CUDA Polynomial Evalutation

Taylor Sasser

November 1, 2023

1 Modeling

1.1 Question: Measure PCI-express latency. (That is the time for an array of size 1.)

Answer: The latency is approximately 0.1 seconds, or 100 milliseconds

1.2 Measure PCI-express Bandwidth. (The initial memory copy for different size of the array.)

Answer: The code is able to move roughly 32 gigabytes per second, using 8 streams. Using more then 8 streams had no effect on my performance.

1.3 Measure GPU memory bandwidth. (Exclude memory copies and use a low degree polynomial.)?

Answer: By using a 4 Gigabyte input array, I am able to evaluate the polynomial in 0.0135 seconds for an approximate bandwidth of: 294.18 GB/s

1.4 Measure GPU flops rating. (Exclude memory copies and use a high degree polynomial.)

Answer: By using an input array of 14 GB, and a polynomial of size 30,000, I am able to achieve approximately 6,500 GFlops