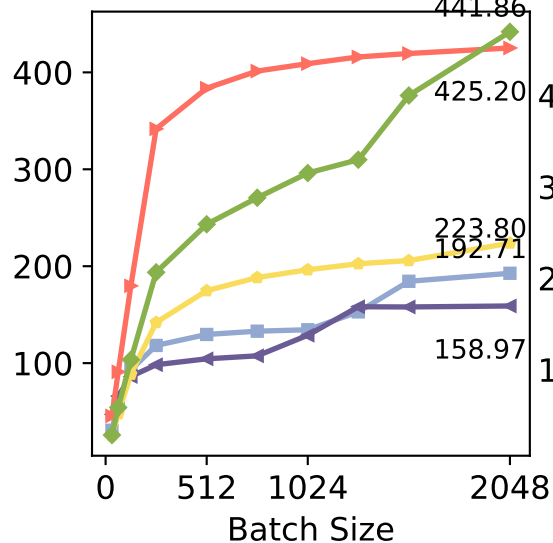
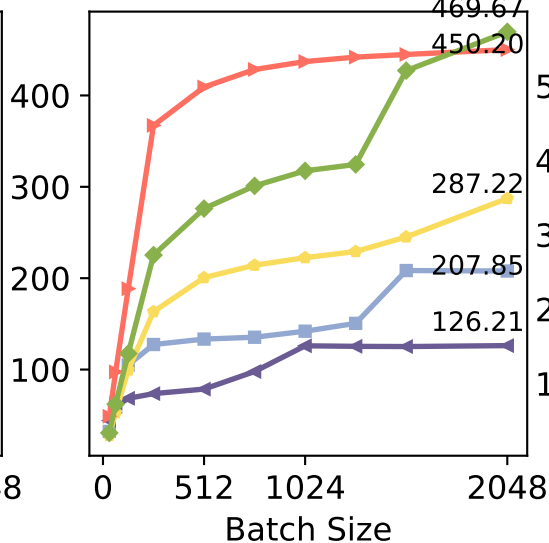


Llama-7B, 4-bit Quantization



Llama-13B, 4-bit Quantization



Llama-70B, 4-bit Quantization

