BPB vs Training FLOPs at Compute Optimal Ratio (Entropy Patching) BLT Entropy ps=4 0.86 BLT Entropy ps=8 LLaMA 2 BPE 0.84 LLaMA 3 BPE Megabyte++ ps=4Bits-per-byte (BPB) 0.0 0.7 0.85 0.9 0.9 Megabyte++ ps=6 0.74 0.72 10²¹ 10²²

Total Training FLOPs