Analyze your results. When does it make sense to use the various approaches?

From the time analysis, it's clear that for smaller matrix sizes, running on the CPU takes much less time. One interesting observation was that the first time I ran each program on the GPU, it took significantly longer due to the need to initialize the CUDA context and set up the GPU environment. This happened for all three GPU algorithms. When I ran them again, the time decreased significantly. As the matrix size increases, it makes more sense to use the GPU since it performs faster with larger matrices.

How did your speed compare with cuBLAS?

It’s clear that as the matrix size increases, cuBLAS performs better and becomes the preferred choice for larger arrays.

What went well with this assignment?

What went well is that I was finally able to understand how each algorithm works.

What was difficult?

The difficult part was that it took a lot of time to understand how tiling works and what cuBLAS is.

How would you approach it differently?

Next time, I would start a little earlier.

Anything else you want me to know?

No, nothing else.