

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

Training Loss

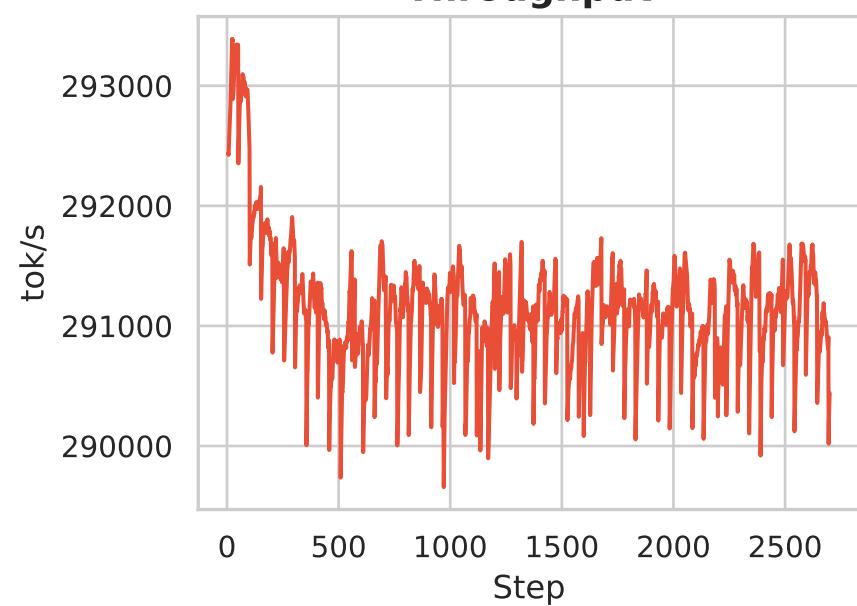


Training Summary

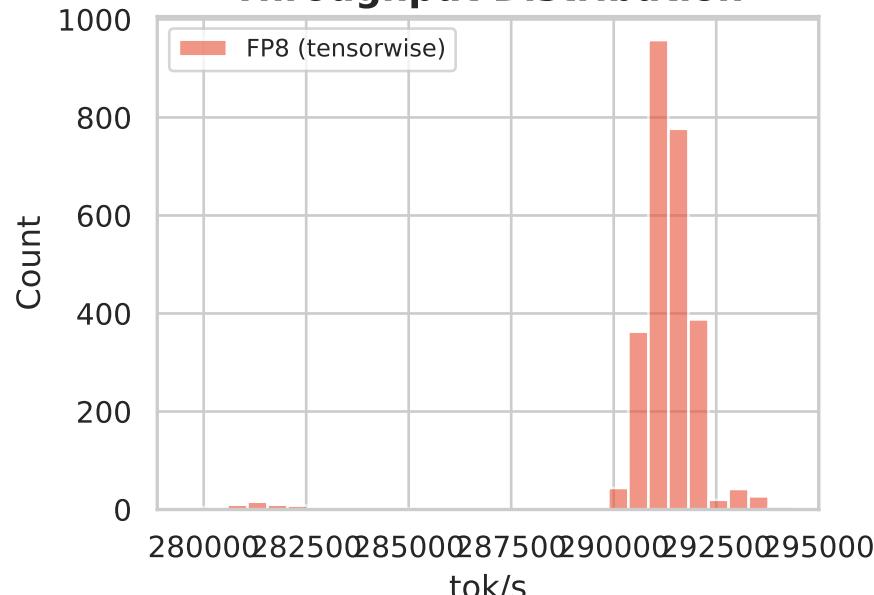
```
=====
```

FP8 (tensorwise)
Final Loss: 2.987
Final PPL: 19.8
Tokens: 5308.4M
Time: 5.09h
Avg tok/s: 291154

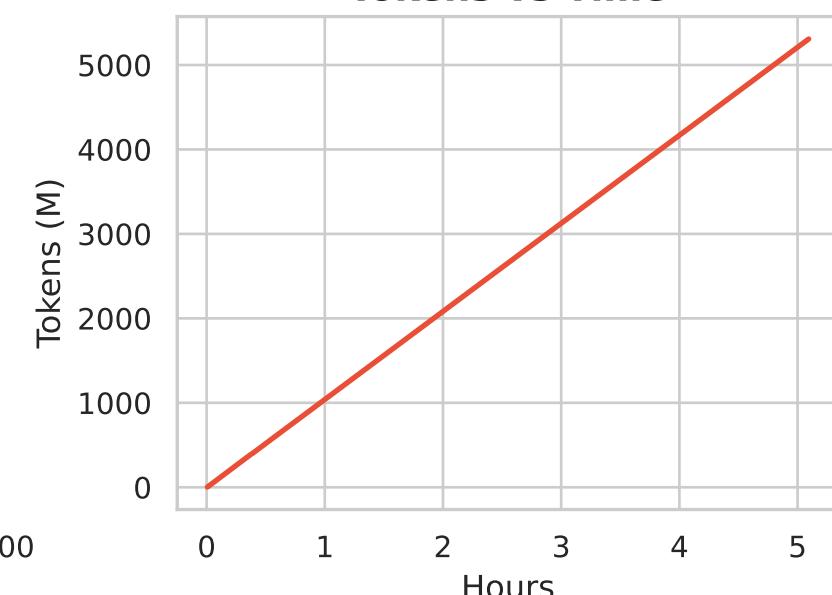
Throughput



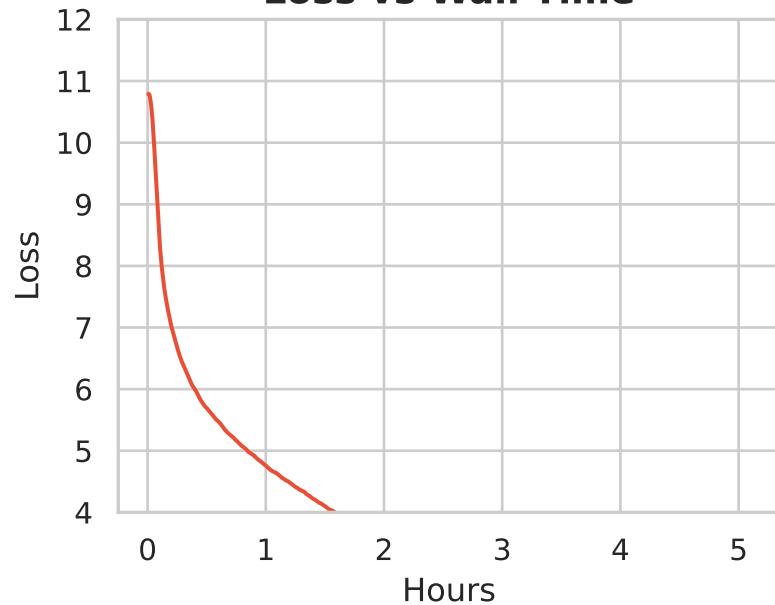
Throughput Distribution



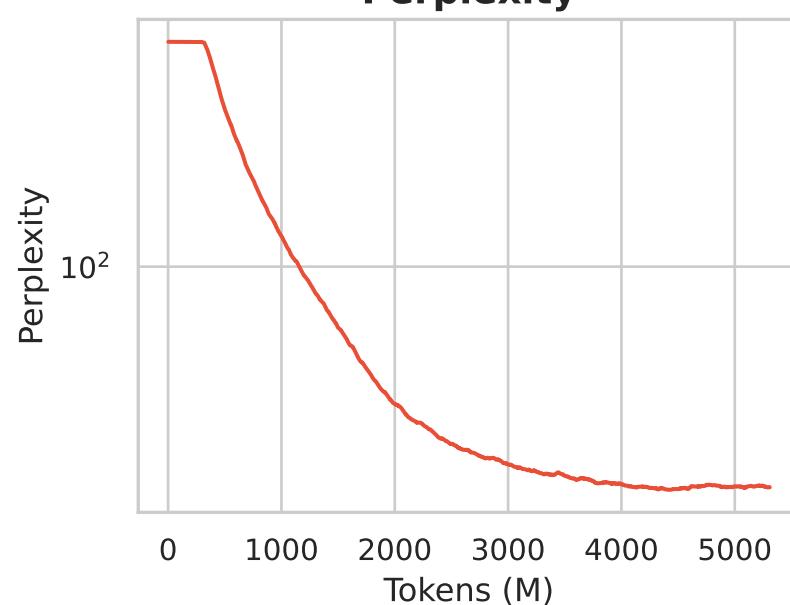
Tokens vs Time



Loss vs Wall Time



Perplexity



Step Time

