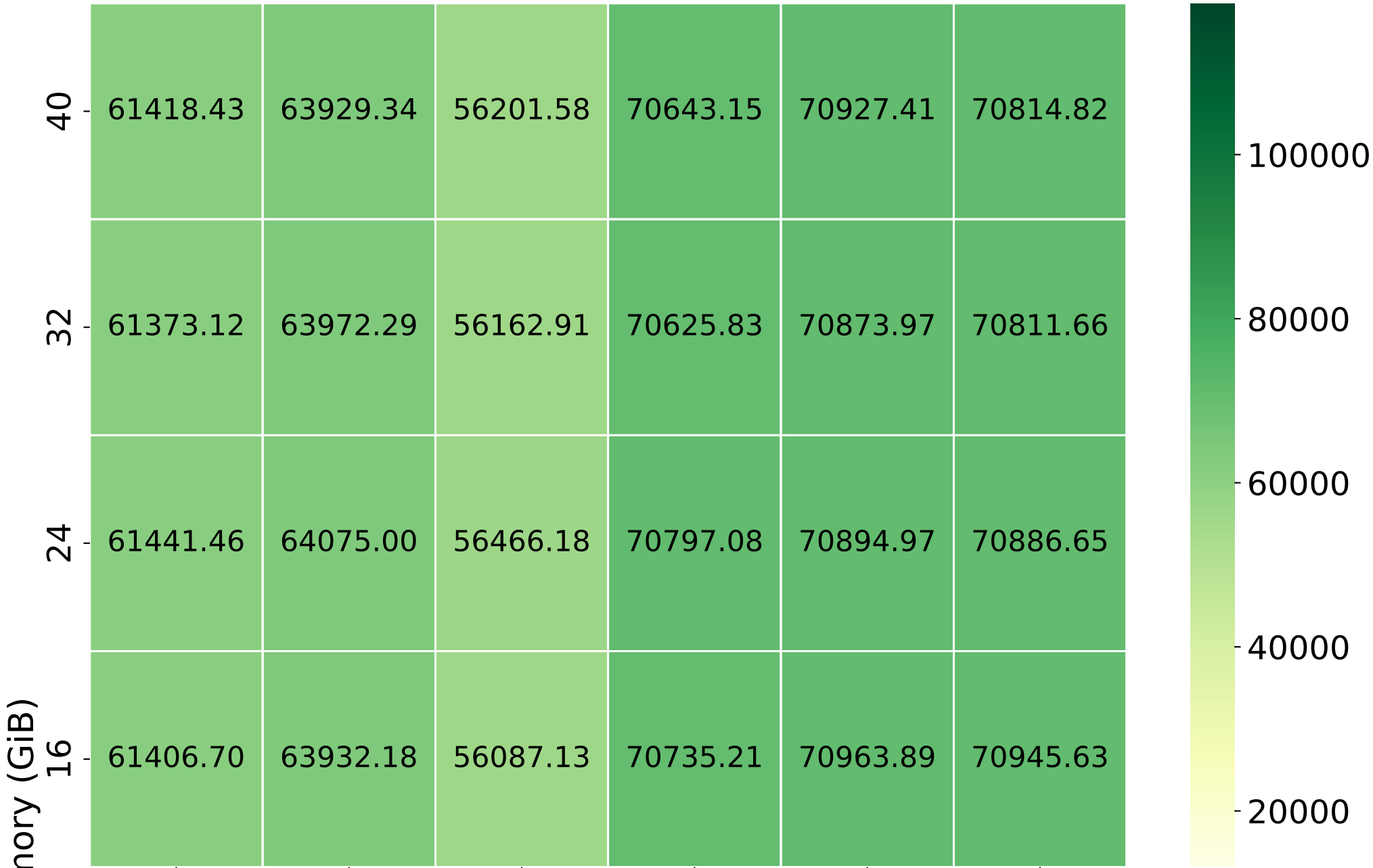
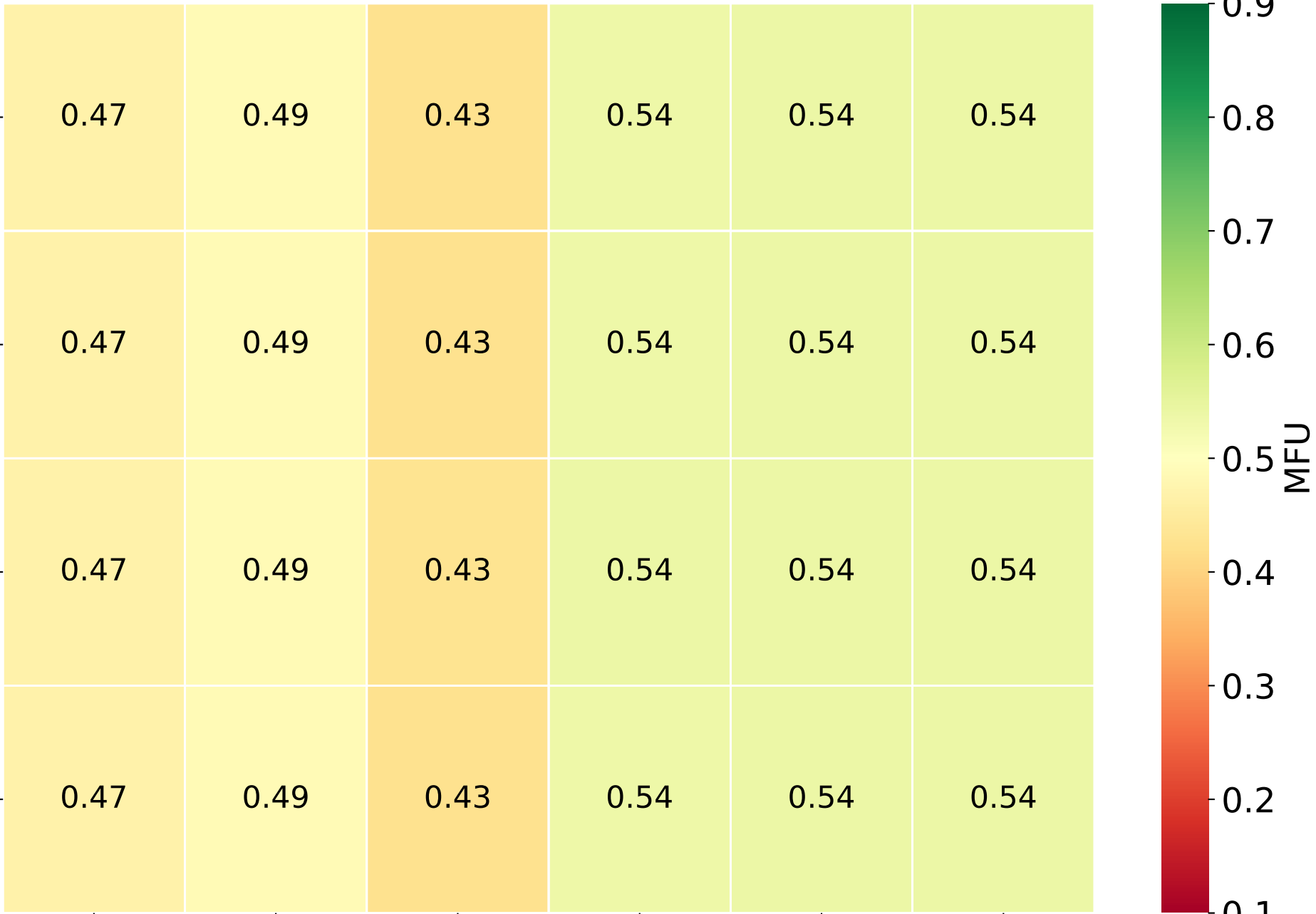


H100 Performance Report  
Model: 1B, Sequence Length: 512

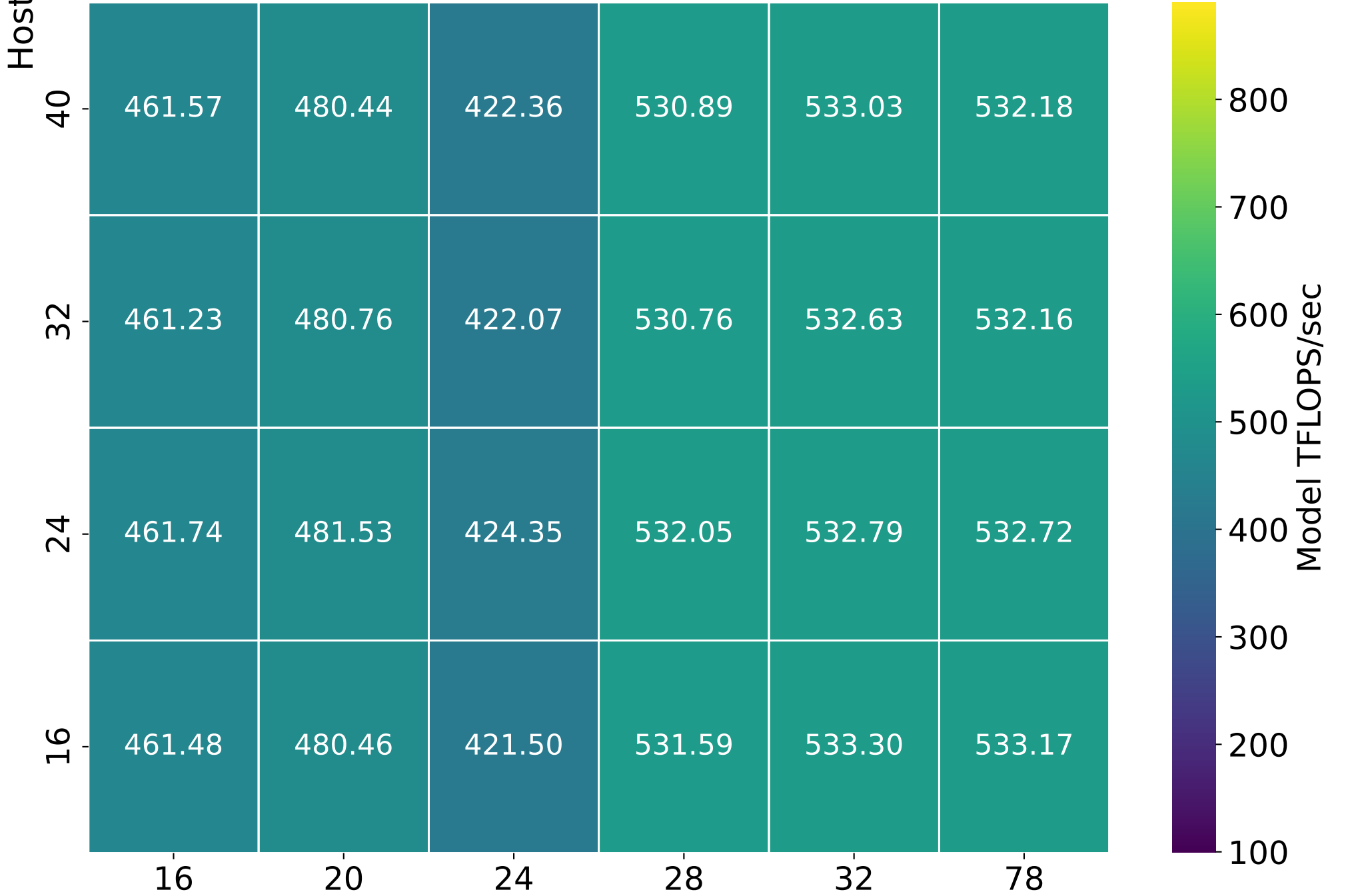
Tokens/sec



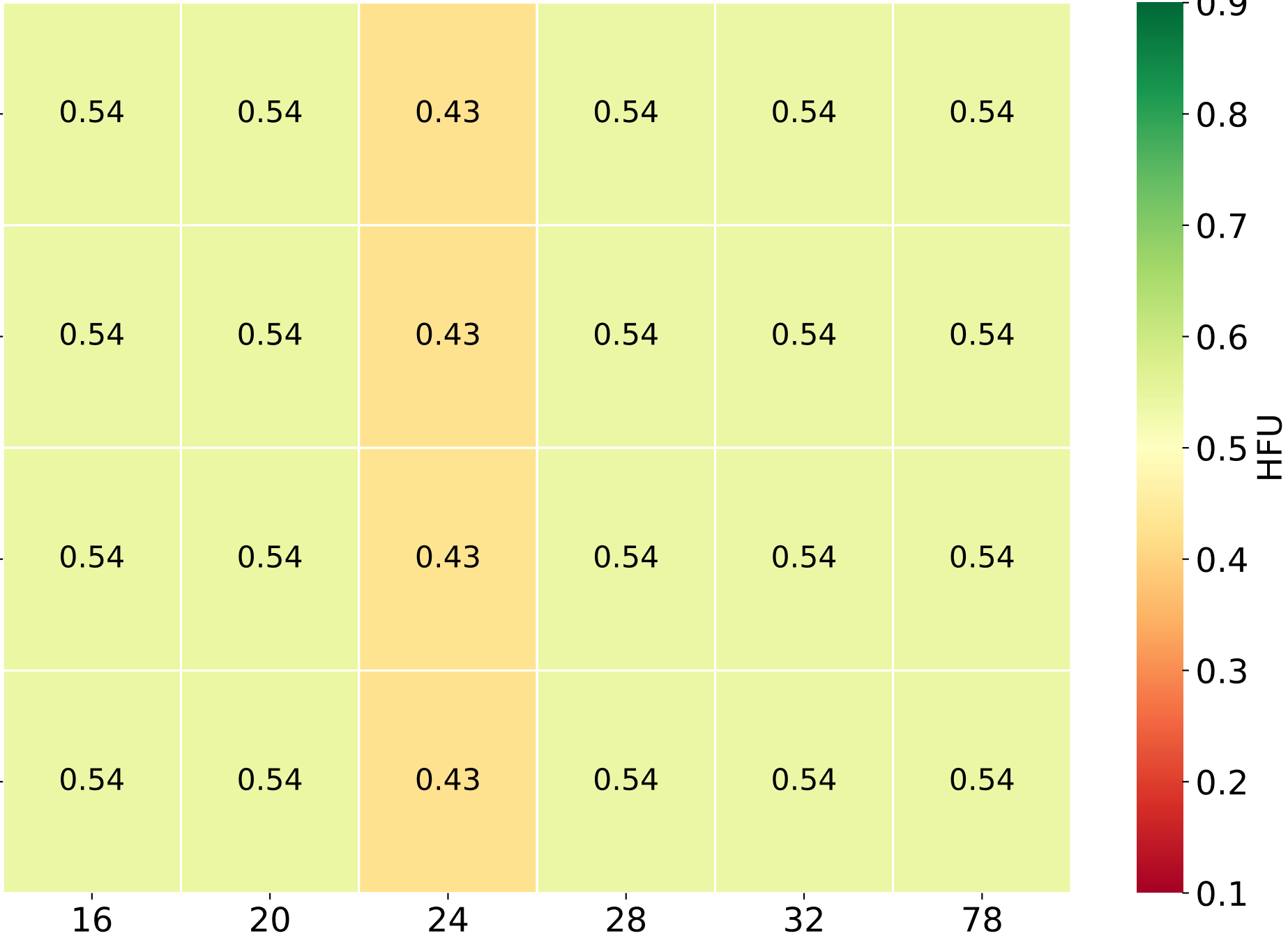
MFU



Model TFLOPS/sec



HFU



Device Memory (GiB)