The original (Non-tiled) version runs on the GPU in 77 milliseconds.

1. My tiled implementation runs in 22 milliseconds, a speedup of about 3x.
2. Using CUBLAS, the calculation is finished after only 0.14milliseconds, a 550x improvement over the original algorithm and a 157x improvement over the tiled implementation.