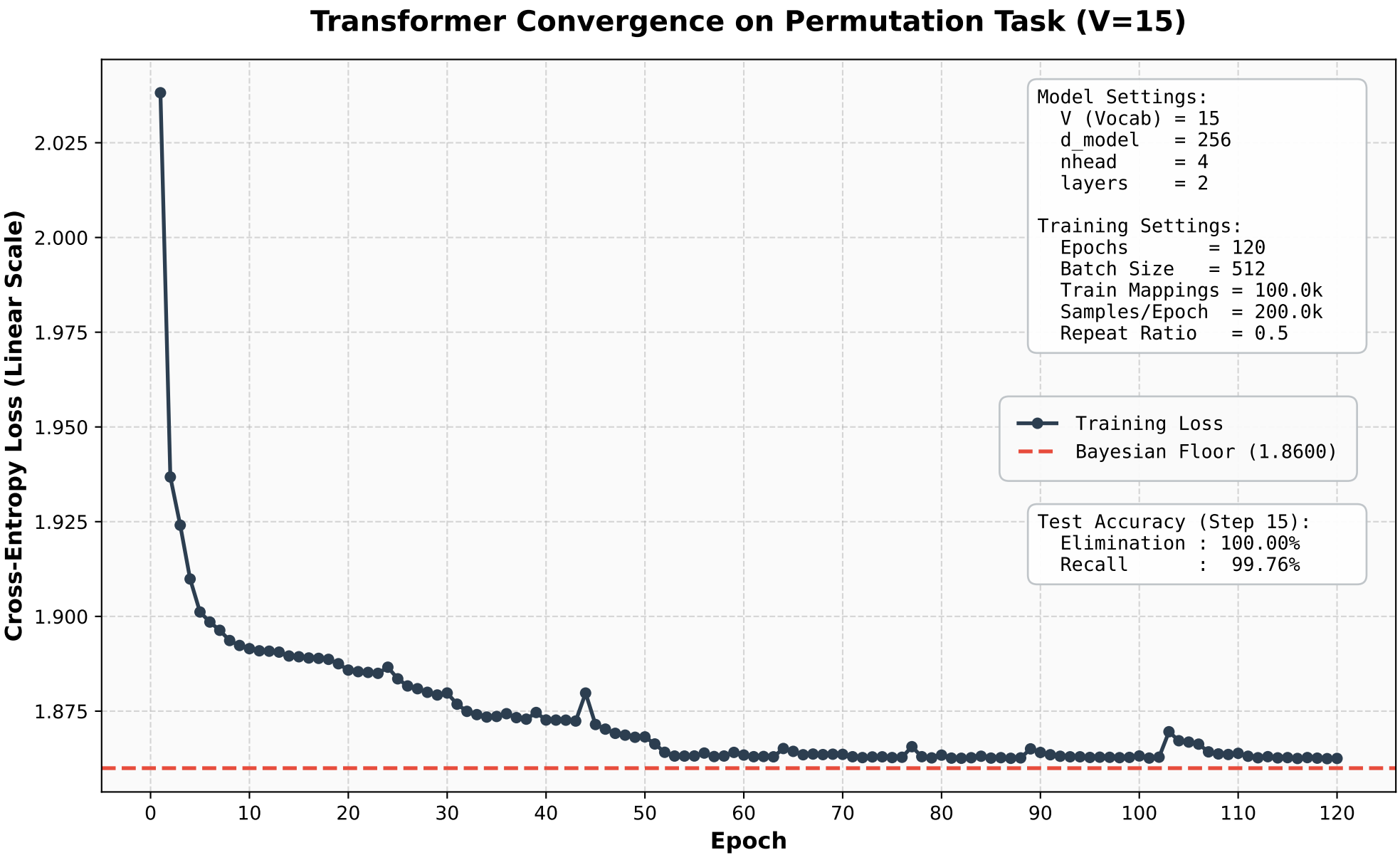
Transformer Convergence on Permutation Task (V=15)
Model Settings:
V (Vocab) = 15
d_model = 256
nhead = 4
layers = 2
Training Settings:
Epochs = 120
Batch Size = 512
Train Mappings = 100.0k
Samples/Epoch = 200.0k
Repeat Ratio = 0.5
Training Loss
Bayesian Floor (1.8600)
Test Accuracy (Step 15):
Elimination : 100.00%
Recall : 99.76%
Cross-Entropy Loss (Linear Scale)
2.025
2.000
1.975
1.950
1.925
1.900
1.875
0
10
20
30
40
50
60
70
80
90
100
110
120
Epoch