

# MAGMA

The Matrix Algebra for GPU and Multicore Architectures (MAGMA) project aims to create a new generation of linear algebra libraries that achieve the fastest possible time to an accurate solution on heterogeneous/hybrid systems, using all available processing power. The main focus is the development of a dense linear algebra library for hybrid systems of homogeneous x86-based multicore accelerators with GPUs. MAGMA is designed to be similar to LAPACK in functionality, data storage, and interface, in order to allow scientists to effortlessly port any of their LAPACK-relying software components to take advantage of the new architectures.

## 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.

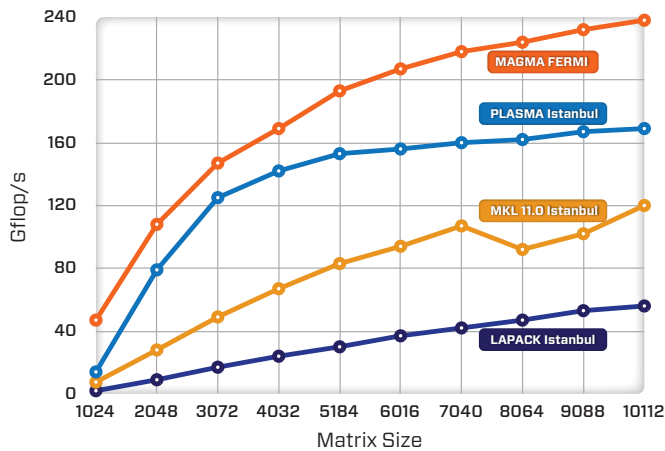
Small non-parallelizable tasks, often on the critical path, are scheduled on the CPU, and larger more parallelizable ones, often Level 3 BLAS, are scheduled on the GPU.

## MAGMA BLAS

MAGMA BLAS targets a subset of BLAS routines for NVIDIA GPUs that are specific to MAGMA and can improve on CUBLAS. It includes routines for Fermi and for the older generation of Tesla GPUs. MAGMA BLAS supports streaming. Its new GEMM for Fermi GPUs is now used in CUBLAS 3.2.

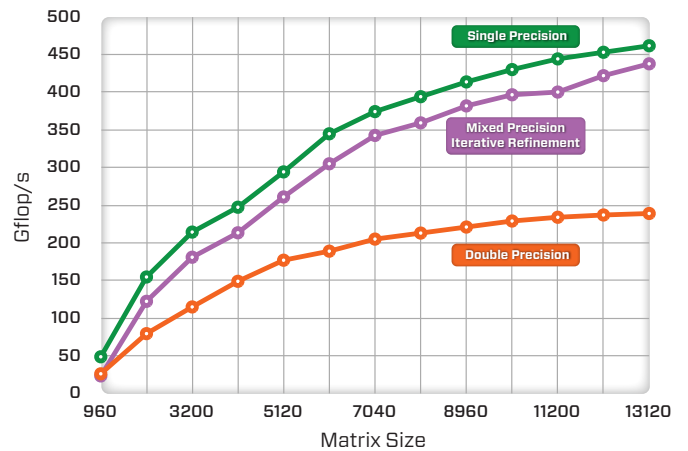
## PERFORMANCE

MAGMA LU in double precision on Fermi (C2050)



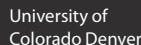
Fermi C2050: 448 Cores @ 1.15 GHz SP/DP peak is 1030/515 Gflop/s  
 Istanbul AMD 8 socket 6 core (48 cores) @ 2.8 GHz SP/DP peak in 1075/538 Gflop/s

MAGMA LU-based solvers on Fermi (C2050)



C2050: 448 Cores @ 1.15 GHz SP/DP peak is 1030/515 Gflop/s

A COLLABORATION OF



WITH SUPPORT FROM



SPONSORED BY

