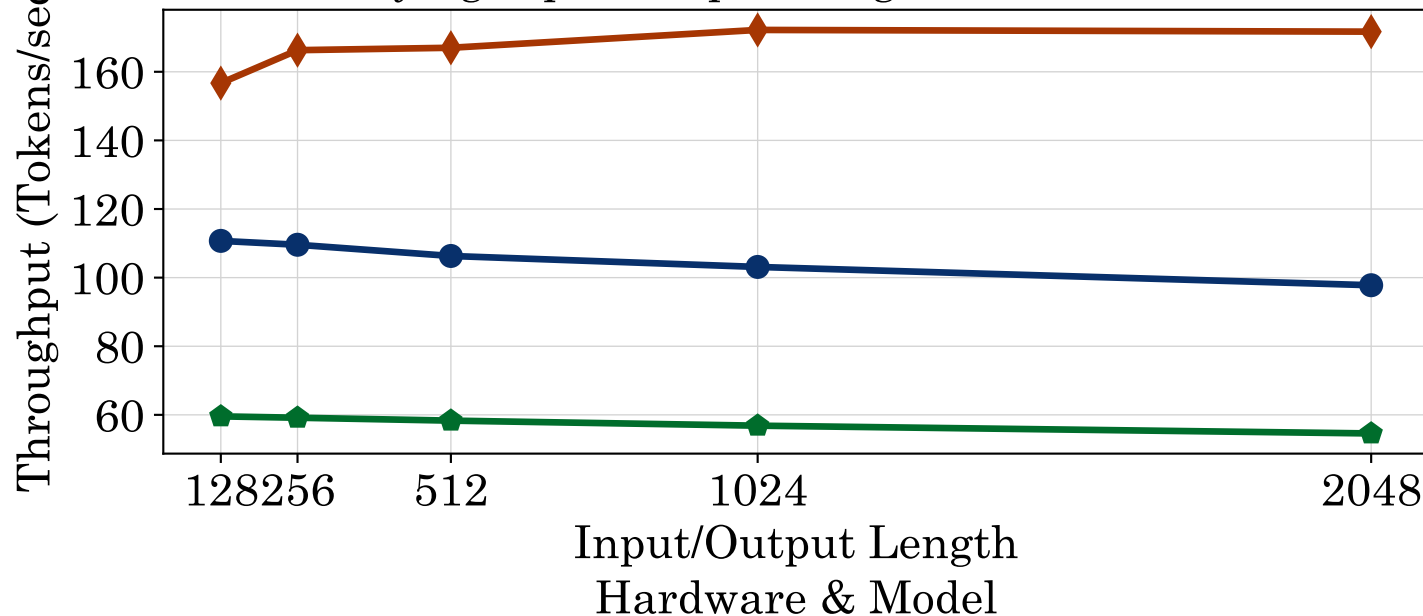


8 SN40L RDUs vs 4 H100 GPU vs 4 A100 GPU: 70B Models
for Varying Input/Output Length and Batch Size = 1



◆ SN40L LLaMA-3-70B ● H100 LLaMA-3-70B ◆ A100 LLaMA-3-70B