16bit×4bit (group=128) mul with 72k×18k matrix on NVIDIA A10 - LOCKED BASE CLOCK