GPU Bench

GPU Performance Details: GeForce GTX 960

Contents:

System Configuration

MATLAB Release: R2016b

Host

Name	Intel(R) Core(TM) i5-6400T CPU @ 2.20GHz
Clock	2201 MHz
Cache	1024 KB
NumProcessors	4
OSType	Windows
OSVersion	Microsoft Windows 10 Home

GPU

Name	GeForce GTX 960
Clock	1.200500e+03 MHz
NumProcessors	8
ComputeCapability	5.2
TotalMemory	4.00 GB
CUDAVersion	8
DriverVersion	6.14.13.6930 (369.30)

Results for MTimes (double)

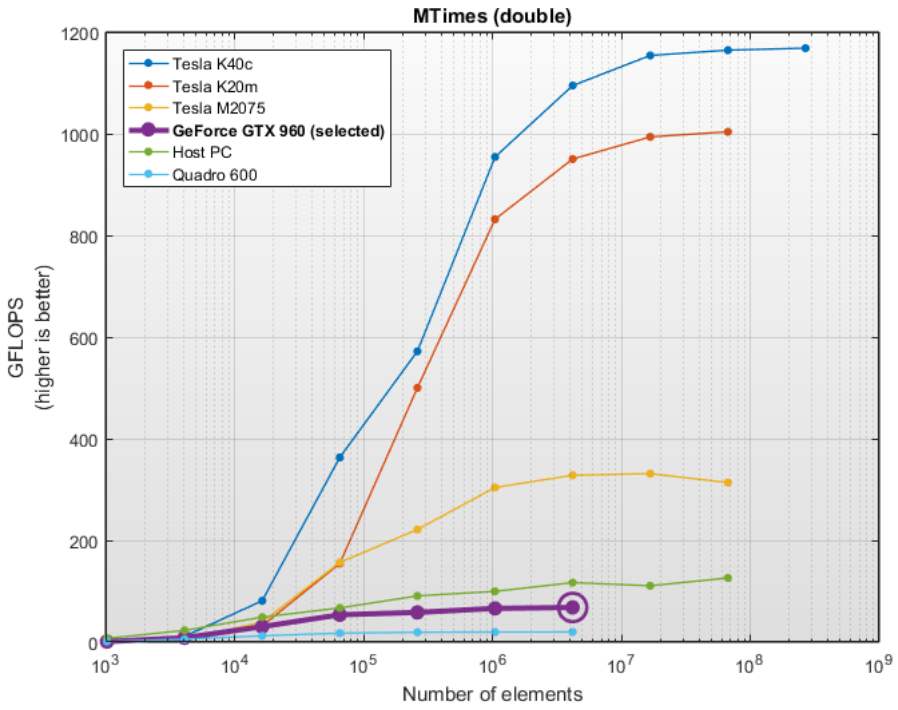
These results show the performance of the GPU or host PC when calculating a matrix multiplication of two NxN real matrices. The number of operation assumed to be `2*N^3 - N^2`.

This calculation is usually compute-bound, i.e. the performance depends mainly on how fast the GPU or host PC can perform floating-point operations.

Raw data for GeForce GTX 960 - MTimes (double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	64,512	0.07	0.91
4,096	520,192	0.06	8.38
16,384	4,177,920	0.14	30.32
65,536	33,488,896	0.62	53.68
262,144	268,173,312	4.60	58.32
1,048,576	2,146,435,072	32.42	66.20
4,194,304	17,175,674,880	252.31	68.07

(N gigaflops = $Nx10^9$ operations per second)



Results for Backslash (double)

These results show the performance of the GPU or host PC when calculating the matrix left division of an `NxN` matrix with an `Nx1` vector. The number o operations is assumed to be `2/3*N^3 + 3/2*N^2`.

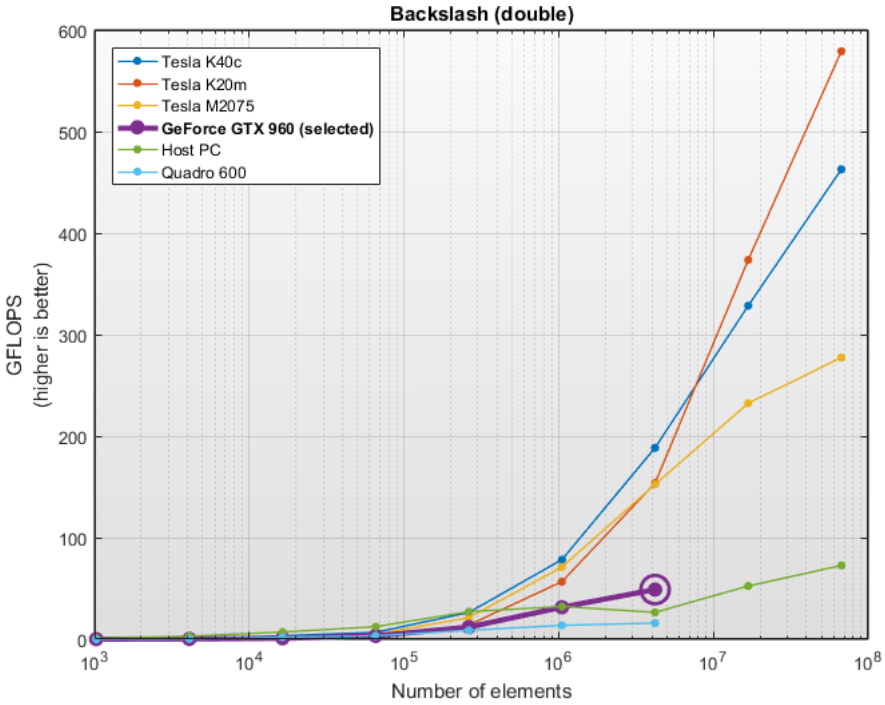
This calculation is usually compute-bound, i.e. the performance depends mainly on how fast the GPU or host PC can perform floating-point operations.

Raw data for GeForce GTX 960 - Backslash

(double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	23,381	0.94	0.02
4,096	180,907	0.96	0.19
16,384	1,422,677	1.62	0.88
65,536	11,283,115	3.51	3.21
262,144	89,871,701	7.57	11.88
1,048,576	717,400,747	22.80	31.46
4,194,304	5,732,914,517	117.32	48.87

(N gigaflops = $Nx10^9$ operations per second)



Results for FFT (double)

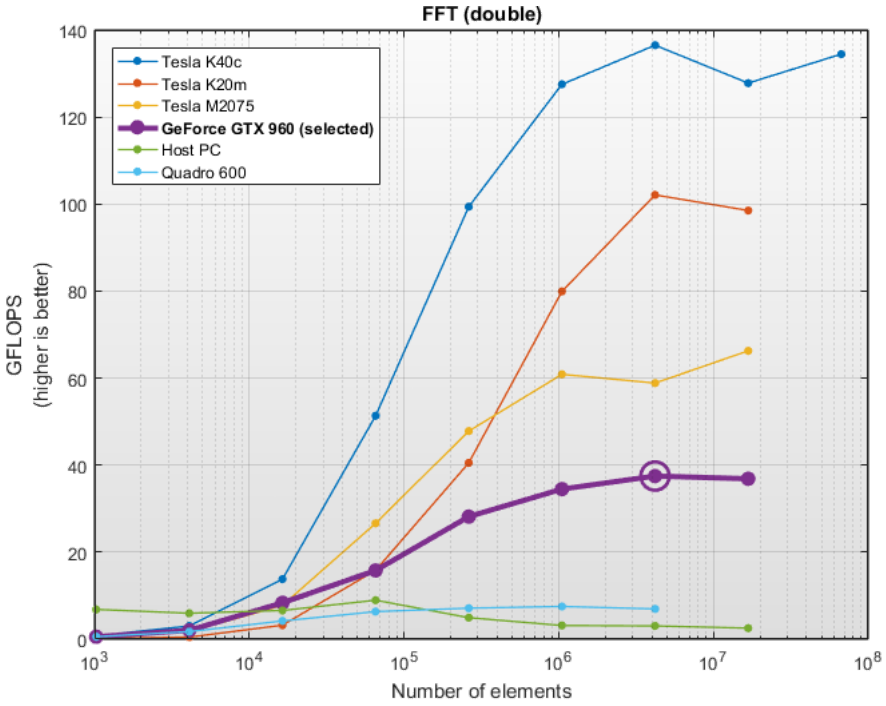
These results show the performance of the GPU or host PC when calculating the Fast-Fourier-Transform of a vector of complex numbers. The number operations for a vector of length `N` is assumed to be `5*N*log2(N)`.

This calculation is usually memory-bound, i.e. the performance depends mainly on how fast the GPU or host PC can read and write data.

Raw data for GeForce GTX 960 - FFT (double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	51,200	0.09	0.56
4,096	245,760	0.12	2.03
16,384	1,146,880	0.14	8.36
65,536	5,242,880	0.33	15.79
262,144	23,592,960	0.84	28.17
1,048,576	104,857,600	3.04	34.49
4,194,304	461,373,440	12.31	37.49
16,777,216	2,013,265,920	54.61	36.87

(N gigaflops = $Nx10^9$ operations per second)



Results for MTimes (single)

These results show the performance of the GPU or host PC when calculating a matrix multiplication of two NxN real matrices. The number of operation assumed to be `2*N^3 - N^2`.

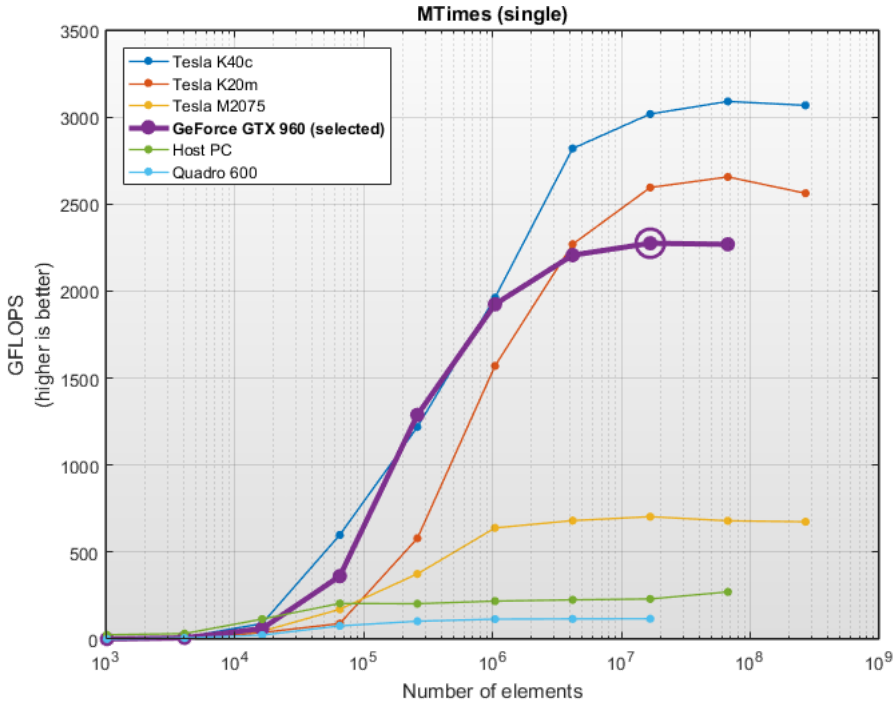
This calculation is usually compute-bound, i.e. the performance depends mainly on how fast the GPU or host PC can perform floating-point operations.

Raw data for GeForce GTX 960 - MTimes (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	64,512	0.08	0.79
4,096	520,192	0.07	7.19
16,384	4,177,920	0.07	60.49
65,536	33,488,896	0.09	362.21

262,144	268,173,312	0.21	1288.35
1,048,576	2,146,435,072	1.12	1924.36
4,194,304	17,175,674,880	7.78	2206.56
16,777,216	137,422,176,256	60.41	2274.96
67,108,864	1,099,444,518,912	484.59	2268.84

(N gigaflops = Nx10^9 operations per second)



Results for Backslash (single)

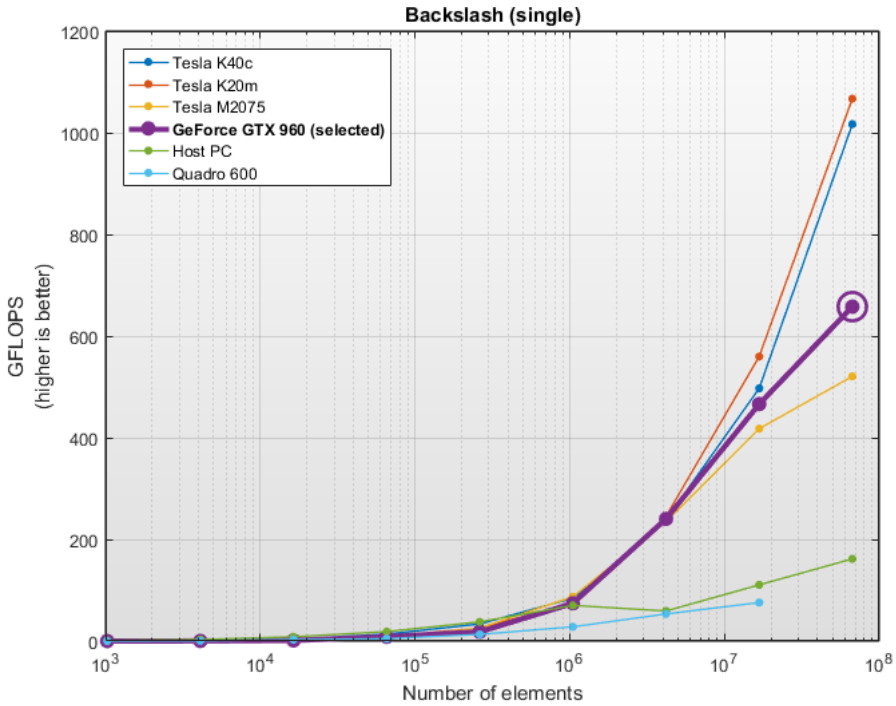
These results show the performance of the GPU or host PC when calculating the matrix left division of an NxN matrix with an Nx1 vector. The number o operations is assumed to be 2/3*N^3 + 3/2*N^2.

This calculation is usually compute-bound, i.e. the performance depends mainly on how fast the GPU or host PC can perform floating-point operations.

Raw data for GeForce GTX 960 - Backslash (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	23,381	0.79	0.03
4,096	180,907	0.93	0.19
16,384	1,422,677	1.25	1.14
65,536	11,283,115	1.34	8.43
262,144	89,871,701	4.95	18.14
1,048,576	717,400,747	9.66	74.25
4,194,304	5,732,914,517	23.84	240.44
16,777,216	45,838,150,315	98.34	466.14
67,108,864	366,604,539,221	557.27	657.86

(N gigaflops = Nx10^9 operations per second)



Results for FFT (single)

These results show the performance of the GPU or host PC when calculating the Fast-Fourier-Transform of a vector of complex numbers. The number operations for a vector of length N is assumed to be 5*N*log2(N).

This calculation is usually memory-bound, i.e. the performance depends mainly on how fast the GPU or host PC can read and write data.

Raw data for GeForce GTX 960 - FFT (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	51,200	0.08	0.61
4,096	245,760	0.09	2.74
16,384	1,146,880	0.12	9.53
65,536	5,242,880	0.14	36.84
262,144	23,592,960	0.33	70.49

1,048,576	104,857,600	1.04	100.66
4,194,304	461,373,440	3.37	136.95
16,777,216	2,013,265,920	12.88	156.26

(N gigaflops = $Nx10^9$ operations per second)