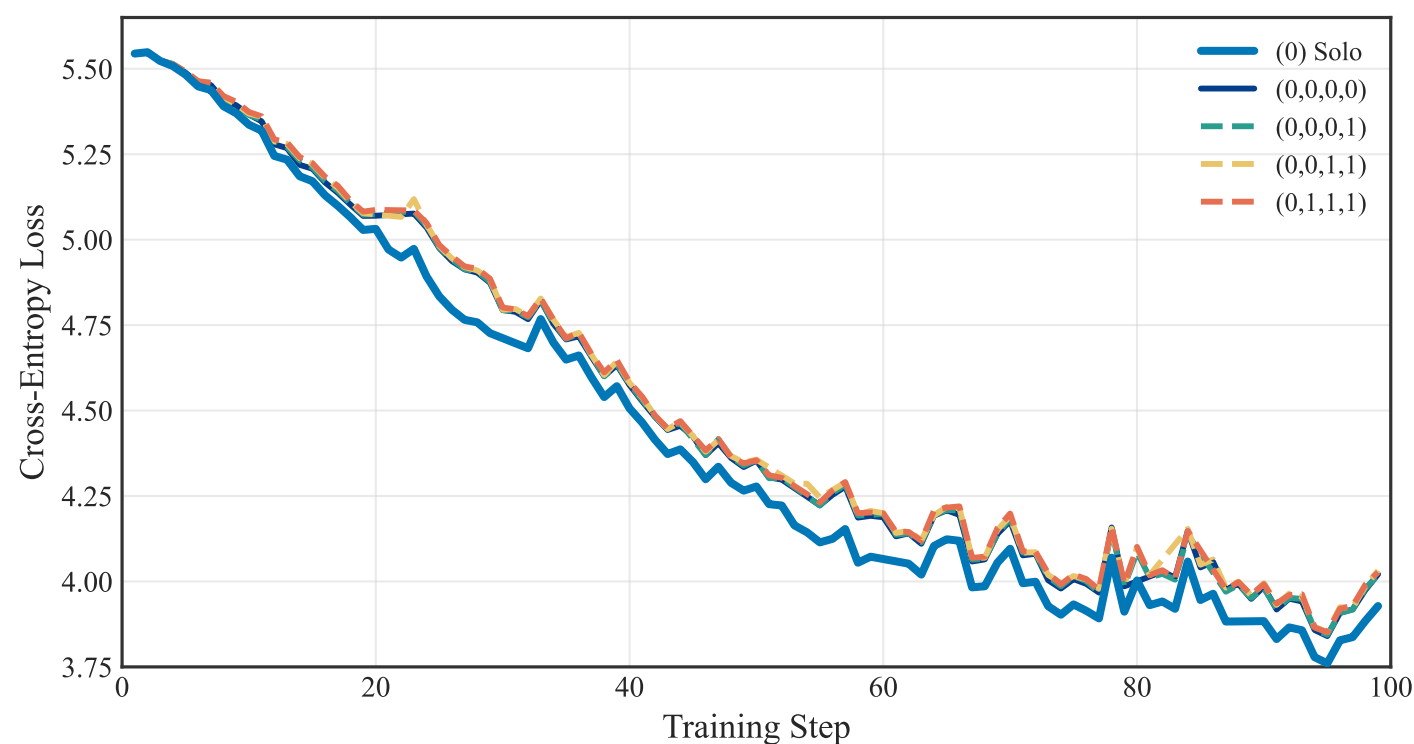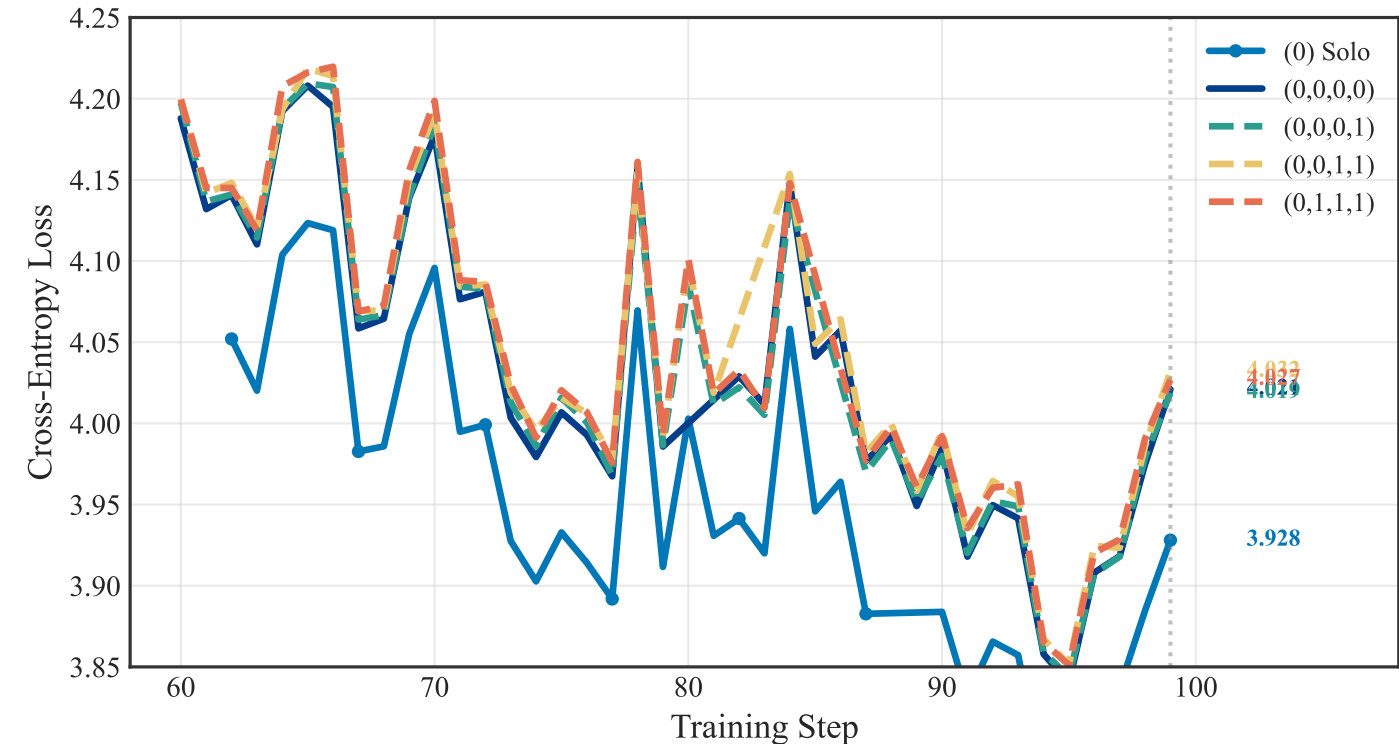**Heterogeneous MatFormer Training: TinyShakespeare 100 Steps**
Comparing Solo, Homogeneous, and Mixed-Tier Distributed Training
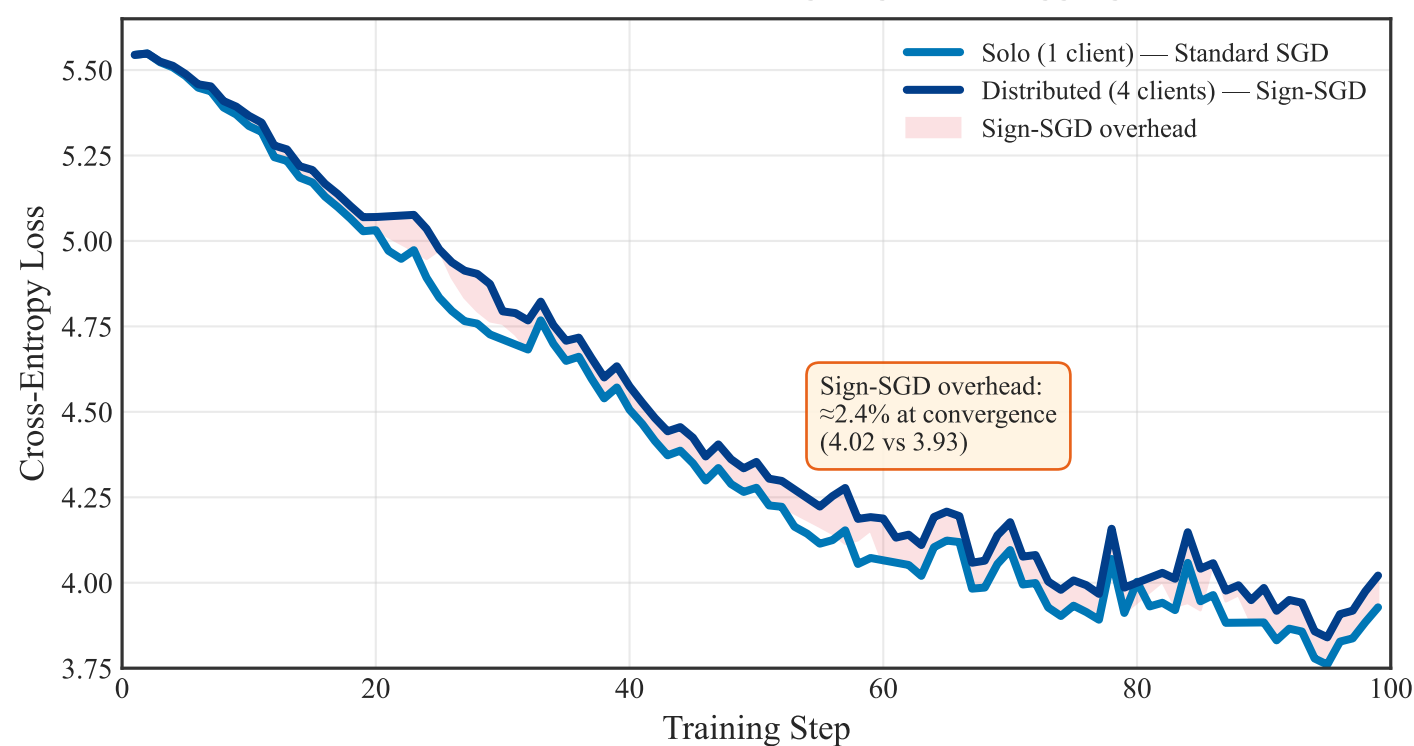
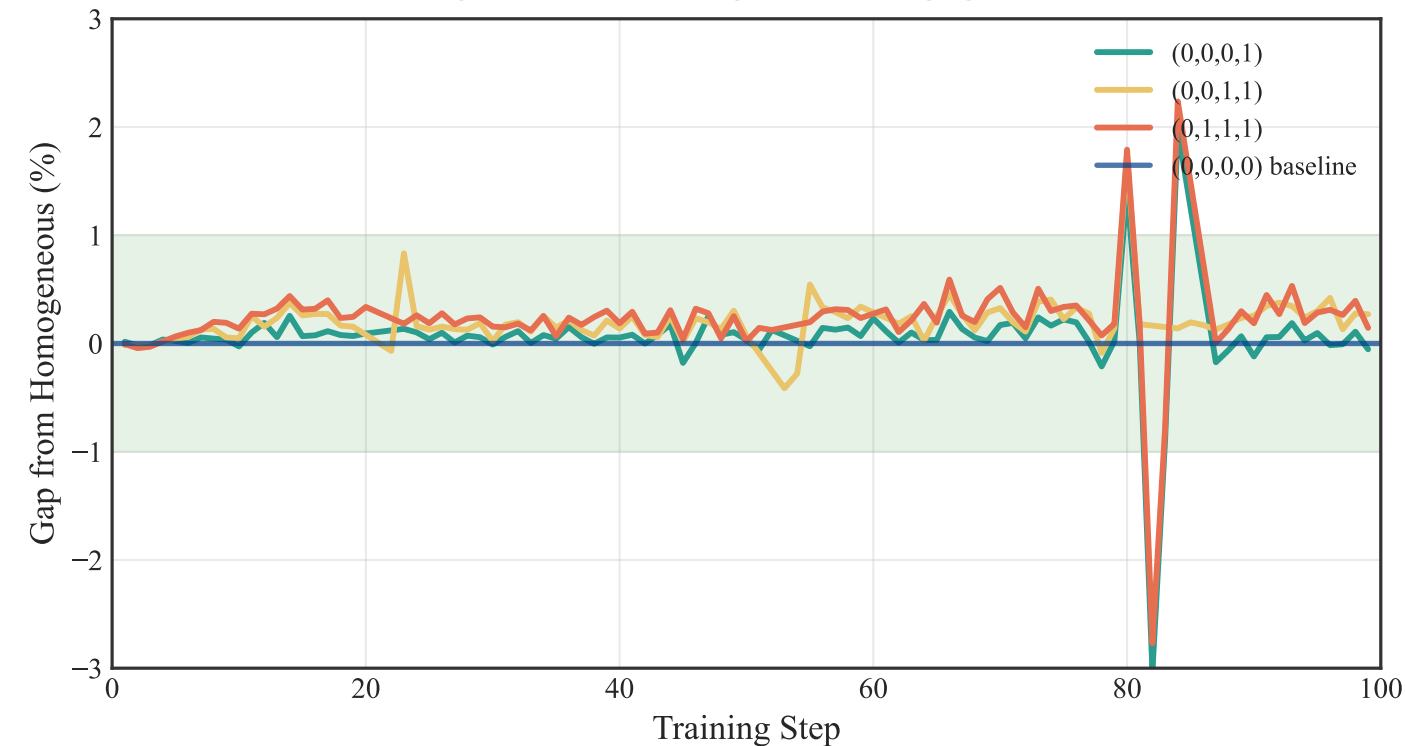**(A) Loss Trajectories Across All Configurations**

(B) Final Convergence (Steps 60–99)

(C) Cost of Distributed Training: Sign-SGD Aggregation

Sign-SGD overhead:
≈2.4% at convergence
(4.02 vs 3.93)

(D) Heterogeneous vs Homogeneous: Negligible Difference

(E) Tier-0 vs Tier-1 Loss Gap During Training

Tier-1 clients track Tier-0 within ±2%
throughout training