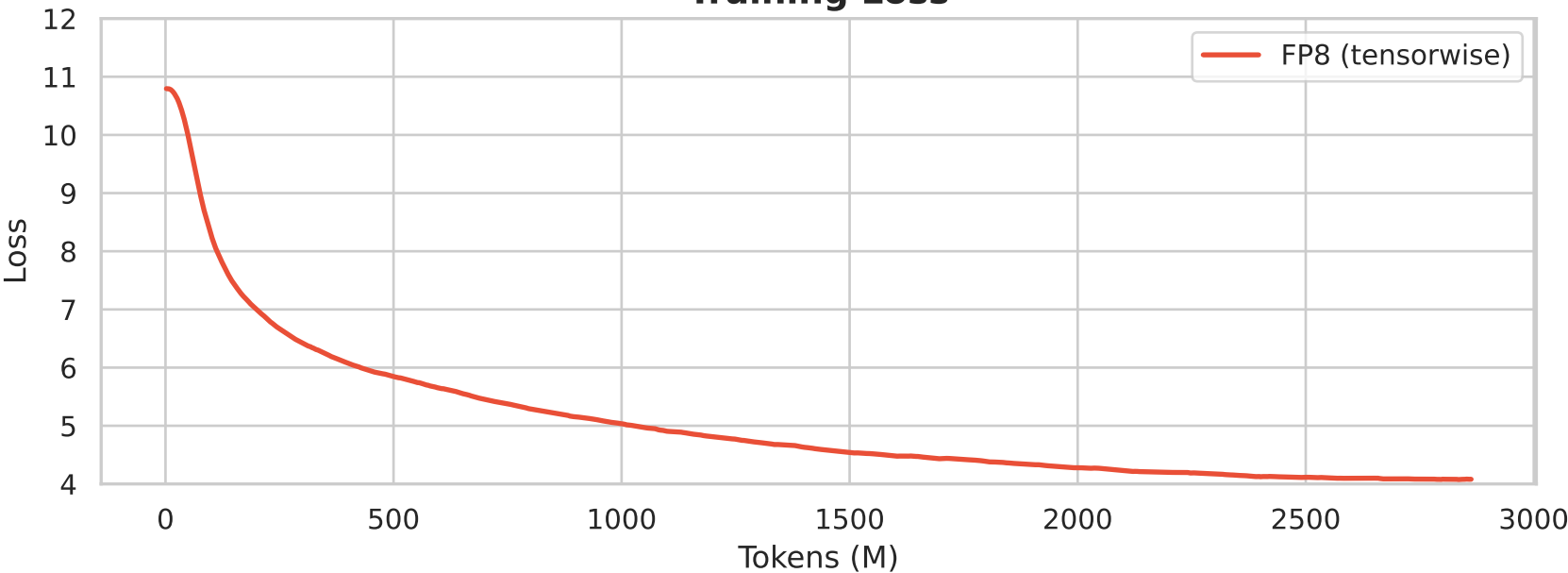


# FP8 vs BF16 Training Comparison - 1.5B Parameter Model on NVIDIA GB10

Training Loss

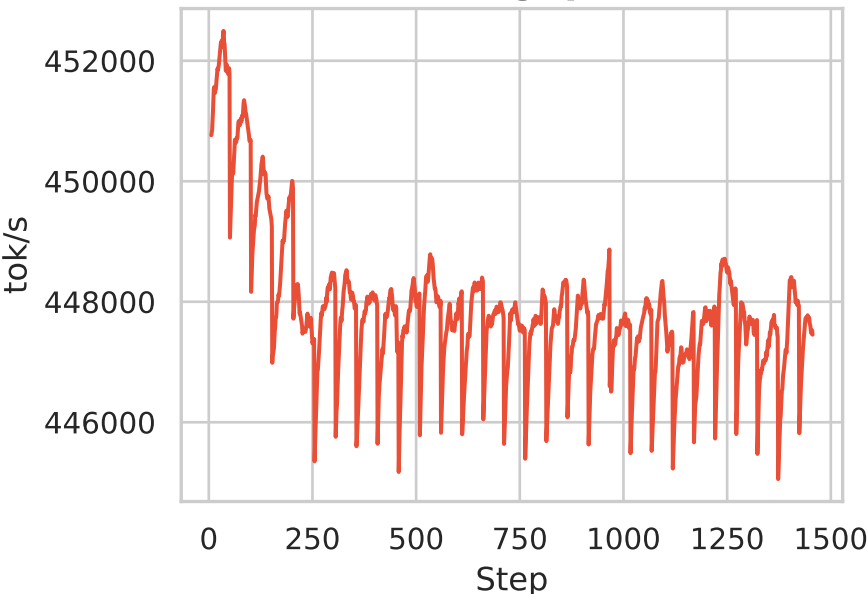


Training Summary

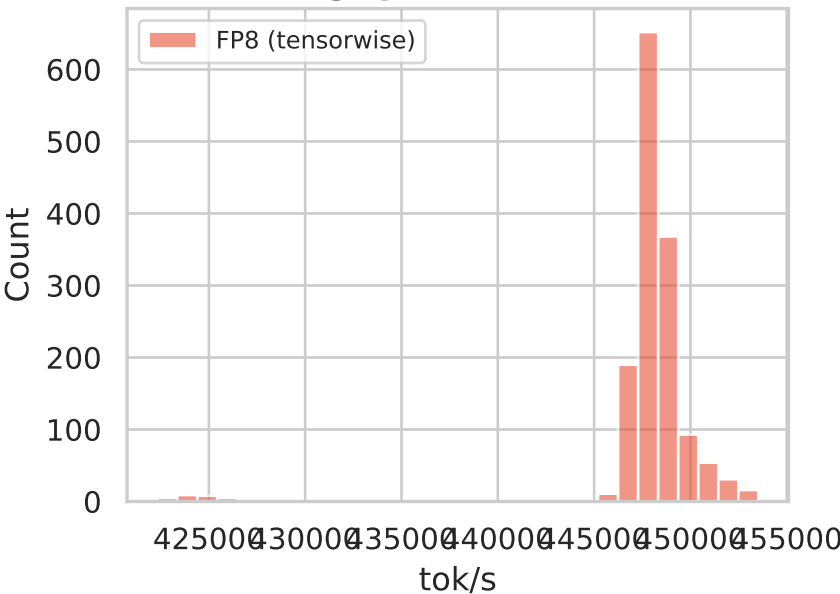
=====

FP8 (tensorwise)  
Final Loss: 4.052  
Final PPL: 57.5  
Tokens: 2862.6M  
Time: 1.79h  
Avg tok/s: 447870

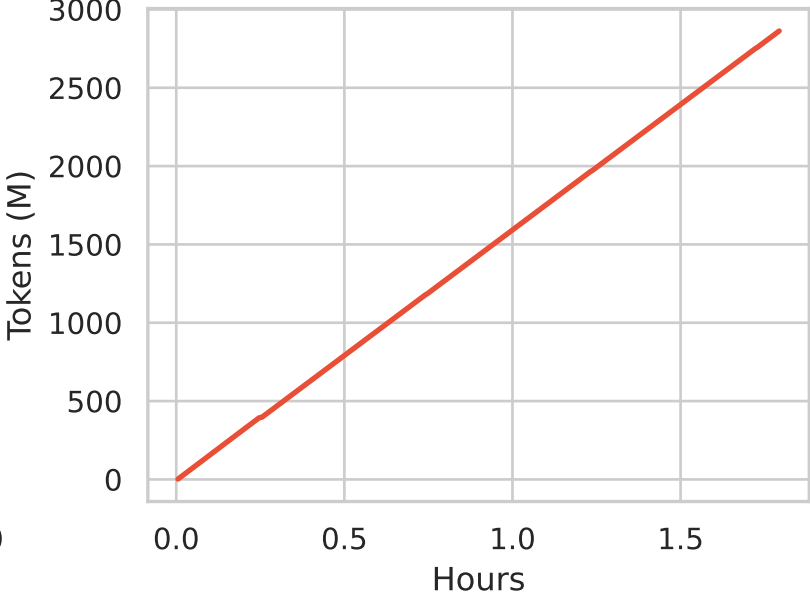
Throughput



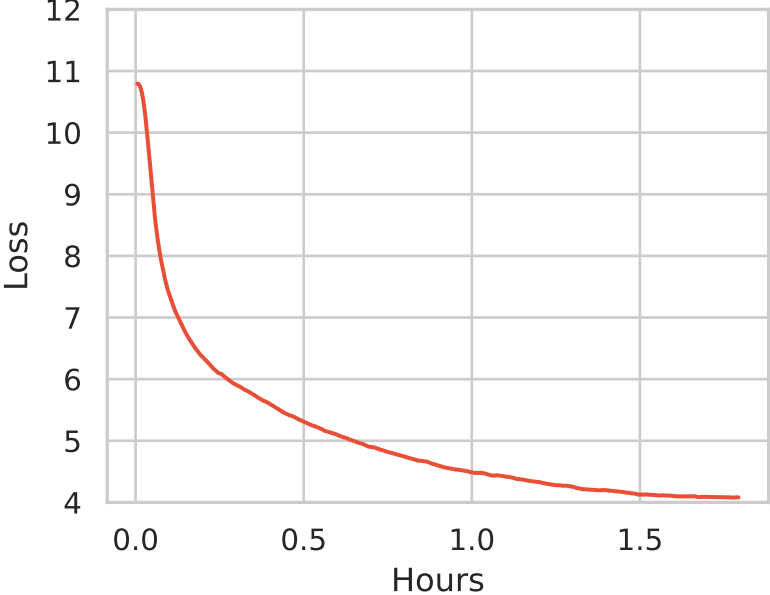
Throughput Distribution



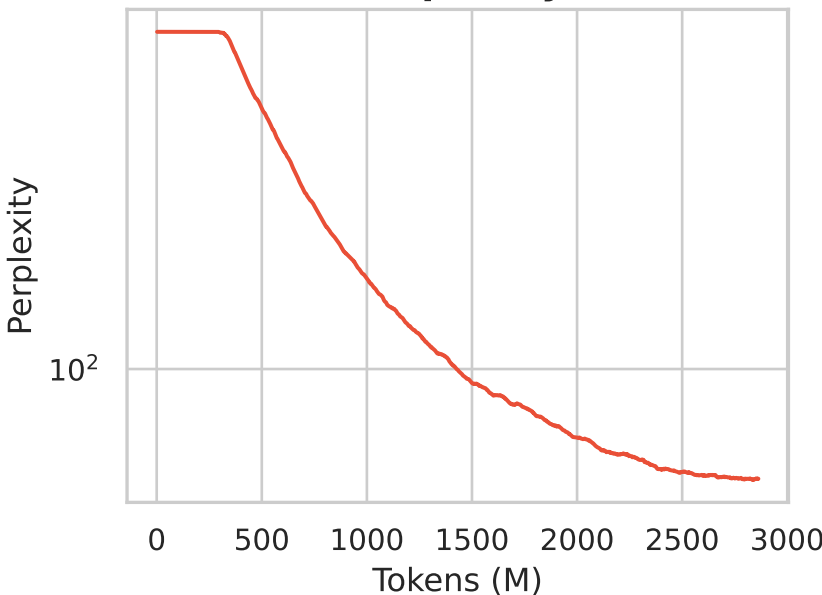
Tokens vs Time



Loss vs Wall Time



Perplexity



Step Time

