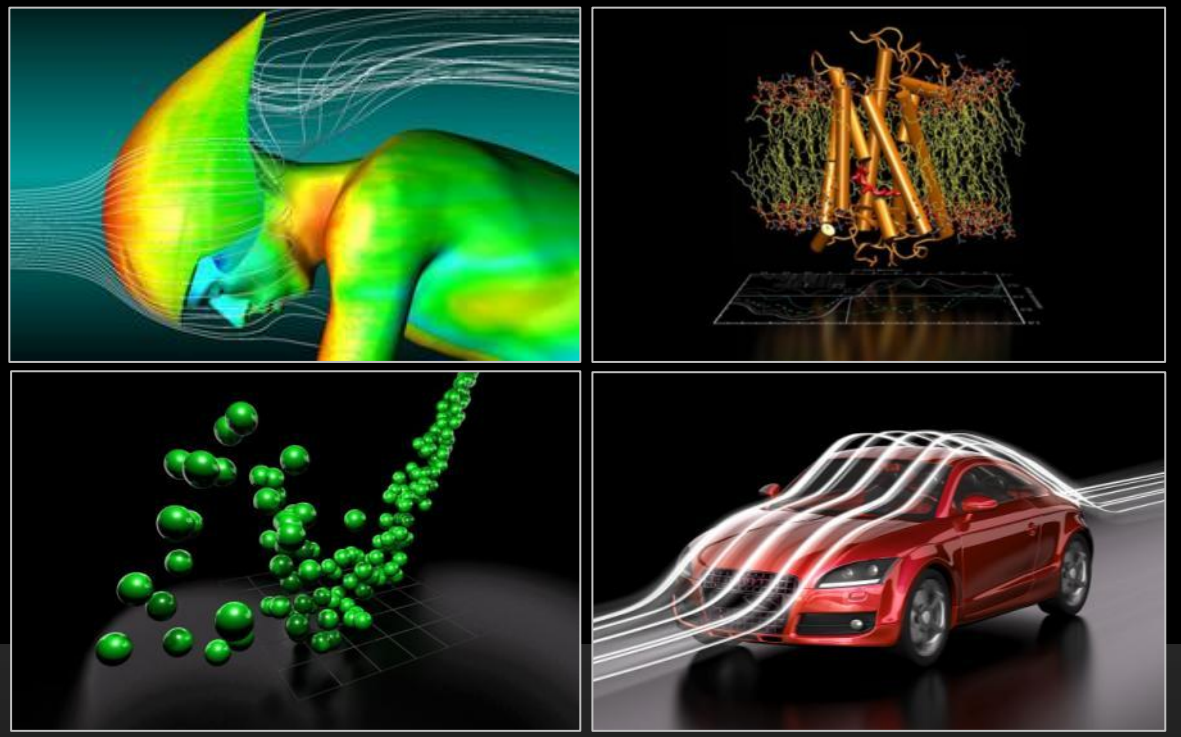
TESLA
GPU Computing



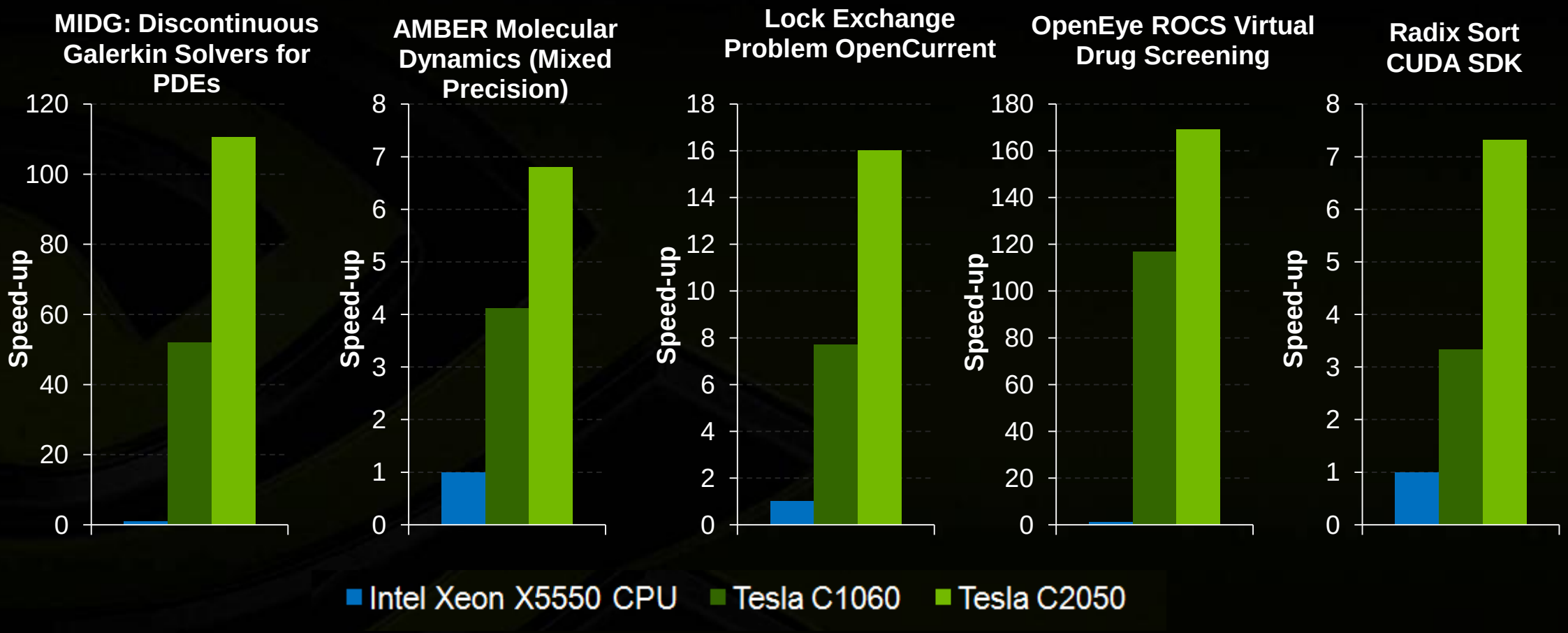
Tesla C2050 Performance Benchmarks

Tesla C-Series Workstation GPUs



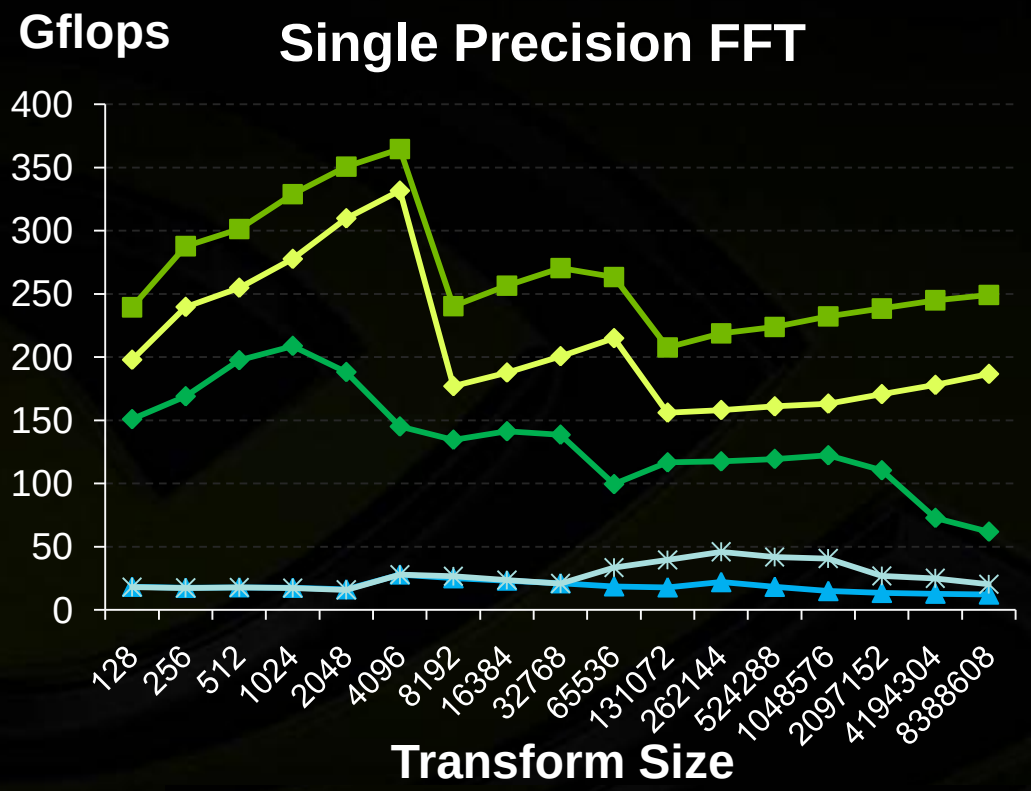
	Tesla C1060	Tesla C2050	Tesla C2070
Architecture	Tesla 10-series GPU	Tesla 20-series GPU	
Number of Cores	240	448	
Caches	16 KB Shared Memory / 8 cores	64 KB L1 cache + Shared Memory / 32 cores, 768 KB L2 cache	
PCI-e DMA Engines	1	2	
Floating Point Peak Performance	933 Gflops (single) 78 Gflops (double)	1030 Gflops (single) 515 Gflops (double)	
GPU Memory	4 GB	3 GB 2.625 GB w/ ECC on	6 GB 5.25 GB w/ ECC on
Memory Bandwith	102 GB/s (GDDR3)	144 GB/s (GDDR5)	
System I/O	PCIe x16 Gen2	PCIe x16 Gen2	
Power	188 W (max)	247 W (max)	247 W (max)
Available	Available now	Available now	Q3 2010

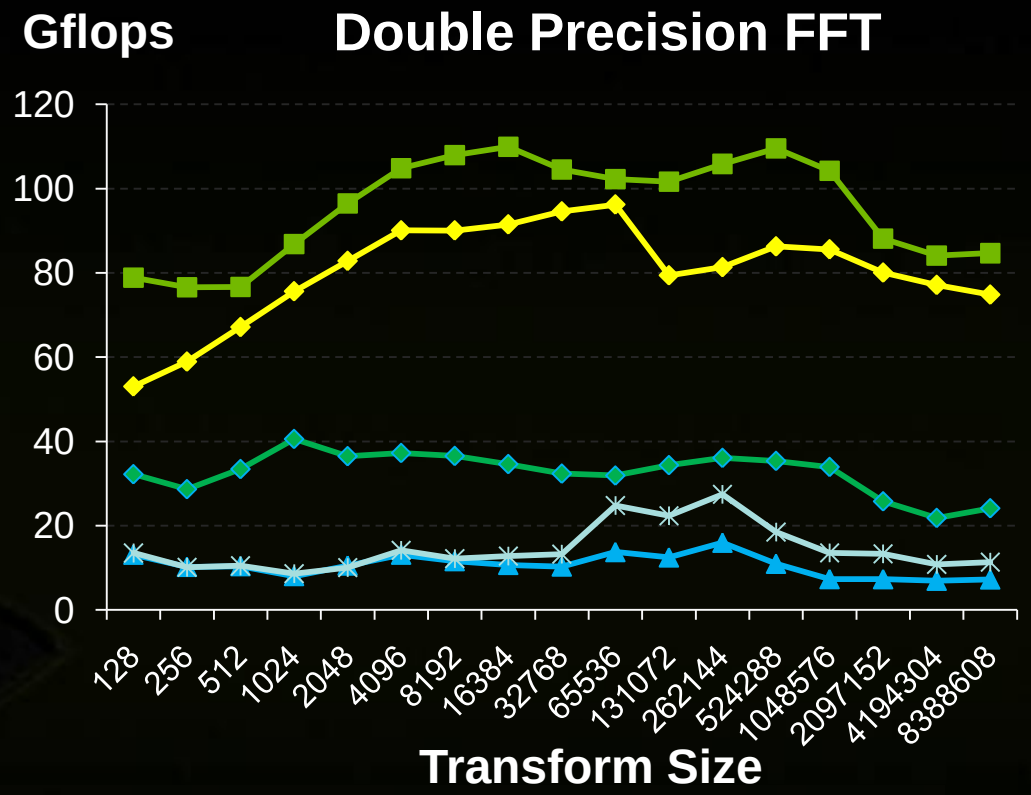
Performance Summary



Preliminary data

Standard FFT Library: cuFFT 3.1



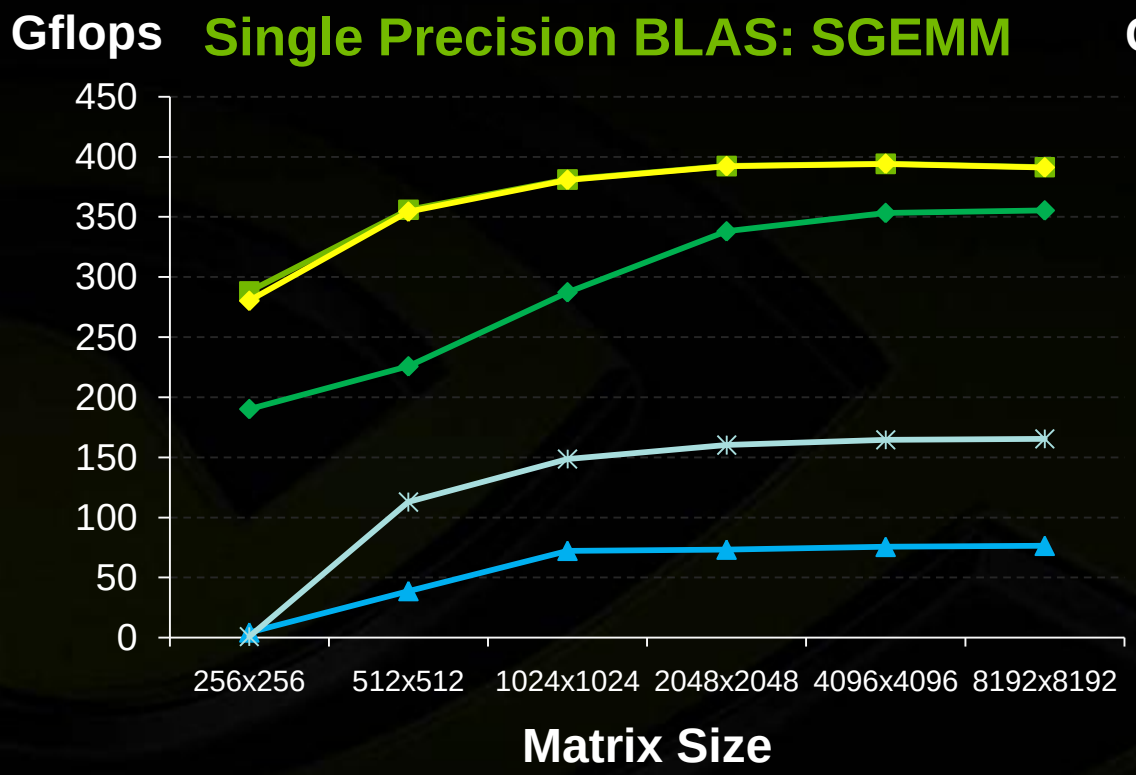


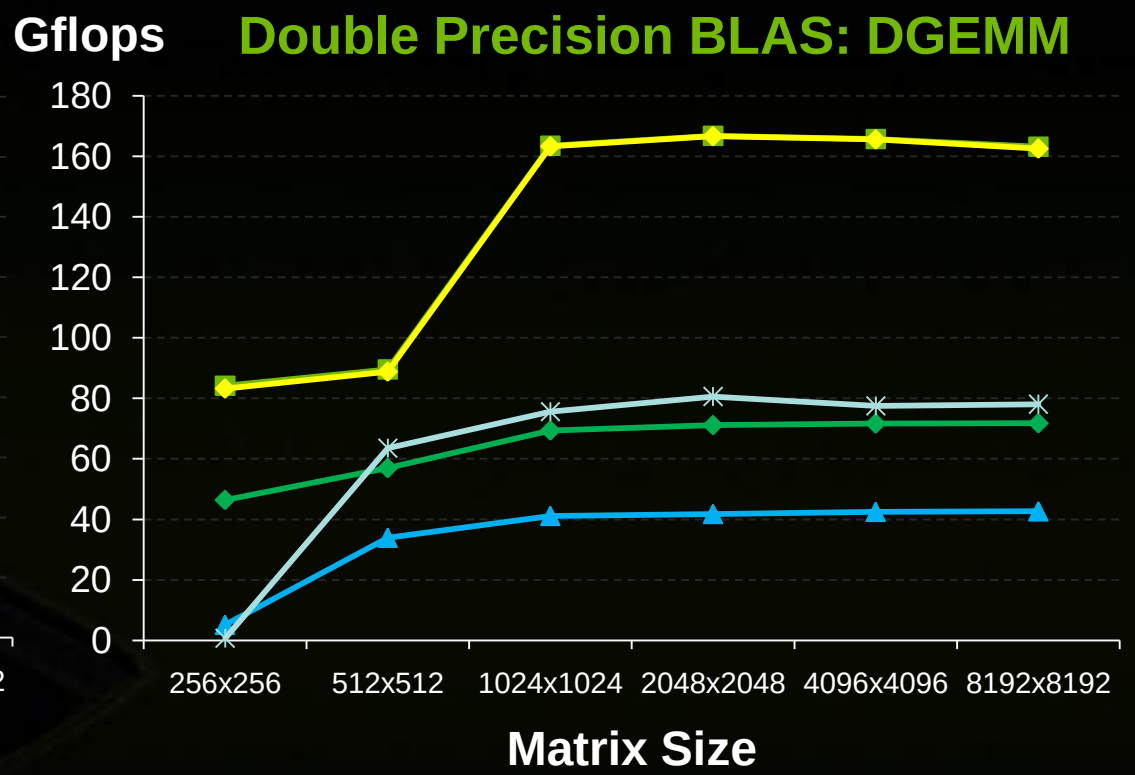
Tesla C2050 (ECC off) — Tesla C2050 (ECC on) — Tesla C1060 — MKL 4 Threads (4 cores) — MKL 8 Threads (8 cores)

cuFFT 3.1: NVIDIA Tesla C1060, Tesla C2050 (Fermi)

MKL 10.2.4.32: Quad-Core Intel Xeon 5550, 2.67 GHz

Standard BLAS Library: cuBLAS 3.1



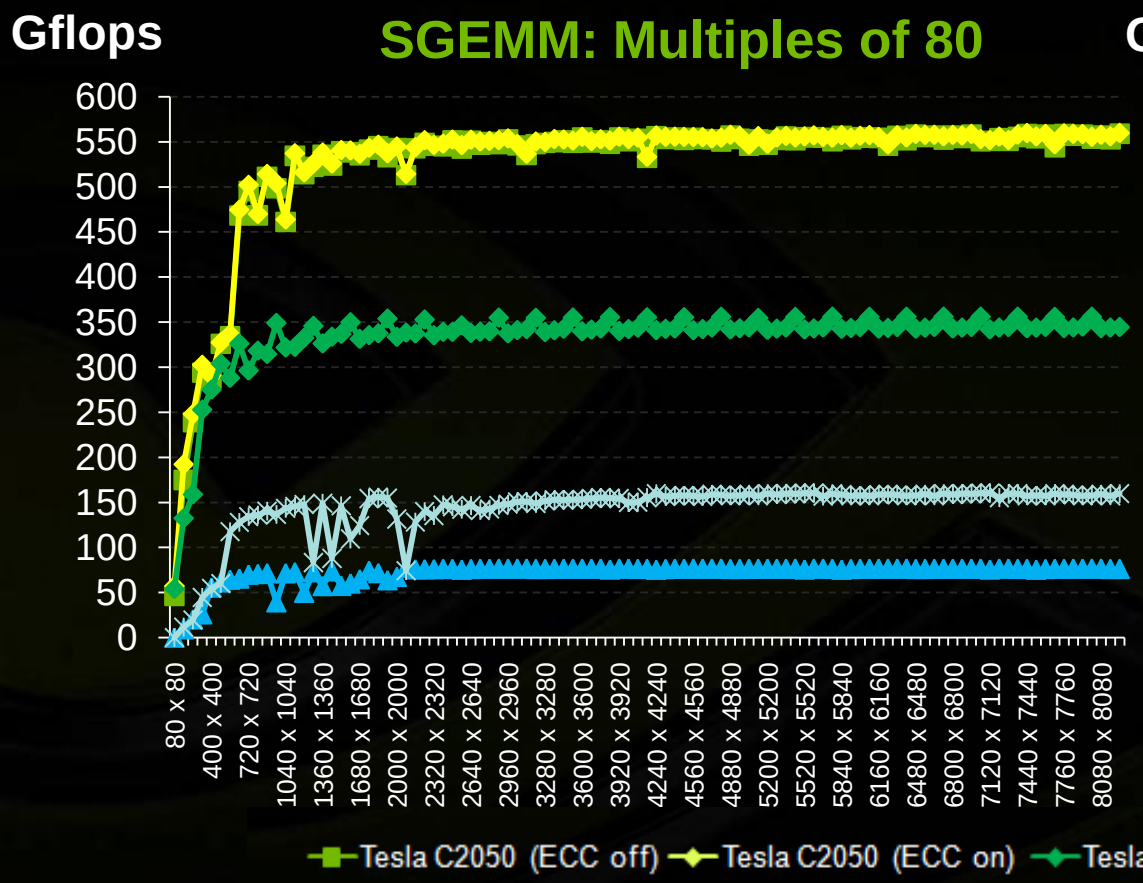


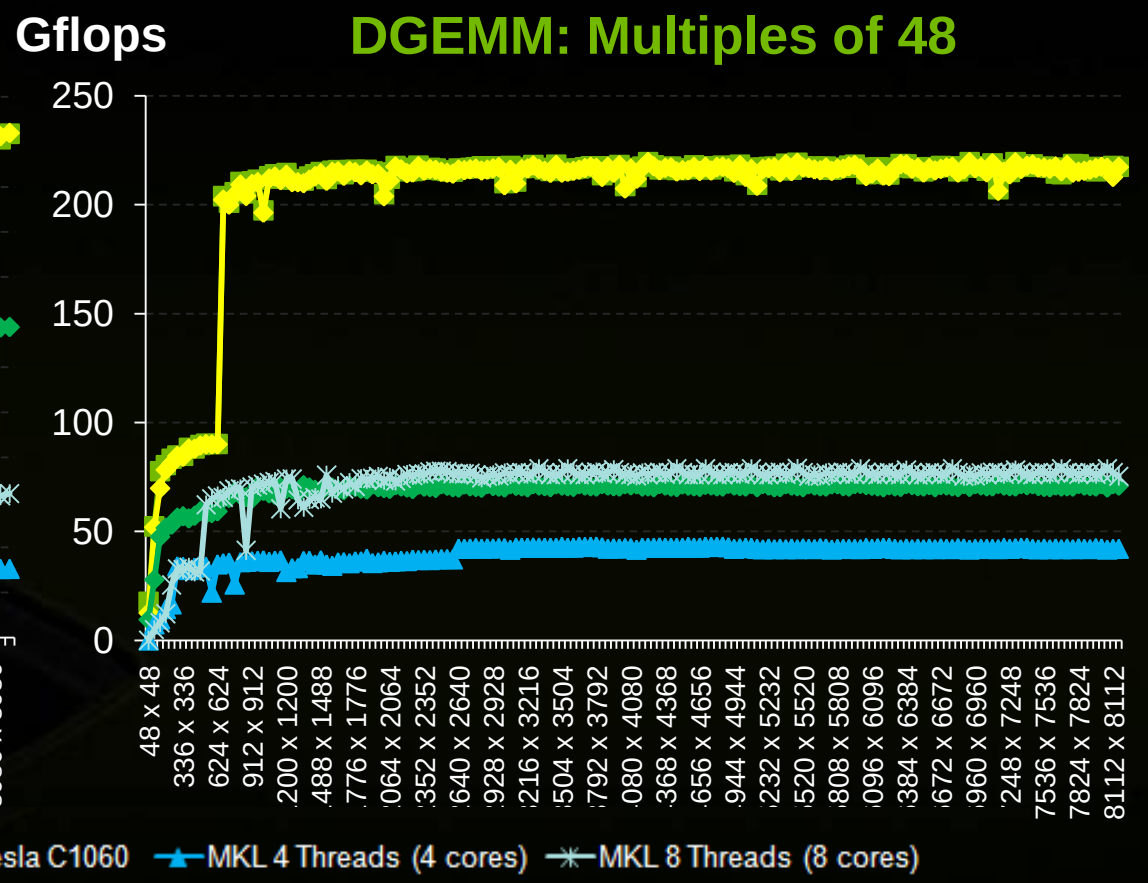
Tesla C2050 (ECC off) Tesla C2050 (ECC on) Tesla C1060 MKL 4 Threads (4 cores) MKL 8 Threads (8 cores)

cuBLAS 3.1: NVIDIA Tesla C1060, Tesla C2050 (Fermi)

MKL 10.2.4.32: Quad-Core Intel Xeon 5550, 2.67 GHz

Matrix Size for Best cuBLAS Performance



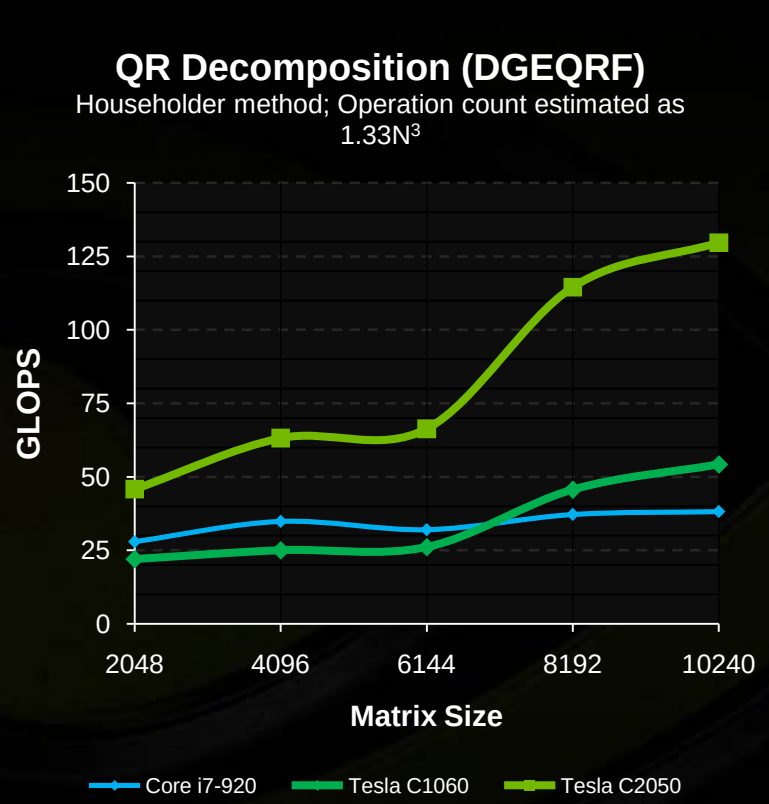


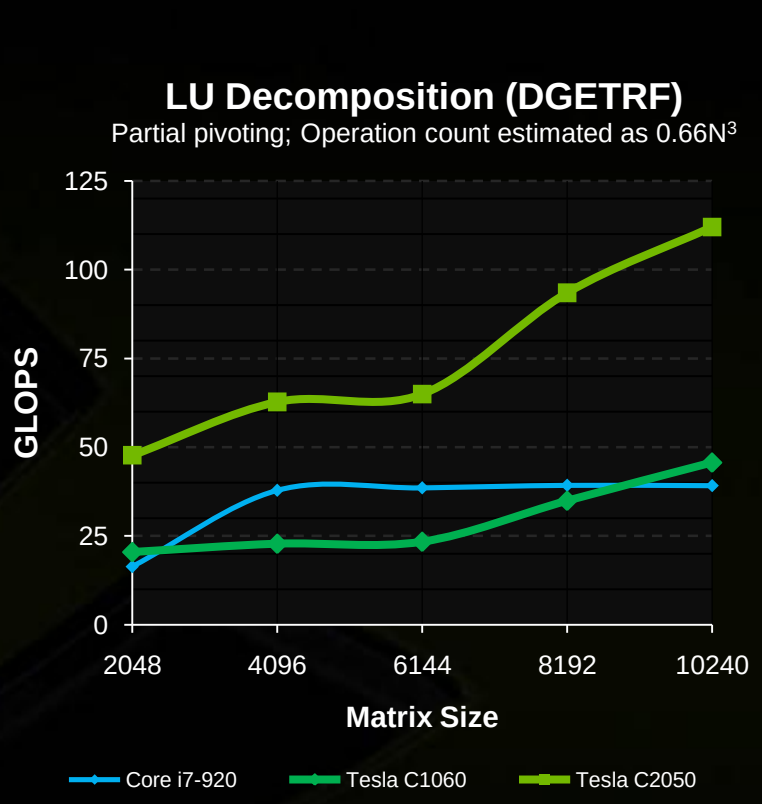
Tesla C2050 (ECC off) — Tesla C2050 (ECC on) — Tesla C1060 — MKL 4 Threads (4 cores) — MKL 8 Threads (8 cores)

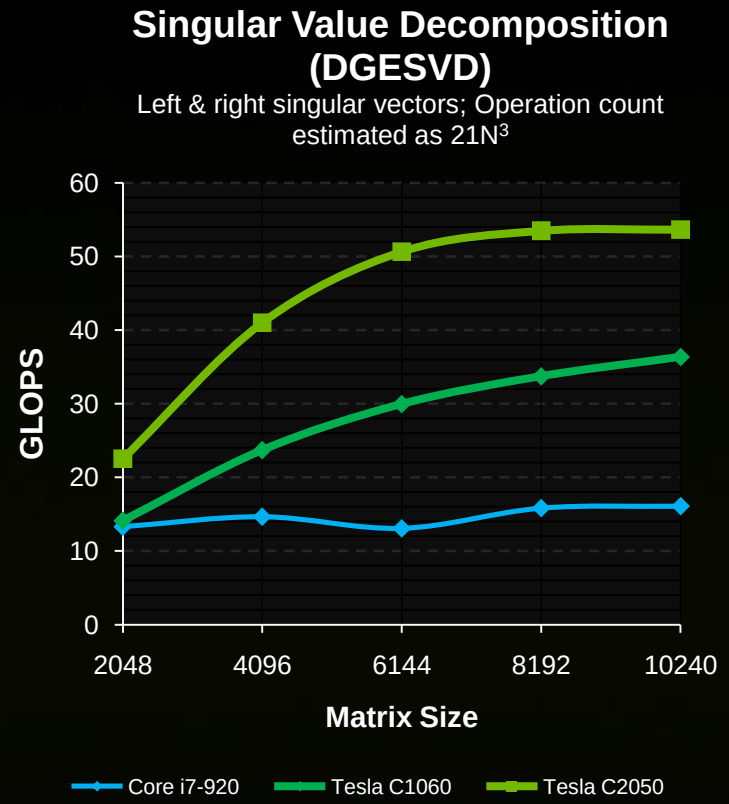
cuBLAS 3.1: NVIDIA Tesla C1060, Tesla C2050 (Fermi)

MKL 10.2.4.32: Quad-Core Intel Xeon 5550, 2.67 GHz

CULA 1.3 LAPACK Library from EM Photonics





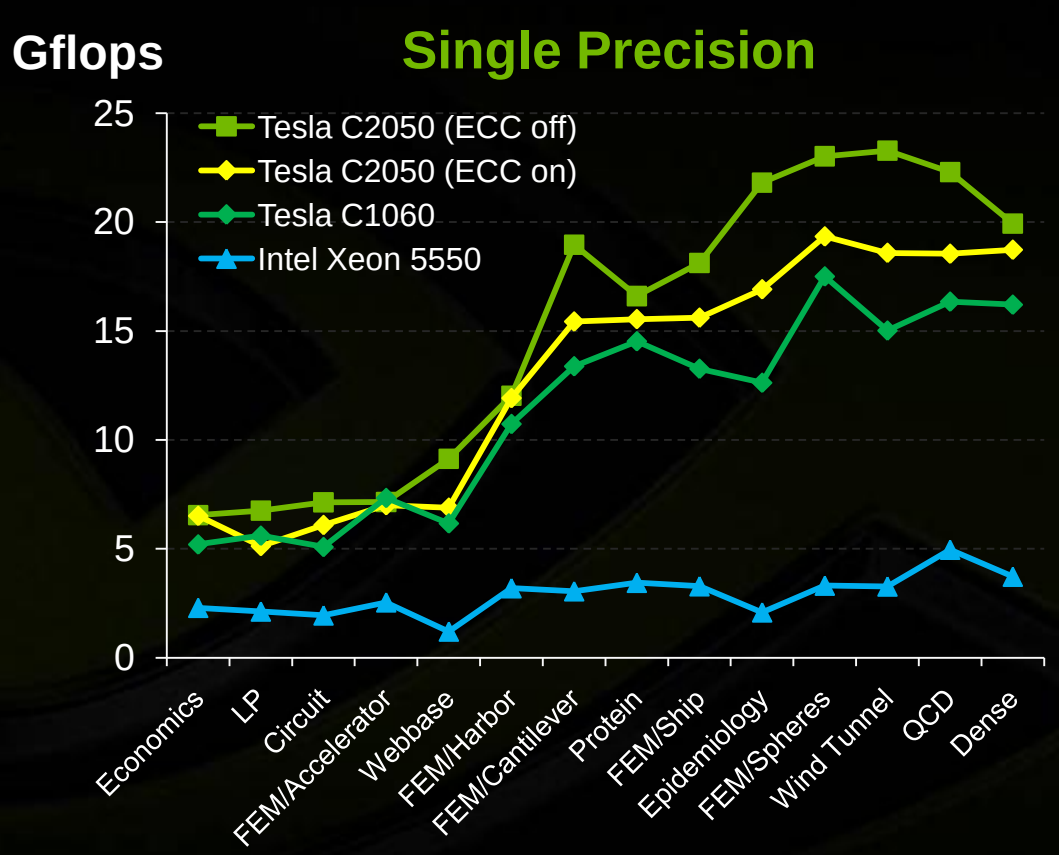


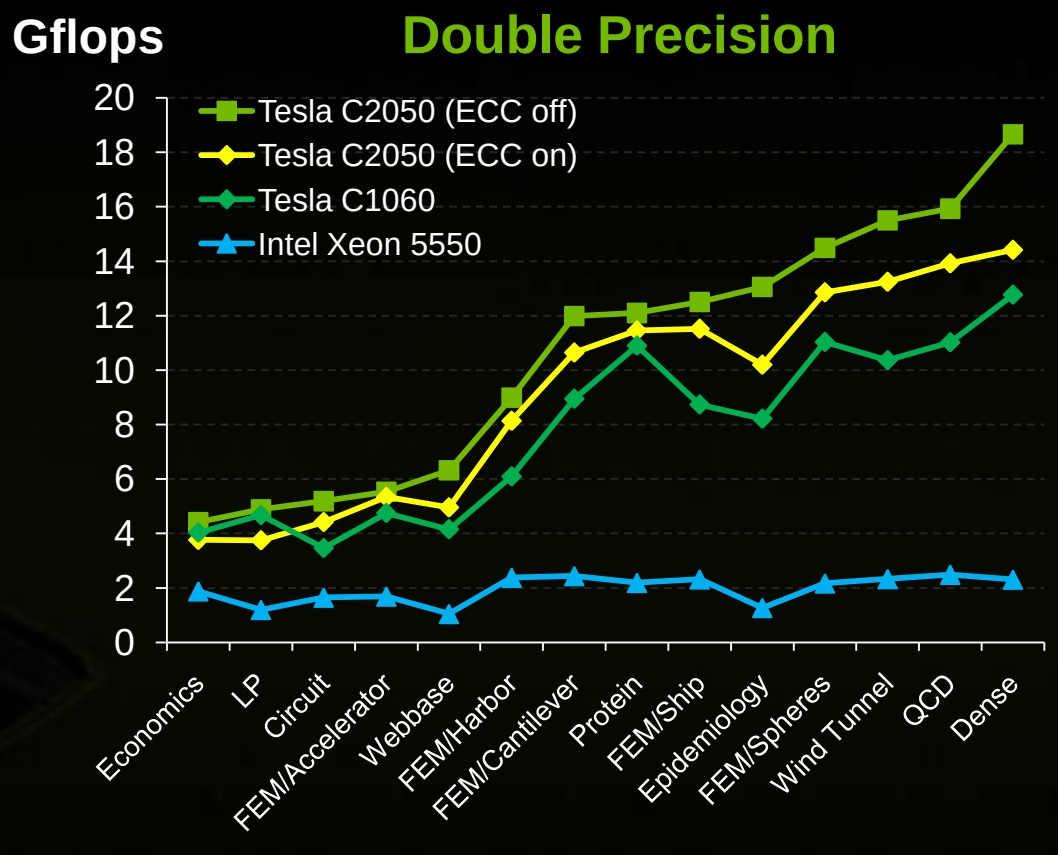
Double Precision Results: CUDA 3.0

Data Courtesy: EM Photonics, Preliminary data

EM Photonics

Sparse Matrix-Vector Multiplication (SpMV)

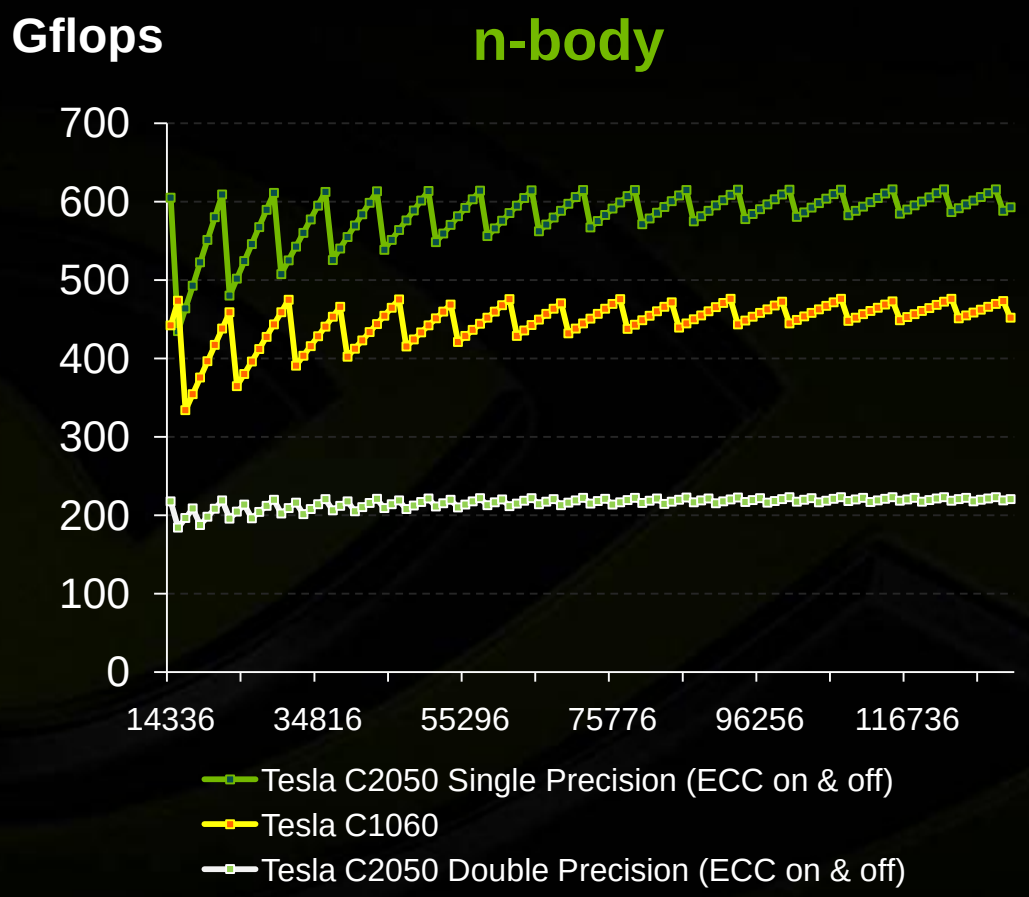


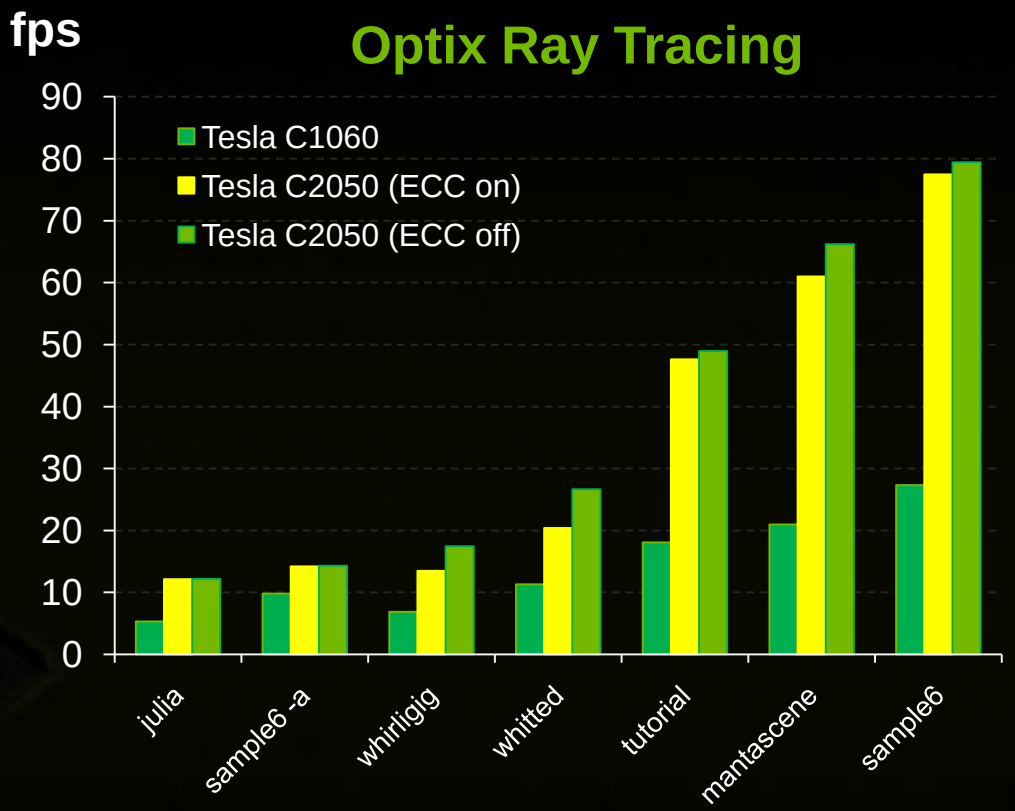


SpMv: CUDA 3.0, Tesla C1060 and Tesla C2050
MKL 10.2: Intel Xeon 5550, 2.67 GHz

Preliminary data

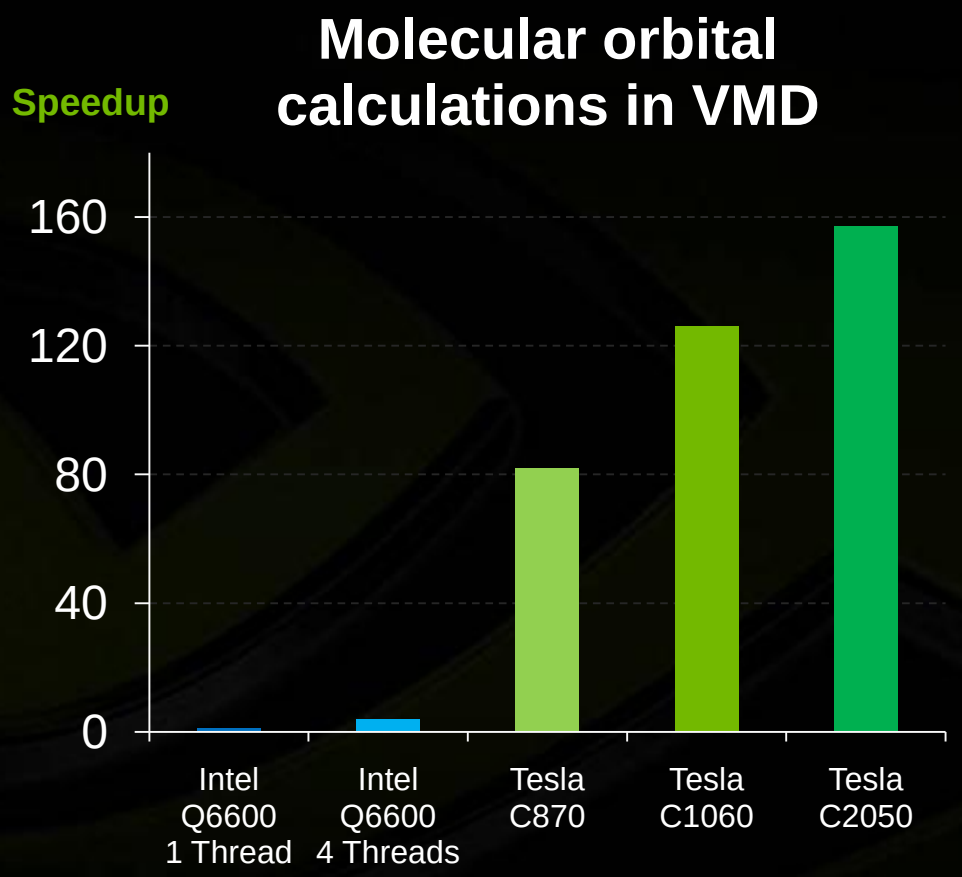
N-body and Ray Tracing Performance

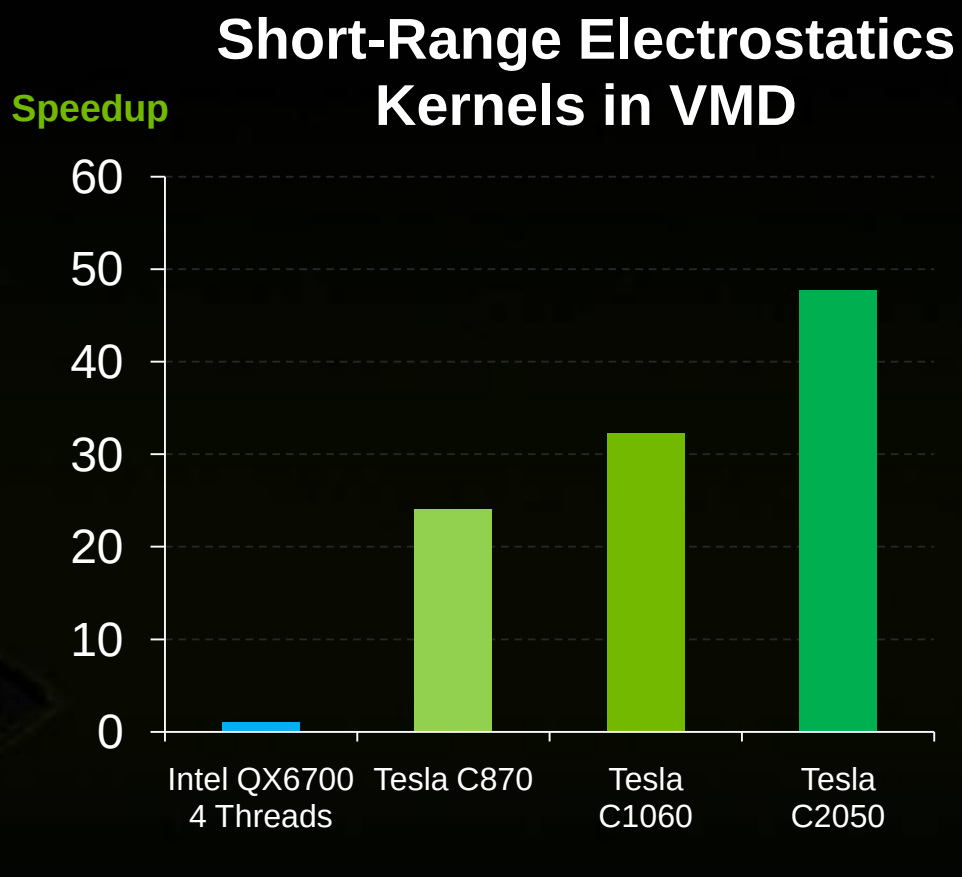




CUDA 3.0, Tesla C1060 and Tesla C2050

Consistent Speedups over 3 Generations

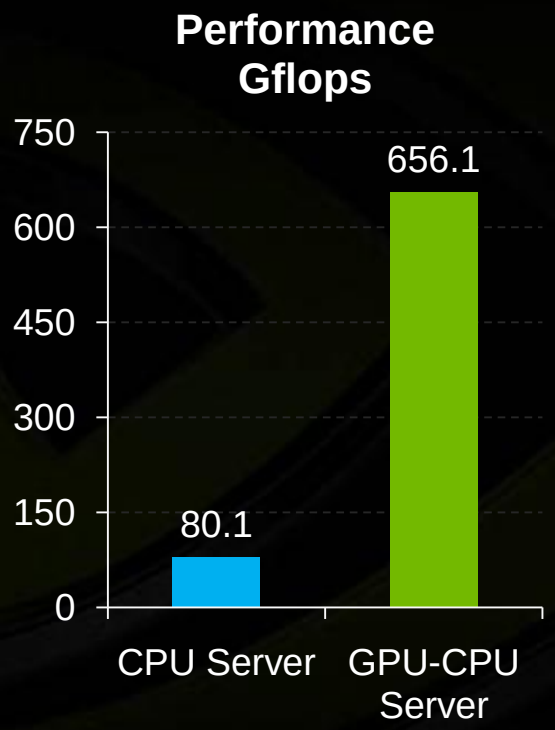


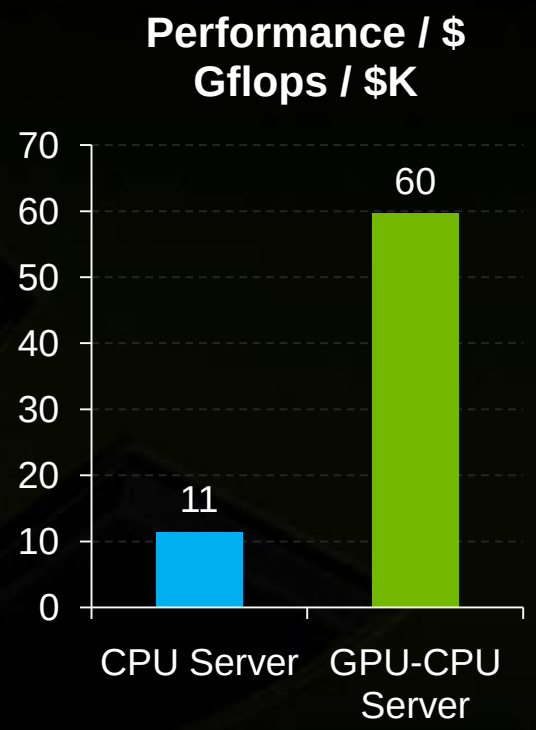


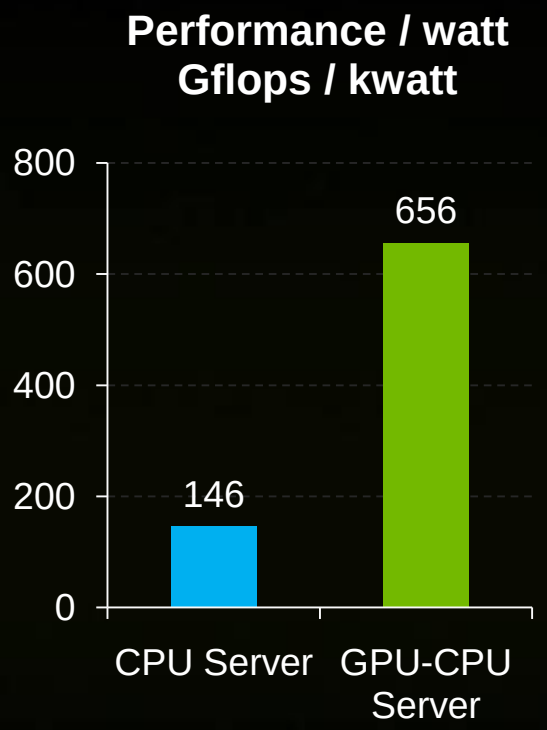
http://www.ks.uiuc.edu/Research/vmd/

Preliminary data: Subject to Change

8x Higher Linpack







CPU 1U Server: 2x Intel Xeon X5550 (Nehalem) 2.66 GHz, 48 GB memory, $7K, 0.55 kw

GPU-CPU 1U Server: 2x Tesla C2050 + 2x Intel Xeon X5550, 48 GB memory, $11K, 1.0 kw

“In testing our key applications, the Tesla GPUs delivered speed-ups that we had never seen before, sometimes even orders of magnitude.”

Satoshi Matsuoka
Professor
Tokyo Institute of Technology

“Future computing architectures will be hybrid systems with parallel-core GPUs working in tandem with multi-core CPUs”

Jack Dongarra
Professor, University of Tennessee
Author of Linpack