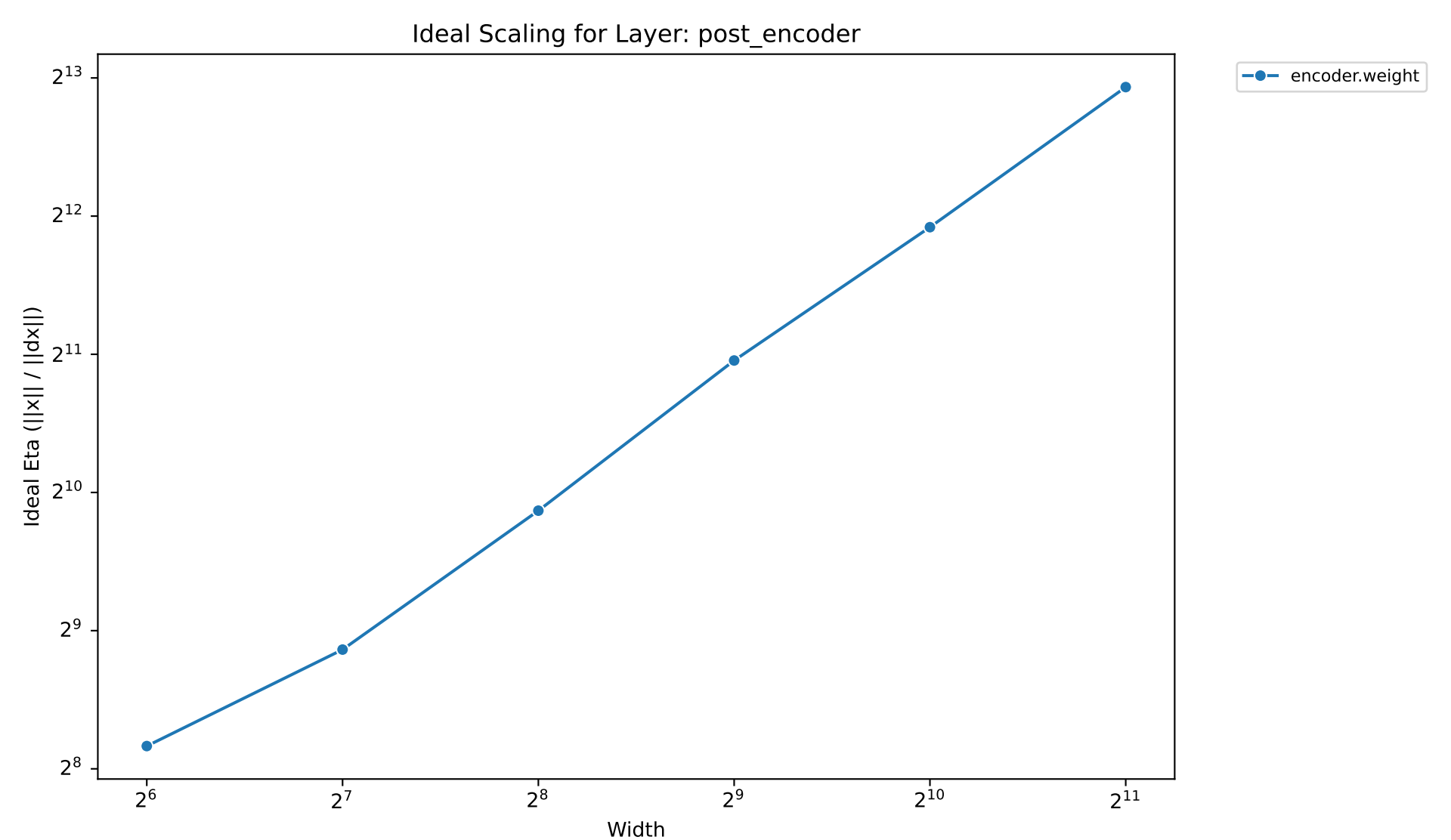
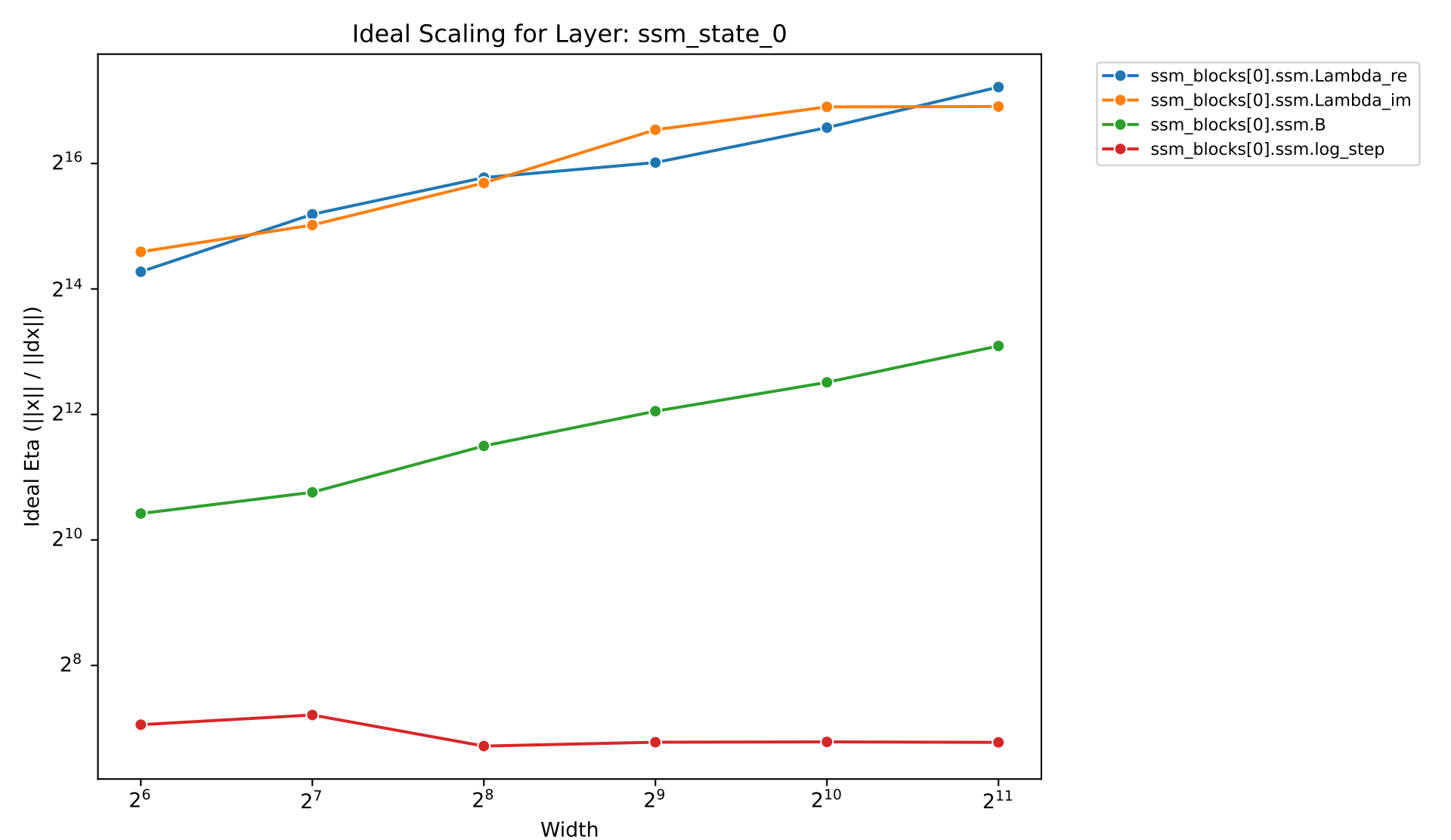
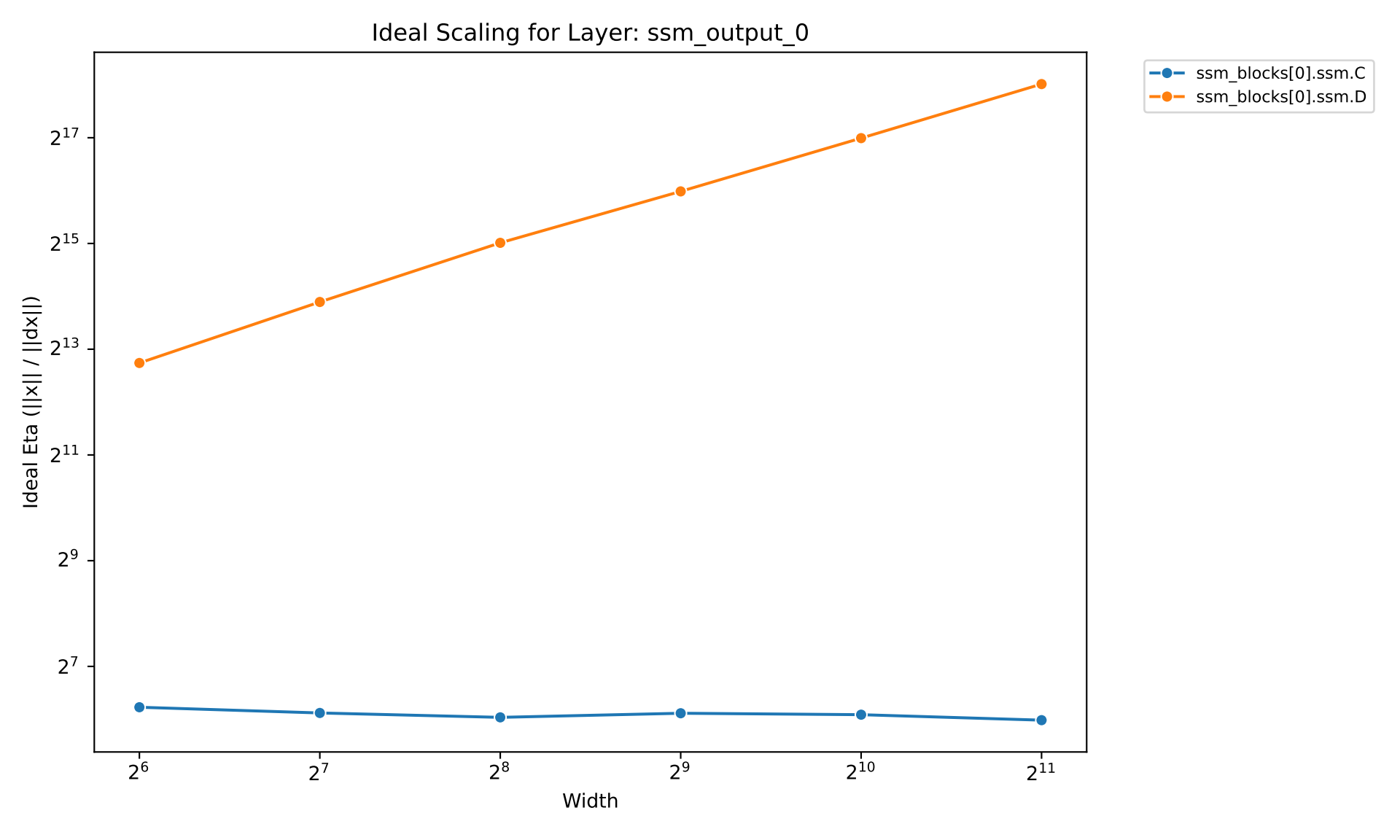
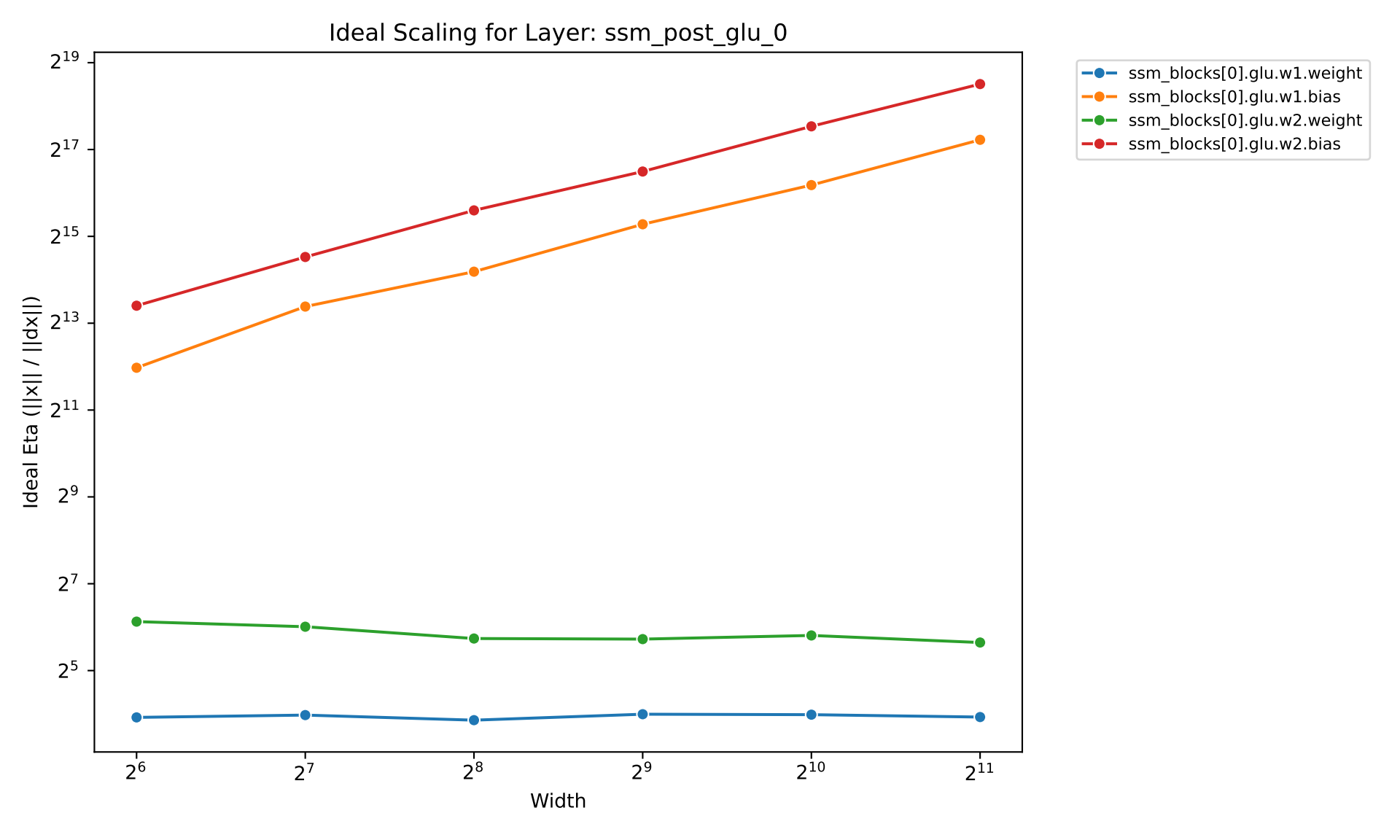
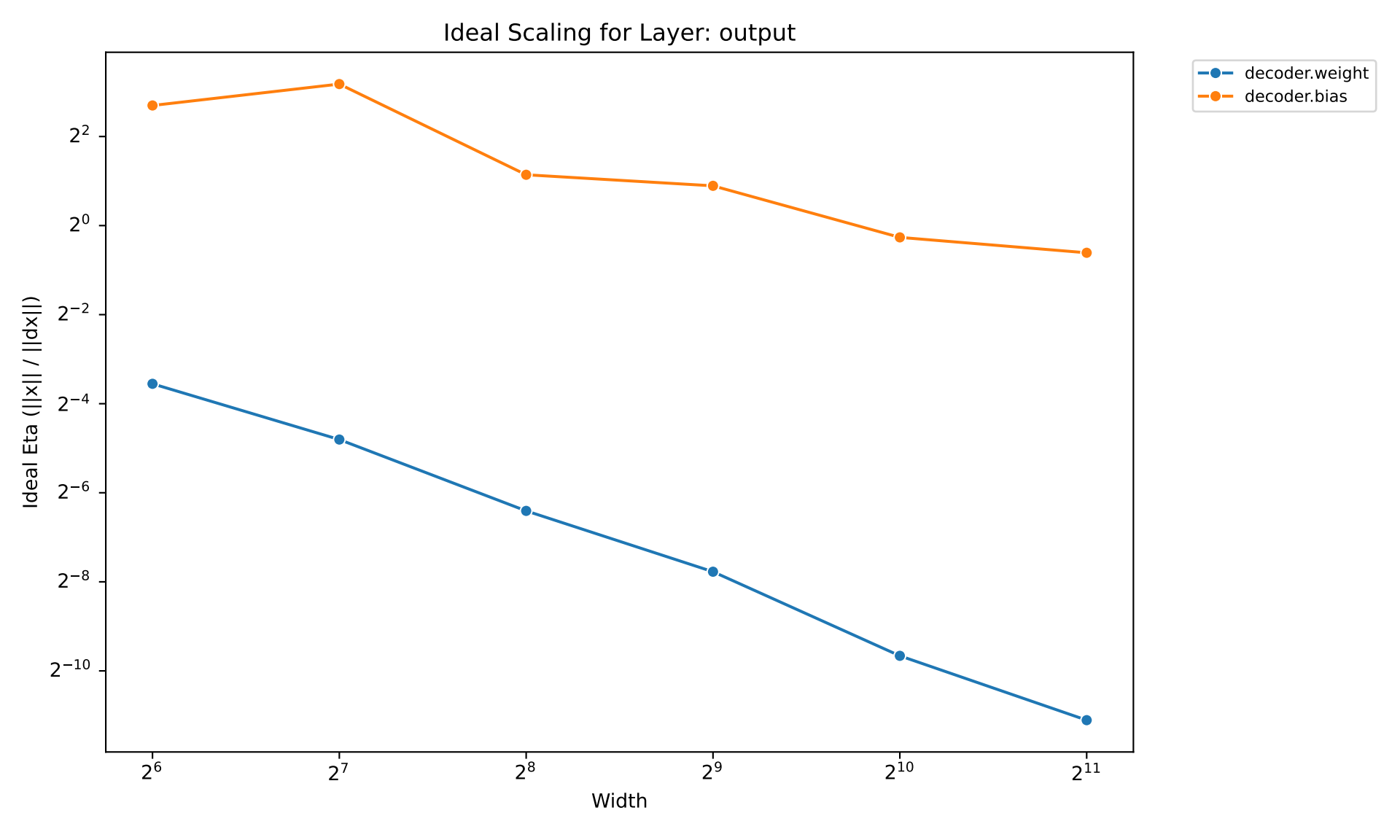

Ideal Scaling for Layer: post_encoder
encoder.weight
Ideal Eta (||x|| / ||dx||)
2^{13}
2^{12}
2^{11}
2^{10}
2^{9}
2^{8}
2^{6}
2^{7}
2^{8}
2^{9}
2^{10}
2^{11}
Width


Ideal Scaling for Layer: ssm_state_0
Ideal Eta (||x|| / ||dx||)
2^{16}
2^{14}
2^{12}
2^{10}
2^{8}
2^{6}
2^{7}
2^{8}
2^{9}
2^{10}
2^{11}
Width
ssm_blocks[0].ssm.Lambda_re
ssm_blocks[0].ssm.Lambda_im
ssm_blocks[0].ssm.B
ssm_blocks[0].ssm.log_step


Ideal Scaling for Layer: ssm_output_0
ssm_blocks[0].ssm.C
ssm_blocks[0].ssm.D
Ideal Eta (||x|| / ||dx||)
2^{17}
2^{15}
2^{13}
2^{11}
2^{9}
2^{7}
2^{6}
2^{7}
2^{8}
2^{9}
2^{10}
2^{11}
Width


Ideal Scaling for Layer: ssm_post_glu_0
Ideal Eta (||x|| / ||dx||)
Width
2^{19}
2^{17}
2^{15}
2^{13}
2^{11}
2^{9}
2^{7}
2^{5}
2^{6}
2^{7}
2^{8}
2^{9}
2^{10}
2^{11}
ssm_blocks[0].glu.w1.weight
ssm_blocks[0].glu.w1.bias
ssm_blocks[0].glu.w2.weight
ssm_blocks[0].glu.w2.bias


Ideal Scaling for Layer: output
decoder.weight
decoder.bias
Ideal Eta (||x|| / ||dx||)
2^{2}
2^{0}
2^{-2}
2^{-4}
2^{-6}
2^{-8}
2^{-10}
2^{6}
2^{7}
2^{8}
2^{9}
2^{10}
2^{11}
Width