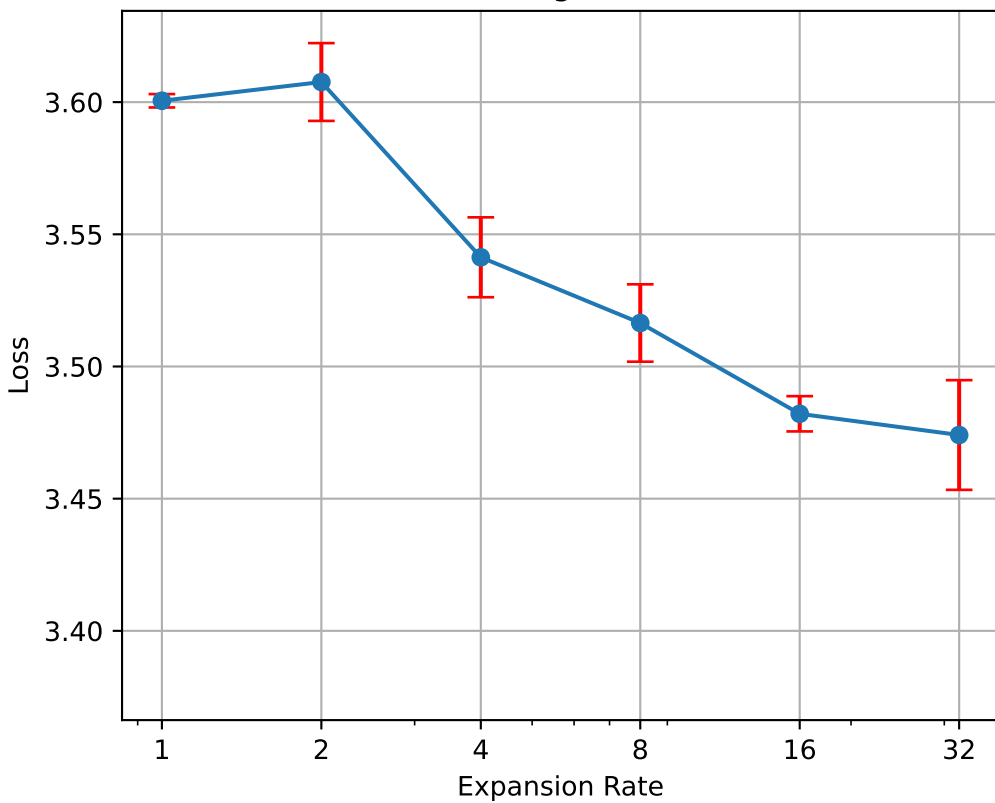
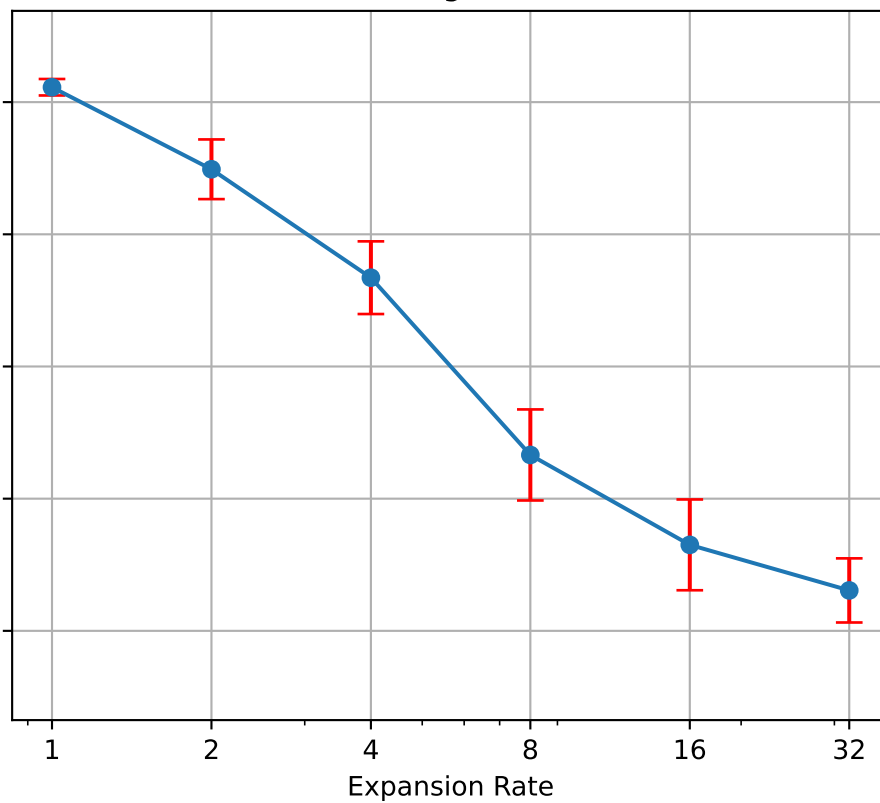


33M model, 6.5B tokens, ratio ~ 200 (token/active), #runs: 3

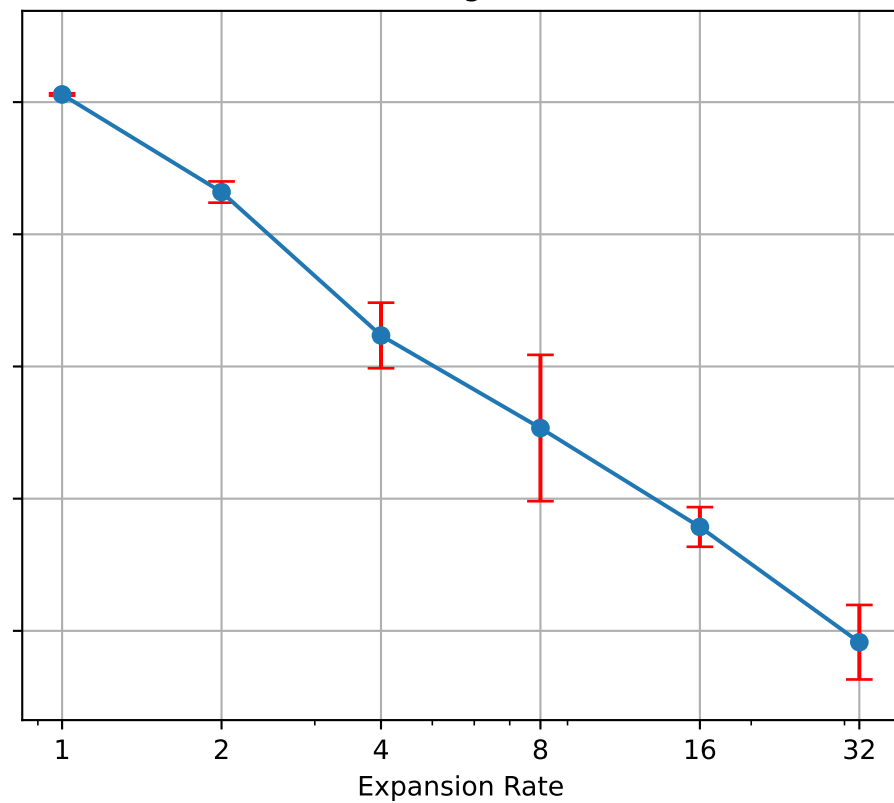
load balancing loss = 0.001



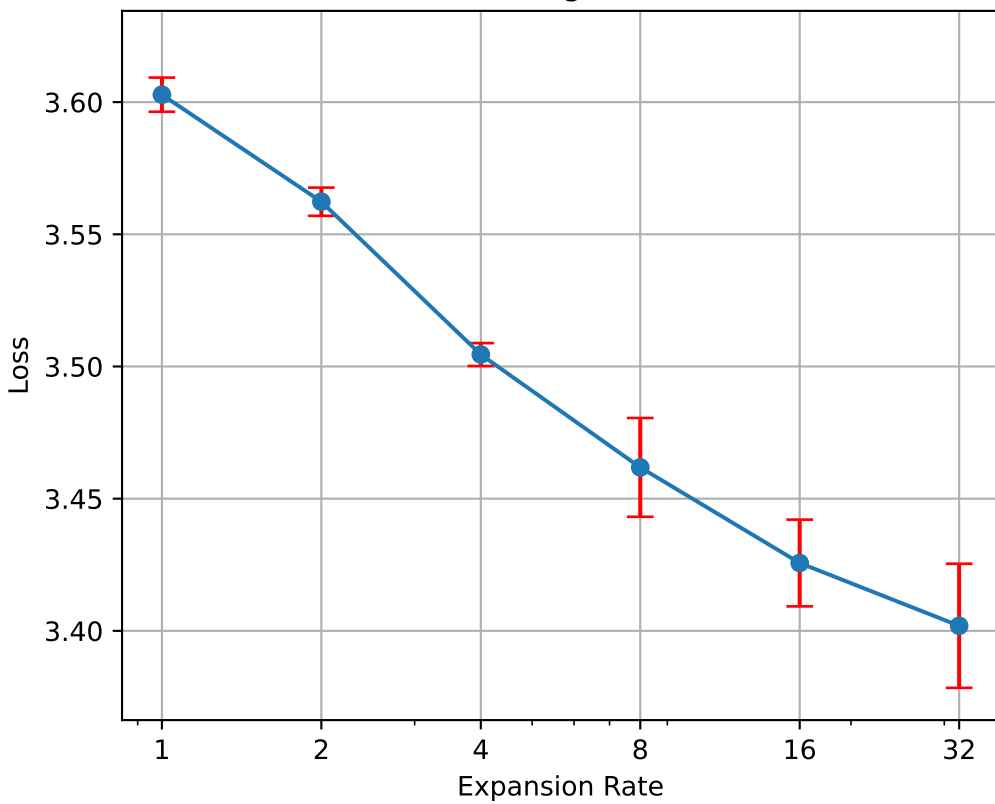
load balancing loss = 0.004



load balancing loss = 0.01



load balancing loss = 0.04



load balancing loss = 0.1

