With MIG GH is less efficient than a100

Avg bandwidth for GH with MIG and a100 at different batch sizes

GH_with_MIG_bandwidth (GB/s)   a100_bandwidth (GB/s)
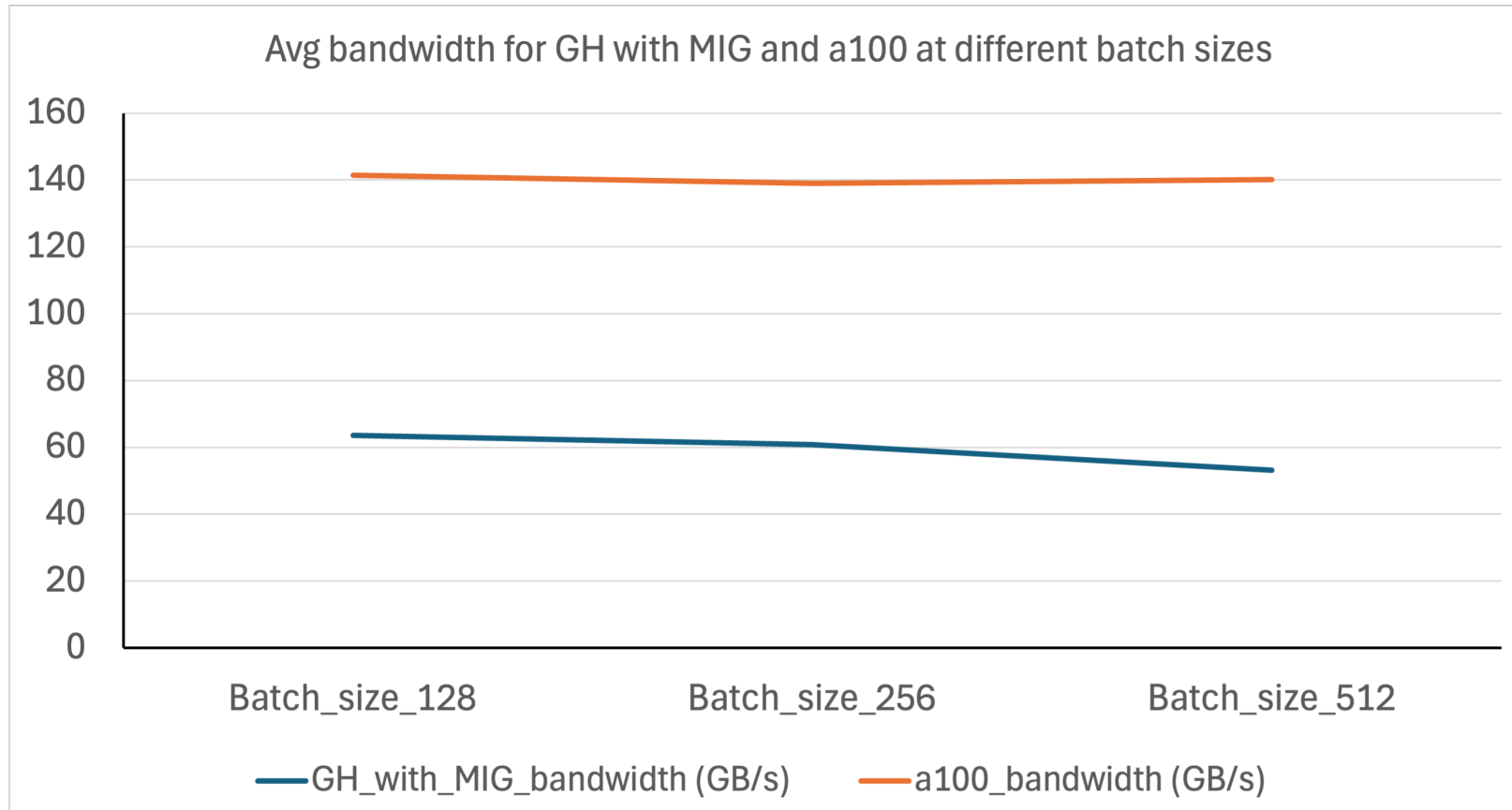
With MIG GH is slower than a100

Kernel opt time for GH with MIG and a100 at different batch sizes
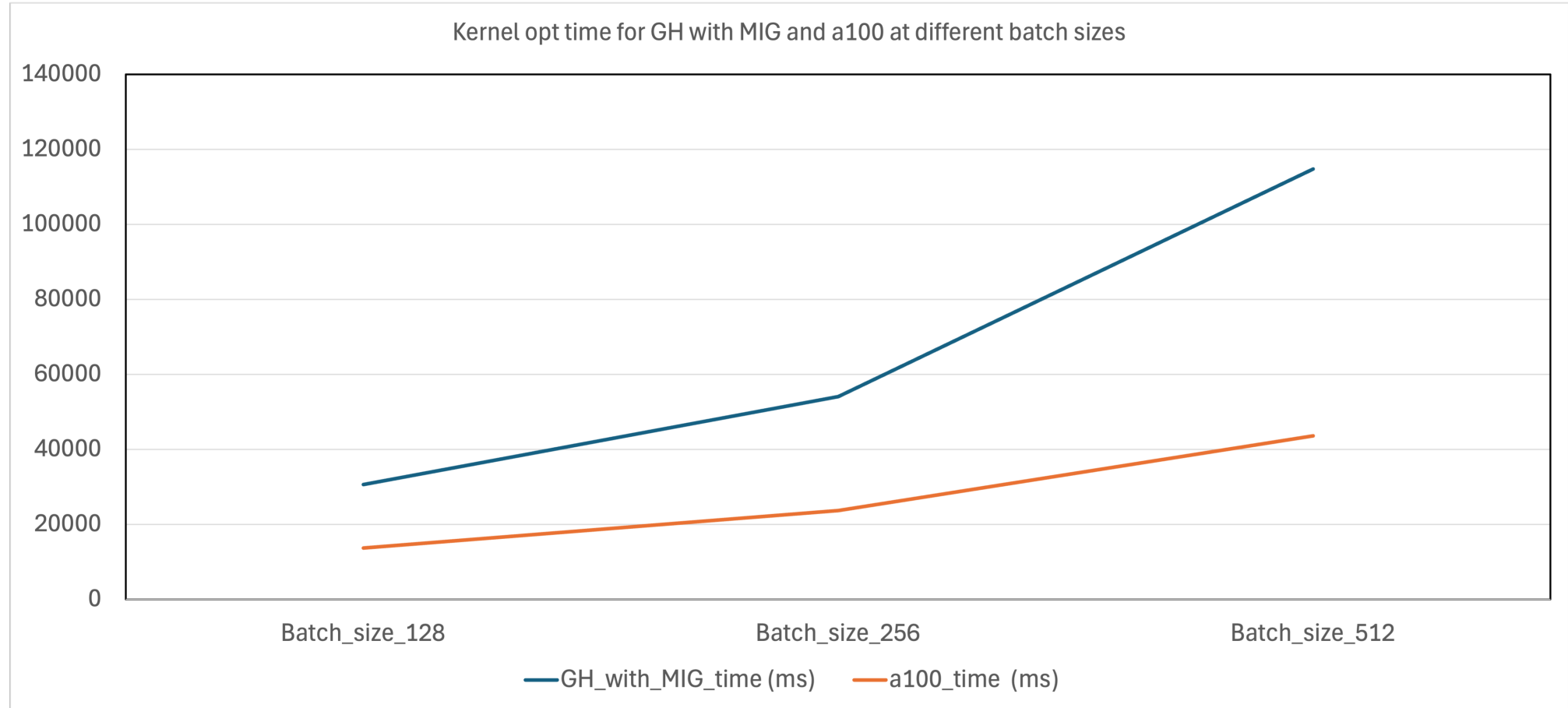
Without MIG, it is a completely different story

Avg bandwidth for GH without MIG and a100 at different batch size

— GH_no_MIG_bandwidth (GB/s) — a100_bandwidth (GB/s)
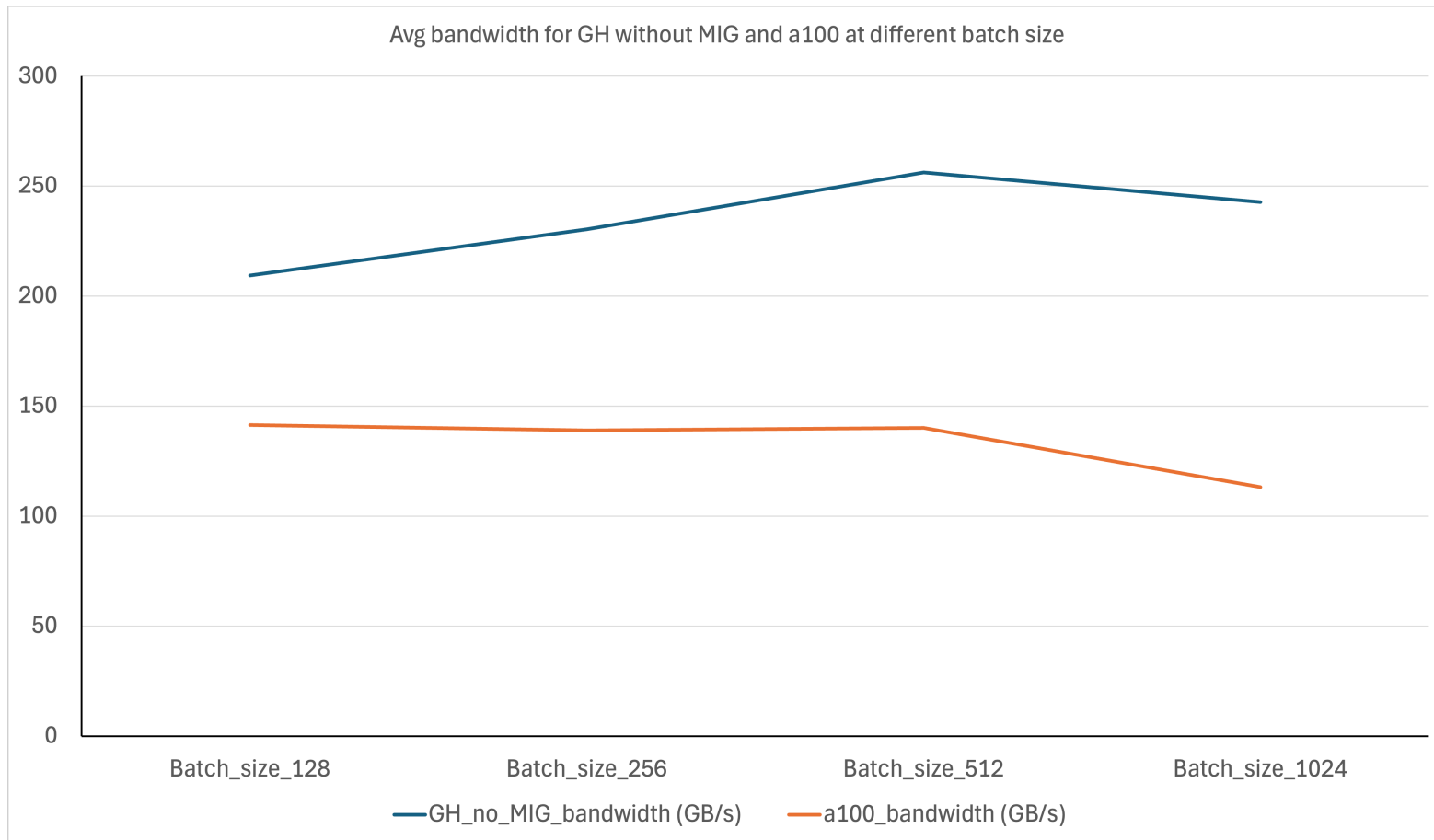
Same for the kernels optimization time

Average kernel optimization times between GH without MIG and A100 at different batch size

GH_no_MIG_time (ms) ― a100_time (ms)