

С. В. Абламейко, С. М. Абрамов, В. В. Анищенко,
Н. Н. Парамонов

Суперкомпьютеры семейства «СКИФ»

Аннотация. Статья посвящена обзору технических характеристик и истории создания в 2000–2004 гг. первых 12 образцов высокопроизводительных систем семейства «СКИФ». Приводится дополнительная информация по теме.

Ключевые слова и фразы: Суперкомпьютерная Программа «СКИФ» Союзного государства, суперкомпьютер, кластер.

1. Введение

Суперкомпьютерная Программа Союзного государства «Разработка и освоение в серийном производстве семейства моделей высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе» (шифр программы — «СКИФ») (далее по тексту Программа) выполняется в соответствии с постановлением Исполнительного комитета Союза Беларуси и России от 22 ноября 1999 г. No 43. Сроки выполнения Программы: 2000–2004 гг (см. [1–3]).

Главной целью Программы «СКИФ» является возрождение компьютерной отрасли двух стран, промышленное производство ряда программно-совместимых моделей суперкомпьютеров с широким спектром производительности — до триллионов операций в секунду, с возможностью выбора конфигурации, оптимальной для заданного применения. В настоящее время для отработки конструкторских решений и программного обеспечения по Программе «СКИФ» выпущено двенадцать (см. Рис. 1) образцов высокопроизводительных вычислительных систем семейства «СКИФ», которые рассматриваются в последующих разделах данной статьи.

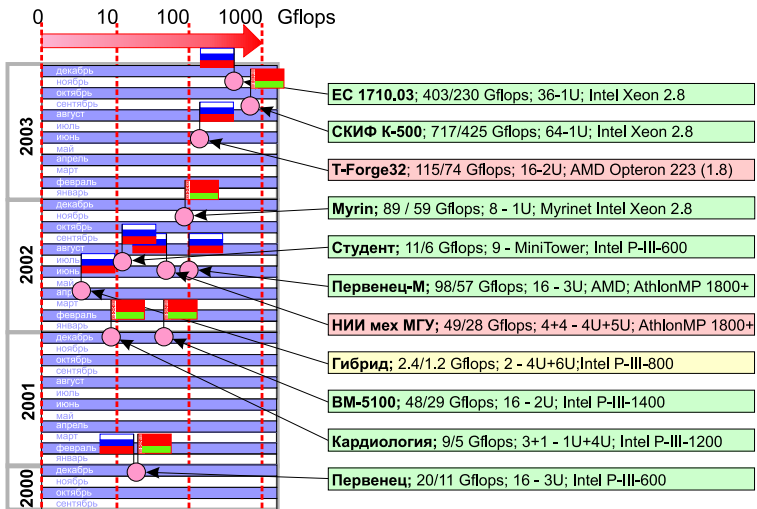


Рис. 1. Образцы высокопроизводительных вычислительных систем семейства «СКИФ»

2. Первые образцы высокопроизводительных кластерных систем семейства «СКИФ»

Фактическое финансирование Программы «СКИФ» началось в сентябре 2000 года, причем объем финансирования составил примерно всего четверть от запланированного: вместо 69 млн. руб. было выделено 17 млн. руб. В этих условиях было проведено совещание основных исполнителей Программы «СКИФ», на котором обсуждалось, каких результатов можно было бы достичь в данных условиях за первый год работы. Было решено не ограничиваться бумажными отчетами, а сделать весомый материальный результат в виде работающей установки. Такая установка была необходима еще и потому, что имелся существенный задел в разработке программного обеспечения и необходимо было его отлаживать на работоспособном образце. Было принято решение до конца года существенную часть финансовых ресурсов, выделяемых на Программу «СКИФ», потратить на создание первых образцов кластерного уровня с емейства «СКИФ». (Далее в статье первые два образца для краткости именуются «Первенцами» (см. Рис. 2).



Рис. 2. Внешний вид первых двух образцов высокопроизводительных кластерных систем семейства «СКИФ» (один установлен в Переславле-Залесском, в ИПС РАН, второй — в Минске, в ОИПИ НАН Беларуси)

В условиях чрезвычайно ограниченных временных рамок для решения поставленной задачи пришлось разрабатывать посуточный сетевой график и выдерживать его очень жестко. В эту работу, помимо основных исполнителей (ОИПИ НАН Беларуси, ИПС РАН, НИЦЭВТ, ИВВиБД) был вовлечен и Постоянный Комитет Союзного государства, от которого требовалось, чтобы финансирование Программы поступало в соответствии с этим графиком.

В результате совместных усилий были выдержаны все условия и сроки согласованного плана-графика, и к концу декабря 2000 года были сделаны два образца высокопроизводительных кластерных систем семейства «СКИФ», один из которых был установлен в Минске в НИИ ЭВМ, второй в Переславле-Залесском в ИПС РАН. Эта

Месяц и год выпуска	декабрь 2000 год
Место расположения	НИИ ЭВМ (Минск) и ИПС РАН (Переславль- Залесский)
Число вычислительных узлов/процессоров	16/32
Тип процессора	Intel Pentium III-600
Частота процессора	600 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	20(11) Gflops
Оперативная память установки	$16 \times 0.5 = 8$ GB
Дисковая память установки	$16 \times 10 = 160$ GB
Тип системной сети	2D-top 4×4 , SCI, D311/312
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	3U

ТАБЛИЦА 1. Первые образцы, технические характеристики

работа запомнилась всем исполнителям как яркий пример сотрудничества, пример четкой кооперации между российскими и белорусскими партнерами. Первые этапы работы — разработка эскизной конструкторской документации — были выполнены в ОАО «НИЦЭВТ» в Москве. Затем на основе этой документации в Минске в НИИ ЭВМ в рекордные сроки была разработана рабочая конструкторская документация и конструктивы. На своем опытном производстве НИИ ЭВМ выпустил 8 стоек и 32 корпуса для узлов вычислительной системы. Затем конструктивы были доставлены в Москву в НИЦЭВТ, где уже был завершён подбор и закупка комплектующих. Здесь была произведена сборка всей аппаратной части и первичное тестирование установок «СКИФ». Было изготовлено два опытных образца. Один из изготовленных образцов был после завершения этих работ отправлен в Минск, в НИИ ЭВМ, а второй — в Переславль-Залесский, в ИПС РАН, где производилась установка базового комплекта программного обеспечения кластерного уровня, окончательная наладка, тестирование и замеры показателей производительности и других технических характеристик. Кроме того, на первые образцы семейства «СКИФ»

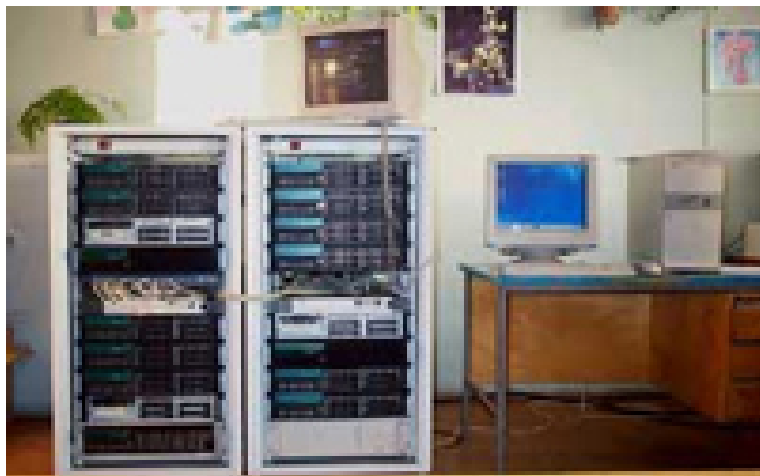


Рис. 3. Кластерная установка VM-5100 семейства «СКИФ»

была проведена установка первой прикладной системы, разработанной в Санкт-Петербурге в Институте высокопроизводительных вычислений и информационных систем (ИВВ ИС). Эта прикладная система предназначена для проектирования химических реакторов.

К 31 декабря 2000 года был выполнен весь комплекс работ и два опытных образца семейства «СКИФ» с полностью установленным программным обеспечением были готовы к работе.

В мае 2001 года была проведена широкая презентация этих образцов с привлечением многочисленных специалистов из Беларуси и России. Так получилось, что уровень пиковой производительности этих образцов — 20 Gflops (см. Таблицу 1) — соответствовал уровню эмбарго на ввоз в Россию и Беларусь высокопроизводительной вычислительной техники, ввоз такой техники требовал оформления особого разрешения.

Месяц и год выпуска	декабрь 2001 год
Место расположения	НИИ ЭВМ (Минск)
Число вычислительных узлов/процессоров	16/32
Тип процессора	Intel Pentium III-1400 MHz
Частота процессора	1400 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	48(29) Gflops
Оперативная память установки	$16 \times 1 = 16$ GB
Дисковая память установки	$16 \times 18 = 288$ GB
Тип системной сети	2D-top 4×4 , SCI, D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	Установлена лицензионная версия пакета LS-DYNA, срок лицензии: 16.01.2004 г.

ТАБЛИЦА 2. Технические характеристики ВМ-5100

3. Высокопроизводительная кластерная система ВМ-5100 семейства «СКИФ»

Второй опытный образец семейства «СКИФ» — ВМ-5100 — был выпущен белорусской стороной в 2001 году (см. Рис. 3). Его отличительными чертами являются более высокопроизводительные (по отношению к первым образцам) процессоры Intel Pentium III-1400 MHz и более компактные конструктивы: 2U вместо 3U. Благодаря процессорам Intel Pentium III-1400 MHz на этом образце были достигнуты более высокие пиковая и реальная производительность: 48 Gflops и 29 Gflops, соответственно (см. Таблицу 2).

Именно этот образец вместе с образцом «Первенец» являлись базой, на которой проводились государственные испытания в феврале 2002 года.



Рис. 4. Аппаратно-программный кардиологический комплекс

4. Аппаратно-программный кардиологический комплекс

В 2001 году в рамках Программы были выполнены работы по созданию аппаратно-программного кардиологического комплекса (АПКК) на основе технических решений «СКИФ». Комплекс предназначен для исследования микроциркуляторного звена сердечно-сосудистой системы методом биомикроскопии.

Первый показ общественности этой разработки (см. Рис. 4 и Таблицу 3) был проведен в рамках государственных испытаний в феврале 2002 года. Кардиологический комплекс является ярким примером создания на базе технических решений программы «СКИФ» законченной прикладной системы. В системе была решена задача совместимости вычислительной и медицинской аппаратуры. Это позволило в режиме реального времени анализировать состояние больных, определять точный диагноз и оптимальный путь лечения. Кардиологическая установка всегда сильно привлекала к себе внимание всех специалистов, поскольку на ней наиболее ярко проявлялись особенности Программы «СКИФ», а именно ориентированность на законченные прикладные системы. На обывательском уровне АПКК это уже не суперкомпьютер, не кластер, а «черный ящик», который выполняет функции кардиологического комплекса (см. Таб. 3). В АПКК используется оригинальная методика диагностики кардиологических заболеваний, разработанная белорусскими коллегами (ОИПИ НАН

Месяц и год выпуска	декабрь 2001 год
Место расположения	ОИПИ НАН Беларуси (Минск)
Число вычислительных узлов/процессоров	$4/3 \times 2 + 1$
Тип процессора	6 Intel Pentium III-1200 MHz + 1 Intel Pentium IV-1800 MHz
Частота процессора	1266 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	9.4(5) Gflops
Оперативная память установки	$4 \times 1 = 4$ GB
Дисковая память установки	$7 \times 18 = 126$ GB
Тип системной и управляющей сети	Fast Ethernet
Конструктив узла (форм-фактор)	3+1-1U+4U

ТАБЛИЦА 3. Технические характеристики АПКК

совместно с Республиканским кардиологическим центром Беларуси). Эта методика защищена патентами. Она основана на том, что снимается и обрабатывается ряд видеоизображений капилляров на сетчатке глаза обследуемого. На такую съемку по физиологическим ограничениям отводится очень малое время, потому что глаз начнет слезиться. Кроме того, жесткое ограничение времени на обследование и постановку диагноза определяется еще и тем, что кардиологический комплекс предполагается использовать во время массовых профилактических осмотров населения.

Российская сторона тоже участвует в этой разработке. В ИВВИС в Санкт-Петербурге разрабатывается программное обеспечение кластерного уровня, которое реализует экспертную систему, позволяющую кардиологу поставить обоснованный диагноз с использованием технологии искусственного интеллекта.

5. Экспериментальная гибридная установка семейства «СКИФ»

Первый показ экспериментального образца гибридной установки семейства «СКИФ» (условное название «Гибрид» (см. Рис. 5) был

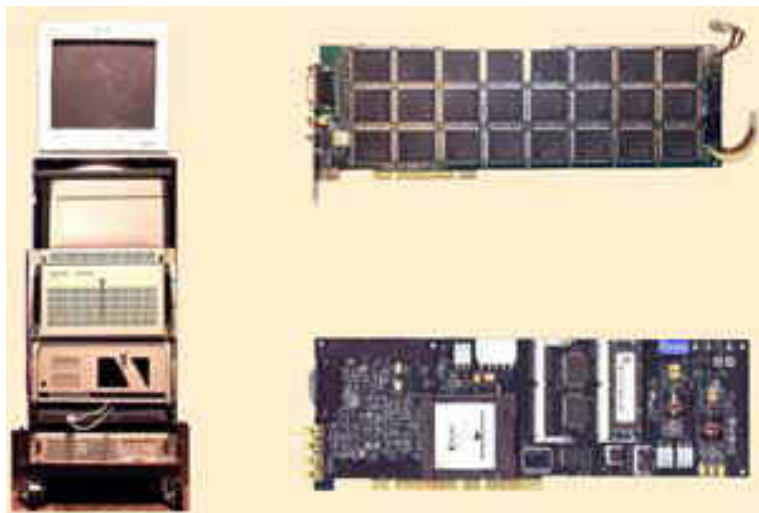


Рис. 5. Экспериментальная гибридная установка семейства «СКИФ» (снизу — управляющая плата ОВС, сверху — плата макета ОВС)

проведен в апреле 2002 года. Этот образец имеет двухуровневую архитектуру: кластерный уровень (КУ) и уровень однородной вычислительной среды (ОВС). Установка «Гибрид» используется разработчиками для отладки аппаратных и программных средств взаимодействия КУ и ОВС.

Во время первой презентации установки «Гибрид» было показано решение задачи обработки изображений. На Т-системе была реализована программа, которая получала цепочку кадров видеоизображений, содержавших шумы. Т-система разбивала видеоряд на кадры, каждый из которых отдавался для обработки в ОВС. На уровне ОВС производилась фильтрация шумов в кадре и результат возвращался в Т-систему. Пользователю показывались два окна в системе X Window. В режиме реального времени в одном окне Т-система предъявляла видеоизображение до обработки, а в другом — изображение после очистки от шумов.

Месяц и год выпуска	апрель 2002 год
Место расположения	Суперкомпьютерные системы (СКС), Москва
Число вычислительных узлов/процессоров	2/3
Тип процессора	Intel Pentium III-800 MHz
Частота процессора	800 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	2.4(1.2) Gflops
Оперативная память установки	$2 \times 0.5 = 1$ GB
Дисковая память установки	$2 \times 80 = 160$ GB
Тип системной сети	1×2 1D-топ, SCI D320
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	4U + 6U
Дополнительно	Подключена экспериментальная реализация модуля ОВС на Xilinx, 216 эл. процессоров, 10 MHz

ТАБЛИЦА 4. Экспериментальная гибридная установка семейства «СКИФ», технические характеристики

В рамках создания данной установки программные мероприятия, связанные с разработкой и производством средств ОВС, как аппаратной, так и программной части, выполняло предприятие «Суперкомпьютерные системы» (СКС). Комплекс работ по обеспечению взаимодействия КУ и ОВС выполнялся СКС совместно с Институтом программных систем РАН (ИПС РАН). В настоящее время установка «Гибрид» используется для отладки уже не экспериментальной версии модуля ОВС, а опытных образцов ОВС, выпускаемых в Минске (НИИ ЭВМ и «Интеграл») (См. Таб. 4).

6. «Первенец-М»: модернизация первого образца кластерной установки семейства «СКИФ»

В марте 2002 года в Москве открылось представительство корпорации AMD. К этому же времени у разработчиков Программы «СКИФ» назрела необходимость опробовать на кластерном уровне



Рис. 6. Кластерная установка «Первенец-М», ИПС РАН, Переславль-Залесский, 2002 г.

суперкомпьютеров семейства «СКИФ» не только решения от фирмы Intel, но и решения от AMD. У сотрудников ИПС РАН установились самые тесные отношения партнерства с сотрудниками московского представительства AMD. В результате от AMD были получены несколько комплектов процессоров, памяти и материнских плат. В то же время НИЦЭВТ предоставил ИПС РАН адаптеры SCI D335. В кратчайшие сроки, буквально за трое суток, сотрудниками ИПС РАН были собраны 12 вариантов пар вычислительных узлов, связанных сетевыми адаптерами SCI D335. На полученных вычислительных узлах были выполнены замеры производительности с целью поиска наиболее оптимальных аппаратных средств для реализации вычислительных узлов на базе существовавших на тот момент решений AMD. По результатам испытаний было принято решение о модернизации кластера «Первенец» и перевода его на новую элементную базу.

Модернизация была выполнена в июле 2002 года. В модернизации «Первенца» (см. Рис. 6 и Таблицу 5) оказала помощь московская

Месяц и год выпуска	июль 2002 год
Место расположения	ИПС РАН (Переславль-Залесский)
Число вычислительных узлов/процессоров	16/32
Тип процессора	AMD AthlonMP 1800+
Частота процессора	1533 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	98(57) Gflops
Оперативная память установки	$2 \times 1 = 16$ GB
Дисковая память установки	$16 \times 40 = 640$ GB
Тип системной сети	4×4 2D-топ, SCI D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	3U
Дополнительно	Установлена лицензионная версия пакета STAR-CD, срок лицензии: 07.07.2004 г.

ТАБЛИЦА 5. Основные технические характеристики вычислительной установки «Первенец-М»

фирма «Storus», которая предоставила платы SCI D335 с целью замены старых, использовавшихся ранее плат SCI D311/312. Старые платы были оставлены ИПС РАН на весьма льготных условиях. Впоследствии они были использованы для сборки кластера с условным названием «Студент». В результате на реализацию каждого узла модернизированного кластера «Первенец-М» было затрачено не более 2500 долларов. Таким образом, при минимальных затратах удалось улучшить основные технические характеристики системы в два-три раза, а производительность — почти в пять раз. В настоящее время установка «Первенец-М» является основной (по мощности) вычислительной мощностью в Переславле-Залесском. Она доступна по сети, и 101 пользователь — потребитель высокопроизводительных вычислений — имеет доступ к ней и решает свои задачи на этой установке.



Рис. 7. Внешний вид вспомогательной кластерной установки «Студент»

7. «Студент»: вспомогательный кластер семейства «СКИФ»

Из электронных компонент, высвободившихся после модернизации установки «Первенец», был собран девятиузловой вспомогательный кластер (см. Рис. 7 и Таблицу 6). Он выполнен в конструктивах MiniTower на стойке, которая была сделана в мастерских ИПС РАН. Условное название «Студент», которое получила эта установка, не случайно. На ней действительно работает много студентов и из МГУ, и из университета города Переславля. Кроме того установка «Студент» используется для отладки программного обеспечения, прежде чем оно устанавливается на основной кластер ИПС РАН — «Первенец-М».

Месяц и год выпуска	июль 2002 год
Место расположения	ИПС РАН (Переславль-Залесский)
Число вычислительных узлов/процессоров	9/18
Тип процессора	Intel Pentium III-600 MHz
Частота процессора	600 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	11(6) Gflops
Оперативная память установки	$16 \times 0.5 = 4.5$ GB
Дисковая память установки	90 Gb
Тип системной сети	3 × 3 2D-топ, SCI SCI D311/312
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	MiniTower

ТАБЛИЦА 6. Кластерная установка «Студент», технические характеристики

8. Кластерная установка семейства «СКИФ» в НИИ механики МГУ и установка «Myrin»

Закупка аппаратных средств для кластера НИИ механики МГУ производилась в то же время, когда проводилась модернизация установки «Первенец». Поэтому по аппаратным решениям этот кластер является близким аналогом «Первенца-М». Отличительной особенностью этого образца является то, что его реализация была выполнена не за счет средств Программы «СКИФ», а целиком за счет средств НИИ механики МГУ (см. Рис. 8 и Таблицу 7).

В том же 2002 г. для изучения технологии Myrinet в Минске (ОИ-ПИ НАН Беларуси и НИИ ЭВМ) был разработан кластер с условным названием «Myrin» (см. Рисунок 9 и Таблицу 8).

9. «T-Forge 32»: первый в СНГ кластер с процессорами AMD Opteron

Сотрудничество с московским представительством корпорации AMD, основы которого были заложены в 2002 году, было продолжено и в 2003 году. По рекомендации AMD в работах ИПС РАН появился



Рис. 8. Внешний вид кластерной установки в НИИ механики МГУ



Рис. 9. Внешний вид установки «Muglin»

Месяц и год выпуска	июль 2002 год
Место расположения	НИИ механики МГУ (Москва)
Число вычислительных узлов/процессоров	8/16
Тип процессора	AMD AthlonMP 1800+
Частота процессора	1533 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	49(28) Gflops
Оперативная память установки	$8 \times 1 = 8$ GB
Дисковая память установки	$8 \times 80 = 640$ GB
Тип системной сети	2×4 2D-топ, SCI D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	$4 \times 4U + 4 \times 5U$
Дополнительно	Кластер оснащен сервисной сетью, разработанной в ИПС РАН

ТАБЛИЦА 7. Установка семейства «СКИФ» в НИИ механики МГУ, технические характеристики

Месяц и год выпуска	ноябрь, 2002 год
Место расположения	ОИПИ НАН Беларуси
Число вычислительных узлов/процессоров	8/16
Тип процессора	Intel Xeon 2.8 GHz
Частота процессора	2.8 GHz
Предельная пиковая (реальная на задаче Linpack) производительность	89(59) Gflops
Оперативная память установки	$8 \times 1 = 8$ GB
Дисковая память установки	$8 \times 40 = 320$ GB
Тип системной сети	Myrinet
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U

ТАБЛИЦА 8. Технические характеристики установки «Myrin»

новый партнер — российская компания «Т-Платформы». ¹ В 2003 году корпорацией AMD был выпущен новый процессор AMD Opteron модели 244, основанный на технологии AMD64. AMD64 — это наименование нового класса компьютерных технологий, предусматривающих расширение стандартной архитектуры набора команд x86 для поддержки как 32-разрядных, так и 64-разрядных платформ. AMD64 — новаторская разработка корпорации AMD, призванная обеспечить полную совместимость существующих 8-, 16- и 32-разрядных решений x86 с высокопроизводительной 64-разрядной средой.

Официально процессор AMD Opteron модели 244 был представлен миру 22 апреля 2003 года. Но уже 5 апреля московское представительство фирмы AMD и компания «Т-Платформы» предоставили для совместной проработки в ИПС РАН первые узлы на процессорах AMD Opteron 244. В ИПС сразу же начались серьезные исследования этой архитектуры, работы по адаптации программного обеспечения для нее. Эти работы потребовали контактов с разработчиком плат SCI D335 компанией Dolphin, поскольку совместная работа плат D335 и платформ с процессорами AMD Opteron оказалось совсем не простым делом.

Уже 26 июня компания «Т-Платформы» с участием ИПС РАН завершили сборку и провели презентацию первого в России кластера на базе процессоров AMD Opteron 244 — «Т-Forge 32» (см. Рис. 10 и Таблицу 9). К этому времени не удалось решить все проблемы, связанные с использованием плат SCI D335, поэтому установка была реализована с системной сетью на базе плат GB Ethernet. Но даже на такой системной сети кластер показал весьма высокие технические характеристики: высокую пиковую производительность и очень хороший КПД (отношение реальной Linpack-производительности к пиковой производительности) — 64%. Более того, в рамках одного узла процессоры Opteron показали еще лучшее соотношение между пиковой и реальной производительностью на задаче Linpack — 80%. Создатели кластера «Т-Forge 32» уверены, что когда удастся решить проблемы системной сети и построить ее на платах D335, на этой установке будут достигнуты еще более привлекательные показатели реальной (Linpack) производительности.

¹Буква «Т» в названии этой компании связана с фирмой «TAYAN», разработчиком и производителем серверных материнских плат, которые широко используются в изделиях «Т-платформы».

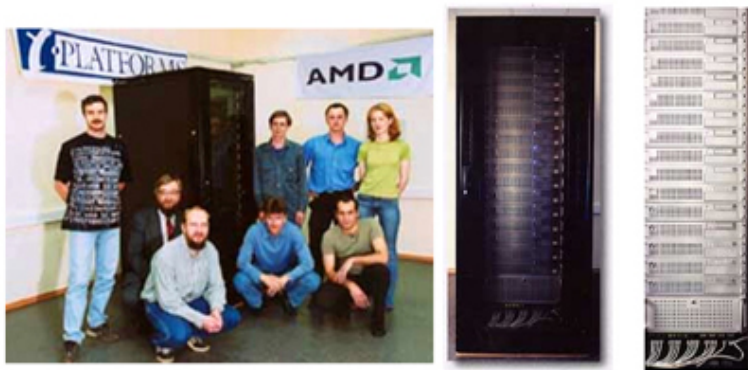


Рис. 10. «Т-Forge 32» на базе процессоров AMD Opteron 244, Москва, компания «Т-Платформы», июнь 2003 г.

Впервые на кластере «Т-Forge 32» была применена технологическая новинка — кластер оснащен сервисной сетью, разработанной и изготовленной в ИПС РАН.

Опыт работы с процессорами Opteron показал, что это действительно очень красивое новое решение, новое слово в компьютерной отрасли, поскольку процессор Opteron предоставляет пользователям реализацию мягкого перехода от 32-разрядных приложений к 64-разрядным приложениям. Процессор поддерживает и 32-разрядный режим работы, и 64-разрядный режим работы, причем 32-разрядный режим поддержан очень эффективно — все 32-разрядные приложения работают с отличной скоростью. Эта работа была важна тем, что впервые в кластерах в России применялись 64-разрядные архитектуры.

Кроме того, что применение 64-разрядные архитектуры дает повышение производительности, существует еще такой аспект высокопроизводительных вычислений, как потребность в больших объемах памяти. Как правило, большие приложения требуют больших объемов памяти и следовательно больших объемов адресного пространства, которые могут обеспечивать только 64-разрядный режим адресации. Только этого обстоятельства достаточно, чтобы сказать, что переход на 64-разрядные архитектуры предreshен.

Месяц и год выпуска	июнь 2003 год
Место расположения	Центр кластерных технологий (ЦКТ), Москва
Число вычислительных узлов/процессоров	16/32
Тип процессора	AMD Opteron 244
Частота процессора	1800 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	115.2(73.8) Gflops
Оперативная память установки	$16 \times 2 = 32$ GB
Дисковая память установки	$16 \times 60 = 960$ GB
Тип системной сети	GB Ethernet
Тип управляющей (вспомогательной) сети	GB Ethernet
Конструктив узла (форм-фактор)	2U
Дополнительно	Linpack 64% (размерность 110,000). Кластер оснащен сервисной сетью, разработанной в ИПС РАН

ТАБЛИЦА 9. Технические характеристики кластера «Т-Forge 32»

Интересно отметить, что создание кластера «Т-Forge 32» — еще один (как это было и с кластером НИИ механики МГУ) пример того, что установка, имеющая важное значение для Программы «СКИФ», была реализована без привлечения средств финансирования Программы «СКИФ». Разработка была полностью профинансирована компанией «Т-платформы».

На базе кластера «Т-Forge 32» компанией «Т-Платформы» совместно с ИПС РАН, НИВЦ МГУ и при технологической поддержке московского представительства корпорации AMD был создан Центр Кластерных Технологий. Основная задача Центра — обеспечение потенциального заказчика в режиме работы «в одно окно» (т.е. в одном месте) всеми необходимыми услугами в области высокопроизводительных вычислений. Заказчик может в Центре выбрать наиболее подходящую для своих задач архитектуру — тип процессора, объем памяти, тип системной сети (SCI, Mirynet, InfiniBand, GB Ethernet и др.), тип вспомогательной сети. Кроме того, в Центре реализуется

политика, при которой заказчик перед покупкой может попробовать, как его задачи решаются на той или другой архитектуре кластера. Заказчик может решить вопрос приобретения необходимого ему кластера, а также запросить установку программного обеспечения на него, установку и наладку вычислительной системы на своей площадке. Заказчик может провести в Центре или на своей площадке подготовку и обучение персонала.

10. Суперкомпьютерная установка «СКИФ К-500»

При планировании работ в 2003 году было принято решение о выпуске первого образца терафлопного диапазона — не менее 0,5 Tflops, который располагался бы в Минске. Данная установка получила название «СКИФ К-500». Окончательная проработка архитектурных особенностей этого образца была завершена в мае 2003 года. Основными исполнителями работ по созданию образца были:

- ОИПИ НАН Беларуси (Минск) — основной заказчик, разработчик технических требований и архитектуры, исполнитель окончательной сборки и наладки установки и место ее размещения;
- УП «НИИ ЭВМ» (Минск) — разработка основных конструктивов, сборка комплекса;
- ИПС РАН (Переславль-Залесский) — установка программного обеспечения;
- Компания «Т-Платформы» (Москва) — разработка и изготовление вычислительных узлов.

Проработка архитектурных особенностей «СКИФ К-500» была завершена в мае 2003 года, а уже в сентябре в Москве были изготовлены 64 узла установки «СКИФ К-500» и выполнена сборка технологической конфигурации (см. Рис. 11).

Технологическая конфигурация установки «СКИФ К-500» была собрана для демонстрации работоспособности вычислительных узлов в комплексе, а также для установки программного обеспечения кластерного уровня.

После испытаний технологической установки она была разобрана и все 64 узла были отправлены в Минск, где была собрана окончательная конфигурация в конструктивах, разработанных УП «НИИ ЭВМ» (см. Рис. 12 и Таблицу 10). Были выполнены замеры ее производительности. Реальная производительность составила 425.2 Gflops

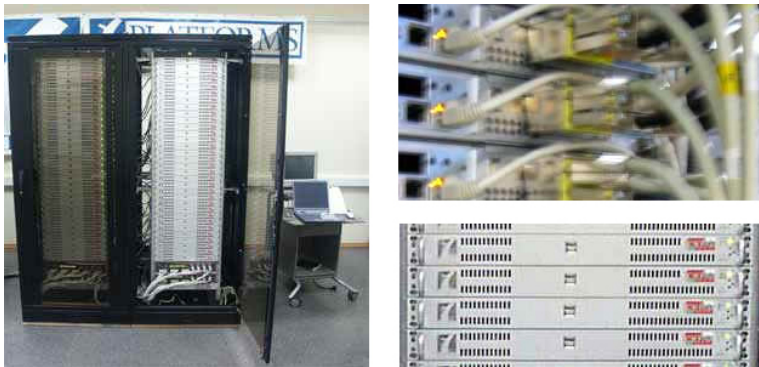


Рис. 11. «СКИФ К-500»: технологическая конфигурация, Москва, «Т-Платформы», сентябрь 2003 года

(на задаче LInpack), пиковая — 716.8 Gflops, КПД — 59%. Помимо высокой производительности следует отметить очень хорошее соотношение производительность/цена: стоимость 1 Тфлор составляет меньше \$1 000 000.

Также отметим еще две технологические новинки. Во-первых, в установке «СКИФ К-500» использованы конструктивы с форм-фактором 1U, что позволило сделать ее очень компактной. Во-вторых, впервые в России и Беларуси было объединено такое большое количество вычислительных узлов (64 узла) системной сетью SCI, причем использовалась топология системной сети SCI 3-х мерный тор (SCI 3D-тор) с адаптерами D336, предназначенными для установки в корпусе с форм-фактором 1U.

Учитывая высокие показатели установки «СКИФ К-500», в конце сентября 2003 года была оформлена заявка на включение суперкомпьютера «СКИФ К-500» в мировой список пятисот наиболее высокопроизводительных вычислительных систем (TOP-500).

16 ноября 2003 года суперкомпьютер «СКИФ К-500» вошел в очередной 22-ой выпуск списка TOP-500 и стал в нем первым суперкомпьютером, разработанным и изготовленным по программе Союзного государства.



Рис. 12. «СКИФ К-500» в окончательной конфигурации, ОИПИ НАН Беларуси, Минск, октябрь 2003 года

В 22-ой выпуск списка TOP-500 включены суперкомпьютеры из 38 стран, но только 14 из них представлены собственными разработками (производителями): Австрия, Беларусь, Германия, Индия, Канада, Китай, Польша, Российская Федерация, США, Тайвань, Швейцария, Швеция, Южная Корея и Япония.

Это замечательный результат программы «СКИФ», тем более что до сего дня в мире было только полтора десятка стран-производителей суперкомпьютеров, чьи изделия удостоены чести быть в списке TOP-500. Теперь к этому элитному клубу принадлежит Беларусь и Союзное государство.

11. Суперкомпьютерная кластерная установка ЕС1710.03

В 2003 году в рамках Программы «СКИФ» было запланировано создать российский образец высокой производительности (около 400

Месяц и год выпуска	сентябрь 2003 год
Место расположения	ОИПИ НАН Беларуси (Минск)
Число вычислительных узлов/процессоров	64/128
Тип процессора	Intel Xeon 2.8 GHz
Частота процессора	2800 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	716.8 (425.2) Gflops
Оперативная память установки	$64 \times 2 = 128$ GB
Дисковая память установки	$64 \times 60 = 3840$ GB
Тип системной сети	3D-top, SCI, D336
Тип управляющей (вспомогательной) сети	GB Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	Linpack: размерность $N = 123\,500$, $N_{\frac{1}{2}} = 25\,000$, КПД $= 425.2/716.8 = 59\%$. «СКИФ К-500» оснащен сервисной сетью, разработанной в ИПС РАН и серийно выпускаемой в УП «НИИ ЭВМ»

ТАБЛИЦА 10. Технические характеристики «СКИФ-К-500»

Gflops), который получил специальное обозначение — EC1710.03 (см. Рис. 13 и Таблицу 11).

Этот образец создавался для отработки типового решения для установок с высокой производительностью. К его сильным сторонам относятся весьма высокая производительность (пиковая производительность 0,43 Tflops), компактность (образец собран в одной стойке, форм-фактор 1U) и, самое главное, системная сеть в установке EC1710.03 выполнена на платах, которые выпускает ОАО «НИЦЭВТ» — SCI N335, аналоги плат SCI N335 компании Dolphin.

Таким образом, произошла частичная замена импортных комплектующих на комплектующие отечественного производства. В 2003 году ОАО НИЦЭВТ была выпущена установочная партия адаптеров



Рис. 13. Суперкомпьютерная установка ЕС1710.03, НИЦЭВТ, Москва, ноябрь 2003 г.



Рис. 14. Адаптер системной сети SCI N335, выпускаемый ОАО НИЦЭВТ

N335 (см. Рис. 14) — 36 плат для 36 узлов вычислительной системы ЕС1710.03.

Месяц и год выпуска	октябрь 2003 год
Место расположения	НИЦЭВТ, Москва
Число вычислительных узлов/процессоров	36/72
Тип процессора	Intel Xeon 2.8 GHz
Частота процессора	2800 MHz
Предельная пиковая (реальная на задаче Linpack) производительность	403(230) Gflops
Оперативная память установки	$36 \times 2 = 72$ GB
Дисковая память установки	$36 \times 60 + 480 = 2640$ GB
Тип системной сети	6×6 2D-top, SCI, N335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	В составе NAS 480 GB

ТАБЛИЦА 11. Технические характеристики ЕС1710.03

Особенностью установки ЕС1710.03 является также то, что в ее состав включено сетевое устройство хранения (Network Attached Storage — NAS) на 480 GB. Наличие такого устройства бывает очень важно для многих приложений суперкомпьютеров.

12. Заключение

Двенадцать разработанных высокопроизводительных вычислительных установок и суперкомпьютеров семейства «СКИФ» в 2000–2004 гг. использовались как для отладки программного обеспечения кластерного уровня, так и для реальных вычислений в интересах предприятий и учреждений России и Беларуси. Эти установки послужили материальной базой для успешного выполнения Программы в 2000–2004 гг. и получения всех основных результатов по Программе:

- (1) Разработана конструкторская документация (КД) Ряда 1 семейства «СКИФ», проведены приемочные (государственные) испытания. По результатам государственных испытаний конструкторской документации присвоена литера «О₁».
- (2) Разработано базовое программное обеспечение (ПО) кластерного уровня (КУ) суперкомпьютеров «СКИФ» Ряда 1, проведены приемочные (государственные) испытания. По

результатам государственных испытаний программным системам ПО КУ «СКИФ» присвоена литера «О₁».

- (3) Разработан ряд прикладных систем для суперкомпьютеров семейства «СКИФ», часть из которых прошла государственные испытания и соответствующей программной документации присвоена литера «О₁».
- (4) В ОАО «НИЦЭВТ» подготовлена необходимая производственная база, проведена разработка КД и освоены в производстве адаптеры (N330, N337, N335) системной сети SCI, которые являются полными функциональными аналогами адаптеров SCI компании Dolphin (D330, D337, D335).
- (5) Предприятием «Суперкомпьютерные системы» разработаны аппаратные и программные средств уровня ОВС, экспериментальный макет ОВС. Первые опытные образцы должны быть разработаны в 2003 году. Приемочные (государственные) испытания средств уровня ОВС предусмотрены в 2004 году.

Помимо этих материальных результатов Программы отметим самый важный технологический и организационный результат: *была создана команда, способная решать самые сложные технические и научные задачи*, что было доказано разработкой суперкомпьютера «СКИФ К-500». В дальнейшем, в 2004 году, в рамках завершения Программы и в последующие годы, в рамках Программы «СКИФ-2» (формирование которой начато), этой командой уже запланирована разработка новых, более мощных образцов суперкомпьютеров семейства «СКИФ».

Благодарности. Авторы благодарны Г. В. Гузиловой за набор текста и М. Г. Химшиашвили за подготовку окончательной версии статьи в системе $\text{\LaTeX} 2_{\epsilon}$.

Список литературы

- [1] Абрамов С. М., Анищенко В. В., Парамонов Н. Н., Чиж О. П. Разработка и опыт эксплуатации суперкомпьютеров семейства «СКИФ»: Материалы I международной конференции «Информационные системы и технологии» (IST'2002). — Минск, 2002, с. 115–117, 5–8 ноября 2002. (Russian) ↑1

- [2] Абрамов С. М., Адамович А. И., Коваленко М. Р., Слепухин А. Ф., Парамонов Н. Н. Кластерные системы семейства суперкомпьютеров «СКИФ»: Научный сервис в сети Интернет: Труды Всероссийской научной конференции. — М.: Изд-во МГУ, 2003, с. 147–151, Новороссийск, 22–27 сентября 2003 г. (Russian) ↑
- [3] Абламейко С. В., Абрамов С. М. Основные результаты суперкомпьютерной программы «СКИФ» Союзного государства: АКИИ'03: Третий расширенный семинар «Использование методов искусственного интеллекта и высокопроизводительных вычислениях и в аэрокосмических исследованиях», Труды семинара ИПС РАН, Переславль-Залесский, ноябрь 2003 г. — М.: Физматлит, 2003, с. 135–140. (Russian) ↑¹

Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ НАН Беларуси)

ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР МУЛЬТИПРОЦЕССОРНЫХ СИСТЕМ ИПС РАН

S. V. Ablamejko, S. M. Abramov, V. V. Anistchenko, N. N. Paramonov. *Supercomputer family "SKIF"*. (in russian.)

ABSTRACT. The article is devoted to the review of technical characteristics and history of creation in 2000–2004 of the first 12 supercomputers "SKIF". The article gives also some additional information on the theme.