3 GB	Fragmentation Overhead (Variable)
6 GB	Temporary Buffers (fp32)
8 GB	Activations (with checkpointing)
6 GB	Optimizer States (fp32 Variance)
6 GB	Optimizer States (fp32 Momentum)
6 GB	Optimizer States (fp32 Parameters)
3 GB	Gradients (fp16)
3 GB	Parameters (fp16)