

Scaling Speedup of BitNET INT8xINT2 on A100

