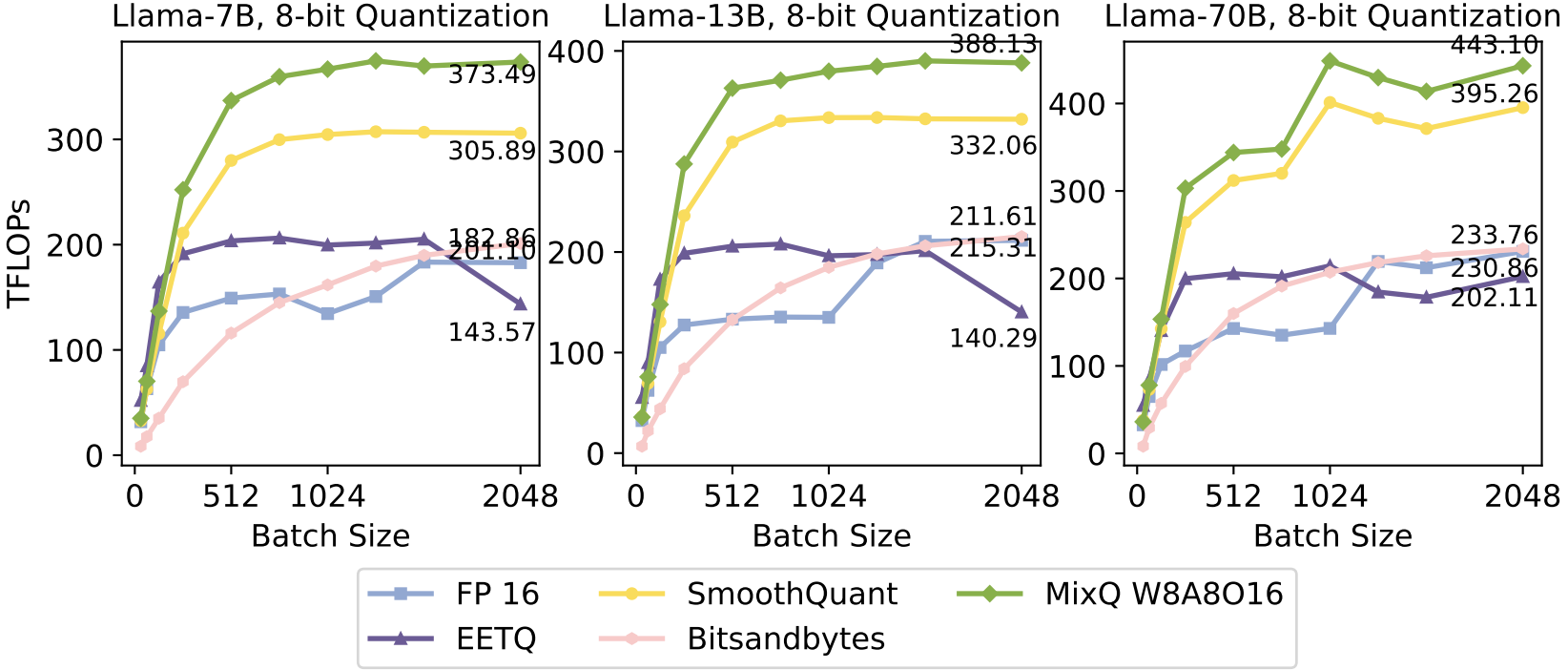
Llama-7B, 8-bit Quantization
Llama-13B, 8-bit Quantization
Llama-70B, 8-bit Quantization
TFLOPs
Batch Size
0
100
200
300
400
512
1024
2048
373.49
305.89
182.86
201.10
143.57
388.13
332.06
211.61
215.31
140.29
443.10
395.26
233.76
230.86
202.11
FP 16
EETQ
SmoothQuant
Bitsandbytes
MixQ W8A8O16