Julia Nelson
CS383 - Homework 5
12/13/2020

Problem 1

a)

2 loads of 8 bytes (a[j] and b[j])
1 operation (multiplication)
1 store of the bytes (c[j])
1 / ((2+1)*8) = 1/24 = 0.04166 = **4.17%**

b)

2 loads of 8 bytes (a[j] and b[j])
1 operation (multiplication)
1 store of the bytes (**a[j],** although already loaded still needs to stores)
1 / ((2+1)*8) = 1/24 = 0.04166 = **4.17%**

c)

1 load of 8 bytes (**b[j]** - only one value loaded from main memory)
1 operation (multiplication)
1 store of the bytes (c[j])
1 / ((1+1)*8) = 1/16 = 0.0625 = **6.25%**

d)

3 loads of 8 bytes (**a[j] and b[j] and c[j]** - loads another value from memory)
2 operation (multiplication **and addition** - performs second operation)
1 store of the bytes (d[j])
2 / ((3+1)*8) = 2/32 = 0.0625 = **6.25%**

e)

2 loads of 8 bytes (**a[j] and b[j],** only needs to load b[j] once)
2 operation (multiplication and addition)
1 store of the bytes (b[j])
2 / ((2+1)*8) = 2/24 = 0.08333 = **8.33%**

Problem 2

Assume a GPU
- 2.5 GHz
- 8 SIMD processors
- 32 single precision FP units.
- supported by a 112 GB/s off-chip memory
- Assume all memory latencies can be hidden

(a) Ignoring memory bandwidth, what is the peak SP FP operation in GFLOPs?

2.5 GHz * 8 SIMD * 32 spfp = **640** GFLOPs

(b) Is this throughput sustainable given the bandwidth for performing SAXPY on large amounts of data? Justify your answer.

SAXPY (from slides) requires 3 4 byte operand input and 1 4byte result for a total of 16 byte access per Flop….

(16 bytes * 640 GFLOP)/sec= 10240 = 10.24 TB/sec

which is way more than 112 GB/s so not sustainable

Problem 3

parallelizable programs are typically accelerated by a factor of 100 on a GPU with 2500 cores
parallel efficiency = 1/# processors *speedup
(1/2500) * 100 = 0.04 = 4%
about 4 % is allocated

Problem 4

consider if we can Partition data into subsets that fit into shared memory.
is data forwarding possible for control