Training 4-layer GPT models on 4xA100 80GB SXM, 500GB RAM PyTorch(FSDP,bf16-true) PyTorch(FSDP,tf32) NNTile(tf32) Fflops/s 27.8 2 8 8 1 2 19.2 1.5 22.15 22.13 3.6 4.5 7.0 6.7 Model size, billions