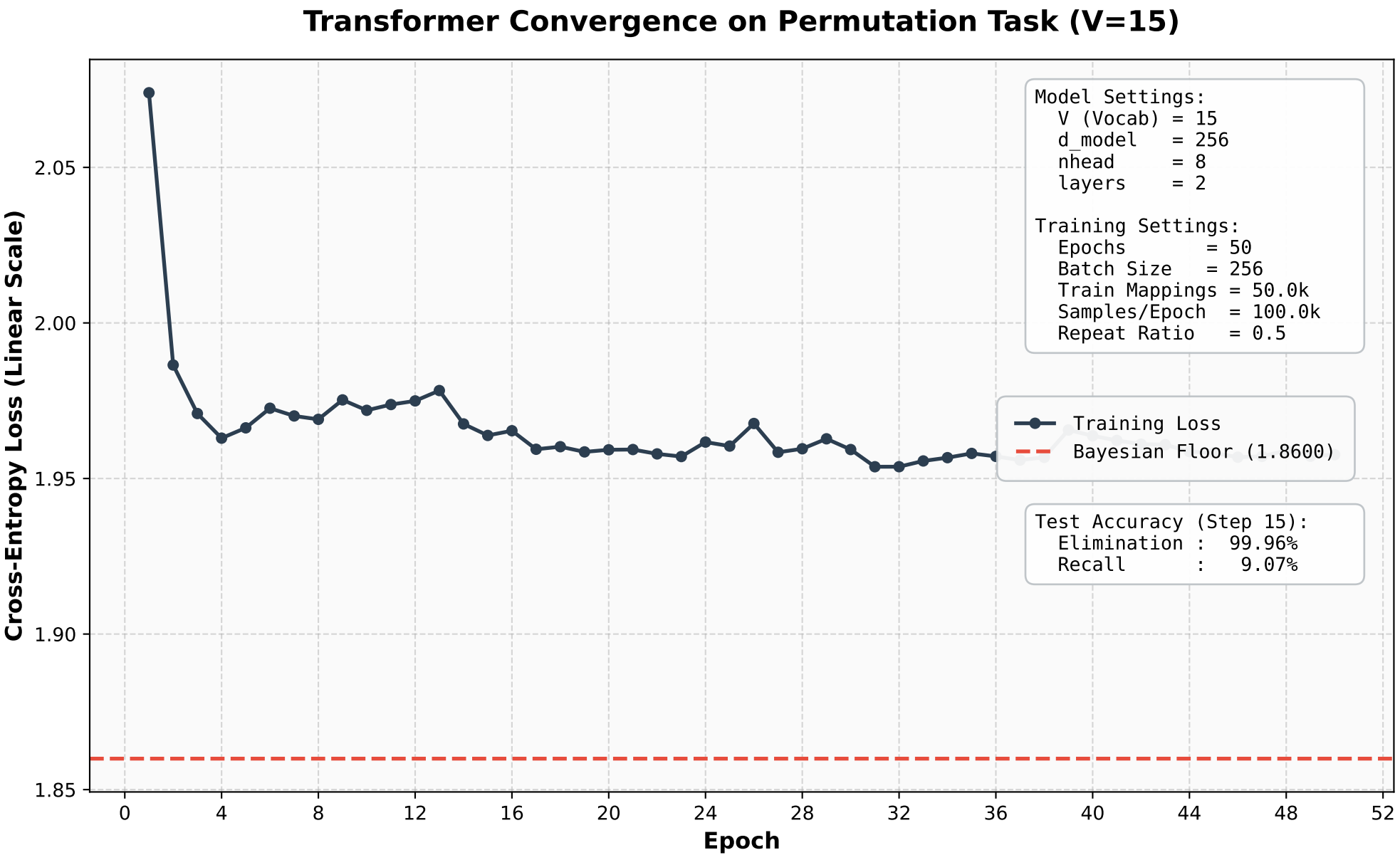
Transformer Convergence on Permutation Task (V=15)
Cross-Entropy Loss (Linear Scale)
Epoch
2.05
2.00
1.95
1.90
1.85
0
4
8
12
16
20
24
28
32
36
40
44
48
52
Model Settings:
V (Vocab) = 15
d_model = 256
nhead = 8
layers = 2
Training Settings:
Epochs = 50
Batch Size = 256
Train Mappings = 50.0k
Samples/Epoch = 100.0k
Repeat Ratio = 0.5
Training Loss
Bayesian Floor (1.8600)
Test Accuracy (Step 15):
Elimination : 99.96%
Recall : 9.07%