

**Output File**  
**Rushikesh Khamkar**

# Part 1 – Getting Started – Execute CPU & GPU File

## 1.1 CPU version of the code -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ g++ matrixMul_cpu.cpp -o matrixMul_cpu.exe
gpu-20-11:/CS5375_GPU_Lecture/Project$ ./matrixMul_cpu.exe
```

```
Size of matrix (N) is 512 by 512.
Starting CPU computation
It took 0.003000 ms on avg.
RUN OK.
```

## 1.2 GPU version of the code (Unoptimized version of the code) -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvcc matrixMul_gpu.cu -o matrixMul_gpu.exe
gpu-20-11:/CS5375_GPU_Lecture/Project$ ./matrixMul_gpu.exe
```

```
Size of matrix (N) is 512 by 512.
Starting unoptimized GPU computation
It took 0.037000 ms on avg.
RUN OK.
```

## Nvprof version for the GPU Code -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvprof ./matrixMul_gpu.exe
Size of matrix (N) is 512 by 512.
==22681== NVPROF is profiling process 22681, command: ./matrixMul_gpu.exe
Starting unoptimized GPU computation
```

```
It took 0.040667 ms on avg.
RUN OK.
```

```
==22681== Profiling application: ./matrixMul_gpu.exe
```

```
==22681== Profiling result:
```

Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	100.00%	312.67us	4	78.167us	30.176us	221.76us	
GPUmatmul(int, float*, float*, float*)							
API calls:	99.40%	201.29ms	3	67.097ms	20.049us	201.23ms	
cudaMallocManaged							
	0.27%	555.86us	1	555.86us	555.86us	555.86us	
cuDeviceTotalMem							
	0.16%	322.96us	4	80.740us	32.964us	223.69us	
cudaDeviceSynchronize							
	0.06%	115.12us	4	28.779us	4.9590us	95.545us	
cudaLaunchKernel							
	0.05%	107.93us	3	35.975us	8.4700us	63.652us	
cudaFree							
	0.05%	92.333us	101	914ns	115ns	37.905us	
cuDeviceGetAttribute							
	0.01%	18.571us	1	18.571us	18.571us	18.571us	
cuDeviceGetName							
	0.00%	1.9160us	1	1.9160us	1.9160us	1.9160us	
cuDeviceGetPCIBusId							
	0.00%	1.1980us	3	399ns	181ns	820ns	
cuDeviceGetCount							
	0.00%	666ns	2	333ns	128ns	538ns	
cuDeviceGet							
	0.00%	278ns	1	278ns	278ns	278ns	
cuDeviceGetUuid							

```
==22681== Unified Memory profiling result:
Device "Tesla V100-PCIE-32GB (0)"
```

Count	Avg Size	Min Size	Max Size	Total Size	Total Time	Name
2	32.000KB	4.0000KB	60.000KB	64.00000KB	9.600000us	Host To Device
2	32.000KB	4.0000KB	60.000KB	64.00000KB	6.848000us	Device To Host
1	-	-	-	-	189.7280us	Gpu page fault

groups

Total CPU Page faults: 2

## Part 2 – Multiple Threads

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvcc matrixMul_gpu_v2.cu -o
matrixMul_gpu_v2.exe
gpu-20-11:/CS5375_GPU_Lecture/Project$ ./matrixMul_gpu_v2.exe
```

```
Size of matrix (N) is 512 by 512.
Starting Optimized GPU computation
It took 0.009333 ms on avg.
RUN OK.
```

### Nvprof version for the GPU Code -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvprof ./matrixMul_gpu_v2.exe
Size of matrix (N) is 512 by 512.
==22745== NVPROF is profiling process 22745, command: ./matrixMul_gpu_v2.exe
Starting Optimized GPU computation
```

```
It took 0.013667 ms on avg.
RUN OK.
```

```
==22745== Profiling application: ./matrixMul_gpu_v2.exe
```

```
==22745== Profiling result:
```

Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	100.00%	135.07us	4	33.767us	2.8480us	126.27us	
GPUmatmul(int, float*, float*, float*)							
API calls:	99.47%	200.72ms	3	66.908ms	17.081us	200.67ms	
cudaMallocManaged							
	0.28%	556.16us	1	556.16us	556.16us	556.16us	
cuDeviceTotalMem							
	0.08%	152.57us	4	38.141us	5.7090us	129.85us	
cudaDeviceSynchronize							
	0.07%	133.72us	4	33.430us	4.8140us	114.34us	
cudaLaunchKernel							
	0.05%	107.47us	3	35.823us	8.6080us	66.796us	
cudaFree							
	0.05%	93.817us	101	928ns	118ns	38.435us	
cuDeviceGetAttribute							
	0.01%	18.258us	1	18.258us	18.258us	18.258us	
cuDeviceGetName							
	0.00%	1.8560us	1	1.8560us	1.8560us	1.8560us	
cuDeviceGetPCIBusId							
	0.00%	1.1210us	3	373ns	162ns	776ns	
cuDeviceGetCount							
	0.00%	732ns	2	366ns	167ns	565ns	
cuDeviceGet							
	0.00%	233ns	1	233ns	233ns	233ns	
cuDeviceGetUuid							

```
==22745== Unified Memory profiling result:
```

```
Device "Tesla V100-PCIE-32GB (0)"
```

Count	Avg Size	Min Size	Max Size	Total Size	Total Time	Name
2	32.000KB	4.0000KB	60.000KB	64.00000KB	9.664000us	Host To Device
2	32.000KB	4.0000KB	60.000KB	64.00000KB	6.848000us	Device To Host
1	-	-	-	-	122.2400us	Gpu page fault

```
groups
```

```
Total CPU Page faults: 2
```

## Part 3 – Multiple Blocks

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvcc matrixMul_gpu_v3.cu -o
matrixMul_gpu_v3.exe
gpu-20-11:/CS5375_GPU_Lecture/Project$ ./matrixMul_gpu_v3.exe
```

```
Size of matrix (N) is 512 by 512.
Starting Optimized GPU computation
It took 0.019000 ms on avg.
RUN OK.
```

### Nvprof version for the GPU Code -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvprof ./matrixMul_gpu_v3.exe
Size of matrix (N) is 512 by 512.
==22764== NVPROF is profiling process 22764, command: ./matrixMul_gpu_v3.exe
Starting Optimized GPU computation
```

```
It took 0.028333 ms on avg.
RUN OK.
```

```
==22764== Profiling application: ./matrixMul_gpu_v3.exe
```

```
==22764== Profiling result:
```

Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	100.00%	181.15us	4	45.287us	18.272us	126.08us	
GPUmatmul(int, float*, float*, float*)							
API calls:	99.40%	200.12ms	3	66.707ms	5.7170us	200.10ms	
cudaMallocManaged							
	0.32%	649.57us	1	649.57us	649.57us	649.57us	
cuDeviceTotalMem							
	0.10%	192.55us	4	48.138us	21.114us	128.83us	
cudaDeviceSynchronize							
	0.06%	130.06us	4	32.516us	4.8170us	111.42us	
cudaLaunchKernel							
	0.05%	108.07us	3	36.024us	8.4180us	67.499us	
cudaFree							
	0.05%	97.652us	101	966ns	123ns	39.093us	
cuDeviceGetAttribute							
	0.01%	18.088us	1	18.088us	18.088us	18.088us	
cuDeviceGetName							
	0.00%	2.4420us	1	2.4420us	2.4420us	2.4420us	
cuDeviceGetPCIBusId							
	0.00%	1.2420us	3	414ns	147ns	895ns	
cuDeviceGetCount							
	0.00%	621ns	2	310ns	151ns	470ns	
cuDeviceGet							
	0.00%	271ns	1	271ns	271ns	271ns	
cuDeviceGetUuid							

```
==22764== Unified Memory profiling result:
```

```
Device "Tesla V100-PCI-E-32GB (0)"
```

Count	Avg Size	Min Size	Max Size	Total Size	Total Time	Name
2	32.000KB	4.0000KB	60.000KB	64.00000KB	9.664000us	Host To Device
2	32.000KB	4.0000KB	60.000KB	64.00000KB	6.848000us	Device To Host
1	-	-	-	-	121.7600us	Gpu page fault

```
groups
```

```
Total CPU Page faults: 2
```

## Part 4 – Optimize

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvcc matrixMul_gpu_v4.cu -o
matrixMul_gpu_v4.exe
gpu-20-11:/CS5375_GPU_Lecture/Project$ ./matrixMul_gpu_v4.exe
```

```
Size of matrix (N) is 512 by 512.
Starting optimized GPU computation
It took 0.308000 ms on avg.
RUN OK.
```

### Nvprof version for the GPU Code -

```
gpu-20-11:/CS5375_GPU_Lecture/Project$ nvprof ./matrixMul_gpu_v4.exe
Size of matrix (N) is 512 by 512.
==22851== NVPROF is profiling process 22851, command: ./matrixMul_gpu_v4.exe
Starting optimized GPU computation
It took 1.033000 ms on avg.
RUN OK.
```

```
==22851== Profiling application: ./matrixMul_gpu_v4.exe
```

```
==22851== Profiling result:
```

Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	100.00%	4.5103ms	4	1.1276ms	796.28us	2.0892ms	
GPUmatmul(int, double*, double*, double*)							
API calls:	96.73%	391.46ms	3	130.49ms	13.399us	391.43ms	
cudaMallocManaged							
	2.82%	11.428ms	4	2.8570ms	2.1883ms	3.0996ms	
cudaDeviceSynchronize							
	0.19%	769.40us	3	256.47us	246.77us	275.30us	
cudaMemPrefetchAsync							
	0.14%	557.18us	1	557.18us	557.18us	557.18us	
cuDeviceTotalMem							
	0.08%	316.26us	3	105.42us	33.964us	146.82us	
cudaFree							
	0.02%	95.173us	101	942ns	117ns	39.326us	
cuDeviceGetAttribute							
	0.01%	50.401us	4	12.600us	4.6700us	30.556us	
cudaLaunchKernel							
	0.00%	19.317us	1	19.317us	19.317us	19.317us	
cuDeviceGetName							
	0.00%	6.9210us	1	6.9210us	6.9210us	6.9210us	
cuDeviceGetPCIBusId							
	0.00%	1.1300us	3	376ns	160ns	775ns	
cuDeviceGetCount							
	0.00%	591ns	2	295ns	126ns	465ns	
cuDeviceGet							
	0.00%	256ns	1	256ns	256ns	256ns	
cuDeviceGetUuid							

```
==22851== Unified Memory profiling result:
```

```
Device "Tesla V100-PCIE-32GB (0)"
```

Count	Avg Size	Min Size	Max Size	Total Size	Total Time	Name
64	64.000KB	4.0000KB	980.00KB	4.000000MB	491.7760us	Host To Device
2	32.000KB	4.0000KB	60.000KB	64.00000KB	6.528000us	Device To Host
5	-	-	-	-	1.254400ms	Gpu page fault

```
groups
```

```
Total CPU Page faults: 2
```