# EE 451 PHW 7 – Report
# Aditya Dhulipala

Problem 1:
The execution time for the global memory based matrix multiplication took around
19 milliseconds

The associated files are p1.cu and the output file is p1.output

Problem 2:
The execution time for the shared memory based matrix multiplication took around
16 milliseconds

The associated files are p2.cu and the output file is p2.output

The shared memory based program was implemented as a block matrix multiplication program due to insufficient memory to hole both the entire 1K X 1K matrices.

Clearly, approach 2 has a better run time than approach 1. This is because we loaded a complete block of elements (of size 32 X 32) into the shared memory (L1 cache) of the cuda blocks. Therefore, each of the threads working inside a particular cudaBlock could access the required matrix elements straight from this cache. This is much faster than accessing from the global memory. This efficiency adds up for the overall computation when we consider the efficiency gain for each of the other such blocks. Hence we have a smaller runtime. In approach 2, we did have to copy the data from global memory to shared memory. However, this happened only once (for each block). Hence the overhead incurred is negligible compared to the throughput.

**Screenshots:**

Problem 1 – Approach 1

```
~/Projects/ee-451/phw7/p1
$ cat cudajob.output
----------------------------------------
Begin PBS Prologue Sun Apr 19 17:57:40 PDT 2015
Job ID:            11959624.hpc-pbs.hpcc.usc.edu
Username:          adhulipa
Group:             csci-ar
Name:              CUDAtest
Queue:             quick
Shared Access:     no
All Cores:         no
Has MIC:           no
Nodes:             hpc3026
TMPDIR:            /tmp/11959624.hpc-pbs.hpcc.usc.edu
End PBS Prologue Sun Apr 19 17:57:41 PDT 2015
----------------------------------------
C[451][451] = 2048
Time - 19.828672
----------------------------------------
Begin PBS Epilogue Sun Apr 19 17:57:41 PDT 2015
Job ID:           11959624.hpc-pbs.hpcc.usc.edu
Username:         adhulipa
Group:            csci-ar
Job Name:         CUDAtest
Session:          29475
Limits:           neednodes=1:ppn=16:gpus=2,nodes=1:ppn=16:gpus=2,walltime=00:08:00
Resources:        cput=00:00:00,mem=0kb,vmem=0kb,walltime=00:00:01
Queue:            quick
Shared Access:    no
Has MIC:          no
End PBS Epilogue Sun Apr 19 17:57:41 PDT 2015
----------------------------------------
```

Problem 2 – Approach 2

```
~/Projects/ee-451/phw7/p1
$ cat ../p2/cudajob.output
----------------------------------------
Begin PBS Prologue Sun Apr 19 18:40:40 PDT 2015
Job ID:            11959677.hpc-pbs.hpcc.usc.edu
Username:          adhulipa
Group:             csci-ar
Name:              CUDAtest
Queue:             quick
Shared Access:     no
All Cores:         no
Has MIC:           no
Nodes:             hpc3025
TMPDIR:            /tmp/11959677.hpc-pbs.hpcc.usc.edu
End PBS Prologue Sun Apr 19 18:40:41 PDT 2015
----------------------------------------
C[451][451] = 2048
Time - 16.662144
----------------------------------------
Begin PBS Epilogue Sun Apr 19 18:40:43 PDT 2015
Job ID:           11959677.hpc-pbs.hpcc.usc.edu
Username:         adhulipa
Group:            csci-ar
Job Name:         CUDAtest
Session:          21137
Limits:           neednodes=1:ppn=16:gpus=2,nodes=1:ppn=16:gpus=2,walltime=00:08:00
Resources:        cput=00:00:00,mem=1900kb,vmem=18816kb,walltime=00:00:01
Queue:            quick
Shared Access:    no
Has MIC:          no
End PBS Epilogue Sun Apr 19 18:40:43 PDT 2015
----------------------------------------
```

The times above are in millseconds.