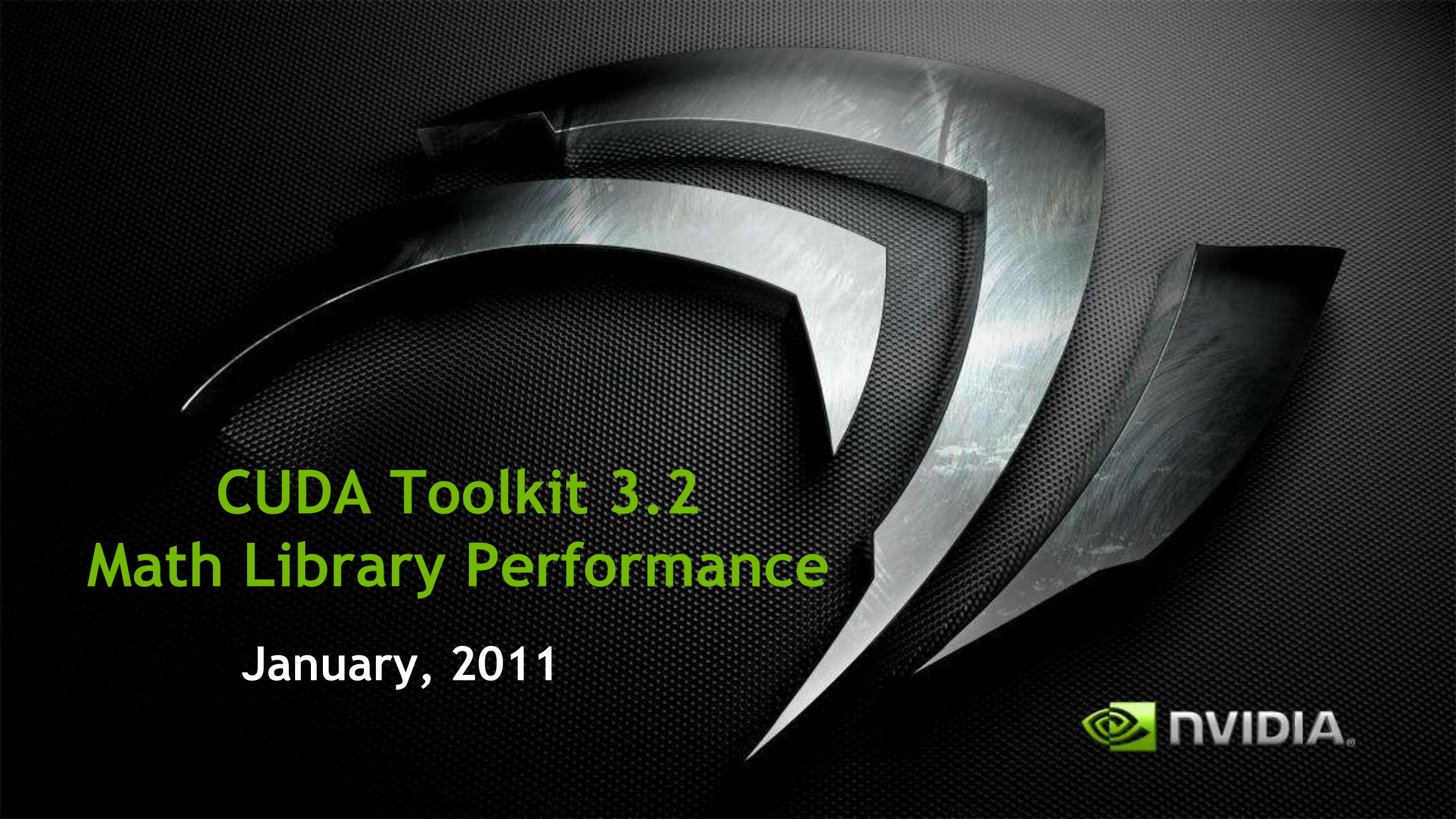

CUDA Toolkit 3.2
Math Library Performance
January, 2011
NVIDIA

CUDA Math Libraries

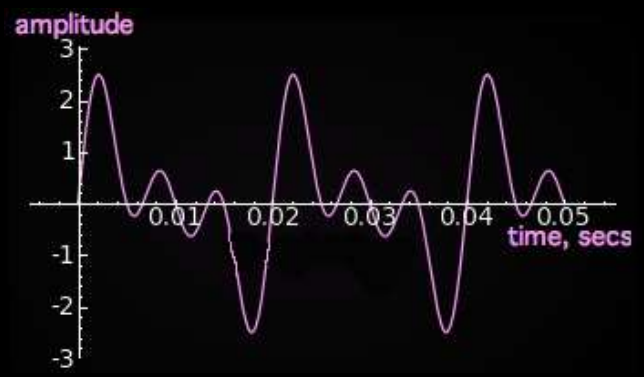
High performance math routines for your applications:

- cuFFT – Fast Fourier Transforms Library
- cuBLAS – Complete BLAS Library
- cuSPARSE – Sparse Matrix Library
- cuRAND – Random Number Generation (RNG) Library

- Included in the CUDA Toolkit (free download)
  - www.nvidia.com/getcuda
- For more information on CUDA libraries:
  - http://www.nvidia.com/object/gtc2010-presentation-archive.html#session2216

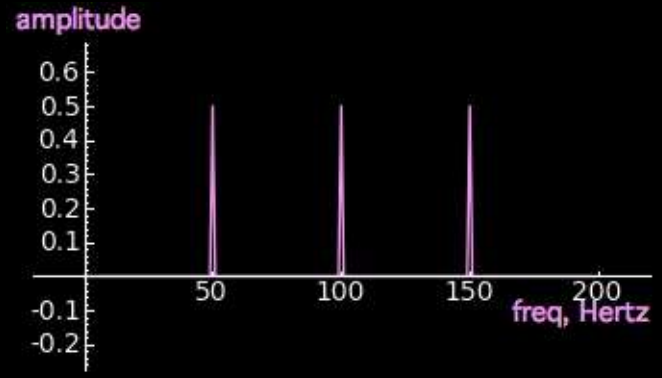
cuFFT

- Multi-dimensional Fast Fourier Transforms
- New in CUDA 3.2:
  - Higher performance of 1D, 2D, 3D transforms with dimensions of powers of 2, 3, 5 or 7
  - Higher performance and accuracy for 1D transform sizes that contain large prime factors



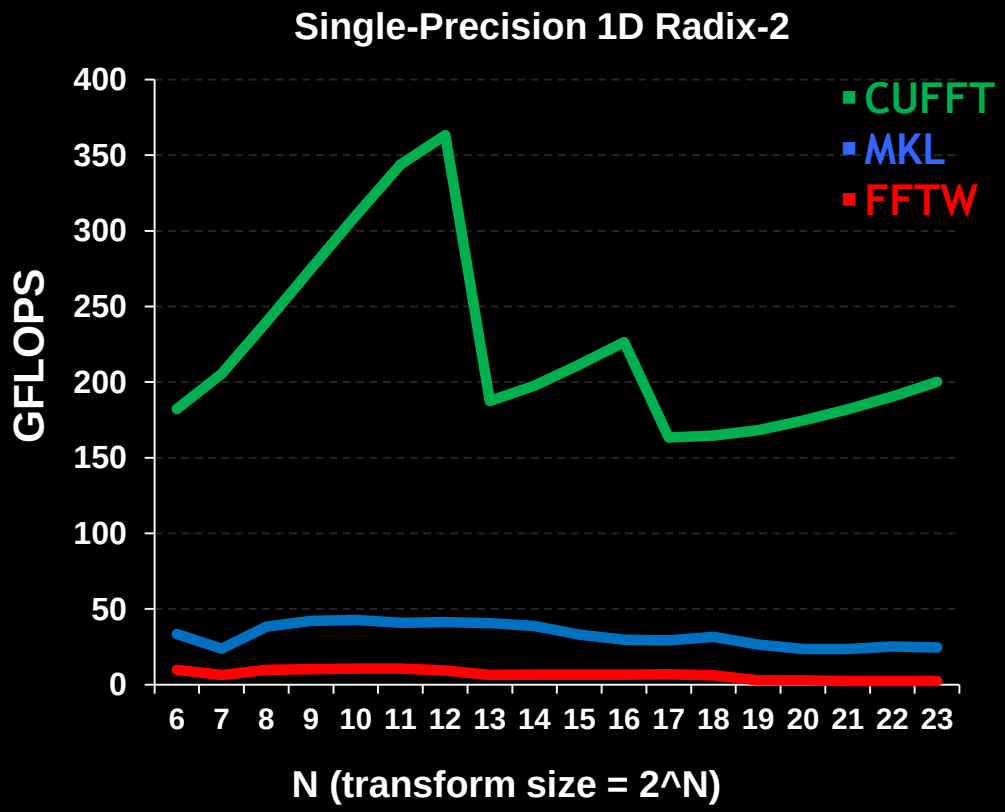
$$F(x) = \sum_{n=0}^{N-1} f(n) e^{-j2\pi(x\frac{n}{N})}$$

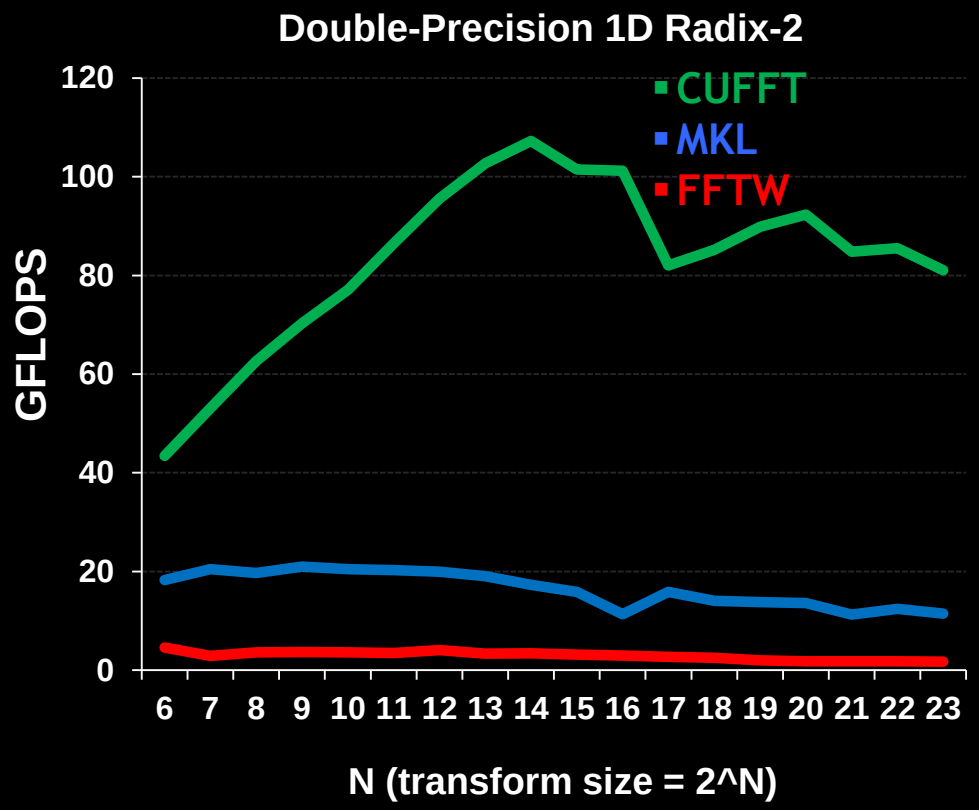
$$f(n) = \frac{1}{N} \sum_{n=0}^{N-1} F(x) e^{j2\pi(x\frac{n}{N})}$$



FFTs up to 8.8x Faster than MKL

1D used in audio processing and as a foundation for 2D and 3D FFTs



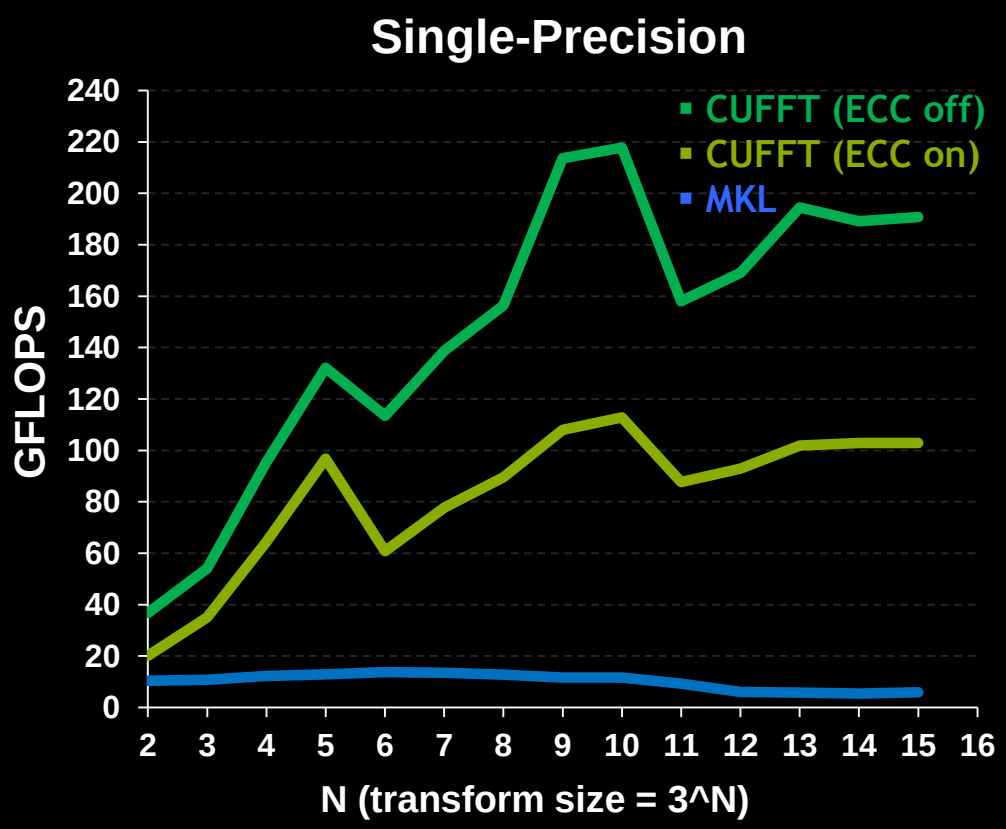


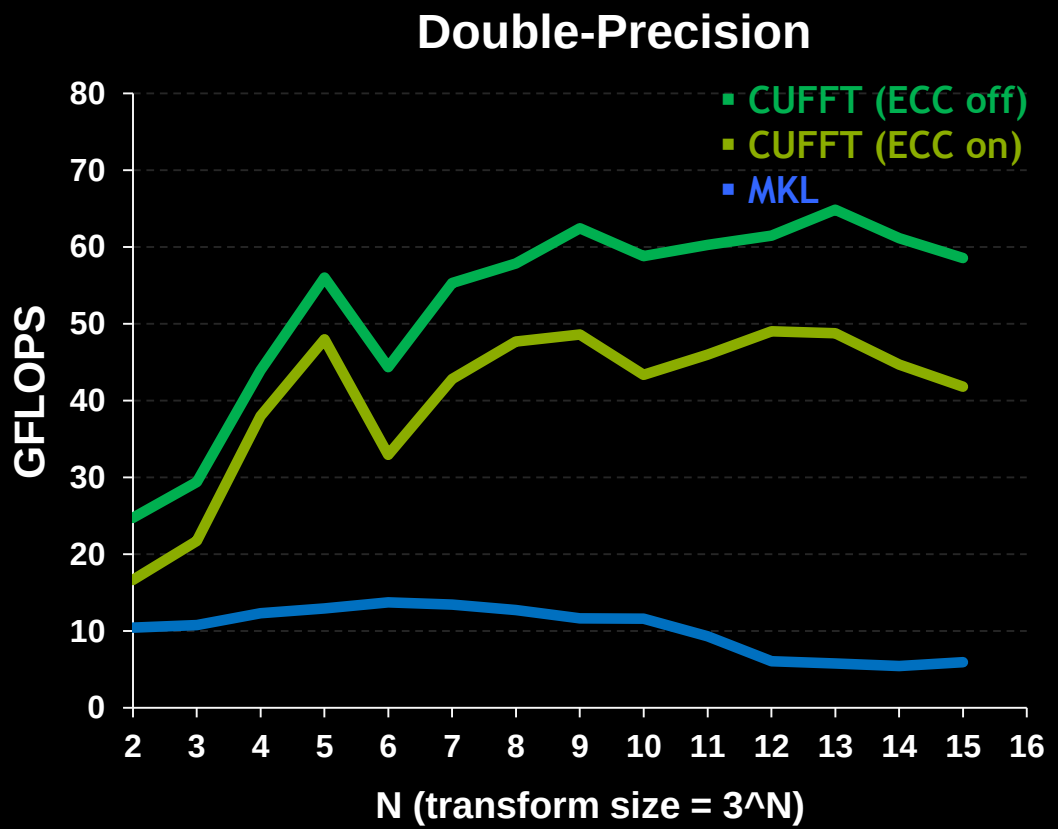
* cuFFT 3.2, Tesla C2050 (Fermi) with ECC on
* MKL 10.1r1, 4-core Corei7 Nehalem @ 3.07GHz
* FFTW single-thread on same CPU

Performance may vary based on OS version and motherboard configuration

cuFFT 1D Radix-3 up to 18x Faster than MKL

18x for single-precision, 15x for double-precision
Similar acceleration for radix-5 and -7





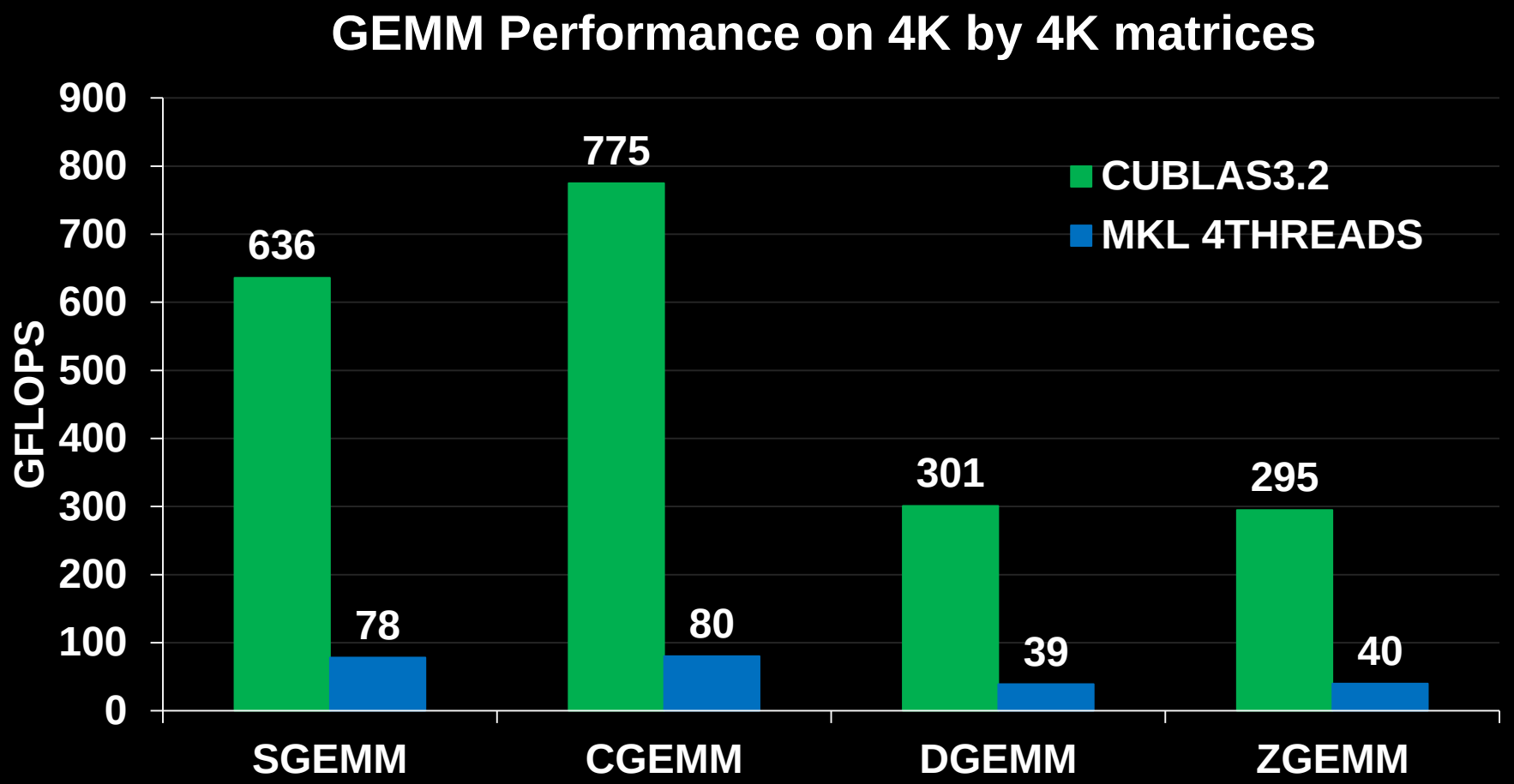
* CUFFT 3.2 on Tesla C2050 (Fermi)
* MKL 10.1r1, 4-core Corei7 Nehalem @ 3.07GHz

Performance may vary based on OS version and motherboard configuration

cuBLAS: Dense Linear Algebra on GPUs

- Complete BLAS implementation
  - Supports all 152 routines for single, double, complex and double complex
- New in CUDA 3.2
  - 7x Faster GEMM (matrix multiply) on Fermi GPUs
  - Higher performance on SGEMM & DGEMM for all matrix sizes and all transpose cases (NN, TT, TN, NT)

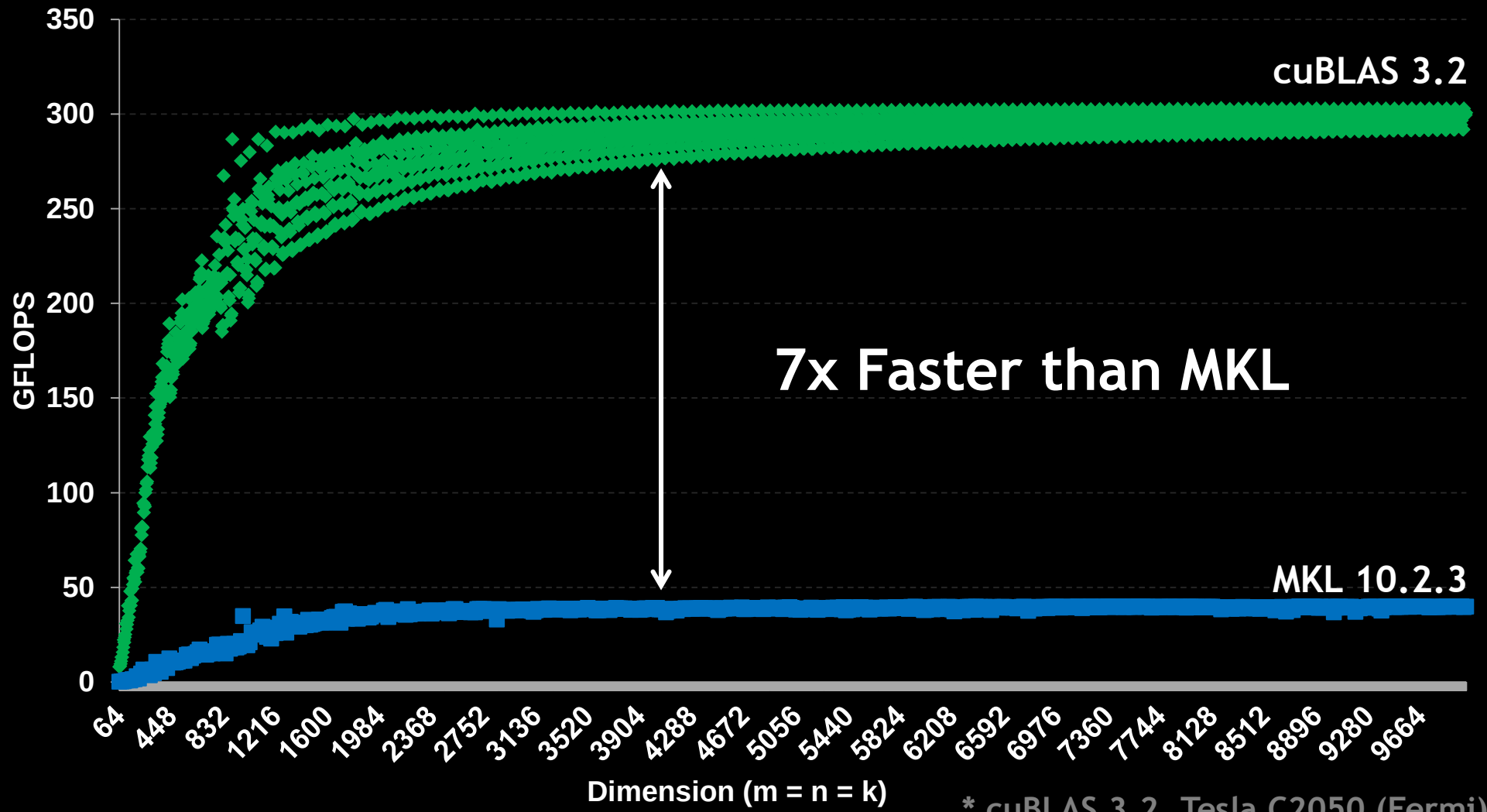
Up to 8x Speedup for all GEMM Types



* cuBLAS 3.2, Tesla C2050 (Fermi), ECC on
* MKL 10.2.3, 4-core Corei7 @ 2.66Ghz

Performance may vary based on OS version and motherboard configuration

7x Faster DGEMM Performance

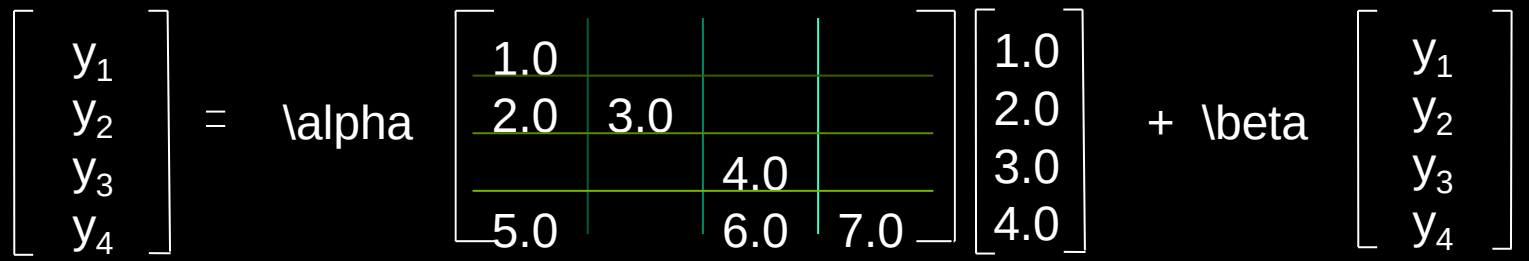


* cuBLAS 3.2, Tesla C2050 (Fermi), ECC on
* MKL 10.2.3, 4-core Corei7 @ 2.66Ghz

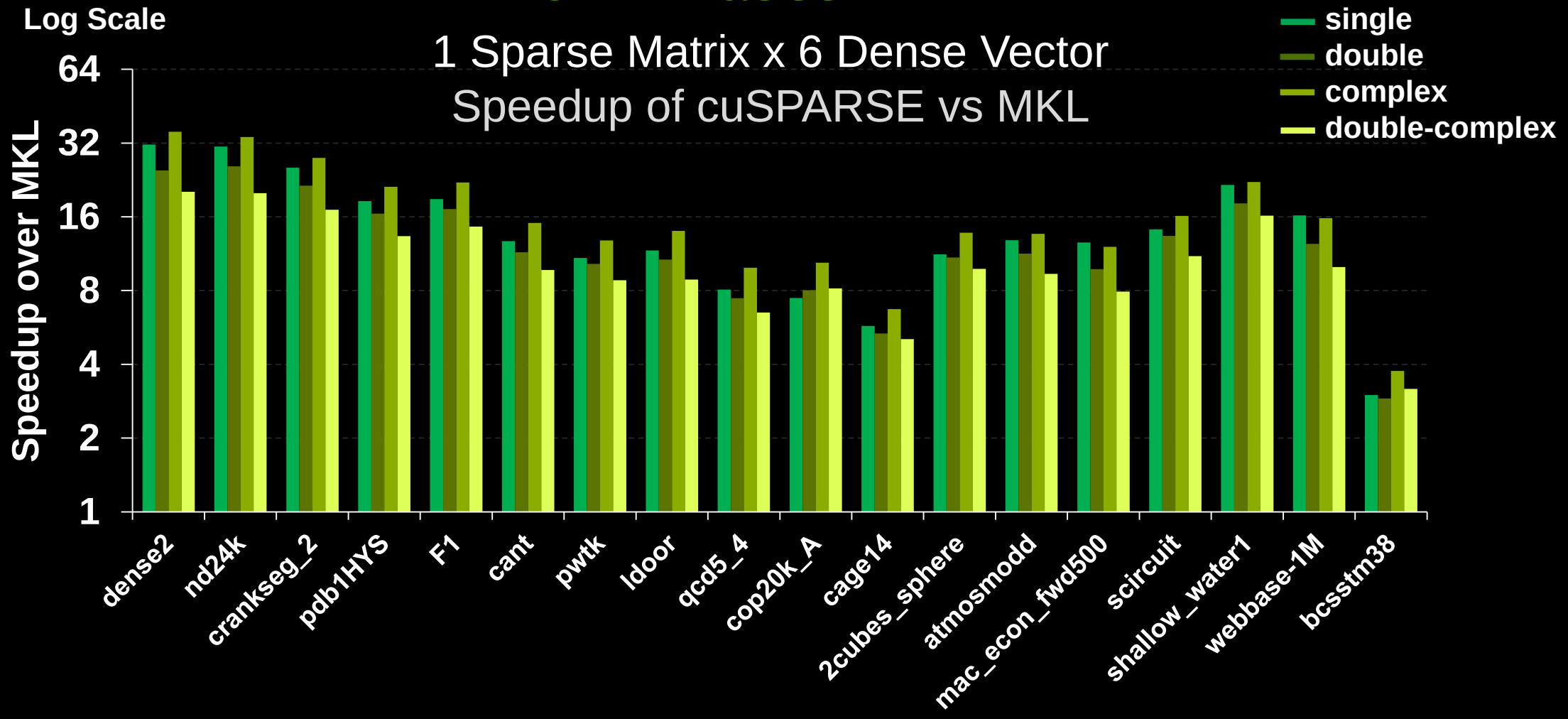
Performance may vary based on OS version and motherboard configuration

cuSPARSE

- New library for sparse linear algebra
- Conversion routines for dense, COO, CSR and CSC formats
- Optimized sparse matrix-vector multiplication



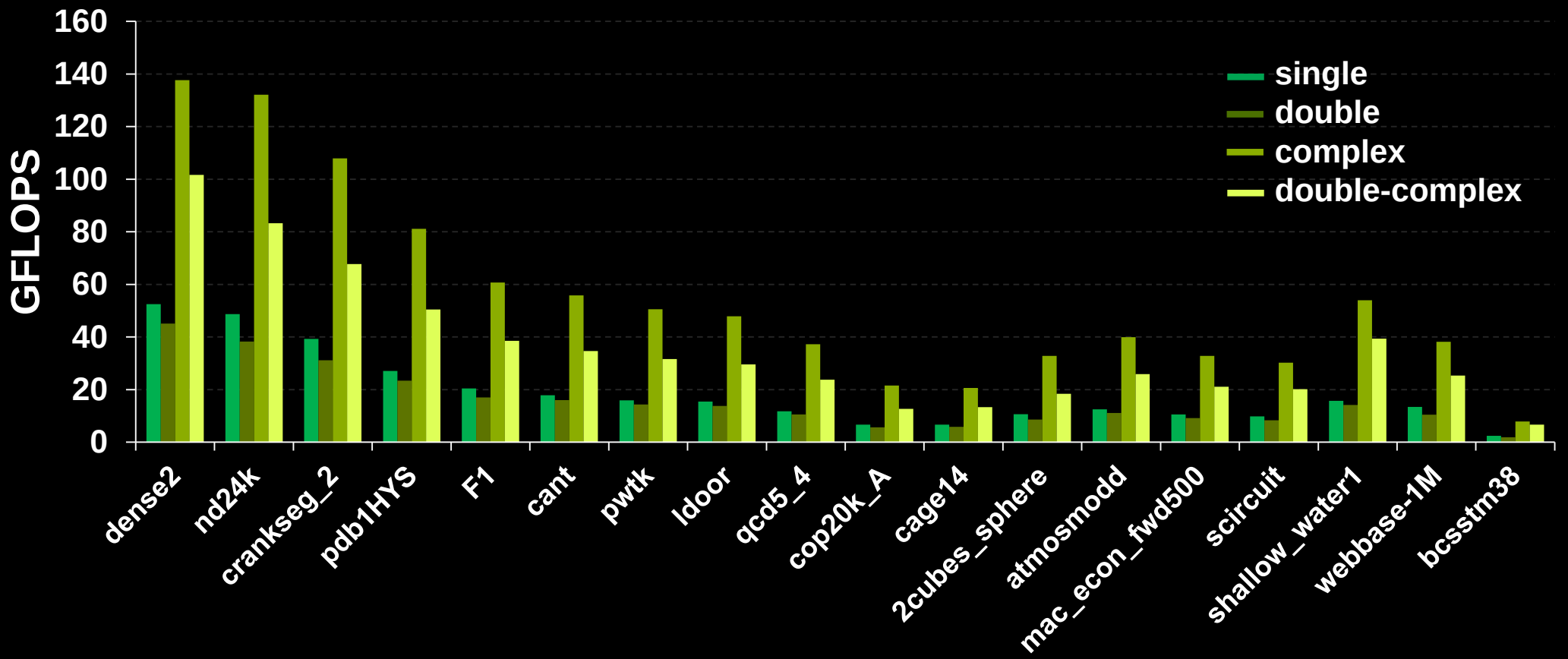
32x Faster



* CUSPARSE 3.2, NVIDIA C2050 (Fermi), ECC on
* MKL 10.2.3, 4-core Corei7 @ 3.07GHz

Performance may vary based on OS version and motherboard configuration

Performance of 1 Sparse Matrix x 6 Dense Vectors

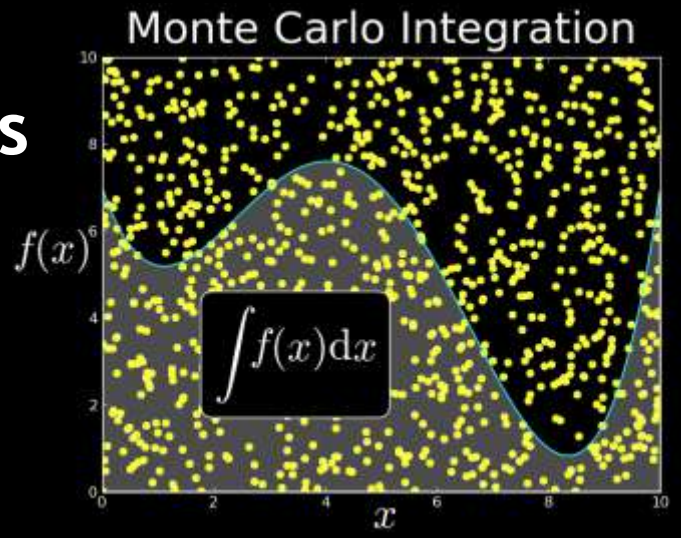


Test cases roughly in order of increasing sparseness

Performance may vary based on OS version and motherboard configuration

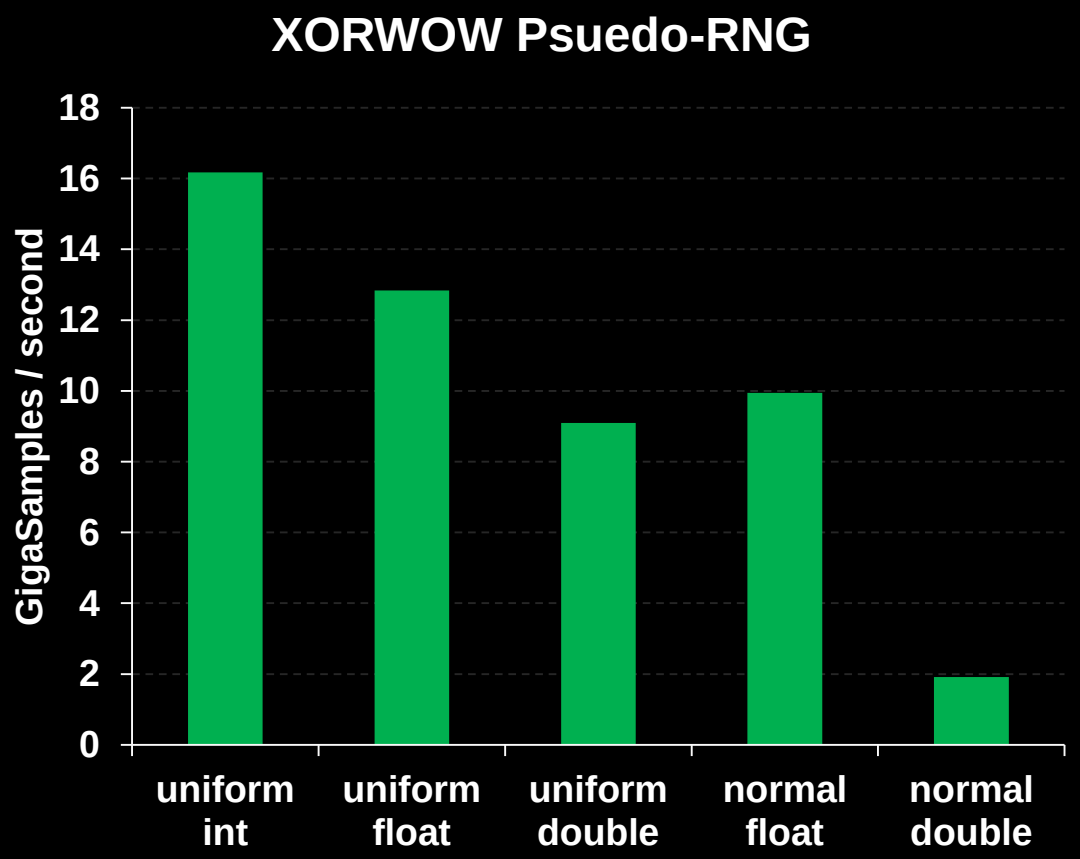
* CUSPARSE 3.2, NVIDIA C2050 (Fermi), ECC on

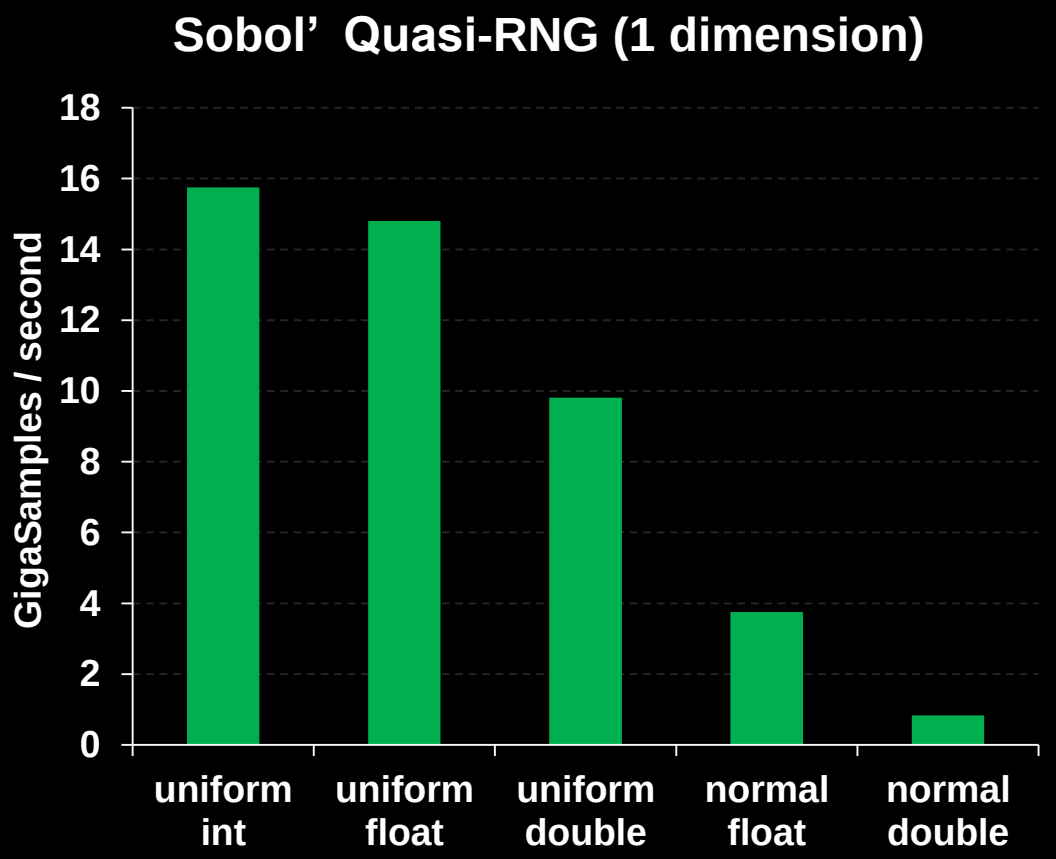
cuRAND



- Library for generating random numbers
- Features supported in CUDA 3.2
  - XORWOW pseudo-random generator
  - Sobol' quasi-random number generators
  - Host API for generating random numbers in bulk
  - Inline implementation allows use inside GPU functions/kernels
  - Single- and double-precision, uniform and normal distributions

cuRAND Performance





Performance may vary based on OS version and motherboard configuration

* CURAND 3.2, NVIDIA C2050 (Fermi), ECC on