

Live Lecture Chat Window

May 17, 2023

15:08:44 How does Warp relate to Block?

A Warp is a group of 32 threads that are in a block. Really you can have any number of threads per block, but logically it makes sense to have a multiple of 32 threads per block.

16:19:51 Work group size doesn't appear to have a massive impact when 32 thread per block to greater.

That would indicate that the GPU code doesn't need to block very often. I.e., there is no big advantage in having another warp ready to go when one warp blocks.

16:23:26 What is the purpose of the "dimindx" on slide 47 again?

It is to indicate if you want the information in the X direction (dimindx=0), Y (1), or Z (2). In CUDA, they did this same thing by taking the .x, .y, and .z components of the built-in variables. Either way, the purpose is to let you compute the **gid**.