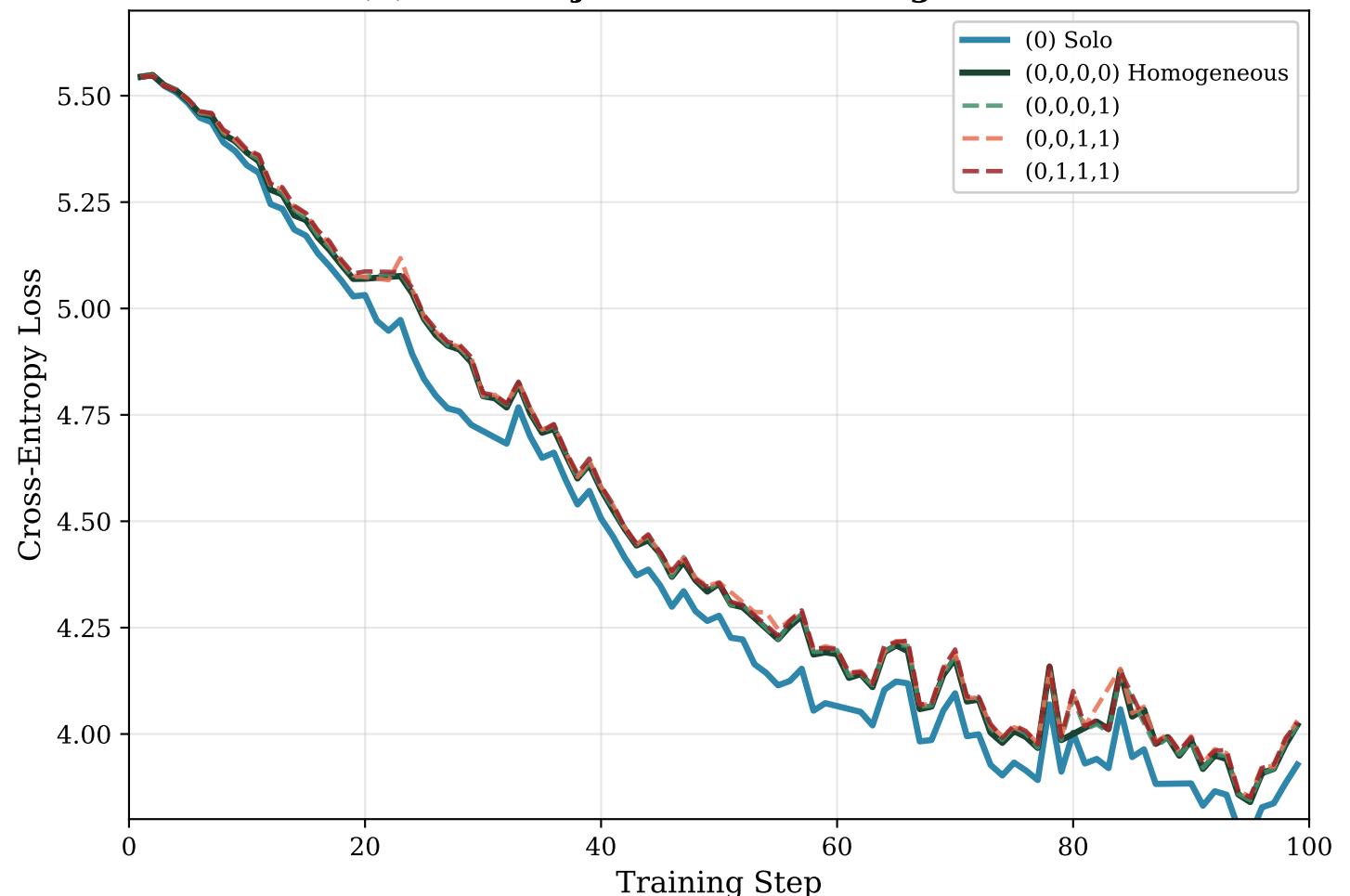


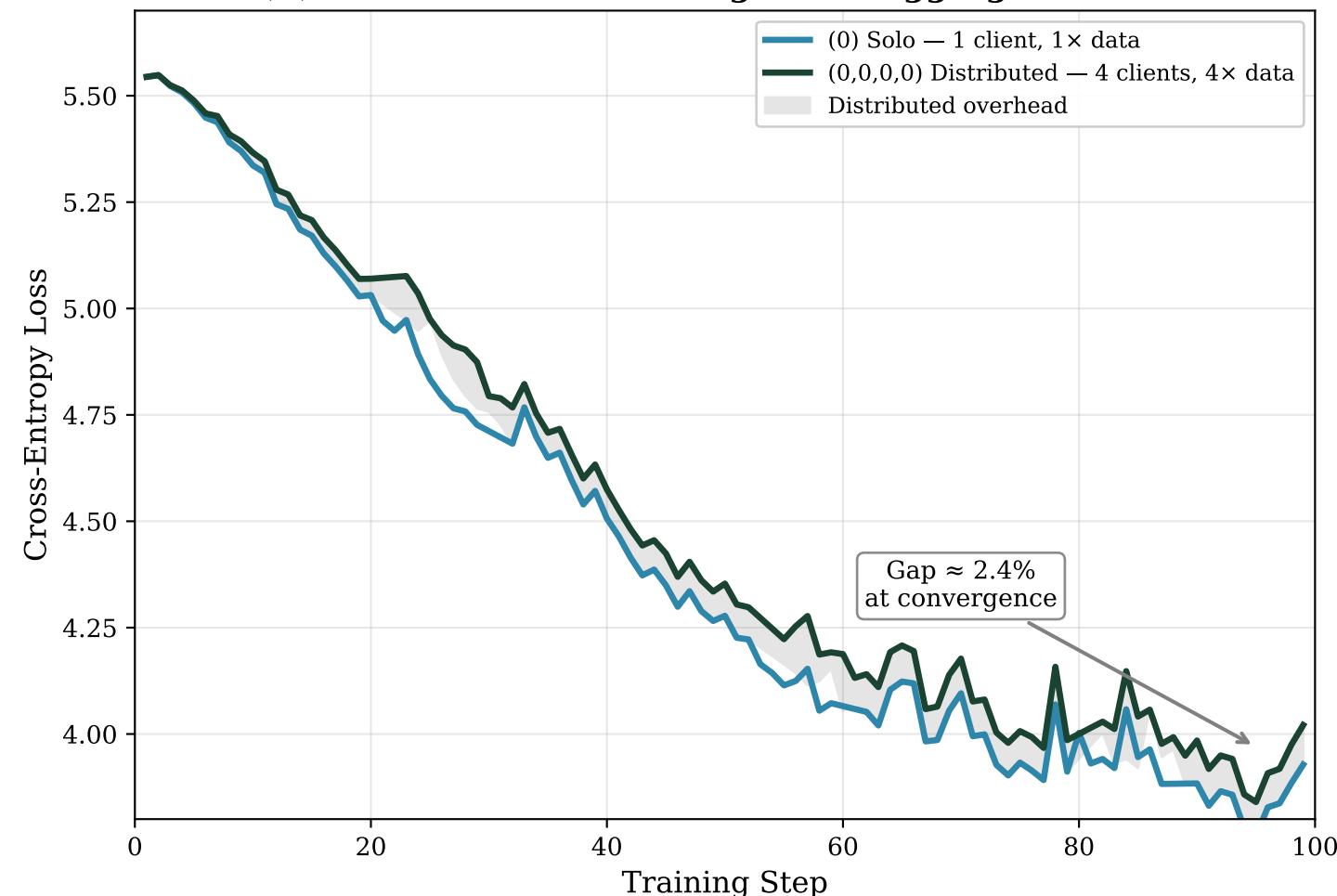
Heterogeneous MatFormer Training on TinyShakespeare (100 Steps)

Sign-SGD Distributed Training with Mixed FFN Widths

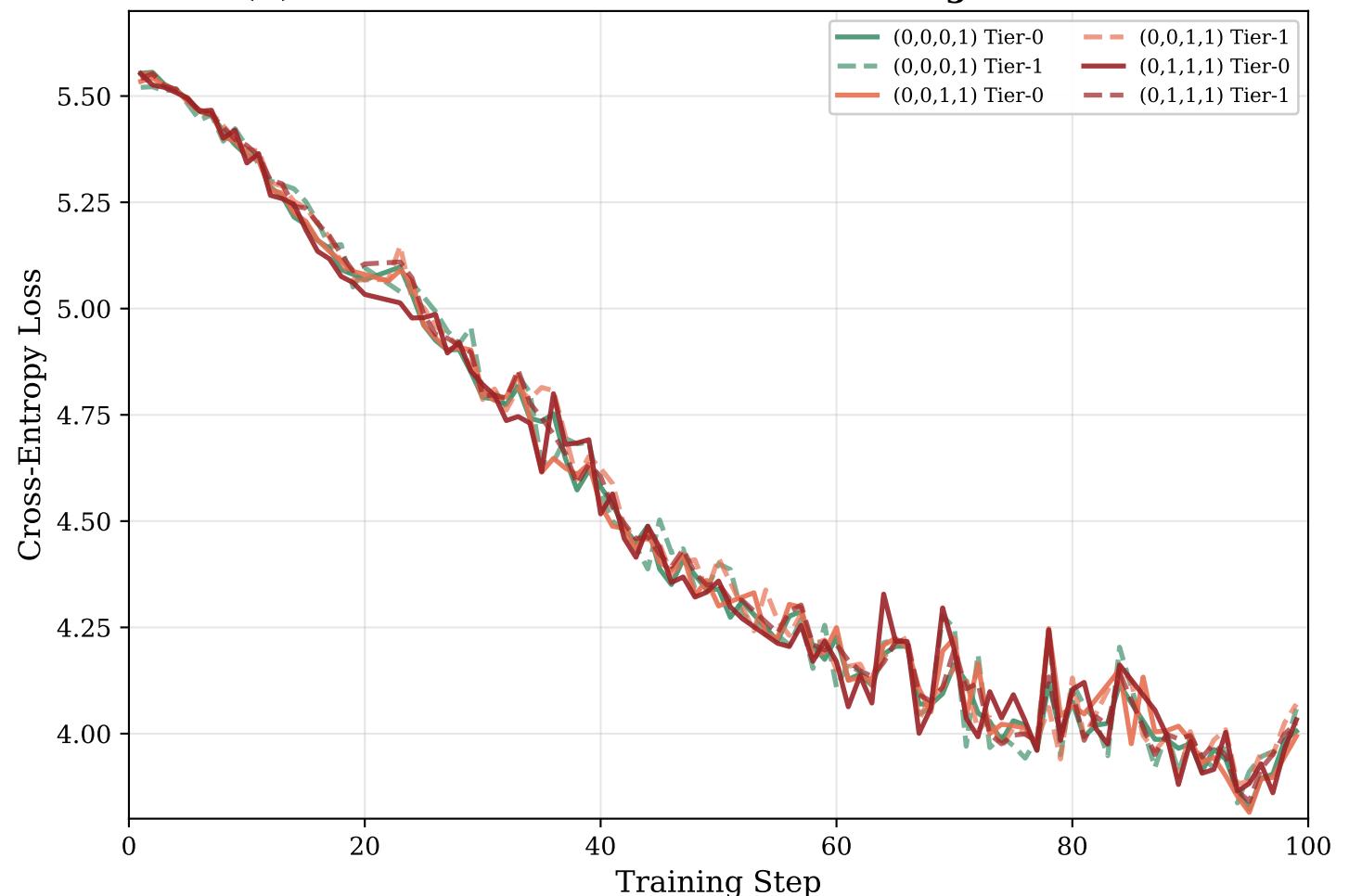
(A) Loss Trajectories: All Configurations



(B) Solo vs Distributed: Sign-SGD Aggregation Cost



(C) Tier-0 vs Tier-1 Loss Within Heterogeneous Runs



(D) Tier Gap Over Training: $(L_{T1} - L_{T0})/L_{T0}$

