

Speedup of GEMV $W_{FP16}A_{FP16}$ on A100

