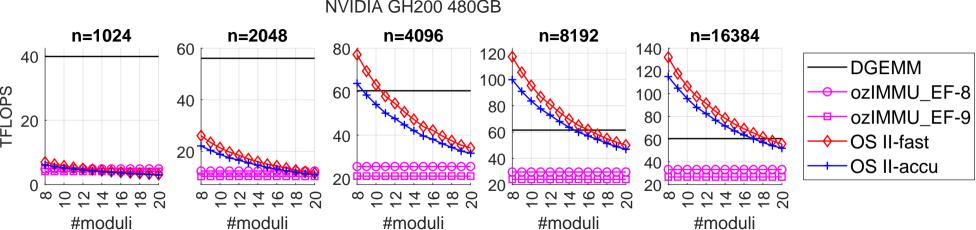
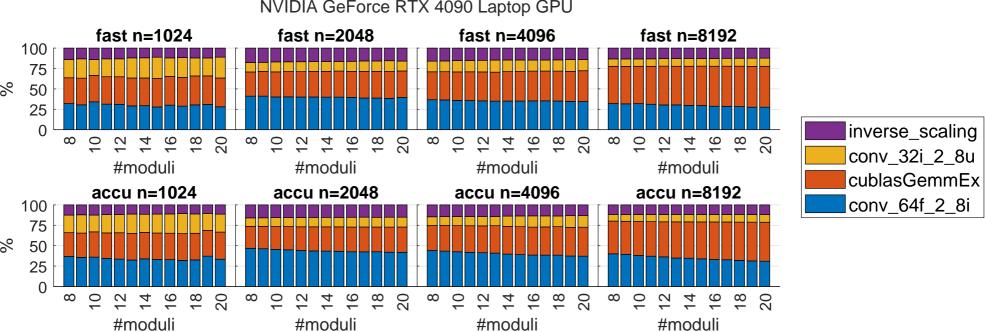
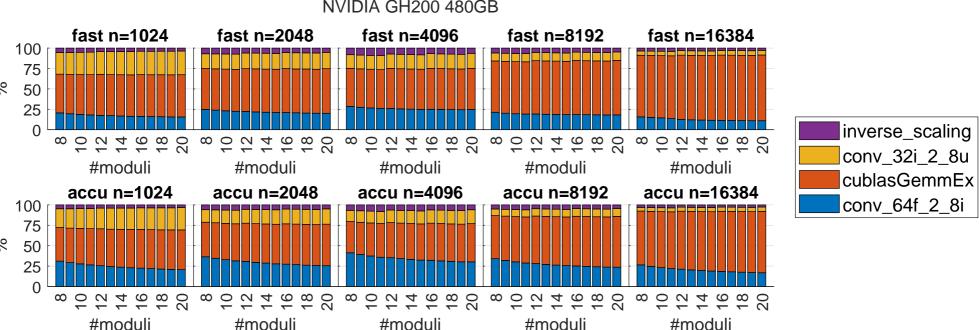


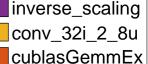
NVIDIA A100 SXM4 80GB n=1024 n=2048 n=4096 n=8192 n=16384 30 🔨 **DGEMM** ozIMMU EF-8 20 ──ozIMMU EF-9 → OS II-fast OS II-accu #moduli #moduli #moduli #moduli #moduli

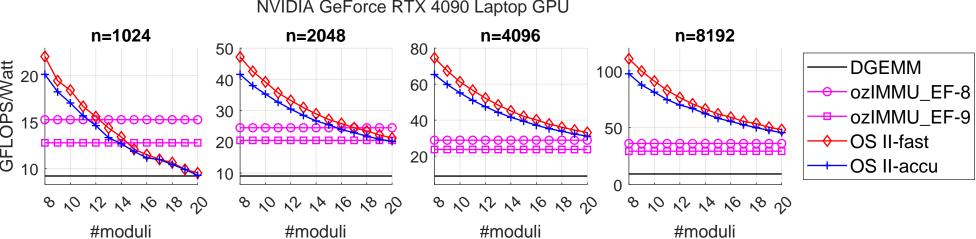


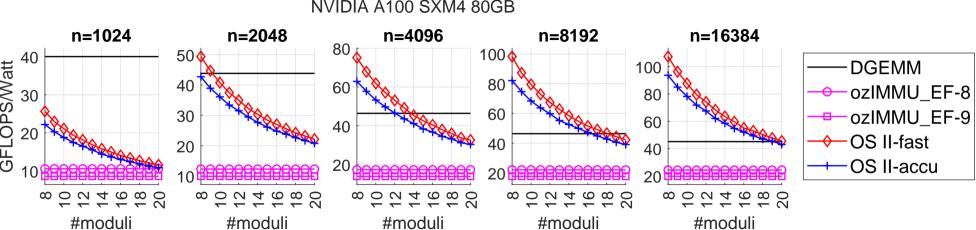


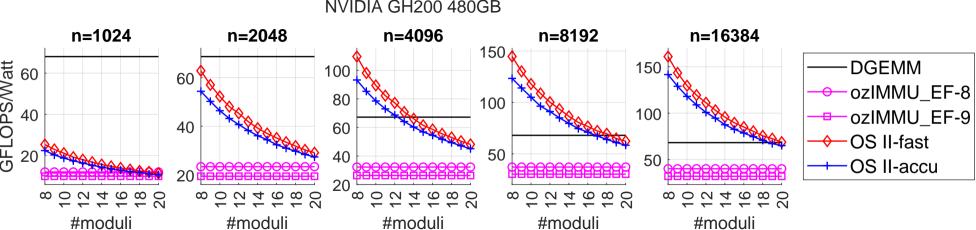
NVIDIA A100 SXM4 80GB fast n=1024 fast n=2048 fast n=4096 fast n=8192 fast n=16384 100 75 50 25 inverse_scaling conv_32i_2_8u #moduli #moduli #moduli #moduli #moduli cublasGemmEx accu n=1024 accu n=2048 accu n=4096 accu n=8192 accu n=16384 conv_64f_2_8i 100 75 50 25 #moduli #moduli #moduli #moduli #moduli

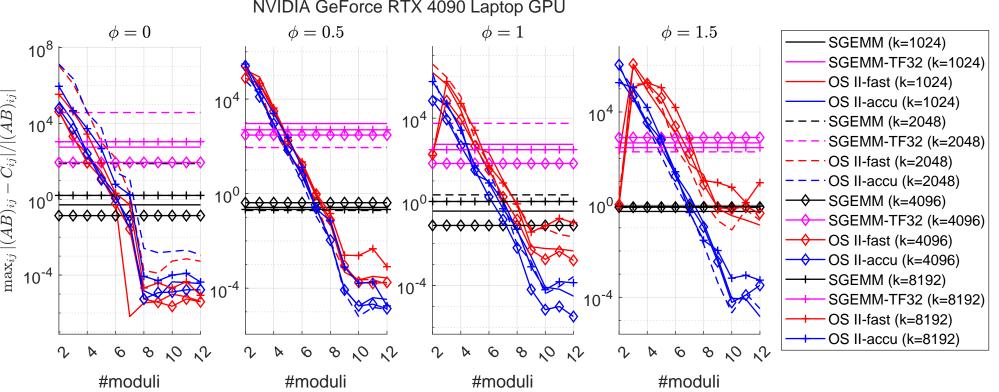


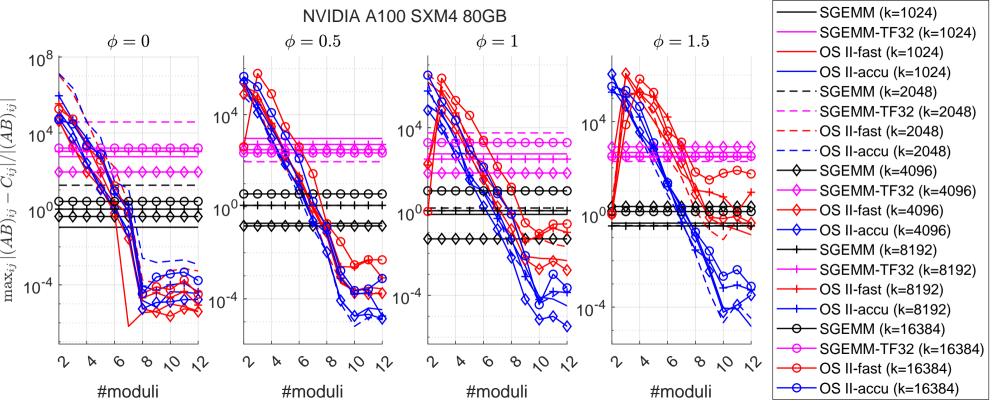


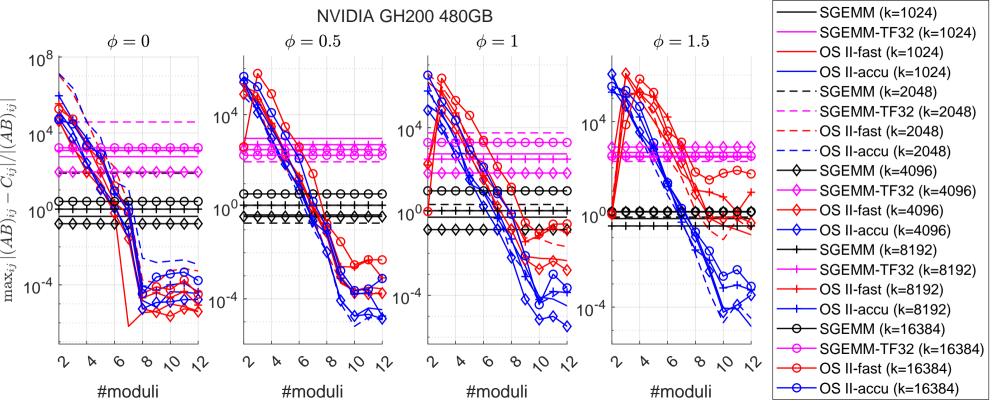


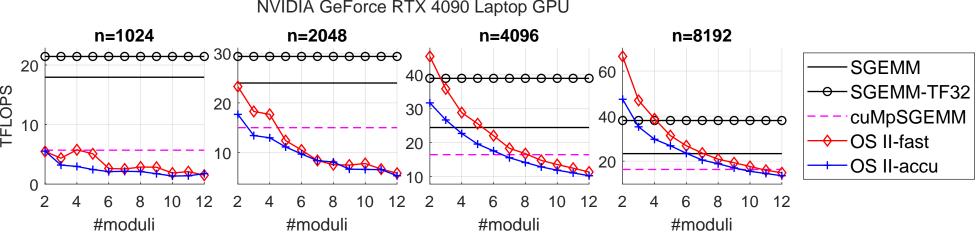


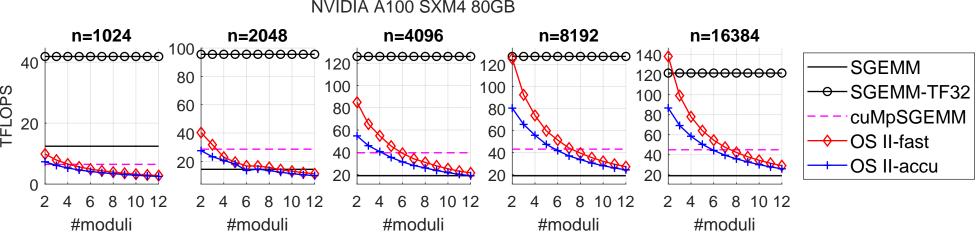


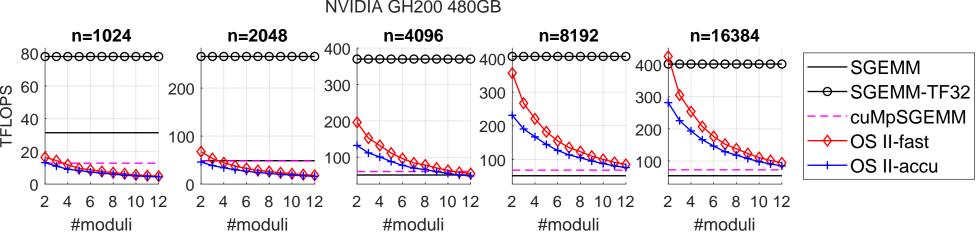


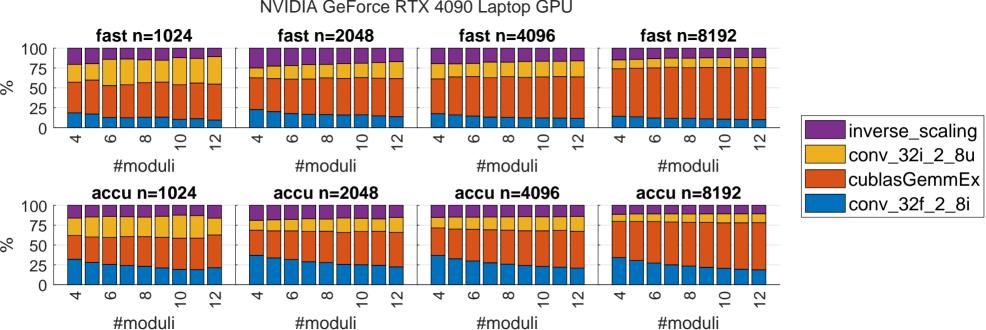


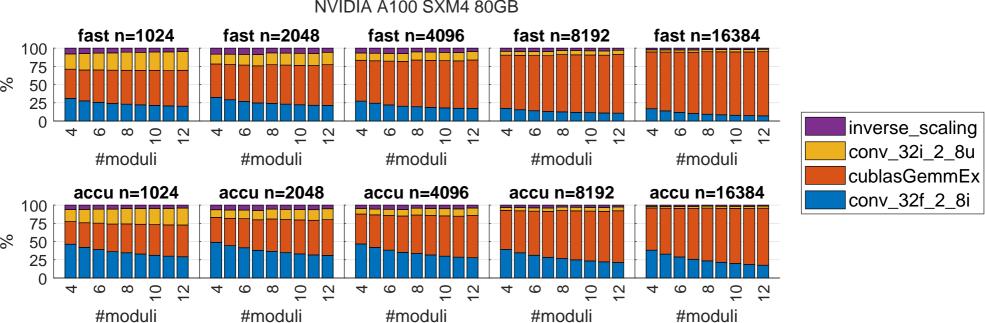


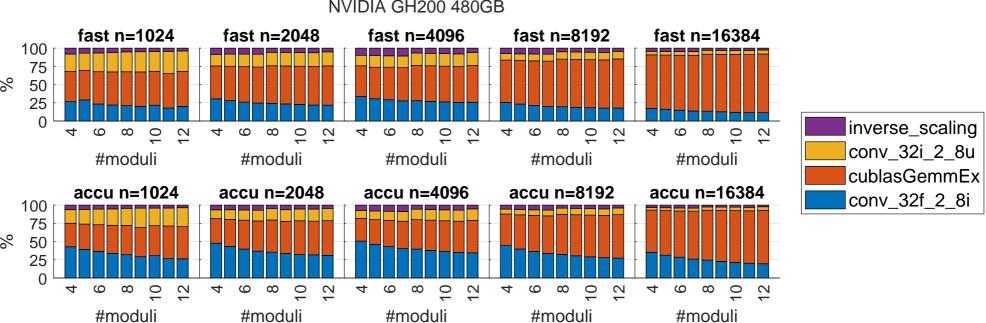


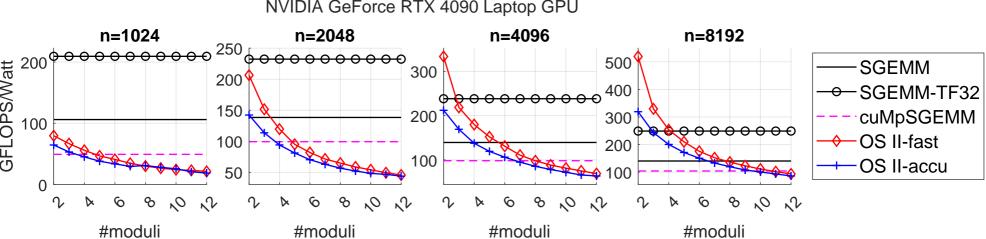


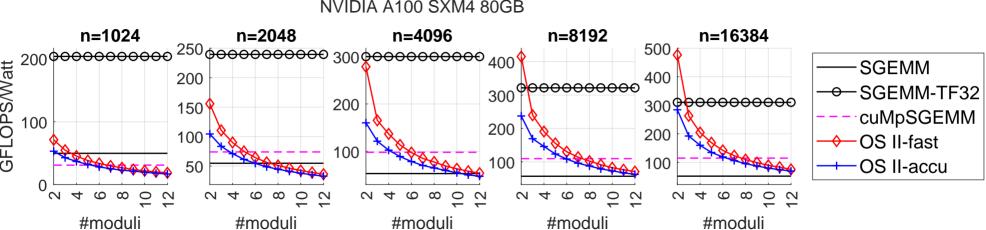












NVIDIA GH200 480GB n=1024 n=2048n = 4096n=8192 n=16384 000000000 8000 SGEMM 400 ^{(*} 0000000000 600 → SGEMM-TF32 200 300 400 000000000 300 cuMpSGEMM 400 200 100 OS II-fast 200 200 200 100 OS II-accu 100 \sim #moduli #moduli #moduli #moduli #moduli