clMAGMA: High Performance Dense Linear Algebra with OpenCL*

Chongxiao Cao
Innovative Computing Laboratory
University of Tennessee
Knoxville, TN
ccao1@utk.edu

Jack Dongarra†
Innovative Computing Laboratory
University of Tennessee
Knoxville, TN
dongarra@cs.utk.edu

Peng Du‡
Amazon.com
Seattle, WA
pdu@utk.edu

Mark Gates
Innovative Computing Laboratory
University of Tennessee
Knoxville, TN
mgates3@utk.edu

Piotr Luszczek
Innovative Computing Laboratory
University of Tennessee
Knoxville, TN
luszczek@eecs.utk.edu

Stanimire Tomov
Innovative Computing Laboratory
University of Tennessee
Knoxville, TN
tomov@cs.utk.edu

ABSTRACT

This paper presents the design and implementation of several fundamental dense linear algebra (DLA) algorithms in OpenCL. In particular, these are linear system solvers and eigenvalue problem solvers. Further, we give an overview of the clMAGMA library, an open source, high performance OpenCL library that incorporates various optimizations, and in general provides the DLA functionality of the popular LAPACK library on heterogeneous architectures. The LAPACK compliance and use of OpenCL simplify the use of clMAGMA in applications, while providing them with portable performance. High performance is obtained through the use of the high-performance OpenCL BLAS, hardware- and OpenCL-specific tuning, and a hybridization methodology, where we split the algorithm into computational tasks of various granularities. Execution of those tasks is efficiently scheduled over the heterogeneous hardware components by minimizing data movements and mapping algorithmic requirements to the architectural strengths of the various heterogeneous hardware components.

*This research was sponsored by the National Science Foundation through the *Keeneland: National Institute for Experimental Computing* grant (award #0910735), the Department of Energy, and AMD.

†Also affiliated with the Oak Ridge National Laboratory, TN, USA and the University of Manchester, UK

‡Research completed while at the University of Tennessee, Knoxville

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IWOCL '14, May 12 - 13 2014, Bristol, United Kingdom

Copyright 2014 ACM 978-1-4503-3007-7/14/05 ...$15.00.

http://dx.doi.org/10.1145/2664666.2664667.

Categories and Subject Descriptors

G.4 [**Mathematical software**]: Algorithm design and analysis, Efficiency, Parallel implementations, Portability; G.1.3 [**Numerical analysis**]: Numerical linear algebra—*linear systems, matrix inversion, eigenvalues and eigenvectors*

1. INTRODUCTION

Solving linear systems of equations and eigenvalue problems is fundamental to scientific computing. The popular LAPACK library [5], and in particular its vendor optimized implementations like Intel's MKL [13] or AMD's ACML [3], have been the libraries of choice to provide these solvers for dense matrices on shared memory systems. This paper considers a redesign of the LAPACK algorithms to facilitate their OpenCL implementation, and to add efficient support for heterogeneous systems of multicore processors with GPU accelerators and coprocessors. This is not the first time that DLA libraries have needed a redesign to be efficient on new architectures – notable examples being the move from LINPACK [10] to LAPACK [5] in the 80's to make the algorithms cache friendly; ScaLAPACK [8] in the 90's to support distributed memory systems, and now the PLASMA and MAGMA libraries [1], that target efficiency on multicore and heterogeneous architectures, respectively.

The development of new high-performance numerical libraries is complex, and requires accounting for the extreme level of parallelism, heterogeneity, and wide variety of accelerators and coprocessors available in current architectures. Challenges vary from new algorithmic designs to choices of programming models, languages, and frameworks that ease development, future maintenance, and portability. This paper addresses these issues while presenting our approach and algorithmic designs in the development of the clMAGMA [9] library.

To provide portability across a variety of GPU accelerators and coprocessors (such as Intel Xeon Phi), clMAGMA uses OpenCL [14]. OpenCL is an open standard for off-loading computations to accelerators, coprocessors, and manycore processors. It is maintained by the Khronos group with the backing of major hardware and software industry vendors. It offers portability across hardware and OS software. Although the use of OpenCL provides portability of

code, cross-device performance portability is not guaranteed. We specifically address this in Section 2.

To deal with the extreme level of parallelism and heterogeneity inherent in current architectures, clMAGMA uses a hybridization methodology, described in Section 3, where we split the algorithms of interest into computational tasks of various granularities, and properly schedule these tasks' execution over the heterogeneous hardware. To do this, we use a Directed Acyclic Graph (DAG) approach to parallelism and scheduling that has been developed and successfully used for dense linear algebra libraries such as PLASMA and MAGMA [1], as well as in general task-based approaches to parallelism with runtime systems like StarPU [6] and SMPSs [7]. Note, however, that we do not use OpenCL to execute the DAG on the CPU but, rather, we use native threading (`pthread` in the case of Linux) combined with our own scheduler called QUARK [22], that was developed before OpenCL gained widespread use.

Besides the general cross-device considerations addressed in Section 2, obtaining high performance in OpenCL depends on a combination of algorithm and hardware-specific optimizations, discussed in Section 4. The implication of this in terms of software is the fact that in order to maintain its performance portability across hardware variations, there is a need to ensure that the algorithmic variations therein are tunable, e.g., at installation time. This is the basis of autotuning, which is an example of these advanced optimization techniques.

A performance study on AMD hardware is presented in Section 5. Besides verifying our approaches and confirming the appeal of OpenCL and accelerators for high-performance DLA, the results open up a number of future work opportunities discussed in our conclusions.

2. CROSS-DEVICE CONSIDERATIONS

A recommended approach to developing a high-performance and easy to maintain DLA library is to express the algorithms of interest in terms of the BLAS standard. Performance portability is then obtained through the use of architecture-specific, highly tuned BLAS implementations (e.g., MKL from Intel or ACML from AMD). LAPACK and ScaLAPACK have demonstrated this over the years, and now we see it in the new MAGMA and PLASMA libraries. The clMAGMA library takes the same approach, and therefore performance portability relies on the availability of portable OpenCL BLAS, discussed in Section 2.1. Specifics related to OpenCL and its implementation are also important for obtaining high-performance and must be addressed while designing and tuning OpenCL algorithms. Well designed microbenchmarks, shown in Section 2.2, can be used to obtain these key OpenCL specifics to achieving high performance.

2.1 Portable OpenCL BLAS

The Automatically Tuned Linear Algebra Software (ATLAS) library [21] is a BLAS implementation for CPUs. ATLAS achieves portable performance across CPUs mainly by relying on empirical autotuning. Still, vendor libraries like MKL and ACML, optimized for their specific architectures, provide higher performing implementations. The same is true with OpenCL BLAS implementations – OpenCL provides software portability, but unless tuned for a particular architecture, optimization opportunities can be missed.

Currently, the most complete OpenCL BLAS implementation is AMD's clAmdBlas, provided through the AMD's Accelerated Parallel Processing Math Libraries (APPML) [2][1]. It can be used on architectures other than AMD, but its tuning, and therefore highest efficiency, will likely be achieved on AMD hardware. The potential of OpenCL to express BLAS algorithms, as opposed to other lower level access to the hardware, while obtaining high performance is evident through the clAmdBlas. Other implementations, e.g., from Nakasato et al. [16, 15], confirm this by obtaining impressive high performance matrix-matrix multiplication (GEMM). In particular, the highest performance that we are aware of has been demonstrated by Matsumoto et al. [15] – their OpenCL DGEMM reaches up to 848 Gflop/s, and SGEMM up to 2,646 Gflop/s, which is 90% and 70% of the double and single precision peak, respectively. The results come from AMD's Tahiti GPU (Radeon HD 7970).

Highly optimized OpenCL BLAS, along with iterative methods and preconditioners, is also available through the open-source Vienna Computing Library (ViennaCL)[2]. Related work on ViennaCL has addressed OpenCL linear algebra programmability and performance issues [17, 18].

In our previous work, we evaluated OpenCL as a programming tool for performance-portable BLAS [11]. Triangular solvers (TRSM) and GEMMs were developed in OpenCL, tuned for a specific device, and compared. The conclusion was that when using OpenCL the overhead associated with environment setup is large and should be minimized, e.g., by preprocessing or localized in library initialization routines. More importantly, the presented performance results confirmed the conclusion above that OpenCL is expressive enough for developing high performance BLAS, so long as architectural specifics are taken into account in the algorithm design. Even though good performance should not be expected from blindly running algorithms on a new platform, autotuning heuristics can help to improve performance on a single platform.

Autotuning mechanisms are already provided in clAmdBlas through a tuning tool that the user can run to produce optimized OpenCL BLAS on the architecture of interest. Thus, as performance portability of OpenCL BLAS can be obtained, organizing higher-level libraries like clMAGMA in terms of OpenCL BLAS can ensure their performance portability as well.

2.2 Microbenchmarks

We developed a number of microbenchmarks to help us gain a better understanding of OpenCL and to guide our algorithm design and tuning. We describe two benchmarks that can be instrumental for performance – kernel launch overhead and CPU-GPU data transfer. To add some context to the reported measurements, we include comparisons with corresponding CUDA measurements.

2.2.1 Kernel launch overhead

The time to asynchronously invoke OpenCL 1.2 AMD-APP (1016.4) kernel on an AMD Tahiti GPU (Radeon HD 7900 Series) varies in the 5.59–8.88μs range. This was measured by asynchronously launching an empty kernel a large number of times and synchronizing at the end. The overhead can increase significantly (it could be two orders of magnitude larger depending on hardware and software configuration) when synchronizing after each kernel invocation, and therefore synchronization should be avoided. Similar benchmarks for CUDA 4.2 [20] showed an overhead of 3–7μs with no synchronization between kernels, and 10–14μs with synchronization between kernels. It should be stressed that this comparison is between two different programming models, software implementations, and hardware configurations. As such, we only mean to suggest efficient usage scenarios.

We also benchmarked the kernel launch overhead for four BLAS functions: DGEMM, DTRSM, DTRMM and DSYRK, which are

[1] AMD's OpenCL BLAS is now available as open source on GitHub https://github.com/clMathLibraries/clBLAS

[2] http://viennacl.sourceforge.net/

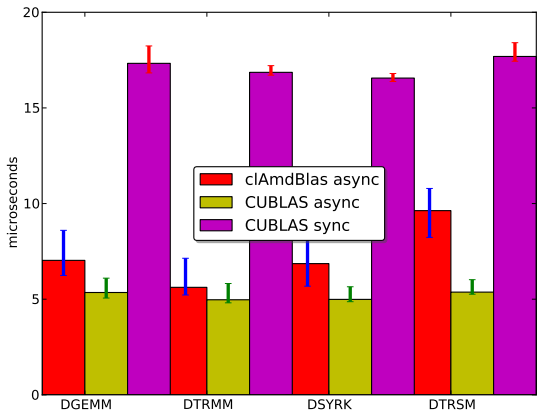

Figure 1: The launch overhead of the GPU BLAS functions for clAmdBlas 1.8.286 with OpenCL 1.2 AMD-APP (1016.4) on Radeon HD 7970 and CUBLAS 4.2 on Tesla S2050, using PCIe 2.0 CPU-GPU interface.

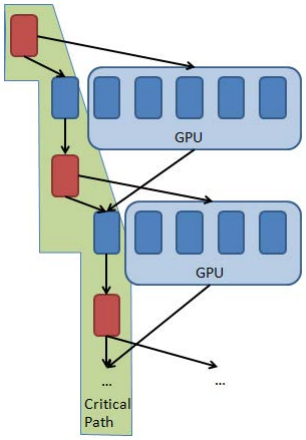

Figure 2: DLA algorithm as a collection of BLAS-based tasks and their dependencies. The algorithm's critical path is, in general, scheduled on the CPUs, and large data-parallel tasks on the GPUs.

used in the double precision LU, Cholesky and QR factorizations. In order to compare with OpenCL, the benchmark for CUDA was tested on an NVIDIA Fermi GPU (Tesla S2050). Results for the kernel launch overhead of OpenCL and CUDA BLAS functions are shown in Figure 1. The OpenCL BLAS functions are from AMD's clAmdBlas 1.8.286 and the CUDA functions are from CUBLAS 4.2. The BLAS functions in clAmdBlas have 6–10 μs asynchronous launch overhead versus 4–5μs in CUBLAS. For synchronous launch overhead, CUBLAS takes only 16–18μs, while clAmdBlas can take significantly longer, and as discussed, synchronization after BLAS kernel invocations should be avoided. Both the CUDA and OpenCL measurements used a PCIe 2.0 interface between CPU and GPU. The hardware and software were different and the comparison is, again, given only to steer the reader toward beneficial programming patterns.

2.2.2 CPU-GPU data transfer overhead

Transfer time for contiguous data between CPU and GPU can be modeled as

$$\text{time} = \text{latency} + \frac{\text{bytes transferred}}{\text{PCIe bandwidth}}. \quad (1)$$

On our system, an AMD Radeon HD 7970 card with a PCIe 2.0 interface, the measured PCIe bandwidth was 2.82 GB/s from CPU to GPU and 3.29 GB/s from GPU to CPU. We found that the latency was 50–60μs from CPU to GPU and 140–150μs from GPU to CPU. Latencies in CUDA are in the 10–17μs range [20], albeit on a different hardware configuration. On our system, that featured an NVIDIA Tesla S2050 with PCIe 2.0 interface, we measured 13–14μs latency in both directions. To avoid the higher latencies associated with synchronizations, algorithms must be designed to use asynchronous data transfers.

3. DENSE LINEAR ALGEBRA IN OPENCL

3.1 Hybridization methodology

The hybridization methodology used in MAGMA [19] is now used in clMAGMA. It is an extension of the task-based approach for parallelism and developing DLA on homogeneous multicore systems [1]. In particular,

- The computation is split into BLAS-based tasks of various granularities, with their data dependencies, as shown in Figure 2.
- Small, latency-bound tasks with significant control-flow are scheduled on the CPUs.
- Large, compute-bound tasks are scheduled on GPUs.

The difference between multicore algorithms and hybridization is the task splitting, which are of various granularities to make different tasks suitable for particular hardware. The scheduling itself is also different.

Challenges with this approach vary from algorithmic designs to tuning for performance portability and balancing work between the CPU cores and the GPUs. Specific algorithms using this methodology, and covering the main classes of DLA, are described in the subsequent sections.

3.2 The clMAGMA design and functionality

The clMAGMA interface is similar to LAPACK. For example, compare LAPACK's LU factorization interface vs. clMAGMA's:

```
lapackf77_dgetrf(&M,&N, hA,   &lda, ipiv, &info)
magma_dgetrf_gpu( M, N, dA,0, ldda, ipiv, &info,
                  queue)
```

Here `hA` is a CPU pointer (`double *`) to the matrix of interest in the CPU memory and `dA` is a pointer in the GPU memory (`magmaDouble_ptr`). The last argument in every clMAGMA call is an OpenCL queue, through which the computation will be streamed on the GPU (`magma_queue_t`).

To relieve the user from knowing OpenCL, all OpenCL data types and main functions, such as BLAS, CPU-GPU data transfers, and memory allocations and deallocations, are redefined in terms of clMAGMA data types and functions. This design allows us to more easily port the MAGMA functionality to clMAGMA, and eventually to merge them together while maintaining a single source. Also, the clMAGMA wrappers are often simpler in syntax than the corresponding OpenCL functions, and provide a comprehensive set of functions for programming hybrid high-performance numerical libraries. Thus, not only the users but also the application developers can choose to use the clMAGMA wrappers without knowing OpenCL. While this might be detrimental from the standpoint of

transparency, we cater for a user base that keeps its focus on the scientific applications and would like to keep using the familiar MAGMA interface, which in turn was largely influenced by LAPACK – the de facto industry standard for interfacing with dense linear algebra software.

clMAGMA provides the standard four floating point arithmetic precisions – single real, double real, single complex, and double complex. There are routines for the so called one-sided factorizations (LU, QR, and Cholesky), two-sided factorizations (Hessenberg, bi-, and tridiagonal reductions), linear system and least squares solvers, matrix inversions, symmetric and non-symmetric standard eigenvalue problems, SVD, and orthogonal transformation routines, all described in the subsections below.

As discussed in [11], compiling an OpenCL kernel from a source file introduces a significant amount of overhead. By caching the Intermediate Representation (IR) resulting from clGetProgramInfo to disk and loading at runtime, overhead can be effectively reduced. AMD and NVIDIA's OpenCL implementations both allow such maneuver, which is essential for the performance of clMAGMA since GPU kernels are repeatedly called in different routines. An efficient way to handle the kernel compiling and caching is required. In clMAGMA, a runtime system is implemented to fulfill this task.

The runtime system, coded in C++ as a singleton class, provides two pieces of functionality depending on the usage phases: during installation, the runtime system compiles OpenCL source files into IRs and stores them to disk; at execution time, the runtime system loads IRs to memory and further builds them into platform specific executables. The runtime system sets up mappings between the name of the OpenCL kernel and its platform specific executables through a series of hash tables. This initialization process executes only once to avoid repeated compiling and allows reusing executables across different higher level routines.

3.3 LU, QR, and Cholesky factorizations

The one-sided factorizations routines implemented and currently available through clMAGMA are:

`magma_zgetrf_gpu` computes an LU factorization of a general M-by-N matrix A using partial pivoting with row interchanges;

`magma_zgeqrf_gpu` computes a QR factorization of a general M-by-N matrix A;

`magma_zpotrf_gpu` computes the Cholesky factorization of a complex Hermitian positive definite matrix A.

Routines in all four standard floating-point precisions are available, following LAPACK's naming convention. Namely, the first letter of the routine name (after the prefix **`magma_`**) indicates the precision, `z`, `c`, `d`, or `s` for double complex, single complex, double real, or single real, respectively. The suffix **`_gpu`** indicates that the input matrix and the output are located in the GPU memory.

The typical hybrid computation and communication pattern for the one-sided factorizations (LU, QR and Cholesky) is shown in Figure 3. At a given iteration, panel dP is copied to the CPU and factored using an LAPACK routine, and the result is copied back to the GPU. The trailing matrix, consisting of the next panel T_1 and submatrix T_2, is updated on the GPU. After receiving dP back from the CPU, T_1 is updated first using dP and the result is sent to the CPU (as being the next panel to be factored there). While the CPU starts the factorization of T_1, the rest of the trailing matrix, T_2, is updated on the GPU in parallel with the CPU factorization of panel T_1. In this pattern, only the data to the right of the current panel is accessed and modified, and the factorizations that use it are known as right-looking. The computation can be organized differently – to access and modify data only to the left of the panel – in which case the factorizations are known as left-looking.

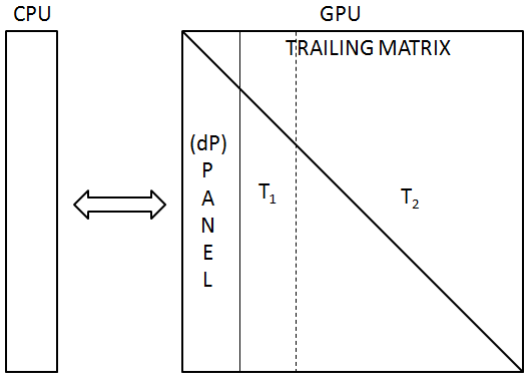

Figure 3: Typical computational pattern for the hybrid one-sided factorizations in clMAGMA.

An example of a left-looking factorization, demonstrating a hybrid algorithm implementation, is given in Figure 4 for the Cholesky factorization. Copying the panel to the CPU, in this case just a square block on the diagonal, is done in line 4. The data transfer is asynchronous, so before we factor it on the CPU (line 8), we synchronize in line 7 to enforce that the data has arrived. Note that the CPU work from line 8 is overlapped with the GPU work in line 6. This is indeed the case because line 6 is an asynchronous call/request from the CPU to the GPU to start the ZGEMM operation. Thus, the control is passed to lines 7 and 8 while the GPU is performing ZGEMM. The resulting factored panel from the CPU work is sent to the GPU in line 11 and used there in line 14, after making sure that it has arrived (the sync in line 13).

```
1  for ( j=0; j<n; j += nb) {
2      jb = min(nb, n – j);
3      magma_zherk( MagmaUpper, MagmaConjTrans,
                    jb, j, m_one, dA(0, j), ldda, one, dA(j, j), ldda, queue );
4      magma_zgetmatrix_async( jb, jb, dA(j,j), ldda, work, 0, jb, queue, &event );
5      if ( j+jb < n )
6          magma_zgemm( MagmaConjTrans, MagmaNoTrans, jb, n-j-jb, j, mz_one,
                        dA(0, j ), ldda, dA(0, j+jb), ldda, z_one,  dA(j, j+jb), ldda, queue );
7      magma_event_sync( event );
8      lapackf77_zpotrf( MagmaUpperStr, &jb, work, &jb, info );
9      if ( *info != 0 )
10         *info += j;
11     magma_zsetmatrix_async( jb, jb, work, 0, jb, dA(j,j), ldda, queue, &event );
12     if ( j+jb < n ) {
13         magma_event_sync( event );
14         magma_ztrsm( MagmaLeft, MagmaUpper, MagmaConjTrans, MagmaNonUnit,
                        jb, n-j-jb, z_one, dA(j, j), ldda, dA(j, j+jb), ldda, queue );
       }
   }
```

Figure 4: Cholesky factorization in clMAGMA.

3.4 Orthogonal transformation routines

The orthogonal transformation routines implemented and currently available through clMAGMA are:

`magma_zungqr[_gpu]` generates an M-by-N matrix Q with orthonormal columns, which is defined as the first N columns of a product of K elementary reflectors of order M as returned by `magma_zgeqrf_gpu`;

`magma_zunmqr[_gpu]` overwrites a general complex M-by-N matrix C with QC or CQ, where Q can also be transposed or not.

The routines are available in all four precisions, and in both CPU (input and output is on the CPU) and GPU interfaces.

Typical uses of the *QR* factorization require computing the product *QC* for some matrix *C* (the `zunmqr` routine). For efficiency, the matrix *Q* is represented implicitly as a product of block Householder reflectors of the form $I - V_i T_i V_i^T$, for $i = 1, \ldots, k$. Instead of forming *Q* explicitly and then performing a matrix-matrix multiplication, it is cheaper to apply the block Householder reflectors directly. Applying each reflector requires three matrix-matrix multiplies, which clMAGMA performs on the GPU. The *V* matrices are tall and skinny, with the upper triangle logically zero, as shown in Figure 5. In LAPACK, the upper triangle of each *V* contains the *R* matrix; in clMAGMA, when *V* is copied to the GPU, the upper triangle is explicitly set to zero. This allows us to simplify the code and improve performance using a single GEMM, instead of a less-efficient triangular multiply (TRMM) and a GEMM. The only work on the CPU is computing the T_i matrices when necessary.

If the *Q* matrix is needed explicitly, clMAGMA can compute it (the `zungqr` routine) by multiplying the implicitly-represented *Q* with identity matrix *I*. This is done in a block-by-block fashion in order to keep it in-place, while overwriting the implicit *Q* (the *V* Householder vectors) with the explicit *Q*.

Similar routines are used by clMAGMA in the eigenvalue and SVD problems, where orthogonal transformations are applied to back-transform the eigenvectors or singular vectors.

3.5 Hessenberg, bi- and tridiagonal reductions

The two-sided factorizations routines currently implemented in clMAGMA are:

`magma_zgehrd` reduces a general matrix *A* to upper Hessenberg form *H* by orthogonal similarity transformations;

`magma_zhetrd` reduces a Hermitian matrix *A* to real symmetric tridiagonal form *T* by orthogonal similarity transformations;

`magma_zgebrd` reduces a general M-by-N matrix *A* to upper or lower bidiagonal form *B* by orthogonal transformations.

The routines are available in all four precisions.

The Hessenberg, bidiagonal, and tridiagonal reductions are two-sided factorizations used in the non-symmetric eigenvalue, symmetric eigenvalue, and SVD problems, respectively. The standard one-stage approach to solving the non-symmetric eigenvalue problem applies an orthogonal transformation *Q* to both sides of the matrix *A* to reduce it to the upper Hessenberg form, $H = QAQ^T$. QR iteration is then used to find the eigenvalues and eigenvectors of *H*; the eigenvalues of *H* are the same as the eigenvalues of *A*, while the eigenvectors can be back-transformed using *Q* to find the eigenvectors of *A*.

Unlike the *QR* factorization, where the panel factorization is independent of the trailing matrix, in the Hessenberg reduction, each column of the panel requires a matrix-vector product (GEMV) with the trailing matrix. We take advantage of the high bandwidth of GPUs to accelerate these memory-bound GEMV operations during the panel factorization. The outline of algorithm is shown in Figure 5. A panel dP_i is copied from the GPU to the CPU (step 1). For each column *j* of the panel, a Householder vector v_j is computed (step 2) and the matrix-vector product $y_j = A_j v_j$ is computed with the trailing matrix on the GPU (step 3). After the panel factorization, the block Householder reflector is applied with several GEMMs to update the trailing matrix, and completed portions of the trailing matrix are copied back to the CPU (step 4). Note that in this pattern the communication-to-computation is in a surface-to-volume ratio –

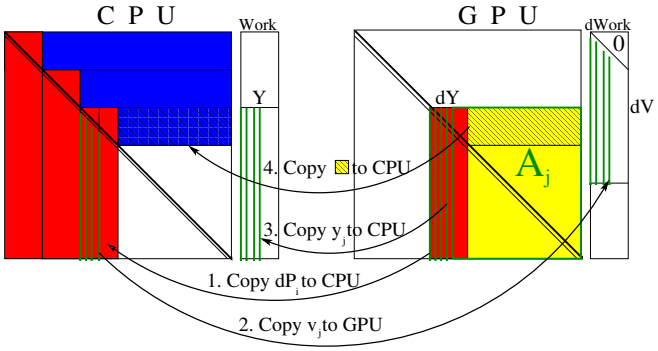

Figure 5: Typical communication pattern for the hybrid two-sided factorizations in clMAGMA.

sending a vector of length *n* is followed by $2n^2$ flops (in the inner loop), and sending a panel of size $n \times nb$ is followed by $O(n^2 \times nb)$ flops (in the outer loop).

Similarly, the symmetric eigenvalue problem involves an initial reduction to tridiagonal form, and the SVD involves an initial reduction to bidiagonal form. The exact details differ from the Hessenberg factorization, but, in a similar fashion, the panel factorization involves matrix-vector products (GEMV or SYMV), which clMAGMA performs on the GPU to take advantage of the high memory bandwidth of the device.

Recent success in MAGMA with two-stage algorithms for the tridiagonal reduction [12] demonstrate that we can recast it using compute-bound Level-3 BLAS SYMM operations, instead of memory-bound Level-2 BLAS SYMV operations. This provides a large speed boost compared to the traditional one-stage algorithm. Future work on clMAGMA involves porting these two-stage algorithms, where we expect a similar speed increase.

3.6 Linear system and eigenproblem solvers

The one- and two-sided factorizations are the major building blocks for developing linear system and eigenproblem solvers, respectively. We have developed the following solvers:

`magma_zposv_gpu` solves a system of linear equations $Ax = B$ with a Hermitian positive definite matrix *A* using the Cholesky factorization of *A*;

`magma_zgesv_gpu` solves a system of linear equations with general N-by-N matrix *A* using the LU factorization of *A*;

`magma_zgels_gpu` solves the overdetermined least squares problem, $\min ||Ax - B||$, using the QR factorization of *A*;

`magma_zheevd` computes all eigenvalues and, optionally, eigenvectors of a complex Hermitian matrix *A*. If eigenvectors are desired, it uses a divide and conquer algorithm;

`magma_zgeev` computes the eigenvalues and, optionally, the left and/or right eigenvectors for an N-by-N complex non-symmetric matrix *A*;

`magma_zgesvd` computes the singular value decomposition (SVD) of a complex M-by-N matrix *A*, optionally computing the left and/or right singular vectors.

The routines are available in all four precisions. The linear solvers use the hybrid clMAGMA one-sided factorization routines and triangular matrix solvers, as provided by their OpenCL BLAS implementations. The eigenproblem solvers use the hybrid clMAGMA

two-sided factorizations, which are the most time consuming parts of the algorithms. The rest is run on the CPUs, using vendor optimized LAPACK.

Related to the linear solvers, clMAGMA provides matrix inversion routines as well:

magma_ztrtri_gpu for computing the inverse of a real upper or lower triangular matrix;

magma_zgetri_gpu for computing the inverse of a matrix using the LU factorization computed by `magma_zgetrf_gpu`;

magma_zpotri_gpu for computing the inverse of a real symmetric positive definite matrix using its Cholesky factorization computed by `magma_zpotrf_gpu`.

The triangular inverse routine is a hybrid, derived from the corresponding block LAPACK algorithm. The diagonal blocks of the matrix are sent and inverted on the CPU, and everything else is done on the GPU. The LU inversion uses `magma_ztrtri_gpu` to invert U and then computes $\text{inv}(A)$ by solving the system $\text{inv}(A)L = \text{inv}(U)$ for $\text{inv}(A)$ (entirely on the GPU). The `magma_zpotri_gpu` also uses `magma_ztrtri_gpu` to invert the upper (U) or lower (L) factor of the Cholesky factorization, and a hybrid code (routine `magma_zlauum_gpu`) to compute the product UU' or $L'L$.

4. ADVANCED OPTIMIZATIONS

We highlight three optimization techniques that are crucial for obtaining high performance. The first one, overlapping CPU-GPU communication with GPU computation, is important because of the slow CPU-GPU interconnect relative to the GPU performance capabilities. For example, sending a few bytes between the CPU and GPU without overlap can result in losing the opportunity to compute $O(10^7)$ double precision flops on the GPU. The second one, overlapping CPU and GPU work, allows us to use the entire system more efficiently. Finally, autotuning is a technique that removes the need for manual tuning and enables cross-device performance portability.

The optimizations described in Sections 4.1 and 4.2 target specifically AMD hardware and AMD's OpenCL implementation, while Section 4.3 addresses the cross-device portability.

4.1 Overlapping CPU-GPU communications with GPU computation

In Section 2, we saw that OpenCL can have higher CPU-GPU data transfer latency overhead than CUDA, which can reduce the effective bandwidth when a small size of data is transferred between the CPU and GPU. Thus, this can become a performance bottleneck, unless it is overlapped with useful GPU work (or minimized with other optimization techniques). Figure 6 shows part of the trace of a double precision LU factorization in clMAGMA: the first row is the CPU work, where the black color represents the time of panel factorization; the second row is the GPU work, where the red color represents DGEMM operations and green color represents DTRSM. Yellow color reflects the time to copy the data from GPU to CPU and grey is copying the data from CPU to GPU. Although computation on the CPU has overlapped with the GPU, communication and computation on the GPU are executed sequentially.

In OpenCL, performing work on a device, such as executing kernels or moving data to and from the device's local memory, is done using a corresponding command-queue [4]. A command-queue is an interface for a specific device and its associated work. A way to overlap CPU-GPU communication and GPU computation

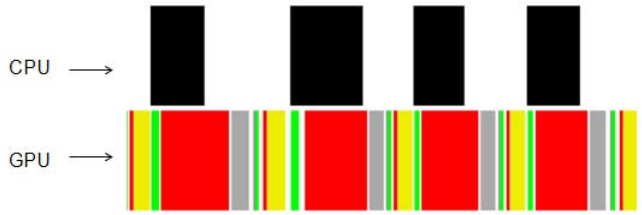

Figure 6: Partial CPU-GPU execution trace of a hybrid LU factorization in clMAGMA. Yellow and gray represent CPU-GPU communication that in this case are not overlapped with the GPU work.

is by creating two command-queues. One queue is used for data transfers and the other is used for kernel computations. Figure 7 shows a part of the trace of double precision LU factorization similar to Figure 6, but here we have applied the optimization of using two queues. The first row is again the CPU work, the second row is the computation work of queue 1 on the GPU, and the third row is the communication work of queue 2. All color definitions are the same as in Figure 7. Note that based on this two-queue technique, we made the communication overlap with the GPU computation work. Experiments showed that this approach lead to about 10% increase of performance for double precision LU factorization.

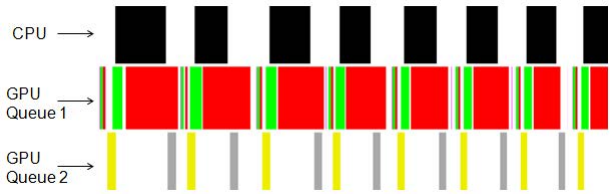

Figure 7: Partial CPU-GPU execution trace of a hybrid LU factorization in clMAGMA based on the two command-queues' optimization, overlapping CPU-GPU data transfers (the yellow and gray transfers in GPU Queue 2) with GPU work (in GPU Queue 1).

From the above two traces, we also notice that there are some blank gaps between different kernels on the GPU. Those represent overheads of kernel switching on the GPU.

4.2 Overlapping CPU and GPU work

In OpenCL, the host creates a data structure called a command-queue to coordinate execution of the kernels on the devices. The host places commands into the command-queue which are then scheduled onto the devices. For example, in Figure 4, line 6 puts a ZGEMM in the command-queue `queue`. The host still must submit the ZGEMM to the device for execution, but this may not happen immediately. As a result, the CPU can start the computation at line 8 while the device has not started the ZGEMM. Thus, although our high-level algorithm is designed to overlap CPU and GPU work, overlap may not happen in practice. In order to force the command-queue to immediately submit the command queued to the appropriate device, one must call `clFlush(queue)` [4]. Therefore, all clMAGMA BLAS wrappers first queue the corresponding OpenCL BLAS and immediately post a `clFlush` to the queue.

The importance of overlapping CPU and GPU work is quantified in Figure 8 for the case of LU factorization in double precision (the DGETRF routine). The blue curve is the performance of DGETRF

without CPU and GPU work overlap. It achieves up to 195 Gflop/s. The red curve is the performance of DGETRF with overlapping CPU and GPU work, using `clFlush`. It achieves up to 280 Gflop/s, i.e., getting about 1.4× speedup.

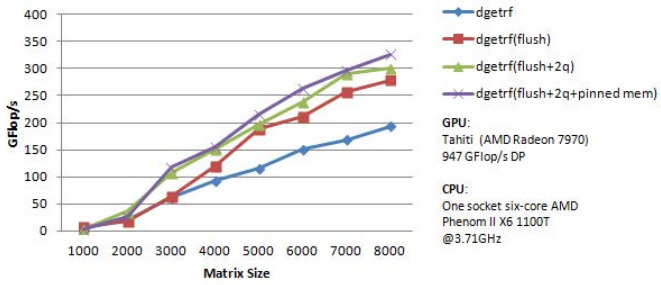

Figure 8: Advanced performance optimizations of DGETRF in clMAGMA.

Figure 8 also shows the effect of further optimizations, and, in particular, the technique of using two queues to overlap CPU-GPU communications with GPU computation (from the previous subsection), and using pinned memory to get higher transfer throughput between the CPU and GPU. Putting all these optimizations together, the performance of dgetrf is shown with the purple curve. It achieves up to 326 Gflop/s, which is almost a 60% speedup compared to the original version without any optimizations.

It is worth noting that OpenCL implementations may differ in their treatment and the effects of `clFlush`. Also important is the fact that the specific behavior of multiple command-queues and their interaction with a single device will likely be different between vendors and their implementations.

Another method for overlap is the use of out-of-order command-queues but they were not supported on the hardware/software combinations that we had available for our tests.

4.3 Autotuning

While functionality of an OpenCL code is portable, the resulting performance often is not. However, it is commonly sufficient to rely on highly optimized BLAS that are provided by the vendor to guarantee transportable efficiency in terms of the peak performance. This is clearly predicated on the fact that the BLAS is of high quality and is capable of providing very efficient execution across a wide range of input parameters including matrix dimensions and data-dependent characteristics such as symmetry or transposition. In practice, this requirement might not be fulfilled, in which case it is necessary to use customized versions of some of the kernels or maybe just one specific instance of the kernel for particular matrix shapes. In our current tests, we did not use autotuning to improve performance.

5. PERFORMANCE STUDY

The performance results that are provided in this section were obtained around February and March of 2013. They use AMD's Radeon HD 7970 discrete card. The multicore host CPU was a single socket, six-core AMD Phenom II X6 1100T running at 3.7 GHz. The kernels executed on the CPU used LAPACK and BLAS implemented by MKL 11.1. The BLAS kernels executed on the GPU are from clAmdBlas 1.8. The OpenCL version was 1.2. We installed AMD-APP 1016.4 as the OpenCL driver. Currently the AMD OpenCL driver for Linux has a 512 MB maximum limit for a single memory allocation on the GPU. Consequently, we tested only matrix sizes up to 8,000 in double precision arithmetic in our experiments.

The performance of double precision LU factorization in clMAGMA is given in Figure 9. It achieves up to 326 Gflop/s for matrices of size 8,000. This corresponds to over 5.7× speedup over the six-core CPU host for this particular size. The break-even point when the performance of the CPU and GPU is the same occurs for matrix sizes of just over 2000. For matrices below this size, we cannot obtain benefits from using the AMD GPU. Also, the performance of the GPU device does not seem to achieve a plateau as the CPU performance curve does. Instead, we see a steady (and nearly linear) increase in performance that might have resulted in even higher speedup over the CPU were it not for the memory limitation mentioned earlier.

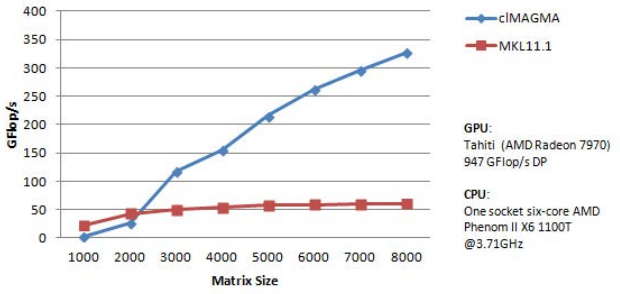

Figure 9: Performance of clMAGMA's LU factorization in double precision compared against MKL 11.1

The performance of the double precision Cholesky factorization from clMAGMA is shown in Figure 10. It achieves up to 344 Gflop/s for matrices of size 8,000, which is about 5.4× speedup over the six-core CPU host. Just as it was the case for the LU factorization, the break-even point of CPU-GPU performance occurs for sizes slightly over 2,000. But unlike for LU, the increase in performance is much more drastic, especially for sizes above 4,000. Still, the plateau of the performance curve for the GPU is not achieved and we have a reason to believe that a higher level of performance is possible with better memory allocation in the GPU driver.

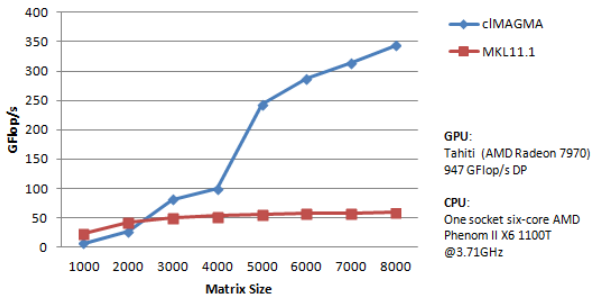

Figure 10: Performance of clMAGMA's Cholesky factorization in double precision compared against MKL 11.1

The performance of the double precision QR factorization from clMAGMA is shown in Figure 11. It achieves up to 347 Gflop/s for matrices of size 8,000, which is about 5.9× speedup over the six-core CPU host. The performance curve looks much like the one for LU with almost a linear ascend and lack of plateau for large matrix sizes. Consequently, the previous comments for the LU performance curve also apply to the QR curve.

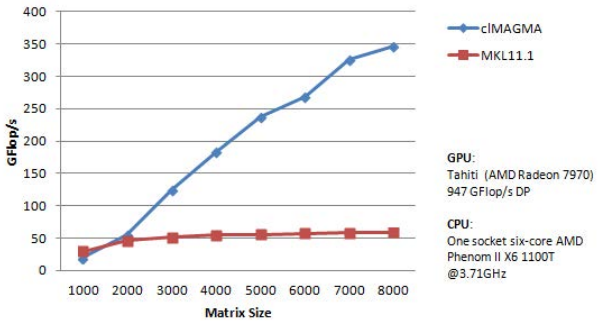

Figure 11: Performance of clMAGMA's QR factorization in double precision compared against MKL 11.1

The performance of the double precision Hessenberg reduction from clMAGMA is shown in Figure 12. It achieves up to 40 Gflop/s for matrices of size 8,000, which is about 5.5× speedup over the six-core CPU host. Hessenberg reduction is the first and the most time consuming step of the process of obtaining eigenvalues of a general non-symmetric matrix. Unlike the Cholesky, LU, and QR factorization, the reduction to the Hessenberg form is memory bound and thus the performance levels for both the GPU and the CPU remain well below 50 Gflop/s. Even with this drastic difference the general trends remain the same: the break-even point is around 2,000 and the linear increase of performance is very clear.

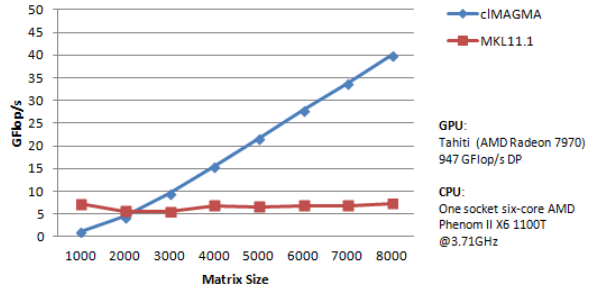

Figure 12: Performance of clMAGMA's Hessenberg factorization in double precision compared against MKL 11.1

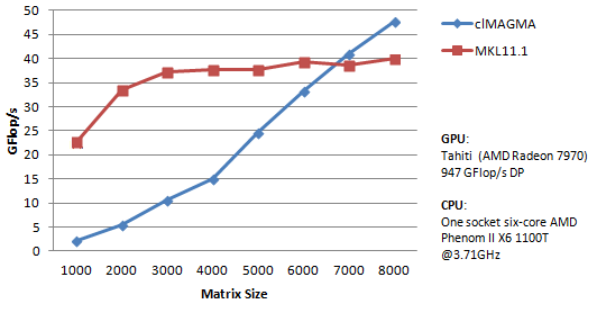

Figure 13: Performance of clMAGMA's Matrix Inversion in double precision compared against MKL 11.1

The performance of the double precision Cholesky matrix inversion in clMAGMA (`magma_zgetri_gpu`) is shown in Figure 13. It achieves up to 48 Gflop/s for matrices of size 8,000, which is about 1.2× speedup over the CPU host. This last performance figure shows the lowest break-even point somewhere between 6,000 and 7,000. All the gains in performance obtained by off-loading the Cholesky factorization to the GPU are negated by the very poorly performing kernels that apply the Cholesky factor to form the matrix inverse. These particular kernels will require the most tuning and optimization in the future as the similar kernels are present in the LU-based inversion and are likely to cause a comparable performance drop.

6. CONCLUSIONS AND FUTURE WORK

We have presented high performance linear algebra routines for a wide range of linear algebra factorizations and reductions. The routines were implemented with efficiency in mind and consequently tuned for a specific family of the AMD's Tahiti GPUs with the use of the implementation of the OpenCL standard and the optimized BLAS routines from the hardware vendor. Our optimization techniques show a wide applicability across the tested routines and yield many-fold performance improvement over highly tuned CPU codes that constitute the state-of-the-art libraries for the current generation of multicore CPUs. With the success we achieved in porting our high performance kernels to the OpenCL implementation on GPUs, we are now encouraged to look into extending our porting efforts to emerging platforms such as the Intel Xeon Phi and ARM's AArch64, as well as the supported editions of the multicore x86 hardware that are targeted by the CPU-oriented implementations of OpenCL.

Porting clMAGMA to mobile or embedded computing platforms is also of interest to us. One direction worth pursuing relies on tuning the blocking sizes, and even further in the future we intend to work on developing autotuning mechanisms to produce software that will use empirical measurement at installation time to automatically tune the library. Yet another direction of our research is to develop implementations that run entirely on the GPU, and thus avoid possible "host" memory restrictions such as the size limitations and the non-uniform time to access data.

Acknowledgments

The authors would like to thank the National Science Foundation (funding a part of this research through awards #0910735 and #1339822), the Department of Energy, and AMD for supporting this research effort.

7. REFERENCES

[1] E. Agullo, J. Demmel, J. Dongarra, B. Hadri, J. Kurzak, J. Langou, H. Ltaief, P. Luszczek, and S. Tomov. Numerical linear algebra on emerging architectures: The PLASMA and MAGMA projects. *J. Phys.: Conf. Ser.*, 180(1), 2009.

[2] AMD. Accelerated Parallel Processing Math Libraries (APPML). Available at http://developer.amd.com/tools/heterogeneous-computing/.

[3] AMD. AMD Core Math Library (ACML). Available at http://developer.amd.com/tools/.

[4] AMD. AMD Accelerated Parallel Processing OpenCL Programming Guide (v2.4), 2012. Available at: http://developer.amd.com.

[5] E. Anderson, Z. Bai, C. Bischof, S. L. Blackford, J. W. Demmel, J. J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. C. Sorensen. *LAPACK User's Guide*. SIAM, Philadelphia, Third edition, 1999.

[6] C. Augonnet, S. Thibault, R. Namyst, and P. Wacrenier. StarPU: A unified platform for task scheduling on

heterogeneous multicore architectures. *Concurrency Computat. Pract. Exper.*, 2010. (to appear).

[7] Barcelona Supercomputing Center. *SMP Superscalar (SMPSs) User's Manual, Version 2.0*, 2008. http://www.bsc.es/media/1002.pdf.

[8] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. SIAM, Philadelphia, PA, 1997. http://www.netlib.org/scalapack/slug/.

[9] Software distribution of clMAGMA version 1.0. http://icl.cs.utk.edu/magma/software/, October 24 2012.

[10] J. Dongarra, J. Bunch, C. Moler, and G. Stewart. *LINPACK Users' Guide*. SIAM, Philadelphia, PA, 1979.

[11] P. Du, R. Weber, P. Luszczek, S. Tomov, G. Peterson, and J. Dongarra. From CUDA to OpenCL: Towards a performance-portable solution for multi-platform GPU programming. *Parallel Comput.*, 38(8):391–407, Aug. 2012.

[12] A. Haidar, S. Tomov, J. Dongarra, R. Solca, and T. Schulthess. A novel hybrid CPU-GPU generalized eigensolver for electronic structure calculations based on fine grained memory aware tasks. *International Journal of High Performance Computing Applications*, September 2012. (accepted).

[13] Intel. Math Kernel Library. Available at http://software.intel.com/en-us/articles/intel-mkl/.

[14] Khronos OpenCL Working Group. The opencl specification, version: 1.0 document revision: 48, 2009.

[15] K. Matsumoto, N. Nakasato, S. Sedukhin, I. Tsuruga, and A. City. Implementing a code generator for fast matrix multiplication in OpenCL on the GPU. Technical Report 2012-002, The University of Aizu, July 2012.

[16] N. Nakasato. A fast GEMM implementation on the Cypress GPU. *SIGMETRICS Performance Evaluation Review*, 38(4):50–55, 2011.

[17] K. Rupp, F. Rudolf, and J. Weinbub. ViennaCL - A High Level Linear Algebra Library for GPUs and Multi-Core CPUs. In *Intl. Workshop on GPUs and Scientific Applications*, pages 51–56, 2010.

[18] P. Tillet, K. Rupp, S. Selberherr, and C.-T. Lin. Towards performance-portable, scalable, and convenient linear algebra. In *Presented as part of the 5th USENIX Workshop on Hot Topics in Parallelism*, Berkeley, CA, 2013. USENIX.

[19] S. Tomov and J. Dongarra. Dense linear algebra for hybrid GPU-based systems. In J. Kurzak, D. A. Bader, and J. Dongarra, editors, *Scientific Computing with Multicore and Accelerators*. Chapman and Hall/CRC, 2010.

[20] V. Volkov and J. Demmel. Benchmarking GPUs to tune dense linear algebra. In *Supercomputing 08*. IEEE, 2008.

[21] R. C. Whaley, A. Petitet, and J. J. Dongarra. Automated empirical optimizations of software and the ATLAS project. *Parallel Computing*, 27(1-2):3–35, 2001.

[22] A. YarKhan, J. Kurzak, and J. Dongarra. QUARK Users' Guide: QUeueing And Runtime for Kernels. *University of Tennessee Innovative Computing Laboratory Technical Report ICL-UT-11-02*, 2011.