State Tensor Memory Usage Increasing with Model Size Compared to SGD-Boost +73.81 GB AdamW Prodigy Adam-mini SGD-Boost +24.33 GB +2.06 GB +17.79+5.93 GB 0 +0.00 6 6 Model Size (Billions of Parameters)

100

80

60

40

20

State Tensor Memory (GB)