

Ускорение инженерных расчетов в ANSYS Mechanical при использовании GPU NVIDIA Tesla

Долгое время графические процессоры (Graphics Processing Unit, GPU), работая в персональных компьютерах и рабочих станциях вместе с центральными процессорами (Central Processing Unit, CPU), использовались только по своему прямому назначению – для ускорения вычислений, связанных с визуализацией данных. С ростом мощности GPU у разработчиков ПО появилось желание переложить на них ряд задач, которые до этого момента выполнялись CPU. Были разработаны библиотеки, позволяющие выполнять программный код на графических процессорах, и подход GPU Computing (использование GPU для вычислений общего назначения) получил очень широкое распространение в самых разных сферах, начиная от астрофизики и заканчивая промышленным дизайном. Более того, несколько лет назад даже появились специальные графические процессоры, предназначенные исключительно для ресурсоемких вычислений. Такие вычислители поддерживают выполнение операций с двойной точностью, механизм коррекции ошибок (Error Code Correction, ECC), обладают большим объемом встроенной памяти для эффективной работы с большими массивами данных.

Современный графический процессор содержит сотни вычислительных ядер, размещенных на одном кристалле. Во многих задачах вычислительные возможности GPU в десятки и сотни раз превосходят самые мощные CPU. Кроме того, графические процессоры оказываются более энергоэффективными, то есть показывают лучшее соотношение потребляемой энергии к производительности системы.



Рис. 1. NVIDIA Tesla C2070

При решении инженерных задач механики деформируемого твердого тела методом конечных элементов (МКЭ) традиционно используется большое количество вычислительных алгоритмов, реализация которых может быть легко делегирована с ядер центральных процессоров на ядра графических процессоров. Таким образом, совместное применение GPU и CPU позволяет эффективно использовать все вычислительные ресурсы рабочей станции и существенно сократить время решения инженерных задач.

В ноябре 2010 года компания ANSYS, лидер рынка компьютерного инжиниринга (Computer-Aided Engineering, CAE) в области численного моделирования и высокопроизводительных вычислений, сообщила о поддержке ускорения расчетов некоторых классов задач посредством специализированных процессоров NVIDIA Tesla (рис. 1). Данная поддержка осуществляется начиная с 13-й версии программной системы инженерного анализа ANSYS, ставшей доступной пользователям в декабре 2010 года.

По словам ведущего разработчика программного обеспечения ANSYS Джеффа Бейшейма (Jeff Beisheim), недавно появившаяся возможность использования графических процессоров при расчетах в ANSYS Mechanical позволяет существенно снизить время для решения задач большой размерности.

Поддерживаемые типы расчетов

Использование GPU при работе с ANSYS 13.0 для задач механики деформируемого твердого тела (МДТТ) помогает ускорить выполнение задач при использовании как прямого метода решения систем уравнений Sparse (для систем с разреженной матрицей), так и итерационных методов, основанных на методе сопряженных градиентов с предварительным улучшением обусловленности матрицы – PCG (preconditioned conjugate gradient) и JCG (Jacobi conjugate gradient). Для задач поиска собственных частот и форм колебаний и задач потери устойчивости возможно достичь ускорения при использовании блочного метода Ланцоша – Block Lanczos и PCG Lanczos.

Выбор платформы и лицензионная политика

В версии ANSYS 13.0 использование GPU в расчетах поддерживается как для систем на базе Windows x64,

так и Linux x64. Максимальное ускорение расчетов обеспечивается поддержкой наиболее мощных процессоров NVIDIA Tesla серии 20 (2050, 2070, 2090), построенных на архитектуре Fermi. Начиная с версии ANSYS 14.0 будет предоставлена поддержка карт NVIDIA Quadro 6000.

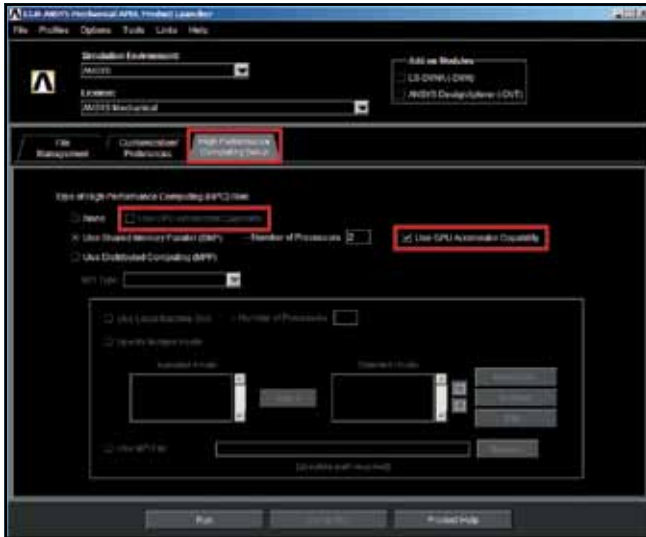


Рис. 2

Согласно лицензионной политике ANSYS, для использования GPU в расчетах не требуется приобретать специальную лицензию. В коммерческих версиях необходимо наличие хотя бы одной лицензии ANSYS HPC Pack, используемой для организации доступа к параллельным вычислениям на CPU. Для пользователей академических лицензий

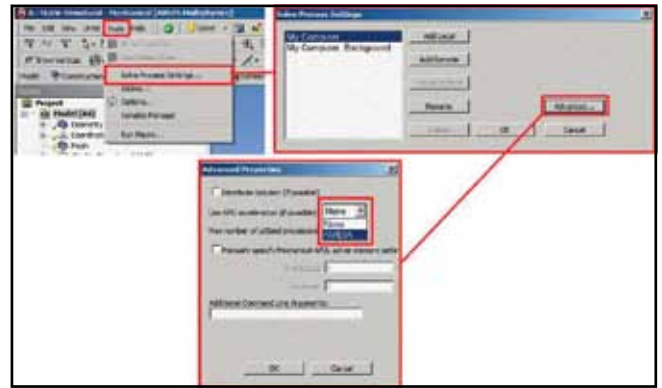


Рис. 3

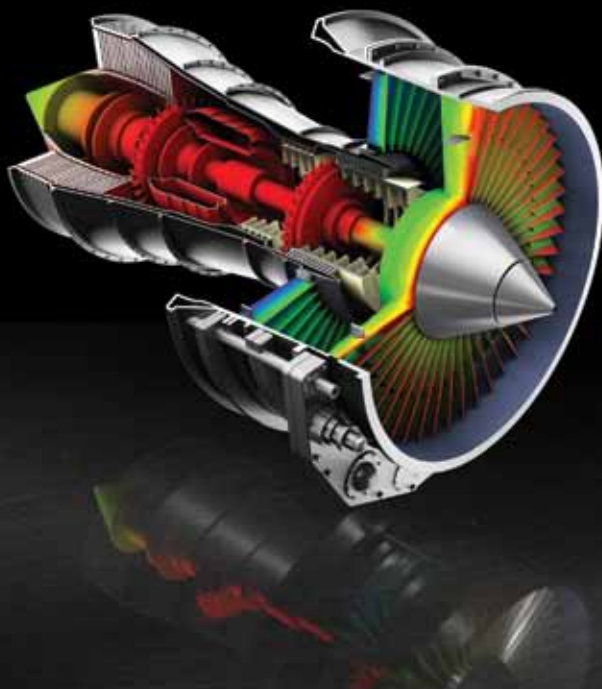
ANSYS Academic Research использование GPU возможно без каких-либо дополнительных лицензий или условий.

Активация GPU ускорения расчетов

В версии ANSYS 13.0 при решении задач МДТТ возможно задействовать один графический процессор GPU, однако в следующих релизах пакета анонсирована поддержка нескольких GPU для расчетов.

Для активации GPU-ускорения в ANSYS Mechanical APDL при запуске через ANSYS MAPDL Launcher необходимо выбрать соответствующую опцию на вкладке High-Performance Setup (рис. 2). Того же эффекта можно добиться, запустив ANSYS MAPDL с ключом `-acc nvidia`.

Пользователи ANSYS Workbench могут активировать возможности GPU в меню Solve Process Settings > Advanced Properties модуля Mechanical (рис. 3).



**МОДЕЛИРУЙТЕ БЫСТРЕЕ
В ANSYS MECHANICAL R13
С NVIDIA® TESLA™**



HPC-T470

ГИБРИДНЫЙ СУПЕРКОМПЬЮТЕР

Пиковая произ-ть GPU: 2.06TFlops

Процессор: 2xIntel Xeon 56xx

Объем памяти CPU: до 96Gb ECC DDR3

Объем памяти GPU: до 24Gb ECC GDDR5

Вычислитель: 4xTESLA C2070



HPC-ENTEROT 810

ГИБРИДНЫЙ КЛАСТЕР

Пиковая произ-ть GPU: 4.12TFlops

Процессор: 10x Intel Xeon 56xx

Опер. память: до 240Gb ECC DDR3

Объем памяти GPU: до 48Gb ECC GDDR5

Вычислитель: 8xTESLA M2070

FORSITE

ПРОФЕССИОНАЛЬНЫЕ ГРАФИЧЕСКИЕ СИСТЕМЫ

190020, Санкт-Петербург, Бумажная 3

Тел. +7 812 622-09-93 Fax: +7 812 622-09-93 ext 105

sale@forsite-company.ru | www.forsite-company.ru/tesla

После активации соответствующих опций ANSYS будет автоматически задействованы возможности GPU для всех этапов решения задач, поддерживающих данный вид ускорения, и дальнейшее вмешательство пользователя не потребуется.

Дополнительные настройки

Для пользователей, желающих получить больший контроль над процессом ускорения задачи с использованием GPU, добавлена APDL-команда ACCOPTION. Эта команда позволяет:

- ▶ отключить заданное при запуске ANSYS использование GPU;
- ▶ задать параметр MinSzThresh, определяющий минимальную размерность задачи, решаемой Sparse-методом, для которой будет активировано GPU-ускорение;
- ▶ задать параметр SPkey для контроля использования одинарной или двойной точности математических операций при использовании прямого метода Sparse.

Кроме того, некоторые дополнительные настройки оборудования могут быть полезны при определенных сценариях:

- ▶ переменные окружения в ANSYS Mechanical APDL позволят избежать одновременного использования одного GPU двумя и более пользователями многопользовательской среды;
- ▶ режим монополии для NVIDIA GPU позволит полностью занять GPU одним процессом (например, одним расчетом ANSYS Mechanical), не выделяя ресурсов для других претендующих на использование GPU процессов;
- ▶ возможность отключения кода коррекции ошибок памяти позволит использовать больший объем памяти GPU и повысит ее общую пропускную способность. Однако для обеспечения точности результатов расчетов этой возможностью пользоваться не рекомендуется.

Когда следует использовать GPU

Выигрыш от использования GPU в процессе расчета будет зависеть от используемого оборудования и типа решаемой задачи. GPU в текущей версии ANSYS

13.0 ускоряют только процесс решения системы алгебраических уравнений, но не операции ее сборки или обработки результатов.

Наиболее благоприятным является сценарий, когда задача полностью помещается в память GPU. Так, NVIDIA Tesla C2050, обладающая 3 Гб оперативной памяти, будет наиболее эффективна для задач размерности от 500 тысяч до 5 млн степеней свободы для прямого метода Sparse и от 500 тысяч до 3000 млн для итерационных методов PCG/JCG. Для GPU, обладающих другим размером оперативной памяти, рекомендуемая размерность задач может отличаться. В случае, когда задача не помещается в память GPU, возможность ускорения ее расчета сохраняется, однако она не будет максимальной.

Для прямого метода Sparse возможно ускорение всех типов расчетов, за исключением процесса создания суперэлементов. GPU-ускорение поддерживает задачи, в процессе решения которых возникают несимметричные матрицы, например модели контактного взаимодействия с трением, использующие команды NROPT и UNSYM. Тем не менее, не поддерживается ускорение расчетов для моделей с использованием конечных элементов со смешанной формулировкой (mixed u-P formulation) или контактных конечных элементов с Лагранжевой формулировкой.

При использовании итерационных методов PCG/JCG возможно ускорение всех поддерживаемых ими типов расчетов. Для метода PCG использование опции экономии памяти (команда MSAVE) деактивирует GPU-ускорение.

Тестирование производительности

В марте-апреле 2011 года сотрудниками лаборатории "Вычислительная механика" (CompMechLab) СПбГПУ, компанией CADFEM CIS – партнером ANSYS в России и компанией "Форсайт" – партнером NVIDIA на Северо-Западе было проведено совместное исследование возможностей сокращения времени расчетов в программной системе ANSYS версий 12.1 и 13.0 при использовании специализированных вычислительных процессоров NVIDIA Tesla и высокоскоростных твердотельных накопителей (Solid State Disk, SSD), позволяющих ускорить процессы ввода-вывода данных в процессе расчетов.

Для проведения тестирования был использован тестовый стенд с двумя вариантами характеристик (таблица).

Операционная система	MS Windows 7 64bit	
Версии ANSYS	ANSYS 12.1 + ANSYS 13.0	
Центральный процессор	Intel Core i7 950 (3,07 ГГц)	
Оперативная память	12 Гб DDR3, 1600 МГц	24 Гб DDR3, 1600 МГц
Материнская плата	ASUS P6T7 WS SUPERCOMPUTER	
Графическая карта	NVIDIA Quadro 2000	
Специализированный процессор	NVIDIA Tesla C2050	
Твердотельные накопители	SSD Intel X25-M	Fast SSD RAID0 на Intel X25-M x2

Технические характеристики процессора NVIDIA Tesla C2050:

- ▶ количество CUDA-ядер – 448;
- ▶ объем специальной памяти GDDR5 – 3 Гб;
- ▶ пропускная способность памяти – 144 Гб/с;
- ▶ производительность – ~ 1 ТФЛОП – одинарная точность и ~ 0,5 ТФЛОП – двойная точность.

Для тестирования производительности рабочей станции при проведении расчетов в ANSYS Mechanical выбран набор тестов ANSYS SP1 BENCH110 Benchmark Suite, который был адаптирован таким образом, чтобы позволить проводить расчеты в ANSYS 12.1 и 13.0 и задействовать возможности NVIDIA Tesla. Данный набор тестов содержит множество представительных задач различной размерности, относящихся к различным классам. Присутствуют линейные/нелинейные, стационарные/нестационарные задачи теории упругости, теории колебаний, теплопроводности и магнитостатики.

1. Статическая задача теории упругости, 850 тыс. степеней свободы, метод Sparse

Данная задача (рис. 4) – средней размерности и должна полностью помещаться в оперативной памяти.

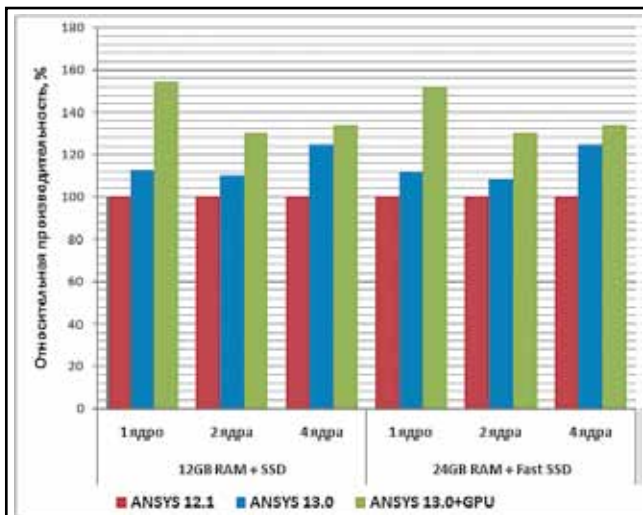
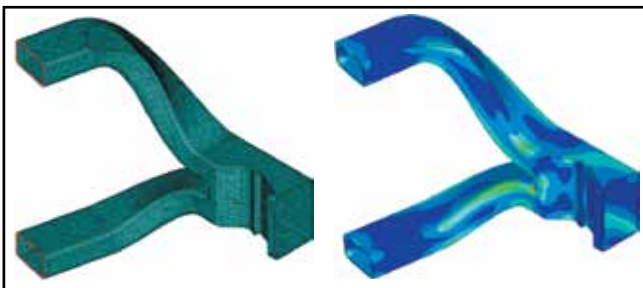


Рис. 4

2. Задача теории колебаний, 760 тыс. степеней свободы, метод Block Lanczos

В ходе решения задачи (рис. 5) проводился расчет 200 частот и форм собственных колебаний. Задача показала сбалансированность работы процессора и жесткого диска. На время решения задачи положительно влияет большой объем оперативной памяти. Выигрыш от ускорения расчета при использовании GPU

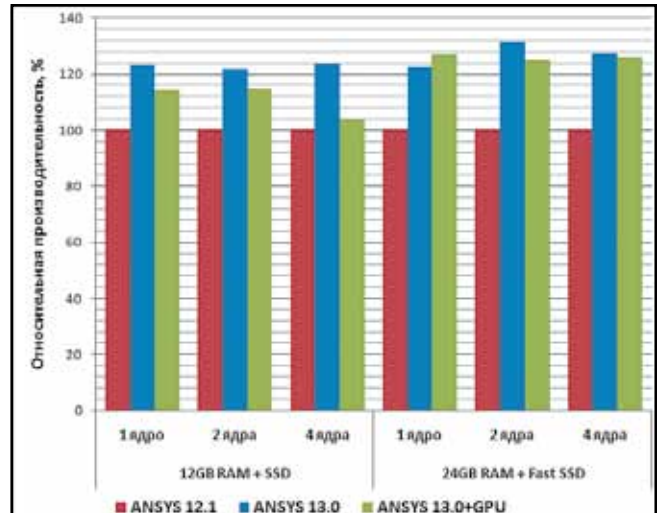
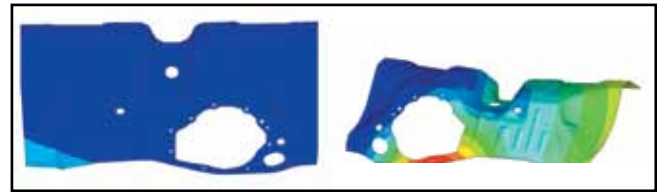


Рис. 5

может нивелироваться на фоне длительного процесса факторизации матриц.

3. Нелинейная контактная задача, 200 тыс. степеней свободы, метод Sparse

Задача (рис. 6) продемонстрировала сбалансированность работы процессора и жесткого диска.

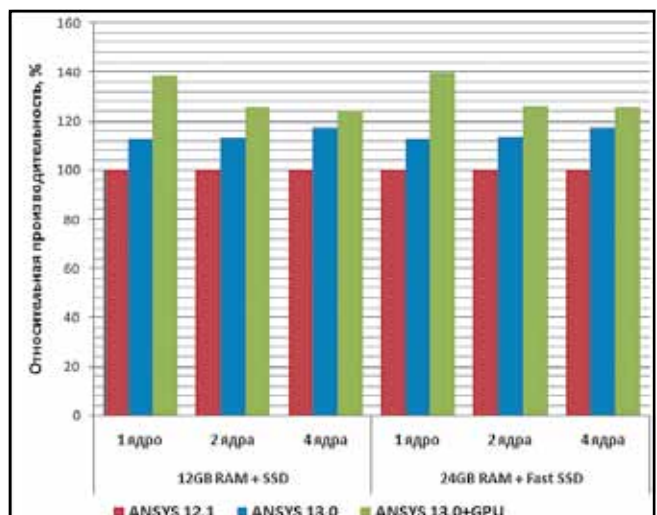
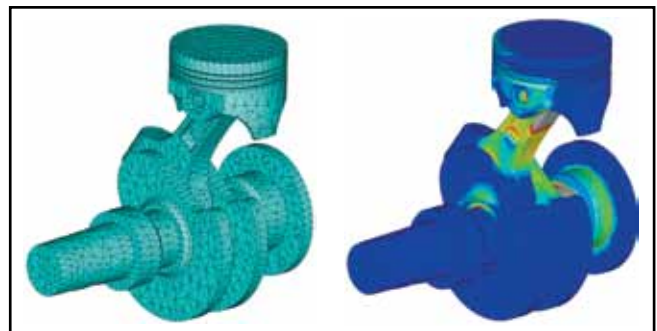


Рис. 6

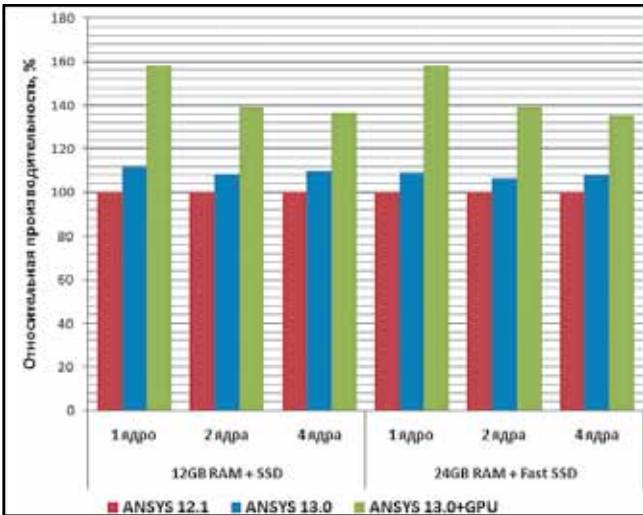
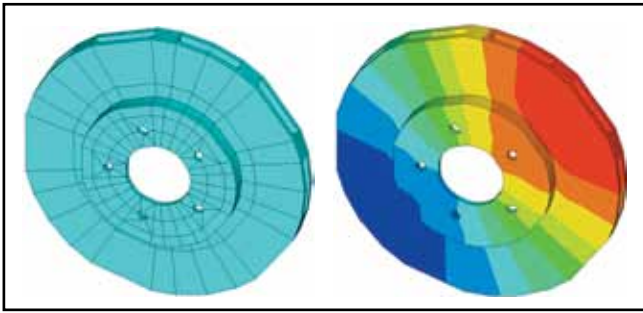


Рис. 7

4. Нестационарная задача теплопроводности, 700 тыс. степеней свободы, метод JCG

Еще одна небольшая задача (рис. 7), которая хорошо решается на любой рабочей станции среднего уровня. Узким местом для итерационного метода, используемого в данной задаче, является пропускная способность памяти.

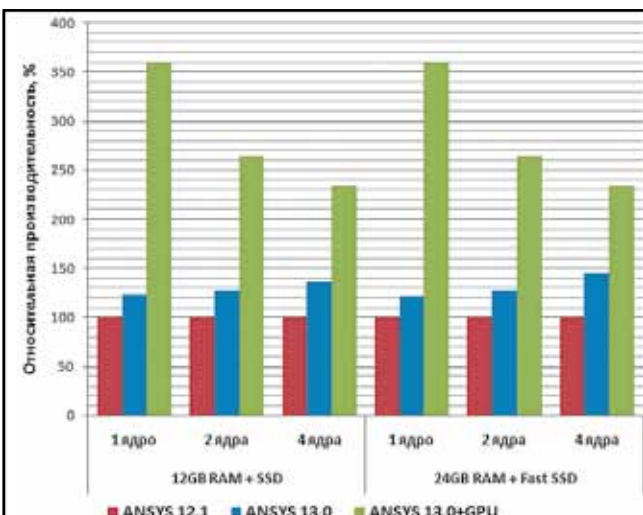
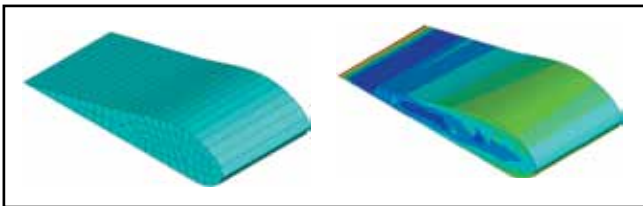


Рис. 8

5. Статическая задача теории упругости, 250 тыс. степеней свободы, метод Sparse

Тест (рис. 8) показывает баланс CPU и I/O-подсистемы. Задача из этого теста решается в памяти практически любой Win32-машины (размерность задачи автоматически подстраивается в тесте). Подпрограммы метода Sparse с двойной точностью создают основную вычислительную нагрузку. Преимущества использования GPU в данной задаче видны наиболее явно.

6. Статическая задача теории упругости, 750 тыс. степеней свободы, метод Sparse

Тест (рис. 9) показывает баланс CPU и I/O-подсистемы. На скорость расчета положительно влияет размер оперативной памяти. Задача требует 16 Гб свободной оперативной памяти.

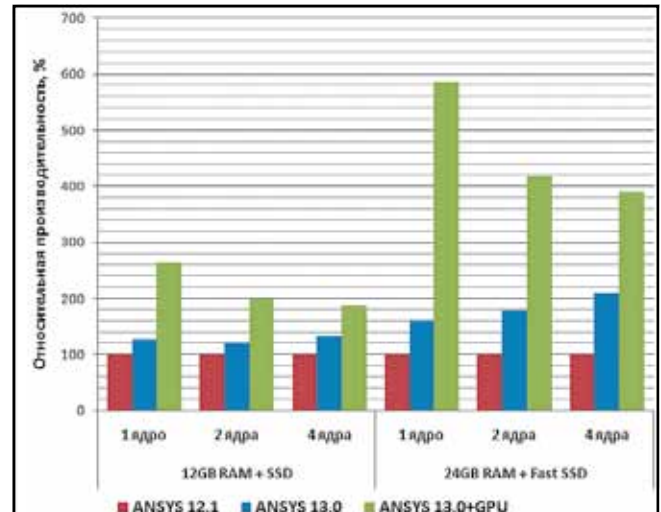
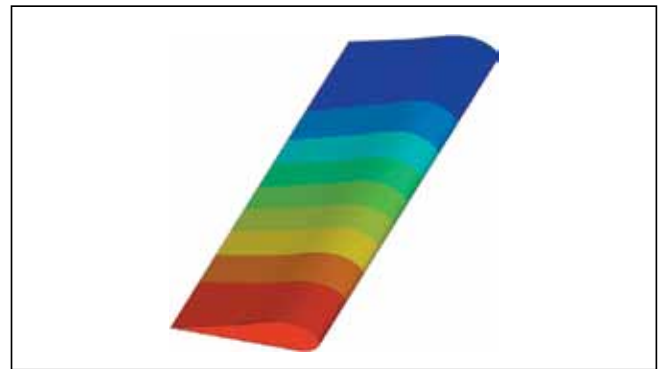


Рис. 9

7. Статическая задача теории упругости, 5 млн степеней свободы, метод PCG

Самая "тяжелая" задача (рис. 10) из данного набора тестов. Лучший тест для пропускной способности подсистемы памяти.

Выводы по итогам тестирования

Результаты проведенных исследований показали существенное сокращение времени расчета задач на GPU в сравнении с результатами на CPU (ANSYS 12.1 на машине с SSD и 12 Гб памяти):

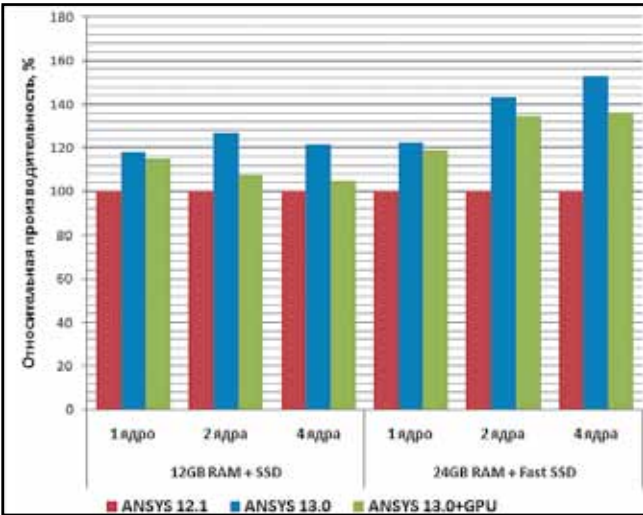
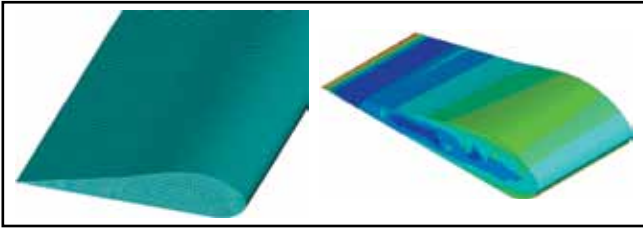


Рис. 10

- ▶ при использовании специализированных вычислительных процессоров NVIDIA Tesla – 20-50 % (до 500 % на некоторых классах задач);
- ▶ при переходе от версии ANSYS 12.1 к ANSYS 13.0 – 10-20 % в среднем (до 100 % на некоторых классах задач).

Увеличение быстродействия дисковой подсистемы и объема доступной для расчета оперативной памяти центрального процессора позволяет значительно повысить эффективность использования GPU. Так, в наиболее ресурсоемком тесте №6 увеличение скорости работы дисковой подсистемы и увеличение оперативной памяти дает обычной машине прирост производительности на 50 %, а машине, использующей GPU, – более чем на 300 %.

На сегодняшний день GPU NVIDIA Tesla позволяют создавать рабочие станции, обладающие максимальной концентрацией вычислительной мощности на объем. Подобные решения являются выигрышными в случае проведения инженерных расчетов для плохо распараллеливаемых задач средней размерности (в настоящий момент – не более 6 миллионов степеней свободы), которые невозможно эффективно решать на классических кластерах. Стоит отметить, что в настоящий момент подобные технологии ускорения расчетов показывают разную эффективность для разных классов задач – наибольший выигрыш получают задачи, решаемые методом Sparse, и задачи поиска собственных частот и форм колебаний.

По итогам тестирования можно однозначно рекомендовать пользователям ANSYS переход на новую версию ANSYS 13, поскольку она позволяет не только получить новые инструменты для решения задач, но и ускорить решение уже имеющихся.

*Евгений Зверев, компания "Форсайт",
Юрий Новожилов, лаборатория
"Вычислительная механика" СПбГПУ,
Дмитрий Михалюк, компания CADFEM CIS*

Открой для себя мир суперкомпьютеров



CRAY

Cray CX1™: высокопроизводительный кластер в компактном корпусе

Персональный суперкомпьютер Cray CX1 позволяет быстро и точно решать самые сложные инженерные задачи в различных отраслях промышленности с использованием программных продуктов ANSYS, Inc. Решение построено на базе ОС Windows® HPC Server 2008 и Red Hat® Enterprise и поддерживает высокопроизводительные вычисления на основе графических процессоров NVIDIA® Tesla® GPU.