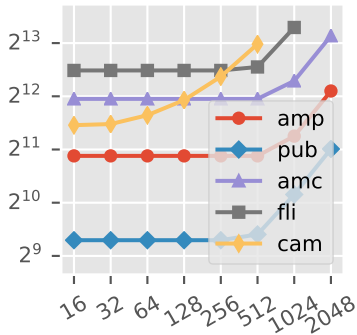
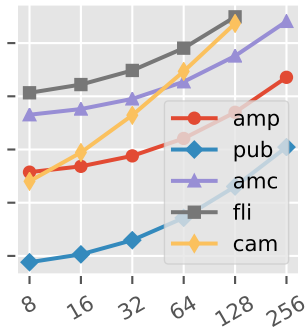


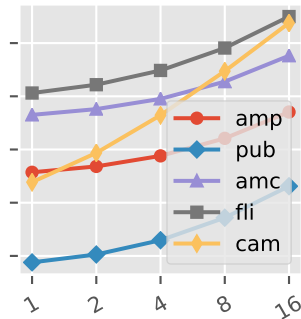
Inference Memory Usage (MB)



Dimension of Hidden Vectors $\dim(\mathbf{h}_x^1)$
(#Head=4, $d_a = d_v = d_m = 32$)



d_a, d_v, d_m
(#Head=4, $\dim(\mathbf{h}_x^1) = 64$)



#Head
($\dim(\mathbf{h}_x^1) = 64, d_a = d_v = d_m = 32$)