

Forward Latency | $T = 16K$ | $k = 2$ | 8 H100s | ↓ is better

