Prefill throughput (tok/s)

	N=256	N=512	N=1024	N=2048	N=4096	N=8192
BS=1	**6.1k** 6.7k	**8.7k** 8.4k	**9.5k** 8.9k	**9.7k** 8.3k	**9.3k** 6.8k	**7.6k** 3.7k
BS=2	**8.3k** 8.1k	**9.2k** 8.8k	**9.5k** 8.6k	**9.5k** 7.5k	**9.5k** 6.1k	**8.1k** OOM
BS=4	**9.7k** 9.5k	**10.0k** 9.8k	**10.4k** 9.1k	**10.0k** 7.8k	**9.4k** 4.9k	**8.2k** OOM
BS=8	**10.3k** 10.1k	**10.5k** 10.1k	**10.5k** 9.3k	**10.1k** 7.7k	**9.5k** 5.0k	**7.7k** OOM