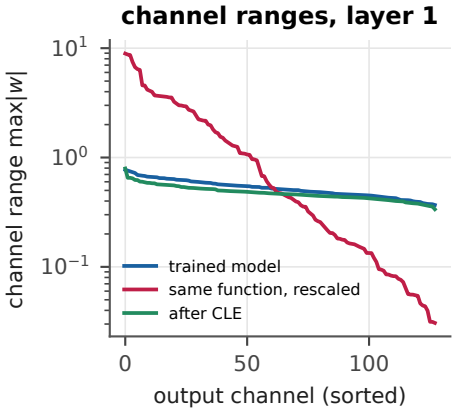
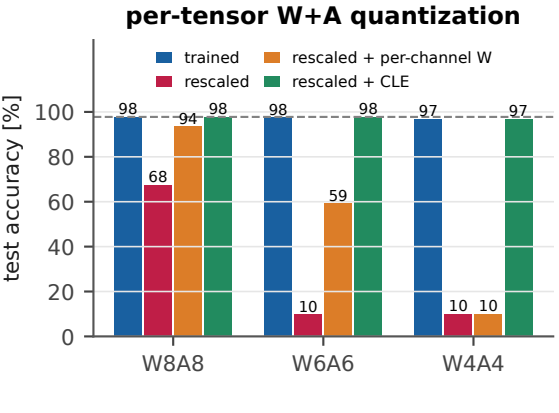

channel ranges, layer 1
channel range max|w|
10^1
10^0
10^{-1}
trained model
same function, rescaled
after CLE
0
50
100
output channel (sorted)


per-tensor W+A quantization
test accuracy [%]
trained
rescaled
rescaled + per-channel W
rescaled + CLE
0
20
40
60
80
100
98
68
94
98
98
10
59
98
97
10
10
97
W8A8
W6A6
W4A4