

Cumulative Error Rate

30
25
20
15
10 10^1 10^2 10^3 10^4

Training TFLOPS

- SM
- + Ens
- UMix
- ★ gEns
- ⋈ gMoE
- ▼ FF

 wait 5