

Define “warp” in the context of CUDA programming.

Consider the following two kernels. Determine whether you think one will run faster than the other, or if both will take approximately the same time. Explain your answer.

```
__global__ void branching1(int n)
{

    int threadMod = threadIdx.x%threadsPerBlock;

    if ( threadMod%32 < 16) {
        //do some stuff
    } else {
        // do some other stuff
    }
}
```

```
__global__ void branching2(int n)
{

    int threadMod = threadIdx.x%threadsPerBlock;

    if ( threadMod%64 < 32) {
        //do some stuff
    } else {
        // do some other stuff
    }
}
```