



### Memory Formula

$$\begin{aligned}\text{Total} &= 2P(\text{FP16 params + grads}) + 2P(\text{FP32 master}) + 2P(\text{FP32 states}) \\ &= 6P \text{ bytes} = 6 \times 110M \times 4 \text{ bytes} \approx \$1.76 \text{ GB}\end{aligned}$$

Mixed precision reduces memory by ~50% while maintaining numerical stability