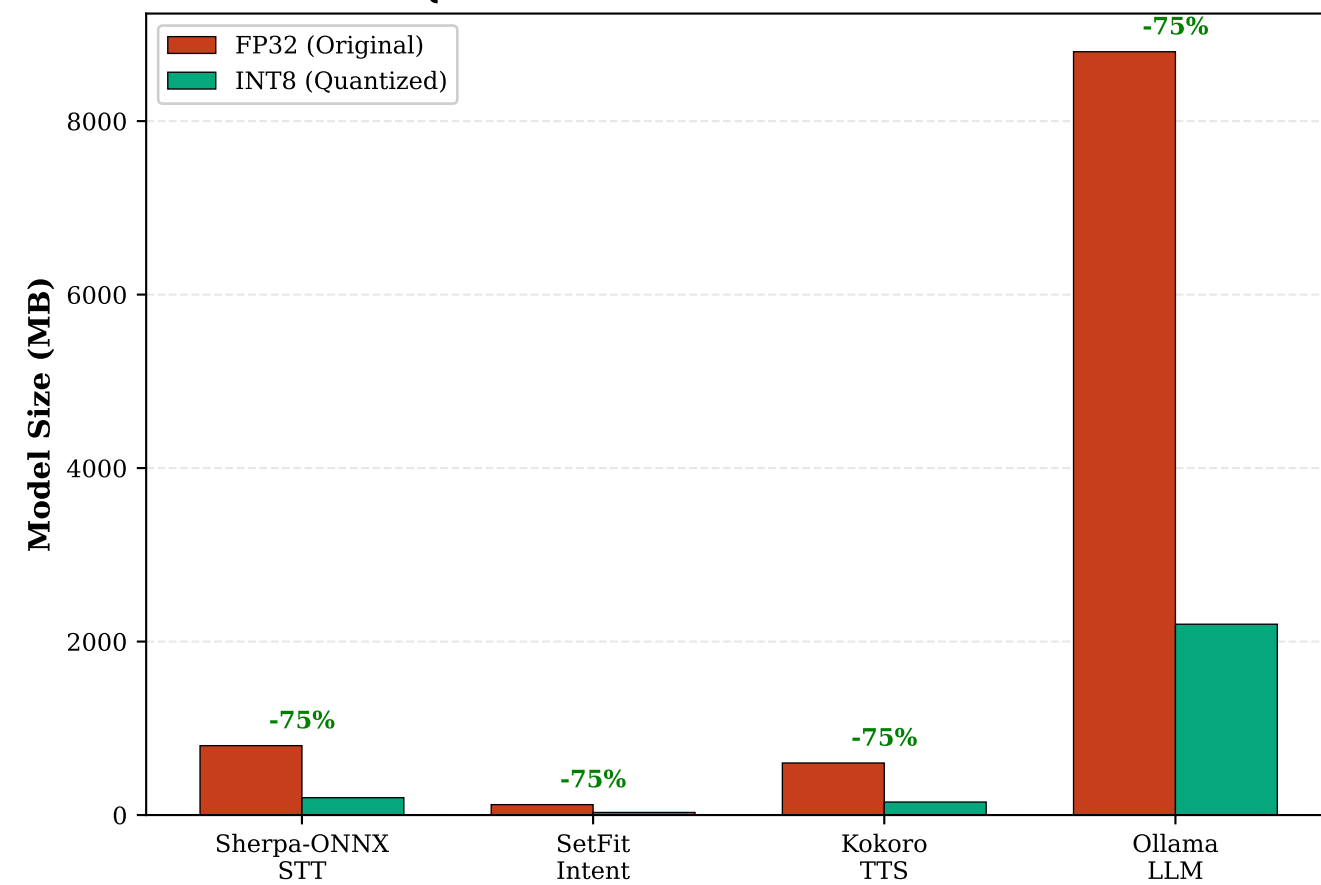
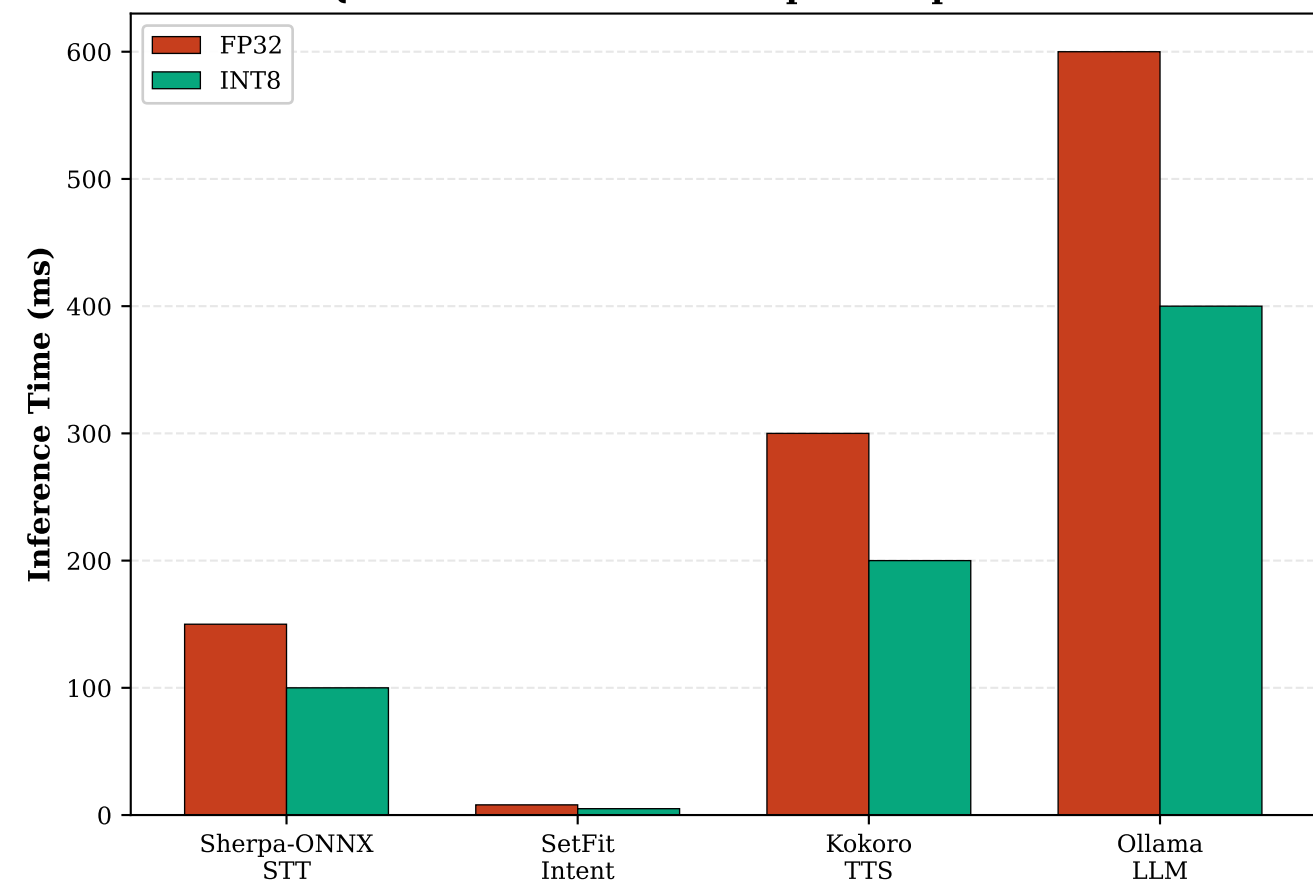


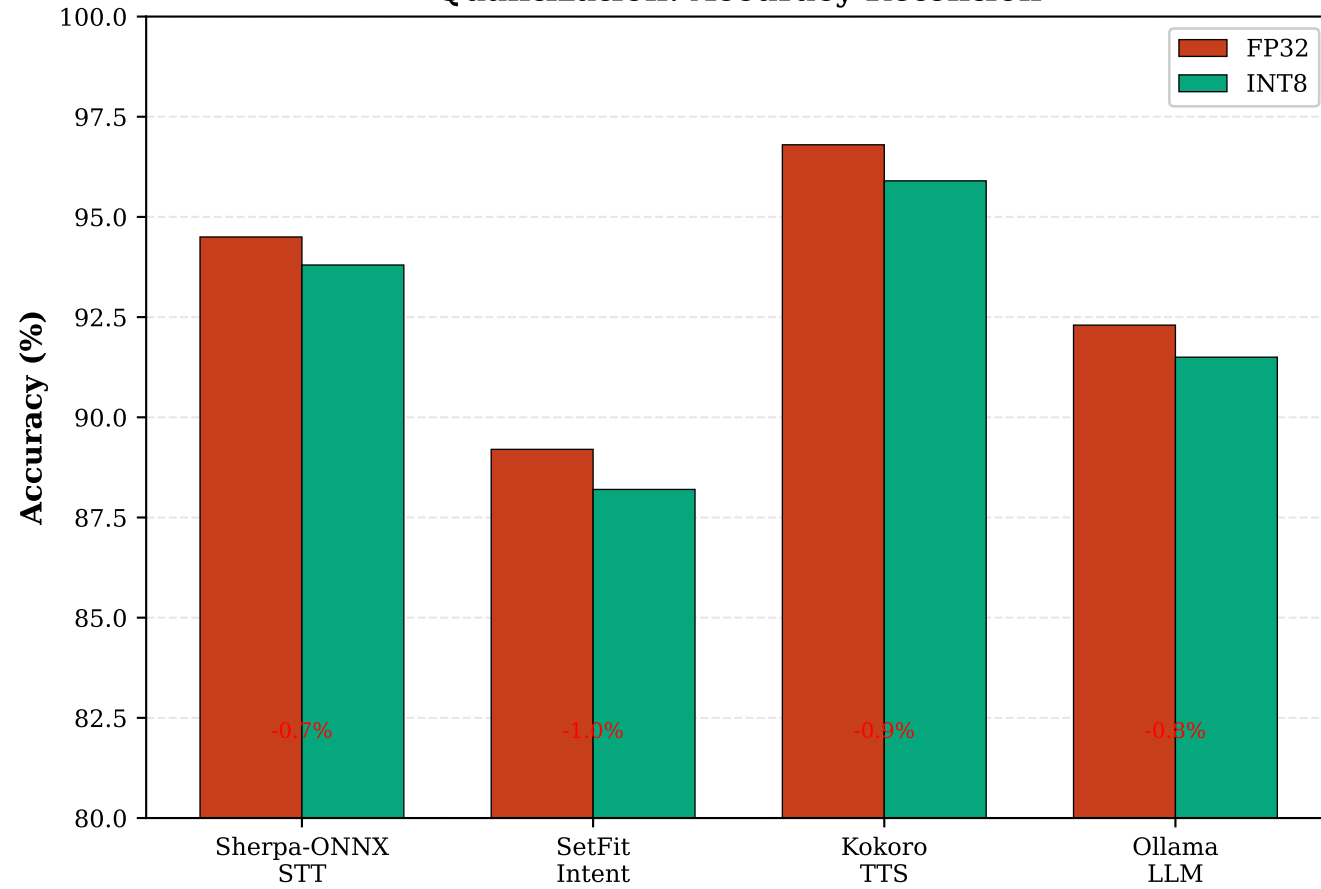
Quantization: Model Size Reduction



Quantization: Inference Speed Improvement



Quantization: Accuracy Retention



Quality-Size Trade-off Curves

