Assignment-4

HPC (CO332)

Arvind Ramachandran – 15CO111
Aswanth P P – 15CO112

Question-2

1. Name 3 applications of reduction.

A. A reduction operation extracts a value from a list of values. This can be sum, average,min,max etc.

2. Are there places in your solution where there is an implicit memory copy between the host and device (a copy that is not from thrust::copy)?

A. Yes, in thrust:: reduce function we are calculating value in gpu memory but it returns the value to a host variable hence there is implicit copy from device to host is happening.

3. If the Thrust version of reduce were not performing as well as you expected, how might you go about investigating and solving the problem?

A. If we knew the number of iterations in a block while compile time we can completelly unroll the reduction.
By using templates we can improve the reduction.
template < unsigned int> blocksize

Question-3

1. Name 3 applications of parallel scan.

A.

- Implement Radix Sort
- Solve Recurrences
- Solve Polynomials of higher order

2. How many floating operations are being performed in your reduction kernel?

A. 2*(log BLOCK_SIZE) Floating operations are being performed.

3. How many global memory reads are being performed by your kernel?

A. 2 Global Reads

4. How many global memory writes are being performed by your kernel?

A. 2 Global Writes

5. What is the minimum, maximum, and average number of real operations that a thread will perform? Real operations are those that directly contribute to the final reduction value.

A. Avg = Min= Max. i.e. Each thread has to perform 2*log(BLOCK_SIZE) operations.

6. How many times does a single thread block synchronize to reduce its portion of the array to a single value?

A. Each thread has to synchronize 2*log(BLOCK_SIZE) + 1.

7. Describe what optimizations were performed to your kernel to achieve a performance speedup.

A. Use of shared memory to optimize the performance.

8. Describe what further optimizations can be implemented to your kernel and what would be the expected performance behaviour?

A. Use of Shared Memory.

9. Suppose the input is greater than 2048*6535, what modifications are needed to your kernel?

A. If input is very large then break it into several parts and the compute an intermediate scan for each of the part then perform scan on intermediate scan generated.

10. Suppose a you want to scan using a a binary operator that's not commutative, can you use a parallel scan for that?

A. Commutative is not necessary.

11. Is it possible to get different results from running the serial version and parallel version of scan?

A. No.

Question-4

1. Name 3 applications of scan.
    - Implement Radix Sort
    - Solve Recurrences
    - Solve Polynomials of higher order

2. Suppose a you want to perform the algorithm using a binary operator that's not commutative, can you use still use parallel scan?

A. Yes, We can use scan operation to perform this task .It is advisable to not use scan if the operator is not associative in nature

3. Is it possible to get different results from running the serial version and parallel version of reduction?

A. Since reduction operation is associative in nature reduction in serial as well as parallel results same answer.