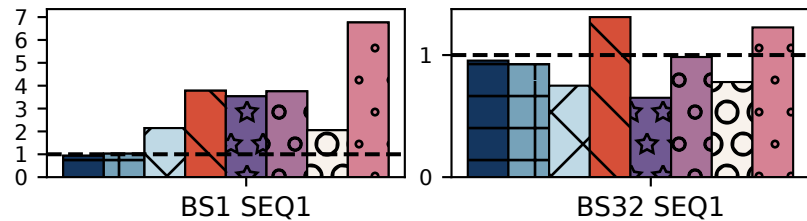
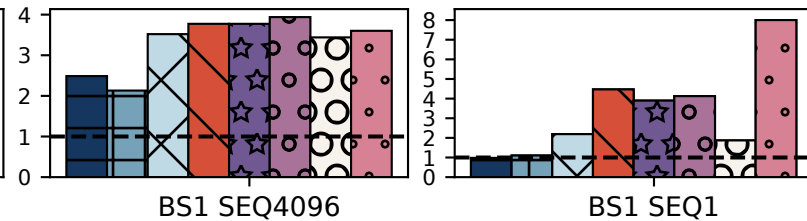


Speedup vs. Welder



(a) LLAMA



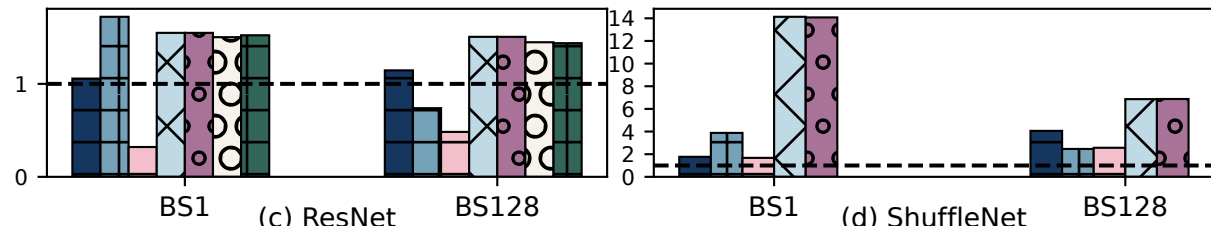
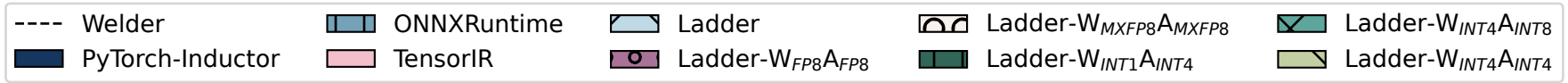
BS1 SEQ1

BS32 SEQ1

BS1 SEQ4096

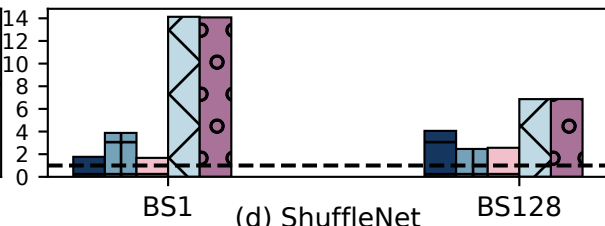
(b) BLOOM

Speedup vs. Welder



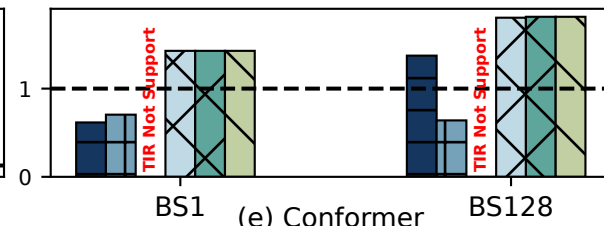
(c) ResNet

BS128



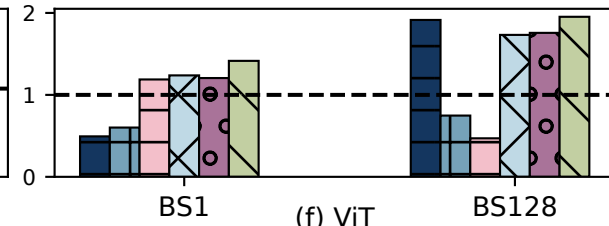
(d) ShuffleNet

BS128



(e) Conformer

BS128



(f) ViT

BS128