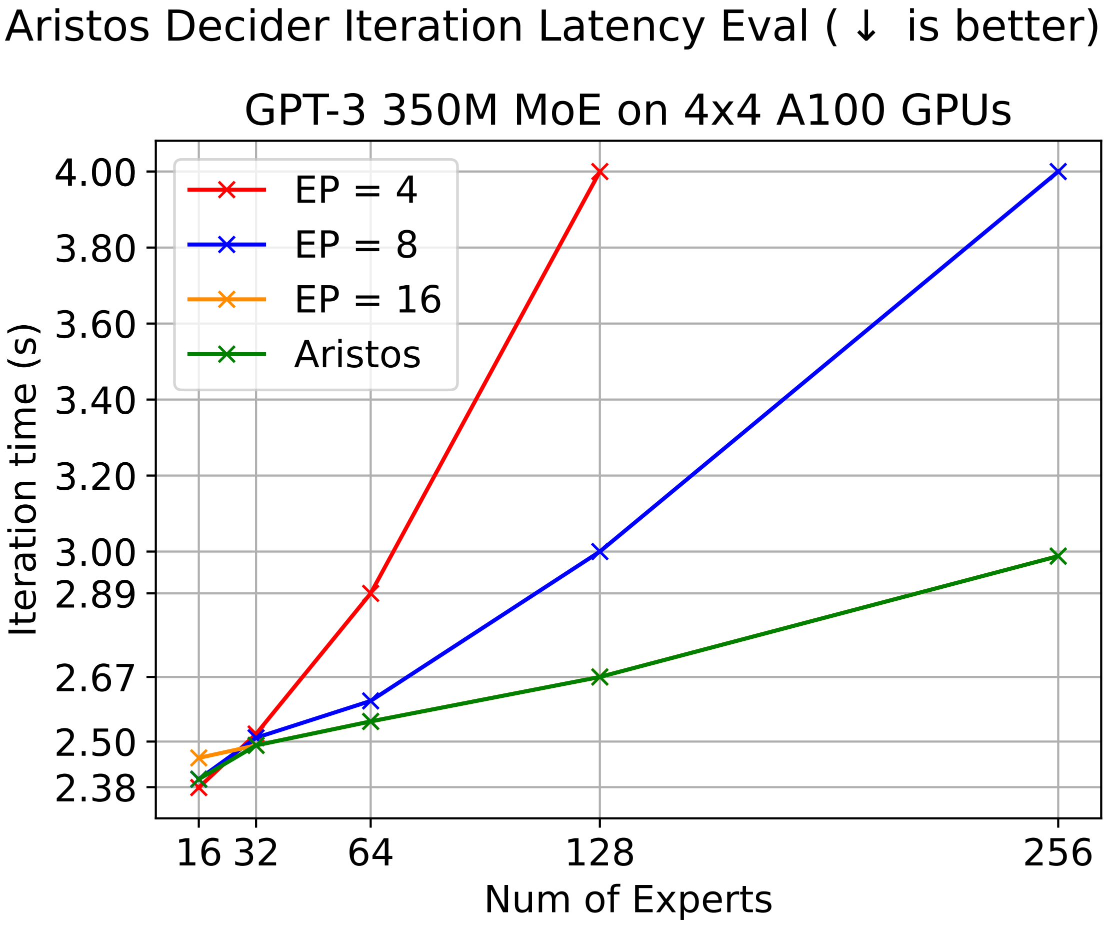

Aristos Decider Iteration Latency Eval (↓ is better)

GPT-3 350M MoE on 4x4 A100 GPUs