CSCI-GA.3033-022 High Performance Computing Assignment 1
Name : Yuqiong Li
NetID : yl5090
Date : 20181008

Part I: Coding questions

C1 + C2.

C1
Time: 0.00 secs
BW: 73242.19 GB/s
FLOPS: 39321.600000 GFLOP/s
C2
Time: 0.00 secs
BW: 73242.19 GB/s
FLOPS: 39321.600000 GFLOP/s

C3 + C4.

Output is 15235693999.999998
Python time : 2.04e-08 secs

Output is 15235694000.0
Numpy time : 2e-10 secs

C5 + C6.

Theory questions

Q1.
Loop unrolling is faster because there are only 1/8 loops, reducing the looping overhead.

Q2.

1. Run the program on multiple cores
2. Use multiple memory banks to move data simultaneously.

Q3.

Q4.
The first layer:

- W1 : 256 * 256 * 4000 = 262144000
- X1 : 256 * 256 = 65536
- Z1 : 4000

The second layer:

- W1 : 4000 * 1000 = 400,000
- X1 : 4000
- Z1 : 1000

Q5.
Double precision : a float is represented by 8 bytes or 64 bits.

Sizes of parameters:
The first layer:

- W1 : 256 * 256 * 4000 * 8 = 2097152000 B = 2048000 KB = 2000 MB ~= 1.95 GB
- X1 : 256 * 256 * 8 = 524288 B = 512 KB
- Z1 : 4000 * 8 = 32000 B = 31.25 KB

The second layer:

- W1 : 4000 * 1000 * 8 = 32,000,000 B = 31250 KB ~= 30.518 MB
- X1 : 4000 * 8 = 32,000 B = 31.25 KB
- Z1 : 1000 * 8 = 8000 B = 7.813 KB