**Assignment 7**

**Performance Modeling for Convolutions on GPU**

**All were run on a single GPU**

Q1) How many Flops needs to be done to compute a convolution of dimension k on a image of size

m\*n?

1. Our Image size is m\*n and our kernel size which would be used for convolution is k\*k
2. In convolution, for a single pixel, we will do 2 times k\*k operations ie with our kernel we will do multiplication first and then addition.
3. So for an image with a size of m\*n, we will do a total of (m\*n\*k\*k\*2) flops.

Q2) ) How much memory needs to be moved to compute a convolution of dimension k on an image of

size n\*m?

* We will bring the image of size m\*n from memory also the kernel of size k\*k.
* So we will need to fetch ((m\*n)+ (k\*k))\*4 bytes from Global memory.
* Also, we will write back (m\*n)\*4 bytes to memory.
* As I am using **const float\* \_\_restrict\_\_ kernel,** the kernel would be cached and we need to bring it only once.

Q3) Assuming the performance numbers you measured in assignment 5 , how long should

computing a convolution of dimension 3 on an image of 1024x768 take?

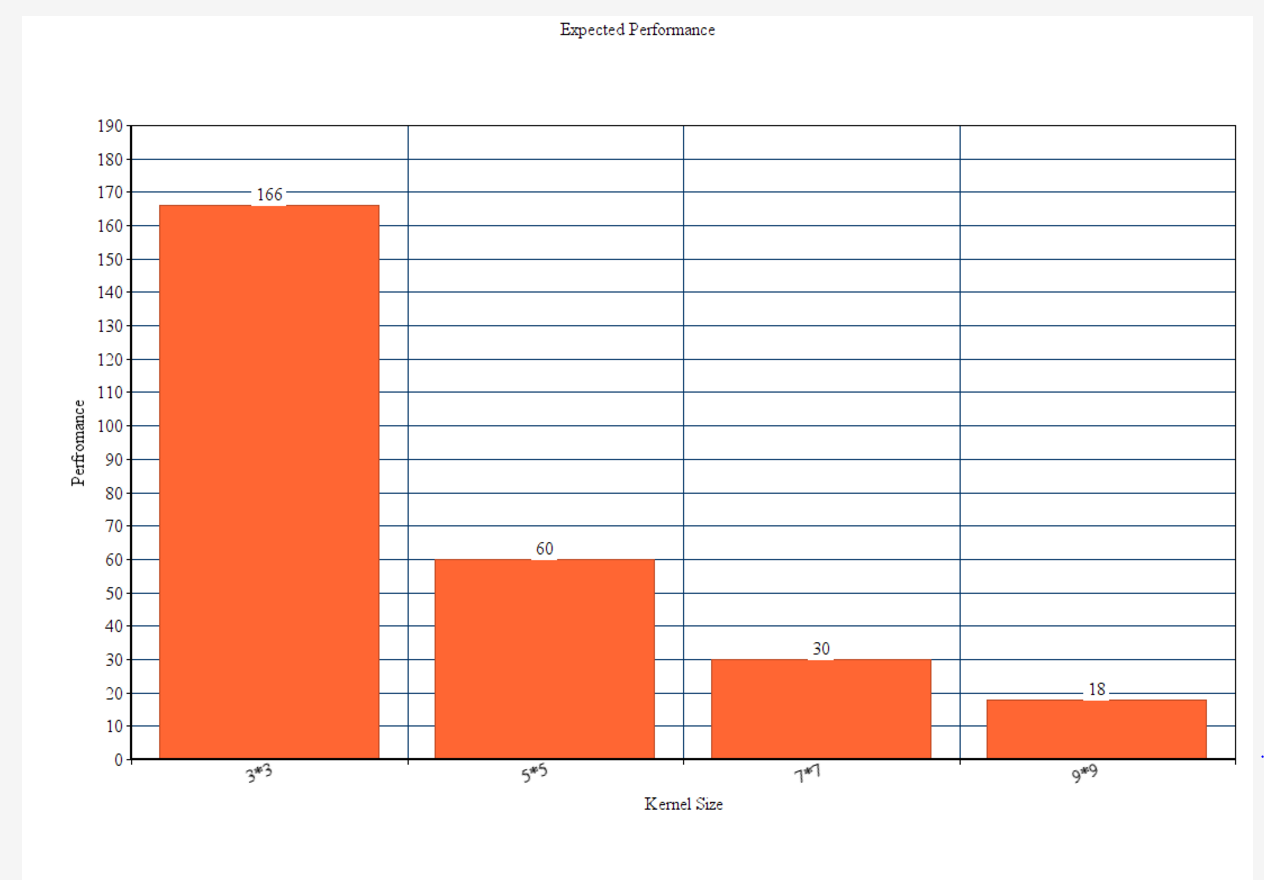
* Considering m\*n= 1024 \* 768 and k\*k=3\*3.
* From Assignment 5, we got our Flop rating as 3TB and Memory Bandwidth as 155GB/sec.
* Flop time for calculation= ((m\*n\*k\*k\*2)/ 3\*10^12))= 4.71 usec.
* Memory time calculation= (2(m\*n)+ (k\*k))\*4)/155\*10^9 = 40.5 usec.

**Performance Model**

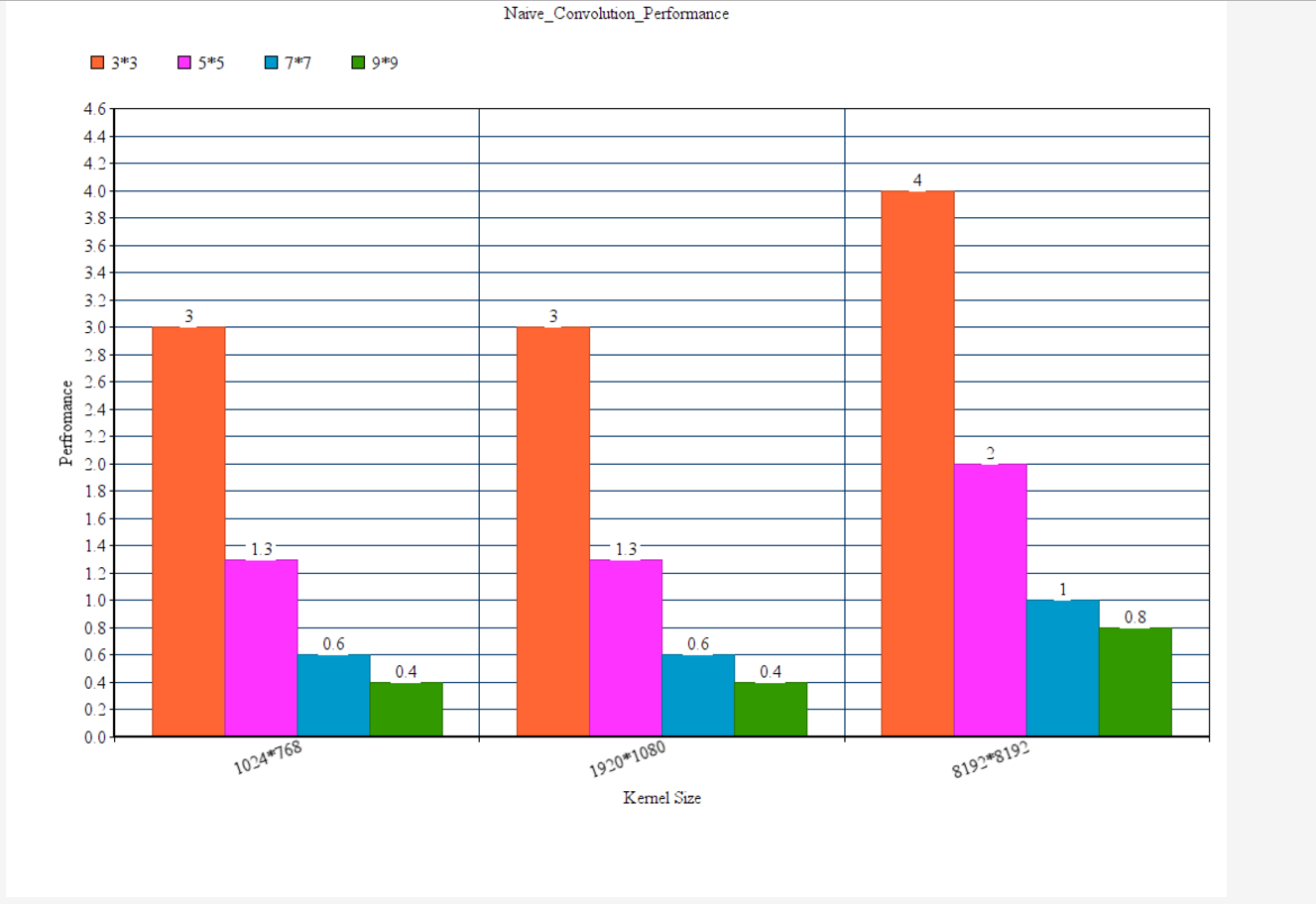
Peak flop= ((1024 \*768)/ 4.71 \* 10^-6) = 166 Gp/sec.

Memory flops= ((1024 \*768)/ 40.5 \* 10^-6) = 19Gp/sec.

**Expected Performance Graph**



Naïve \_Convolution Perfromance



Performance Reported for various Image Sizes and Kernels

1. M=Rows and N=Cols

M=1024, N-768

Kernel- 3\*3

Flop Bound- 166Gp/sec.

Mem Bound- 19Gp/sec.

Naïve Code Performance- 3Gp/sec

Shared Mem Code Performance-

K=5\*5

The flop Bound Perfromance is60

The Mem Bound Performance is 18

The actual Performance is 1105094977

1.1

K=7\*7

The flop Bound Perfromance is30

The Mem Bound Performance is 18

The actual Performance is 573635155

K=9\*9

The flop Bound Perfromance is18

The Mem Bound Performance is 18

The actual Performance is 348342825

M=1920 N=1080

K=3

The flop Bound Perfromance is166

The Mem Bound Performance is 18

The actual Performance is 2791719621

K=5

The flop Bound Perfromance is 60

The Mem Bound Performance is 18

The actual Performance is 1099398077

K=7

The flop Bound Perfromance is30

The Mem Bound Performance is 18

The actual Performance is 614308991

K=9

For array size 1080 The time required for convolution is 0.005546

The flop Bound Perfromance is18

The Mem Bound Performance is 18

The actual Performance is 373920151

~

M=8192 N=8192

K=3

The flop Bound Perfromance is166

The Mem Bound Performance is -12

The actual Performance is 4041710240

K=5

The flop Bound Perfromance is60

The Mem Bound Performance is -12

The actual Performance is 1651914849

K=7

The flop Bound Perfromance is30

The Mem Bound Performance is -12

The actual Performance is 866366484\

K=9

The flop Bound Perfromance is18

The Mem Bound Performance is -12

The actual Performance is 528856497