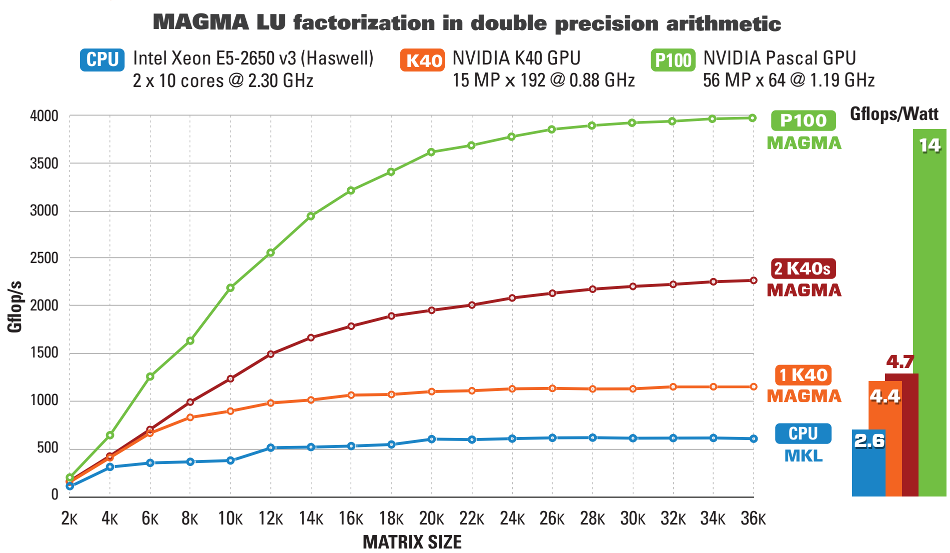
MAGMA

MAGMA (Matrix Algebra on GPU and Multicore Architectures) is a collection of next generation linear algebra libraries for heterogeneous architectures. MAGMA is designed and implemented by the team that developed LAPACK and ScaLAPACK, incorporating the latest developments in hybrid synchronization- and communication-avoiding algorithms, as well as dynamic runtime systems. Interfaces for the current LAPACK and BLAS standards are supported to allow computational scientists to seamlessly port any linear algebra reliant software components to heterogeneous architectures. MAGMA allows applications to fully exploit the power of current heterogeneous systems of multi/many-core CPUs and multi-GPUs to deliver the fastest possible time to accurate solution within given energy constraints.

HYBRID ALGORITHMS

MAGMA uses a hybridization methodology where algorithms of interest are split into tasks of varying granularity and their execution scheduled over the available hardware components. Scheduling can be static or dynamic. In either case, small non-parallelizable tasks, often on the critical path, are scheduled on the CPU, and larger more parallelizable ones, often Level 3 BLAS, are scheduled on the GPU.

PERFORMANCE & ENERGY EFFICIENCY



FEATURES AND SUPPORT

- **MAGMA 2.2** FOR **CUDA**
- **clMAGMA 1.4** FOR **OpenCL**
- **MAGMA MIC 1.4** FOR **Intel Xeon Phi**

CUDA	OpenCL	Intel Xeon Phi	
●	●	●	Linear system solvers
●	●	●	Eigenvalue problem solvers
●	●		Auxiliary BLAS
●			Batched LA
●		●	Sparse LA
●	●	●	CPU Interface
●	●	●	GPU Interface
●	●	●	Multiple precision support
●			Non-GPU-resident factorizations
●	●	●	Multicore and multi-GPU support
●	●	●	LAPACK testing
●	●	●	Linux
●	●		Windows
●	●		Mac OS

INDUSTRY COLLABORATION

NVIDIA GPU CENTER OF EXCELLENCE

NVIDIA's GPU Center of Excellence Program recognizes universities expanding the frontier of massively parallel computing using CUDA.

intel Intel Parallel Computing Center

The objective of the Innovative Computing Laboratory's IPCC is the development and optimization of numerical linear algebra libraries and technologies for applications, while tackling current challenges in heterogeneous Intel® Xeon Phi™ coprocessor-based High Performance Computing.

AMD

Long-term collaboration and support on the development of clMAGMA, the OpenCL™ port of MAGMA.

ICL INNOVATIVE COMPUTING LABORATORY

IN COLLABORATION WITH
Berkeley University of California; University of Colorado Denver; Inria; KAUST

WITH SUPPORT FROM
AMD; intel; MathWorks; NVIDIA

SPONSORED BY
U.S. Department of Defense; ECP; National Science Foundation SILAS Award

INNOVATIVE COMPUTING LABORATORY

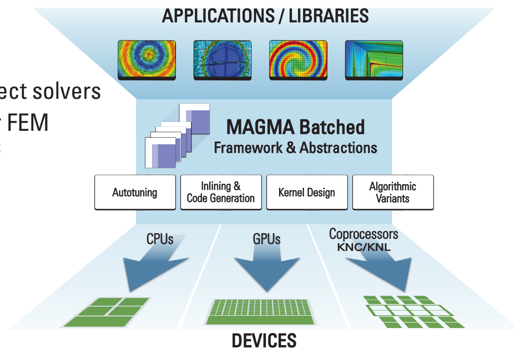
MAGMA

MAGMA BATCHED

BATCHED FACTORIZATION OF A SET OF SMALL MATRICES IN PARALLEL

Numerous applications require factorization of many small matrices

- Deep learning
- Structural mechanics
- Astrophysics
- Sparse direct solvers
- High-order FEM simulations

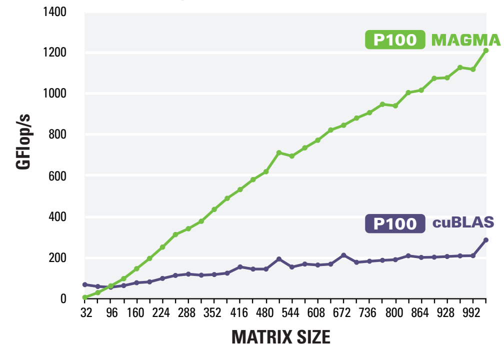


ROUTINES

LU, QR, and Cholesky	✓
Solvers and matrix inversion	✓
All BLAS 3 (fixed + variable)	✓
SYMV, GEMV (fixed + variable)	✓

PERFORMANCE OF BATCHED LU

in double precision arithmetic on 1000 matrices



MAGMA 2.2 DRIVER ROUTINES

	MATRIX	OPERATION	ROUTINE	INTERFACES CPU	INTERFACES GPU
LINEAR EQUATIONS	GE	Solve using LU	{sdcz}gesv	✓	✓
		Solve using MP	{zc,ds}gesv		✓
	SPD/HPD	Solve using Cholesky	{sdcz}posv	✓	✓
		Solve using MP	{zc,ds}posv		✓
LEAST SQUARES	GE	Solve LLS using QR	{sdcz}gels		✓
		Solve using MP	{zc,ds}geqrsv		✓
STANDARD EVP	GE	Compute e-values, optionally e-vectors	{sdcz}geev	✓	
	SY/HE	Computes all e-values,	{sd}syevd	✓	✓
		optionally e-vectors	{cz}heevd	✓	✓
		Range (D&C)	{cz}heevdx		✓
		Range (B&I It.)	{cz}heevx	✓	✓
		Range (MRRR)	{cz}heevr	✓	✓
STAND. SVP	GE	Compute SVD,	{sdcz}gesvd	✓	
		optionally s-vectors	{sdcz}gesdd	✓	
GENERALIZED EVP	SPD/HPD	Compute all e-values,	{sd}sygvd	✓	
		optionally e-vectors	{cz}hegvd	✓	
		Range (D&C)	{cz}hegvdx	✓	
		Range (B&I It.)	{cz}hegvx	✓	
		Range (MRRR)	{cz}hegvr	✓	

ABBREVIATIONS

GE	General
SPD/HPD	Symmetric/Hermitian Positive Definite
TR	Triangular
D&C	Divide & Conquer
B&I It	Bisection & Inverse Iteration
MP	Mixed-precision Iterative Refinement

NAMING CONVENTION

magma_{routine name}[_gpu]

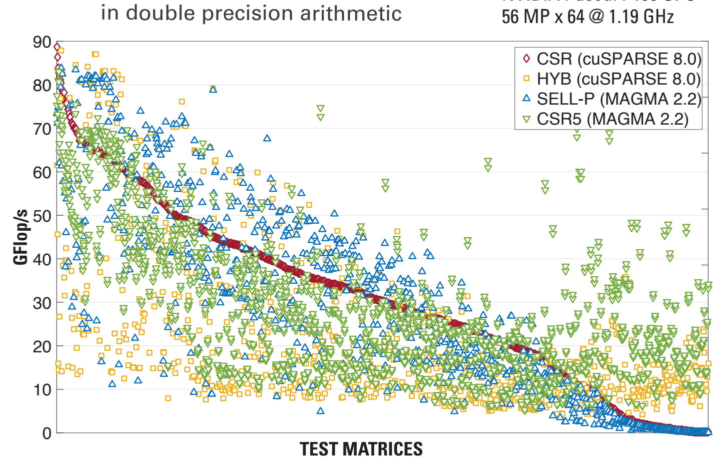
MAGMA SPARSE

ROUTINES	BiCG, BiCGSTAB, Block-Asynchronous Jacobi, CG, CGS, GMRES, IDR, Iterative refinement, LOBPCG, LSQR, QMR, TFQMR
PRECONDITIONERS	ILU / IC, Jacobi, ParILU, ParILUT, Block Jacobi, ISAI
KERNELS	SpMV, SpMM
DATA FORMATS	CSR, ELL, SELL-P, CSR5, HYB

PERFORMANCE

Sparse matrix - vector product (SpMV) in double precision arithmetic

NVIDIA Pascal P100 GPU
56 MP x 64 @ 1.19 GHz



from The University of Florida Sparse Matrix Collection http://www.cise.ufl.edu/research/sparse/matrices/

ICL INNOVATIVE COMPUTING LABORATORY

IN COLLABORATION WITH
Berkeley University of California · University of Colorado Denver · Inria · KAUST

WITH SUPPORT FROM
AMD · intel · MathWorks · NVIDIA

SPONSORED BY
U.S. Department of Defense · U.S. Department of Energy · ECP · National Science Foundation SILAS Award

CITR CENTER FOR INFORMATION TECHNOLOGY RESEARCH

FIND OUT MORE AT http://icl.cs.utk.edu/magma