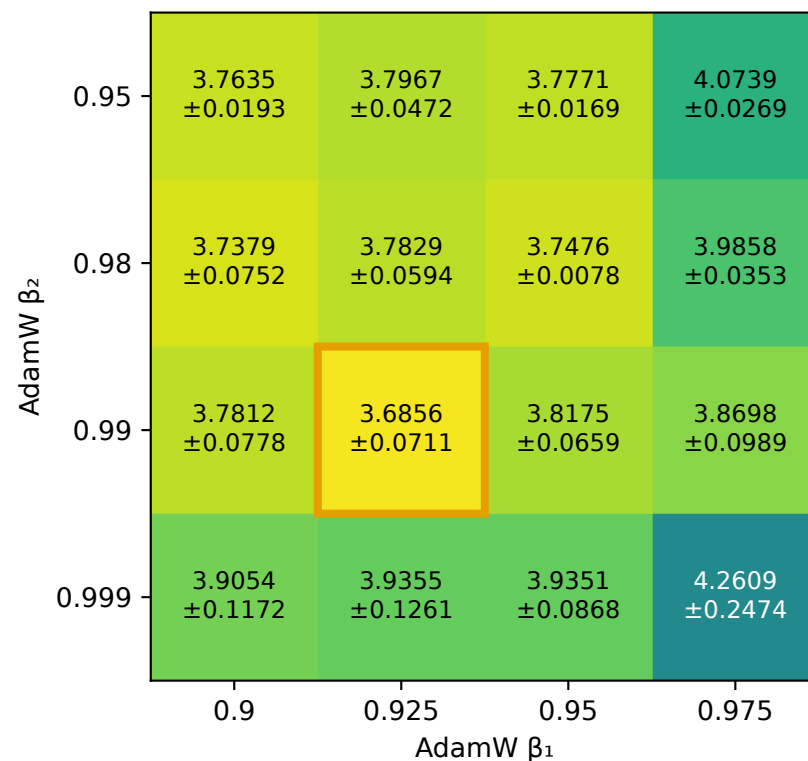


AWC · four 20M workers · no clipping · 131,072 targets per batch
Complete LR × β_1 × β_2 grid · 80 configurations · 240 runs

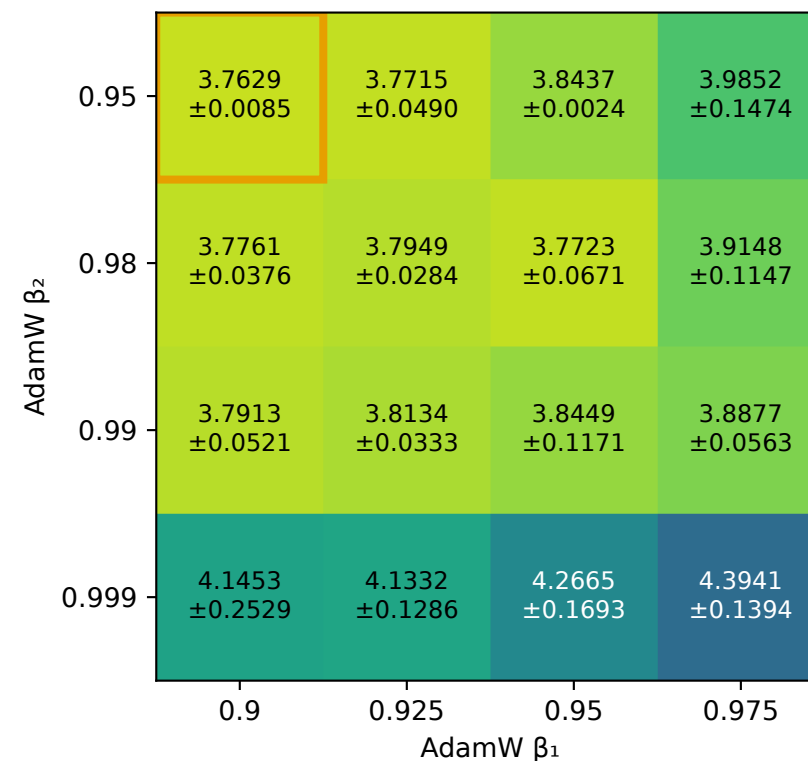
LR 0.0056



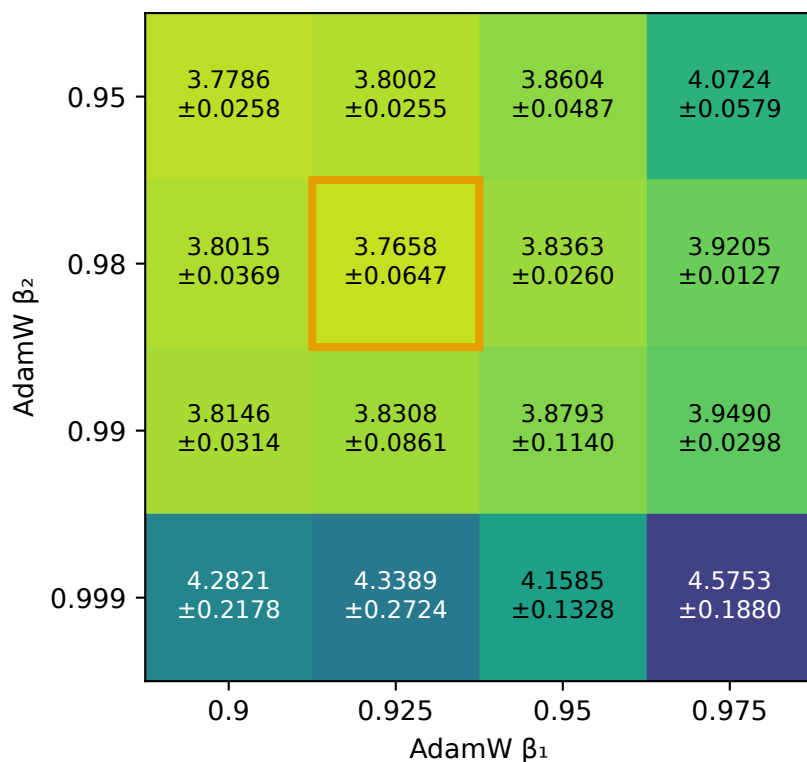
LR 0.008



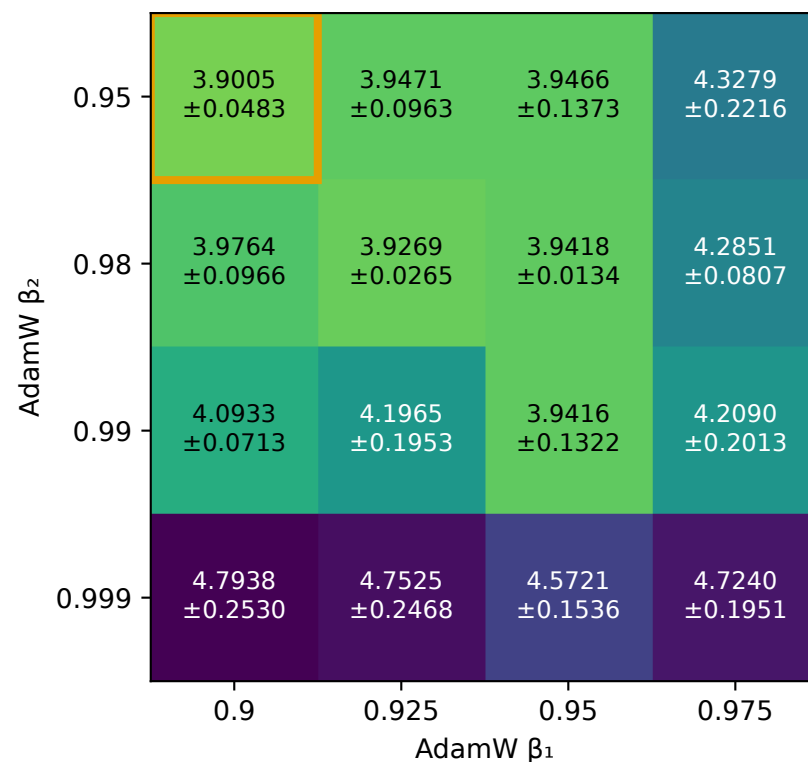
LR 0.01



LR 0.0112



LR 0.016



Best tested recipe
LR 0.0056, β_1 0.9, β_2 0.98
3.668404 ± 0.046502 nats

Orange outline: lowest mean at each LR
Shared color scale across all panels
Every cell: mean ± sample SD, three seeds
Lower loss is better

