SAURAV RAI
17558

AUM SRI SAI RAM

MATRIX - MATRIX MULTIPLICATION (GLOBAL AND SHARED MEMORY:

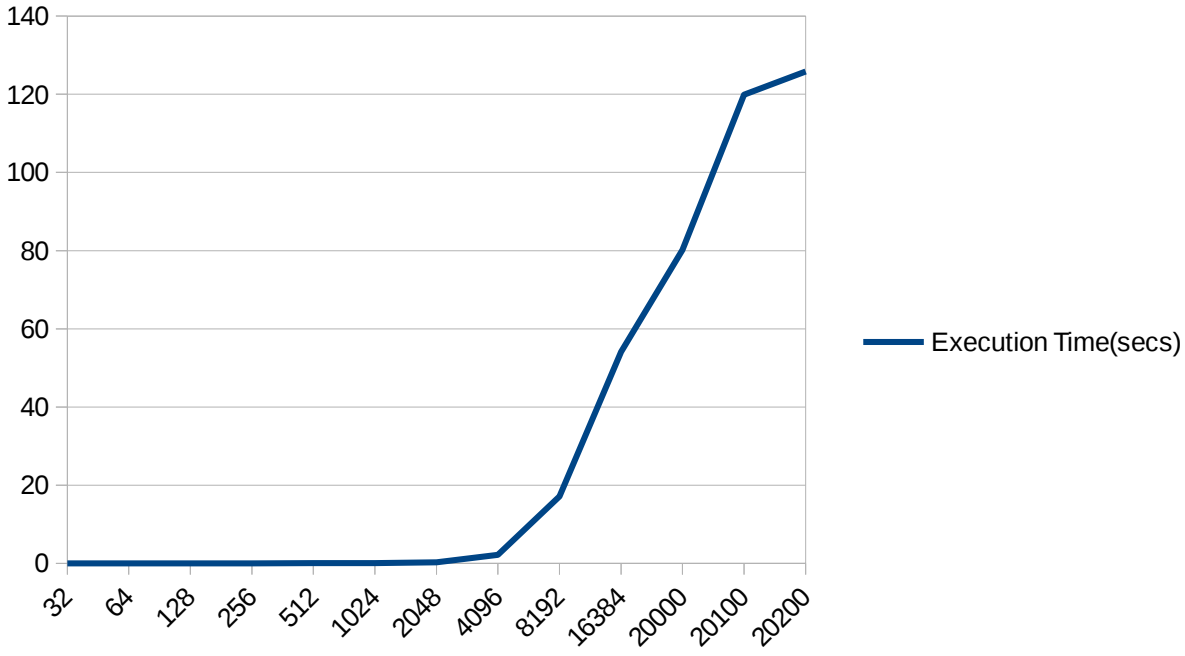
• I have wrote two programs for matrix multiplication. One program which uses the global memory and other one uses the share memory.

• **NAME :** Tesla K20c
Memory:
Size: 4742Mb , **Type:** GDDR5 SDRAM
Bus: 320 bit , **Bandwidth:** 208.0 GBps

1. Global memory matrix multiplication:

• In this implementation all the elements of the matrix are stored in the global memory.
• Whenever threads in a block requires an element for the multiplication it gets from the global memory. The following table shows the performance of the matrix multiplication.

Sl.no	**Size**	Execution Time(secs)
1.	32	0.0000
2	64	0.0001
3.	128	0.0002
4.	256	0.0008
5.	512	0.0200
6	1024	0.0300
7	2048	0.2800
8	4096	2.1500
9	8192	17.0800
10	16384	54.0888
11	20000	80.1945
12	20100	119.9010
13	20200	125.8012
14	20300	Out of memory



REPORT :

From the graph we can observe that as the size of the matrix for multiplication increases , the execution time also increases exponentially.

2. Shared (Tiled) matrix multiplication:

• In case of shared matix multiplication , it uses the shared matrix efficiently.

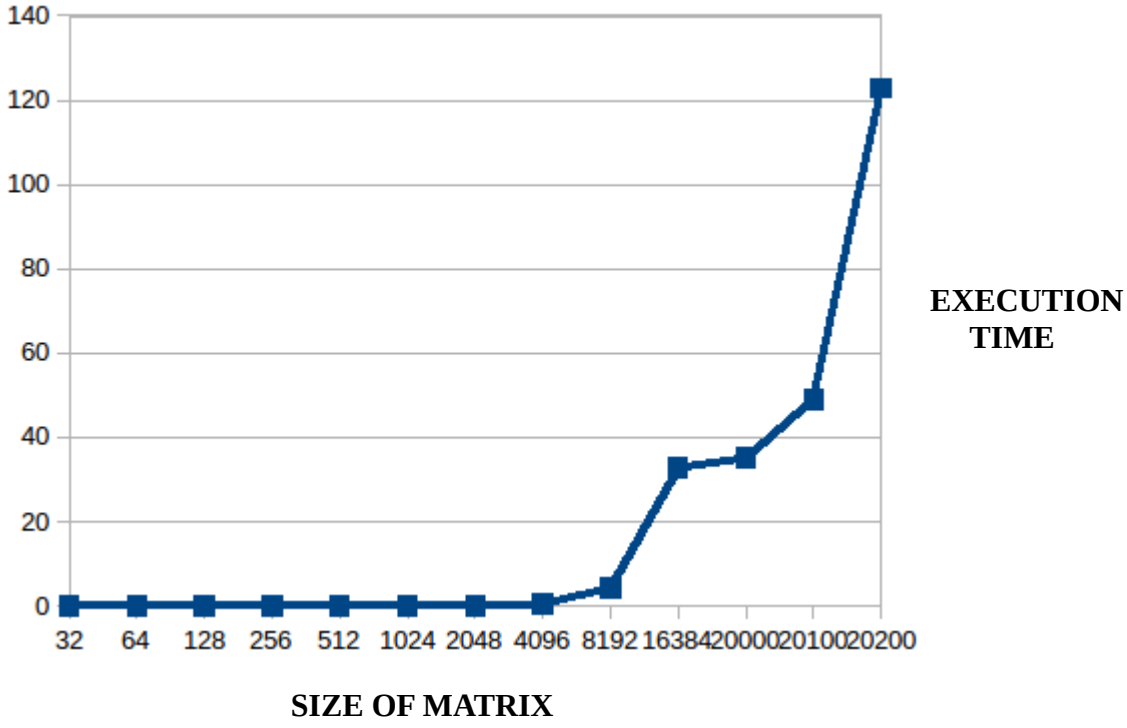
• Whenever threads accesses the elements from global memory, it copies a tile of elements in shared memory.
• As a result, the thread first sees elements in the shared memory. It does not need to access from the global memory , which then results in the performance gain in the matrix multiplication.

• Gain in performance is observed below:

Sl.no	**Size of**	**Sequential GFLOPS**	**SEQUENTIAL EXECUTION TIMES (secs)**	**Parllel GFLOPS**	**PARALLEL EXECUTION TIMES (secs)**
1	32	0.1	0.00056	0.6	0.00010
2	64	0.1	0.00596	2.8	0.00019
3	128	0.1	0.03318	17.6	0.00024
4	256	0.1	0.26974	47.5	0.00071
5	512	0.1	2.44514	91.9	0.00292

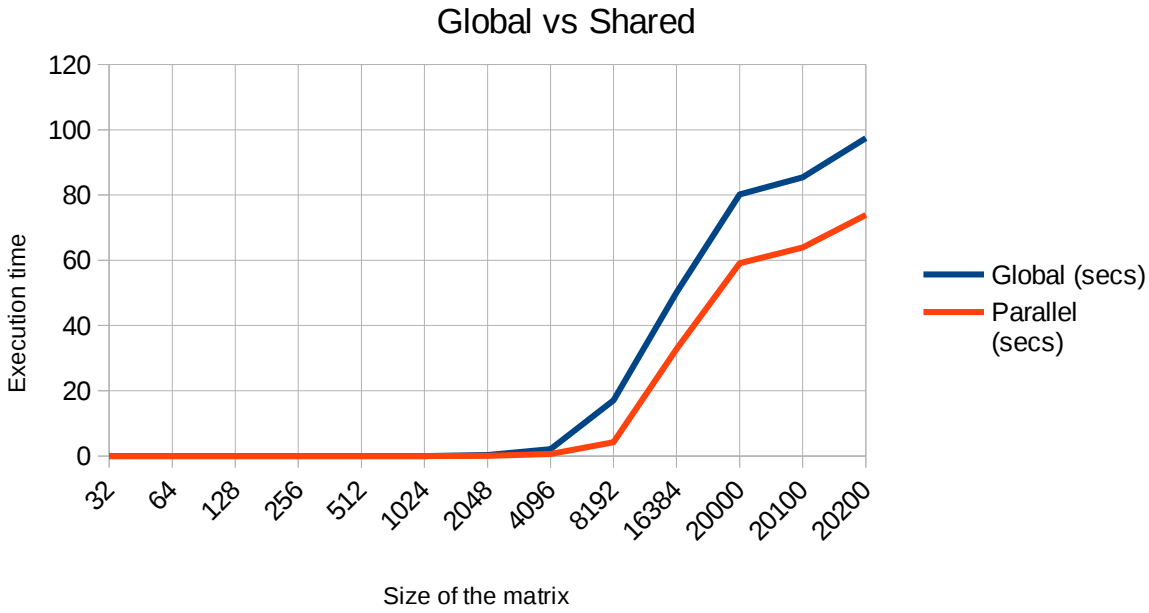
6	1024	0.1	29.53870	92.1	0.02332
7	2048	0.1	473.58276	223.9	0.07674
8	4096	0.16	-	226.2	0.60722
9	8192	0.07	-	260.5	4.22044
10	16384	0.03	-	268.6	32.75022
11	20000	0.02	-	270.5	35.14555
12	20100	0.01	-	254.2	48.90349
13	20200	0.01	-	223.2	59.85013
14	20300	0.001	-	Out of Mem	Out of Mem

EXECUTION TIME ON SHARED _ MATRIX_MULTIPLICATION



COMPARISION BETWEEN SHARED AND GLOBAL MATRIX MULTIPLICATION:

• Below is the comparision between the shared and global matrix multiplication.



RESULT :

• The above result shows as the size of matrix goes on increasing ,the execution time for both matrix multiplication increases.

• The execution time for shared matrix multiplication is less compared to global matrix multiplication . This is due to coalesced memory access and tiling which results that the access to global memory access is reduced drastically improving execution time.