

# Моделирование российского суперкомпьютера «Ангара» на суперкомпьютере





## 18 Моделирование российского суперкомпьютера «Ангара» на суперкомпьютере

Для оценки производительности на тестовых программах российского суперкомпьютера «Ангара» на языке Charm++ разработана параллельная потактовая имитационная модель. Модель хорошо масштабируется при использовании до 256 узлов MVS-100K, а результаты моделирования позволили перейти к аппаратной реализации в ОАО «НИЦЭВТ» мультитредового микропроцессора и маршрутизатора коммуникационной сети в составе «Ангара».

### АВТОРЫ:

Л. К. Эйсымонт — начальник отдела, канд. физ.-мат. наук, ОАО «НИЦЭВТ»,  
[e-mail: verger@nicevt.ru](mailto:verger@nicevt.ru)

А. С. Семенов — начальник сектора, ОАО «НИЦЭВТ», [e-mail: semenov@nicevt.ru](mailto:semenov@nicevt.ru),

А. А. Соколов — научный сотрудник, ОАО «НИЦЭВТ», [e-mail: sokolov@nicevt.ru](mailto:sokolov@nicevt.ru)

А. С. Фролов — начальник сектора, ОАО «НИЦЭВТ», [e-mail: frolov@nicevt.ru](mailto:frolov@nicevt.ru),

А. Б. Шворин — инженер-программист, ИПС РАН, [e-mail: shvorin@gmail.com](mailto:shvorin@gmail.com)

В ОАО «НИЦЭВТ» ведется проект создания суперкомпьютера стратегического назначения (СКСН) «Ангара» с мультитредово-потокковой архитектурой и аппаратной поддержкой глобально адресуемой памяти [1]. Этот суперкомпьютер должен на 2—3 порядка эффективнее по отношению к обычным суперкомпьютерам решать задачи с интенсивной нерегулярной работой с памятью и не уступать им при решении задач другого типа.

В процессе создания СКСН «Ангара» была разработана идеализированная архитектурная потактовая модель (ИДМ). Главная цель создания ИДМ — получение оценок производительности СКСН «Ангара» на тестовых программах, отработка созданных принципов работы и вариантов их аппаратной реализации, создание и отработка принципов построения системного программного обеспечения, получение навыков использования разных вычислительных моделей программ.

При постановке работ по ИДМ одним из основных требований являлась необходимость моделирования суперкомпьютера «Ангара» в полном объеме (до 32 тысяч вычислительных узлов), что требует значительных вычислительных ресурсов, оперативной и дисковой памяти и приводит к необходимости распараллеливания модели. В качестве языка реализации выбран Charm++ с объектноориентированной моделью вычислений и управляемой передачей сообщений. Charm++ использовался при моделировании суперкомпьютера IBM BlueGene/L [2]. Используемая реализация языка Charm++ позволяет работать с десятками тысяч объектов — моделирующих процессов, с эффективно выполняемой барьерной синхронизацией и автоматической балансировкой загрузки процессоров вычислительного кластера.

ИДМ состоит из модели коммуникационной сети, в узлах которой располагаются многоядерные мультитредово-потокковые микропроцессоры. Модель включает в себя различные варианты реализации топологии сети: многомерные торы с адаптивной бездедлоковой передачей пакетов, возможно, модифицированные в соответствии с графами Кэли, многостадийные сети Клоса. В статье рассматриваются два варианта микропроцессора — двухъядерный J7 и восьмиядерный J10.

Процессор и сеть моделируются потактово, причем их тактовые частоты могут различаться. Идеализированность модели состоит в следующем соглашении: все команды, кроме команд работы с памятью и коммуникационной сетью, выполняются за один такт, выполнение команд работы с памятью и коммуникационной сетью максимально приближено к реальности.

Модель параметризована, можно задавать характеристики разных блоков и линий связи. В настоящее время это примерно 100 параметров. Имеются богатые средства выдачи трасс и накопление статистики.

Сложности создания таких имитационных моделей известны — реальный па-

раллелизм моделирующих процессов требует принятия мер по синхронизации этих процессов во времени моделируемой системы. В случае ИДМ был изначально применен метод с потактовой барьерной синхронизацией всех моделирующих процессов. Позже для повышения эффективности была введена асинхронная схема моделирования, позволяющая экономить на барьерных синхронизациях, используя тот факт, что моделируемая задержка передачи сообщений между узлами составляет сотни тактов. В этом случае все процессоры выполняют сразу несколько тактов, их число определяется параметром и не должно превышать задержку передачи сообщения между независимо моделируемыми объектами. После этого процессоры обмениваются накопленными за это время сообщениями, таким образом экономятся коммуникации, так как выполняется несколько крупных посылок вместо сотни более мелких. Затем происходит синхронизация на глобальном барьере.

При работе с ИДМ используется разработанное базовое программное обеспечение СКСН «Ангара», включающее ассемблер, редактор связей и загрузчик. Скорость моделирования одного такта одного узла J7 при выполнении на процессоре Intel Xeon 5365 составляет около 30 мкс и 100 мкс для J10.

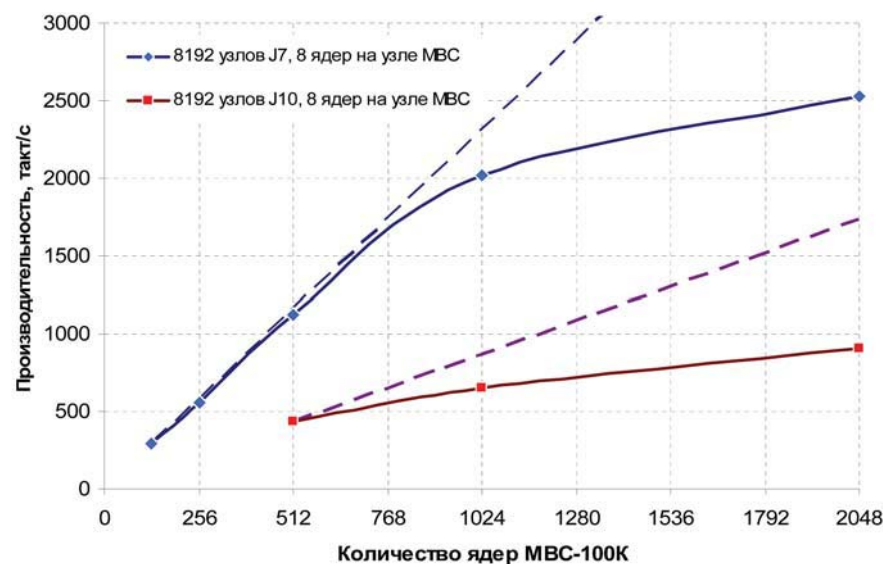


Рис. 1. Зависимость производительности имитационной модели СКСН «Ангара» от количества используемых ядер на МВС-100К

На рис. 1 изображена зависимость производительности моделирования теста RandomAccess на 8192 узлах J7 и J10 СКСН «Ангара» от количества используемых ядер на МВС-100К при использовании асинхронной схемы моделирования. Для данных конфигураций модели требования к объему оперативной памяти составляют приблизительно 39 Гбайт для J7 и 180 Гбайт для J10, поэтому модель невозможно было запустить менее чем на 16 и 64 узлах МВС соответственно для J7 и J10. Пунктирными линиями изображено линейное масштабирование производительности. Производительность для J7 вырастает в 8,75 раз при выполнении на 256 узлах МВС по сравнению с выполнением на 16 узлах, а для J10 — в 2,09 раза при переходе от 64 к 256 узлам МВС. Данный пример иллюстрирует необходимость использования суперкомпьютера для моделирования, с одной стороны, из-за требований к используемой памяти, с другой стороны — для ускорения счета.

При помощи ИДМ при выполнении на МВС-100К получены оценки производительности СКСН «Ангара» на тестах RandomAccess, умножения плотнозаполненных матриц, 1D FFT, поиска вширь в графе, умножения разреженной матрицы на вектор. На рис. 2 и 3 представлены оценки производительности микропроцессоров J7 и J10 на тестах поиска вширь в графе и умножения разреженной матрицы на вектор в сравнении с различными процессорами. Оценки доказали возможность решения главной задачи проекта — получения высокой реальной производительности при решении задач с интенсивной нерегулярной работой с памятью. Моделирование также позволило проверить решения, принятые при разработке принципов работы, алгоритмов и архитектуры маршрутизатора коммуникационной сети.

В целом все это позволило более подготовленно перейти к аппаратной реализации в ОАО «НИЦЭВТ» мультитредово-поточкового микропроцессора и маршрутизатора коммуникационной сети.

В настоящее время уже разработана микроархитектурная параллельная модель, в которой аппаратура моделируется без упрощений. Планируется разработка другого варианта ИДМ для скоростной отладки операционной системы и системы обеспечения отказоустойчивости.

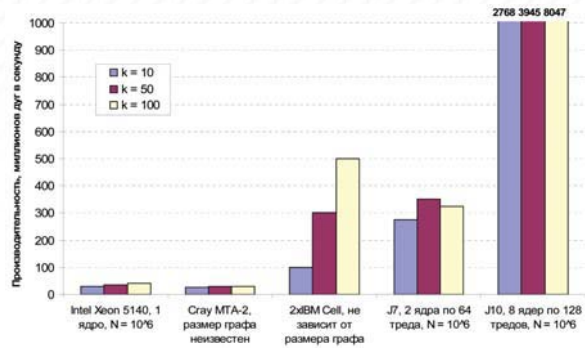


Рис. 2.

Производительность разных процессоров на тесте поиска вширь в графе в миллионах пройденных дуг в секунду,  $k$  - средняя арность графа,  $N$  – количество вершин в графе

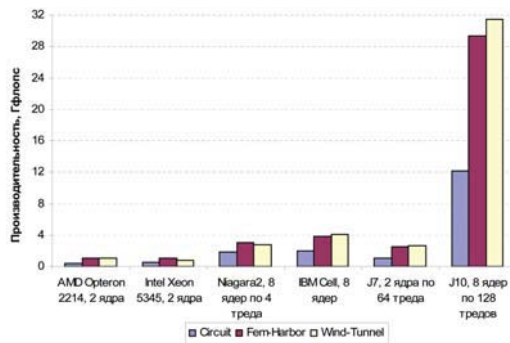


Рис. 3.

Производительность разных процессоров на задаче умножения разреженной матрицы на вектор. Circuit, Fem-Harbor, Wind-Tunnel – матрицы с разным количеством ненулевых элементов

## СПИСОК ЛИТЕРАТУРЫ

1. Слущин А., Эйсымонт Л. Российский суперкомпьютер с глобально адресуемой памятью // Открытые системы. 2007. №9. С. 42–51.
2. Heng Z., et al. Performance Modeling and Programming Environments for Petaflops Computers and the Blue Gene Machine. Parallel and Distributed Processing Symposium, International. Vol. 11. P. 197a. 18th International Parallel and Distributed Processing Symposium (IPDPS'04) – Workshop 10, 2004.