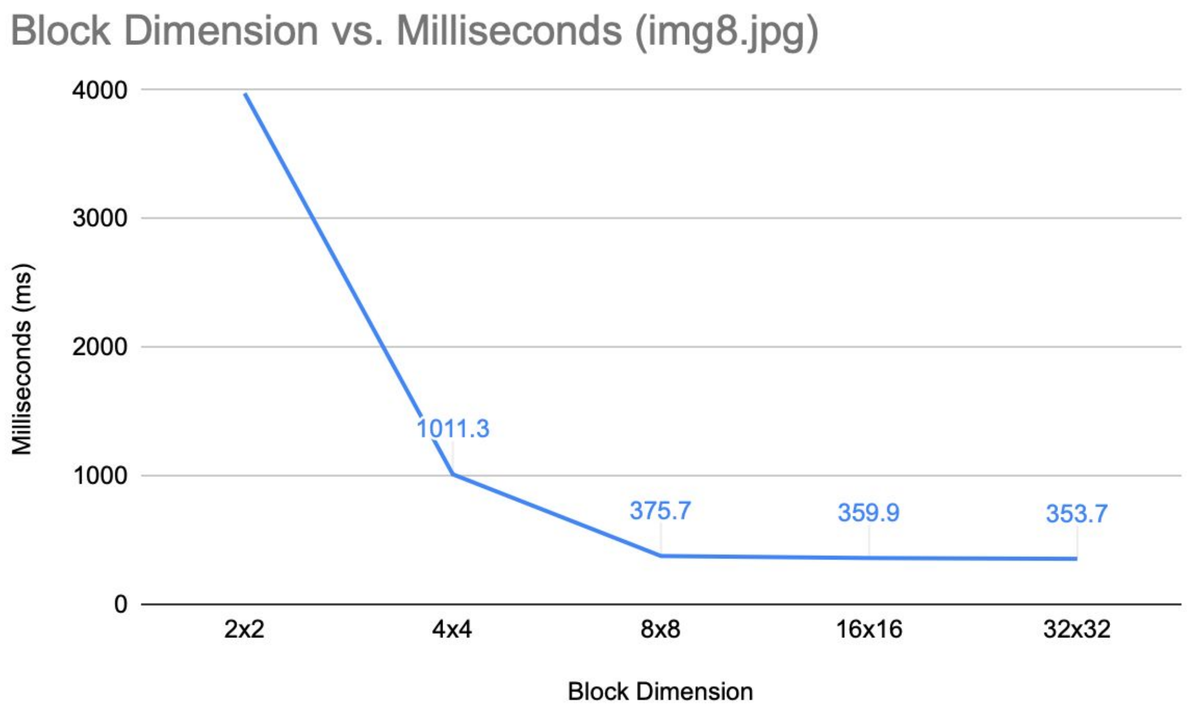
20je8
Programming Assignment 3

Part 1



There are two critical observations here:

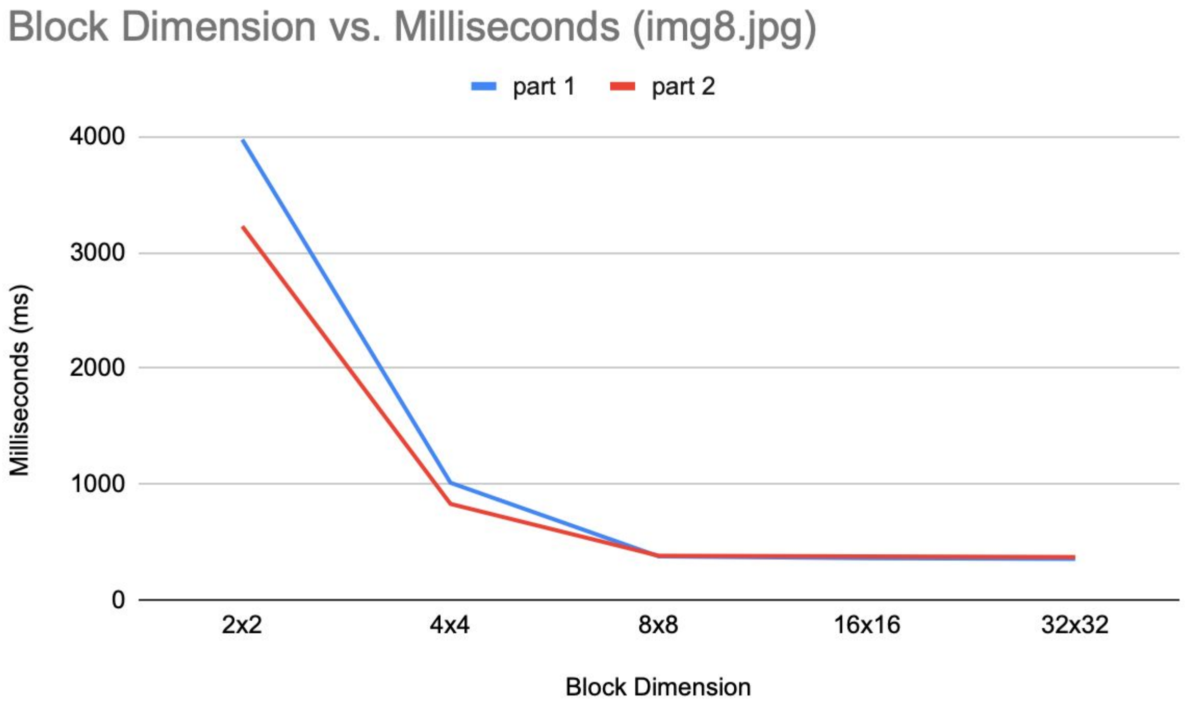
1) As block dimensions get bigger, the time to blur the image decreases.

2) Performance reaches an asymptote.

The first observation could be due to the number of threads an SM executes at once. With bigger block sizes, warps represent more threads and SMs can execute more threads at once.

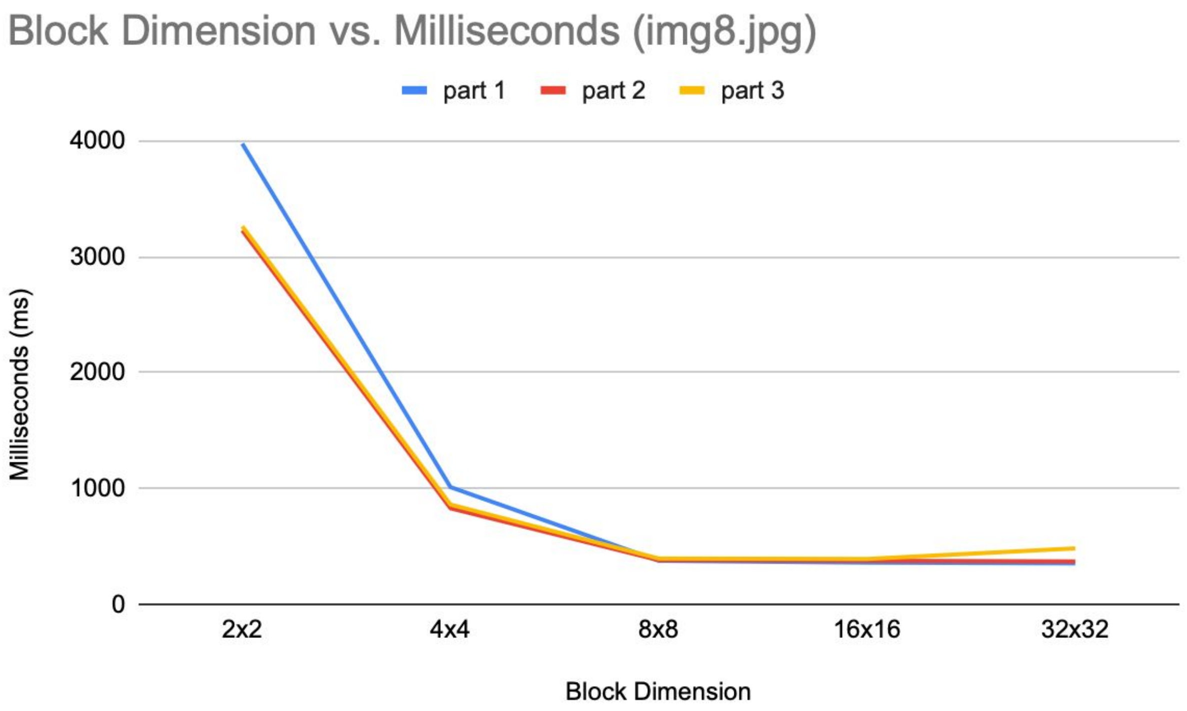
The second observation could be explained by the fact that the number of registers used by a thread and the amount of shared memory used by threads in a block can impact the number of blocks that can run in parallel on a GPU device. Generally, a 32x32 block will execute more code in parallel than a 16x16. However, if there are not enough resources then not all of these extra threads can run in parallel.

Part 2



Branching reduces the time it takes an sm to execute a piece of code. This is because threads within a single warp take different paths. Different execution paths are serialized, so a lot of parallelism is lost. We can observe a slight decrease in time needed to blur the image but, again, there are limiting factors in the number of blocks that can execute in parallel.

Part 3



There is a slight increase in the amount of time it took to blur images with the addition of the precomputed weights and division factor. It is worth discussing how these two parameters may affect the performance of the code. The array of precomputed weights requires the host to copy the array over to the GPU memory. It is stored under global memory which means that whenever a thread needs to access some value in it it has to do a load request to shared memory, which is slow! Hence, having this array as a parameter increases the time needed to blur.

GPUs are really good at doing computation, so precomputing the division factor does not really add any performance benefits. The time added by accessing the weight array in global memory undermines the small -- maybe even insignificant -- benefits that come from parametrizing the division factor