Speedup of GEMM on GTX3090 (nf4)

