

GPU Bench

GPU Performance Details: Tesla K20m

- Contents:**
- [System Configuration](#)
 - Results for datatype double
 - [MTimes \(double\)](#)
 - [Backslash \(double\)](#)
 - [FFT \(double\)](#)
 - Results for datatype single
 - [MTimes \(single\)](#)
 - [Backslash \(single\)](#)
 - [FFT \(single\)](#)

System Configuration

⚠ Note that this is previously stored data and does not reflect your system configuration.

MATLAB Release: R2016a

Host

Name	Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz
Clock	2201 MHz
Cache	2048 KB
NumProcessors	16
OSType	Windows
OSVersion	Microsoft Windows 7 Enterprise

GPU

Name	Tesla K20m
Clock	7.055000e+02 MHz
NumProcessors	13
ComputeCapability	3.5
TotalMemory	4.69 GB
CUDAVersion	7.5
DriverVersion	8.17.13.5390 (353.90)

Results for MTimes (double)

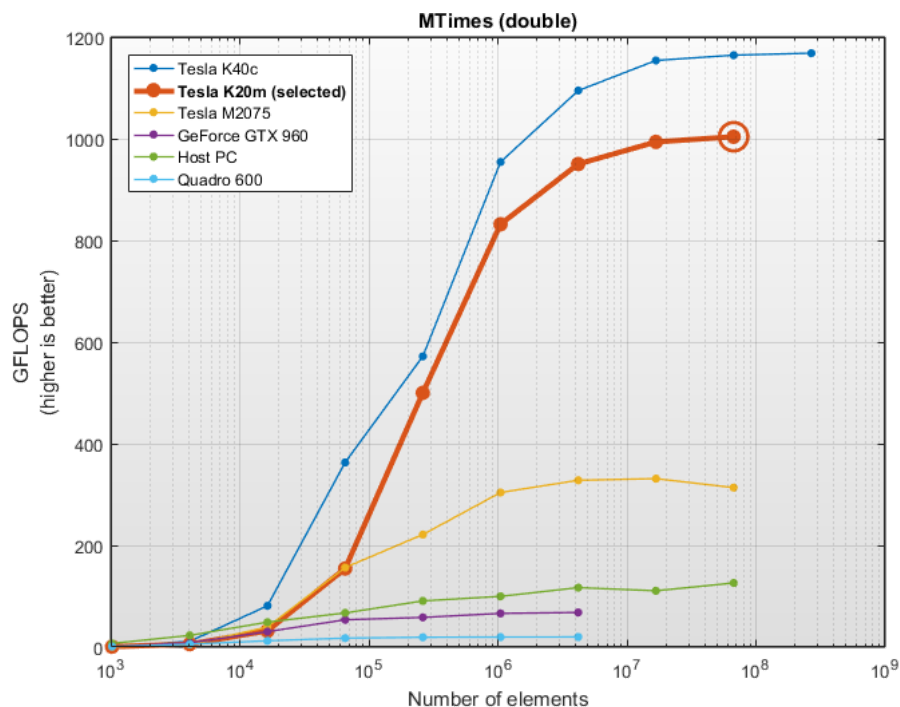
These results show the performance of the GPU or host PC when calculating a [matrix multiplication](#) of two NxN real matrices. The number of operation assumed to be $2 \times N^3 - N^2$.

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 Tesla K20m - MTimes (double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	64,512	0.10	0.68
4,096	520,192	0.09	5.73
16,384	4,177,920	0.13	31.94
65,536	33,488,896	0.22	154.40
262,144	268,173,312	0.54	500.15
1,048,576	2,146,435,072	2.58	831.82
4,194,304	17,175,674,880	18.07	950.30
16,777,216	137,422,176,256	138.29	993.74
67,108,864	1,099,444,518,912	1095.15	1003.93

(N gigaflops = $N \times 10^9$ operations per second)



Results for Backslash (double)

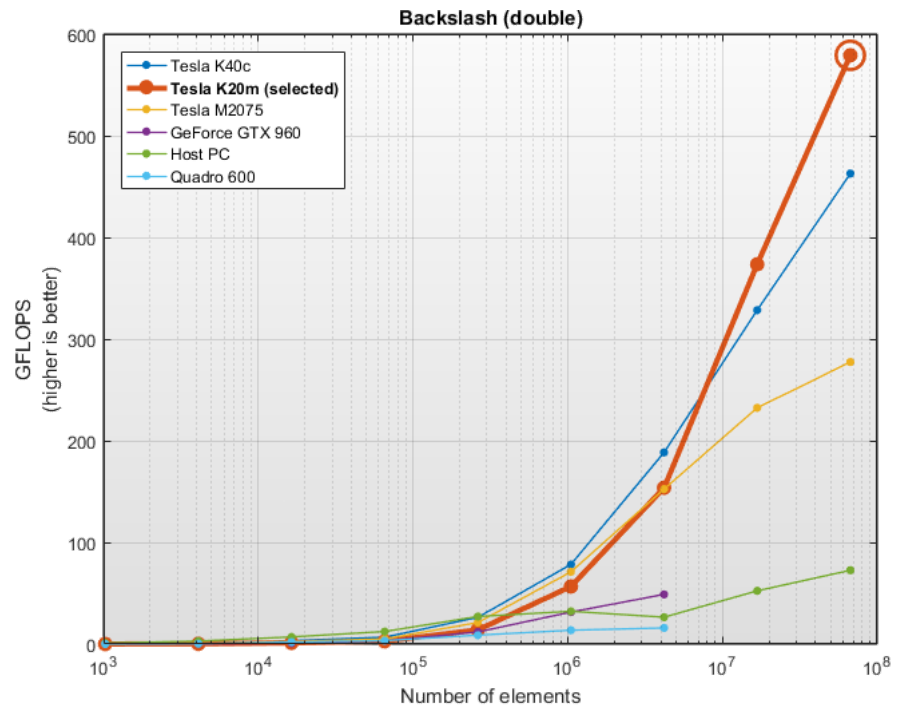
These results show the performance of the GPU or host PC when calculating the [matrix left division](#) of an NxN matrix with an Nx1 vector. The number of operations is assumed to be $\frac{2}{3} \times N^3 + \frac{3}{2} \times N^2$.

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 Tesla K20m - Backslash (double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	23,381	0.53	0.04
4,096	180,907	0.86	0.21
16,384	1,422,677	1.24	1.15
65,536	11,283,115	4.11	2.74
262,144	89,871,701	6.44	13.95
1,048,576	717,400,747	12.68	56.60
4,194,304	5,732,914,517	37.29	153.72
16,777,216	45,838,150,315	122.70	373.57
67,108,864	366,604,539,221	633.05	579.11

(N gigaflops = $N \times 10^9$ operations per second)



Results for FFT (double)

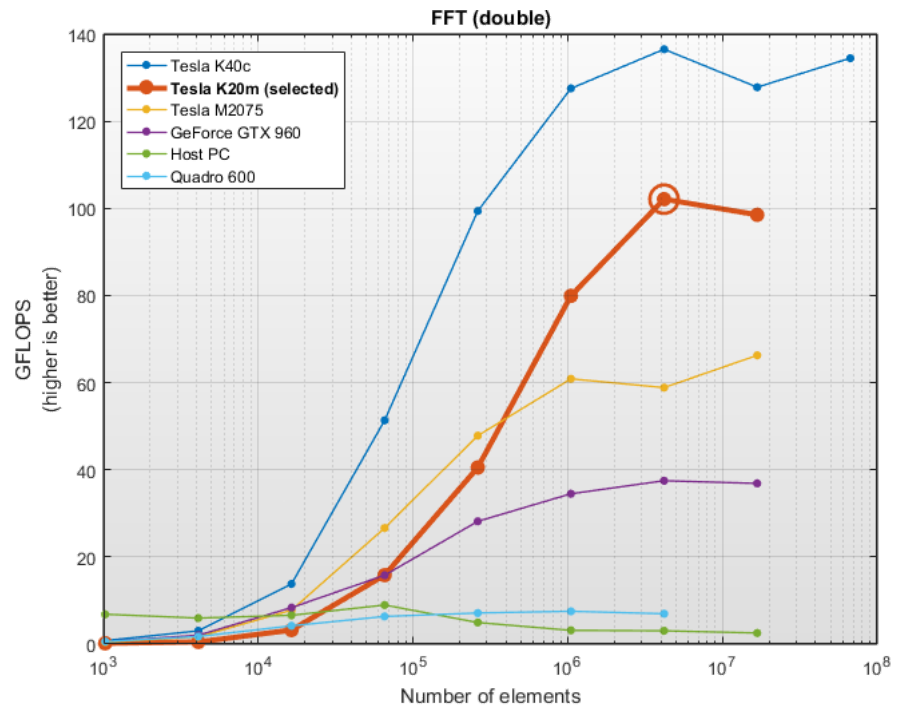
These results show the performance of the GPU or host PC when calculating the [Fast-Fourier-Transform](#) of a vector of complex numbers. The number operations for a vector of length N is assumed to be $5 \cdot N \cdot \log_2(N)$.

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 Tesla K20m - FFT (double)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	51,200	0.28	0.18
4,096	245,760	0.52	0.47
16,384	1,146,880	0.36	3.21
65,536	5,242,880	0.33	15.82
262,144	23,592,960	0.58	40.52
1,048,576	104,857,600	1.31	79.92
4,194,304	461,373,440	4.52	102.11
16,777,216	2,013,265,920	20.44	98.52

(N gigaflops = $N \times 10^9$ 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 $2 \cdot N^3 - N^2$.

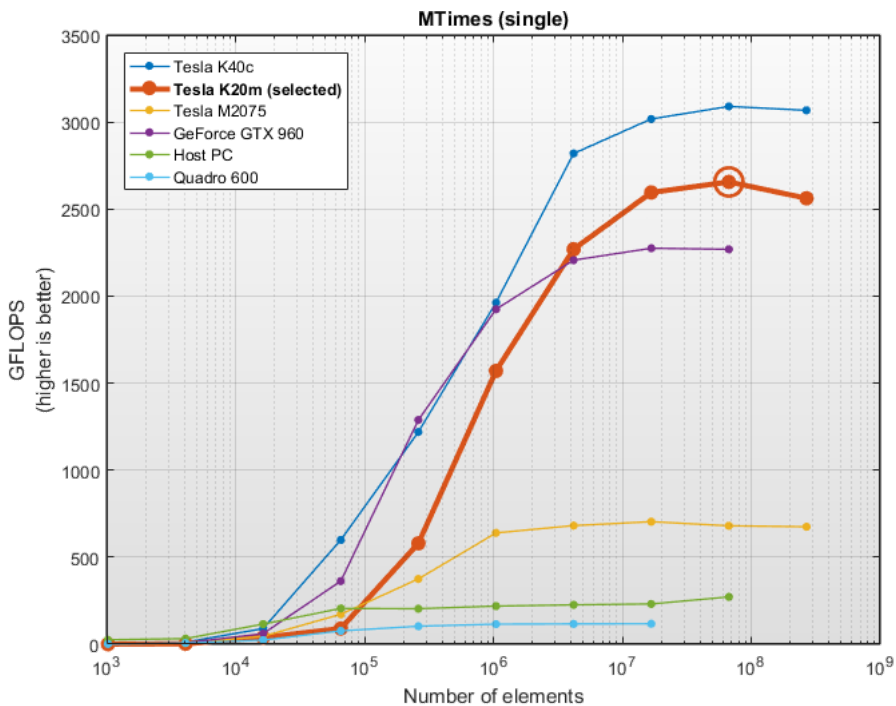
This calculation is usually compute-bound, i.e. the performance depends mainly on how fast the GPU or host PC can perform floating-point operations.

Raw data for Tesla K20m - MTimes (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	64,512	0.35	0.18
4,096	520,192	0.23	2.29
16,384	4,177,920	0.11	38.03

65,536	33,488,896	0.37	89.83
262,144	268,173,312	0.46	579.65
1,048,576	2,146,435,072	1.37	1570.18
4,194,304	17,175,674,880	7.57	2269.35
16,777,216	137,422,176,256	52.95	2595.38
67,108,864	1,099,444,518,912	413.96	2655.92
268,435,456	8,795,824,586,752	3433.08	2562.08

(N gigaflops = $N \times 10^9$ operations per second)



Results for Backslash (single)

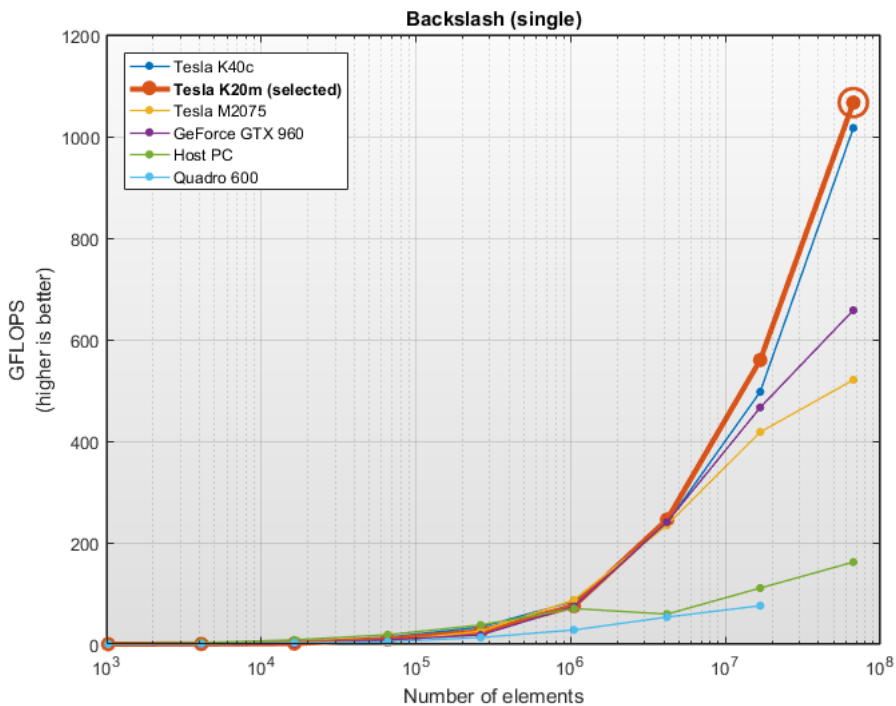
These results show the performance of the GPU or host PC when calculating the [matrix left division](#) of an $N \times N$ matrix with an $N \times 1$ vector. The number of operations is assumed to be $\frac{2}{3} * N^3 + \frac{3}{2} * N^2$.

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 Tesla K20m - Backslash (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	23,381	0.94	0.02
4,096	180,907	0.60	0.30
16,384	1,422,677	1.06	1.34
65,536	11,283,115	1.19	9.45
262,144	89,871,701	3.56	25.24
1,048,576	717,400,747	9.55	75.15
4,194,304	5,732,914,517	23.36	245.41
16,777,216	45,838,150,315	81.90	559.68
67,108,864	366,604,539,221	343.62	1066.88

(N gigaflops = $N \times 10^9$ 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 of operations for a vector of length N is assumed to be $5 * N * \log_2(N)$.

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 Tesla K20m - FFT (single)

Array size (elements)	Num Operations	Time (ms)	GigaFLOPS
1,024	51,200	0.38	0.13
4,096	245,760	0.31	0.79
16,384	1,146,880	0.14	8.00
65,536	5,242,880	0.14	36.45
262,144	23,592,960	0.20	120.69

1,048,576	104,857,600	0.82	128.44
4,194,304	461,373,440	2.38	194.17
16,777,216	2,013,265,920	8.67	232.15
67,108,864	8,724,152,320	40.17	217.16

(N gigaflops = $N \times 10^9$ operations per second)

