HP-SEE
Using CUDA numerical libraries

www.hp-see.eu

Dušan Stanković, Petar Jovanović
Scientific Computing Laboratory
Institute of Physics Belgrade
dusan.stankovic@ipb.ac.rs, petarj@ipb.ac.rs

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

The HP-SEE initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 261499

Agenda

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- CUDA Showroom
- cuBLAS
- cuFFT
- cuRAND

CUDA Showroom

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

GPU Applications: http://www.nvidia.com/object/gpu-applications.html

GPU APPLICATIONS

Most Popular | Search | 1 of 1

AMBER • 89.44 ns/day JAC NVE Speed up* Supported Features • PMEMD: Explicit and Implicit Solvent Molecular Dynamics	GROMACS • 165 ns/day DHFR Speed up* Supported Features • Implicit (5x) • Explicit(2x) Solvent Molecular Dynamics	LAMMPS • 3.5 - 15x Speed up* Supported Features • Lennard-Jones • Gay-Berne Molecular Dynamics	NAMD • 6.44 ns/day STMV585x 2050s Speed up* Supported Features • Non-Bond Force calculation • 100M atom capable Molecular Dynamics
GeoVation CGG/Veritas RTM • Speed up non-disclosed Supported Features • RTM algorithm	Paradigm Echos RTM • 14x Speed up* Supported Features • RTM algorithm	Schlumberger WesternGeco Omega2 RTM • Non-disclosed Supported Features • Multiple algorithms (RTM, etc)	Tsunami RTM • 5x Speed up* Supported Features • RTM algorithm
NWChem • 3-10x Speed up* Supported Features • Triples part of Reg-CCSD(T) • CCSD and EOMCCSD task schedulers • In development Computational Chemistry	ANSYS Mechanical • 2-3x Speed up* Supported Features • Direct & iterative solver Computational Structural Mechanics	Abaqus/Standard • 1.5 - 2.5x Speed up* Supported Features • Direct sparse solver Computational Structural Mechanics	Impetus Afea • 10x SPH, 2x total Speed up* Supported Features • linear equation solver Computational Structural Mechanics
MSC Nastran • 1.5 Speed up* Supported Features • Linear equation solver Computational Structural Mechanics	MathWorks MATLAB • 2-20x Speed up* Supported Features • GPU acceleration for over 200 common MATLAB functions Computational Finance	VASP • 2x Speed-up* 2 GPUs vs. 128 CPU cores Supported Features • Hybrid Hartree-Fock DFT functionals including exact exchange	Chroma • 8-9x Speed up* 768 GPUs vs. 768 (16-core) CPUs Supported Features • Wilson-clover fermions, Krylov solvers, Domain-decomposition

1 of 1

cuBLAS(1)

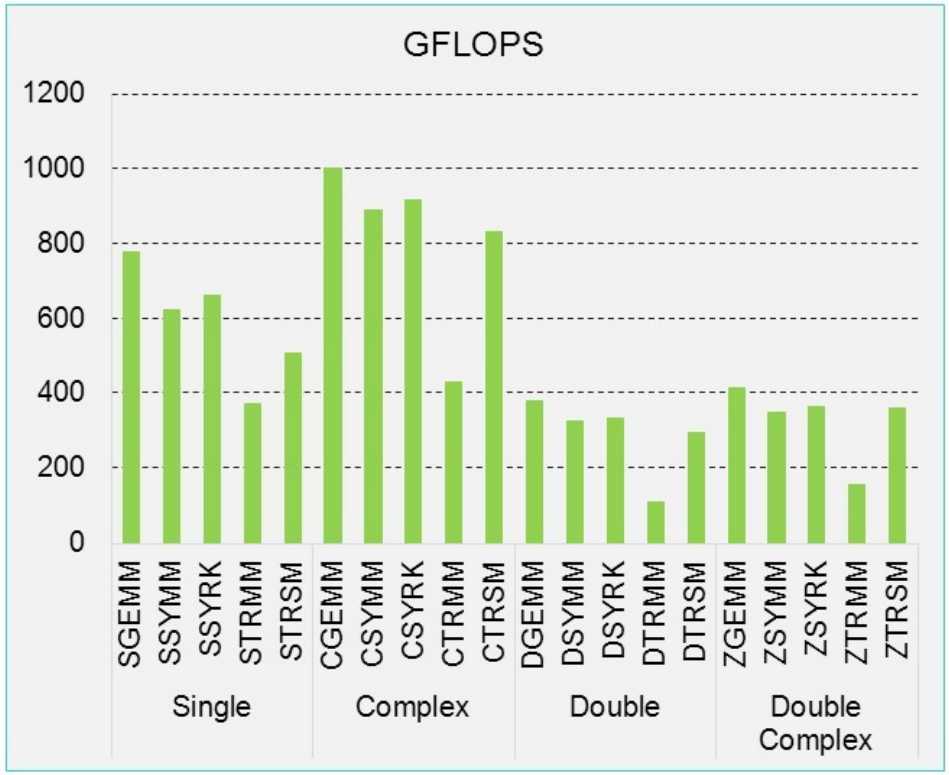
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Key features
  - Complete support for all 152 standard BLAS routines
  - Single, double, complex and double complex data types
  - Support for CUDA streams
  - Fortran bindings
  - Support for multiple GPUs and concurrent kernels
  - Batched GEMM API
  - Device API that can be called from CUDA kernels
  - Batched LU factorization API

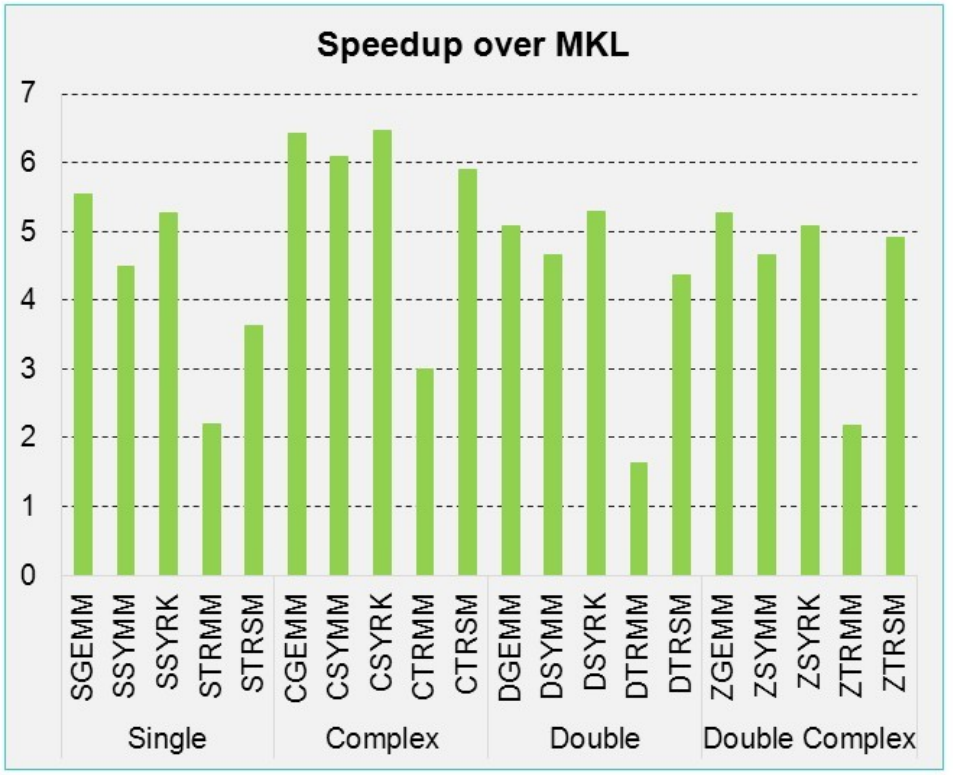
cuBLAS(2)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance



- 4Kx4K matrix size
- cuBLAS 4.1, Tesla M2090 (Fermi), ECC on

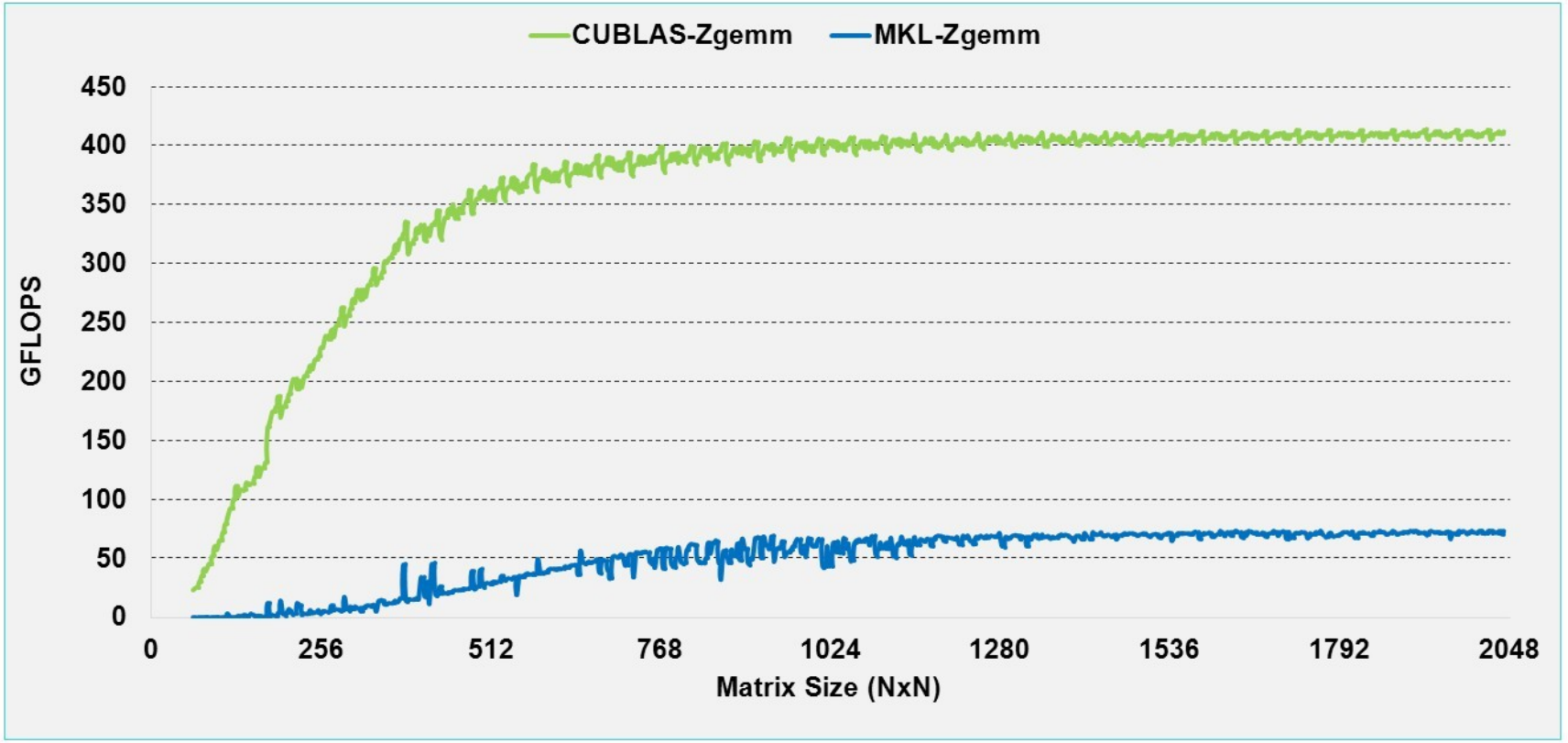


- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz
- Performance may vary based on OS ver. and motherboard config

cuBLAS(3)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance



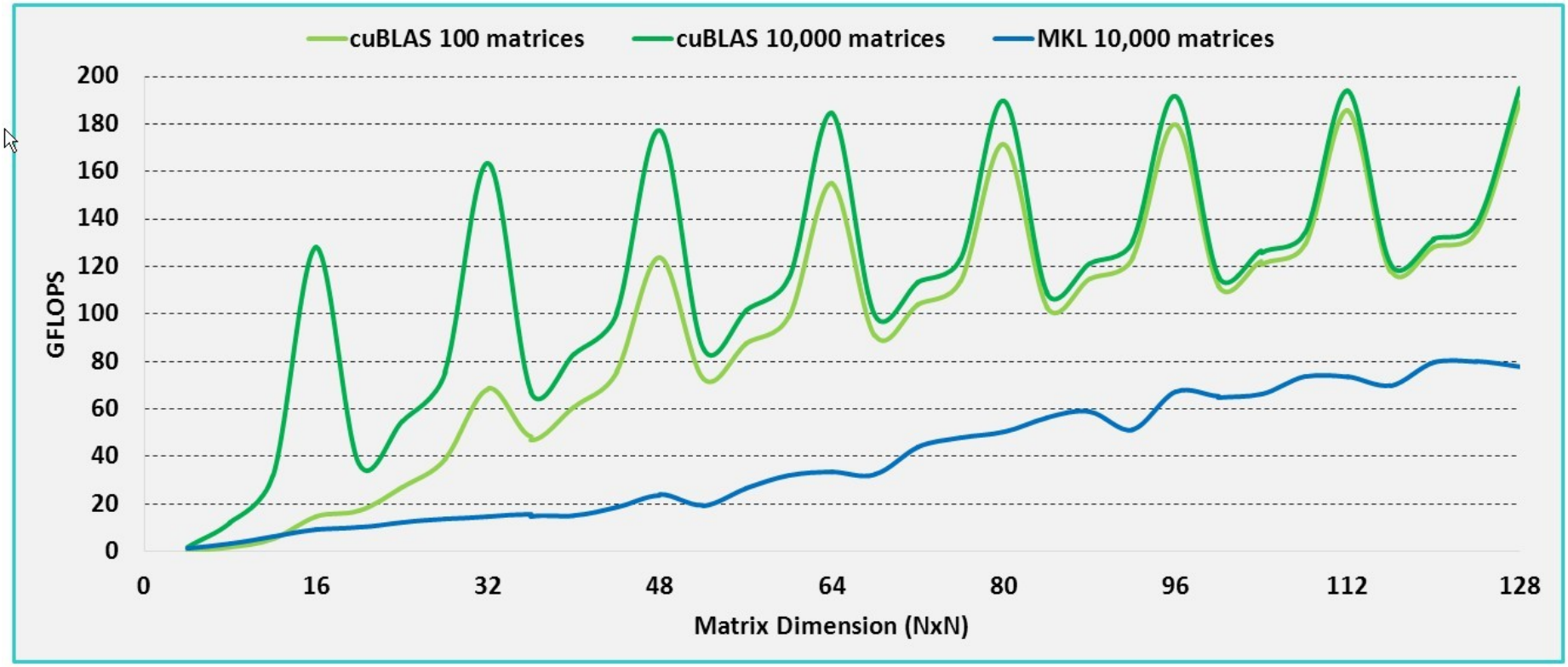
- cuBLAS 4.1 on Tesla M2090, ECC on
- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz

- Performance may vary based on OS ver. and motherboard config.

cuBLAS(4)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance



- cuBLAS 4.1 on Tesla M2090, ECC on
- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz

- Performance may vary based on OS ver. and motherboard config.

cuFFT(1)

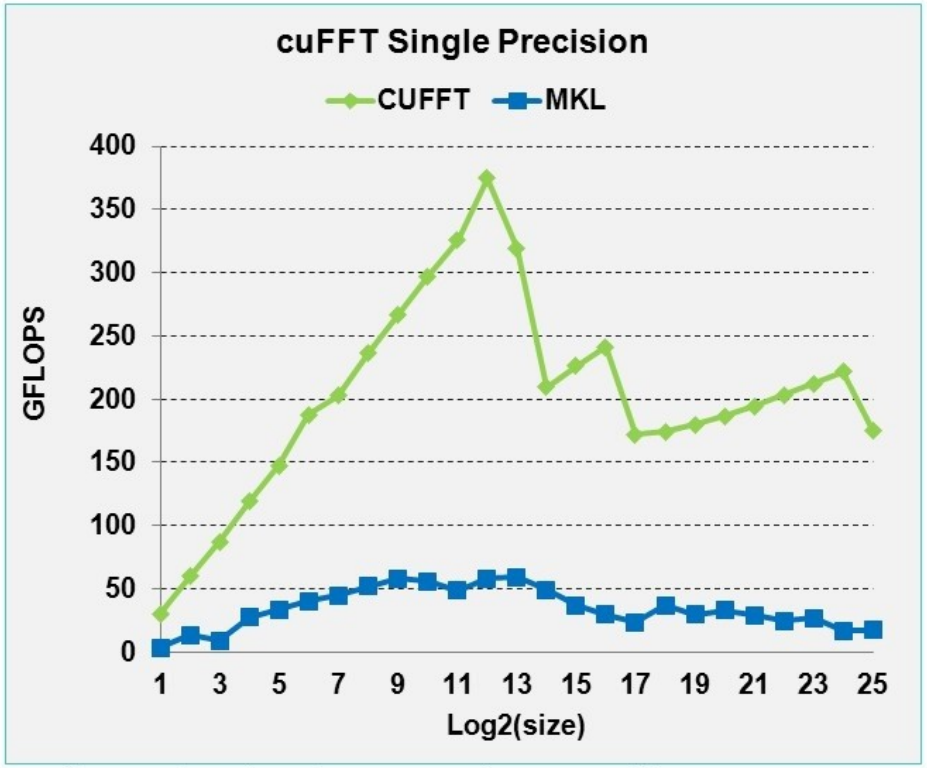
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Key features:
  - 1D, 2D, 3D transforms of complex and real data types
  - 1D transform sizes up to 128 million elements
  - Flexible data layouts by allowing arbitrary strides between individual elements and array dimensions
  - FFT algorithms based on Cooley-Tukey and Bluestein
  - Familiar API similar to FFTW Advanced Interface
  - Streamed asynchronous execution
  - Single and double precision transforms
  - Batch execution for doing multiple transforms
  - In-place and out-of-place transforms
  - Flexible input & output data layouts, similar to FFTW Advanced Interface
  - Thread-safe & callable from multiple host threads

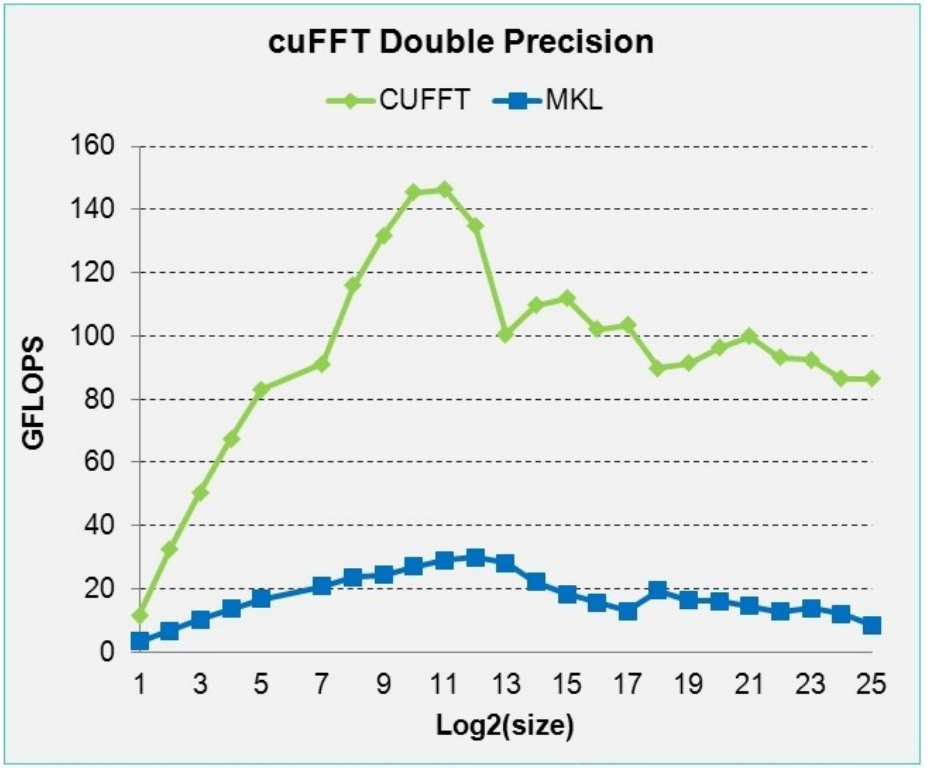
cuFFT(2)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance



- Measured on sizes that are exactly powers-of-2
- cuFFT 4.1 on Tesla M2090, ECC on
- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz

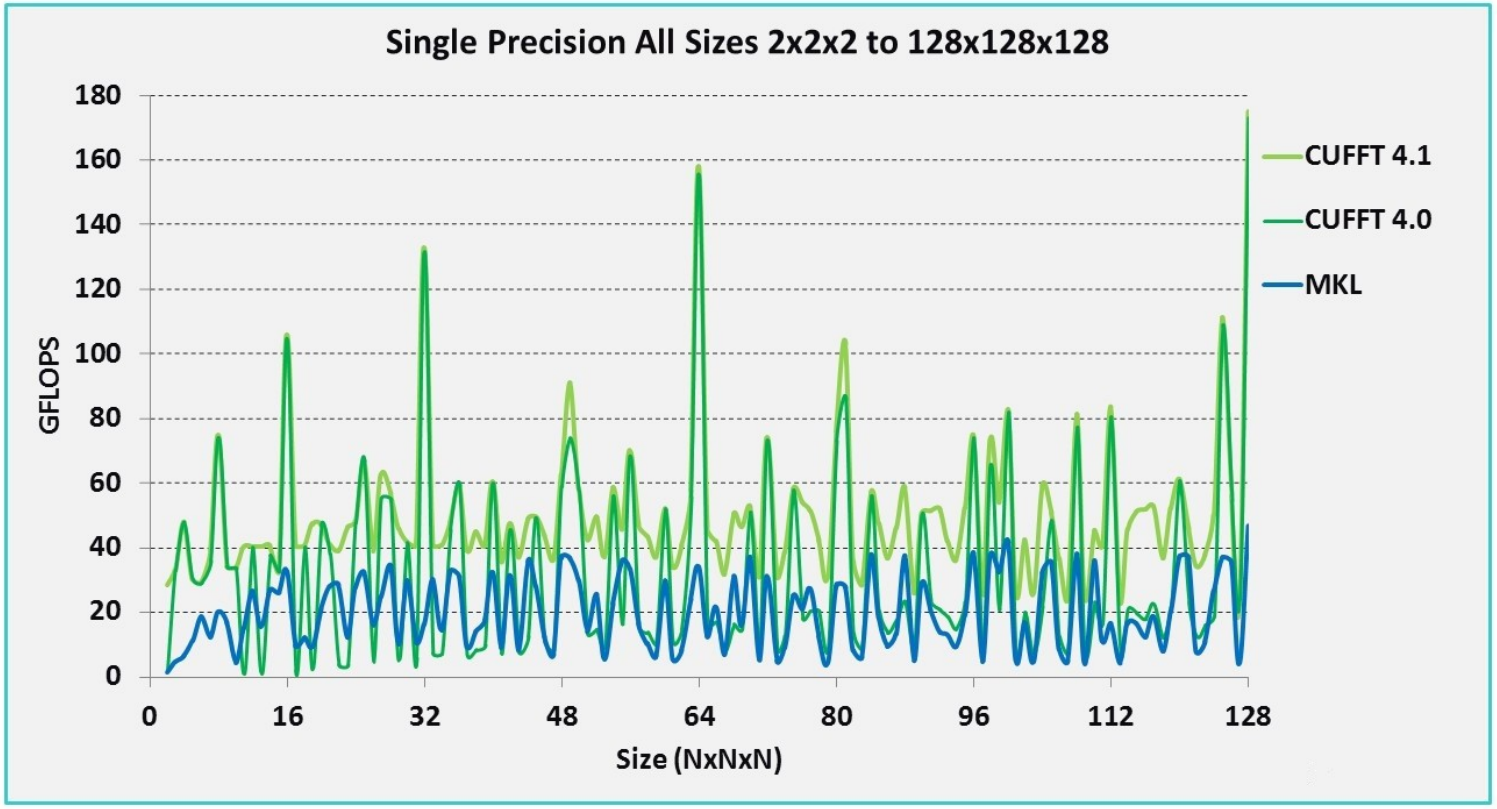


- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz
- Performance may vary based on OS version and motherboard configuration

cuFFT(3)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance



- cuFFT 4.1 on Tesla M2090, ECC on
- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz

•Performance may vary based on OS ver. and motherboard config.

cuRAND(1)

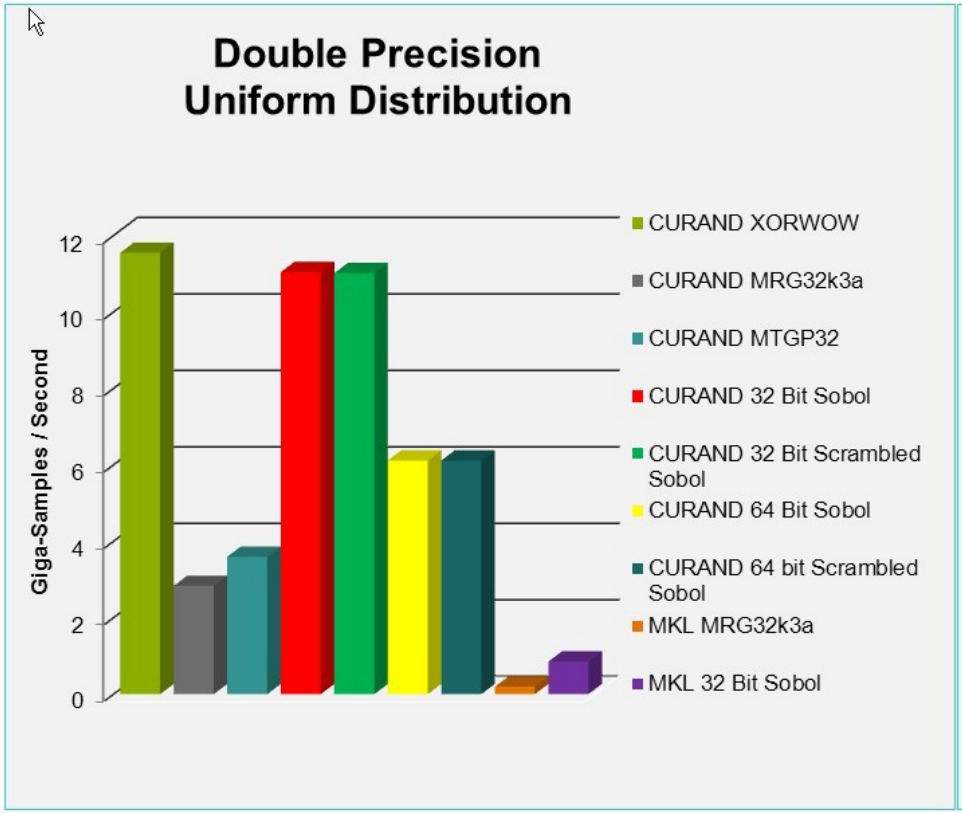
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

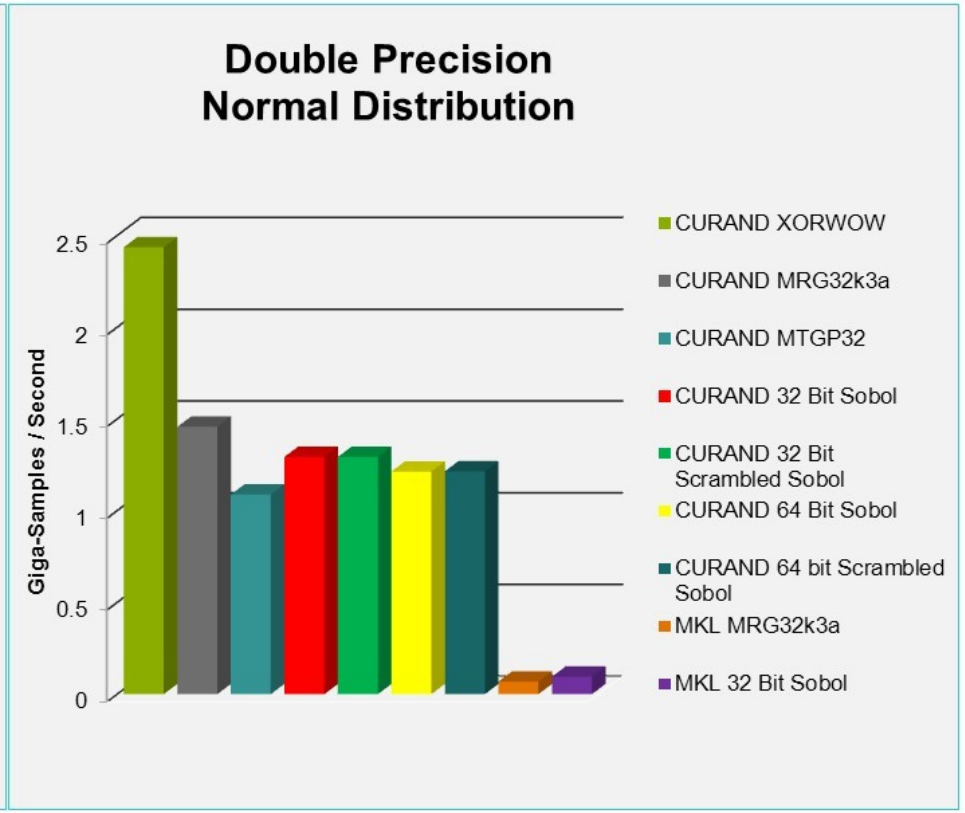
- Key features:
  - Flexible Usage Model
    - Host API for generating random numbers in bulk on GPU
    - Inline implementation allows use inside GPU functions/kernels, or in your host code
  - Four High-Quality RNG Algorithms
    - MRG32k3a
    - MTGP Merseinne Twister
    - XORWOW pseudo-random generation
    - Sobol' quasi-random number generators, including support for scrambled and 64-bit RNG
  - Multiple RNG Distribution Options
    - Uniform, normal, log-normal, Poisson distribution
    - Single or double precision

cuRAND(2)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Performance





- cuRAND 4.1 on Tesla M2090, ECC on
- MKL 10.2.3, TYAN FT72-B7015 Xeon x5680 Six-Core @ 3.33 GHz

- Performance may vary based on OS ver. and motherboard config.