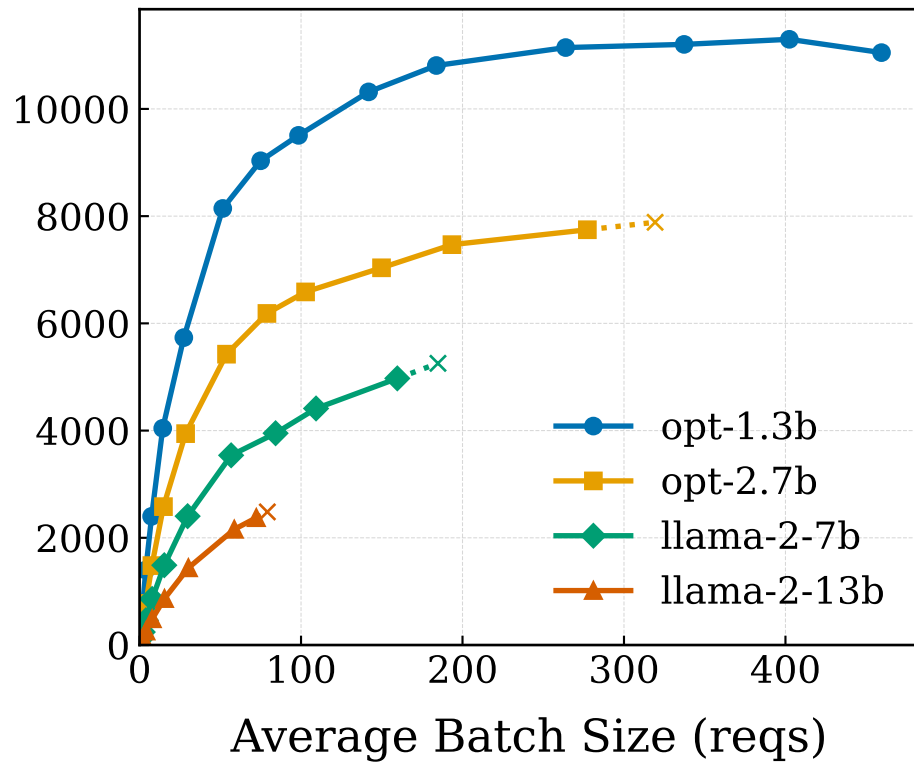


Throughput (tokens/s)



Latency (ms)

