To calculate the time of each program in this lab use the Linux command `time`. So, if your program is called `myprog` then

```
linux> time myprog
```

And report all three numbers (real, usr, and sys)

1. Assume a reduction algorithm that finds the maximum of an array of 1024 integers.
   A. Write the sequential version of the program in C.
   B. Write a CUDA version of the program that does not take thread divergence into account.
   C. Update the version in B to take thread divergence into account.
   D. Update the program in C to make use of shared memory to reduce global memory bandwidth.

Draw a bar graph that compares the execution time of each of the above 4 versions. That is, x-axis contains the 4 versions (for each one report the real, user, and sys) and the y-axis contains the time. So, we expect to see 12 bars (4 versions and 3 timing each).

2. Repeat problem 1 with an array of 8192 elements.

3. What can we conclude from the results of problems 1 and 2 regarding the optimizations and the problem size?