

# Speedup of GEMM $W_{INT8}A_{INT8}$ on A100

Speedup vs cuBLAS- $W_{INT8}A_{INT8}$

