CSE 491 HW 4

Block size and the way the version was implemented had a big impact on how the code performed. First of all v1 was the worst performing at any block size. This was due to having to launch kernels for every loop that was written. This has a lot of overhead and brings down performance. Any kernel using an 8x8 thread block did terrible performing as well. Similar to the reason v1 did poorly, creating that many thread blocks has too high of an overhead to be efficient. The best running code turned out to be v4 with a 16x16 block size. Version 4 was the most optimized with all kernels fused, and utilization of shared memory. Cutting down on the number of global accesses of memory was enough to make it the most efficient implementation. Block size did not make too much of a difference between 16x16 and 32x32.

| Version number | Block size 8 | Block size 16 | Block size 32 |
| --- | --- | --- | --- |
| **1** | 49.7832 | 65.7629 | 68.12 |
| **2** | 78.7722 | 104.276 | 105.864 |
| **3** | 84.8302 | 104.416 | 107.916 |
| **4** | 82.5713 | 108.154 | 107.386 |