

Sparsity=0.95 Seq\_len=8192, num\_h=8

Latency(ms)

140  
120  
100  
80  
60  
40  
20  
0

batchsize=2

batchsize=8

