Will Hall
GPU Computing
Homework 4

Change the main() function to invoke the kernel reduce3() you implemented, and measure the kernel execution time. Then you have to compare the kernel reduce3() with the reduce2() by filling out the following table.

Input size	1048576	16777216	67108864	134217728
Block Dimensions	1024	1024	1024	1024
T1:time cost for reduce2 (ms)	0.377	2.881	9.275	17.635
T2:time cost for reduce3 (ms)	0.384	0.796	1.901	3.034
Speedup = T1 / T2	0.982	3.619	4.879	5.812

Do you identify any advantage of reduce3() over reduce2() kernel? Or vice versus? And why ? (hint: in terms of performance, such as bank conflicts, condition divergence or idle threads.)

Yes, the larger the input size becomes, the larger the improvement. Presumably this is because it uses less threads and overall less resources on the GPU.