Performance for Model: 8B, Seq Len: 512 on H100 176 0.35 0.40 0.47 0.54 0.56 0.57 8.0 160 0.35 0.40 0.47 0.54 0.56 0.57 -0.7Host Memory (GB) 12 128 144 0.35 0.40 0.47 0.54 0.57 0.57 - 0.6 0.57 0.35 0.40 0.47 0.54 0.57 - 0.5 0.35 0.40 0.47 0.54 0.57 0.57 0.4 96 0.57 0.35 0.40 0.47 0.54 0.57 0.3 80 0.51 0.57 0.52 0.53 0.55 0.57 40 78 30 50 60 70 Device Memory (GB)