

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ РЕШЕНИЯ ИНЖЕНЕРНЫХ ЗАДАЧ ПРИ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ MSC.SOFTWARE И ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ ARBYTE С ГРАФИЧЕСКИМИ ПРОЦЕССОРАМИ NVIDIA

Александр Федорович Георгиев,
старший технический эксперт MSC.Software RUS
Дмитрий Анатольевич Якунин,
руководитель направления САПР ARBYTE

Введение

В последние годы достаточно активно развивается направление по использованию графических процессоров (Graphics Processing Unit - GPU) в решении инженерных задач. Однако далеко не все типы графических карт позволяют проводить вычисления с использованием GPU. Для этих целей можно использовать только профессиональные видеокарты NVIDIA® серии Quadro™ и специализированные GPU-ускорители NVIDIA® Tesla™.

Компания MSC.Software - мировой лидер в разработке программного обеспечения для решения инженерных задач в области механики, термодинамики, гидрогазодинамики и др. - разработала версии программного обеспечения (MSC Nastran 2012 и Marc 2012), которые поддерживают вычисления на GPU. В MSC Nastran 2012 решение задач с использованием GPU может быть осуществлено в следующих последовательностях решений:

- SOL 101 - решение задач линейной статики;
- SOL 103 - решение задачи поиска собственных значений;
- SOL 110 - решение задачи поиска комплексных собственных значений в модальной постановке;
- SOL 111 - решение задач частотного отклика в модальной постановке;
- SOL 112 - решение задач линейной динамики в модальной постановке;
- SOL 400 - при решении задач нелинейной и линейной статики;
- SOL 200 - решение задач оптимизации (только с использованием перечисленных последовательностей решений: SOL 101, 103, 110, 111, 112, 400).

При решении вышеперечисленных задач, GPU используются только для математических операций, связанных с факторизацией разреженных матриц. Все остальные вычислительные операции производятся центральным процессором (Central Processing Unit - CPU) компьютера.

Оценки производительности вычислительной системы с устройствами GPU с использованием технологии NVIDIA® Maximus™ производились на тестовом стенде, предоставленном компанией ARBYTE, разработчиком и поставщиком решений, продуктов и услуг в области информационных технологий.

Описание конфигурации тестового стенда и режимов проведения тестирования

Рабочая станция собрана на базе модели средней линейки WS 479, которая позволяет как работать со сложными 3D моделями и проектами, так и проводить инженерный анализ конструкций.

Характеристики тестового стенда приведены в таблице 1.

Тестирование производительности вычислительной системы производилось на следующих задачах:

1. Задача линейной статики (SOL 101): рассматривалось статическое нагружение картера коробки передач грузового автомобиля.
 2. Задача нелинейной статики (SOL 400): имитация испытаний колеса на долговечность.
 3. Задачи поиска собственных значений (SOL 103):
 - модель картера коробки передач грузового автомобиля;
 - модель коленчатого вала двигателя;
 - модель "черного" кузова легкового автомобиля.
- Решение задач производилось на следующих вариантах конфигурации оборудования:

1. Только на CPU;
2. CPU + 1 GPU NVIDIA Tesla C2075;
3. CPU + 1 GPU NVIDIA Quadro 6000;
4. CPU + 1 GPU NVIDIA Quadro 6000 + 1 GPU NVIDIA Tesla C2075;
5. CPU + 2 GPU NVIDIA Tesla C2075;
6. CPU + 1 GPU NVIDIA Quadro 6000 + 2 GPU NVIDIA Tesla C2075.

Необходимо отметить, что при установке двух GPU в тестовый стенд, каждое устройство работает на шине PCI-Express со скоростью 16x. При установке трех GPU - только одно устройство может работать со скоростью 16x, а остальные два устройства работают с пониженной скоростью шин - 8x.

Тестирование показало, что если задействовать два GPU, работающих на пониженной скорости (8x), то производительность тестового стенда понижается на 5...10 % относительно производительности этого же стенда, когда GPU работают на шинах с полной пропускной способностью (16x). Конфигурации № 2 и № 3, так же, как и конфигурации № 4 и № 5, дают сопоставимые результаты, соответственно. Поэтому в данной статье приводятся результаты тестирования для следующих конфигураций:

1. Только на CPU - на графиках обозначено как "CPU";
2. CPU + 1 GPU NVIDIA Tesla C2075 - обозначение "CPU + 1 GPU";

Таблица 1

Характеристики тестового стенда	
Модель рабочей станции	ARBYTE CAD Station WS 479
CPU	Intel Core i7 3960X, 3,30 ГГц
RAM	64 Гб DDR3 1600 МГц (PC3-12800)
GPU #1	NVIDIA Quadro 6000
GPU #2	NVIDIA Tesla C2075
GPU #3	NVIDIA Tesla C2075
Твердотельный накопитель (SSD)	60 GB
Жесткий диск (HDD)	300 GB 10K об/мин
Операционная система	Microsoft Windows 7 Профессиональная 64 bit, версия 6.1.7601 Service Pack 1
Программное обеспечение	MSC Nastran 2012

3. CPU + 1 GPU NVIDIA Quadro 6000 + 1 GPU NVIDIA Tesla C2075 - обозначение "CPU + 2 GPU";

4. CPU + 1 GPU NVIDIA Quadro 6000 + 2 GPU NVIDIA Tesla C2075 - обозначение "CPU + 3 GPU".

Также исследовано влияние режимов запуска решения задач в MSC Nastran на производительность. Рассмотрены следующие режимы запуска решений задач:

1. Без использования распараллеливания в режиме SMP (Shared Memory Parallel) - на графиках обозначено как "1 поток".

2. С распараллеливанием в режиме SMP на два потока (параметр запуска parallel = 2) - обозначение "2 потока".

3. С распараллеливанием в режиме SMP на четыре потока (параметр запуска parallel = 4) - обозначение "4 потока".

Оценка производительности вычислительной системы при решении задачи линейной статики (SOL 101) размерностью 2,2 млн степеней свободы

В качестве задачи для тестирования была взята задача анализа напряженно-деформированного состояния картера коробки передач грузового автомобиля при воздействии статической нагрузки. Модель насчитывает порядка 2,2 млн степеней свободы. Модель представлена на рис. 1.

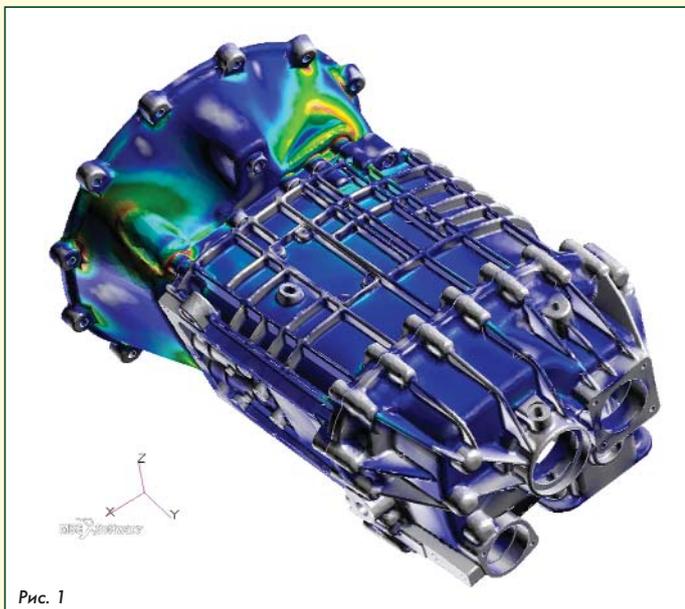


Рис. 1

Результаты тестирования (рис. 2) показывают, что при решении задачи только на CPU с использованием распараллеливания на 4 потока в SMP режиме дает до 50 % прироста производительности. Однако необходимо отметить, что в этом случае работать на компьютере (использовать другие программы) не представляется возможным из-за полной загруженности CPU.

При использовании в решении задач ресурсов GPU можно получить почти трехкратный прирост производительности (на конфигурации CPU + 1 GPU NVIDIA Quadro 6000 + 2 GPU NVIDIA Tesla

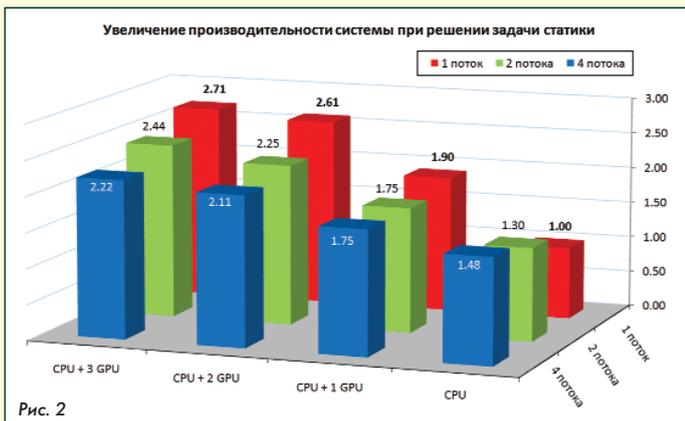


Рис. 2

C2075) и при этом загрузка CPU будет неполной, что позволит пользователю работать с другими программными продуктами - например, обрабатывать результаты решения или работать с расчетной моделью в Patran.

Использование одного GPU дало почти двухкратный прирост производительности при решении задачи статики средней размерности (2,2 млн степеней свободы).

Таким образом, если у инженера есть рабочая станция хотя бы с одним GPU, то целесообразно запускать решение задачи с использованием этого устройства, чтобы сократить вдвое время расчета и иметь возможность во время проведения расчетов эффективно работать на компьютере с другими программами.

Оценка производительности вычислительной системы при решении нелинейных задач статики (SOL 400)

Для тестирования была взята задача имитации испытаний колеса на долговечность (рис. 3). Задача решается в статической постановке. Размерность задачи составляет более 785 тыс. степеней свободы. В модели имеются контакты и большие перемещения.

Результаты тестирования с одним и двумя GPU показывают,

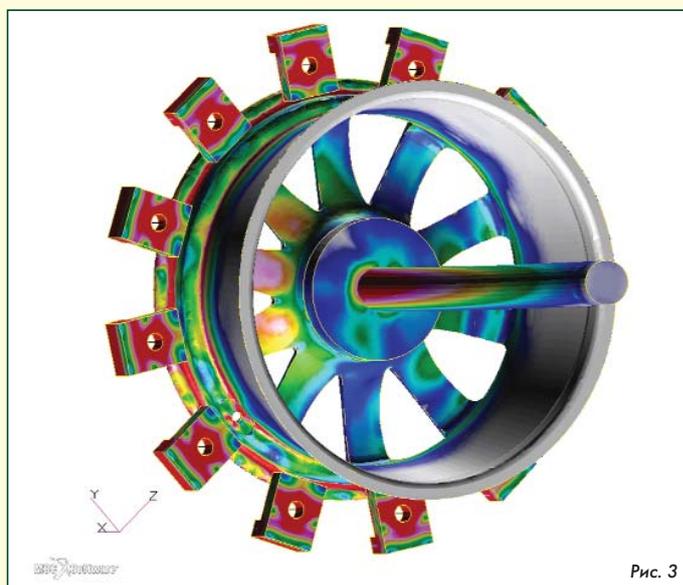


Рис. 3

что использование распараллеливания в SMP режиме при использовании GPU не является эффективным (рис. 4). Поэтому тестирование с использованием трех GPU проводилось только на CPU (без распараллеливания в SMP режиме).

В общем, характер изменения производительности вычислительной системы при решении задач нелинейной статики схож с характером изменения производительности при решении задач линейной статики: максимальное увеличение производительности (до 45 %) наблюдается при использовании трех GPU, даже использование одного GPU без распараллеливания задачи дает больший прирост производительности, чем решение задачи с распараллеливанием в SMP режиме на 4 потока.

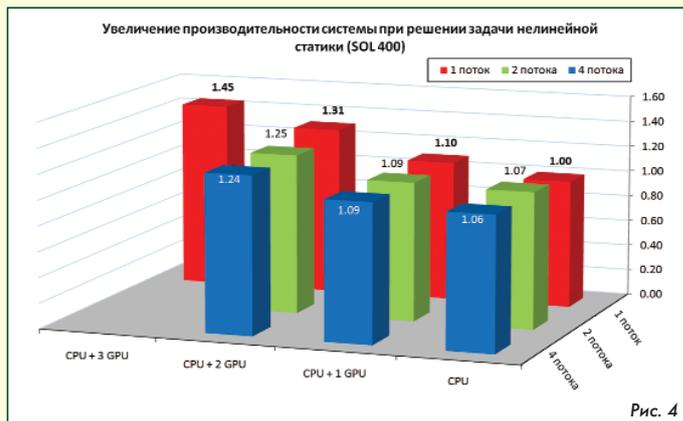
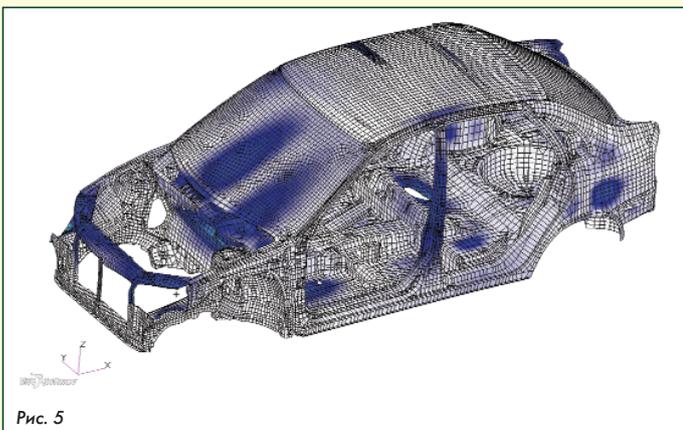


Рис. 4

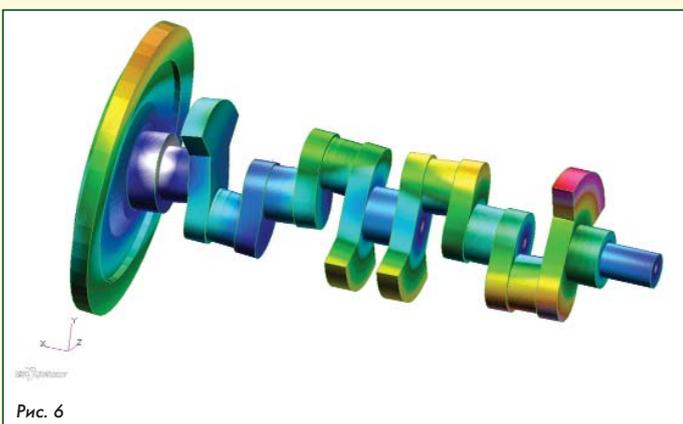
Оценка производительности вычислительной системы при решении задач поиска собственных значений (SOL 103)

Оценка производительности вычислительной системы при решении задач поиска собственных значений на GPU производилась на трех задачах:

1. Задача малой размерности. Рассмотрены собственные колебания "черного" кузова легкового автомобиля до 500 Гц. Модель насчитывает более 316 тыс. степеней свободы (рис. 5).



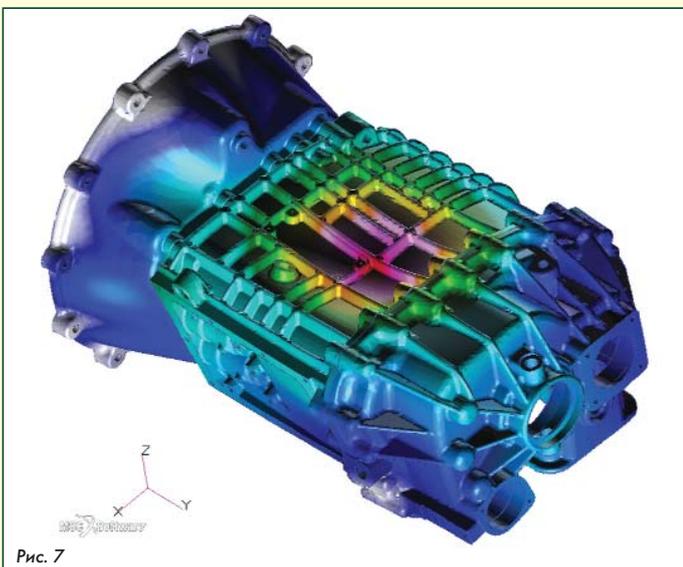
2. Задача средней размерности. Рассмотрены собственные колебания коленчатого вала двигателя легкового автомобиля. Модель насчитывает более 1,5 млн степеней свободы (рис. 6).



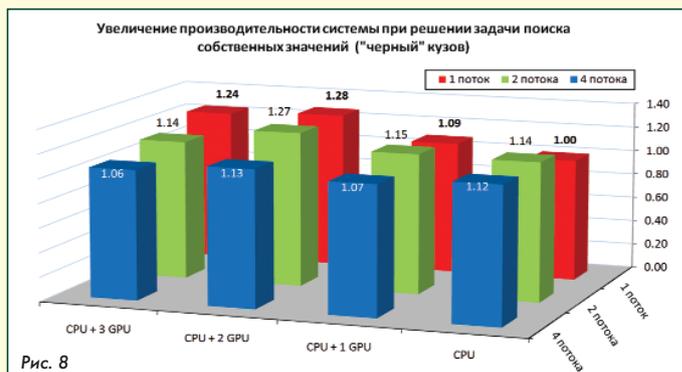
3. Задача большой размерности. Рассмотрены собственные колебания картера коробки передач грузового автомобиля. Модель насчитывает около 2,2 млн степеней свободы (рис. 7).

Результаты тестирования производительности показали:

- Использование GPU для задач малой размерности (менее

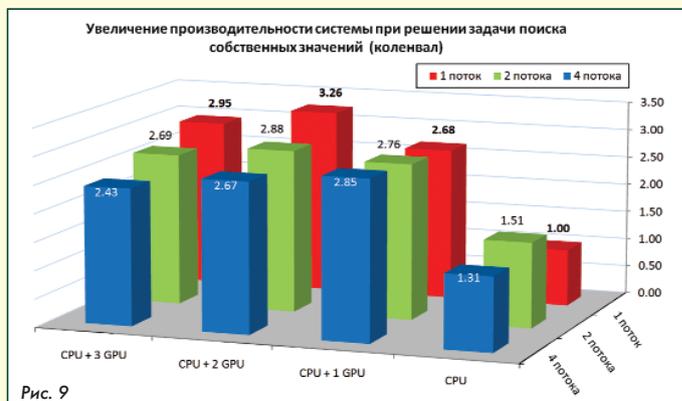


1 млн степеней свободы) неэффективно (рис. 8). Это обусловлено



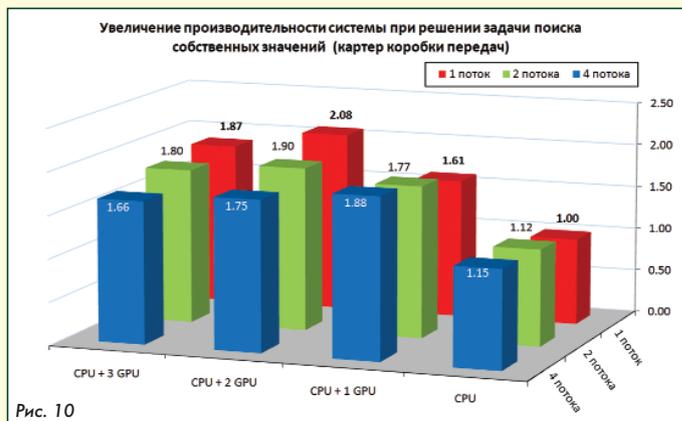
тем, что суммарное время, затраченное на факторизацию матриц, а также передачу данных из основной памяти в память GPU и обратно, сопоставимо со временем факторизации матриц с использованием CPU.

- Использование GPU для задач средней размерности (от 1 до 2 млн степеней свободы) очень эффективно. Наблюдается почти трехкратный прирост производительности при использовании одного GPU и более при использовании двух GPU (рис. 9). Некоторое



падение производительности при использовании трех GPU устройств по отношению к производительности с двумя GPU устройствами объясняется тем, что в конфигурации тестового стенда с тремя GPU два из них работали на пониженной скорости шины PCI Express (8x вместо 16x). При этом, в отличие от задач статики, количество вычислений и операций с матрицами при поиске собственных значений значительно больше.

- Для решения задач большой размерности (свыше 2 млн степеней свободы) также эффективно использовать GPU (рис. 10). При



распараллеливании задачи на четыре потока с использованием одного GPU можно получить прирост производительности почти в 2 раза. Оптимальным режимом использования рабочей станции будет решение задач с распараллеливанием на два потока, т.к. в этом случае инженер имеет возможность использовать свободные

ресурсы для запуска других программ и при этом производительность выполнения расчета повысится почти на 80 %.

При использовании двух GPU производительность системы повышается в два раза. При использовании трех GPU, как и в предыдущем случае, наблюдается небольшое падение производительности, обусловленное работой двух GPU на более низкой скорости шины.

Выводы

Результаты тестирования показывают, что использование GPU позволяет высвободить ресурсы вычислительной системы и при этом добиться значительного повышения производительности. При решении задач малой размерности, как показывает тестирование,

GPU дают небольшой прирост производительности.

Решая задачи средней и большой размерности, даже при наличии одного GPU в системе, производительность повышается почти в 2,7 раза (рис. 9).

Эффективным, с точки зрения роста производительности вычислительной системы, является использование двух GPU. В такой конфигурации производительность может увеличиваться практически в 3,3 раза (рис. 9).

Таким образом, если в компьютере имеется хоть один GPU, инженер может более эффективно решать инженерные задачи и при этом одновременно работать с другими программами, например, обрабатывать результаты, создавать новые расчетные модели, писать отчеты и т.п. 



Графическая станция USN IGLAx 1104

Рекомендуется для работы в сфере инженерных расчетов в приложениях ANSYS, MSC.Software и др.



Графический процессор
NVIDIA® QUADRO® 4000

Центральный процессор
Intel Core i5-3570K 3.4 ГГц

Жесткий диск
120 Гб SSD + 2 Тб HDD

Оперативная память 16 Гб

www.usn.ru

www.nvidia.ru

ИНФОРМАЦИЯ

12 ноября NVIDIA объявила о том, что в этом году поддержку GPU-ускорения получили еще 70 популярных приложений, общее число которых превысило 200.

Среди новых приложений можно отметить:

- ANSYS® Fluent® - позволяет инженерам создавать автомобили и самолеты с улучшенным аэродинамическим качеством, а также улучшить теплоотвод и надежность интегральных микросхем. ANSYS Fluent добавляет бета-версию нового решателя с поддержкой одного GPU к списку ведущих приложений для NVIDIA® CUDA®, таких, как ANSYS Mechanical™.

- MSC® Nastran® - используется для оптимизации уровня шума и вибраций.

- CHARMM - широко используется для изучения биологических процессов на молекулярном уровне, а GPU-ускорение позволяет лучше исследовать свойства белков, вовлеченных в процесс болезни, а также их взаимодействие с лекарствами, что позволяет создавать более эффективные лекарственные препараты.

"На начальном этапе вычисления на GPU привлекали внимание ученых, которые могли использовать CUDA для ускорения собственных приложений, используемых для проведения исследований и совершения на-

учных открытий", - отметил Эддисон Снелл (Addison Snell), исполнительный директор Intersect 360 Research. - *Сейчас мы вошли в новую эпоху, где многие коммерческие приложения получили оптимизацию под GPU, обеспечивая ускорение массовых инженерных и бизнес задач*".

Для автоматизированной разработки предлагается приложения, получившие поддержку GPU-ускорения: Abaqus/Standard, Agilent ADS и EMPro, ANSYS Mechanical, CST MWS, MSC Nastran, Marc, библиотеки решателя OpenFOAM, RADIOSS™. Для научных расчетов: AMBER, CHARMM, Chroma, FastROCS, GAMESS, GROMACS и др. 