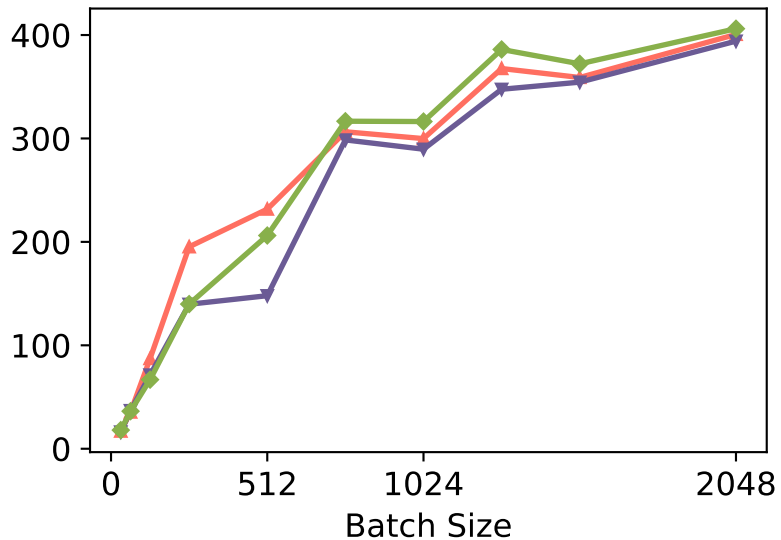
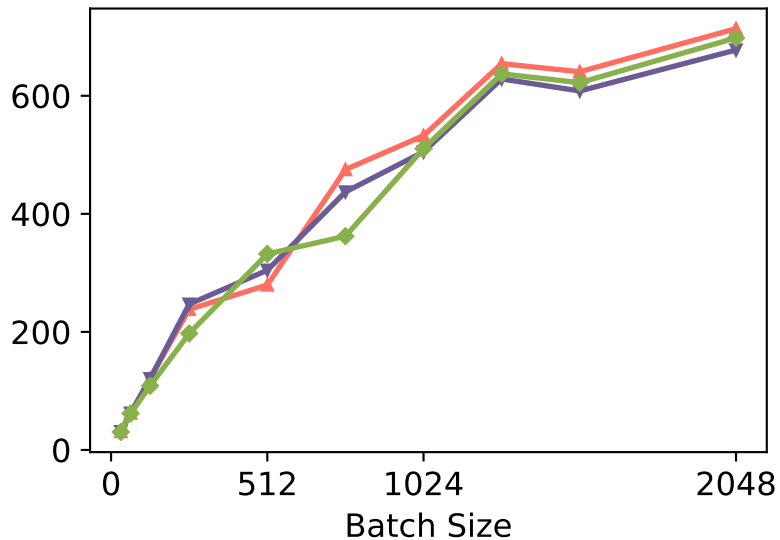


Llama-13B, MixQ (W8A8O16)



Llama-13B, MixQ (W4A4O16)



—▲— down_proj layer 0
 —▼— down_proj layer 20
 —◆— down_proj layer 39