

УДК 004.051:004.75

DOI: 10.24160/1993-6982-2018-6-87-95

Базовые библиотеки линейной алгебры для высокопроизводительных расчетов

И.С. Кружилов, М.Б. Кузьминский, А.М. Чернецов, О.Ю. Шамаева

Рассмотрены библиотеки линейной алгебры (BLAS, LAPACK, ScaLAPACK, MKL, ATLAS и др), поддерживающие высокопроизводительные вычисления на современных архитектурах и использующиеся как в известных тестах оценки производительности, так и в различных приложениях.

В большинстве приложений наиболее трудоемкие стадии вычислений реализуются путем вызова подпрограмм из подобных библиотек, поэтому актуальным является оптимальный выбор библиотеки для организации расчетов. Основная цель представленного обзора — описание «инвариантных» характеристик библиотек для достижения высокой производительности приложений.

Приведен краткий обзор HPC-вычислений в различных областях знаний. Предложена классификация библиотек линейной алгебры с точки зрения их функциональности и используемых высокопроизводительных архитектур. Рассмотрена базовая низкоуровневая библиотека BLAS, реализованная для всех HPC-архитектур. Указано, что библиотека BLAS поддерживает распараллеливание на системах с общим полем памяти, для чего применяются средства, такие как OpenMP или OpenACC, а для систем с распределенной памятью используется параллельный вариант PBLAS, поддерживающий обмен сообщениями между узлами с помощью стандарта MPI.

Описаны библиотеки более высокого уровня, основанные на BLAS, например, библиотека LAPACK, содержащая большой набор разных программ для линейной алгебры. Представлены библиотека Scalapack для модели распределенной памяти, основанная на LAPACK и PBLAS, и ее современное развитие — библиотека Intel MKL.

Для поддержки эффективного функционирования гибридных систем анализируются принципиально новые библиотеки — MAGMA и PLASMA, особенностью которых является оптимизация линейно-алгебраических расчетов небольшой размерности.

Рассмотрены библиотеки, поддерживающие решение задач на собственные значения, такие как EISPACK, PeIGS и ряд других. Указано, что в новой библиотеке ELPA, ориентированной на суперкомпьютеры, могут использоваться как средства OpenMP, так и MPI. Отмечено, что операции над разреженными матрицами, особенно умножение матриц, весьма актуальны для многих прикладных областей науки, при этом базовым стандартом для них можно считать библиотеку Sparse BLAS.

Сделан вывод о том, что оптимальный выбор библиотеки существенно зависит как от конкретного приложения, так и от используемой вычислительной архитектуры.

Ключевые слова: библиотеки линейной алгебры, HPC-вычисления, BLAS, ScaLapack, MAGMA, MKL, ELPA, PLASMA.

Для цитирования: Кружилов И.С., Кузьминский М.Б., Чернецов А.М., Шамаева О.Ю. Базовые библиотеки линейной алгебры для высокопроизводительных расчетов // Вестник МЭИ. 2018. № 6. С. 87—95. DOI: 10.24160/1993-6982-2018-6-87-95.

Basic Linear Algebra Libraries for High Performance Computing

I.S. Kruzhilov, M.B. Kuzminsky, A.M. Chernetsov, O.Yu. Shamayeva

The article considers linear algebra libraries such as BLAS, LAPACK, ScaLAPACK, MKL, and ATLAS, which support high-performance computing (HPC) in modern architectures and are used both in well-known performance tests and in various applications.

In the majority of applications, the most time-consuming computation stages are implemented by calling subroutines from such libraries; therefore, the optimal choice of a library is an important issue in setting up computations. The main aim of this review is to describe the "invariant" characteristics of libraries to achieve high performance of applications.

High performance computations used in different fields of knowledge are briefly reviewed. Classification of linear algebra libraries in terms of their functionality and applied high-performance architectures is suggested. The basic low-level BLAS library implemented for all HPC architectures is demonstrated. It is pointed out that the BLAS library supports dividing of the entire computation process into several parallel flows in systems with a common memory field; for such systems, tools such as OpenMP or OpenACC are used. In the case of systems with distributed memory, the parallel version of this library, called PBLAS is used, which supports exchange of messages between nodes using the MPI standard.

Higher-level libraries based on the BLAS, e.g., the LAPACK library, which contains a large set of different programs for linear algebra, are described. The ScaLAPACK library for the distributed memory model, which is based on the LAPACK and PBLAS libraries, as well as the Intel MKL library, which is its modern development, are presented.

To support efficient operation of hybrid systems, the fundamentally new libraries MAGMA and PLASMA involving features for optimizing linear-algebraic computing of small dimension are analyzed.

Libraries supporting solution of eigenvalue problems, such as the EISPACK, PeigS, and a number of other libraries, are investigated. It is pointed out that in the new ELPA library oriented to supercomputers, both OpenMP and MPI tools can be used.

It is noted that operations on sparse matrices, especially multiplication of matrices, are very relevant for many applied fields of science; in this regard, the SparseBLAS library can be considered to be the basic standard for them.

It is concluded that the optimal choice of a library depends essentially on both the particular application and on the used computing architecture.

Key words: linear algebra libraries, high-performance computing, BLAS, ScaLapack, MAGMA, MKL, ELPA, PLASMA.

For citation: Kruzhilov I.S., Kuzminsky M.B., Chernetsov A.M., Shamayeva O.Yu. Basic Linear Algebra Libraries for High Performance Computing. MPEI Vestnik. 2018;6:87—95. (in Russian). DOI: 10.24160/1993-6982-2018-6-87-95.

Сложные вычислительные задачи и HPC-технологии

Высокопроизводительные вычисления (HPC — High Performance Computing), необходимые для решения сложных задач большой размерности, стали одной из основных областей применения суперкомпьютеров. Задачи, связанные с изучением человека и окружающего его мира (моделированием работы мозга, созданием новых лекарственных форм, экологическими прогнозами, геной инженерии и др.) определяют инновационное развитие общества, стимулируют создание экзафлопных компьютеров и развитие HPC-технологий.

Одной из важнейших областей применения HPC-вычислений являются задачи вычислительной химии. Например, в США сложные задачи вычислительной химии совместно с молекулярной биологией составляют порядка 54% всех высокопроизводительных приложений [1]. Работы в области химии и материалов охватывают 56% всех расчетов, проводимых в Швейцарском национальном суперкомпьютерном центре [2]. Схожая картина наблюдается и в Российской Федерации. В 2014 г. использование суперкомпьютерного комплекса МГУ для HPC-расчетов составило: для химических приложений — 31,8%, для физических — 30,1% [3]. В обзоре классических HPC-задач линейной алгебры указано, что развитие аппаратных средств получает мотивацию от расчетов электронной структуры больших молекулярных систем [4].

Сложные квантово-химические расчеты выполняются практически на всех суперкомпьютерах из первой десятки списка TOP500 самых мощных вычислительных систем мира [5]. Мир современных компьютеров — это многопроцессорность и многоядерность, требующие перехода к применению параллельных высокоэффективных технологий, которые должны поддерживаться базовыми стандартными библиотеками, в частности, библиотеками линейной алгебры для реализации HPC-вычислений.

Для обеспечения высокой производительности программ для решения сложных задач используются библиотеки линейной алгебры, такие как BLAS (Basic Linear Algebra Subprograms) — низкоуровневая библиотека линейной алгебры [6], LAPACK (Linear Algebra Package) [7] и ScaLAPACK (Scalable LAPACK) [8] — наборы программ для линейной алгебры высокого функционального уровня, основанные на BLAS, MKL (Math Kernel Library) [9] — библиотеке оптимизированных реализаций BLAS для архитектуры x86-64 Intel, ATLAS (Automatically Tuned Linear Algebra Software) [10] — библиотека в исходных текстах и др.

HPC обычно ориентированы на вычисления с двойной точностью, т. е. на обработку 64-разрядных форматов представления чисел. HPC позволяют эффективно вести обработку комплексных чисел, необходимых при решении многих задач электротехники, гидродинамики, картографии, квантовой механики, теории колебаний, теории хаоса и др.

Очевидно, что HPC-область вычислений в первую очередь ориентирована на решение задач большой размерности N . В последние годы, помимо размерности, важнейшими учитываемыми параметрами стали струк-

тура представления данных и степень ее разреженности, так как во многих областях возникают задачи, требующие обработки разреженных матриц большой размерности, в том числе и блочно-разреженных структур.

Библиотеки линейной алгебры для HPC

Наиболее вычислительно трудоемкими являются операции, связанные с умножением матриц общего вида и реализуемые семейством функций *gemm* из пакета BLAS. Их трудоемкость составляет $O(N^3)$. Оптимизация операций матричного умножения определяется производительностью функций *dgemm* (для вычислений с двойной точностью) и *zgemm* (для обработки комплексных данных).

Имеется множество библиотек математических программ для HPC, но в настоящей статье речь идет в основном о библиотеках линейной алгебры, широко применяемых как в оценочных тестах производительности, так и в различных приложениях. Для большого количества приложений вызов некоторых подпрограмм из таких библиотек определяет время расчета важнейших стадий вычисления.

Библиотеки можно объединить в различные группы по следующим принципиальным показателям:

- относятся ли они к традиционной «плотной» линейной алгебре (Dense Linear Algebra или DLA) или к работе с разреженными (Sparse) векторами и матрицами;
- идет ли речь о библиотеках с открытым исходным текстом, прилагаемым к различным аппаратным реализациям и часто ставшими международными стандартами, или об оптимизированных (обычно объектные файлы в машинных кодах) для определенных классов процессоров или ускорителей, в том числе графических процессоров GPU;
- каковы они по функциональности модулей: базовые BLAS или более высокого уровня;
- для каких вычислительных архитектур — с общей или распределенной памятью, и как реализуется распараллеливание — с помощью технологий OpenMP [11], MPI [12], гибридной или какой-либо еще.

Рассмотрим библиотеки для работы с DLA. Известно, что операции над плотными матрицами широко применяются в квантово-химических методах расчета электронной структуры молекулярных систем и обеспечивают линейное масштабирование числа требуемых операций с плавающей запятой от размера молекулы (числа атомов) [13].

Базовая библиотека BLAS

BLAS — основная (низкоуровневая) библиотека линейной алгебры, существующая несколько десятков лет [6]. В нее входят базовые подпрограммы, например, умножение векторов и матриц. BLAS первого уровня (BLAS L1) включает операции сложения — умножения над векторами, второго уровня (BLAS L2) — векторно-матричные операции, третьего уровня (BLAS L3) — матрично-матричные. BLAS является стандартом, и его

подпрограммы вызываются из других библиотек DLA, имеющих более высокий функциональный уровень. К средствам BLAS обращаются библиотеки линейной алгебры более высокого уровня, поэтому BLAS в первую очередь оптимизируют для конкретных архитектур процессоров, например, для архитектуры Intel.

Библиотека предполагает распараллеливание на системах с общим полем памяти (многопроцессорных серверах и ускорителях), для чего используются такие средства, как OpenMP или OpenACC (как в GPU Nvidia) [14]. Также возможна работа с низкоуровневой библиотекой OpenCL [15]. Для систем с распределенной памятью, где основой организации вычислительного процесса является обмен сообщениями между вычислительными узлами, существует параллельный вариант PBLAS [16], в котором применяется MPI.

Современные высокопроизводительные процессоры и ускорители используют многоуровневые системы памяти от кэшей разных уровней до обычной памяти, иногда также не одного уровня. Поэтому в BLAS L3 активно используется тонкая оптимизация с различными многоуровневыми буферными блоками памяти. Вызовы подпрограмм из BLAS L3, например, вызовы *dgemm*, часто определяют время работы HPC-приложения в целом.

Библиотека BLAS постоянно развивается. Многие библиотеки, оптимизированные для различных архитектур процессоров и ускорителей и функционально более широкие, включают реализацию BLAS. Отметим только библиотеку Global Arrays (GA) [17], включающую модуль *ga_dgemm* для умножения матриц, который использует эффективный алгоритм умножения матриц SRUMMA (Shared and Remote memory access based Universal Matrix Multiplication Algorithm) [19]. Рассмотрим базовые DLA-библиотеки с открытым исходным кодом.

Современное развитие библиотек для HPC

Библиотека LAPACK (Linear Algebra Package), написанная на языке Fortran 90 с вариантом на Fortran 95, содержит большой набор разных программ для линейной алгебры более высокого функционального уровня, основанный на BLAS [7]. Первоначально она была создана для последовательной реализации, однако возможность вызова подпрограмм BLAS позволила выполнять распараллеливание для систем с общим полем памяти [16, 20].

Из большого количества оптимизированных реализаций BLAS и LAPACK приведем наиболее известные и широко распространенные, доступные для различных архитектур процессоров, в том числе для архитектуры x86-64: Intel MKL [9], AMD ACML (AMD Core Math Library) [21], в настоящее время не поддерживается), ATLAS (Automatically Tuned Linear Algebra Software — библиотека в исходных текстах [10]), GotoBLAS [22] и др. Перейдя на создание универсальных библиотек с

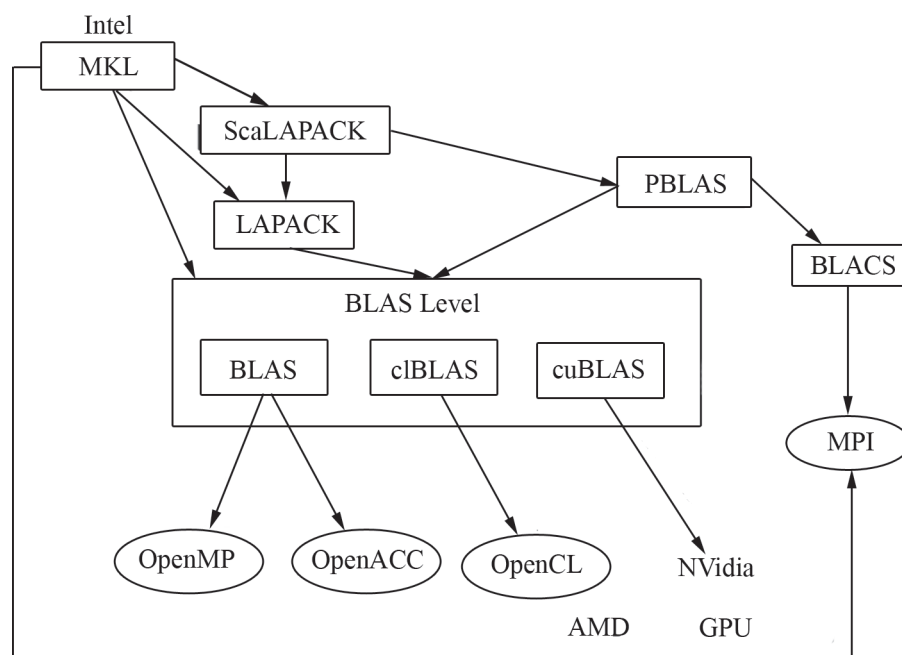
открытым кодом на OpenCL как для своих процессоров, так и для GPU, AMD прекратила развитие ACML. Библиотека GotoBLAS, ограниченная поддержкой только BLAS, также перестала развиваться. Работы продолжились в рамках GotoBLAS2, ориентированной на некоторые аппаратные платформы, в частности, из класса x86-64, но в настоящее время и этот программный продукт активно не развивается [23].

Появилась библиотека OpenBLAS — библиотека открытых текстов, основанная на GotoBLAS2 [24], но ее двоичная версия пригодна только для Windows [25]. Для процессоров Intel Xeon, наиболее распространенных в суперкомпьютерах, библиотека MKL является, вероятно, наиболее эффективной и опережает библиотеку ATLAS на умножении матриц — *dgemm* [26]. Библиотека MKL также содержит и функции библиотеки ScaLAPACK с возможностью распараллеливания как в OpenMP, так и в MPI [27]. Фирма IBM для своих процессоров Power/PowerPC предлагает библиотеку ESSL (Engineering and Scientific Subroutine Library [28]), для GPU Nvidia имеется библиотека cuBLAS [29] и т. д.

Для высокопроизводительных архитектур с распределенной памятью, таких как кластеры и массово-параллельные системы, где основная модель вычислений — модель передачи сообщений, появилась библиотека ScaLAPACK (Scalable LAPACK) — библиотека программ линейной алгебры, основанная на LAPACK и PBLAS [8]. PBLAS — реализация BLAS уровней 2 и 3 поддерживает работу в модели распределенной памяти. В ней по-прежнему используются последовательные функции BLAS первого уровня. Также PBLAS работает со средствами BLACS (Basic Linear Algebra Communication Subprograms) для межузловых коммуникаций, совместимыми с различными низкоуровневыми средствами распараллеливания, в том числе MPI. Схема зависимостей описанных библиотек с указанием поддерживаемых средств параллельной обработки представлена на рисунке.

Отметим также, что и ScaLAPACK может оказаться замененным в будущем на новую библиотеку, поскольку появилось предложение по переходу к применению библиотеки SLATE для следующих поколений суперкомпьютеров [30].

Кроме продвижения от LAPACK в сторону ScaLAPACK, современные DLA-библиотеки развивались и в другом направлении. Появились новые решения в архитектуре вычислительных узлов HPC-систем, такие как применение многоядерных процессоров с большим числом ядер (например, как в Intel KNL или китайских SW2410) или ускорителей, в частности, графического процессора GPU. Для подобных узлов характерна многоуровневая память и использование архитектур типа SMP или NUMA [31]. Для поддержки эффективного функционирования таких компьютерных систем потребовались принципиально новые библиотеки — MAGMA (Matrix Algebra on GPU and Multicore Architectures) [32] и PLASMA (Parallel Linear



Основные библиотеки класса DLA

Algebra Software for Multicore Architectures) [33]. Важной особенностью указанных библиотек является оптимизация линейно-алгебраических расчетов небольшой размерности [16].

Библиотеки для ускорителей

Библиотека PLASMA ориентирована на SMP или NUMA-системы с общей памятью, в ней реализована большая часть функциональности LAPACK и используется распараллеливание с помощью средств OpenMP. Также в пакет входит оптимизированный подкласс BLAS L3 для эффективной работы с кэш-памятью и процессорами с большим числом ядер [20]. Библиотеку можно эффективно применять при работе с многоядерными процессорами Intel KNL, интегрировавшими в себе некоторые особенности ускорителей. В ней возможно распараллеливание на более высоком уровне, чем в BLAS, поскольку доступно распараллеливание на основе новых моделей планировщиков работы из OpenMP версий 3 и 4. PLASMA поддерживает поднабор LAPACK, включая решения задач поиска собственных значений, систем линейных уравнений и др. задач DLA [16]. Достижимая производительность PLASMA сильно зависит от настраиваемых параметров выполнения, подбираемых в зависимости от применяемых аппаратных средств [34].

Из-за ограниченной функциональности библиотека не является полной заменой библиотек LAPACK и ScaLAPACK, поскольку ориентирована только на вычислительные системы с общим полем памяти.

Библиотека MAGMA ориентирована на работу в гетерогенных системах, где совместно могут функционировать обыкновенные и графические процессоры. MAGMA использует сложные алгоритмы, в которых

вычисления разбиваются на задачи с различной степенью детализации, и их выполнение планируется статически или динамически на разных аппаратных компонентах. Библиотека применима не только для GPU Nvidia, но и для других ускорителей, в том числе, Xeon Phi. Существует версия, использующая распараллеливание через OpenCL — cMAGMA [35].

Указанная выше библиотека MKL [36] может использоваться современными процессорами Intel Xeon Phi. Для GPU имеются собственные варианты таких библиотек [37]. Так, AMD вместо ACML рекомендует общие для ее процессоров и GPU библиотеки с исходным текстом на OpenCL, в частности cBLAS [38]. У NVIDIA поддержку BLAS включают библиотеки CUDA SDK — cuBLAS и cuBLAS-XT (последняя обеспечивает распределение нагрузки сразу на несколько GPU для BLAS третьего уровня) [39].

Рассмотрим современные библиотеки линейной алгебры, поддерживающие решение задач на собственные значения. Классический пример такой библиотеки — EISPACK [40], включающая набор подпрограмм на языке Fortran, скопированных позднее в библиотеку LAPACK. Поскольку обязательным требованием для HPC-библиотек является поддержка средств распараллеливания, то с этой точки зрения классической библиотекой считается ScaLAPACK.

Кроме того, следует упомянуть библиотеку PeIGS [41], основанную на эффективном методе распараллеливания GA [42] и применяемую в квантово-химическом комплексе NWChem [43].

Имеется много подобных библиотек, представленных в [44], оптимальный выбор которых для конкретного приложения зависит от используемых аппаратных средств.

Отметим отдельно использование на стадии работы уже с трехдиагональной матрицей более современного

алгоритма MRRR, требующего только $O(N^2)$ операций и применяемого в библиотеках LAPACK/ScaLAPACK [45].

Кроме эффективности указанных алгоритмов расчета и средств распараллеливания, следует обращать внимание и на оптимизацию библиотеки для конкретных аппаратных средств, например, применение библиотеки MKL [46, 47] для аппаратных средств Intel или PESSL — для аппаратуры от IBM [45].

В последнее время для решения задач на собственные значения появилась новая библиотека ELPA (Eigenvalue solvers for Petaflop-Applications) [48,49], которая с самого начала ориентируется на высокопроизводительные системы суперкомпьютерного уровня. Для распараллеливания в ELPA можно применять средства OpenMP и MPI.

Библиотеки для матриц разреженной структуры

Операции над разреженными матрицами, особенно умножение разреженных матриц, актуальны для многих прикладных областей науки [50, 51].

Базовым стандартом для решения таких задач считается библиотека Sparse BLAS [52]. Однако современные общематематические библиотеки линейной алгебры также работают с разреженными матрицами, например, все операции третьего уровня BLAS в MKL [36].

В настоящее время появилось немало специальных библиотек линейной алгебры, эффективно реализующих обработку разреженных матриц, например, DBCSR (Distributed Block Compressed Sparse Row) — библиотека для умножения разреженных матриц, используемая в квантово-химическом программном комплексе CP2K [53]. Соответствующие программные модули реализуются и на ускорителях, в том числе на GPU [53]. Компания Nvidia предлагает для своих GPU библиотеку cuSPARSE [39]. MAGMA тоже работает с разреженными

матрицами, а GPU от NVIDIA — через cuSPARSE. AMD предлагает для своих GPU библиотеку clSPARSE на OpenCL [38, 54].

В этой области, включая умножение разреженных матриц, ведутся активные исследования [55, 56], и появляются новые более эффективные библиотеки, в которых обеспечивается распараллеливание и реализованы возможности применения на различных ускорителях.

Классической библиотекой для поиска собственных значений в разреженных матрицах следует считать библиотеку PARPACK [57].

Создана также новая библиотека LIBXSMM для оптимизации умножения матриц невысокой размерности, которая может использоваться для умножения разреженных матриц и применяется в комплексе CP2K для расчетов задач молекулярной динамики [58, 59].

Заключение

Представленная в статье информация показывает, что при создании оптимальных по производительности программ целесообразно пользоваться библиотеками. Особенно это важно для HPC-области, где для достижения высокой производительности приложений необходимо распараллеливание процесса обработки. Оптимальный выбор и применение той или иной библиотеки кардинально зависит от приложения и используемых аппаратных средств, способа хранения матриц (плотный/разреженный), применяемых средств распараллеливания. Однако, при написании программного кода необходимо, в первую очередь, работать со стандартными библиотеками BLAS, Lapack и др., поскольку остальные библиотеки базируются именно на них.

В заключение отметим, что средства математических библиотек постоянно развиваются и оптимальный выбор библиотеки для реализации конкретного приложения вполне может быть отдельным исследованием.

Литература

1. **Hopkins B.W.** Chemistry in Parallel Computing [Электрон. ресурс] <http://studylib.net/doc/9480097/parallel-computing-in-chemistry> (дата обращения 17.10.2017).
2. **Annual Rep.** Swiss National Supercomputing Centre (CSCS) [Официальный сайт] http://www.cscs.ch/uploads/tx_factsheet/Annual_Report_2014.pdf (дата обращения 16.10.2017).
3. **High Research Computing Center** in Moscow State University [Официальный сайт] <http://hpc.msu.ru/?q=apps> (дата обращения 17.09.2017).
4. **Luszczek P., Kurzak J., Dongarra J.** Looking Back at Dense Linear Algebra Software // J. Parallel and Distributed Comp. 2014. V. 74. No. 7. Pp. 2548—2560.
5. **Список TOP500** [Официальный сайт] <http://www.top500.org> (дата обращения 10.10.2017).
6. **Basic Linear Algebra Subprograms BLAS** [Официальный сайт] <http://www.netlib.org/blas/> (дата обращения 10.10.2017).

References

1. **Hopkins B.W.** Chemistry in Parallel Computing [Elektron. Resurs] <http://studylib.net/doc/9480097/parallel-computing-in-chemistry> (Data Obrashcheniya 17.10.2017).
2. **Annual Rep.** Swiss National Supercomputing Centre (CSCS) [Ofits. Sayt] http://www.cscs.ch/uploads/tx_factsheet/Annual_Report_2014.pdf (Data Obrashcheniya 16.10.2017).
3. **High Research Computing Center** in Moscow State University [Ofits. Sayt] <http://hpc.msu.ru/?q=apps> (Data Obrashcheniya 17.09.2017).
4. **Luszczek P., Kurzak J., Dongarra J.** Looking Back at Dense Linear Algebra Software. J. Parallel and Distributed Comp. 2014;4;7:2548—2560.
5. **Spisok TOP500** [Ofits. Sayt] <http://www.top500.org> (Data Obrashcheniya 10.10.2017).
6. **Basic Linear Algebra Subprograms BLAS** [Ofits. Sayt] <http://www.netlib.org/blas/> (Data Obrashcheniya 10.10.2017).

7. **Математическая** библиотека LAPACK [Официальный сайт] <http://www.netlib.org/lapack> (дата обращения 10.10.2017).
8. **Scalable** Linear Algebra PACKage SCALAPACK [Официальный сайт] <http://www.netlib.org/scalapack/> (дата обращения 10.10.2017).
9. **Математическая** библиотека MKL [Официальный сайт] <https://software.intel.com/en-us/intel-mkl> (дата обращения 10.10.2017).
10. **Automatically** Tuned Linear Algebra Software a ATLAS [Официальный сайт] <http://math-atlas.sourceforge.net> (дата обращения 10.10.2017).
11. **OpenMP** [Официальный сайт] <http://openmp.org> (дата обращения 15.10.2017).
12. **MPI** Forum [Официальный сайт] <http://www.mpi-forum.org/docs/> (дата обращения 15.10.2017).
13. **Niklasson A.M.N.** Density Matrix Methods in Linear Scaling Electronic Structure Theory // Linear-Scaling Techniques in Computational Chemistry and Phys. Springer Netherlands, 2011. Pp. 439—473.
14. **OpenACC** [Официальный сайт] <https://www.openacc.org/> (дата обращения 15.10.2017).
15. **Описание** стандарта OpenCL [Официальный сайт] <https://www.khronos.org/opencv/> (дата обращения 15.10.2017).
16. **Dongarra J. e. a.** Parallel Programming Models for Dense Linear Algebra on Heterogeneous Systems // Supercomputing Frontiers and Innovations. 2016. V. 2. No. 4. Pp. 67—86.
17. **Описание** библиотеки GlobalArrays [Официальный сайт] <http://hpc.pnl.gov/globalarrays/> (дата обращения 15.10.2017).
18. **PGAS** [Официальный сайт] <http://www.pgas.org/> (дата обращения 24.10.2017).
19. **Krishnan M., Nieplocha J.** SRUMMA: a Matrix Multiplication Algorithm Suitable for Clusters and Scalable-Shared Memory Systems // Proc. 18 Intern. Symp. Parallel and Distributed Proc. 2004. P. 70.
20. **YarKhan A. e. a.** Porting the PLASMA Numerical Library to the OpenMP Standard // Intern. J. Parallel Programming. 2017. V. 45. No. 3. Pp. 612—633.
21. **AMD** Developer Central [Официальный сайт] <http://developer.amd.com/tools-and-sdks/archive/acml-archive-downloads/> (дата обращения 10.10.2017).
22. **High-Performance** BLAS by Kazushige Goto [Официальный сайт] http://www.cs.utexas.edu/users/flame/goto/signup_first.html (дата обращения 10.10.2017).
23. **GOTOBLAS2** [Официальный сайт] <https://www.tacc.utexas.edu/research-development/tacc-software/gotoblas2> (дата обращения 10.10.2017).
24. **Документация** математической библиотеки OpenBLAS [Официальный сайт] <https://github.com/xianyi/OpenBLAS/wiki> (дата обращения 10.10.2017).
25. **OpenBLAS** [Официальный сайт] <http://www.openblas.net/> (дата обращения 10.10.2017).
26. **Naveen G.V.** Accelerating Deep Learning and Machine Learning to a New Level // CMG — Computer
7. **Matematicheskaya** Biblioteka LAPACK [Ofits. Sayt] <http://www.netlib.org/lapack> (Data Obrashcheniya 10.10.2017).
8. **Scalable** Linear Algebra PACKage SCALAPACK [Ofits. Sayt] <http://www.netlib.org/scalapack/> (Data Obrashcheniya 10.10.2017).
9. **Matematicheskaya** Biblioteka MKL [Ofits. Sayt] <https://software.intel.com/en-us/intel-mkl> (Data Obrashcheniya 10.10.2017).
10. **Automatically** Tuned Linear Algebra Software a ATLAS [Ofits. Sayt] <http://math-atlas.sourceforge.net> (Data Obrashcheniya 10.10.2017).
11. **OpenMP** [Ofits. Sayt] <http://openmp.org> (Data Obrashcheniya 15.10.2017).
12. **MPI** Forum [Ofits. Sayt] <http://www.mpi-forum.org/docs/> (Data Obrashcheniya 15.10.2017).
13. **Niklasson A.M.N.** Density Matrix Methods in Linear Scaling Electronic Structure Theory. Linear-Scaling Techniques in Computational Chemistry and Phys. Springer Netherlands, 2011:439—473.
14. **OpenACC** [Ofits. Sayt] <https://www.openacc.org/> (Data Obrashcheniya 15.10.2017).
15. **Opisanie** Standarta OpenCL [Ofits. Sayt] <https://www.khronos.org/opencv/> (Data Obrashcheniya 15.10.2017).
16. **Dongarra J. e. a.** Parallel Programming Models for Dense Linear Algebra on Heterogeneous System. Supercomputing Frontiers and Innovations. 2016;2;4: 67—86.
17. **Opisanie** Biblioteki GlobalArrays [Ofits. Sayt] <http://hpc.pnl.gov/globalarrays/> (Data Obrashcheniya 15.10.2017).
18. **PGAS** [Ofits. Sayt] <http://www.pgas.org/> (Data Obrashcheniya 24.10.2017).
19. **Krishnan M., Nieplocha J.** SRUMMA: a Matrix Multiplication Algorithm Suitable for Clusters and Scalable-Shared Memory Systems. Proc. 18 Intern. Symp. Parallel and Distributed Proc. 2004:70.
20. **YarKhan A. e. a.** Porting the PLASMA Numerical Library to the OpenMP Standard. Intern. J. Parallel Programming. 2017;45;3:612—633.
21. **AMD** Developer Central [Ofits. Sayt] <http://developer.amd.com/tools-and-sdks/archive/acml-archive-downloads/> (Data Obrashcheniya 10.10.2017).
22. **High-Performance** BLAS by Kazushige Goto [Ofits. Sayt] http://www.cs.utexas.edu/users/flame/goto/signup_first.html (Data Obrashcheniyaya 10.10.2017).
23. **GOTOBLAS2** [Ofits. Sayt] <https://www.tacc.utexas.edu/research-development/tacc-software/gotoblas2> (Data Obrashcheniya 10.10.2017).
24. **Dokumentatsiya** Matematicheskoy Biblioteki OpenBLAS [Ofits. Sayt] <https://github.com/xianyi/OpenBLAS/wiki> (Data Obrashcheniya 10.10.2017).
25. **OpenBLAS** [Ofits. Sayt] <http://www.openblas.net/> (Data Obrashcheniya 10.10.2017).
26. **Naveen G.V.** Accelerating Deep Learning and Machine Learning to a New Level // CMG — Computer

Measurement Group India [Электрон. ресурс] http://www.cmgindia.org/wp-content/uploads/2016/12/Intel_CMG_India_Keynote_2016.pdf (дата обращения 10.10.2017).

27. **Intel Math Kernel Library** [Электрон. ресурс] https://software.intel.com/sites/default/files/managed/e0/9d/mkl-11.3.2-developer-reference-fortran_0.pdf (дата обращения 10.10.2017).

28. **Quintero D. e. a.** High-Performance Computing Guide // IBM Redbooks [Электрон. ресурс] <http://www.redbooks.ibm.com/redbooks/pdfs/sg10371.pdf> (дата обращения 01.09.2017).

29. **cuBLAS** [Электрон. ресурс] <http://docs.nvidia.com/cuda/cublas/index.html> (дата обращения 10.10.2017).

30. **Kurzak J. e. a.** Designing SLATE: Software for Linear Algebra Targeting Exascale // Innovative Computing Laboratory [Электрон. ресурс] <http://www.icl.utk.edu/files/publications/2017/icl-utk-980-2017.pdf> (дата обращения 15.10.2017).

31. **Воеводин В.В., Воеводин В.В.** Параллельные вычисления. СПб.: БХВ-Петербург, 2002.

32. **MAGMA** [Электрон. ресурс] <https://developer.nvidia.com/magma> (дата обращения 10.10.2017).

33. **PLASMA** [Электрон. ресурс] <http://icl.cs.utk.edu/plasma/> (дата обращения 10.10.2017).

34. **Agullo E. e. a.** Numerical Linear Algebra on Emerging Architectures: The PLASMA and MAGMA projects // J. Physics. 2009. V. 180. No. 1. P. 012037.

35. **MAGMA** [Электрон. ресурс] <http://icl.cs.utk.edu/magma/> (дата обращения 10.10.2017).

36. **Intel Developer Zone** [Официальный сайт] <https://software.intel.com/sites> (дата обращения 10.10.2017).

37. **Dongarra J. e. a.** Accelerating Numerical Dense Linear Algebra Calculations with GPUs // Numerical Computations with GPUs. Springer Intern. Publ., 2014. Pp. 3—28.

38. **CLSPARSE** [Электрон. ресурс] <http://gropen.com/compute-product/clsparse/> (дата обращения 10.10.2017).

39. **Перечень математических библиотек** [Электрон. ресурс] <http://www.nvidia.ru/object/tesla-gpu-accelerated-libraries-ru.html> (дата обращения 10.10.2017).

40. **EISPACK** [Электрон. ресурс] <http://www.netlib.org/eispack/> (дата обращения 15.10.2017).

41. **PEIGS** [Электрон. ресурс] <http://hpc.pnl.gov/globalarrays/peigs.shtml> (дата обращения 15.10.2017).

42. **Global Arrays** [Электрон. ресурс] <http://hpc.pnl.gov/globalarrays/> (дата обращения 23.10.2017).

43. **Kendall R. A. e. a.** High Performance Computational Chemistry: an Overview of NWChem a Distributed Parallel Application // Computer Phys. Communications. 2000. V. 128. No. 1—2. Pp. 260—283.

44. **Sunderland A.G., Breitmoser E.Y.** An Overview of Eigensolvers for HPCx // Complexity. 2003. V. 2. Pp. 3—12.

Measurement Group India [Elektron. Resurs] http://www.cmgindia.org/wp-content/uploads/2016/12/Intel_CMG_India_Keynote_2016.pdf (Data Obrashcheniya 10.10.2017).

27. **Intel Math Kernel Library** [Elektron. Resurs] https://software.intel.com/sites/default/files/managed/e0/9d/mkl-11.3.2-developer-reference-fortran_0.pdf (Data Obrashcheniya 10.10.2017).

28. **Quintero D. e. a.** High-Performance Computing Guide // IBM Redbooks [Elektron. Resurs] <http://www.redbooks.ibm.com/redbooks/pdfs/sg10371.pdf> (Data Obrashcheniya 01.09.2017).

29. **cuBLAS** [Elektron. Resurs] <http://docs.nvidia.com/cuda/cublas/index.html> (Data Obrashcheniya 10.10.2017).

30. **Kurzak J. e. a.** Designing SLATE: Software for Linear Algebra Targeting Exascale // Innovative Computing Laboratory [Elektron. Resurs] <http://www.icl.utk.edu/files/publications/2017/icl-utk-980-2017.pdf> (Data Obrashcheniya 15.10.2017).

31. **Voevodin V.V., Voevodin V.V.** Parallel'nye Vychisleniya. SPb.: BHV-Peterburg, 2002. (in Russian).

32. **MAGMA** [Elektron. Resurs] <https://developer.nvidia.com/magma> (Data Obrashcheniya 10.10.2017).

33. **PLASMA** [Elektron. Resurs] <http://icl.cs.utk.edu/plasma/> (Data Obrashcheniya 10.10.2017).

34. **Agullo E. e. a.** Numerical Linear Algebra on Emerging Architectures: The PLASMA and MAGMA projects. J. Physics. 2009;180;1:012037.

35. **MAGMA** [Elektron. Resurs] <http://icl.cs.utk.edu/magma/> (Data Obrashcheniya 10.10.2017).

36. **Intel Developer Zone** [Ofits. Sayt] <https://software.intel.com/sites> (Data Obrashcheniya 10.10.2017).

37. **Dongarra J. e. a.** Accelerating Numerical Dense Linear Algebra Calculations with GPUs. Numerical Computations with GPUs. Springer Intern. Publ., 2014:3—28.

38. **CLSPARSE** [Elektron. Resurs] <http://gropen.com/compute-product/clsparse/> (Data Obrashcheniya 10.10.2017).

39. **Perechen' Matematicheskikh Bibliotek** [Elektron. Resurs] <http://www.nvidia.ru/object/tesla-gpu-accelerated-libraries-ru.html> (Data Obrashcheniya 10.10.2017). (in Russian).

40. **EISPACK** [Elektron. Resurs] <http://www.netlib.org/eispack/> (Data Obrashcheniya 15.10.2017).

41. **PEIGS** [Elektron. Resurs] <http://hpc.pnl.gov/globalarrays/peigs.shtml> (Data Obrashcheniya 15.10.2017).

42. **Global Arrays** [Elektron. Resurs] <http://hpc.pnl.gov/globalarrays/> (Data Obrashcheniya 23.10.2017).

43. **Kendall R. A. e. a.** High Performance Computational Chemistry: an Overview of NWChem a Distributed Parallel Application. Computer Phys. Communications. 2000;128;1:260—283.

44. **Sunderland A.G., Breitmoser E.Y.** An Overview of Eigensolvers for HPCx. Complexity. 2003;2:3—12.

45. **Hammarling S.** New Developments in LAPACK and ScaLAPACK [Электрон. ресурс] <http://www.maths.bath.ac.uk/~masrs/ma50177/pdfs/bath-25apr07.pdf> (дата обращения 15.10.2017).
46. **Sparse Matrix Storage Formats for Sparse BLAS Routines** [Электрон. ресурс] <https://software.intel.com/en-us/mkl-developer-reference-c-sparse-matrix-storage-formats-for-sparse-blas-routines> (дата обращения 15.10.2017).
47. **Introduction** to the Intel MKL Extended Eigensolver [Электрон. ресурс] <https://software.intel.com/en-us/articles/introduction-to-the-intel-mkl-extended-eigensolver> (дата обращения 15.10.2017).
48. **ELPA** [Официальный сайт] <https://elpa.mpcdf.mpg.de/> (дата обращения 15.10.2017).
49. **Marek A.** The ELPA Library — Scalable Parallel Eigenvalue Solutions for Electronic Structure Theory and Computational Science // *J. Physics*. 2014. V. 26. P. 213201.
50. **Bock N., Challacombe M.** An Optimized Sparse Approximate Matrix Multiply // *CoRR*. 2012. Pp. 1—16.
51. **Bock N., Challacombe M.** An Optimized Sparse Approximate Matrix Multiply for Matrices With Decay // *SIAM J. Scientific Comp.* 2013. V. 35. No. 1. Pp. 72—98.
52. **Sparse BLAS** [Электрон. ресурс] <http://www.netlib.org/sparse-blas/> (дата обращения 10.10.2017).
53. **Borstnik U. e. a.** Sparse Matrix Multiplication: the Distributed Block-compressed Sparse Row Library // *Parallel Comp.* 2014. V. 40. No. 5. Pp. 47—58.
54. **Greathouse J.L. e. a.** cLSPARSE: a Vendor-optimized Open-source Sparse BLAS Library // *Proc. IV Intern. Workshop on OpenCL*. 2016. P. 7.
55. **Piccolo A., Soodla J.** Performance of Parallel Sparse Matrix-matrix multiplication // *SCRIBD* [Электрон. ресурс] <http://www.diva-portal.org/smash/get/diva2:821031/FULLTEXT01.pdf> (дата обращения 15.10.2017).
56. **Rubensson E.H., Rudberg E.** Locality-aware Parallel Block-sparse Matrix-matrix Multiplication Using the Chunks and Tasks Programming Model // *Parallel Computing*. Elsevier, 2016.
57. **PARPACK** [Электрон. ресурс] www.caam.rice.edu/~kristyn/parpack_home.html (дата обращения 15.10.2017).
58. **Heinecke A. e. a.** LIBXSMM: Accelerating Small Matrix Multiplications by Runtime Code Generation // *Proc. High Performance Computing, Networking, Storage and Analysis Intern. Conf.* 2016. Pp. 981—991.
59. **Zhao Z. e. a.** Performance of Hybrid MPI/OpenMP VASP on Cray XC40 Based on Intel Knights Landing Many Integrated Core Architecture [Электрон. ресурс] https://cug.org/proceedings/cug2017_proceedings/includes/files/pap134s2-file1.pdf (дата обращения 10.10.2017).
45. **Hammarling S.** New Developments in LAPACK and ScaLAPACK [Elektron. Resurs] <http://www.maths.bath.ac.uk/~masrs/ma50177/pdfs/bath-25apr07.pdf> (Data Obrashcheniya 15.10.2017).
46. **Sparse Matrix Storage Formats for Sparse BLAS Routines** [Elektron. Resurs] <https://software.intel.com/en-us/mkl-developer-reference-c-sparse-matrix-storage-formats-for-sparse-blas-routines> (Data Obrashcheniya 15.10.2017).
47. **Introduction** to the Intel MKL Extended Eigensolver [Elektron. Resurs] <https://software.intel.com/en-us/articles/introduction-to-the-intel-mkl-extended-eigensolver> (Data Obrashcheniya 15.10.2017).
48. **ELPA** [Ofits. Sayt] <https://elpa.mpcdf.mpg.de/> (Data Obrashcheniya 15.10.2017).
49. **Marek A.** The ELPA Library — Scalable Parallel Eigenvalue Solutions for Electronic Structure Theory and Computational Science. *J. Physics*. 2014;26:213201.
50. **Bock N., Challacombe M.** An Optimized Sparse Approximate Matrix Multiply. *CoRR*. 2012:1—16.
51. **Bock N., Challacombe M.** An Optimized Sparse Approximate Matrix Multiply for Matrices With Decay. *SIAM J. Scientific Comp.* 2013;35;1:72—98.
52. **Sparse BLAS** [Elektron. Resurs] <http://www.netlib.org/sparse-blas/> (Data Obrashcheniya 10.10.2017).
53. **Borstnik U. e. a.** Sparse Matrix Multiplication: the Distributed Block-compressed Sparse Row Library. *Parallel Comp.* 2014;40;5:47—58.
54. **Greathouse J.L. e. a.** cLSPARSE: a Vendor-optimized Open-source Sparse BLAS Library. *Proc. IV Intern. Workshop on OpenCL*. 2016:7.
55. **Piccolo A., Soodla J.** Performance of Parallel Sparse Matrix-matrix multiplication // *SCRIBD* [Elektron. Resurs] <http://www.diva-portal.org/smash/get/diva2:821031/FULLTEXT01.pdf> (Data Obrashcheniya 15.10.2017).
56. **Rubensson E.H., Rudberg E.** Locality-aware Parallel Block-sparse Matrix-matrix Multiplication Using the Chunks and Tasks Programming Model. *Parallel Computing*. Elsevier, 2016.
57. **PARPACK** [Elektron. Resurs] www.caam.rice.edu/~kristyn/parpack_home.html (Data Obrashcheniya 15.10.2017).
58. **Heinecke A. e. a.** LIBXSMM: Accelerating Small Matrix Multiplications by Runtime Code Generation. *Proc. High Performance Computing, Networking, Storage and Analysis Intern. Conf.* 2016:981—991.
59. **Zhao Z. e. a.** Performance of Hybrid MPI/OpenMP VASP on Cray XC40 Based on Intel Knights Landing Many Integrated Core Architecture [Elektron. Resurs] https://cug.org/proceedings/cug2017_proceedings/includes/files/pap134s2-file1.pdf (Data Obrashcheniya 10.10.2017).

Сведения об авторах:

Кружилов Иван Сергеевич — кандидат технических наук, доцент кафедры прикладной математики НИУ «МЭИ», e-mail: bearsubj13@yandex.ru

Кузьминский Михаил Борисович — кандидат химических наук, старший научный сотрудник Института органической химии им. Н.Д. Зелинского РАН, e-mail: kus@free.net

Чернецов Андрей Михайлович — кандидат технических наук, доцент кафедры прикладной математики НИУ «МЭИ», научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, e-mail: ChernetsovAM@mpei.ru; an@ccas.ru

Шамаева Ольга Юрьевна — кандидат технических наук, доцент кафедры прикладной математики НИУ «МЭИ», e-mail: shamayevaou@mpei.ru

Information about authors:

Kruzhilov Ivan S. — Ph.D. (Techn.), Assistant Professor of Applied Mathematics Dept., NRU MPEI, e-mail: bearsubj13@yandex.ru

Kuzminsky Mikhail B. — Ph.D. (Chem.), Senior Researcher of N.D. Zelinsky Institute of Organic Chemistry Russian Academy of Sciences, e-mail: kus@free.net

Chernetsov Andrey M. — Ph.D. (Techn.), Assistant Professor of Applied Mathematics Dept., NRU MPEI, Research Assistant of Dorodnicyn Computing Centre FRC CSC RAS, e-mail: ChernetsovAM@mpei.ru; an@ccas.ru

Shamayeva Olga Yu. — Ph.D. (Techn.), Assistant Professor of Applied Mathematics Dept., NRU MPEI, e-mail: shamayevaou@mpei.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов

Conflict of interests: the authors declare no conflict of interest

Статья поступила в редакцию: 07.11.2017

The article received to the editor: 07.11.2017