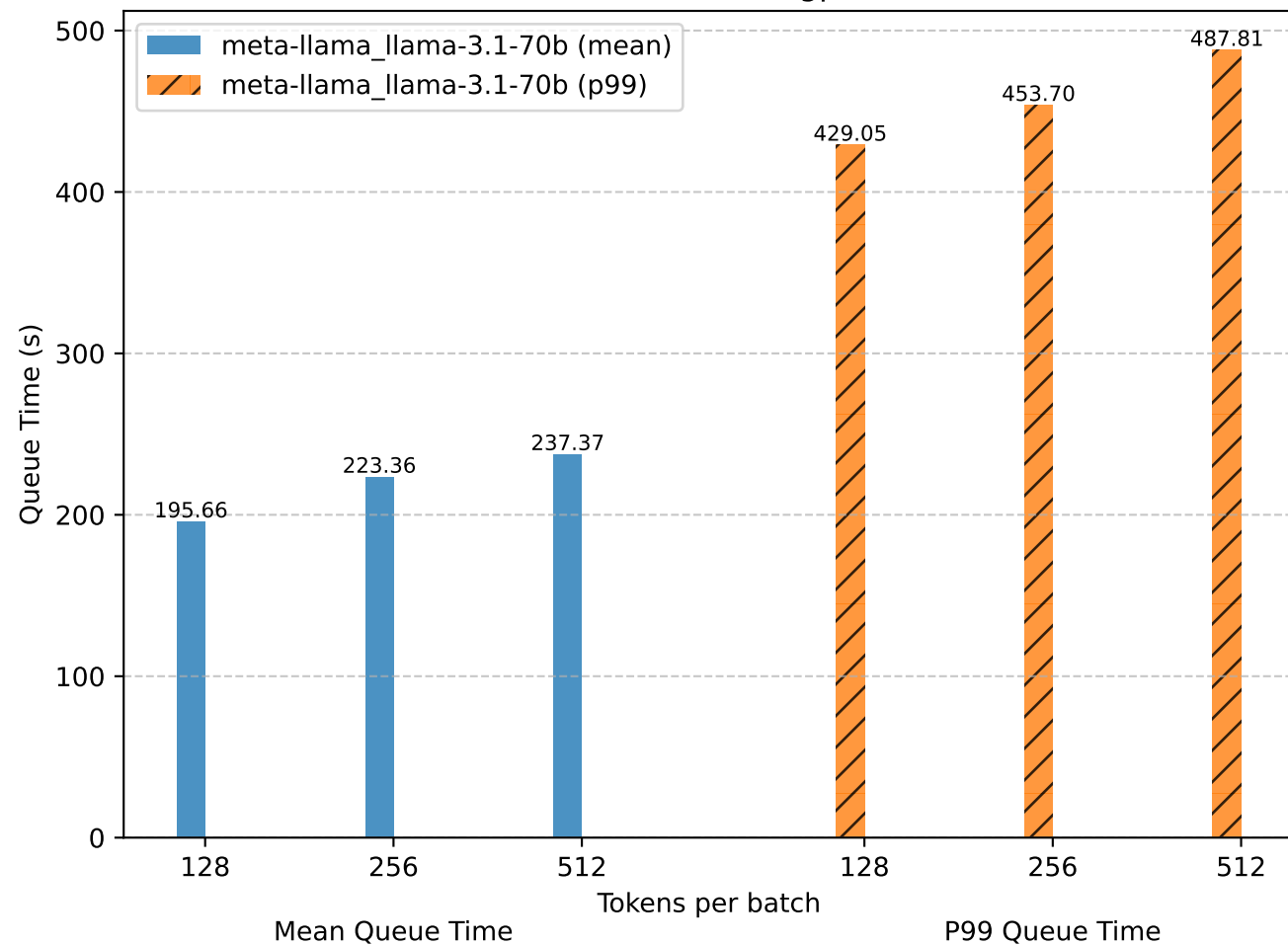


Queueing Time  
meta-llama\_llama-3.1-70b (TP=4)  
Batch Size: 8  
Dataset: sharegpt



Queueing Time  
meta-llama\_llama-3.1-70b (TP=4)  
Batch Size: 8  
Dataset: wildchat

