

Библиотека для умножения матриц написана в рамках цикла исследовательских работ, проводимых специалистами компании «*Базовые Проекты*». Главной мотивацией и аргументом для использования языка ассемблера, является повышение производительности. Высокопроизводительные вычисления или *HPC (High Performance Computing)* и их практическое применение для моделирования сложных систем являются одним из основных направлений деятельности нашего коллектива. Многолетний опыт программирования вычислительных задач накоплен нашими специалистами еще со времен математического сопроцессора Intel 8087 и первых персональных компьютеров IBM PC. Новые разработанные теоретические методы и принципы компьютерного моделирования обеспечивают решение задач компьютерного моделирования «больших систем» и сложной техники, которые ранее считались нерешаемыми. См. также информацию в конце файла.

Операция умножения матриц реализована с использованием 256-битного векторного расширения AVX и оптимизирована для выполнения на многопроцессорных серверных платформах. Текущая версия библиотеки предназначена для операционных систем семейства Microsoft Windows и процессоров Intel Sandy Bridge. Проект находится в непрерывном развитии. Планируется поддержка новых функциональных расширений (AVX512) и операционных систем семейства Linux, а также расширение набора выполняемых вычислительных операций.

Библиотека разработана двумя специалистами:

Парфус Александр Владимирович. Начальник департамента IT. Кандидат технических наук. Разработал вычислительные алгоритмы и технологию «экстремальной оптимизации» производительности многоуровневой кэш-памяти для эффективной реализации процедуры умножения матриц на современных мультипроцессорных системах с многоуровневой кэш-памятью. Разработал модель разделения функциональности и интерфейс между программой верхнего уровня и низкоуровневой вычислительной библиотекой.

Манусов Илья Вадимович. Главный специалист по аппаратному обеспечению и системному программированию. Выполнял исследование критических факторов производительности. Запрограммировал и отладил алгоритмы на языке ассемблера с использованием векторного расширения AVX256. Реализовал бенчмарки для вычислительных процедур и типовых операций чтения и записи кэш-памяти различных уровней и оперативной памяти.

Функции библиотеки выполняются как один поток, вместе с тем, оптимизированы для многопоточковой обработки на мультипроцессорной

системе. Предполагается, что управление параллельным запуском потоков выполняется в программе верхнего уровня.

Библиотека написана на языке ассемблера для Intel x64. Результатом трансляции являются файлы DLL, предназначенные для выполнения в среде 64-битных ОС семейства Microsoft Windows. В качестве транслятора используется Flat Assembler (FASM) версии 1.71.17. Отладчик уровня ассемблерного кода FDBG. Среда разработки FASM Editor 2.0. Для правильного отображения табуляций и цветового выделения комментариев рекомендуется использовать данную среду разработки при работе с исходными текстами программы.

Библиотека использует 256-битное векторное расширение Intel AVX первого поколения и на данном этапе разработки оптимизирована для процессоров Intel с ядром Sandy Bridge. Производительность, близкая к пиковой, соответствующей выполнению двух векторных операций (одного сложения и одного умножения) каждым ядром за каждый такт, достигнута для Sandy Bridge. При этом для ряда тестовых примеров, удалось превзойти показатели производительности известной математической библиотеки Intel MKL на единицы процентов. В ближайшее время будут опубликованы результаты комплексного тестирования. На других процессорах, поддерживающих AVX256, быстродействие текущей версии библиотеки может быть не оптимальным. В разработке находится библиотека, использующая векторное расширение AVX512 для процессоров Intel Xeon на основе ядра Skylake.

Библиотека использует два базовых алгоритма, выбираемых программой верхнего уровня в зависимости от размерности умножаемых матриц с целью достижения максимальной производительности применительно к конкретной задаче. Более детальное описание находится в документе [*15-09-27_AVX_256_Specific_Specification_v200.DOC*](#).

В документе [*15-10-13_AVX_OS_DLL_Specification_v202.DOC*](#) находится описание вспомогательной библиотеки конфигурирования вычислительной платформы, предназначенной для детектирования функциональных расширений, поддерживаемых процессором, размеров кэш-памяти и других параметров.

Описание основных файлов библиотеки конфигурирования вычислительной платформы AVX_OS.

Транслируемый модуль:

AVX_OS.asm

Модули для отладки функций:

Debug_DLL_DetectCache.asm

Debug_DLL_DetectVXLib.asm

Подключаемые модули:

DetectCache.inc

DetectVXLib.inc

**Описание основных файлов вычислительной библиотеки
AVX_256_Specific.**

Транслируемый модуль:

Specific_AVX1_256_1.asm

Модули для отладки функций:

Debug_DLL_InitLibrary.asm

Debug_DLL_PackA.asm

Debug_DLL_PackASmall.asm

Debug_DLL_PackB.asm

Debug_DLL_PackBSmall.asm

Debug_DLL_UnPackA.asm

Debug_DLL_UnPackASmall.asm

Debug_DLL_UnPackB.asm

Debug_DLL_UnPackBSmall.asm

Debug_MultiplyMatrixBig.asm

Debug_MultiplyMatrixSmall.asm

Подключаемые модули:

<DIR> MMB

<DIR> MMS

Debug_Context.inc

InitLibrary.inc

PackASmall.inc

PackA_TRANSP.inc

PackB.inc

PackBSmall.inc

UnPackA.inc

UnPackASmall.inc

UnPackB.inc

UnPackBSmall.inc

Кратко «о нас».

Костяк коллектива разработчиков составляют выходцы из институтов системы АН бывшего СССР, высококвалифицированные специалисты, доктора и кандидаты наук, сфера деятельности которых – решение проблем моделирования «больших систем». Это представители научной школы акад. Георгия Евгеньевича Пухова, а также другие специалисты.

В качестве примера рассмотрим задачи моделирования газотранспортных систем (ГТС). В первую очередь речь идет о решении таких основных задач:

- идентификация фактических режимов работы ГТС и параметров её элементов (решение задачи вариационного исчисления);
- решение задачи оптимального управления режимами работы ГТС.

В обеих задачах рассматриваются реальные нестационарные режимы работы ГТС. Могут рассматриваться ГТС любого уровня сложности, вплоть до ГТС континентального уровня.

При решении подобных задач нужно быть готовым к следующим размерностям системы: количество степеней свободы (количество переменных) может составлять величину «миллионы и десятки миллионов неизвестных»; количество учитываемых технологических ограничений оборудования ГТС может составлять величину «сотни тысяч и миллионы».

«Проклятие размерности» делает невозможным решение таких задач известными математическими методами.

Термин «Проклятие размерности» отражает лавинообразный рост объема необходимых вычислений, возникающий по мере увеличения объекта моделирования. Рассматриваемые задачи относятся к задачам экспоненциальной сложности.

До недавнего времени задача «Оптимизация режимов работы ГТС» не поддавалась расчёту. Это справедливо вообще для задач оптимизации (а также обратных задач) так называемых «Больших систем».

В основу нашей разработки заложены новые принципы и методы создания точных численных моделей ГТС (и вообще сложных систем), позволяющие решать в ускоренном масштабе времени задачи компьютерного моделирования реальных объектов и систем практически любого уровня сложности. В том числе те задачи, которые в настоящее время считаются нерешаемыми на базе современного уровня развития вычислительной техники. **Обеспечивается снятие «Проклятия размерности»,** довлеющего над задачами моделирования и оптимизации «больших систем».

Разработанные методы и алгоритмы «высокого уровня», в сочетании с методиками «экстремальной оптимизации» низкоуровневых вычислительных алгоритмов, позволяют решать задачи суперкомпьютерного уровня, используя вычислительные платформы массового применения.

Опубликованные исходные тексты вычислительной библиотеки являются одним из примеров такой оптимизации.

Некоторые разработанные алгоритмы аппаратного ускорения вычислений предполагается включить в данный проект по мере его развития.

* * *

Контакты:

Руководитель разработки Александр Парфус

E-mail: AVP@basic-projects.com

Главный специалист по аппаратному обеспечению Илья Манусов

E-mail: im1969@yandex.ru