СУПЕРКОМПЬЮТЕРНЫЕ КОНФИГУРАЦИИ «СКИФ»

УДК 004.38

Рецензенты:
заведующий кафедрой математического обеспечения ЭВМ
Белорусского государственного университета доктор технических наук, профессор
М.К. Буза,
заведующий лабораторией
государственного научного учреждения «Объединенный институт проблем
информатики Национальной академии наук Беларуси» доктор технических наук,
профессор Р.Х. Садыхов

Абламейко С.В., Абрамов С.М., Анищенко В.В. , Медведев С.В., Парамонов Н.Н., Чиж О.П.

Суперкомпьютерные конфигурации «СКИФ» / Мн.:ОИПИ НАН Беларуси, 2005.-195 с.

Публикация посвящена отражению основополагающих принципов создания суперкомпьютерных систем «СКИФ», практических результатов комплексной реализации программы Союзного государства «СКИФ» и перспективности использования суперкомпьютерных технологий. В ней последовательно изложены общие сведения о программе «СКИФ», концептуальные принципы создания суперкомпьютеров «СКИФ», описаны прикладные комплексы и системы на базе суперкомпьютерных платформ, отмечены перспективы развития и внедрения наукоёмких суперкомпьютерных технологий.

Материалы будут полезны студентам соответствующих специальностей при изучении высокопроизводительных вычислительных систем с параллельной архитектурой, а также специалистам в области суперкомпьютерных технологий, включая разработчиков средств параллельного программирования и прикладных задач с большим объемом вычислений.

УДК 004.38

СОДЕРЖАНИЕ

Введение

В современном обществе высокие информационные технологии стали фундаментальной инфраструктурой, подобно энергетике, дорожным коммуникациям и другим жизненно важным для экономики государства системам. Научное знание и информация становятся определяющим фактором общественной жизни и производства. В таком обществе, основанном на экономике знаний, значительная часть валового национального продукта создается в отраслях, непосредственно производящих новые знания, информационные блага и услуги, а также оборудование для передачи и обработки информации.

В последнее десятилетие в мире наблюдается лавинообразное увеличение объема информации – каждые три-четыре года он удваивается. Синхронно с этим процессом в ряде областей науки, техники и управления народно-хозяйственным комплексом появляется все больше задач, требующих для своего эффективного решения принципиально новых технологий обработки данных с предельно достижимыми значениями быстродействия средств вычислительной техники.

Передовые информационно-коммуникационные технологии (ИКТ) являются краеугольным камнем развития инфраструктуры современного государства. Нет ни одной отрасли науки, экономики, государственного управления и безопасности, для которых не были бы исключительно важны разработки и реализации перспективных ИКТ. В первую очередь это относится к суперкомпьютерным технологиям и прикладным комплексам на их основе. Без использования современных средств ИКТ эффективное, конкурентоспособное производство становится невозможным ни в одной отрасли.

При любом уровне развития компьютерной техники всегда имеются задачи, которые требуют для своего решения использования вычислительной мощности в десятки, сотни или тысячи раз превышающей ту, которая может быть реализована на одном процессоре. В связи с этим в настоящее время в мире наблюдается своеобразный бум в области стратегически важного направления по созданию высокопроизводительных вычислительных систем с параллельной архитектурой или суперкомпьютеров. Обладание все большими вычислительными мощностями имеет стратегическое значение для развитых государств, сравнимое со значением ракетно-ядерного потенциала. В связи с этим, среди ведущих промышленных стран идет острое соперничество за обладание все более совершенными и сверхпроизводительными компьютерными технологиями, как важным стратегическим ресурсом обеспечения развития страны. Практически все развитые страны Запада имеют сегодня национальные программы создания компьютеров сверхвысокой производительности.

Именно для решение этой проблемы правительством США принята стратегическая инициатива по развитию вычислительных мощностей (Accelerated Strategic Computing Initiative /ASCI/), которая, как декларируется в программных документах, станет инструментом обеспечения национальной безопасности США и основой для сохранения позиций США как мирового научно-технического и промышленного лидера. Американские эксперты не скрывают, что успех этой национальной программы позволит решить главную геостратегическую задачу – сделать глобальным контроль Соединённых Штатов над информационным пространством в масштабах всей планеты.

В рамках европейской программы ESPRIT многочисленные европейские проекты по высокопроизводительным вычислительным и сетевым технологиям (High Performance Computing and Networking /HPCN/) ежегодно получают обильное финансирование правительств стран Европейского Союза. При этом отбор проектов, финансируемых в рамках ESPRIT, в сильной степени основывается на учете применимости в промышленности результатов реализации этих проектов. Правительство Японии, в соответствие с проводимой политикой целенаправленного развития информационно-коммуникационной инфраструктуры (Advanced Info-Communications Infrastructure /ICI/), создает правительственные лаборатории и поддерживает сотрудничающие с ними университетские центры, работающие в области параллельных супервычислений, в частности, путем предоставления приобретенных у фирм NEC, Hitachi и Fujitsu мультикомпьютеров с параллельной архитектурой.

Стратегическое значение отрасли параллельных вычислительных технологий обусловлено тем, что десятки, сотни или тысячи современных мощных процессоров, объединенных в единую вычислительную систему, дают новое качество – инструмент обработки информации, мощь которого сравнима с мощью человеческого мозга, а по некоторым параметрам (например, по скорости вычислений) во много раз превосходит его. Такой интеллектуальный ассистент становится незаменимым инструментом или даже мощным оружием, способным дать значительное, если не подавляющее технологическое и военное превосходство над другими государствами. Именно это и стало причиной проведения ограничительной политики в области экспорта высокопроизводительных параллельных вычислительных установок, проводимой правительствами индустриальных держав и соответствующими фирмами. Такая политика, наряду с явными запретами на экспорт высокопроизводительных вычислительных систем, включает в себя неявные ограничения, заключающиеся, например, в установлении сверхвысокого уровня цен на экспортируемые готовые параллельные суперкомпьютеры. Уровень цен таких изделий, по сути, исключает возможность активного использования па-

раллельных суперкомпьютеров в различных отраслях промышленности и науки, для обеспечения национальной безопасности и т.д. Многие развивающиеся страны осознают необходимость овладения технологиями высокопроизводительных параллельных вычислений с целью освобождения от упомянутой выше зависимости и обеспечения развития промышленности этих государств.

Мировые тенденции в области разработки и использования технологий высокопроизводительных параллельных вычислений показывают, что задача развития суперкомпьютерных вычислений является актуальнейшей задачей для России и Беларуси. Экономический прогресс, политическая стабильность и национальная безопасность обеих стран существенно зависят от успехов в решении данной задачи.

В связи с вышеизложенным, в конце 90-х годов прошлого столетия возникла идея разработки и создания единого семейства относительно недорогих моделей суперкомпьютеров с параллельной архитектурой с собственным оригинальным программным обеспечением, способных перекрыть диапазон производительности от миллиардов до триллионов операций в секунду.

Обе стороны – российская и белорусская – обладали необходимыми экспериментально апробированными научно-техническими и конструкторско-технологическими решениями высокой степени новизны. С направлением суперкомпьютеров самым тесным образом связаны такие наукоемкие сектора промышленности как микроэлектроника, оптическое приборостроение, точная механика, средства отображения информации, коммуникационная техника, производство программных продуктов и др. Именно в этих секторах Республика Беларусь и Российская Федерация имеют значительный научно-технический потенциал, поддерживаемый необходимыми фундаментальными и прикладными исследованиями, целевое использование которого позволило в сравнительно короткие сроки при относительно небольших затратах выйти на собственный альтернативный, практически независимый от Запада путь развития отечественной конкурентоспособной высокопроизводительной вычислительной техники, уровень которой будет соответствовать прогнозируемым требованиям со стороны широкой категории конечных пользователей.

Исходя из этих предпосылок, в 2000 году на базе научно-технического задела в Беларуси и России, была разработана и принята к исполнению программа Союзного «Разработка и освоение в серийном производстве семейства высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе» (шифр программы – «СКИФ»).

Головные исполнители Программы – Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ НАН Беларуси) и Институт программных систем Российской академии наук (ИПС РАН).

В реализации программных мероприятий приняло участие около 20 предприятий от Республики Беларусь и Российской Федерации в том числе:

– от Республики Бедарусь: УП «Научно-исследовательский институт электронных вычислительных машин» (УП «НИИ ЭВМ») и УП «Белмикросистемы» (НПО «Интеграл») Минпрома Республики Беларусь, УП «Национальный центр информационных ресурсов и технологий» НАН Беларуси, НИРУП «Геоинформационные системы» НАН Беларуси, БГУИР и БГУ Минобразования Республики Беларусь, РНПЦ «Кардиология» Минздрава Республики Беларусь, Институт тепло и массообмена НАН Беларуси и др.;

– от Российской Федерации: ОАО «Научно-исследовательский центр электронно-вычислительной техники» (ОАО «НИЦЭВТ»), АНО «Институт высокопроизводительных вычислений и информационных систем» (АНО «ИВВиИС»), Центр научных телекоммуникаций и информационных технологий РАН, ГНУ «Российский научно-исследовательский институт региональных проблем» Министерства образования Российской Федерации (ГНУ «РосНИИ РП»), Научно-исследовательский институт механики Московского государственного университета им. М.В. Ломоносова и др.

Главной целью настоящей публикации является отражение основополагающих принципов создания суперкомпьютерных систем «СКИФ», практических результатов комплексной реализации программы Союзного государства «СКИФ» и перспективности использования суперкомпьютерных технологий.

Результаты комплексной реализации программы «СКИФ» являются существенным научно-техническим и организационным заделом для дальнейшего развития суперкомпьютерного направления, в том числе для формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ».

При подготовке книги использовались в основном материалы, разработанные головными исполнителями программы «СКИФ» – ОИПИ НАН Беларуси (генеральный директор, научный руководитель программы от РБ – С.В. Абламейко) и ИПС РАН (директор, научный руководитель программы от Российской Федерации – С.М. Абрамов). В книге использованы также материалы, разработанные в рамках программы «СКИФ» предприятиями ГНУ «ИТМО им. А.В.Лыкова» НАН Беларуси (директор – С.А. Жданок), НИРУП «Национальный центр информаци-

онных ресурсов и технологий» НАН Беларуси (директор – М.М. Маханек), Белорусского государственного университета (первый проректор – С.К. Рахманов), УП «НИИЭВМ» (директор П.И. Сидорик) и др.

Благодарности. Авторы благодарны всем участникам программы «СКИФ», а особенно: Д.Б. Жаворонкову, В.П. Качкову, Е.Э. Константиновой, Л.И. Кульбаку, В.А. Лапицкому, А.Г. Рымарчуку, В.К. Фисенко. Авторы выражают особую благодарность О.В. Учкуронис за большой вклад в подготовку окончательной редакции книги.

1. Концептуальные принципы создания суперкомпьютеров «СКИФ»

Основополагающими архитектурными принципами создания суперкомпьютерных конфигураций «СКИФ» являются: базовая кластерная архитектура, иерархические кластерные конфигурации (метакластеры) и универсальная двухуровневая архитектура.

1.1. Базовая кластерная архитектура

Концепция создания моделей семейства суперкомпьютеров «СКИФ» базируется на масштабируемой кластерной архитектуре, реализуемой на классических кластерах из вычислительных узлов (рис. 1.1) на основе компонент широкого применения (стандартных микропроцессоров, модулей памяти, жестких дисков и материнских плат, в том числе с поддержкой SMP).

Кластерный архитектурный уровень – это тесносвязанная сеть (кластер) вычислительных узлов, работающих под управлением ОС Linux - одного из клонов широко используемой многопользовательской универсальной операционной системы Unix. Для организации параллельного выполнения прикладных задач на данном уровне используются:

– разработанная в рамках программы оригинальная система поддержки параллельных вычислений - Т-система, реализующая автоматическое динамическое распараллеливание программ;

– классические системы поддержки параллельных вычислений, обеспечивающие эффективное распараллеливание прикладных задач различных классов (как правило, задач с явным параллелизмом): MPI, PVM, Norma, DVM и др. В семействе суперкомпьютеров «СКИФ» в качестве базовой классической системы поддержки параллельных вычислений выбран MPI, что не исключает использование других средств.

На кластерном уровне с использованием Т-системы и MPI эффективно реализуются фрагменты со сложной логикой вычисления, с крупноблочным (явным статическим или скрытым динамическим) параллелизмом. Фрагменты же с простой логикой вычисления, с конвейерным или мелкозернистым явным параллелизмом, с большими потоками информации, требующими обработки в реальном режиме времени, на кластерных конфигурациях реализуются менее эффективно. Для организации параллельного исполнения задач с подобными фрагментами наиболее адекватна модель потоковых вычислений (data-flow).

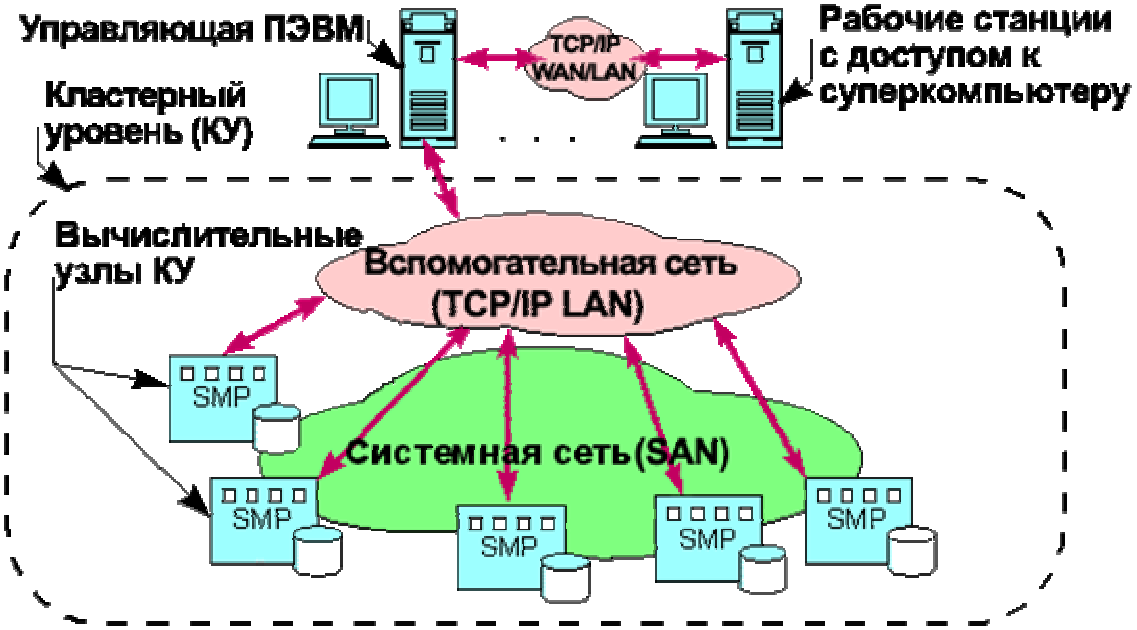

Рис. 1.1. Базовая кластерная архитектура

Кластерная архитектура является открытой и масштабируемой, т.е. не накладывает жестких ограничений к программно-аппаратной платформе узлов кластера, топологии вычислительной сети, конфигурации и диапазону производительности суперкомпьютеров.

Для организации взаимодействия вычислительных узлов суперкомпьютера в его составе используются различные сетевые (аппаратные и программные) средства, в совокупности образующие две системы передачи данных:

Системная сеть кластера (СС) или **System Area Network (SAN)** объединяет узлы кластерного уровня в кластер. Данная сеть поддерживает масштабируемость кластерного уровня суперкомпьютера, а также пересылку и когерентность данных во всех вычислительных узлах кластерного уровня суперкомпьютера. Системная сеть кластера строится на основе специализированных высокоскоростных линков класса SCI, Myrinet, Infiniband и др., предназначенных для эффективной поддержки кластерных вычислений на уровне ОС Linux и систем организации параллельных вычислений (Т-система, MPI).

Вспомогательная сеть суперкомпьютера (ВС) с протоколом TCP/IP объединяет узлы кластерного уровня в обычную (TCP/IP) локальную сеть (TCP/IP LAN). Данная сеть может быть реализована на основе широко используемых сетевых технологий класса Fast Ethernet, Gigabit Ethernet, ATM и др. Данная сеть предназначена для управления системой, подключения рабочих мест пользователей, интеграции суперкомпьютера в локальную сеть предприятия и/или в глобальные сети.

Кроме того, данный уровень может быть использован и системой организации параллельных кластерных вычислений (Т-система, MPI) для вспомогательных целей (основные потоки информации, возникающие при организации параллельных кластерных вычислений, передаются через системную сеть кластера).

В некоторых случаях аппаратура системной сети, например, Myrinet, позволяет без ущерба для реализации кластерных вычислений поддержать на этой же аппаратуре реализацию сети TCP/IP. В этих случаях аппаратные части обеих сетей (SAN и TCP/IP LAN) могут быть совмещены.

Кластерные конфигурации на базе только вспомогательной сети TCP/IP без использования дорогостоящих специализированных высокоскоростных линков класса SCI могут быть реализованы в рамках семейства «СКИФ» в виде самостоятельных изделий (TCP/IP кластеры). Программное обеспечение таких кластеров – ОС Linux, Т-система и соответствующая версия MPI. Реализация сравнительно недорогих TCP/IP кластеров на базе «масштабирования вниз» архитектурных решений «СКИФ» (дополнительный или вторичный эффект) существенно расширяет область применения результатов выполнения Программы.

Кластерные конфигурации на базе только вспомогательной сети могут быть реализованы как на базовых конструктивах «СКИФ», так и путем кластеризации имеющихся у пользователей ПЭВМ («персональные кластеры» или «супер ПЭВМ»).

1.2. Базовое (системное) программное обеспечение суперкомпьютеров

В качестве базовой операционной системы (ОС) в универсальном кластерном суперкомпьютере используется операционная система Linux, являющаяся клоном ОС Unix. Операционная система Linux является одной из самых надежных, эффективных и перспективных операционных систем, которую сегодня многие коммерческие и государственные организации выбирают в качестве базовой для приложений и перспективных разработок в области параллельных вычислений. ОС Linux распространяется свободно (бесплатно) с исходными текстами. Это дает возможность модифицировать и вносить изменения, необходимые для реализации поставленной задачи.

Функциональные возможности ОС Linux и ее утилит развиваются огромной армией добровольных программистов-разработчиков (сегодня 7-10 миллионов установок ОС Linux в мире), что обеспечивает непрерывность ее тестирования и корректировки ошибок в исходных текстах. Распространение ОС Linux не подвержено каким-либо ограничениям каких-либо стран или фирм. ОС Linux является открытой, то есть она реализована не только для платформ класса IBM PC, но и для многих дру-

гих аппаратных платформ.

1.3. Иерархические кластерные конфигурации (метакластеры)

Отдельные кластеры могут быть объединены в единую кластерную конфигурацию – кластер высшего уровня или метакластер (metacluster). Метакластерный принцип позволяет создавать распределенные метакластерные конфигурации на базе локальных или глобальных сетей передачи данных. При этом, естественно, уменьшается степень связности подкластеров метакластерной конфигурации.

Системное программное обеспечение метакластера обеспечивает возможность реализации **гетерогенных систем**, включающих подкластеры различной архитектуры на различных программно-аппаратных платформах.

Одним из перспективных программных продуктов, с использованием которого возможна реализация метакластерных конфигураций на подкластерах с различными программно-аппаратными платформами, является **IMPI** (Interoperable Message Passing Interface). IMPI реализует стандартизованный протокол, обеспечивающий взаимодействие различных реализаций MPI. Это позволяет выполнять общую задачу на различной аппаратуре с использованием настраиваемых поставщиком (vendor-tuned) различных реализаций MPI на каждом узле кластерной конфигурации соответствующего уровня иерархии. Такая возможность полезна в случаях, когда объем вычислений задачи слишком велик для одной системы или когда разные части задачи оптимальнее выполняются на разных реализациях MPI.

IMPI определяет только протоколы, необходимые для взаимодействия различных реализаций MPI, а также может использовать собственные высокопроизводительные протоколы этих реализаций. Существуют свободно распространяемые (открытые) версии IMPI, например, на базе LAM/MPI.

1.4. Универсальная двухуровневая архитектура

Для оптимизации организации на суперкомпьютерах «СКИФ» параллельного счета задач как с крупноблочным (явным статическим или скрытым динамическим) параллелизмом, так и с конвейерным или мелкозернистым явным параллелизмом, с большими потоками информации, требующими обработки в реальном режиме времени, предусмотрена реализация **универсальной двухуровневой архитектуры** суперкомпьютеров (рис. 1.2):

– 1-й уровень - базовый (кластерный) архитектурный уровень;

– 2-й уровень - потоковый архитектурный уровень, реализующий модель потоковых вычислений (data-flow).

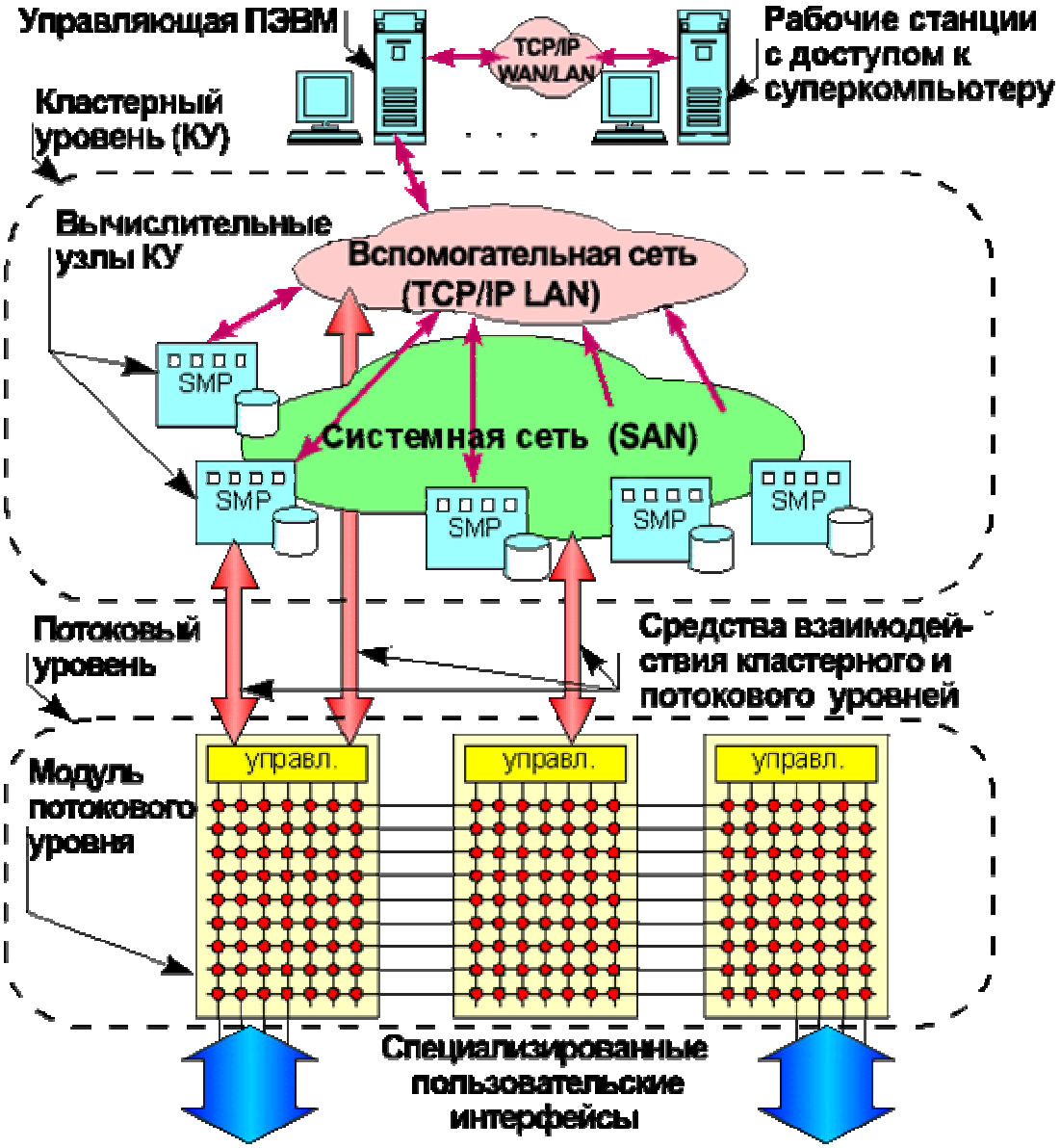

Рис. 1.2. Универсальная двухуровневая архитектура

Концепция создания суперкомпьютеров «СКИФ» предусматривает реализацию потокового архитектурного уровня как на базе однородной вычислительной среды (ОВС) с использованием оригинальных СБИС ОВС, разработанных в рамках программы «СКИФ», так и на базе других (альтернативных) структурных и технических решений (например, на базе нейроструктур, FPGA типа XILINX и др.). По сути, вычислительные модули потокового уровня являются сопроцессорами вычислительных ресурсов кластерной конфигурации.

Предпосылкой объединения двух программно-аппаратных решений (кластерного и потокового) для организации параллельной обработки в рамках одной вычислительной системы, является то, что эти два подхода, как уже отмечалось, своими сильными сторонами компенсируют недостатки друг друга. Тем самым, в общем случае, каждая прикладная проблема может быть разбита на:

– фрагменты со сложной логикой вычисления, с крупноблочным (явным статическим или скрытым динамическим) параллелизмом, эффективно реализуемые на кластерном уровне с использованием Т-системы и других (классических) систем поддержки параллельных вычислений;

– фрагменты с простой логикой вычисления, с конвейерным или мелкозернистым явным параллелизмом, с большими потоками информации, требующими обработки в реальном режиме времени, эффективно реализуемые на потоковом уровне.

На потоковом уровне может быть эффективно реализован высокоскоростной потоковый обмен со стандартной компьютерной периферией и/или с нестандартными устройствами-датчиками, например, с датчиками медицинских и других приборов.

1.5. Программные средства сопряжения кластерного и потокового архитектурных уровней

Средства взаимодействия двух уровней суперкомпьютера обеспечивают возможность взаимодействия между кластерным и потоковым уровнями суперкомпьютера и реализуются в рамках сетей SAN или TCP/IP LAN. Следовательно, при реализации в модулях потокового уровня соответствующих сетевых интерфейсов, эти модули, в принципе, могут выступать в качестве устройств системной сети (SAN) и/или вспомогательной сети суперкомпьютера (TCP/IP LAN).

Программные средства сопряжения в части кластерного уровня должны включать в себя:

– набор драйверов устройств, обеспечивающих сопряжение кластерного и потокового уровней;

– базовую библиотеку стандартных примитивов обмена информацией и управления потоковым уровнем;

– библиотеку прикладных задач и подпрограмм, реализуемых с использованием потокового уровня;

– структуры данных и программные механизмы, обеспечивающие:

а) передачу Т-процесса, из которого осуществляется взаимодействие с модулем потокового уровня, в один из вычислительных узлов кластерного уровня, имеющих физический интерфейс с модулем потокового уровня;

б) осуществление удаленного вызова функции/прикладной задачи из вычислительного узла кластерного уровня, не имеющего интерфейса с модулем потокового уровня с использованием механизмов, предназначенных для распределенной работы с файлами.

В части потокового уровня программные средства сопряжения должны включать в себя реализованный в виде специализированной библиотеки набор фрагментов программного кода, предназначенных для загрузки из кластерной компоненты в потоковую. Каждый из фрагментов непосредственно реализует на потоковом архитектурном уровне ту или иную прикладную задачу или фрагмент вычислений, в частности:

– получает из кластерного уровня наборы входных данных;

– организует и осуществляет выполнение вычислений в модуле потокового уровня в соответствии с алгоритмом решения соответствующей прикладной задачи;

– передает из потокового в кластерный уровень наборы данных, содержащие результаты вычислений.

Описанный набор программных средств, структур данных и механизмов поддерживает возможности:

– передачи фрагмента решаемой задачи из Т-программы на вычисление в модуль потокового уровня;

– передачи фрагмента решаемой задачи из выполняемого в модуле потокового уровня кода на вычисление в кластерную компоненту.

1.6. Отличительные особенности архитектуры семейства суперкомпьютеров «СКИФ»

Предложенная многоуровневая схема реализации архитектурных принципов обладает рядом особенностей и преимуществ (по сравнению с аналогичными разработками), позволяющими достичь современный мировой уровень в суперкомпьютерной отрасли:

1) **В части Т-системы:** обеспечивается автоматическое динамическое распараллеливание программ, что освобождает программиста от большинства трудоемких аспектов разработки параллельных программ, свойственных различным системам ручного статического распараллеливания:

– обнаружение готовых к выполнению фрагментов задачи (процессов);

– их распределение по процессорам;

– их синхронизацию по данным.

Все эти (и другие) операции выполняются в Т-системе автоматически и в динамике (во время выполнения задачи). Тем самым при более низких затратах на разработку параллельных программ обеспечивается более высокая их надежность.

По сравнению с использованием распараллеливающих компиляторов, Т-система обеспечивает более глубокий уровень параллелизма во время выполнения программы и более полное использование вычислительных ресурсов мультипроцессоров. Это связано с принципиальными алгоритмическими трудностями (алгоритмически неразрешимыми проблемами), не позволяющими во время компиляции (в статике) выполнить полный точный анализ и предсказать последующее поведение программы во время счета.

Кроме указанных выше принципиальных преимуществ Т-системы перед известными сегодня методами организации параллельного счета, в реализации Т-системы имеется ряд технологических находок, не имеющих аналогов в мире, в частности:

– реализация понятия «неготовое значение» и поддержка корректного выполнения некоторых операций над неготовыми значениями. Тем самым поддерживается возможность выполнение счета в некотором процессе-потребителе в условиях, когда часть из обрабатываемых им значений еще не готова, т. е. не вычислена в соответствующем процессе-поставщике. Данное техническое решение обеспечивает обнаружение более глубокого параллелизма в программе;

– оригинальный алгоритм динамического автоматического распределения процессов по процессорам. Данный алгоритм учитывает особенности неоднородных распределенных вычислительных сетей. По сравнению с известными алгоритмами динамического автоматического распределения процессов по процессорам (например, с диффузионным алгоритмом и его модификациями), алгоритм Т-системы имеет существенно более низкий трафик межпроцессорных передач. Тем самым, Т-система обеспечивает снижение накладных расходов на организацию параллельного счета и предъявляет менее жесткие требования к пропускной способности аппаратуры объединения процессорных элементов в кластер;

2) **В части потокового уровня:** архитектура вычислительных модулей потокового уровня позволяет использовать естественный параллелизм решаемой задачи вплоть до битового уровня, то есть уровня структуры обрабатываемых данных, а также позволяет строить конвейеры произвольной глубины. Потоковый уровень предоставляет возможность одновременной обработки множества независимых некогерентных потоков.

Фактически, при решении конкретной функции или самостоятельной задачи, на вычислительных модулях потокового уровня путем ввода соответствующей программы организуется **спецпроцессор**, реализующий решаемую функцию или задачу с наибольшей эффективностью. На матрице модулей потокового уровня одновременно могут решаться не-

сколько независимых задач и функций, причем механизм перезагрузки сегментов потокового уровня позволяет перезагружать часть матрицы без остановки выполнения еще незавершенных задач. Потоковый уровень обладает высокой гибкостью и перестраиваемостью, в частности, полной аппаратной и программной масштабируемостью, что позволяет строить на его основе вычислительные системы с большим быстродействием. Производительность матрицы модулей потокового уровня, теоретически, растет линейно с увеличением рабочей частоты поля и площади вычислительной матрицы.

Вычислительные модули потокового уровня позволяют создавать системы с высоким уровнем надежности и отказоустойчивости.

Предложенные архитектурные принципы позволяют эффективно реализовывать любые виды параллелизма в том числе нейросетевые алгоритмы. **Архитектура является открытой и масштабируемой**, то есть не накладывает жестких ограничений к программно-аппаратной платформе узлов кластера, топологии вычислительной сети, конфигурации и диапазону производительности суперкомпьютеров. Вычислительные системы, создаваемые на базе основополагающих концептуальных архитектурных принципов могут оптимально решать как классические вычислительные задачи математической физики и линейной алгебры, так и специализированные задачи обработки сигналов, моделирования виртуальной реальности, задачи управления сложными системами в реальном времени и другие приложения.

1.7. Конструкторско-технологические решения

Универсализация и унификация интерфейсов аппаратно-программных средств ПЭВМ и серверных платформ привела к возможности сборки кластерных вычислительных систем из свободно продаваемых компонентов высокой степени готовности и интеграции.

В рамках программы «СКИФ» разработано семейство базовых конструктивно-вычислительных модулей, предназначенных для построения моделей первого ряда семейства суперкомпьютеров кластерного уровня.

Семейство базовых конструктивно-вычислительных модулей кластерного уровня (БКВМ КУ) включает:

– базовые вычислительные модули (узлы кластера) на серверной платформе с габаритами 1U, 2U, 3U или 4U;

– базовые конструктивные стойки (шкафы) стандарта 19″ высотой 15U, 18U, 24U, 28U и др., предназначенные для размещения вычислительных узлов, коммутационных средств системной и управляющей сети, источника бесперебойного питания, системы вентиляции и др.

Принятые технические решения по БКВМ КУ обеспечивают возможность создания на их основе различных конфигураций вычисли-

тельных систем кластерного уровня с широким диапазоном производительности с использованием готовых изделий массового применения, широко представленных на компьютерном рынке, в частности, стандартных двухпроцессорных системных плат SMP, микропроцессоров известных фирм производителей типа Intel, AMD (или совместимых микропроцессоров других фирм), стандартных модулей памяти (с объемом 0,5 – 4,0 Гбайт), стандартных жестких дисков (с объемом – 10 Гбайт и выше), адаптеров для подключения вычислительных узлов к системной сети кластера типа SCI, Infiniband, MYRINET, Gigabit Ethernet и вспомогательной сети типа Fast Ethernet, Gigabit Ethernet.

Вычислительные системы кластерного типа семейства «СКИФ» могут наращиваться объединением стоек по горизонтали и/или по вертикали в зависимости от количества и габаритов используемых базовых конструктивно-вычислительных модулей.

1.8. Базовые конфигурации суперкомпьютерных систем

Базовые конфигурации суперкомпьютерных систем (БКСС) отображают весь спектр возможных прикладных суперкомпьютерных реализаций на базе основополагающего концептуального принципа создания семейства суперкомпьютеров – двухуровневой открытой, масштабируемой архитектуры.

Конфигурация вычислительной системы (по определению) – это совокупность функциональных частей вычислительной системы и связей между ними, обусловленная основными техническими характеристиками этих частей, а также характеристиками решаемых задач обработки данных.

В состав базовых конфигураций входят базовые вычислительные модули (БВМ), сетевые средства поддержки взаимодействия вычислительных узлов и соответствующее базовое (общесистемное) и прикладное программное обеспечение.

Количество типов БКСС определяется с учетом следующих основных моментов:

– диапазон производительности создаваемых на базе БКСС прикладных суперкомпьютерных систем и характерные области их применения;

– принятый порядок разработки суперкомпьютерных реализаций, включая разработку конструкторской документации и проведение соответствующих испытаний;

– освоение суперкомпьютерных конфигураций в производстве, включая подготовку производства, создание необходимого производственного задела, проведение различных типов испытаний, потребитель-

ский спрос, поставку конкретных прикладных суперкомпьютерных конфигураций;

– техническое обслуживание у потребителей суперкомпьютерных изделий.

Основные концептуальные принципы создания семейства суперкомпьютеров (открытая масштабируемая архитектура, набор базовых вычислительных модулей и конфигураций и др.) позволяют оптимальным способом создавать для каждой конкретной прикладной проблемы адекватную суперкомпьютерную конфигурацию. В связи с этим фактически отождествляются понятия модель и прикладная конфигурация.

Модели (прикладные конфигурации) идентифицируются в соответствии с **классификатором семейства суперкомпьютеров**.

Идентификатор модели суперкомпьютера, включающий идентифицирующие признаки, определяет:

– диапазон производительности;

– тип архитектуры - кластерная, метакластерная, потоковая, двухуровневая;

– типы используемых вычислительных узлов;

– тип базовой конфигурации суперкомпьютерных систем;

– тип средств сопряжения вычислительных узлов и т.п.

Учитывая наличие естественного диапазона влияния идентифицирующих признаков, практически каждая конкретная модель суперкомпьютера может включать несколько конкретных прикладных суперкомпьютерных конфигураций (модификации моделей).

Специфика моделей и их модификаций отражается в эксплуатационной документации.

Основополагающие принципы Концепции позволяют создавать прикладные системы, соответствующие требованиям конкретного заказчика, оптимально использовать производственные мощности предприятия- изготовителя с учетом специфики рынка сбыта высокопроизводительных вычислительных систем.

1.9. Конструкторская и программная документация.

Конструкторская документация. Конструкторская документация разрабатывается на базовые модули, имеющие самостоятельную поставку, и на базовые конфигурации суперкомпьютерных систем.

Конструкторская документация выполняется в едином для всех исполнении БВМ или БКСС групповом варианте в соответствии с действующими стандартами.

В соответствии с групповым принципом документация содержит постоянные и переменные данные. Постоянные данные – это технические параметры и характеристики, являющиеся общими для всех испол-

нений каждого типа БВМ или БКСС. Переменные данные отражают отличительные характеристики конкретных исполнений БВМ или конкретных конфигураций, реализуемых на основе данного типа БКСС (то есть отражают специфику моделей суперкомпьютеров и их модификаций).

Конструкторские документы (спецификации, формуляры, паспорта, комплекты монтажных частей, программного обеспечения и т.п.) являются открытыми, то есть содержат разделы, заполняемые для конкретных реализаций.

Технические условия также оформляются по групповому принципу – в них указываются диапазоны изменения параметров и технических характеристик, допускаемые конфигурации, комплектность и т.д. для БВМ и БКСС соответствующего типа, а также предусматривается возможность для указания конкретных значений этих характеристик для конкретных реализаций.

Групповое построение конструкторской документации адекватно отражает возможности архитектурной идеологии (открытость, масштабируемость), позволяя оптимальным способом организовать серийное производство широкой номенклатуры моделей суперкомпьютеров, наиболее полно удовлетворяющих предъявляемым пользовательским требованиям.

Программная документация. Программная документация разрабатывается на базовое программное обеспечение кластерного уровня в соответствии с требованиями комплекса стандартов ЕСПД и включает программную документацию на программные компоненты:

1) Ядро ОС Linux, адаптированное для работы на суперкомпьютере «СКИФ».

2) Программное обеспечение поддержки параллельных вычислений.

1.10. Программное обеспечение кластерного уровня

1.10.1. Состав программного обеспечения кластерного уровня

Программное обеспечение (ПО) кластерного уровня (КУ) моделей суперкомпьютеров семейства «СКИФ» включает:

а)низкоуровневое программное обеспечение кластерного уровня:

– система мониторинга FLAME;

– параллельная файловая система PVFS-SKIF;

б) программное обеспечение поддержки параллельных вычислений:

– Т-система с открытой архитектурой и ее окружение (ядро Т-системы с открытой архитектурой, компилятор TG языка Т++, транслятор TF2TC с языка программирования T-Fortran в язык ТС);

– распределенная программная система интерактивной отладки MPI-программ – ПС TDB.

в) прикладные программные системы.

Высокие технико-экономические показатели программного обеспечения обеспечиваются возможностью использования ядра Т-системы для автоматического динамического распараллеливания программ, что дает положительный экономический эффект за счет:

– упрощения процесса создания высокопроизводительных программ (не требуется явное ручное распараллеливание и ручное распределение свободных вычислительных ресурсов);

– обеспечения для ряда задач более высокого коэффициента утилизации вычислительной мощности суперкомпьютера, чем при ручном статическом распараллеливании;

– снижение общих затрат, необходимых для приобретения и эксплуатации программно-аппаратной платформы суперкомпьютеров семейства «СКИФ», за счет предоставления Т-системы, как отечественной альтернативы зарубежным коммерческим средствам поддержки параллельных вычислений.

Возможность использования ПС TDB для отладки MPI-программ и, следовательно, отказа от закупок аналогичных дорогостоящих коммерческих продуктов (таких, как средство интерактивной отладки MPI-программ TotalView фирмы Etnus) должна дать положительный экономический эффект путем снижения общих затрат, необходимых для приобретения и эксплуатации программно-аппаратной платформы суперкомпьютеров семейства «СКИФ», и, таким образом, повышения ее привлекательности для потенциальных потребителей изделий семейства «СКИФ».

1.10.2. Система мониторинга FLAME

Система мониторинга FLAME обеспечивает автоматический контроль состояния программных и аппаратных средств кластерного уровня суперкомпьютеров семейства «СКИФ». Система может быть использована для мониторинга вычислительных систем, работающих под управлением операционной системы (ОС) Linux. Кроме того, FLAME предоставляет интерфейс для удаленного управления кластером посредством сервисной сети.

Использование системы мониторинга FLAME позволяет:

– автоматизировать большую часть операций, связанных с мониторингом и управлением параметрами аппаратных и программных средств суперкомпьютеров семейства «СКИФ»;

– отказаться от закупок аналогичных дорогостоящих коммерческих продуктов (таких, как системы управления сетями Sun Net Manager и HP OpenView).

Система мониторинга FLAME выполняет следующие функции:

– сбор;

– обработку;

– отображение (визуализацию);

– предоставление информации о состоянии базовых вычислительных модулей кластерного уровня. Состояние – это совокупность значений всех контролируемых параметров (метрик) вычислительной системы в определенный момент времени.

Набор метрик должен включать следующие параметры:

– количество контролируемых вычислительных узлов (от агентов которых поступает информация);

– показания датчиков температуры и скорости вращения вентиляторов;

– количество процессоров;

– загруженность процессоров;

– использование оперативной памяти и swap;

– количество процессов;

– количество пользователей;

– список процессов;

– состояние (статус) линков системной сети (SCI);

– загруженность управляющей сети (как правило, семейства Ethernet, с поддержкой TCP/IP);

– сообщения (логи) в заданных файлах.

Сбор информации о вычислительной системе осуществляется посредством опроса узлов по протоколам SNMP и HTTP через определенные интервалы времени. При этом, в ответ на поступивший запрос, на вычислительных узлах запускаются программы-агенты, результаты выполнения которых передаются по заданному протоколу на управляющую ЭВМ, на которой установлена и функционирует система мониторинга FLAME.

Полученная от агентов информация преобразуется во внутренний XML формат в соответствии с правилами, заданными в файле конфигурации вычислительной системы, который также имеет формат XML. В файле конфигурации задаются интервалы опроса устройств и повторных вызовов функций.

Консоль оператора системы мониторинга FLAME отображает, посредством экранных форм, информацию, поступившую в заранее известном формате XML, используя для этого все доступные средства и

способы наглядного представления информации: в виде текста, иконок, таблиц, графиков, списков. При этом оператор может задавать интервал времени обновления информации отдельно для каждой экранной формы и «ускорять» обновление информации, нажав кнопку вызова обновления экранной формы. Система мониторинга FLAME предоставляет по запросу информацию о состоянии вычислительной системы в текстовом виде.

В системе FLAME предусмотрен механизм выявления и предупреждения о возможности возникновения аварийных ситуаций посредством автоматического мониторинга критических параметров системы и сопоставления их с критичными значениями. При выполнении определенных условий система мониторинга FLAME автоматически выполняет предписанные действия, направленные на поддержание работоспособности кластера.

Система мониторинга FLAME предоставляет графический пользовательский интерфейс к утилитам сервисной сети кластера. Интерфейс обеспечивает:

– получение информации о состоянии (статусе) вычислительного узла: узел включен или выключен;

– отображение информации о состоянии узла;

– управление питанием: плавное (с задержкой) включение и выключение всей установки, стойки или узла;

– аппаратный сброс всей установки, стойки или узла;

– управление загрузкой и перезагрузкой: выбор операционной системы, а для операционной системы Linux - выбор ядра и параметров ядра;

– предоставление сериальной консоли для взаимодействия с узлом;

– сбор логов сериальных консолей узлов: постоянное считывание буфера сериальной консоли каждого узла и сохранение полученной информации в лог-файлах;

– запись в лог-файл интерактивной сессии сериальной консоли с узлом: сохранение обмена данными в ходе сеанса работа в сериальной консоли;

– управление записью логов сериальных консолей узлов: выбор файла для записи логов, включение и выключение записи сообщений и записи интерактивных сессий;

– отображение информации о записанных логах сериальных консолей узлов: выборка логов, удовлетворяющих заданным условиям;

– мониторинг логов сериальных консолей узлов: информирование оператора о возникновении критических событий на основании анализа логов сериальных консолей узлов.

Система мониторинга FLAME функционирует в составе ПО КУ на аппаратной платформе кластерного уровня суперкомпьютеров «СКИФ», включающей:

– управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура, мышь);

– вычислительные узлы кластерного уровня (БВМ КУ);

– системную сеть, объединяющую вычислительные узлы (базовые вычислительные модули кластерного уровня, БВМ КУ);

– управляющую сеть (как правило, семейства Ethernet, с поддержкой TCP/IP), объединяющую управляющую ЭВМ и вычислительные узлы;

– сервисную сеть, объединяющую управляющую ЭВМ и вычислительные узлы.

1.10.3. Параллельная файловая система PVFS (PVFS-SKIF)

Параллельная файловая система PVFS (PVFS-SKIF) предназначена для обеспечения приложений возможностью организации на локальных дисках БВМ КУ суперкомпьютеров «СКИФ» распределенного хранения данных и высокоскоростного параллельного доступа к этим данным.

Программная система PVFS-SKIF является адаптацией свободного программного обеспечения «Файловая система PVFS» к аппаратным и программным средствам кластерного уровня суперкомпьютеров «СКИФ». Возможность использования на суперкомпьютерах «СКИФ» файловой системы PVFS позволяет отказаться от закупок дорогостоящих коммерческих продуктов.

PVFS-SKIF обеспечивает возможность прозрачного доступа к файловой системе PVFS для приложений, использующих эту файловую систему. В частности, программная система PVFS-SKIF предоставляет:

– возможность определения конфигурации оборудования;

– возможность конфигурирования файловой системы PVFS;

– возможность корректного запуска файловой системы PVFS;

– возможность предоставления приложениям доступа к файловой система PVFS с использованием системной сети (SCI);

– возможность корректного завершения работы.

Нормальное функционирование PVFS-SKIF в составе ПО КУ обеспечивается на аппаратной платформе кластерного уровня суперкомпьютеров «СКИФ», включающей:

– управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура);

– вычислительные узлы кластерного уровня (базовые вычислительные модули кластерного уровня, БВМ КУ);

– системную сеть кластера (SCI), объединяющую вычислительные узлы;

– вспомогательную сеть (как правило, семейства Ethernet, с поддержкой TCP/IP), объединяющую управляющую ЭВМ и вычислительные узлы.

В части совместимости с операционной системой реализована возможность установки PVFS-SKIF на программно-аппаратную платформу суперкомпьютеров «СКИФ» с использованием стандартного средства управления программными пакетами RPM.

1.10.4. Ядро Т-системы с открытой архитектурой

Ядро Т-системы с открытой архитектурой, предназначенно для реализации базовой функциональности Т-системы в части поддержки динамического распараллеливания программ, динамического распределения вычислительных ресурсов, а также обеспечения корректного взаимодействия параллельных вычислительных процессов как между собой, так и с базовым низкоуровневым ПО КУ суперкомпьютера «СКИФ».

Ядро Т-системы реализует базовые функции для автоматического динамического распараллеливания программ на языке ТС во время их работы на кластерном уровне суперкомпьютера «СКИФ». В частности, интерфейсы и библиотека Т-ядра обеспечивают:

– возможность компиляции программ на языке ТС для обычного последовательного исполнения стандартным компилятором для языка С;

– возможность компиляции и компоновки программ на языке ТС для параллельного исполнения при помощи входящего в комплект ПО КУ «СКИФ» специализированного компилятора TGCC;

– возможность корректного запуска полученного после компиляции и компоновки исполняемого модуля с помощью общепринятого средства запуска MPI-программ mpirun;

– поддержку в части динамического распараллеливания программ на языке ТС: в процессе порождения значительного количества параллельно работающих Т-функций;

– обеспечение автоматического распределения порождаемых гранул параллелизма на те процессоры и узлы суперкомпьютера, загруженность которых минимальна;

– поддержку явного указания контекста, в котором должна исполняться Т-функция, а именно номера вычислительного узла и типа вычислительного процесса (счетный / системный);

– поддержку в части обменов данными и синхронизации параллельных процессов при доступе к неготовым данным: в случае обращения процессов-потребителей к неготовым данным, исполнение послед-

него должно быть приостановлено до того момента, пока данные не будут объявлены процессом-поставщиком как готовые;

– выдачу статистической информации по окончании счета: печать реального времени, потраченного на ожидание результата и количества вызванных Т-функций на каждом узле суперкомпьютера.

Ядро Т-системы предоставляет возможность для пользователя производить:

– профилировку;

– трассировку исполняемой программы.

Нормальное функционирование ядра Т-системы обеспечивается в составе ПО КУ «СКИФ» на аппаратной платформе кластерного уровня суперкомпьютеров «СКИФ», включающей:

– управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура);

– вычислительные узлы кластерного уровня (базовые вычислительные модули кластерного уровня, БВМ КУ);

– системную сеть кластера, объединяющую вычислительные узлы;

– вспомогательную сеть (как правило семейства Ethernet, с поддержкой TCP/IP), объединяющую управляющую ЭВМ и вычислительные узлы.

В части совместимости с операционной системой реализована возможность установки ядра Т-системы на программно-аппаратную платформу суперкомпьютеров «СКИФ» с использованием стандартного средства управления программными пакетами RPM.

1.10.5. Компилятор TG языка T++

Компилятор TG языка T++ предназначен для компиляции программ, написанных на языке T++, который является одним из входных языков системы автоматического динамического распараллеливания программ (Т-системы с открытой архитектурой) и создания исполняемого кода, способного работать в системе автоматического динамического распараллеливания программ (Т-системе с открытой архитектурой) на суперкомпьютерах семейства «СКИФ».

Компилятор TG++ устанавливается на управляющей ЭВМ суперкомпьютера семейства «СКИФ».

В случае обнаружения ошибок в исходном коде программы, написанной на языке T++, компилятор TG++ выдает на стандартный вывод краткие сообщения об ошибках.

Компилятор TG++ имеет возможность снабжать сгенерированный им исполняемый код информацией, необходимой для дальнейшей отладки программы, работающей на мультипроцессорной системе.

Нормальное функционирование компилятора TG++ обеспечивается

в составе ПО КУ «СКИФ» на аппаратной платформе кластерного уровня суперкомпьютеров семейства «СКИФ», а именно на управляющей ЭВМ, оснащенной стандартными средствами, необходимыми для работы оператора (монитор, клавиатура), функционирующей под управлением ОС Linux.

В части совместимости с операционной системой реализована возможность установки компилятора TG++ на программно-аппаратную платформу суперкомпьютеров семейства «СКИФ» с использованием стандартного средства управления программными пакетами RPM.

При реализации компилятора TG++ используются только свободно распространяемые в исходных текстах программные средства.

1.10.6. Транслятор TF2TC с языка программирования T-Fortran в язык TC

Транслятор TF2TC с языка программирования T-Fortran в язык TC, предназначенный для трансляции программ, написанных на языке программирования T-Fortran, в программы на языке TC, который является одним из входных языков системы автоматического динамического распараллеливания программ (T-системы), а также последующую компиляции полученного кода на языке TC средствами компилятора T++ для создания исполняемого кода, способного работать в системе автоматического динамического распараллеливания программ (T-системе) на суперкомпьютерах семейства «СКИФ».

Транслятор TF2TC устанавливается на управляющей ЭВМ суперкомпьютера семейства «СКИФ».

Транслятор позволяет компилировать и собирать с ядром T-системы и с другими библиотеками синтаксически правильные (в соответствии с описанием языка программирования T-Fortran) программы на языке T-Fortran. В результате получается исполняемый код, пригодный для запуска и исполнения на кластерном уровне суперкомпьютеров «СКИФ».

В случае обнаружения ошибок в исходном коде программы, написанной на языке T-Fortran, компилятор TF2TC выдает на стандартный вывод краткие сообщения об ошибках.

Нормальное функционирование транслятора TF2TC обеспечивается в составе ПО КУ «СКИФ» на аппаратной платформе кластерного уровня суперкомпьютеров семейства «СКИФ», а именно на управляющей ЭВМ, оснащенной стандартными средствами, необходимыми для работы оператора (монитор, клавиатура), функционирующей под управлением ОС Linux.

В части совместимости с операционной системой реализована возможность установки транслятора TF2TC на программно-аппаратную

платформу суперкомпьютеров семейства «СКИФ» с использованием стандартного средства управления программными пакетами RPM.

1.10.7. Распределенная программная система интерактивной отладки MPI-программ (ПС TDB)

ПС TDB предназначена для обеспечения интерактивной отладки MPI-программ, реализованных на языках семейства C/C++, в том числе и программ, реализованных с использованием программного комплекса Т-система, на кластерном уровне суперкомпьютера «СКИФ». В частности, ПС TDB предоставляет:

– возможность запуска MPI-программ в режиме отладки с использованием одного или нескольких (не менее 16) вычислительных узлов суперкомпьютера;

– возможность формирования групп процессов отлаживаемого MPI-приложения и выполнения команд ПС TDB на таких группах процессов,

– возможности установки и отмены точек останова в процессах, входящих в отлаживаемую MPI-программу (MPI-процессов);

– возможность просмотра значений переменных, расположенных в приостановленных MPI-процессах;

– возможность вычисления значения отдельной функции в приостановленном MPI-процессе;

– возможности установки положения и просмотра текущего фрейма стека в приостановленном MPI-процессе;

– возможность приостановки процессов задачи по интерактивно заданной команде приостановки;

– возможность продолжения выполнения приостановленного MPI-процесса;

– возможность останова отдельного MPI-процесса по исключительной ситуации;

– возможность выполнения специальных групповых команд, таких как: групповые точки останова и групповой вариант автоматического отображения значений переменных и структур данных.

В случае, если для поддержки проведения интерактивной отладки при компиляции или компоновке (сборке) предназначенных к отладке MPI-программ необходимо использование специальных дополнительных заголовочных файлов или программных библиотек, ПС TDB обеспечивает возможность компиляции и/или компоновки данных MPI-программ путем предоставления соответствующих заголовочных файлов и программных библиотек.

ПС TDB включает в свой состав графический клиентский компо-

нент, который обеспечивает функциональность, гарантирующую использование планируемых к реализации функциональных свойств ПС TDB в полном объеме.

Нормальное функционирование ПС TDB обеспечивается в составе ПО КУ на аппаратной платформе кластерного уровня суперкомпьютеров «СКИФ», включающей:

– управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура);

– вычислительные узлы кластерного уровня (базовые вычислительные модули кластерного уровня, БВМ КУ);

– системную сеть кластера, объединяющую вычислительные узлы;

– вспомогательную сеть (как правило, семейства Ethernet, с поддержкой TCP/IP), объединяющую управляющую ЭВМ и вычислительные узлы.

В части совместимости с операционной системой реализована возможность установки ПС TDB на программно-аппаратную платформу суперкомпьютеров «СКИФ» с использованием стандартного средства управления программными пакетами RPM.

1.11. Показатели надежности и отказоустойчивости кластерных конфигураций семейства «СКИФ»

Принципы и методику расчета показателей надежности суперкомпьютеров «СКИФ» разработал старший научный сотрудник ОИПИ НАН Беларуси кандидат технических наук Кульбак Л.И.

1.11.1. Показатели надежности и отказоустойчивости суперкомпьютеров

Отличительной особенностью суперкомпьютера является параллельное участие в вычислительном процессе большого количества вычислительных средств. Исключение из вычислительного процесса одного или нескольких вычислительных средств не препятствует продолжению использования суперкомпьютера по назначению, а лишь снижает его потенциальную производительность. Эффективность использования объекта (в т.ч. и суперкомпьютера) по назначению – важное эксплуатационное свойство объекта. Очевидно, чем реже объект вынужден исключаться из производительной работы и чем меньше ущерб от вынужденного простоя, тем выше эффективность его использования.

Частота вынужденного простоя объекта зависит от его безотказности, а ущерб от вынужденного простоя объекта зависит от его ремонтопригодности. Следовательно, показатели оценки этих свойств объекта, в той или иной мере, характеризуют эффективность его использования.

Отказоустойчивость – это свойство объекта сохранять возмож-

ность использования его по назначению при возникновении в нем в процессе работы отказов составных частей (СЧ).

Предлагается различать следующие последствия отказов СЧ суперкомпьютера:

– вычислительный процесс при отказе СЧ суперкомпьютера не прервался – такое положение считается как отказоустойчивость первого вида;

– вычислительный процесс при отказе СЧ суперкомпьютера прервался для реконфигурации системы и затем, без процедуры восстановления отказавшей СЧ суперкомпьютера, продолжился с места его прерывания с незначительным повтором части вычислений – такое положение считается как отказоустойчивость второго вида;

– вычислительный процесс при отказе СЧ суперкомпьютера прервался для реконфигурации системы и затем, без процедуры восстановления отказавшей СЧ суперкомпьютера, стало возможным выполнение задания лишь сначала – такое положение считается как отказоустойчивость третьего вида.

Выбор показателя надежности (ПН) следует производить с учетом соответствующих нормативных документов, регламентирующих выбор номенклатуры ПН. С учетом этого изделия можно разделить на два вида. Изделия вида 1 – это те изделия, которые по работоспособности могут находиться только в двух состояниях – работоспособном (РС) и неработоспособном (НРС). Изделия вида 2 – это изделия, которые могут находиться в нескольких частично работоспособных состояниях.

Очевидно, суперкомпьютер в целом следует отнести к изделиям вида 2, так как кластеры могут использоваться по назначению и при снижении производительности из-за отказов некоторых СЧ суперкомпьютера. Для изделий вида 2 в качестве основного ПН рекомендуется использовать коэффициент сохранения эффективности. Примем в качестве меры эффективности суперкомпьютера пиковую производительность, которая пропорциональна числу вычислительных узлов (ВУ) в кластерной вычислительной системе (КВС), доступных программному обеспечению (ПО) для использования в вычислительном процессе. Допускается изделия вида 2 приводить к изделиям вида 1 путем установления критерия их отказа. Для изделий вида 1 рекомендуется в качестве ПН использовать коэффициент готовности, среднюю наработку на отказ, среднее время восстановления. В качестве основного ПН принят коэффициент сохранения эффективности (производительности), а в качестве дополнительных ПН – средняя наработка до снижения производительности ниже установленного уровня ТО, коэффициент готовности КГ и средняя наработка на неисправность ТН, которая характеризует потребность суперкомпьютера в текущих ремонтах и запасных частях.

В отличие от ПН, показатели отказоустойчивости не регламентированы в нормативных документах.

Показатели безотказности суперкомпьютера косвенным образом характеризуют его отказоустойчивость. Недостатком оценки отказоустойчивости суперкомпьютера по показателям безотказности является то, что эти показатели в явном виде не содержат характеристики устойчивости суперкомпьютера к отказам. Представляется целесообразным для характеристики устойчивости суперкомпьютера к отказам ввести такие показатели, которые бы непосредственно оценивали поведение суперкомпьютера при наличии в нем определенного количества отказавших СЧ и могли характеризовать эффективность придания свойства отказоустойчивости суперкомпьютера.

Наиболее естественной качественной характеристикой свойства отказоустойчивости суперкомпьютера является перечень СЧ суперкомпьютера, к отказам которых он должен быть устойчив. Задание качественных характеристик отказоустойчивости не позволяет сравнивать различные модели суперкомпьютера по этим свойствам и не учитывает реальные возможности обеспечения отказоустойчивости.

Для восстанавливаемых объектов, в том числе и суперкомпьютера, накопление более одного отказа в СЧ имеет очень малую вероятность. Поэтому предлагается оценивать устойчивость суперкомпьютера только к одному отказу в любой его СЧ. Для вывода показателей устойчивости суперкомпьютера к одному отказу в любой его СЧ рассматривалась следующая модель суперкомпьютера:

– суперкомпьютер состоит из n функционально необходимых СЧ; отказы СЧ – события независимые; каждая СЧ характеризуется простейшим потоком отказов с известным параметром;

– отказ любой СЧ суперкомпьютера, в случае отсутствия избыточности для неё, приводит к отказу суперкомпьютера в целом;

– при отказе СЧ суперкомпьютера вводится избыточность, которая с определенной вероятностью успешно компенсирует отказ СЧ (обеспечивает безотказное функционирование суперкомпьютера).

1.11.2. Структурная схема надежности суперкомпьютера

Сложный объект (к числу которых следует отнести и суперкомпьютер) состоит из большого числа СЧ, которые могут находиться во множестве состояний, и тем самым увеличивать число состояний объекта в целом, до размеров, которые плохо поддаются математическому описанию. С целью упрощения математического описания объекта производится его декомпозиция на составные части (элементы). Декомпозиция объекта на элементы производится таким образом, чтобы можно было по показателям надежности элементов вычислить показатели надежно-

сти объекта в целом. В качестве элементов декомпозиции объекта выбираются СЧ объекта, отказы которых независимы друг от друга.

Из элементов декомпозиции объекта строится структурная схема надежности объекта, которая является логико-вероятностной схемой расчета безотказности объекта. Последовательное соединение элементов означает логическое И, а параллельное соединение элементов означает логическое ИЛИ. Структурная схема надежности наглядно представляет взаимосвязь показателей надежности объекта с показателями надежности элементов.

Установлено, что декомпозицию суперкомпьютера следует начинать с оценки влияния отказов СЧ суперкомпьютера на основной ПН. С этой целью все СЧ суперкомпьютера целесообразно разделить на три группы. К группе 1 следует отнести СЧ, отказы которых приводят к снижению производительности суперкомпьютера до нуля или до уровня ниже допустимого. К группе 2 следует отнести СЧ, отказы которых приводят к снижению производительности суперкомпьютера в пределах допустимого уровня. К группе 3 следует отнести СЧ, отказы которых не влияют на производительность суперкомпьютера.

В качестве СЧ суперкомпьютера, отказ которых влияет на производительность суперкомпьютера, рекомендуется использовать технические средства (ТС), а иногда и функциональные части ТС. Например, коммутаторы следует разделять на две части: общую часть всех портов и отдельные порты коммутатора. Такое деление определяется тем, что при отказе общей части коммутатора недоступными ПО суперкомпьютера станут все изделия, подключенные к его портам, а при отказе порта коммутатора недоступным ПО суперкомпьютера окажется лишь изделие, подключенное к этому порту. Порт коммутатора и кабель, связывающий его с изделием, следует включить в это изделие при оценке его надежности. Например, при оценке надежности ВУ кроме надежности вычислительного устройства следует учесть надежность портов коммутаторов вспомогательной и системной сетей, соответствующих им адаптеров и кабелей, соединяющих порты с адаптерами.

Назовем совокупность СЧ группы 1 ядром СК, а совокупность СЧ группы 2 совокупностью вычислительных средств (СВС). При принятых обозначениях, наиболее обобщенная структурная схема надежности суперкомпьютера примет вид, приведенный на рис.1.3.

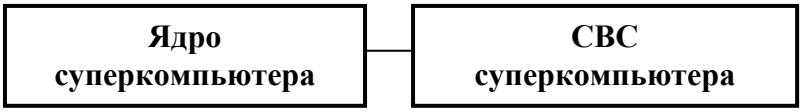

Рис. 1.3. Обобщенная структурная схема надежности суперкомпьютера

В СВС включают совокупность вычислительных узлов и возможно общие части коммутаторов вспомогательной и системной сетей. Коммутаторы включают в СВС только в том случае, если допустимое число недоступных ПО суперкомпьютера ВУ больше числа ВУ, подключенных к данному коммутатору. Исключением является корневой коммутатор вспомогательной сети, ибо при его отказе недоступными станут все ВУ. Остальные СЧ суперкомпьютера, отказы которых влияют на его производительность, относят к ядру суперкомпьютера. Следует заметить, что не всегда однозначно назначение ТС суперкомпьютера в состав ядра КВС. Так например, для привилегированного пользователя кластера отказ файл-сервера не приведет к отказу КВС и его следует отнести к вспомогательному оборудованию. В то время как для непривилегированного пользователя отказ файл-сервера лишит их доступа к КВС и в этом случае файл-сервер следует отнести к ядру КВС.

1.11.3. Математические модели показателей надежности суперкомпьютера

Для расчета показателей надежности и отказоустойчивости необходимо разработать математические модели показателей надежности суперкомпьютера. Условимся под математической моделью показателей надежности понимать совокупность способов и методов формализации расчета показателей надежности. Под расчетным методом определения надежности понимается метод, основанный на вычислении показателей надежности по справочным данным о надежности компонентов и комплектующих элементов объекта. Очевидно, расчет показателей надежности производится по соответствующим формулам, которые выводятся на основании математической модели соответствующего показателя надежности.

В качестве модели для оценки надежности объекта (под объектом будем понимать как суперкомпьютер, так и его СЧ) примем модель состояний объекта в процессе его эксплуатации. При этом допустим, что процесс изменения состояний объекта при его эксплуатации является дискретным марковским процессом с конечным числом состояний и непрерывным временем, а потоки перевода объекта из одного состояния в другие состояния являются простейшими, т.е. имеют показательное распределение.

Допущение того, что потоки отказов и восстановления являются простейшими, общепринято в инженерной практике. В общем случае, распределение времени восстановления элемента не является показательным, однако, заменяя истинное неизвестное нам распределение времени восстановления показательным, мы можем только занизить надеж-

ность, что вполне допустимо, так как при этом истинная надежность будет только немного выше предсказуемой. В большинстве задач прикладного характера замена непуассоновских потоков событий пуассоновскими с теми же интенсивностями приводит к получению решения, которое мало отличается от истинного, а иногда и вовсе не отличается. Специальное моделирование различных задач, проведенное методом Монте-Карло, показало, что в большинстве случаев эта погрешность ограничена 3-5% , и лишь в редких случаях доходит до 10-12%.

Структурная схема надежности ядра суперкомпьютера состоит из последовательных и параллельных (в смысле надежности) элементов. Построить структурную схему надежности из последовательно–параллельных элементов для СВС не представляется возможным. Для оценки ПН СВС предлагается использовать модель состояний СВС в процессе ее эксплуатации (далее модель состояний). Модель состояний графически выглядит в виде ориентированного размеченного графа. Узлами графа являются состояния объекта в процессе его эксплуатации, ориентированные ребра изображают пути перехода из одного состояния в другое, а разметка ребер показывает интенсивности переходов. Используем ранее принятые допущение о том, что процесс изменения состояний СВС при ее эксплуатации является дискретным Марковским процессом с конечным числом состояний и непрерывным временем, а потоки перевода СВС из одного состояния в другие состояния являются простейшими, т.е. имеют показательное распределение.

Заметим, что граф состояний СВС зависит от допустимого значения снижения производительности суперкомпьютера и используемой стратегии восстановления отказавших ВУ в процессе эксплуатации суперкомпьютера. Стратегии восстановления ВУ будут сказываться на показателях надежности СВС и будут изменять структуру графа состояний СВС. Восстановление суперкомпьютера потребуется при отказах его ядра и, может потребоваться, при отказах ВУ в СВУ, или отказах коммутаторов вспомогательной и системной сетей.

Совершенно очевидно, что при отказе ядра суперкомпьютера или коммутатора к восстановлению следует приступать немедленно и проводить его достаточно быстро. Иначе обстоят дела при отказах ВУ в СВС. Здесь возможны различные стратегии восстановления. К числу возможных стратегий восстановления суперкомпьютера при отказах ВУ следует отнести следующие:

а) стратегия 1 – восстановление СВС начинается не сразу после отказа ВУ, оно откладывается на определенный срок (часы, сутки, неделю) и производится в процессе продолжения использования суперкомпьютера по назначению с уменьшенной производительностью и производится в порядке очередности отказов ВУ. При восстановлении ВУ он немед-

ленно вводится в конфигурацию суперкомпьютера и становится доступным для ПО суперкомпьютера. В случае накопления в СВС отказавших ВУ более установленного числа производится экстренное восстановление ВУ.

б) стратегия 2 – восстановление СВС откладывается до накопления определенного количества отказавших ВУ. Затем производится восстановление всех накопившихся отказавших ВУ по одному в рациональном порядке. Для восстановления СВС в исходное состояние прерывается работа суперкомпьютера.

1.11.4. Практическое применение методики расчета показателей надежности

Используя приведенные модели, были получены формулы расчета показателей надежности СВС при различных стратегиях восстановления СВС и допустимости или недопустимости отказов коммутаторов вспомогательной и системной сетей.

Разработанная методика расчета показателей надежности суперкомпьютеров нашла практическое применение при разработке конструкторской документации в виде расчетов надежности на кластер ВМ5100, суперкомпьютерные конфигурации «СКИФ К-500» и «СКИФ К-1000», аппаратно-программный кардиологический комплекс на основе суперкомпьютерных вычислительных модулей для исследования микроциркулярного звена сердечно-сосудистой системы методом биомикроскопии.

1.12. Проведение испытаний и организация серийного производства.

Суперкомпьютерные конфигурации на различных этапах разработки и производства подвергаются ряду проверок и испытаний:

– проверка качества разработки (предварительные, приемочные испытания);

– проверка качества подготовки и освоения серийного производства (квалифицированные, сертификационные испытания);

– проверка стабильности технологического процесса (периодические испытания);

– проверка качества каждой изготовленной прикладной реализации (приемосдаточные испытания).

Предварительные, приемочные, квалификационные, сертификационные и периодические испытания проводятся для каждого типа БВМ, а также для каждого типа базовых конфигураций суперкомпьютерных систем.

На этих испытаниях проверяются параметры базовых исполнений, а также возможность реализации базовых характеристик в диапазонах,

указанных в соответствующей конструкторской докуметации.

На приемосдаточных испытаниях проверяются параметры конкретной прикладной реализации на соответствие требованиям конкретного пользователя.

Для обеспечения проведения всех типов испытаний предусмотрено, наряду с базовым (общесистемным) программным обеспечением, соответствующее тестовое прикладное программное обеспечение.

2. Программа Союзного государства «СКИФ»

2.1. Общие сведения о программе «СКИФ»

Создание собственной передовой высокопроизводительной вычислительной техники имеет для Беларуси важное стратегическое, политическое и экономическое значение. Предпосылки для достижения этой цели были заложены еще во времена Советского Союза, когда Беларусь являлась одним из ведущих разработчиков и производителей компьютеров в странах социалистического лагеря. Белорусские предприятия «НИИЭВМ», МПО ВТ и «Интеграл» во многом определяли стратегию и темпы развития компьютерной отрасли Советского Союза. Предприятие «НИИЭВМ» было одним из основных звеньев в государственной программе СССР по созданию вычислительной техники. Здесь разрабатывались достаточно известные не только в СССР, но и во всем социалистическом лагере вычислительные машины ЕС ЭВМ и системное программное обеспечение для них. Эти вычислительные машины стали основой для компьютеризации отраслей народного хозяйства Советского Союза. Производством занималось другое белорусское предприятие – МПО ВТ.

С разработкой и производством высокопроизводительной техники самым тесным образом связаны такие наукоемкие сектора промышленности как микроэлектроника, оптическое приборостроение, точная механика, средства отображения информации, коммуникационная техника, производство программных продуктов и др. Именно в этих секторах Республика Беларусь сохраняет значительный научно-технический потенциал, поддерживаемый необходимыми фундаментальными и прикладными исследованиями, целевое использование которого позволило в сравнительно короткие сроки при относительно небольших затратах выйти на собственный альтернативный, практически независимый от Запада путь развития отечественной конкурентоспособной высокопроизводительной вычислительной техники, уровень которой соответствует текущим требованиям со стороны широкой категории пользователей.

Научно-технические исследования в области разработки методов и средств моделирования интеллектуальных процессов, создания новых информационных и телекоммуникационных технологий, программно-технических комплексов и систем в Республике Беларусь выполнялись в рамках Государственных программам фундаментальных и прикладных исследований, Государственных научно-технических программ (ГНТП) и отраслевых программ: Государственная программа ориентированных фундаментальных исследований «Теоретические основы новых информационных технологий» (шифр «Инфотех»), Государственная програм-

ма информатизации в Республике Беларусь на 2001-2005 гг., ГНТП «Передовые информационные и телекоммуникационные технологии на 2001-2005 гг.» (ГНТП «Информационные технологии»), Государственная отраслевая научно-техническая программа «Компьютерные технологии проектирования и производства новой продукции» на 2001-2005 гг., Программа работ по развитию Научно-информационной компьютерной сети (НИКС) Республики Беларусь на 2001-2003 гг., Государственная программа «Электронная Беларусь», Государственная программа «Импортозамещение».

В рамках ГНТП «Информатика» выполнялись проекты по созданию высокопроизводительных систем на основе нейрокомпьютеров и нейросетей, которые относятся к вычислительным системам с высоким параллелизмом.

Таким образом, Республика Беларусь в конце девяностых годов прошлого столетья обладала существенным научно-техническим потенциалом для создания, освоения, использования и развития средств высокопроизводительной вычислительной техники. Практические работы в этом направлении были начаты в рамках программы Союзного государства «Разработка и освоение в серийном производстве семейства высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе» (шифр «СКИФ»), которая была утверждена постановлением Исполнительного Комитета Союза Беларуси и России № 43 от 22 ноября 1999 года.

Программа Союзного государства «СКИФ» с учетом продления на один год в соответствии с постановлением Совета Министров Союзного государства от 29.12.2003 года № 29 была рассчитана на 5 лет – 2000-2004 гг.

В соответствии с постановлением Совета Министров Союзного государства от 29.12.2003 года № 29 государственный заказчик-координатор Программы от Республики Беларусь – Национальная академия наук Беларуси, государственный заказчик Программы от Российской Федерации – Министерство промышленности, науки и технологий Российской Федерации. В соответствии с постановлением Совета Министров Союзного государства от 11.10.2004 года № 20 государственный заказчик Программы от Российской Федерации – Федеральное агентство по науке и инновациям.

Головные исполнители Программы – Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ НАН Беларуси) и Институт программных систем Российской академии наук (ИПС РАН).

Система программных мероприятий включает 21 задание, в кото-

рых предусмотрены работы по созданию базовых конструктивно-вычислительных модулей, элементной базы и системного программного обеспечения.

Область применения моделей семейства суперкомпьютеров отражена в заданиях программы, предусматривающих создание ряда пилотных прикладных комплексов, включая инструментальные средства проектирования интеллектуальных суперкомпьютерных систем.

В реализации программных мероприятий участвовало около 20 предприятий от Республики Беларусь и Российской Федерации.

Главная цель программы «СКИФ» – возрождение компьютерной отрасли двух стран, промышленное производство ряда программно-совместимых моделей суперкомпьютеров с широким спектром производительности – до триллионов операций в секунду.

Для достижения этой цели в рамках Программы был реализован комплексный подход, включающий:

– разработку Концепции создания моделей семейства суперкомпьютеров «СКИФ»;

– создание опытных образцов базовых конфигураций суперкомпьютерных систем (БКСС) «СКИФ», разработку литерной конструкторской и программной документации, проведение комплексных предварительных и приемочных (государственных) испытаний опытных образцов БКСС «СКИФ»;

– создание единого информационного и телекоммуникационного пространства участников Программы с возможностью удаленного доступа к суперкомпьютерным ресурсам;

– создание пилотных прикладных комплексов на базе суперкомпьютеров «СКИФ»;

– подготовку и переподготовку кадров для работы с суперкомпьютерными технологиями;

– организацию промышленного выпуска и системы технического обслуживания моделей суперкомпьютеров «СКИФ» в широком диапазоне производительности.

Практическая реализация комплексного подхода состояла из двух этапов.

ЭТАП 1 (2000 – 2002 гг.). Разработка и организация производства моделей первого ряда (**модели Ряда 1**) семейства суперкомпьютеров «СКИФ».

На этапе 1 были отработаны основные концептуальные принципы и реализованы суперкомпьютерные конфигурации, обеспечивающие возможность создания семейства моделей суперкомпьютеров среднего класса (модели Ряда 1) с пиковой производительностью до 300 – 400

миллиардов операций в секунду – суперсерверы «СКИФ». Разработаны комплекты конструкторской и программной документации и созданы образцы базовых конфигураций суперкомпьютерных систем с пиковой производительностью до 100 миллиардов операций в секунду. Проведены государственные испытания опытного образца БКСС кластерного уровня, комплекты документации подготовлены для организации промышленного выпуска суперкомпьютерных конфигураций. Проведена подготовка специалистов по разработке программного обеспечения и использованию суперкомпьютеров «СКИФ», начаты разработки пилотных прикладных систем, созданы необходимые заделы для разработки второго ряда моделей семейства суперкомпьютеров.

ЭТАП 2 (2003 – 2004 гг.). Разработка и организация производства моделей второго ряда (**модели Ряда 2**) семейства суперкомпьютеров «СКИФ» и создание прикладных систем на их основе.

На этапе 2 отработаны основные концептуальные принципы и созданы суперкомпьютерные конфигурации с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду) – модели суперкомпьютеров «СКИФ» Ряда 2. Модели Ряда 2 реализуются на базе кластерной архитектуры, включая сетевые (метакластерные) конфигурации. На этом этапе также разработаны пилотные прикладные программно-аппаратные комплексы на базе суперкомпьютеров «СКИФ».

2.2. Основные результаты комплексной реализации программы «СКИФ»

Важнейший практический результат выполнения программы – выпуск 16 образцов кластерных суперкомпьютеров с пиковой производительностью в диапазоне от десятков миллиардов до нескольких триллионов операций в секунду, которые использовались как для отработки программного обеспечения кластерного уровня, так и для реальных вычислений в интересах предприятий и учреждений России и Беларуси.

Эти модели суперкомпьютеров построены на основе конструктивных компонент – базовых конструктивно-вычислительных модулей кластерного уровня (БКВМ КУ).

На базе БКВМ КУ создаются высокопроизводительные кластерные системы различных конфигураций, представляющие собой самостоятельные вычислительные системы, реализованные в виде тесно связанной сети (кластера) вычислительных узлов, работающих под управлением многопользовательской универсальной операционной системы Linux.

Специфика моделей и их модификаций отражается в эксплуатационной документации. Основополагающие принципы Концепции позволяют создавать прикладные комплексы, соответствующие требованиям

конкретного заказчика, оптимально использовать производственные мощности предприятия-изготовителя с учетом специфики рынка сбыта высокопроизводительных вычислительных систем.

УП «НИИ ЭВМ» разработана конструкторская документация на базовые модули, имеющие самостоятельную поставку, и на базовые конфигурации суперкомпьютерных систем. Конструкторская документация выполнена в едином для всех исполнений суперкомпьютерных конфигураций групповом варианте в соответствии с действующими стандартами.

Групповое построение конструкторской документации адекватно отражает возможности архитектурной идеологии (открытость, масштабируемость), позволяя оптимальным способом организовать серийное производство широкой номенклатуры моделей суперкомпьютеров, наиболее полно удовлетворяющих предъявляемым пользовательским требованиям.

При создании конкретной кластерной конфигурации основной упор делается на выборе технических решений, в частности определении оптимальной конфигурации вычислительного узла, основанного на базе определенного процессорного элемента. При построении кластерных суперкомпьютерных конфигураций семейства «СКИФ» в качестве процессорных элементов для вычислительных узлов кластера были выбраны процессоры с лучшими показателями цена/производительность для предполагаемого круга прикладных задач. Технологией производства высокопроизводительных процессорных элементов владеют только несколько корпораций в мире (Intel, AMD, IBM, NEC). Основной задачей программы «СКИФ» являлось освоение суперкомпьютерных вычислительных технологий и предоставление отечественным потребителям возможности использования современной высокопроизводительной вычислительной техники. Именно в таком направлении развиваются и западные корпорации, вкладывающие огромные средства в построение центров коллективного использования суперкомпьютерных ресурсов.

В 2003 году в рамках программы «СКИФ» был создан кластер «СКИФ К-500» с пиковой производительностью 716,6 Gflops на базе процессоров Intel Xeon 2.8 Ghz. Создание этого кластера явилось качественно новым результатом, позволившим вплотную приблизиться к терафлопному диапазону.

Реальная производительность кластера в 425,2 миллиарда операций в секунду, достигнутая на тесте LinPack, стала основанием для его включения под номером 407 в 22-ой выпуск списка 500 самых производительных компьютерных систем в мире top500 (рис.2.1).

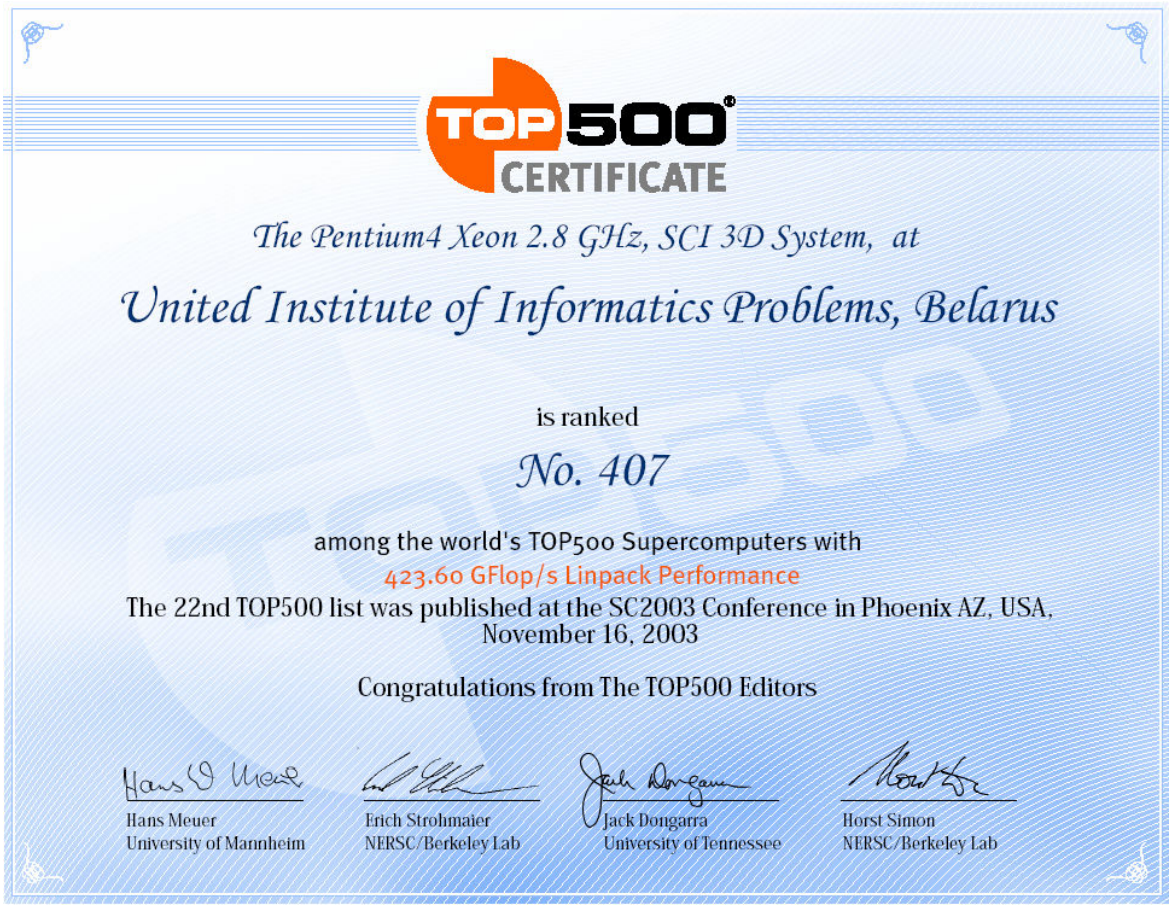

Рис. 2.1. Международный сертификат 22 выпуска списка top500

Включение кластера «СКИФ К-500» в список пятисот наиболее мощных вычислительных установок в мире означало достижение уже в 2003 году важного прямого политического эффекта – Республика Беларусь и Россия наравне с США, Японией и еще несколькими странами стали обладателями критической суперкомпьютерной технологии, повысив свой престиж, как разработчика этой технологии.

В 2004 году создан кластер «СКИФ К-1000» с пиковой производительностью 2,5 триллиона операций в секунду. Большинство из технических решений, использованных в суперкомпьютере «СКИФ К-1000», являются передовыми для суперкомпьютерной отрасли, в частности, были использованы 64-битовые процессоры AMD Opteron 248 (2200 MHz).

9 ноября 2004 года суперкомпьютерная конфигурация «СКИФ К-1000» включена в очередной 24-й выпуск списка top-500 под номером 98 (рис.2.2). В первую сотню рейтинга от 9 ноября 2004 года вошли суперкомпьютерные установки 16 стран, из них установки собственных разработчиков представили только 4: США, Япония, Китай и Союзное го-

сударство. Суперкомпьютер «СКИФ К-1000» на 01.01.2005 г. являлся самым мощным компьютером на территориях СНГ и Восточной Европы. Создание кластера «СКИФ К-1000» с пиковой производительностью 2,5 триллиона операций в секунду подтвердило выход на собственный путь развития конкурентоспособной высокопроизводительной вычислительной техники, уровень которой соответствует современным мировым требованиям со стороны широкой категории пользователей.

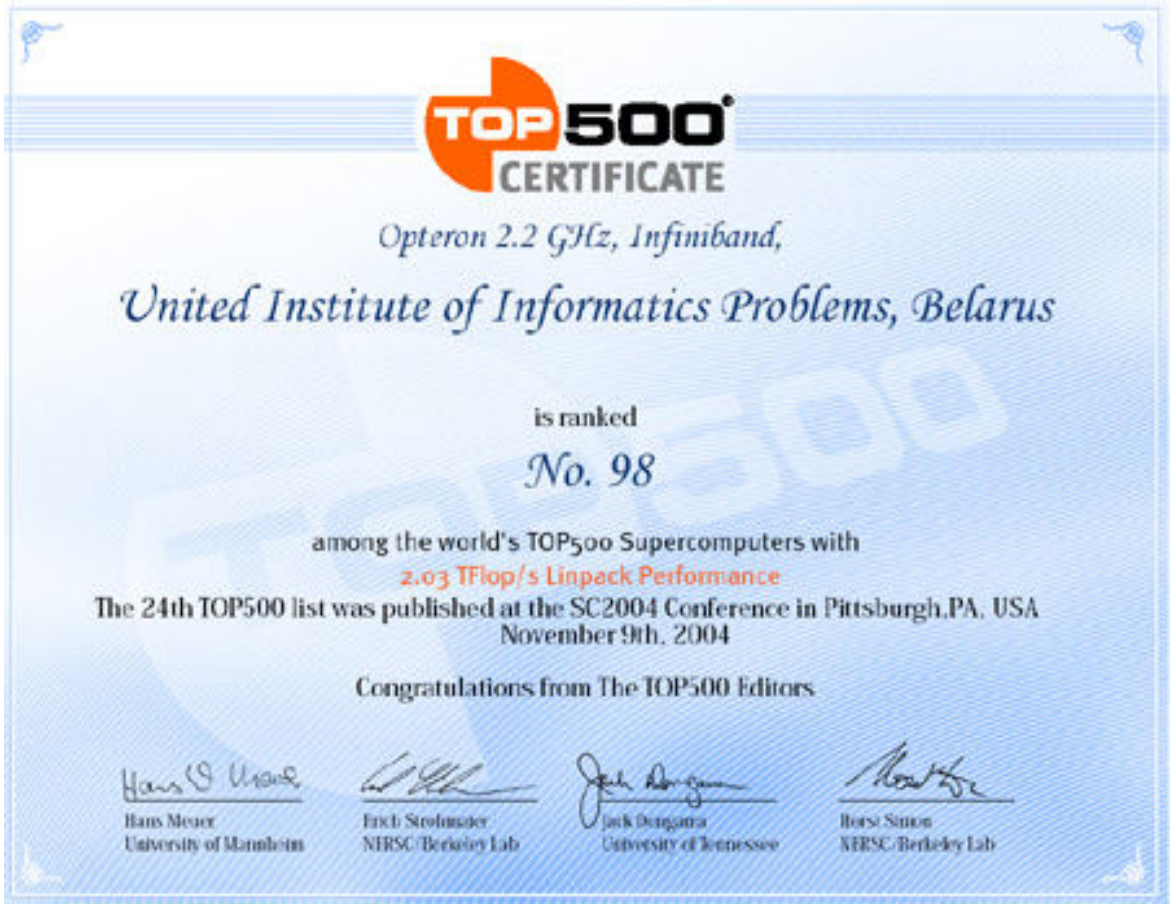

Рис.2.2. Международный сертификат 24 выпуска списка top500

17-19 ноября 2004 года были проведены приемочные испытания суперкомпьютерных конфигураций «СКИФ К-500» и «СКИФ К-1000». Комплектам конструкторской и программной документации в установленном порядке присвоена литера «О1». Предприятие УП «НИИ ЭВМ» по состоянию технической оснащенности и отработанности технологического процесса было признано готовым к производству кластеров типа «СКИФ К-500» и «СКИФ К-1000». По состоянию на 01.01.2005 г. четыре белорусских суперкомпьютера входили в текущую редакцию списка 50 самых мощных компьютеров СНГ, занимая 1 (СКИФ К-1000), 4

(СКИФ К-500), 22 и 32 места.

В ходе выполнения программы «СКИФ» разработаны системное программное обеспечение и языковые средства для моделей суперкомпьютеров «СКИФ». Отличительной особенностью суперкомпьютеров «СКИФ» является использование отечественной оригинальной системы поддержки параллельных вычислений – Т-системы, реализующей автоматическое динамическое распараллеливание программ и имеющей ряд технологических находок, не имеющих аналогов в мире.

Создана телекоммуникационная сеть, объединяющая участников программы «СКИФ», потенциальных потребителей и вычислительные ресурсы суперкомпьютерных конфигураций «СКИФ».

Разработан ряд пилотных прикладных программно-аппаратных комплексов с использованием технологий высокопроизводительных вычислений, которые были проведены на суперкомпьютерных конфигурациях «СКИФ»:

1) Программно-аппаратный комплекс для численного моделирования процессов в задачах радиационной газодинамики.

2) Экспериментальная вычислительная система на базе специализированных вычислительных модулей однородной вычислительной среды (ОВС) с параллельной архитектурой для обработки и распознавания рукописных символов.

3) Макет экспериментальной системы для обработки радиолокационных сигналов, распознавания объектов и моделирования широкополосных радиолокационных сигналов с использованием вычислительных модулей ОВС.

4) Программный комплекс моделирования фильтрационных процессов в условиях блочных нарушений породного массива, обусловленных воздействием крупномасштабных техногенных нагрузок.

5) Программный комплекс расчета зонной структуры твердых тел.

6) Макет программного комплекса оптимизации частотно – территориальных планов радиоэлектронных средств.

7) Аппаратно-программный кардиологический комплекс на основе вычислительных модулей «СКИФ» для исследования микроциркуляторного звена сердечно-сосудистой системы методом биомикроскопии. В 2004 году в РНПЦ «Кардиология» Министерства здравоохранения Республики Беларусь завершена клиническая апробация данного комплекса, реализующего запатентованный метод биомикроскопии. На кардиологический комплекс получено 4 патента Республики Беларусь на полезную модель, поданы 1 заявка на получение патента Республики Беларусь на изобретение и 1 заявка на получение патента Российской Федерации на изобретение.

8) Экспериментальный программно-информационный комплекс

оперативного прогноза ветрового переноса загрязнений при чрезвычайных ситуациях.

9) Макет системы оперативной идентификации личности по акустико-фонетическим признакам на основе произвольного фрагмента речи.

В рамках программы были проведены мероприятия по подготовке и переподготовке кадров для использования новых информационных технологий на базе суперкомпьютеров:

– разработаны рабочие и методические материалы для обучения и переподготовки пользователей по MPI и ОС Linux;

– подготовлены практические и наглядные пособия для базовых курсов по MPI и ОС Linux и проведено чтения курсов тестовой группе слушателей;

– разработаны материалы учебного пособия и программы курса «Практическое использование средств разработки приложений в ОС «Linux» в составе «Методы и средства параллельной организации процессов», «Практическое использование высокоуровневых средств параллельного программирования MPI, PETSC и DIPC», «Практическое использование средств разработки приложений в ОС Linux» и проведено чтения курсов тестовой группе слушателей;

– разработана русифицированная документация на библиотеку параллельного программирования PETSC;

– выпущено учебное пособие по технологиям параллельного программирования «Средства параллельного программирования для ОС Linux»;

– разработаны русскоязычные версии документов «Руководство пользователя LAM/MPI» и «Руководство по установке LAM/MPI» для системы параллельного программирования LAM/MPI;

– разработана программа специального курса по математическим основам и современным технологиям высокопроизводительных вычислений, занятия по которой велись на кафедре вычислительной математики МГУ им. М.В. Ломоносова в 2002–2004 гг.;

– разработана программа основного университетского курса «Высокопроизводительные вычисления», чтение которого ведется в Московском Государственном Индустриальном Университете (в качестве программно-аппаратной базы для практических занятий курса используются (в режиме удаленного доступа) суперкомпьютерные установки «СКИФ»);

– проведена летняя студенческая школа с 25 июня по 10 июля 2004 года по тематике «Высокопроизводительные вычисления на кластерных и GRID-архитектурах».

На сайтах http://www.skif.bas-net.by и http://skif.pereslavl.ru отражены ход выполнения и реализации программы Союзного государства «СКИФ».

2.3. Экономический эффект от реализации программы «СКИФ»

Комплексное выполнение мероприятий программы завершено в 2004 году. Одним из важнейших результатов программы является формирование команды исполнителей с тесными кооперационными связями, способного решать технические задачи любой сложности в суперкомпьютерной отрасли. Создание суперкомпьютерных конфигураций «СКИФ К-500» и «СКИФ К-1000» позволяет утверждать, что сегодняшний научный и технологический уровень коллектива исполнителей программы соответствует мировому уровню и имевшееся отставание в этой части, фактически, ликвидировано.

Комплексная реализация мероприятий по развитию суперкомпьютерных технологий позволит выйти на собственный путь развития конкурентоспособной высокопроизводительной вычислительной техники, уровень которой будет соответствовать прогнозируемым требованиям со стороны широкой категории пользователей.

Конкретные практические результаты, достигнутые по программе «СКИФ» в 2000 – 2004 годах, подтверждают высокую квалификацию российских и белорусских специалистов и высокую эффективность использования средств по программе. Так, в выпущенном образце «СКИФ К-1000» достигнуты:

– пиковая производительность 2,5 TFlops, реальная на задаче Linpack –2,0 TFlops, что позволило данной установке войти в первую сотню (№ 98) престижного рейтинга top-500;

– удельная стоимость/производительность: менее $1 000 000 за TFlops, что в разы (и даже на порядок) лучше импортных аналогов (HP, IBM и др.);

– высокий КПД (отношение реальной производительности к пиковой) – 80% .

Важнейшим практическим внедрением результатов реализации программы «СКИФ» является создание в ОИПИ НАН Беларуси суперкомпьютерного центра коллективного пользования с возможностью удаленного доступа к его вычислительным ресурсам (рис. 2.3).

Рис. 2.3. Суперкомпьютерного центра коллективного пользования

Создание суперкомпьютерного центра в ОИПИ НАН Беларуси для развития и внедрения в республике наукоемких информационных технологий позволяет предоставлять услуги для решения наукоёмких задач, возникающих в промышленности и в других областях народного хозяйства, требующих компьютерных и информационных ресурсов, владение которыми недоступно или экономически нецелесообразно для отдельных организаций. Ориентировочно прямой экономический эффект только от создания суперкомпьютерного центра можно оценить исходя из стоимости его суперкомпьютерных установок «СКИФ К-500» и «СКИФ К-1000», которая составила около 500 тыс.долларов США и 2000 тыс. долларов США соответственно. В зависимости от аппаратно-программной конфигурации это в 3-10 раз меньше, чем аналогичные разработки зарубежных производителей.

Интегральный экономический и политический эффект от комплексной реализации Программы «СКИФ» обеспечивается тем, что ее результаты будут способствовать форсированному технологическому перевооружению ключевых отраслей промышленности стран-участниц Союзного государства, их реформированию с целью достижения мирового уровня качества продукции на базе новейших наукоёмких инфор-

мационных технологий и суперкомпьютерных конфигураций «СКИФ».

Практическое использование результатов программы «СКИФ» и развитие в Республике Беларусь суперкомпьютерного направления позволит получить:

а) экономический эффект:

– предоставление вычислительных ресурсов Суперкомпьютерного центра коллективного пользования ОИПИ НАН Беларуси предприятиям республики;

– поставки суперкомпьютерной техники и программного обеспечения собственного производства в заинтересованные предприятия и организации Беларуси и России;

– сокращение средств на импорт аппаратных средств параллельных высокопроизводительных вычислений;

– сокращение средств на закупку программного обеспечения для организации параллельных высокопроизводительных вычислений (годовые потребности Беларуси и России оцениваются в несколько миллионов долларов);

– экспортные возможности поставок собственных аппаратных и программных средств для организации параллельных высокопроизводительных вычислений в развивающиеся страны, которые проявляют большой интерес к обладанию суперкомпьютерными технологиями, и испытывают ограничения в доступе к таким технологиям со стороны высокоразвитых стран.

Результаты программы «СКИФ» (2000–2004 гг.) на сегодняшний день широко используются в интересах различных отраслей Российской Федерации и Республики Беларусь.

б) социальный эффект:

– поддержка собственных разработчиков аппаратных и программных средств для параллельных высокопроизводительных вычислений;

– создание рабочих мест в наукоемких отраслях, сдерживание «утечки мозгов», подготовка и переподготовка кадров для суперкомпьютерной отрасли.

Ключевым фактором успешного развития перспективного сектора национальной экономики – разработки и экспорта информационных технологий (ИТ) – является подготовка, повышение квалификации и переподготовка кадров. Стимулирование высококлассного ИТ образования и последующего закрепления специалистов в этой области должно быть направлено на решение стратегически важных задач: укрепление и развитие национальной экономики, защиту государственных и привлечение дополнительных инвестиций в образование и науку. Качественное массовое ИТ образование повышает конкурентоспособность национальных

ИТ продуктов и услуг на мировых рынках, в целом благоприятно сказывается на конкурентоспособности национальной экономики.

В настоящее же время фактически отсутствует система повышения квалификации и переподготовки кадров в области суперкомпьютерных технологий. Большинство ВУЗ-ов не располагают необходимыми материально-технической базой и кадровым потенциалом в области суперкомпьютерного образования, отсутствует отвечающая современным требованиям учебная и методическая литература в области суперкомпьютерных технологий. Поэтому растущий спрос на повышение квалификации и переподготовку кадров в этой области не удовлетворяется должным образом, при этом особую озабоченность вызывает отсутствие учебных профилей повышения квалификации и переподготовки специалистов в области суперкомпьютерных технологий. Решение же этих проблем имеет особое значение для развития разработки и экспорта информационных технологий. Именно приоритетное развитие этих направлений позволит создать в республике экономику, ориентированную не только на импорт и потребление информационных технологий и услуг, но и на их производство и экспорт.

В процессе выполнения программы «СКИФ» было подготовлено более 100 высококвалифицированных специалистов, которые имеют достаточный опыт для успешного выполнения любых по сложности разработок и консалтинговых услуг в области высокопроизводительных вычислений.

в) политический эффект:

– обладание критической технологией;

– уменьшение зависимости Беларуси от внешних поставок суперкомпьютерного оборудования и программного обеспечения;

– повышение престижа страны, как разработчика суперкомпьютерных технологий.

г) технологический эффект; дополнительные (вторичные, косвенные) эффекты:

– использование суперкомпьютерных технологий в различных областях науки, техники и промышленности позволит осуществить технологический прорыв в этих важнейших направлениях, что приведет к дополнительным (вторичным, косвенным) экономическим, политическим и социальным эффектам.

3. Кластерные конфигурации «СКИФ» Ряда-1

Как уже отмечалось, на этапе 1 реализации программы «СКИФ» были отработаны основные концептуальные принципы и реализованы кластерные конфигурации, обеспечивающие возможность создания семейства моделей «СКИФ» среднего класса (модели Ряда 1) с пиковой производительностью до 300 – 400 миллиардов операций в секунду – суперсерверы «СКИФ». В этот период были разработаны и изготовлены несколько кластерных установок семейства «СКИФ».

3.1. Базовая конфигурация системы кластерного уровня «Первенец»

За первые три месяца реализации Программы к концу декабря 2000 года были созданы два образца кластерных конфигураций «СКИФ» («Первенцы», рис. 3.1), один из которых был установлен в Минске в УП «НИИЭВМ», второй – в Переславле-Залесском в ИПС РАН. Эта работа явилась ярким примером сотрудничества, примером 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 УП «НИИЭВМ» в сжатые сроки была разработана рабочая конструкторская документация и базовые конструктивы. На своем опытном производстве УП «НИИЭВМ» выпустило 8 стоек и 32 корпуса для узлов вычислительной системы. Затем конструктивы были доставлены в Москву в ОАО «НИЦЭВТ», где уже был завершен подбор и закупка комплектующих. Здесь была произведена сборка всей аппаратной части и первичное тестирование установок «СКИФ». В ИПС РАН были произведены установка базового комплекта программного обеспечения кластерного уровня, окончательная наладка, тестирование и замеры показателей производительности и других технических характеристик. Кроме того, на первые образцы суперкомпьютера «СКИФ» была проведена установка первой прикладной системы, разработанной в Санкт-Петербурге в Институте высокопроизводительных вычислений и информационных систем (ИВВиИС). Эта прикладная система предназначена для проектирования химических реакторов.

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. табл. 3.1) соответствовал уровню эмбарго того времени на ввоз в Россию и Беларусь высокопроизводительной вычислительной техники, ввоз такой техники требовал оформления особого

разрешения.

Рис. 3.1. Внешний вид первых двух образцов кластеров семейства «СКИФ»

Таблица 3.1
Кластерная установка «Первенец»: Основные технические характеристики семейства

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 Linpak) 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

3.2. Кластер «СКИФ» ВМ-5100

Второй опытный образец семейства «СКИФ» - ВМ-5100 – был выпущен белорусской стороной в 2001 году (см. рис. 3.2).

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 Linpak) 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.

Рис. 3.2. Суперкомпьютерная установка ВМ-5100 семейства «СКИФ»

Его отличительными чертами являются более высокопроизводительные (по отношению к первым образцам) процессоры Intel Pentium III-1400 MHz и более компактные конструктивы: 2U вместо 3U. Благо-

даря процессорам Intel Pentium III-1400 MHz на этом образце были достигнуты более высокие пиковая и реальная производительность: 48 Gflops и 29 Gflops, соответственно.

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

3.3. «Первенец-М»: модернизация первого образца суперкомпьютера 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 суперкомпьютеров семейства «СКИФ» не только решения от фирмы Intel, но и решения от AMD. У ИПС РАН установились отношения партнерства с сотрудниками московского представительства AMD. В результате от AMD были получены несколько комплектов процессоров, памяти и материнских плат. В то же время ОАО «НИЦЭВТ» предоставил ИПС РАН адаптеры SCI D335. 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» (рис.3.3) существенную помощь оказала московская фирма «Storus», которая предоставила платы SCI D335 с целью замены старых, использовавшихся ранее плат SCI D311/312. Старые платы впоследствии были использованы для сборки кластера с условным названием «Студент». В результате модернизации при минимальных затратах удалось улучшить основные технические характеристики системы в два-три раза, а производительность – почти в пять раз. В свое время «Первенец-М» являлся основной (по мощности) вычислительной установкой в Переславле-Залесском. Был организован удаленный доступ к ее ресурсам для более чем 100 пользователей, решавших свои задачи на этой установке.

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 Linpak) 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-тор, 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.

Рис. 3.3. Кластерная установка «Первенец-М», Переславль-Залесский, ИПС РАН, 2002 г.

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

Из электронных компонент, высвободившихся после модернизации установки «Первенец», был собран девятиузловой вспомогательный кластер (рис. 3.4). Кластер 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 работало много студентов и из МГУ, и из университета города Переславля. Кроме того, установка «Студент» использовалась для отладки программного обеспечения того времени, прежде чем оно устанавливалось 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 Linpak) 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-тор, SCI SCI D311/312
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	MiniTower

Рис. 3.4. Вспомогательная кластерная установка «Студент»

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

Кластер НИИ механики МГУ (рис. 3.5) по аппаратным решениям этот кластер является близким аналогом установки «Первенец-М».

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 Linpak) 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-тор, 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

Рис. 3.5. Кластерная установка в НИИ механики МГУ

В том же 2002 году для отработки технологии Myrinet в Минске (ОИПИ НАН Беларуси и УП «НИИЭВМ») был разработан кластер с условным названием «Myrin» (рис. 3.6).

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 Linpak) 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

Рис. 3.6. Внешний вид установки «Myrin»

3.6. Старшая модель семейства «СКИФ» Ряда-1 ЕС1710.03

В 2003 году в рамках Программы «СКИФ» было запланировано создать российский образец высокой производительности (около 400 Gflops), который получил специальное обозначение – ЕС1710.03 (рис. 3.7).

Этот образец создавался для отработки типового решения для установок с высокой производительностью. К его сильным сторонам относятся весьма высокая для своего времени производительность (пиковая производительность 0,43 Tflops), компактность (образец собран в одной стойке, форм-фактор 1U) и, самое главное, системная сеть в установке ЕС1710.03 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 N335 (рис. 3.8) – 36 плат для 36 узлов вычислительной системы 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.

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 Linpak) 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-тор, SCI, N335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	В составе NAS 480 GB

Рис. 3.7. Кластер ЕС1710.03, НИЦЭВТ, Москва, ноябрь 2003 г.

Рис. 3.8. Адаптер системной сети SCI N335, выпускаемый ОАО НИЦЭВТ

4. Суперкомпьютерные конфигурации «СКИФ» терафлопсного диапазона (модели Ряда-2)

4.1. Общие принципы создания моделей семейства «СКИФ» Ряда-2

На этапе 2 реализации программы «СКИФ» (2003-2004 гг.) были отработаны основные концептуальные принципы и созданы суперкомпьютерные конфигурации для отработки принципов построения семейства моделей суперкомпьютеров с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду) – модели суперкомпьютеров «СКИФ» Ряда 2. Модели Ряда 2 реализованы на базе кластерной и гибридной архитектур, включая сетевые (метакластерные) конфигурации. На этом этапе были созданы суперкомпьютерная кластерная конфигурация терафлопсного диапазона производительности, сетевые (метакластерные) суперкомпьютерные конфигурации и прикладные программно-аппаратные комплексы на базе суперкомпьютеров «СКИФ». Можно выделить следующие основные отличительные признаки моделей Ряда 2:

– комплексная реализация архитектурных концептуальных принципов, позволяющая создавать конфигурации на базе кластерной, потоковой и гибридной архитектур, включая сетевые (метакластерные) конфигурации;

– широкий спектр производительности – до триллионов операций в секунду;

– использование передовых технологий – вычислительные платформы на базе 64-разрядных процессоров, конструктивы 1U;

– сетевые решения – Infiniband, SCI (Scalable Coherent Interface).

– разработка принципов повышения отказоустойчивости (живучести) суперкомпьютерных конфигураций;

– расширенное программное обеспечение, включая сервис для создания прикладных комплексов для конкретных пользователей.

Все современные кластерные суперкомпьютеры устроены приблизительно одинаково. Но реальная производительность такого суперкомпьютера зависит от того, насколько удачно взаимодействуют между собой аппаратная и программная компоненты. Только после тщательного анализа различных аспектов вычислительного процесса определяется окончательный вариант реализации суперкомпьютера. Принимая во внимание факт, что аналогичные по мощности вычислительные установки могут отличаться по стоимости на порядок, подобные исследования приобретают особое значение. В связи с этим для создания супер-

компьютерных конфигураций «СКИФ» Ряда-2 необходимо было провести исследования по целому ряду направлений:

1) Сравнительный анализ перспективных 64-разрядных вычислительных платформ.

2) Оценка результатов тестов производительности при выборе узлов суперкомпьютера.

3) Выбор системной и вспомогательной сетей для суперкомпьютерных конфигураций.

4) Выбор конфигурации систем внешней памяти.

5) Оценка методов установки операционной системы Linux для суперкомпьютеров «СКИФ».

6) Выбор перспективных конструктивно-технологических решений.

7) Предварительная оценка затрат на создание высокопроизводительных кластеров в условиях отечественного производства и др.

Достигнутые в рамках программы «СКИФ» практические результаты показали эффективность проведенных предпроектных исследований и техническую корректность принятых организационно-технических решений.

Ключевой работой в 2003 году было создание экспериментального образца суперкомпьютерной конфигурации кластерного уровня «СКИФ-К-500» для отработки основных системных принципов создания моделей Ряда 2 суперкомпьютеров «СКИФ» с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду). В экспериментальном образце реализована архитектура 3D тор и использованы современные вычислительные платформы с технологией Hyper Threading. Пиковая производительность кластера «СКИФ К-500» – более 700 миллиардов операций в секунду.

Основные цели и задачи создания экспериментальной кластерной конфигурации «СКИФ К-500»):

– практическая отработка принципов реализации кластеров с 3D-архитектурой;

– освоение возможностей технологии Hyper Threading;

– отработка конструктивно-технологических решений на типоразмерах 1U;

– отработка тепловых режимов;

– исследование и отработка принципов мониторинга состояния узлов кластера;

– исследование, анализ и тестирование кластера в различных режимах работы;

– отработка метакластерных конфигураций и режимов удаленного

доступа к ресурсам кластера;

– отработка приложений и демонстрационных задач;

– отработка прикладных комплексов для конкретных пользователей;

– отработка концептуальных принципов создания суперкомпьютерного центра коллективного пользования.

В 2004 году экспериментальный образец суперкомпьютерной конфигурации кластерного уровня «СКИФ К-500» был доработан до опытного образца, в том же году была создана старшая модель семейства «СКИФ» – суперкомпьютер «СКИФ К-1000» с пиковой производительностью более 2,5 триллионов операций в секунду.

4.2. Сравнительный анализ 64-разрядных вычислительных платформ

Одной из важнейших проблем (возможно, ключевой проблемой) при создании кластера триллионного диапазона является выбор вычислительной платформы для реализации его вычислительных узлов. В свою очередь, ключевым моментом при выборе вычислительной платформы является принятие решения по выбору архитектуры процессора и его типа.

Появление 64-разрядных процессоров является решением, направленным на самую прибыльную часть мирового рынка серверов – системы старшего класса (число процессоров более 4, стоимость – свыше 25 тысяч USD). На эти системы в стоимостном выражении приходится около 60% мирового рынка серверов. Рынок систем старшего класса включает в себя как часть, относящуюся к традиционным коммерческим приложениям (СУБД, системы управления ресурсами предприятий и т. д.), так и часть, связанную с научно-техническими (в т.ч. высокопроизводительными) приложениями.

Для «коммерческой» части характерен высокий консерватизм в выборе программно-аппаратных платформ. Поэтому необходимость переноса ПО на новую 64-разрядную платформу I A-64 замедляет, например, продвижение систем на базе Itanium 2. К тому же целочисленная производительность этого процессора уступала не только 64-разрядным IBM Power4 и HP Alpha 21364, но и 32-разрядным Pentium 4, Xeon, а также AMD Athlon MP и Opteron. Следует также учитывать, что в этой области рынка вычислениями с данными в формате с плавающей точкой, где Itanium 2 очень хорош, практически не интересуются.

В последнее время в этих приложениях уменьшилась доля RISC-серверов, но это связано не с увеличением доли систем на базе Itanium 2, а с переходом на более дешевые сервера на базе процессоров Xeon, обладающие хорошим соотношением цена/производительность. Главный

недостаток таких реализаций – отсутствие 64-разрядной адресации. Поэтому для приложений, требующих большой емкости оперативной памяти, они не подходят. Для таких приложений хорош 64-разрядный AMD Opteron (представлен в апреле 2003 года). Но весь путь, связанный с переносом ПО на платформу Opteron (для тех, кто решит отказаться от RISC), у AMD в 2003 году только начинался. В то же время, в Intel считали, что для архитектуры IA-64 теперь уже имеется все необходимое ПО, обеспечивающее работу на этой платформе основных СУБД (Oracle, IBM DB2, Microsoft SQL Server).

Для научно-технических приложений применение Itanium 2 достаточно перспективно. Эта область не столь консервативна и имеет давние традиции обеспечения мобильности приложений для разных аппаратных платформ. По вычислениям с плавающей точкой Itanium 2 сначала был в лидерах, затем его обогнали Alpha 21364 и Power4+, но новый Madison вернул все на круги своя. Тенденция в этой части рынка однозначна – доля кластеров быстро растет. В узлах кластеров используются разные процессоры, включая Itanium 2. Однако чаще пока применяются процессоры Pentium 4, Xeon и Athlon. Для приложений же, требующих большой емкости памяти (если, например, память узла, доступная приложению, должна быть больше 4 Гбайт), требуются 64-разрядные процессоры. Основные конкуренты – Itanium 2 и Opteron.

Таким образом, конкуренция среди процессорных платформ на рынке высокопроизводительных систем весьма велика. В одних случаях лучше Xeon, в других – Itanium 2, Opteron или IBM Power4 (например, для DB2). Очень сложно привести все показатели к одному знаменателю. У понятия «конкурент» слишком много переменных – конфигурация, стоимость, соотношение стоимости и производительности, используемые приложения и т. д. При этом фирмы приводят различные довольно убедительные доводы в пользу своих решений. Например, Intel считает, что сочетание IA-32 и Itanium – это мощная идея на уровне совокупности аппаратных систем с разной архитектурой. Для 32-разрядных платформ выбран путь увеличения производительности за счет использования технологии Hyper-Threading. Этот путь позволяет поднять производительность на 20-25%, а на некоторых приложениях-даже на 30%.

В конце 2003 года можно было сделать вывод, что среди 32-разрядных процессоров Intel Xeon является предпочтительным вариантом. Intel Xeon обладает рядом преимуществ по сравнению с альтернативными вариантами, в частности он использует перспективную технологию Hyper Threading для увеличения производительности параллельно выполняющихся задач. Кроме того, существует возможность построения вычислительных SMP-узлов с использованием до 4-х процессоров Xeon.

Этот процессор уже имел хорошую репутацию среди создателей кластерных систем.

Если же рассматривать вариант вычислительного кластера на базе 64-х разрядных процессоров, то необходимо тщательное обоснование выбора из возможных вариантов (например, Intel Itanium 2 или AMD Opteron).

При анализе возможных вариантов необходимо учитывать, что архитектура IA-64 выходит за пределы подходов RISC и CISC путем применения явных команд параллельных вычислений (EPIC – Explicit Parallel Instruction Computing), объединяя вычислительные ресурсы с работой компиляторов, что позволяет осуществить явное распараллеливание на процессоре. Архитектура с высокой параллельностью вычислений EPIC позволяет процессору выполнять до 20 операций за один такт. Внутренние ресурсы процессора комбинируются с предикативными и спекулятивными алгоритмами, что позволяет осуществлять оптимизацию высокопроизводительных приложений, запускаемых в большинстве операционных систем, включая версии Microsoft Windows, HP-UX и Linux.

Архитектура IA-64 не является ни 64-х разрядным расширением 32-х разрядной архитектуры x86 компании Intel, ни переработкой 64-разрядной архитектуры PA-RISC компании HP. IA-64 представляет собой нечто абсолютно новое: архитектуру, использующую длинные слова команд (long instruction words – LIW), предикаты команд (instruction predication), устранение ветвлений (branch elimination), предварительную загрузку данных (speculative loading) и другие способы максимального извлечения параллелизма из кода программ.

Теоретическая пиковая производительность процессора при работе с числами с плавающей точкой 64-х разрядной точности составляет 4 Гфлоп/с при тактовой частоте 1 ГГц. При использовании же 32-х разрядной точности пиковая производительность удваивается.

Совместимость с 32-х разрядным кодом обеспечивается специальным блоком декодирования и управления IA-32. По сути, выполнение 32-х разрядного кода происходит в режиме эмуляции, что отрицательно сказывается на производительности.

Для перехода на 64-разрядную архитектуру компания AMD предложила расширение существующей архитектуры i386 в отличие от Intel с ее кардинально новым решением IA-64. При этом процессоры AMD сохраняют непосредственную совместимость с 32-х разрядными приложениями, в то время как процессоры на базе IA-64 вынуждены использовать специальный режим эмуляции, заметно снижающий производительность таких приложений.

В процессорах с ядром Hammer в 64-х разрядном режиме применяется «плоская» модель памяти, количество регистров общего назначения

расширено до 16. Процессор имеет несколько режимов работы: кроме стандартных, существовавших еще в i386, введен особый режим – Long mode. Когда он включен (бит LME выставлен в единицу), существует два частных режима работы процессора. В одном из них процессор находится в режиме совместимости, во втором – в полном 64-х разрядном режиме. Два частных режима необходимы для одновременной поддержки (в случае использования 64-х разрядной ОС) как 32-х (в этом случае и нужен режим совместимости), так и 64-х разрядных приложений. При этом переключения частных режимов Long mode происходят весьма быстро, в отличие от переключения режимов работы процессора.

В принципе, необходимо учитывать, что решение на новых 64-х разрядных процессорах сопровождается определенными рисками, поэтому нужно тщательно обосновать, действительно ли используемым приложениям необходима 64 разрядность.

4.3. Оценка результатов тестов производительности при выборе узлов суперкомпьютера

В качестве тестов производительности использовались тесты компании Standard Performance Evaluation Corporation (SPEC), созданной в 1988 году с целью разработки и поддержки широкого спектра программ для измерения производительности компьютерных систем. Интернет-сайт компании находится по URL http://www.spec.org. Там же представлены данные тестирования различных платформ, выполненные специалистами компании и производителями техники. Для тестирования используется тест SPEC CPU2000, ориентированный на измерение производительности центрального процессора (или процессоров) компьютеров. Однако, учитывая факт, что компьютер состоит не только из процессора, более правильно было бы считать, что производится тестирование вычислительной системы в целом с использованием задач с большой интенсивностью вычислений. Все тесты предоставляются как набор исходных текстов программ на языках C, C++ и Fortran. При этом, рассматривая результаты тестов, необходимо учитывать также выбор компиляторов, а также опции компиляции и создания исполняемых файлов.

В целом тесты SPEC CPU2000 разделены на две группы: CINT2000 для тестирования скорости работы с целочисленными значениями и CFP2000 для тестирования скорости работы системы при вычислениях с плавающей точкой. Тесты предоставляют скорее «точку отсчета» для оценки производительности различных систем, нежели истину «в последней инстанции», так как результаты работы конкретных программ могут значительно отличаться от ожидаемых результатов, сделанных на основе только данных тестов. Свой отпечаток могут наложить скорость работы с памятью, особенности реализации алгоритмов (многопоточ-

ность) и прочие факторы. Необходимо отметить также, что в тестах CPU2000 предусмотрено деление на base и peak тесты, в первом случае накладываются более строгие ограничения на компиляцию кода, во втором разрешается вносить больше оптимизаций на этапе получения исполняемого кода. При проведении тестов необходимо также выбирать метрику – существуют две возможности – speed и rate. В первой производится оценка скорости выполнения одной задачи и отражается результат в процентах от скорости базовой системы (например, значение 120 означает, что система работает в 1.2 раза быстрее базовой системы). Базовая система – это рабочая станция Sun Ultra 5/10 (процессор UltraSPARC II с тактовой частотой 300 МГц). На данной машине прогон всех тестов CPU2000 занимает примерно двое суток (48 ч). Вторая метрика дает ответ в количестве выполняемых задач в час. Соответственно метрику speed лучше использовать для оценки однопроцессорных систем, в то время как метрику rate – для многопроцессорных.

Все используемые приложения разделены на две группы. CINT2000 включает в себя 12 приложений, в основном оперирующих целочисленными значениями (а также логическими операторами). Одиннадцать из них написаны на чистом C, а одно – на C++. Второй набор – CFP2000, состоит из 14 приложений (6 Fortran-77, 4 Fortran-90 и 4 на C), интенсивно использующих вычисления с плавающей точкой. Итоговые оценки строятся из результатов измерения времени работы этих приложений. Тесты для целочисленной арифметики, включают, например, версию популярной утилиты компрессии gzip. Тест выполняет несколько операций сжатия/распаковки над набором файлов, общим объемом 28МБ. В набор входит изображение в формате tiff, логфайл веб сервера, приложение в двоичных кодах, файл со случайными данными и tar файл исходных текстов самого gzip. Все операции над данными производятся в оперативной памяти, что позволяет уменьшить влияние дисковой системы, а также повысить зависимость от скорости оперативной памяти.

Приложение для расчета FPGA-кристаллов VPR (Versatile Place and Route) используется для решения задач расположения и связи различных микроблоков FPGA (Field-Programmable Gate Array) чипов для достижения высокой скорости и работы. Во время теста решаются задачи выбора, расположения и связи блоков схемы для выполнения заданного алгоритма ее работы. Программа для игры в шахматы, благодаря своей сильно нелинейной структуре, может использоваться для проверки эффективности механизма предсказания ветвлений в современных процессорах. Во время теста решаются 5 шахматных комбинаций, строится дерево возможных ходов и выбирается лучший. При этом устанавливается различная глубина поиска. В приложении для решения задачи потока минимальной стоимости в сети решается проблема достижения мини-

мальной стоимости и составляется расписание перевозок с указанием времени прибытия и т.п.

При тестировании синтаксического разбора для естественного языка используется словарь из более 60000 словоформ. Анализируется входное множество фраз объемом 770кБ. При трассировке лучей используется вероятностный метод для построения изображения трехмерного объекта. В качестве тестовой задачи используется построение картинки стула, стоящего в углу комнаты. Для решения последовательно применяются три разных алгоритма.

Для языка Perl используется сокращенная версия интерпретатора этого языка скриптов для решения четырех задач: исполнение скрипта преобразования e-mail в HTML, работа с specdiff (используемый в самом SPEC скрипт с некоторыми изменениями), программа нахождения совершенных чисел и последняя – генерация последовательности случайных чисел. Для вычислительной задачи из теории групп в качестве теста используются несколько комбинаторных задач, работа с группами перестановок и другие. При тестировании объектно-ориентированной базы данных моделируется работа с тремя взаимосвязанными базами (почтовой рассылки, списком устройств и геометрических данных). Программа была специально модифицирована для уменьшения влияния дисковой системы и большинство операций проходит только в оперативной памяти. Используются операции вставки, удаления и поиска по базам. Тест запускает последовательно три различных шаблона работы с информацией.

Для утилиты сжатия данных bzip2 проверяется еще один вариант программы компрессии. В качестве исходных файлов используются изображение, программа и исходный текст. Общий объем данных – почти 20МБ. Для задачи позиционирования и маршрутизации используется оригинальная программа при создании литографических шаблонов для изготовления микрочипов. Решаются задачи взаимного расположения групп транзисторов и прокладки маршрутов между ними.

Тесты CFP2000 с интенсивными вычислениями с плавающей точкой включают, например, задачу квантовой хромодинамики (Fortran 77) с решением одного из важнейших уравнений теории сильного взаимодействия кварков – lattice-Dirac методом BiCGStab.

В гидродинамической задаче моделирования для «мелкой» воды (Fortran 77) решается разностное уравнение мелководья. Тест использовался для сравнения мощности суперкомпьютеров. Многосеточная «решалка» для трехмерного потенциального поля (Fortran 77) выполняет расчет трехмерного потенциального поля. Также относится к стандартным тестам суперкомпьютеров. Тест для параболических/эллиптических дифференциальных уравнений (Fortran 77) решает систему из пяти не-

линейных уравнений в частных производных на трехмерной сетке. Используется неявный метод.

В тесте на языке C для трехмерной графической библиотеки (Mesa3D) реализуется набор функций, аналогичный популярному интерфейсу OpenGL.

Выполняются гидродинамическая задача анализа колебательной нестабильности (Fortran 90) и численные вычисления параметров движения жидкости в закрытом пространстве (C) для моделирования нейронной сети. Используется модель нейронной сети для распознавания образов.

При моделировании землетрясения методом конечных элементов (C) рассчитывается распространение сейсмических волн на больших неоднородных пространствах методом конечных элементов.

Предусмотрены также:

1) Задача распознавания лиц на графических изображениях на Fortran 90.

2) Решение задач молекулярной динамики (C) для систем из большого количества молекул, характерных для биологии.

3) Задача теории чисел с проверкой простоты чисел Мерсенна методом Лукаса-Лехмера (Fortran 90).

4) Моделирование crash-тестов методом конечных элементов (Fortran 90). Метод конечных элементов используется для моделирования переходных характеристик при импульсных нагрузках на трехмерные тела.

5) Моделирование ускорителя элементарных частиц (Fortran 77). Решается задача моделирования работы ускорителя частиц и расчета динамической апертуры для проверки устойчивости луча.

6) Атмосферная задача с учетом температуры, ветра и загрязнений (Fortran 77). Рассчитывается распространение загрязняющего вещества в зависимости от погодных условий.

Более подробную информацию о тестах можно получить на вебсайте www.spec.org. Графически эти результаты проиллюстрированы на рисунках 4.1 – 4.4.

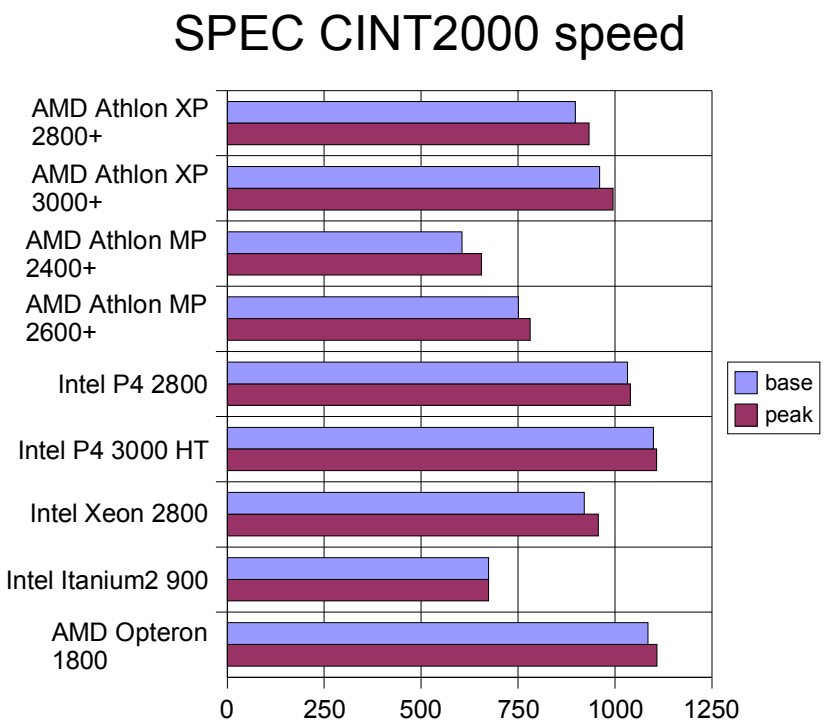

Рис. 4.1. Тест SPEC CINT2000 speed

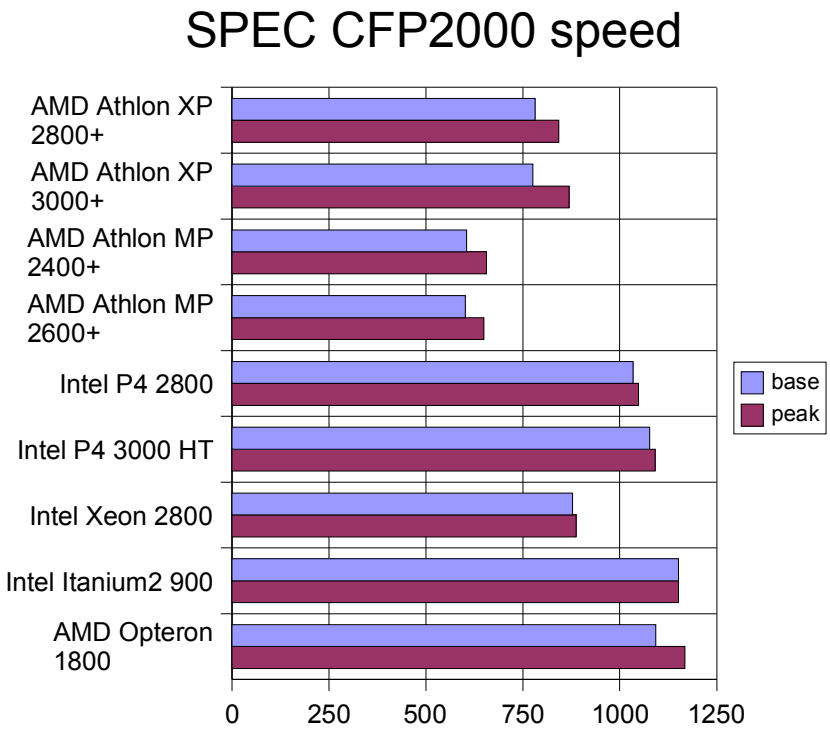

Рис. 4.2. Тест SPEC CFP2000 speed

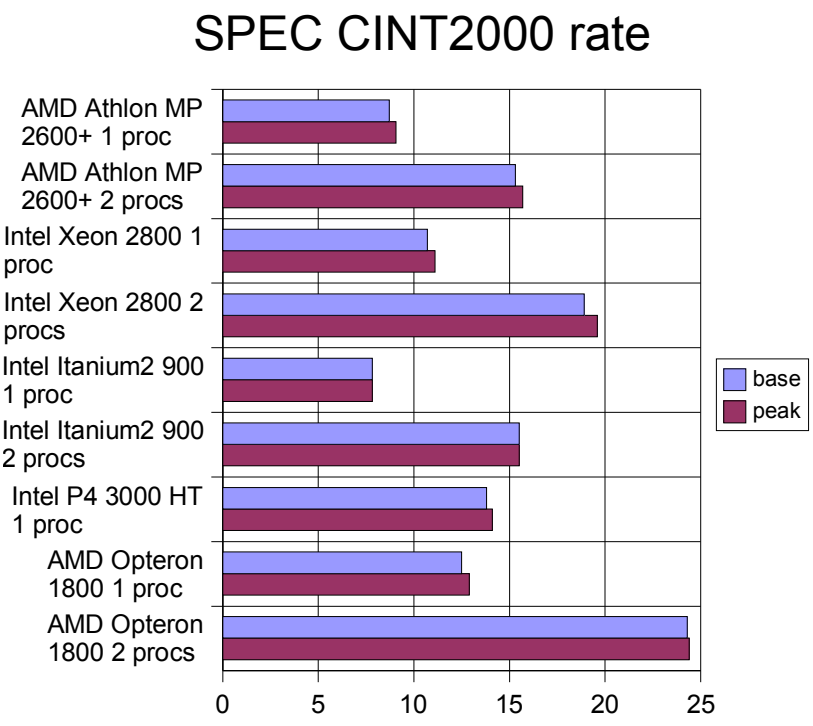

Рис. 4.3. Тест SPEC CINT2000 rate

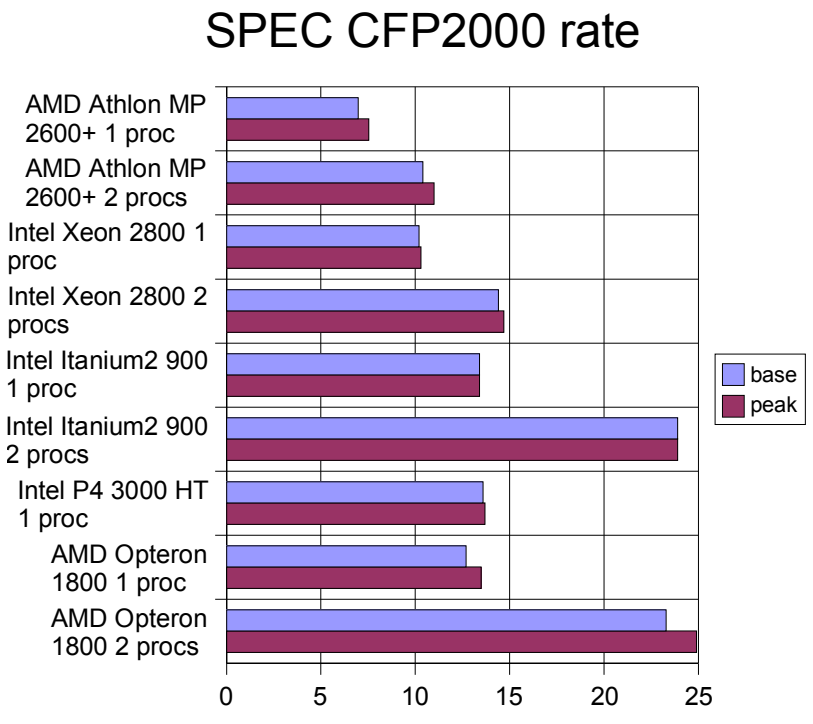

Рис. 4.4. Тест SPEC CFP 2000 rate

4.4. Выбор системной и вспомогательной сетей для суперкомпьютерных конфигураций терафлопсного диапазона

В качестве системной сети в экспериментальных и опытных образцах кластеров семейства СКИФ, созданных в 2000-2003 годах, выбрана технология SCI. Специализированная технология класса SCI (Scalable Coherent Interface, масштабируемый когерентный интерфейс) принят как стандарт в 1992 году. Интерфейс обеспечивает достижение высоких скоростей передачи с малым временем задержки, обеспечивая при этом масштабируемую архитектуру, позволяющую строить системы, состоящие из множества блоков. SCI представляет собой комбинацию шины и локальной сети, обеспечивает реализацию когерентности кэш-памяти, размещаемой в узле SCI, посредством механизма распределения директорий, который улучшает производительность, скрывая затраты на доступ к удаленным данным в модели с распределенной разделяемой памятью.

Производительность передачи данных обычно находится в пределах от 200МБ/с до 1000МБ/с на расстояниях десятков метров с использованием электрических кабелей и километров с использованием оптоволокна. SCI уменьшает время межузловых коммуникаций по сравнению с традиционными схемами передачи данных в сетях путем устранения обращений к программным уровням операционной системы и библиотекам времени выполнения. Коммуникации представляются как часть простой операции загрузки данных процессором (командами load или store).

Обычно обращение к данным, физически расположенным в памяти другого вычислительного узла и не находящимся в кэше, приводит к формированию запроса на удаленный узел для получения необходимых данных, которые в течение нескольких микросекунд доставляются в локальный кэш, и выполнение программы продолжается. Старый подход требовал формирования пакетов на программном уровне с последующей передачей их аппаратному обеспечению. Точно также происходил и прием, в результате чего задержки были в сотни раз больше, чем у SCI. Однако для совместимости SCI имеет возможность переносить пакеты других протоколов. Другое преимущество SCI – использование простых протоколов типа RISC (Reduced Instruction Set Computer), которые обеспечивают большую пропускную способность. Узлы с адаптером SCI могут соединяться в кольцо или же использовать для соединения коммутаторы. В отличие от технологии HIPPI (High Performance Parallel Interface, технология соединения типа точка-точка), данная технология оптимизирована для работы с динамическим трафиком, однако может быть менее

эффективна при работе с большими блоками данных.

Наиболее популярными сегодня коммуникационными технологиями для построения суперкомпьютеров на базе кластерных архитектур являются: Myrinet, Infiniband, Virtual Interface Architecture, SCI (Scalable Coherent Interface), QsNet (Quadrics Supercomputers World), Memory Channel , а также Fast Ethernet и Gigabit Ethernet. Выбор конкретной коммуникационной среды зависит от многих факторов: параметров быстродействия; особенностей класса решаемых задач, объемов финансирования, необходимости последующего расширения кластера и т.п. Основными характеристиками быстродействия коммуникационной среды являются латентность (latency) и пропускная способность (bandwidth). Под пропускной способностью сети понимают количество информации, передаваемой между узлами сети в единицу времени (байт в секунду). Реальная пропускная способность снижается программным обеспечением за счет передачи разного рода служебной информации. Латентностью (задержкой) называется время, затрачиваемое программным обеспечением и сетевыми устройствами на подготовку к передаче информации по данному каналу. Полная латентность складывается из программной и аппаратной составляющих.

Различают следующие виды пропускной способности сети:

– пропускная способность однонаправленных пересылок («точка-точка», uni-directional bandwidth), равная максимальной скорости, с которой процесс на одном узле может передавать данные другому процессу на другом узле;

– пропускная способность двунаправленных пересылок (bi-directional bandwidth), равная максимальной скорости, с которой два процесса могут одновременно обмениваться данными по сети.

Значения пропускной способности выражаются в мегабайтах в секунду (МБ/сек), значения латентности – в микросекундах (мкс = 10-6 с). При выборе конкретной коммуникационной среды важны не столько пиковые характеристики быстродействия, заявляемые производителем, сколько реальные, достигаемые на уровне пользовательских приложений, например, на уровне MPI-приложений.

Для оценки сетевых решений были измерены латентность и пропускная способность на доступных в тот момент сетевых средах для различных реализаций MPI: SCI (ScaMPI), Myrinet (MPICH), Fast Ethernet (LAM) и Gigabit Ethernet (LAM). Кроме того в средах ScaSCI и Myrinet существует возможность использования протокола IP для передачи пакетов данных между узлами, связанными сетями SCI и Myrinet соответсвенно, – эмуляция Ethernet. Приведены также результаты для этих реализаций – IP over SCI (LAM) и IP over Myrinet (LAM). Измерения проводились на тестах Alltoall и Bandwidth из пакета ScaMPItst входящего в

состав дистрибутива програмного обеспечения Scali – SSP-3.

Программа Bandwidth предназначена для измерения пропускной способности MPI при передаче сообщений различных размеров между двумя процессами. Сначала измеряется односторонняя пропускная способность и задержка для сообщений размером 0 байт, затем измеряется двухсторонняя пропускная способность и задержка. Программа Alltoall запускается на заданном количестве узлов и измеряет пропускную способность при передаче сообщений различной длины.

Тестирование производилось на трех вычислительных узлах с установленной операционной системой RedHat Linux 7.3 следующей конфигурации:

– серверная системная плата Intel® SCB2, 2xP-III 1,4 ГГц; ОЗУ 512Мбайт;

– операционная система RedHat Linux 7.3;

– MPI: ScaMPI 1.13.8, MPICH-GM 1.6, LAM 6.5.6.

– параметры коммуникационного оборудования приведены в таблице 4.1.

Таблица 4.1.

Параметры коммуникационного оборудования

	Fast Ethernet	**Gigabit Ethernet**	**Myrinet**	**SCI**
Производитель оборудования	**Intel**	**Intel, 3 COM**	**Myricom**	**Dolphin**
Состав тестируемого сетевого оборудования	1. Сетевой адаптер Intel PRO/100+ Fast Ethernet 2. Коммутатор Intel Express 410T	1. Сетевой адаптер Intel PRO/1000 XT Gigabit Server Adapter 2. Коммутатор 3. Com SuperStack 3 Switch 4900	1. Сетевой адаптер Myrinet-2000-Fiber/PCI interfaces M3F-PCI64C 2. Коммутатор M3F-SW8	1. Сетевой адаптер PCI-64/66 / D335 SCI Adapter Card

Графически эти результаты отображены на рисунках 4.5 – 4.9.

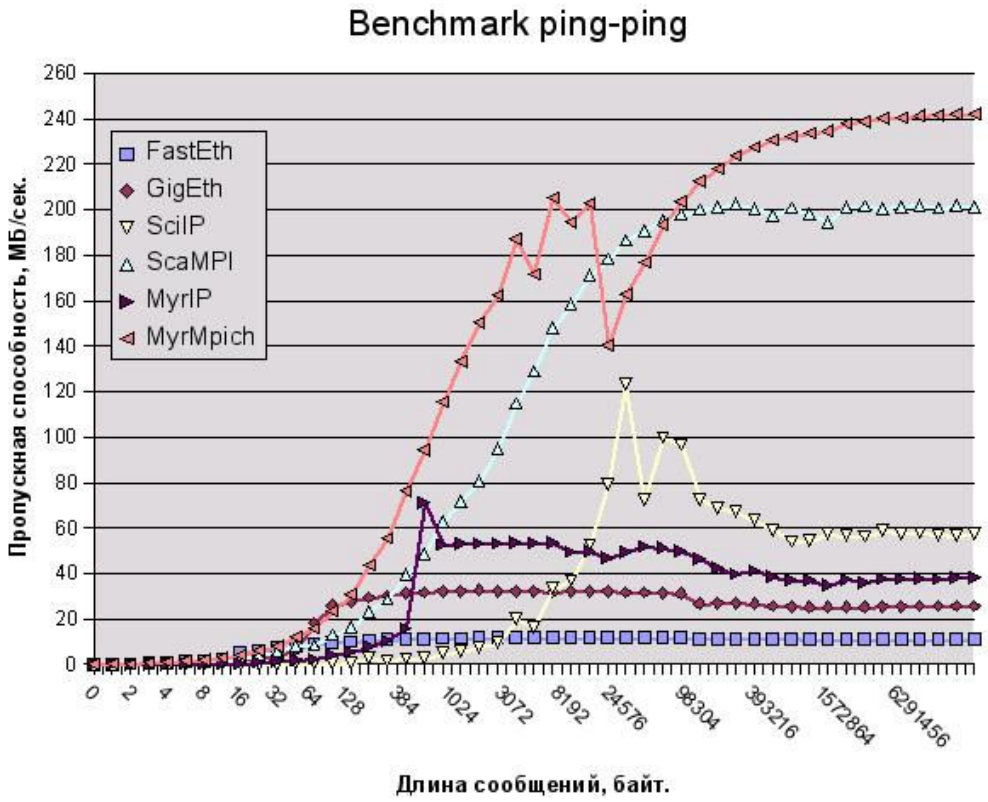

Рис. 4.5. Benchmark ping-ping

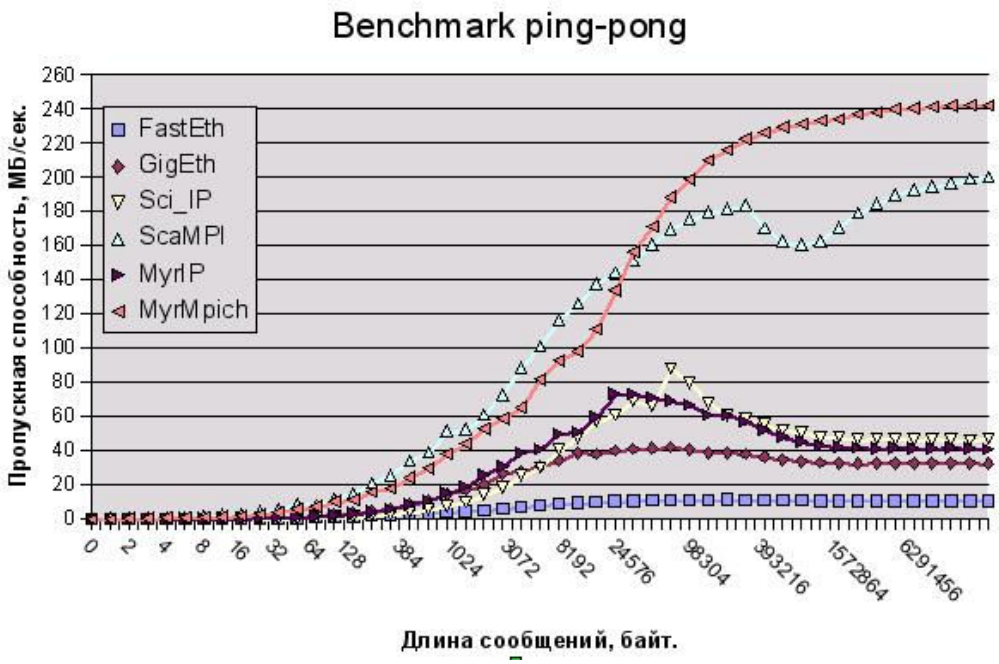

Рис. 4.6. Benchmark ping-pong

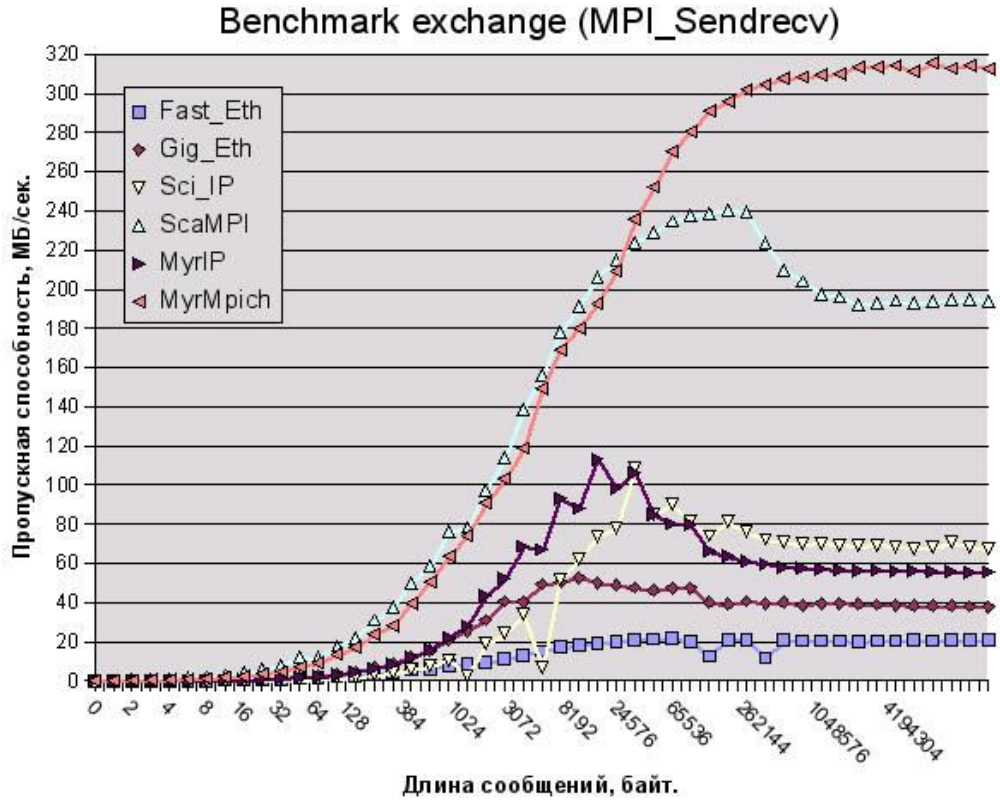

Рис. 4.7. Benchmark exchange

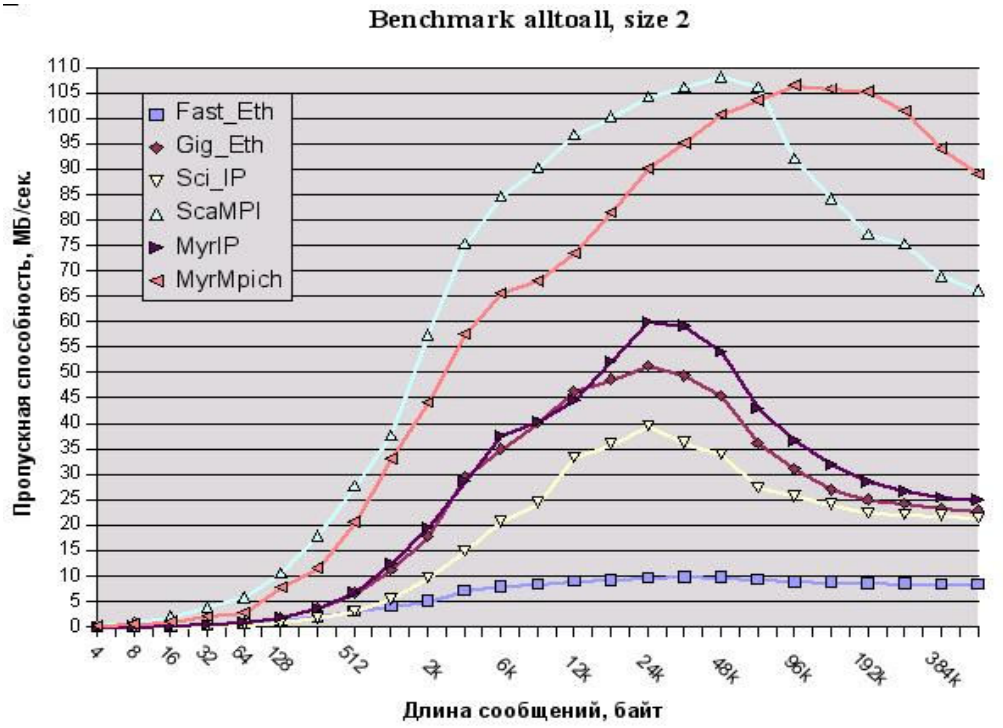

Рис. 4.8. Benchmark Alltoall, размер 2

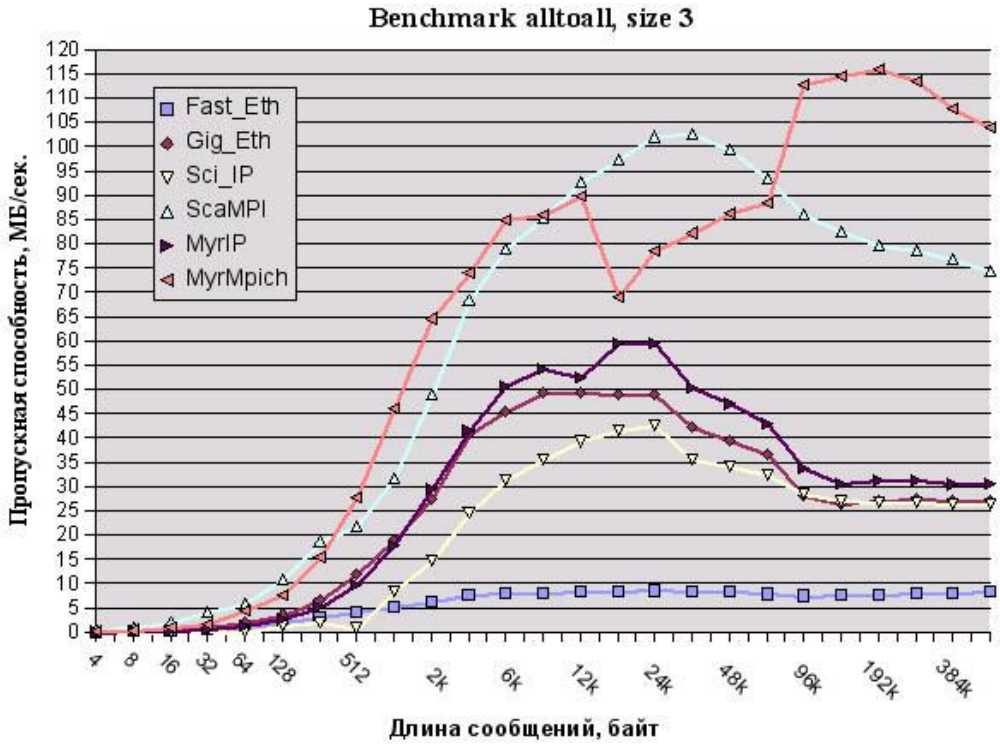

Рис. 4.9. Benchmark Alltoall, размер 3

Как видно из результатов тестирования производительность сетевых решений Myrinet и SCI значительно превосходит Fast Ethernet и Gigabit Ethernet. Программное обеспечение СС SAN (SCI) по составу включает:

а) драйверы аппаратуры СС SAN (SCI), обеспечивающие:

– совместимость с выбранными аппаратными средствами базового вычислительного элемента кластерного уровня (БВМ КУ);

– совместимость с разрабатываемым дистрибутивом ОС Linux.

б) ПО поддержки MPI на платформе СС SAN (SCI),обеспечивающее:

– совместимость со стандартом MPI 1.2;

– совместимость со стандартом MPI 1.2 + ROMIO.

в) ROMIO реализация системы параллельного ввода-вывода MPI-IO обеспечивает:

– расширенную совместимость MPI с SMP;

– ПО поддержки модели общей памяти на платформе СС (SCI), позволяющее напрямую (без поддержки со стороны ОС и/или MPI) выполнять операции доступа на чтение и запись к памяти различных базовых вычислительных модулей (БВМ) КУ.

Разделяемая (общая) память (Shared Memory) – это модель взаимо-

действия процессоров внутри параллельной системы. В системах с физически разделяемой между несколькими процессорами памятью значение, записанное в память одним из процессоров, может быть напрямую доступно для другого процессора. Для передачи данных между процессорами не требуется делать обращение к сетевой поддержке. Если в системе каждый процессор имеет свою собственную память, то возможна реализация так называемой логически разделяемой памяти (logically shared memory). Этот способ реализуется посредством трансляции обращений к памяти по нелокальному адресу в соответствующий вид межпроцессорных коммуникаций. Физически разделяемая память может иметь крайне высокую пропускную способность (bandwidth) и очень низкую задержку (latency) при передаче информации между процессорами, но при условии, что не происходит одновременного обращения нескольких процессоров к одному и тому же элементу памяти.

Еще одной технологией, заслуживающей внимания, является технология Infiniband. Infiniband архитектура была предложена в качестве следующего поколения соединений для операций ввода-вывода и межпроцессорного взаимодействия. В Infiniband, вычислительные узлы и узлы ввода - вывода связаны с switched fabric через адаптеры (Host Channel Adapters). Infiniband обеспечивает Verbs интерфейс, который является надмножеством Virtual Interface Architecture (VIA). Этот интерфейс используется операционной системой для связи с адаптером Host Channel Adapter. NBCL/OSU MVAPICH это реализация MPI на уровне Verbs Infiniband.

Данная реализация основана на MPICH и MVICH. MVICH – это реализация MPI, основанная на MPICH для Virtual Interface Architecture (VIA), разработанной Berkeley Lab. В данной реализации MPICH поверх IBA, интерфейс ADI2 осуществлен непосредственно поверх уровня Verbs. Эта реализация базируется на Mellanox Technologies VAPI интерфейсе. Данная реализация была проверена на Mellanox InfiniHost MT23108 DualPort 4X HCA адаптерах, соединенных через InfiniScale MT43132 Eight 4x Port Infiniband Switch.

Тесты производительности. Тест на время ожидания проводился методом пинг-понг. Отправитель посылает сообщение с некоторым размером данных получателю и ждет ответ от получателя. Получатель получает сообщение от отправителя и посылает назад ответ с тем же самым размером данных. Тест проводился с большим количеством итераций для получения усредненного времени ожидания. В этом тесте использовались блокирующие версии функций MPI (MPI_Send и MPI_RECV). Программа для теста времени ожидания доступна на http://nowlab.cis.ohio-state.edu/projects/mpi-iba/mpi_latency.c.

Тест на ширину полосы пропускания. Тест проводился при нали-

чии отправителя, посылающего большое количество (1000) сообщений получателю и ожидающего ответ от получателя. Получатель посылает ответ только после получения всех (1000) сообщений. Полоса пропускания рассчитана на основании прошедшего времени (с момента, когда отправитель посылает первое сообщение до времени, когда получает ответ назад от получателя) и числа байтов, посланных отправителем. Цель этого теста состоит в том, чтобы определить максимум скорости передачи данных, которая может быть достигнута на сетевом уровне. В этом тесте использовались неблокирующие версии функций MPI (MPI_Isend и MPI_IRECV). Программа для теста полосы пропускания доступна на http://nowlab.cis.ohio-state.edu/projects/mpi-iba/mpi_bandwidth.c.

Имеется много различных способов измерения полосы пропускания MPI. Могут использоваться различные типы запросов MPI (блокирующие либо неблокирующие), может изменяться общее количество сообщений, представляющих одну итерацию, число итераций, и т.д. Тесты, результаты которых представлены, предназначены для определения максимальной однонаправленной скорости передачи данных, которая может быть поддержана на сетевом уровне приложением MPI. Другие способы измерять полосу пропускания могут давать различные результаты.

MVAPICH уровень позволяет достичь одностороннее время ожидания 6.8 микросекунд и однонаправленной полосы пропускания до 829 мегабайт в секунду.

Тесты проводились на кластерной системе, состоящей из узлов на материнской плате SuperMicro SUPER P4DL6. На каждом узеле было установлено по два процессора Intel Xeon 2.40 ГГц с 512 кБ кэша второго уровня на 400 МГц FSB. Машины были связаны Mellanox InfiniHost MT23108 DualPort 4X HCA через InfiniScale MT43132 Eight 4x Port InfiniBand Switch. HCA адаптеры использовали PCI-X 64-bit 133 МГц интерфейсы. Версия Mellanox InfiniHost HCA SDK - thca-x86-0.1.2-build-001. Версия firmware - fw-23108-rel-1_17_0000-rc12-build-001. Также использовались компьютеры, связаные Myrinet сетью, использующие M3F-PCIXD-2 адаптеры с 225 МГц Lanai-XP процессорами через 8-port Myrinet-2000 M3F-PCIXD-2 коммутатор. Myrinet адаптеры использовали 64-bit 133 МГц PCI-X шину для экспериментов. Quadrics Elan3 QM-400 адаптеры были использованы на тех же компьютерах. Они были связаны друг с другом через коммутатор Elite16. QM-400 плата использовала 64-bit 66 МГц PCI слот. Использовалась операционная система Linux Red Hat 7.2.

Результаты тестов приведены на рисунках 4.10 – 4.12.

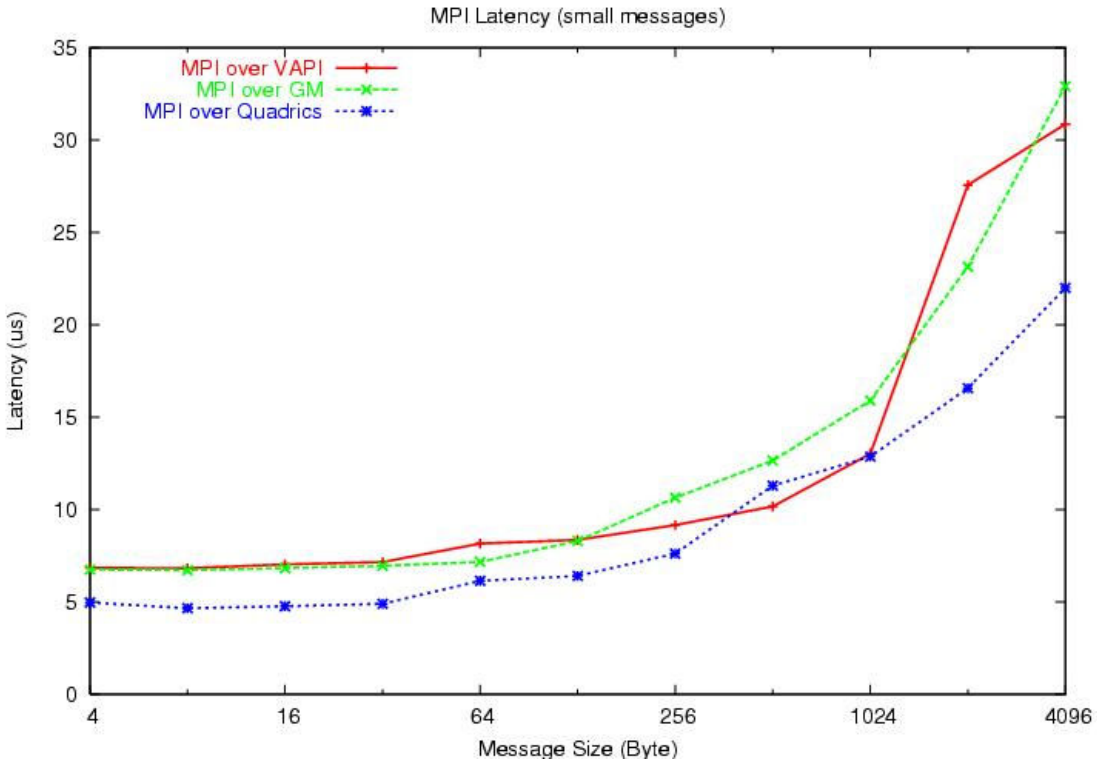

Рис. 4.10. MPI латентность на коротких сообщениях

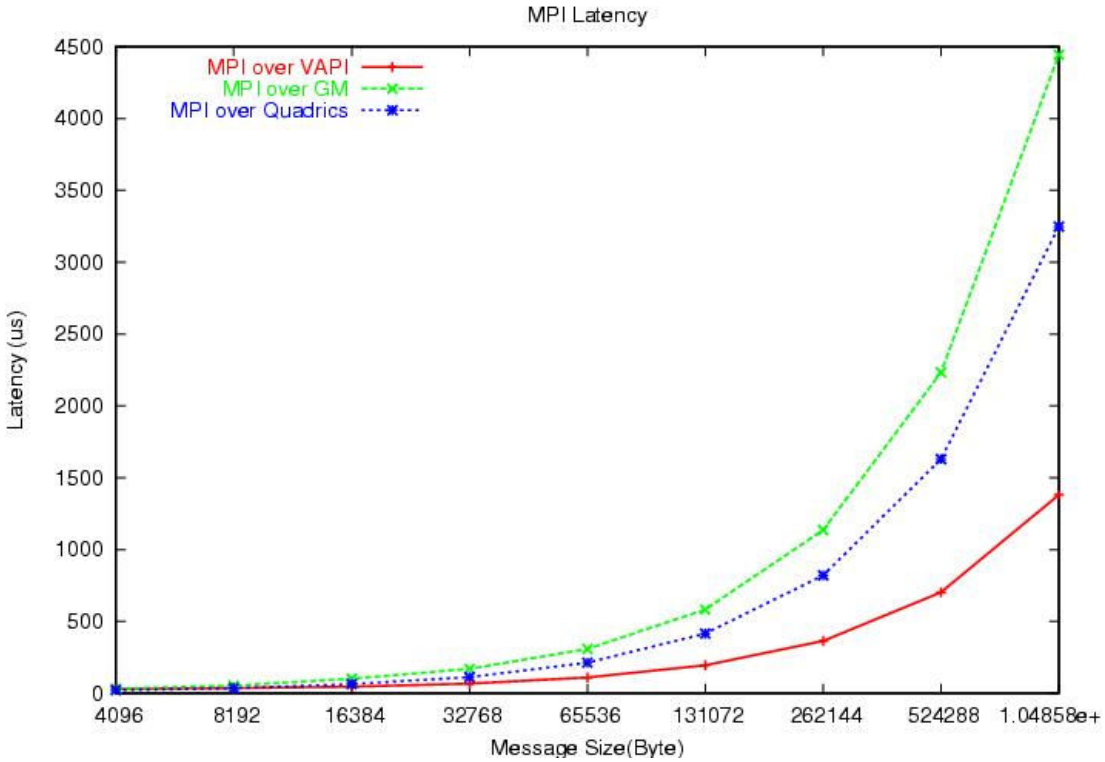

Рис. 4.11. MPI латентность на длинных сообщениях

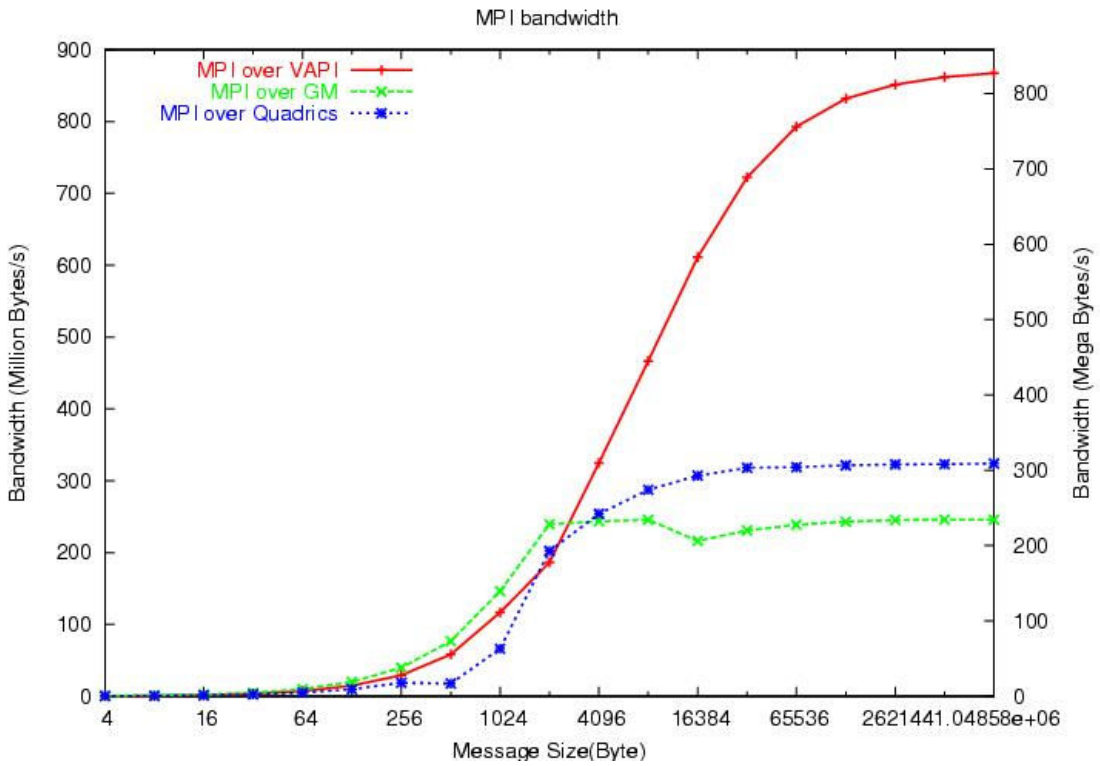

Рис. 4.12. MPI пропускная способность

Вспомогательная сеть суперкомпьютера терафлопной производительности. Для вспомогательной сети стандартным решением является Fast Ethernet. В тех случаях, когда экономически нецелесообразно использовать специализированную системную сеть, пользователь может использовать вспомогательную сеть для организации параллельных вычислений. Вспомогательная сеть суперкомпьютера с протоколом TCP/IP объединяет узлы кластерного уровня в обычную локальную сеть (TCP/IP LAN). Данная сеть предназначена для управления системой, подключения рабочих мест пользователей, интеграции суперкомпьютера в локальную сеть предприятия и/или в глобальные сети. Кроме того, данный уровень может быть использован и системой организации параллельных кластерных вычислений (T-система, MPI) для вспомогательных целей (основные потоки информации, возникающие при организации параллельных кластерных вычислений, передаются через системную сеть кластера).

Вспомогательная сеть кластера терафлопной производительности может быть реализована на основе широко используемых сетевых технологий класса Fast Ethernet, Gigabit Ethernet, ATM и др. До последнего времени при построении вспомогательной сети кластеров терафлопной производительности применялись оба решения: Fast Ethernet (суперком-

пьютер Monolith, Швеция), Gigabit Ethernet или их комбинация (например, суперкомпьютер Межведомственного Суперкомпьютерного Центра, Россия). Выбор обуславливался классом решаемых задач и, конечно, стоимостью. Для терафлопного кластера значение производительности вспомогательной сети возрастает. При распределении большой вычислительной задачи на значительном количестве узлов высокая пропускная способность гигабитной сети позволит уменьшить задержки.

На рисунках 4.13-4.16 приведены сравнительные диаграммы пропускной способности сетей Gigabit Ethernet, Fast Ethernet, а также Fast Ethernet со связыванием каналов для реализации LAM MPI на тестах Alltoall и Bandwidth из пакета ScaMPItst. Измерения проводились на вычислительных узлах кластера BM5100 с серверной платой Intel SCB2. Связывание двух портов Fast Ethernet, присутствующих на большинстве серверов, позволяет вдвое увеличить пропускную способность канала, но не влияет на величину задержки сети.

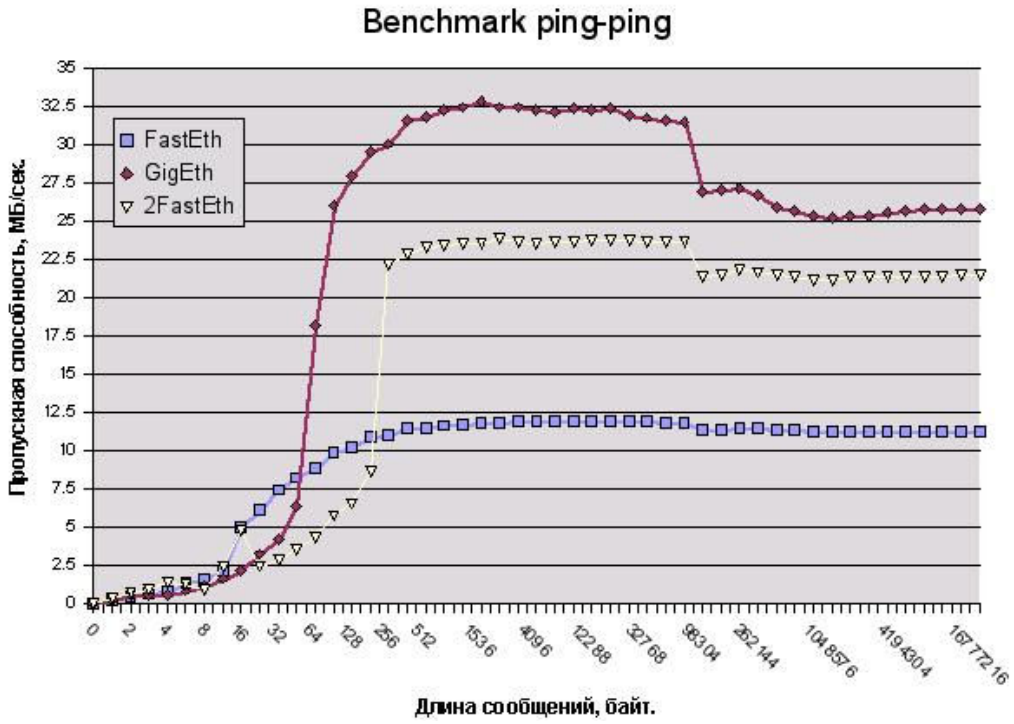

Рис. 4.13. Результаты выполнения теста ping-ping

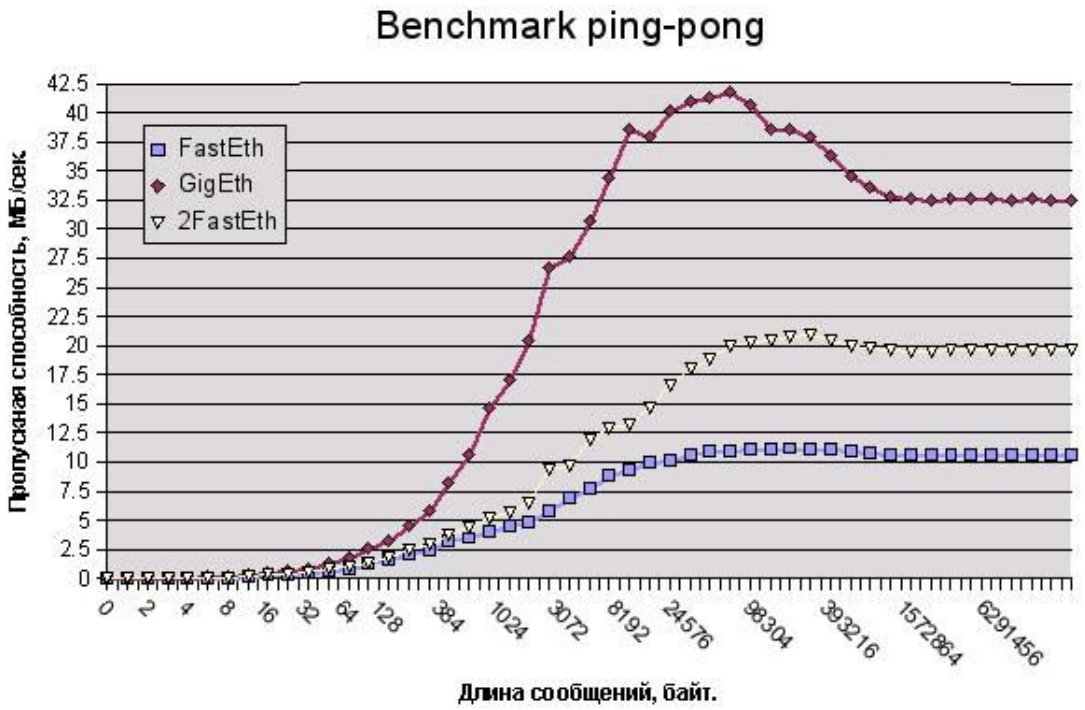

Рис. 4.14. Результаты выполнения теста ping-pong

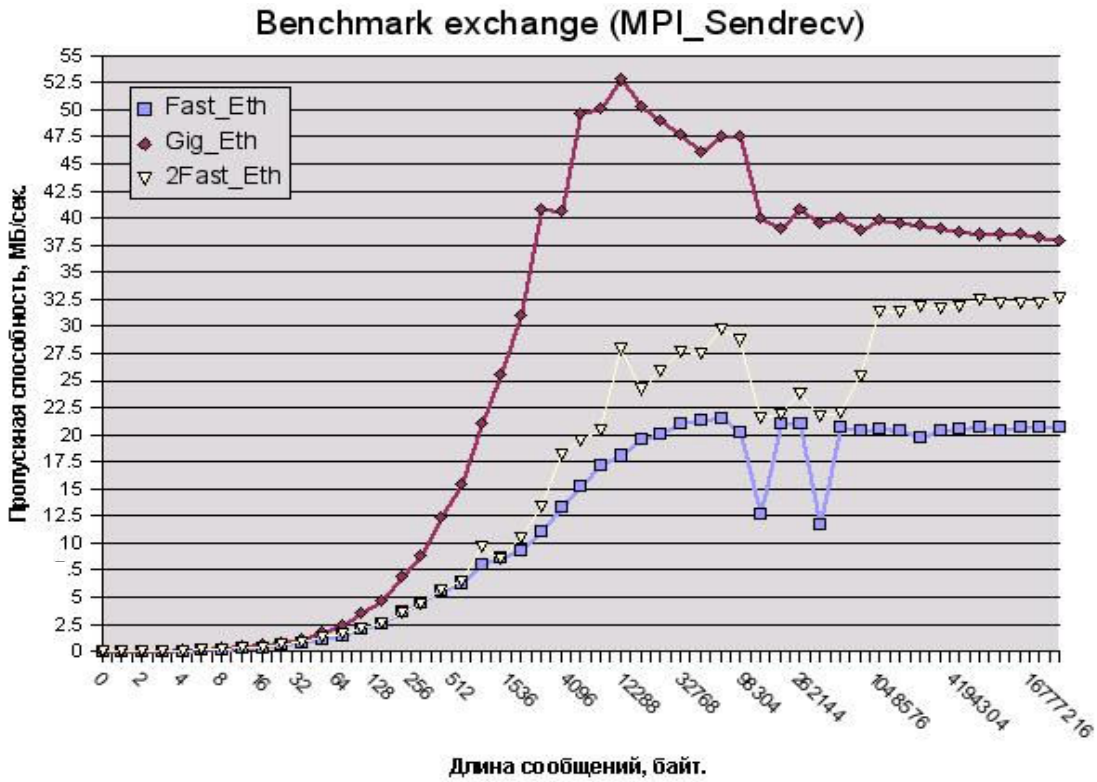

Рис. 4.15. Результаты выполнения теста MPI_Sendrecv

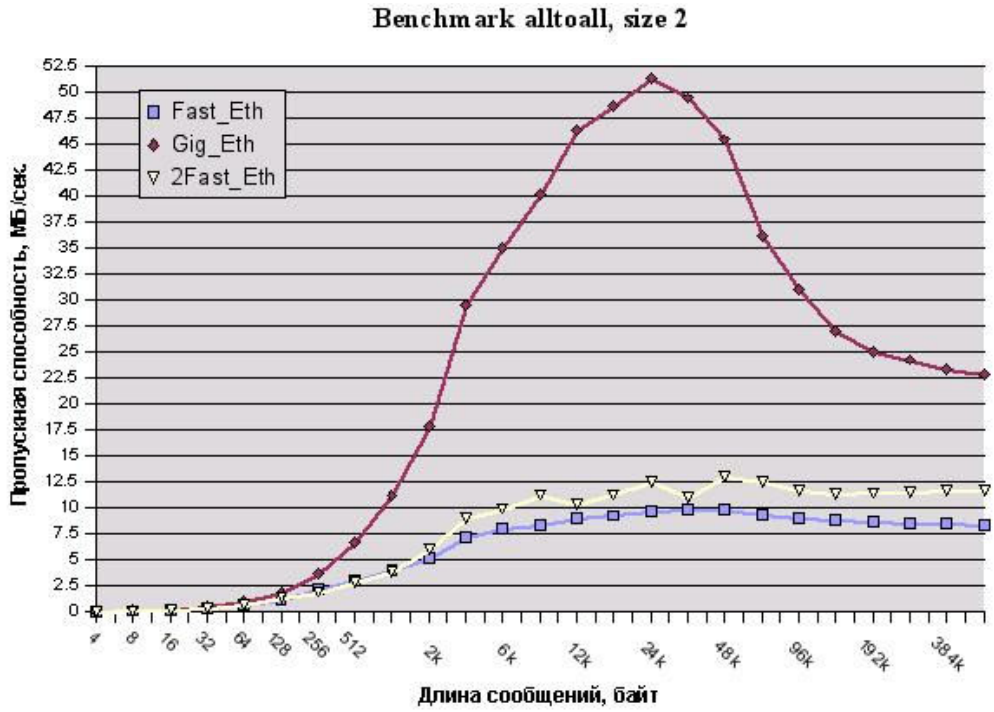

Рис. 4.16. Результаты выполнения теста Alltoall

Для достижения терафлопной производительности число узлов кластера SCI на современных серверных сетевых платах с процессорами Intel Itanium или AMD Opteron должно быть 200-300. Благодаря значительному удешевлению решений Gigabit Ethernet в последнее время, стоимость вспомогательной сети такого кластера будет незначительно превышать аналогичное решение на Fast Ethernet. При этом следует учесть, что подавляющее большинство современных серверных сетевых плат имеет два встроенных гигабитных сетевых адаптера (например Intel7501wv2 или Tyan Thunder i7501Pro), несильно влияющих на общую стоимость сервера. Таким образом, основное ценовое отличие приходиться на коммутаторы вспомогательной сети и оно не является определяющим в стоимости всего кластера. Например, для нового семейства коммутаторов Cisco Catalyst 3750 стоимость порта для варианта Fast Ethernet составляет порядка 210-230 $, для варианта Gigabit Ethernet – 250-260 $.

Можно предложить различные схемы вспомогательной сети суперкомпьютера. Они будут отличаться в зависимости от используемого коммутационного оборудования: управляемые, неуправляемые, стекируемые коммутаторы, количество портов в коммутаторах и т.д.

4.5. Выбор конфигурации систем внешней памяти

С повседневным усложнением сетевых компьютерных систем и глобальных корпоративных решений мир начал требовать технологий, которые бы дали толчок к возрождению корпоративных систем хранения информации (storage-system, сторедж-система). Ниже приведена терминология и технологические основы систем хранения информации, а также их необходимость в разных ситуациях. Различают три основных варианта организации доступа к системам хранения:

– SAS (Server Attached Storage) – сторедж, присоединенный к серверу (второе на-звание DAS (Direct Attached Storage));

– NAS (Network Attached Storage) – сторедж, подсоединенный к сети;

– SAN (Storage Area Network) - сеть хранения данных.

Ниже рассмотрены топологии соответствующих сторедж-систем и их особенности. **SAS** – сторедж-система, присоединенная к серверу. Это традиционный способ подключения системы хранения данных к высокоскоростному интерфейсу в сервере, как правило, к параллельному SCSI интерфейсу. Основное преимущество стореджа, подсоединенного к серверу, в сравнении с другими вариантами – низкая цена и высокое быстродействие из расчета один сторедж для одного сервера. Такая топология является самой оптимальной в случае использования одного сервера, через который организуется доступ к массиву данных. Однако существует ряд проблем, которые побудили проектировщиков искать другие варианты организации доступа к системам хранения данных.

К особенностям SAS можно отнести:

– доступ к базе данных зависит от ОС и файловой системы;

– сложность организации систем с высокой готовностью;

– низкая стоимость;

– высокое быстродействие в рамках одного узла;

– уменьшение скорости отклика при загрузке сервера, который обслуживает сторедж.

SAS/DAS – это достаточно простой традиционный способ подключения, который подразумевает прямое (отсюда и DAS) подсоединение системы хранения к одной или нескольким хост-системам через высокоскоростной канальный интерфейс. Часто в таких системах, для подсоединения накопителя к хосту используется такой же интерфейс, который используется для доступа к внутренним дискам хост-системы, что в общем случае обеспечивает высокое быстродействие и простое подключение. SAS-систему можно рекомендовать к использованию в случае, если имеется потребность в высокоскоростной обработке данных больших объемов на одной или нескольких хост-системах. Это, например, может

быть файл-сервер, графическая станция или отказоустойчивая кластерная система, состоящая из двух узлов.

NAS – сторедж-система, подсоединенная к сети. Основным преимуществом этого способа является удобство интеграции дополнительной системы хранения данных в существующие сети, но сам по себе он не привносит сколько-нибудь радикальных улучшений в архитектуру сторедж. Фактически NAS есть чистый файл-сервер, и сегодня можно встретить немало новых реализаций сторедж типа NAS на основе технологии тонкого сервера (Thin Server).

Особенности NAS:

- выделенный файл-сервер;
- доступ к данным не зависит от ОС и платформы;
- удобство администрирования;
- максимальная простота установки;
- низкая масштабируемость;
- конфликт с трафиком LAN/WAN.

Сторедж, построенный по технологии NAS, является идеальным вариантом для дешевых серверов с минимальным набором функций. NAS – накопитель, который подсоединен к сети и обеспечивает файловый (файловый, а не блочный) доступ к данным для хост-систем в сети LAN/WAN. Клиенты, которые работает с NAS, для доступа к данным, обычно используют протоколы NFS (Network File System) или CIFS (Common Internet File System). NAS интерпретирует команды файловых протоколов и исполняет запрос к дисковым накопителям в соответствии с используемым в нём канальным протоколом. Фактически, архитектура NAS – это эволюция файловых серверов. Главным преимуществом такого решения является быстрота развёртывания и качество организации доступа к файлам.

Исходя из сказанного, NAS можно рекомендовать для использования в случае, если нужен сетевой доступ к файлам и достаточно важными факторами являются: простота решения, его сопровождения и установки. Прекрасным примером является использование NAS в качестве файл-сервера в офисе небольшой компании, для которой важна простота установки и администрирования. Но в то же время, если вам нужен доступ к файлам с большого количества хост-систем, мощный NAS-накопитель, благодаря отточенному специализированному решению, способен обеспечить интенсивный обмен трафиком с огромным пулом серверов и рабочих станций при достаточно низкой стоимости используемой коммуникационной инфраструктуры (например, коммутаторов Gigabit Ethernet и медной витой пары).

Файловые протоколы в современных NAS.

CIFS (Common Internet File System) – это стандартный протокол, который обеспечивает доступ к файлам и сервисам на удаленных компьютерах (в том числе и в Интернет). Протокол использует клиент-серверную модель взаимодействия. Клиент создает запрос к серверу на доступ к файлам или передачу сообщения программе, которая находится на сервере. Сервер выполняет запрос клиента и возвращает результат своей работы. CIFS – это открытый стандарт, который возник на основе SMB-протокола (Server Message Block Protocol), разработанного Microsoft, но, в отличие от последнего, CIFS учитывает возможность возникновения больших таймаутов, так как ориентирован на использование в том числе и в распределённых сетях. SMB-протокол традиционно использовался в локальных сетях с ОС Windows для доступа к файлам и печати. Для транспортировки данных CIFS использует TCP/IP протокол. CIFS обеспечивает функциональность, похожую на FTP (File Transfer Protocol), но предоставляет клиентам улучшенный (похожий на прямой) контроль над файлами. Он также позволяет разделять доступ к файлам между клиентами, используя блокирование и автоматическое восстановление связи с сервером в случае сбоя сети.

NFS (Network File System) – это стандарт IETF, который включает в себя распределенную файловую систему и сетевой протокол. Протокол NFS разработан компанией Sun Microsystem Computer Corporation и первоначально использовался только в Unix-системах. NFS, как и CIFS, использует клиент-серверную модель взаимодействия. Он обеспечивает доступ к файлам на удаленном компьютере (сервере) для записи и считывания так, как если бы они находились на компьютере пользователя. В ранних версиях NFS для транспортирования данных использовался UDP-протокол, в современных – используется TCP/IP. Для работы NFS в Интерент компанией Sun был разработан протокол WebNFS, который использует расширения функциональности NFS для его корректной работы во всемирной сети.

DAFS (Direct Access File System) – это стандартный протокол файлового доступа, который базируется на NFSv4. Он позволяет прикладным задачам передавать данные в обход операционной системы и ее буферного пространства напрямую к транспортным ресурсам, сохраняя семантику, свойственную файловым системам. DAFS использует преимущества новейших технологий передачи данных по схеме память-память. Его использование обеспечивает высокие скорости файлового ввода-вывода, минимальную загрузку CPU и всей системы, благодаря значительному уменьшению количества операций и прерываний, которые обычно необходимы при обработке сетевых протоколов. Особенно эффективным является использование аппаратных средств поддержки

VI (Virtual Interface).

DAFS проектировался с ориентацией на использование в кластерном и серверном окружении для баз данных и разнообразных Интернет-приложений, ориентированных на непрерывную работу. Он обеспечивает наименьшие задержки доступа к общим файловым ресурсам и данным, а также поддерживает интеллектуальные механизмы восстановления работоспособности системы и данных, что делает его очень привлекательным для использования в High-End NAS-накопителях.

Сети хранения данных (SAN) начали интенсивно развиваться и внедряться лишь с 1999-го года. Основой SAN является отдельная от LAN/WAN сеть, которая служит для организации доступа к данным серверов и рабочих станций, занимающихся их прямой обработкой. Такая сеть создается на основе стандарта Fibre Channel, что дает сторедж-системам преимущества технологий LAN/WAN и возможности по организации стандартных платформ для систем с высокой готовностью и высокой интенсивностью запросов. Почти единственным недостатком SAN на сегодня остается относительно высокая цена компонент, но при этом общая стоимость владения для корпоративных систем, построенных с использованием технологии сетей хранения данных, является довольно низкой.

К основным преимуществам SAN можно отнести:

– независимость топологии SAN от сторедж-систем и серверов;

– удобное централизованное управление;

– отсутствие конфликта с трафиком LAN/WAN;

– удобное резервирование данных без загрузки локальной сети и серверов;

– высокое быстродействие;

– высокая масштабируемость;

– высокая гибкость;

– высокая готовность и отказоустойчивость.

Однако технология эта на данном этапе находится на стадии разработки и в будущем она пройдет немало усовершенствований в области стандартизации управления и способов взаимодействия SAN подсетей. SAN – это сеть хранения данных. Обычно в SAN используется блочный доступ к данным, хотя возможно подключение к сетям хранения данных устройств, предоставляющих файловые сервисы, например NAS. В современных реализациях сети хранения данных как правило используют протокол Fibre Channel, но в общем случае это не является обязательным, в связи с чем, принято выделять отдельный класс Fibre Channel SAN (сети хранения данных на основе Fibre Channel).

Основой SAN является отдельная от LAN/WAN сеть, которая слу-

жит для организации доступа к данным серверов и рабочих станций, непосредственно занимающихся обработкой. Такая структура делает построение систем с высокой готовностью и высокой интенсивностью запросов относительно простой задачей. Несмотря на то, что SAN сегодня является дорогой реализацией, TCO (общая стоимость владения) для средних и больших систем, построенных с использованием технологии сетей хранения данных, является довольно низкой. Описание способов снижения TCO корпоративных систем хранения данных благодаря SAN можно найти на страницах ресурса techTarget:

http://searchstorage.techtarget.com

Кластеризация серверов (Server Clustering). Одной из типичных задач, для которых эффективно используется SAN, является кластеризация серверов. Поскольку один из ключевых моментов в организации высокоскоростных кластерных систем, которые работают с данными – это доступ к сторедж, то с появлением SAN построение многонодовых кластеров на аппаратном уровне решается простым добавлением сервера с подключением к SAN (это можно сделать, даже не выключая системы, поскольку свичи FC поддерживают режим hot-plug). При использовании параллельного SCSI интерфейса, возможности по подсоединению и масштабируемость которого значительно хуже, чем у FC, затруднено создание кластеров, ориентированных на обработку данных, с количеством узлов больше двух. Коммутаторы параллельного SCSI-интерфейса весьма сложные и дорогие устройства, а для FC это стандартный компонент. Для создания кластера, который не будет иметь ни единой точки отказов, достаточно интегрировать в систему зеркальную SAN (технология DUAL Path).

Основная идея сети хранения (Storage Area Network, SAN) заключается в возможности отделения аппаратных средств хранения данных от сервера и, главное, от сетевой операционной системы. Например, распространенной является ситуация, при которой сеть предприятия содержит в целом достаточное количество дискового пространства, но все свободные блоки оказываются в разделах, не доступных для использования. С помощью сети хранения эту проблему можно разрешить, объединив все пространство хранения независимо от конфигурации серверов и разделов и их состояния.

Сложными являются также вопросы, связанные с резервированием, тестированием и добавлением дисков и массивов для типичной комбинации различных аппаратных платформ. Сеть хранения предполагает возможность применения набора отказоустойчивых, стандартизованных, централизованно управляемых массивов таким образом, чтобы все потребности предприятия в ресурсах хранения удовлетворялись полностью. Технология SAN, избавляя от необходимости изучать разнообраз-

ные интерфейсы управления RAID, резервировать многочисленные диски всевозможных типов и следить за тем, чтобы свободное дисковое пространство не оказалось целиком закреплено за неподходящим сервером, разделом или сетевой ОС, позволяет создать полностью независимую и стандартизованную систему хранения. Однако на практике решения в виде сетей хранения за те деньги, в которые обходится типичная конфигурация, не всегда оправдывают надежды. Зачастую они требуют радикального обновления и капитальной переделки существующей модели хранения и потому становятся нерентабельными. В таких ситуациях можно обратиться к альтернативным технологиям, тем более что они обеспечивают по разумной цене практически те же возможности, что и SAN, и удачно вписываются в традиционные модели хранения, ориентированные на серверы.

Эти решения можно разбить на две категории. Продукт первого типа – много-портовый массив дисков, где используется интерфейс SCSI/SCSI для обеспечения независимости от сетевой ОС и достижения максимальной централизации и стандартизации. Ко второму типу относятся подключаемые к сети устройства серверного типа, они используют сетевую ОС в минимальной конфигурации и поддерживают один или несколько стандартных протоколов файловых служб. Фактически они представляют собой NAS устройства хранения данных.

Применение SAN в качестве системы хранения позволяет достичь наибольшей скорости и масштабируемости. В имеющейся конфигурации – 64-х узловой кластер с использованием вспомогательной сети на Gigabit Ethernet – применение NAS выглядит предпочтительней ввиду отсутствия необходимости в масштабируемости кластерной системы и простоты в настройке и эксплуатации, а также меньшей стоимости NAS.

Одна из возможных реализаций NAS-системы – это сервер сетевого хранения данных с использованием дисковой системы RAID.

RAID-системы. Основные задачи, которые позволяют решить RAID – это обеспечение отказоустойчивости дисковой системы и повышение ее производительности. Отказоустойчивость достигается за счет избыточности. В RAID объединяются больше дисков, чем необходимо для получения требуемой емкости. Производительность дисковой системы повышается благодаря тому, что современные интерфейсы (в частности, SCSI) позволяют осуществлять операции записи и считывания фактически одновременно на нескольких дисках. Поэтому можно рассчитывать на то, что скорость записи или чтения, в случае применения RAID, увеличивается пропорционально количеству дисков, объединяемых в RAID.

Существует несколько способов организовать RAID-систему, первое – это программным способом (на рынке существует большое коли-

чество программного обеспечения для этих целей), второе – это аппаратным способом, т.е. с помощью RAID-контроллера. При использовании аппаратных реализаций, RAID-система «видна» как один целый диск. Такой дисковый массив конфигурируется с помощью программных средств, предоставляемых производителем. Ниже описаны примеры настройки программного RAID.

Уровни RAID. Текущие RAID патчи для Linux поддерживают несколько режимов.

Линейный режим. Два или более диска объединяются в одно устройство. Диски «добавляются» один к другому, таким образом, запись на устройство RAID будет заполнять сначала диск 0, затем диск 1 и так далее. Диски не обязательно должны быть одного размера. Фактически, размер здесь вообще не имеет значения. На этом уровне нет избыточности. Если один диск отказывает, то, с большой вероятностью, произойдет потеря данных. Однако, существует вероятность удачного восстановления части данных, так как в файловой системе будет просто отсутствовать один большой последовательный кусок данных.

Производительность чтения и записи не увеличивается для одиночных операций считывания/записи. Но если несколько пользователей используют устройство, может возникнуть ситуация, при которой один пользователь может фактически использовать первый диск, а другой пользователь обращаться к файлам на втором диске. Если это произойдет, будет получен прирост производительности.

RAID-0 (режим «stripe»). Аналогичен линейному режиму, исключая то, что чтение и запись производятся параллельно с двух устройств. Устройства должны иметь приблизительно одинаковый размер. Так как весь доступ производится параллельно, устройства заполняются одинаково. Если одно устройство больше, чем другие, это дополнительное пространство все еще используется в RAID устройстве, но при записи в самом конце RAID устройства, получается доступ только к этому одному диску, что, конечно, снижает производительность.

Как и в линейном режиме, на этом уровне нет никакой избыточности. В отличие от линейного режима, невозможно восстановить никакие данные при отказе диска. Если удалить диск из RAID-0 набора, в RAID устройстве будет не просто отсутствовать последовательный кусок данных, оно будет заполнено маленькими дырочками по всему устройству. Производительность чтения и записи увеличится, так как чтение и запись будут выполняться параллельно на дисках. Обычно, это главная причина использования этого уровня RAID.

RAID-1. Это первый режим, который реализует избыточность. RAID-1 может использоваться на двух или более дисках с нулем или более резервными дисками. Этот режим поддерживает точную копию ин-

формации одного диска на всех дисках. Диски должны быть одного размера. Если один из дисков больше другого, RAID будет иметь размер наименьшего диска. Если N-1 диск удален (или отказал), все данные все еще целы. Если имеются резервные диски, и если система (SCSI драйвера или IDE чипсет и т.п.) пережили отказ, после обнаружения отказа, начинается немедленная реконструкция зеркала на резервные диски.

Производительность записи немного хуже, чем у одного диска, так как на каждый диск массива должны быть посланы идентичные копии записанных данных. Производительность чтения обычно достаточно хорошая, начиная с ядра 2.4, так как в нем реализована улучшенная стратегия балансировки чтения.

RAID-4. Этот уровень RAID не часто используется. Он может быть использован с тремя или более дисками. Вместо полной зеркализации информации, он сохраняет информацию о четности на отдельном диске, и записывает данные на другой диск подобным, используемому в RAID-0 образом. Так как один диск зарезервирован для информации четности, размер массива будет (N-1)*S, где S – размер наименьшего устройства в массиве. Как и в RAID-1, диски должны быть либо одного размера, либо S, в формуле (N-1)*S, должно быть размером наименьшего диска в массиве. Если один диск откажет, информация о четности может быть использована для восстановления всех данных. Если два диска откажет – все данные будут потеряны.

Причина нечастого использования этого уровня – информация о паритете хранится на одном диске. Эта информация должна быть обновлена каждый раз, когда ведется запись на один из других дисков. Таким образом, диск с паритетом становится узким местом, если он не намного быстрее остальных дисков. Однако если так случилось, что есть много медленных дисков и один очень быстрый – этот уровень RAID может быть очень полезен.

RAID-5. Это режим RAID для тех случаев, когда необходимо соединить несколько физических дисков, и к тому же сохранить избыточность. RAID-5 может быть использован на трех или более дисках, с нулем или более резервных дисков. Размер результирующего RAID-5 устройства будет (N-1)*S, как и в RAID-4. Главное отличие между RAID -5 и -4 в том, что распределением информации о паритете по всем устройствам, избегается проблема узкого места.

Если один из этих дисков отказывает, все данные все еще не повреждены, благодаря информации о паритете. Если имеются резервные диски, при отказе диска немедленно начинается реконструкция. Если отказывает два диска одновременно – все данные потеряны. RAID-5 может пережить отказ одного диска, но не двух или более. Обычно увеличивается производительность как чтения, так и записи, но проблематич-

но предсказать насколько.

Резервные диски. Резервные диски – это диски, которые не являются частью RAID тома, пока один из активных дисков откажет. Когда обнаруживается отказ диска, он маркируется как «плохой» и, если имеются резервные диски, немедленно начинается реконструкция. Таким образом, резервные диски добавляют дополнительную безопасность, особенно к RAID-5 системам, где, возможно, тяжело достичь этого (физически). Это позволяет работать системе некоторое время, с отказавшим диском, так как вся избыточность – это наличие резервных дисков.

Установка программного обеспечения RAID. Для любых уровней RAID необходимы ядро и RAID утилиты. Все необходимые компоненты включены в большинство дистрибутивов Linux. Если система поддерживает RAID, то должен существовать файл /proc/mdstat. В нем указано, какие зарегистрированы RAID режимы, и какие устройства RAID уже активны.

Следующим шагом необходимо создать разделы, которые будут включены в RAID набор. Ниже рассмотрена специфика режимов.

Линейный режим. Резервные диски в данном случае не поддерживаются. Если диск выйдет из строя, массив выйдет из строя вместе с ним. Не существует информации для помещения на резервный диск.

Для создания массива необходимо запустить команду: mkraid /dev/md0. Эта команда инициализирует массив, записывает отдельные суперблоки, и запускает массив.

Следующим этапом необходимо создать файловую систему (как и на любом другом устройстве), смонтировать ее, включить в файл fstab и продолжить работу.

RAID-0. Применяется в случае, когда имеются два или более устройств, приблизительно одного размера, и необходимо объединить их емкость и производительность путем параллельного доступа. RAID-0 не имеет избыточности, поэтому в случае, если диск выйдет из строя, массив выйдет из строя вместе с ним.

Запускается командой: mkraid /dev/md0 для инициализации массива. Это позволяет инициализировать суперблок и запустить устройство.

RAID-1. Применяется в случае, когда есть два устройства приблизительно одного размера, и необходимо их зеркалировать. Кроме этого можно использовать большее количество дисков в качестве запасных, если одно из активных устройств выйдет из строя, то запасной диск автоматически станет частью дискового массива.

Следующим шагом происходит инициализация RAID. Зеркало должно быть сконструировано, содержимое двух дисков должно быть синхронизировано. Должно появиться сообщение о том, что зеркало начало реконструироваться, а также определено оценочное время завер-

шения реконструкции. Реконструкция делается в периоды отсутствия ввода-вывода. При этом система должна обладать малым временем отклика, хотя индикатор дисковой активности должен почти непрерывно светиться. Процесс реконструкции незаметен, так что допускается использовать зеркало, несмотря на реконструкцию. Допускается также форматировать устройство при запущенной реконструкции. Также разрешено монтировать его и использовать в процессе реконструкции. Необходимо учесть тот факт, что если неисправный диск разрушается при реконструкции, то возникнут проблемы его использования.

RAID-4. Применяется в случае, когда есть три или более, приблизительно одного размера, диска, один значительно быстрее других, и требуется скомбинировать их все в одно большое устройство, которое содержит немного избыточной информации. При этом допускается присутствие несколько устройств, которые используются как резервные диски.

RAID-5. Применяется в случае, когда есть три или более дисков приблизительно одного размера, нужно скомбинировать их в большое устройство, содержащее некоторую степень избыточности. Допускается использование еще нескольких дисков в качестве резервных, которые не будут являться частью массива до отказа другого устройства. Если используется N дисков, то размер всего массива будет (N-1)*S, где S – размер наименьшего. Данное пространство не включает дисковое пространство, используемое для информации о четности (избыточности). В результате применения данного типа RAID, в случае, если любой диск отказывает, все данные остаются целыми. Однако необходимо учитывать тот факт, что если два диска отказывают одновременно, то все данные будут потеряны. В случае наличия резервных дисков, они добавляются согласно спецификациям raid-disk.

Размер куска (chunk-size) в 32 кБ является оптимальным начальным значением для многих общих применений файловой системы. Массив, на котором используется вышеуказанный raidtab, содержит файловую систему ext2 с размером блока 4 кБ. Если файловая система намного больше или на ней будут храниться очень большие файлы, то необходимо установить больший размер куска и размер блока файловой системы. Массив не устойчив, пока фаза реконструкции не завершена. Однако массив полностью функционален (кроме обработки дисковых отказов), и его допускается форматировать и использовать, в то время пока он реконструируется. Перед форматированием массива, необходимо учесть секцию специальных опций. После того, как RAID устройство запущенно, имеется возможность остановить его или снова запустить. Вместо помещения команд в init-файлы и многократных перезагрузок необходимо запустить автодетектирование.

Отдельный суперблок. В старых версиях, raidtools необходимо было прочитать /etc/raidtab файл и затем инициализировать массив. Однако это требовало наличия файловой системы, на которой был смонтирован /etc/raidtab. Это являлось недопустимым для загрузки с RAID. Старый подход приводил также к сложностям при монтировании файловой системы на RAID устройствах. Не допускалось вставлять их в /etc/fstab файл, а приходилось монтировать из скриптов инициализации. Отдельный суперблок позволяет решить эти проблемы. При инициализации с опцией persistent-superblock в файле /etc/raidtab в начале всех дисков массива записывается специальный суперблок. Это позволяет ядру читать конфигурацию устройств RAID прямо с затрагиваемых дисков вместо чтения конфигурационного файла, который может быть не всегда доступен. Однако необходимо поддерживать целостность файла /etc/raidtab, так как он может понадобиться позже при реконструкции массива. При необходимости автоматического детектирования RAID устройств при загрузке наличие отдельного суперблока является обязательным.

Размер кусков. Если имеются два диска и необходимо записать байт, то, фактически, нужно записать четыре бита на каждый диск, каждый второй бит записывается на диск 0, а другие на диск 1. Аппаратно данная возможность не поддерживается. Вместо этого выбирается некоторый размер куска, который определяется как наименьшая «атомарная» порция данных, которые могут быть записаны на диски. Запись 16 кБ с размером куска в 4 кБ, приведет к записи первой и третьей порции данных размером 4 кБ на первый диск, а второй и четвертой на второй диск (в случае RAID-0 из двух дисков). Такой режим для длинных записей уменьшает накладные расходы при довольно больших размерах кусков, в то время как массивы, которые содержат небольшие файлы, имеют преимущество при небольших размерах куска. Размеры куска должны быть указаны для всех уровней RAID, включая линейный режим. Для линейного режима этот параметр игнорируется. Для оптимальной производительности необходимо определить это значение экспериментальным образом, также как и размер блока файловой системы, которая создается в массиве.

RAID-0. В данном режиме данные записываются на диски массива почти в параллельном режиме. Фактически же, chunk-size байт записываются на каждый диск последовательно. Если указан размер куска в 4 кБ и пишется 16 кБ на массив из трех дисков, RAID система произведет запись 4 кБ на диски 0, 1 и 2 параллельно, а оставшиеся 4 кБ запишет на диск 0.

Размер куска в 32 кБ является оптимальным начальным значением для большинства массивов. Но оптимальное значение сильно зависит от

количества дисков, содержимого файловой системы на массиве и многих других факторов.

RAID-1. На скорость записи размер куска не влияет, так как все данные должны быть записаны на все диски. Однако для чтения размер куска указывает сколько данных читаются последовательно с участвующих дисков. Так как все диски массива содержат одинаковую информацию, то чтение может быть произведено параллельно, подобно RAID-0.

RAID-4. Когда сделана запись на массив RAID-4, обновляется информация о паритете на паритетном диске. Размер куска – размер паритетных блоков. Если байт записывается на массив RAID-4, потом chunk-size байт считываются с N-1 дисков, вычисляется информация о паритете и chunk-size байт записываются на паритетный диск. Размер куска влияет на производительность чтения также как и в RAID-0, так как считывания с RAID-4 делаются аналогично.

RAID-5. Размер куска имеет такое же значение, как и в RAID-4. Разумный размер куска для RAID-5 массива – 128 кБ, однако допускается корректировать его экспериментальным путем. Так же на производительность RAID-5 влияют опции форматирования файловой системы (секция специальных опций mke2fs).

Опции mke2fs. Существует специальная опция форматирования RAID-4 и RAID-5 устройств с mke2fs. Опция R stride=nn позволяет утилите mke2fs лучше размещать различные ext2 специфичные структуры данных на устройство RAID. Если размер куска 32 кБ, это значит, что 32 кБ последовательных данных будут лежать на одном диске. Если нужно создать ext2 файловую систему с размером блока в 4 кБ, получится, что восемь блоков файловой системы будут в одном куске. Эта информация указывается для утилиты mke2fs при создании файловой системы mke2fs -b 4096 -R stride=8 /dev/md0.

Производительность RAID-{4,5} строго зависит от этой опции. Размер блока ext2fs определяет производительность файловой системы. Желательно использовать размер блока 4кБ на любой файловой системе более чем нескольких сот мегабайт, если не предусматривается помещать очень большое число маленьких файлов на нее.

Автодетектирование. Автодетектирование позволяет ядру автоматически распознавать устройства RAID при загрузке, сразу после завершения обычного детектирования разделов. Для этого требуется выполнение следующих пунктов:

- обеспечить поддержку автодетектирования в ядре;
- создать RAID устройства, используя отдельный суперблок;

– установить в 0xFD (запустить fdisk и установить тип в "fd") тип раздела устройств, используемых в RAID. Перед сменой типа раздела нужно удостовериться, что RAID не запущен.

После выполнения указанных выше трех пунктов устанавливается авто-детектирование. После загрузки системы, в /proc/mdstat указывается, что RAID запущен. При загрузке на экран выдаются диагностические сообщения. Автоматически стартующие устройства также автоматически останавливаются при выключении. Нет необходимости указывать данные устройства в init скриптах. Устройства /dev/md используются как любые другие /dev/sd или /dev/hd устройства. В init-скриптах могут оказаться некие raidstart/raidstop команды. Они часто имеются в стандартных RedHat init скриптах. Данные команды используются для RAID старого типа и не используются в RAID нового типа с авто-детектированием. Эти строки допускается удалить.

Корневая файловая система на RAID. В случае загрузки системы с RAID корневая файловая система (/) монтируется на устройство RAID. Некоторые из дистрибутивов поддерживают установку корневой файловой системы на устройство RAID, что существенно упрощает начальную конфигурацию RAID системы.

С учетом изложенного можно сделать вывод, что минимальная конфигурация системы внешней памяти для кластерной установки должна состоять из стандартных дисков с применением программного RAID массива. Более предпочтительным является решение на основе NAS с применением аппаратного RAID массива.

4.6. Методы автоматической установки операционной системы Linux для суперкомпьютеров семейства «СКИФ»

Для ускорения процесса установки предпочтительно пользоваться автоматическими методами установки операционной системы Linux на компьютеры. Этот метод установки является также наиболее подходящим для подготовки вычислительных узлов кластеров. В Red Hat Linux такой метод называется Kickstart Installation. Используя этот метод, системный администратор создает один файл, содержащий ответы на все вопросы, которые обычно задаются во время установки. Затем этот файл размещается на серверной машине и используется при установке операционной системы на другие машины (вычислительные узлы кластера).

Автоматическая установка для Red Hat Linux позволяет автоматизировать следующие этапы установки:

– выбор языка;
– конфигурация мыши;
– конфигурация клавиатуры;
– установка программы начальной загрузки;

– разбиение жесткого диска;
– конфигурация сетевых интерфейсов;
– конфигурация NIS, LDAP, Kerberos, Hesiod и Samba;
– конфигурация фаервола;
– выбор пакетов для установки;
– конфигурация графической системы X Window System.

Выполнение автоматической установки. Автоматическая установка выполняется с использованием локального CD-ROMa, локального жесткого диска или посредством NFS, FTP или HTTP. Для использования метода автоматической установки необходимо:

– создать файл автоматической установки (kickstart file);
– создать загрузочный диск или сделать файл автоматической установки доступным для компьютера по сети;
– выполнить установку с использованием файла автоматической установки.

Создание файла автоматической установки. Файл автоматической установки – это обычный текстовый файл, содержащий список элементов, каждый из которых идентифицируется ключевым словом. Он создается редактированием файла sample.ks, находящегося на диске Red Hat Linux Documentation, с использованием приложения Kickstart Configurator или должен быть создан самостоятельно. Программа установки Red Hat Linux также создает образец файла автоматической установки, основываясь на опциях, которые были выбраны во время установки Linux. Этот файл находится в /root/anaconda-ks.cfg. Это обычный текстовый файл, который можно отредактировать с помощью любого текстового редактора. Создавая файл автоматической установки самостоятельно необходимо учитывать следующее:

– элементы должны указываться в определенном порядке – секция команд, секция пакетов, секции преустановки и постустановки (последние две секции могут идти в любом порядке и не являются обязательными);
– элементы, которые не являются обязательными, могут быть опущены;
– пропуск любого обязательного элемента приведет к тому, что программа установки запросит недостающие параметры у пользователя во время установки;
– строки, начинающиеся со знака ‘#’ распознаются как комментарии и игнорируются;
– для апгрэйдов требуются следующие элементы: язык, метод установки, спецификация устройства (если оно необходимо для установки), настройки клавиатуры, ключевое слово upgrade, конфигурация

LILO. Если для апгрэйда указываются другие элементы, то они будут проигнорированы (включая выбор пакетов).

При использовании графического интерфейса для создания файла автоматической установки необходимо использовать приложение Kickstart Configurator.

Опции файла автоматической установки устанавливают опции аутентификации для системы. Значения этой опции аналогичны команде authconfig, которая может быть запущена после выполнения установки. По умолчанию пароль шифруется и не используется файл /etc/shadow. Опции также обеспечивают поддержку NIS.

Поддержка LDAP в файле /etc/nsswitch.conf позволяет системе получать информацию о пользователях (UID, домашние директории, оболочки, и т.д.) из LDAP каталога. Для аутентификации и изменения паролей используется LDAP каталог. Опция TLS (Transport Layer Security) поиск позволяет LDAP отправлять шифрованные имя пользователей и пароли серверу LDAP перед аутентификацией. Возможно испольование Kerberos 5 для аутентификации пользователей. Сам Kerberos не имеет информации о домашних директориях, UID или оболочках. Поэтому, если используется Kerberos, то необходимо использовать LDAP, NIS, или Hesiod, или создавать пользовательские аккаунты.

Поддержка Hesiod используется для поиска информации о пользовательских домашних директориях, UID, оболочках. Hesiod – это расширение DNS, которое использует записи DNS для хранения информации о пользователях, группах и т.д.

Опция Hesiod LHS («left-hand side») изменяет опции в файле /etc/hesiod.conf. Эта опция используется библиотекой Hesiod для определения имени для поиска в DNS, аналогично тому, как в LDAP используется базовый DN.

Опция Hesiod RHS («right-hand side») изменяет опции в файле /etc/hesiod.conf. Эта опция используется библиотекой Hesiod для определения имени для поиска в DNS, аналогично тому, как в LDAP используется базовый DN.

Возможно использование поддержки аутентификации пользователей через SMB сервер (обычно Samba или Windows сервер). SMB аутентификация не позволяет получить информацию о домашних директориях, UID или оболочках. Поэтому, если используется этот тип аутентификации, необходимо использовать LDAP, NIS, или Hesiod, или создавать пользовательские аккаунты. Для SMB аутентификации, если указывается более, чем один сервер, имена указываются через запятую.

Nscd сервис кеширует информацию о пользователях, группах и другую информацию. Кеширование особенно полезно при распространении информации о пользователях и группах через сеть, при использо-

вании NIS, LDAP или hesiod.

Опции bootloader (required) указывает настройки загрузчика и какой использовать LILO или GRUB. Если используется GRUB, то устанавливается пароль для GRUB в mypassword. Установка пароля позволяет ограничить доступ к командной строке GRUB, где могут быть указаны различные опции ядра. Если используется GRUB, то пароль mypassword должен быть уже зашифрованным.

Опция обновить существующую конфигурацию загрузчика доступна только при обновлении системы. Для большинства PCI систем программа установки автоматически правильно определяет сетевые и SCSI карты. Для некоторых PCI систем при автоматической установке необходима подсказка, чтобы правильно определить устройства. Во время автоматической установки могут быть использованы дискеты с драйверами. Необходимо скопировать их содержимое в корневую директорию раздела на жестком диске.

Опциия Install сообщает программе установки о необходимости установки системы заново вместо обновления. Этот режим используется по умолчанию. Опция Lang (required) устанавливает язык, используемый во время установки. Если нужен только один язык, необходимо указать его. Если устанавливается поддержка нескольких языков, то необходимо указать язык, используемый по умолчанию. Например, устанавливается поддержка русского и английского языка, причем английский используется по умолчанию.

В опциях Network (optional) указывается информация о сетевых настройках. Если установка требует использование сети (установка через NFS, HTTP или FTP) и информация о сетевых настройках не указана в конфигурационном файле, то программа установки пытается использовать интерфейс eth0 с динамическим IP адресом (пытается получить адрес от BOOTP или DHCP сервера) и конфигурирует устанавливаемую систему для автоматического определения IP адреса.

Строка для статических настроек сети более сложная, так как в этой строке необходимо указать все параметры сетевых настроек. Следует указать IP адрес, маску подсети, IP адрес шлюза, IP адрес DNS сервера. При использовании метода STATIC, возникает два ограничения. Вся информация должна быть указана в одной строке; нельзя разбивать строку, используя, например, символ обратной наклонной черты. При этом допускается указывать только один сервер DNS. Тем не менее, можно использовать раздел %post конфигурационного файла автоматической установки для добавления дополнительных серверов.

Опция Part или partition (обязательна при установке, игнорируется при обновлении) предназначена для управления разделами жестких дисков. Если более чем одна установка Red Hat есть на разных разделах, то

программа установки запросит пользователя, какую из них обновлять.

Опция Raid (optional) указывает программе установки, что необходимо создать программный RAID. Имя RAID устройства может быть от md0 до md7, и может использоваться только один раз. Запасные диски будут использованы для восстановления массива в случае сбоя.

Опция reboot (optional) позволяет выполнить перезагрузку после установки. Без этой опции программа установки после установки выводит сообщение и ждет нажатия клавиши от пользователя.

Опция Rootpw (required) устанавливает пароль суперпользователя в <password>.

Если используется опция Timezone (required), то устанавливается часовой пояс, который может быть любым из списка в программе timeconfig. Если присутствует опция utc, то время часов компьютера будет выставлено в соответствии со временем нулевого мередиана (UTC).

Опция resolution <res> определяет разрешение для X Window System по умолчанию. Значение может быть 640x480, 800x600, 1024x768, 1152x864, 1280x1024, 1400x1050, 1600x1200. Необходимо предварительно убедиться, что заданное разрешение поддерживается монитором и видеокартой.

Опция depth <cdepth> определяет глубину цвета для X Window System по умолчанию. Значение может быть 8, 16, 24 или 32. Необходимо убедиться, что заданная глубина цвета поддерживается монитором и видеокартой.

Если указана опция zerombr с аргументом yes, то все найденные неверные таблицы разделов будут переписаны. Эта операция уничтожит все содержимое дисков с неверными таблицами разделов. Эта команда должна быть в формате zerombr yes. Другие форматы не поддерживаются.

Опция %packages конфигурационного файла автоматической установки начинает раздел, в котором расположен список пакетов, которые требуется установить (этот раздел используется только для установки, а при обновлении системы выбор пакетов не поддерживается).

Опция %pre позволяет задать команды, которые будут выполнены сразу после разбора файла конфигурации ks.cfg. Этот раздел должен быть в конце конфигурационного файла и должен начинаться с опции %pre. Несмотря на наличие доступа к сети к моменту выполнения раздела %pre, DNS сервис еще не сконфигурирован в этот момент и поэтому можно использовать только IP адреса.

Опция %post позволяет задать команды, которые будут выполнены после завершения установки системы. Этот раздел должен быть в конце конфигурационного файла и должен начинаться с опции %post. Режим удобно использовать для установки дополнительных программ конфи-

гурации дополнительных DNS серверов. Если сеть сконфигурирована со статическими IP адресами, включая DNS сервер, то имеется доступ к сети и DNS серверу во время выполнения раздела %post. Если сеть была сконфигурирована через DHCP, то файл /etc/resolv.conf не будет заполнен, когда программа установки выполняет раздел %post. Хотя имеется доступ к сети, тем не менее, DNS сервис еще не сконфигурирован в этот момент, поэтому можно использовать только IP адреса. Необходимо отметить, что раздел %post выполняется в окружении с измененным корнем, поэтому выполнение задач таких, как копирование скриптов или RPM пакетов с установочного носителя работать не будет.

Опция %include используется, чтобы включить в конфигурационный файл содержимое другого файла. Файл автоматической установки необходимо поместить в одно из двух предопределенных мест – на загрузочную дискету или на сетевой диск. Наиболее распространенным является сетевой подход, так как большинство автоматических установок выполняется на удаленных компьютерах (в данном случае таковыми являются вычислительные узлы кластера).

В случае использования загрузочной дискеты, файл автоматической установки должен иметь имя ks.cfg и должен находиться в каталоге верхнего уровня дискеты. Использование файла автоматической установки на сетевом диске является обычным режимом, так как позволяет автоматизировать процесс установки на многих компьютерах, соединенных сетью. Для данного подхода необходимо иметь в сети BOOTP/DHCP сервер и NFS сервер. Первый используется для того, чтобы дать клиенту всю необходимую сетевую информацию, в то время как сами файлы являются доступными с помощью NFS сервера. Часто оба сервера физически находятся на одном компьютере, но это не является обязательным.

Для выполнения автоматической установки по сети необходимо иметь в сети BOOTP/DHCP сервер и он должен содержать конфигурационную информацию для компьютера, на котором будет произведена установка. BOOTP/DHCP сервер предоставляет клиенту сетевую информацию и данные о местоположении файла автоматической установки.

Если файл автоматической установки указывается с помощью BOOTP/DHCP сервера, клиент пытается смонтировать указанный диск как NFS диск и скопировать файл к себе, используя его затем как файл автоматической установки. Точные установки сервера зависят от используемого вами BOOTP/DHCP сервера. Необходимо подставить свои значения после ключевых слов filename (указать имя файла автоматической установки или путь к нему) и next-server (указать имя NFS сервера). Если имя файла, возвращаемое BOOTP/DHCP сервером оканчивается на

слэш (/), то оно интерпретируется только как путь. В этом случае клиент монтирует путь как NFS диск и делает поиск файла. Для начала автоматической установки необходимо загрузиться с загрузочной дискеты или CD-ROMa и ввести специальную загрузочную команду на начальный запрос при загрузке.

Проведенные исследования позволили создать в рамках программы «СКИФ» высокопроизводительные кластерные конфигурации «СКИФ К-500» и «СКИФ К-1000», которые на момент создания по своим техническим характеристикам соответствовали мировому уровню в области высоких суперкомпьютерных технологий. Это подтверждается результатами тестов, на основании которых эти суперкомпьютеры были включены в соответствующие выпуски списка 500 самых мощных вычислительных систем мира.

4.7. Суперкомпьютерные конфигурации «СКИФ К-500» и «СКИФ К-1000»

Суперкомпьютерная установка «СКИФ К-500» (экспериментальный образец) с пиковой производительностью 716,8 миллиардов операций в секунду была создана в 2003 году. Кластер «СКИФ К-500» (рис.4.17) создан для отработки принципов построения моделей суперкомпьютеров «СКИФ» с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду).

Технические характеристики суперкомпьютера «СКИФ К-500»:

Предельная пиковая (реальная на задаче Linpack) производительность:	716,8(474,2)Gflops
Тип процессора:	Intel Xeon 2.8 Ghz
Число вычислительных узлов/процессоров:	64/128
Оперативная память узла/установки:	64*2=128 GB
Дисковая память установки:	64*60=3840 GB
Тип системной сети:	3D-тор, SCI, D336
Тип вспомогательной сети:	GB Ethernet
Конструктив узла (форм-фактор):	1U

Рис. 4.17. Суперкомпьютерная конфигурация «СКИФ К-500»

Кластер «СКИФ К-500» особенно эффективен при решении задач с интенсивным межузловым обменом.

Структурная схема «СКИФ К-500» представлена на рис. 4.18. Кластер «СКИФ К-500» состоит из:

– 64 вычислительных узлов, расположенных в шкафах блоками по 16 узлов в каждом;
– управляющей ПЭВМ;
– дополнительной дисковой памяти (файл-сервер);
– системного сетевого интерфейса SCI;
– вспомогательного сетевого интерфейса GB Ethernet;
– сервисной сети RS-485;

– 4-х 24 портовых коммутаторов GB Ethernet;
– 4-х канального коммутатора KVM;
– источников бесперебойного питания.

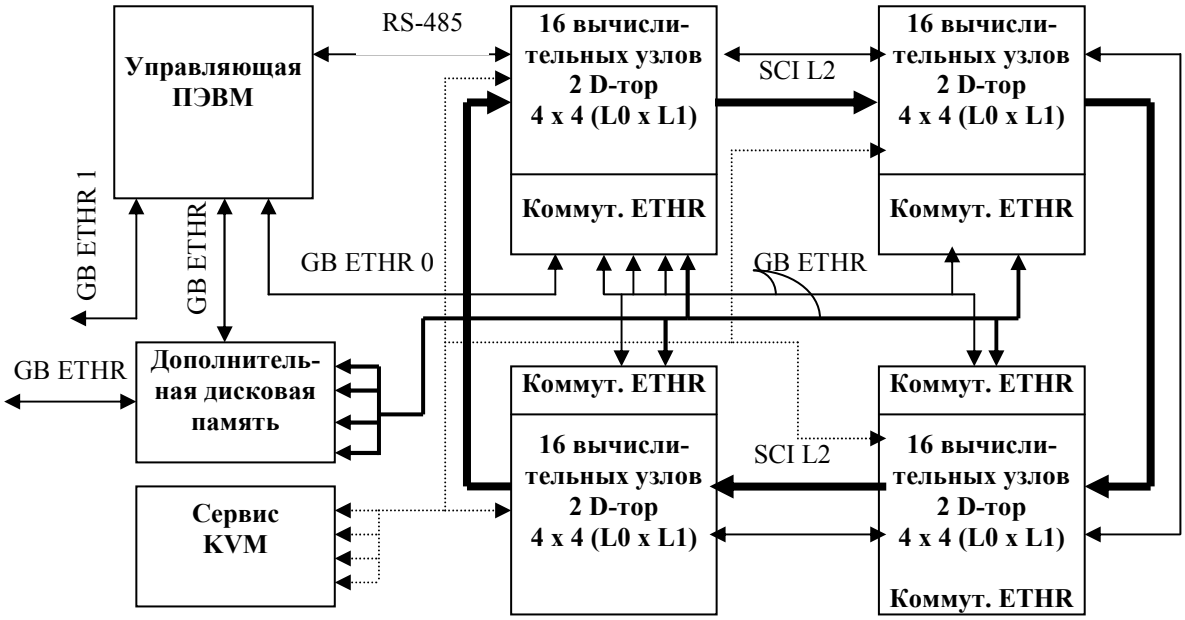

Рис. 4.18. Структурная схема «СКИФ К-500»

Каждый вычислительный узел кластера включает:
– 2 процессора Intel Xeon 2,8 ГГц;
– оперативная память – 2 Гбайт;
– дисковую память – 60 Гбайт HDD;
– адаптер SCI фирмы Dolphin 236 64 бит / 66 МГц.

Пиковая производительность каждого узла – 11,2 Гфлопс. SCI интерфейс обеспечивает задержку при передаче сообщений в соответствии со стандартом MPI – 3 мкс, скорость обмена узел-узел – 263 Мбайт/с.

Вычислительные узлы соединяются высокоскоростной сетью SCI и образуют трехмерный тор (3D-тор) 4 x 4 x 4 (L0 x L1 x L2) – кубической структуры. Блоки по 16 узлов в отдельных шкафах соединены в двухмерный тор (2D-тор 4 x 4). 3D-тор образуется с помощью 16 колец внешних связей SCI интерфейса между соответствующими вычислительными узлами разных шкафов по линку L2.

Вспомогательный сетевой интерфейс GB Ethernet 0 предназначен для загрузки программ, данных, управления и мониторинга. Он имеет звездообразную топологию, в которой к коммутатору соединенному с управляющей ПЭВМ подключены 3 коммутатора остальных шкафов. Вычислительные узлы одного шкафа подключаются к своему коммута-

тору.

GB Ethernet 1 предназначен для доступа из внешней локальной сети к управляющей ПЭВМ. Дополнительная дисковая память (файловый сервер) подключена к каждому из 4-х коммутаторов отдельным GB Ethernet каналом. С управляющей ПЭВМ она связана по GB Ethernet перекрестным кабелем. Она использует технологию хранения данных RAID5 и имеет общую емкость дисковой памяти около 800 Гбайт.

Сервисная сеть RS-485 предназначена для выполнения включения/выключения электропитания вычислительного узла, аппаратного сброса узла и выполнения взаимодействия с вычислительным узлом в консольном режиме с управляющей ПЭВМ.

KVM обеспечивает подключение любого из вычислительных узлов каждого шкафа к общей клавиатуре, видеомонитору, «мыши» для сервисного обслуживания.

Программное обеспечение кластера:

– операционная система LINUX RED HAT с поддержкой SMP;

– система управления, администрирования кластером и поддержка стандарта MPI 1.2 фирмы Scali (SSP 3.0.1);

– компиляторы C, C++;

– система программирования и управления выполнением вычислений – T-система.

Суперкомпьютер «СКИФ К-1000» (рис. 4.19) создавался в 2004 году с целью комплексной реализации принципов построения моделей суперкомпьютеров «СКИФ» Ряда 2 кластерного уровня с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду), включая сетевые (метакластерные) конфигурации. Суперкомпьютер «СКИФ К-1000» является старшей моделью семейства «СКИФ». При выборе технических параметров суперкомпьютера «СКИФ К-1000» ставилась задача обеспечить его включение в очередной выпуск списка Top-500.

Всесторонне прорабатывались требования к конструкции суперкомпьютера «СКИФ К-1000». В частности, базовые конструктивы должны были быть реализованы с использованием серийных изделий массового применения широко распространенных на рынке персональных компьютеров – микропроцессоров, материнских плат, модулей ОЗУ, НМД, коммуникационных средств и т.д.

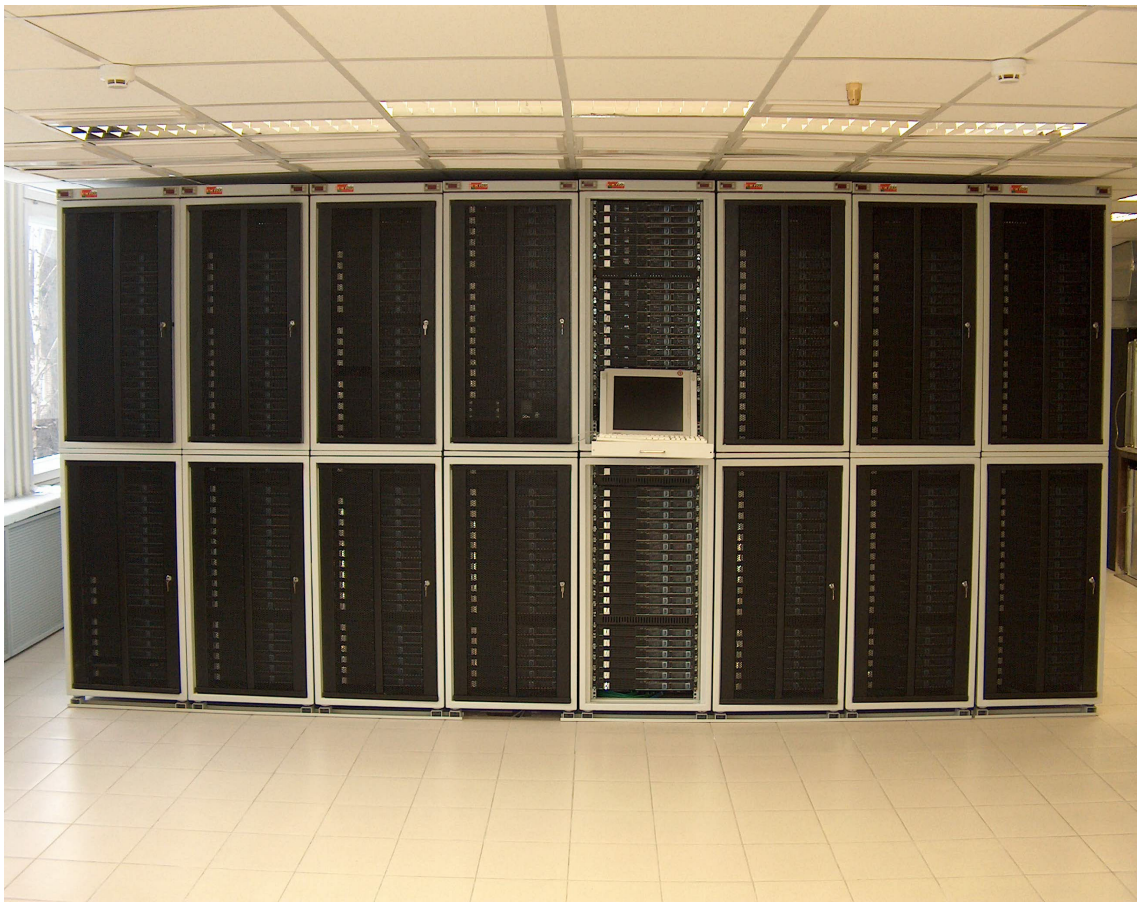

Рис. 4.19. Суперкомпьютерная конфигурация «СКИФ К-1000»

Технические характеристики суперкомьютерной конфигурации «СКИФ К-1000»:

Число вычислительных узлов/процессоров	288/576
Тип процессора	AMD Opteron™ 2.2 Ггц
Пиковая производительность	2.534 Tflops
Производительность на тесте Linpack	2.032 Tflops (80.1% от пиковой)
Тип системной сети	Infiniband (пропускная способность на уровне MPI ~800Mb/sec; задержка на уровне MPI ~5.5 μsec)
Тип управляющей (вспомогательной) сети	Gigabit Ethernet
Тип сервисной сети	СКИФ-ServNet v.2.0

Оперативная память	288 x (8 x 0.5 GB) = 1 152 GB
Дисковая память	288 x 80 GB = 23 040 GB
Потребляемая мощность в режиме максимальной нагрузки	89КВт
Потребляемая мощность в режиме простоя	73КВт
Уровень шума	84 Дб
Производительность системы охлаждения	16000 $м^3$/час
Вес установки	6.5 Т
Суммарная длина кабельных соединений	более 2км

Конструкция «СКИФ К-1000», построенная по блочному принципу, содержит:

– корпус типа Rack-mounted для размещения одного базового вычислительного модуля (БВМ КУ) или дополнительной дисковой системы;

– базовый конструктивный модуль (шкаф) для размещения в нем БВМ КУ и вспомогательных средств (коммутаторов, средств контроля за состоянием системы, системы электропитания, системы вентиляции и др.).

Большинство из технических решений, использованных в суперкомпьютере «СКИФ К-1000», являлись передовыми для суперкомпьютерной отрасли того времени, например:

– были использованы 64-битовые процессоры AMD Opteron 248 (2200 MHz);

системная сеть базируется на технологии Infiniband 4x, что обеспечивает высокие показатели при работе MPI-приложений: скорость MPI-обменов достигает 830 Мбайт/сек, задержка передачи коротких пакетов составляет около 5 мкс.

Структура суперкомпьютера «СКИФ К-1000» включает:

1) Вычислительную подсистему:

– вычислительный узел (8x36=288 шт.): 1U, 2xAMD Opteron (2.2 Ghz), RAM: 4GB, HDD IDE 80GB,IB 4x Mellanox HCA MHXL-CF128 (подключен к “Leaf” IB-коммутатору), 2xGbEthernet (одна линия к “Leaf” коммутатору), СКИФ-Servnet v.2;

– управляющий узел (1 шт.): 2U, CPU:2xAMD Opteron 248 (2.2 Ghz), RAM: 4GB, HDD;2x36GB, SCSI 10K RPM,HotSwap, DVD/CD,

FDD, Moxa CP104UL 4 port RS232 LP Universal card + 4 ведущие платы СКИФ-Sernet v.2, 2xGbEthernet (одна линия к одному из “Core” Eth-коммутаторов).

2) Системную сеть Infiniband 4x (MPI:~830 MB/s, ~ 5μs)

“Core” IB-коммутатор Mellanox MTS-2400 (2x6=12 штук): 1U, 24 порта, по одной линии на каждый из 24 “Leaf” IB-коммутаторов;

“Leaf” IB-коммутатор Mellanox MTS-2400 (3x8=24 шт.): 1U, 24 порта, 12 линий на вычислительные узлы, 12 линий на “Core” коммутаторы (по одной линии на каждый из них).

3) Вспомогательную (управляющую) сеть GbEthernet

“Core” Eth-коммутатор D-Link DGS-3324SR (2 шт.): 1U, 24 порта, по 2 порта (2 Gbps trunk) на 8 “Leaf” Eth-коммутаторов, 40Gbps между “Core” Eth-коммутаторами;

“Leaf” Eth коммутаторов F-Link DGS-1224T (2x8=16 шт.): 1U (установлены по 2 шт. в 1U), 24 порта, 18 линий на вычислительные узлы, 2 порта (2 Gbps trunk) на “Core” Eth-коммутатор.

4) Сервисную сеть СКИФ-Servnet v.2. Платы Servnet в вычислительных узлах связаны в линию RS-485 (два шкафа: 2x36=72 штуки в линии) и подключены к одной из четырех ведущих плат Servnet в управляющем узле.

5. Прикладные комплексы и системы на базе суперкомпьютеров «СКИФ»

5.1. Аппаратно-программный кардиологический комплекс

Аппаратно-программный кардиологический комплекс для выявления лиц с повышенным риском развития ишемической болезни сердца и артериальной гипертензии методом конъюнктивальной биомикроскопии разработан ОИПИ НАН Беларуси совместно с Республиканским научно-практическим центром «Кардиология» (г. Минск). Руководитель темы – Анищенко В.В., основные исполнители работы – Лапицкий В.А., Константинова Е.Э., Спиридонов С.В.

Основной целью выполненной работы была разработка опытного образца аппаратно-программного комплекса для анализа и обработки данных в реальном масштабе времени в медицинских системах неинвазивной диагностики состояния сердечно-сосудистой системы на базе технических решений суперкомпьютеров семейства «СКИФ». Были предложены методы и алгоритмы предварительной обработки, сегментации биомикроскопических изображений и количественной оценки микроциркуляторного звена сердечно-сосудистой системы методом биомикроскопии на основе применения нейросетевых методов и алгоритмов. Произведен анализ алгоритмов и выбор подходов для их распараллеливания и реализации на вычислительной системе с параллельной архитектурой.

Проблема роста смертности от сердечно-сосудистых заболеваний, основное место в структуре которых занимает ишемическая болезнь сердца (ИБС), сохраняет глобальный характер. С учетом того факта, что «омоложение» данного заболевания имеет место в большинстве промышленно развитых стран, решение задачи повышения эффективности первичной профилактики ИБС имеет международное значение. Анализ результатов многоцентровых исследований указывает на то, что основным условием предупреждения развития и прогрессирования данного заболевания является активное выявление предпатологических состояний, позволяющее своевременно провести необходимые профилактические или, в ряде случаев, лечебные мероприятия. С этих позиций разработка диагностического подхода, направленного на оценку степени риска развития ИБС на доклинической стадии, является актуальной задачей.

Одним из наиболее перспективных методов исследования терминального сосудистого русла является конъюнктивальная биомикроскопия. Современный уровень развития цифровой и аналоговой техники, компьютерных методов сбора, обработки и хранения видеоинформации открывает реальные возможности для решения данных задач. Так, бла-

годаря увеличению быстродействия вычислительной техники, позволяющей использовать сложные, критичные во времени алгоритмы, и появлению цветных телевизионных датчиков высокого разрешения можно получать на персональном компьютере качественные изображения микроциркуляторного русла и проводить их автоматизированную обработку.

В процессе создания аппаратно-программного кардиологического комплекса (АПКК) были выполненны следующие работы:

1) Разработано специализированное программное обеспечение количественной оценки микроциркуляторного звена сердечно-сосудистой системы методом биомикроскопии с учетом методов параллельной обработки данных.

2) Разработана медицинская методика количественной оценки изображений бульбарной конъюнктивы в диагностике состояния микроциркуляции при патологии сердечно-сосудистой системы.

3) Создан опытный образец АПКК.

4) Проведена клиническая апробация опытного образца АПКК на базе РПНЦ «Кардиология».

5.1.1. Методы оценки функционального состояния микроциркуляции с учетом методов параллельной обработки данных

Для оценки изменений микроциркуляторного русла при проведении нагрузочных проб по предлагаемым методикам в программном обеспечении должна быть предусмотрена реализация одновременно возможностей совмещения кадров по характерным признакам и сравнения получаемых изображений в динамике (интервал, ориентировочно, 10-20 сек в течение 5-10 минут) до момента появления изменений, являющихся точкой прекращения воздействия. То есть при проведении нагрузочной пробы должна быть реализована следующая последовательность действий:

1) Регистрация одного и того же участка конъюнктивы на протяжении всего исследования.

2) Совмещение кадров серии (5-10) между точками оценки при воздействии. Расчет параметров в течение 10-20 секунд (до следующей серии).

Сравнение последующих результатов с предыдущими в процессе пробы каждые 10-20 секунд до момента регистрации изменений.

На рис. 5.1 приведены изображения бульбарной конъюнктивы (БК) до проведения пробы и в момент (через 7 минут от начала пробы) появления признаков проявления антитромбогенной активности сосудистой стенки у здорового человека.

На рис. 5.2-5.4. представлены изображения бульбарной конъюнктивы пациента, соответственно, до начала исследования, в момент достижения максимальной ЧСС и через минуты после прекращения нагрузки.

Рис. 5.1. Изображения БК до проведения (А) пробы и через 7 минут (Б).

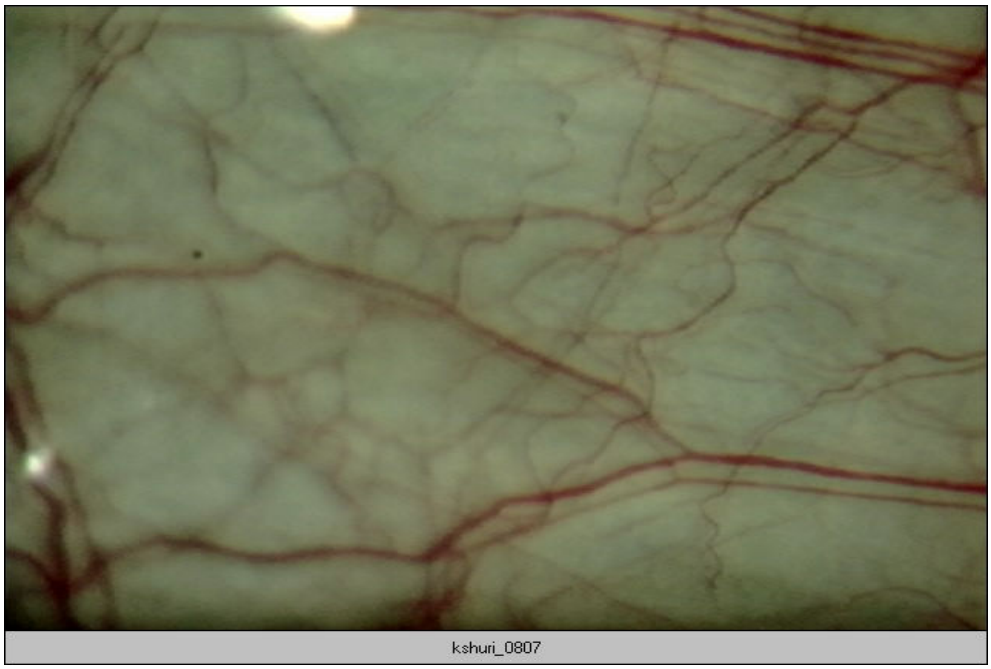

Рис. 5.2. Изображение бульбарной конъюнктивы пациента до начала ВЭП

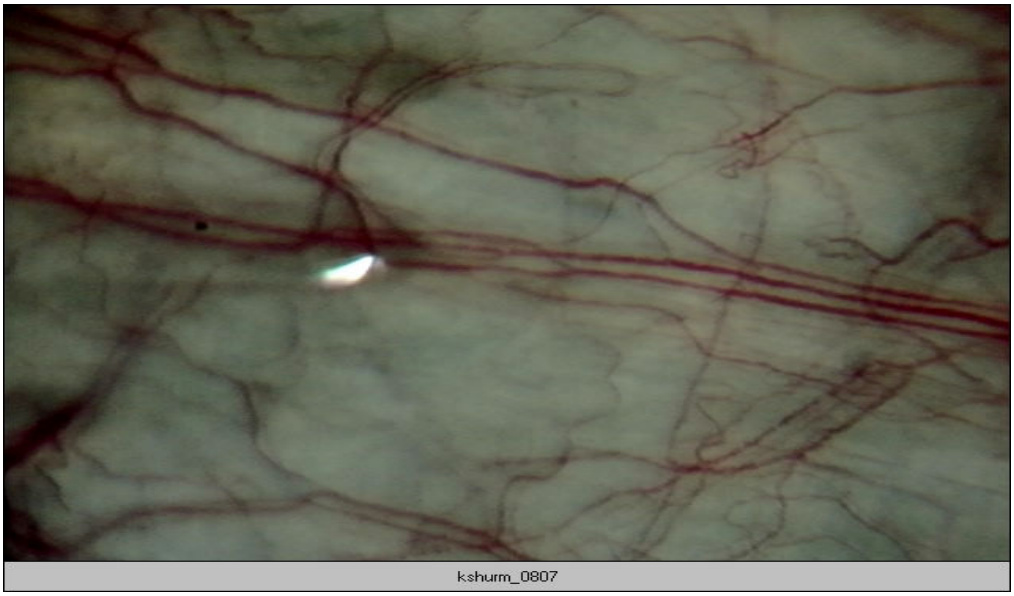

Рис. 5.3. Изображение бульбарной конъюнктивы пацинта в момент достижения максимальной ЧСС при выполнении ВЭП

Рис. 5.4. Изображение бульбарной конъюнктивы пациента через 3 минуты после прекращения нагрузки при проведении ВЭП

Отметим, что разработанный способ оценки состояния МР позволяет четко дифференцировать структуры, реагирующие на физическую нагрузку и количественно оценить степень выраженности изменений, что значительно повышает информативность и точность исследования.

5.1.2. Функциональные возможности опытного образца аппаратно-программного кардиологического комплекса

АПКК на базе суперкомпьютерной конфигурации «СКИФ» предназначен для поддержки проведения диагностических исследований методом конъюнктивальной биомикроскопии и получения количественной оценки риска развития ишемической болезни сердца и артериальной гипертензии.

Использование метода конъюнктивальной биомикроскопии в качестве метода исследования состояния сердечно-сосудистой системы требует оценки сложных многофакторных объектов, для получения надежной информации о которых, обеспечения высокой точности и диагностической информативности результатов исследований необходимо обеспечение сбора и цифровой обработки изображений микроциркуляторного русла в реальном масштабе времени.

Основные возможности АПКК (рис.5.5.):

– создание и ведение единого архива электронных медицинских карт и диагностических протоколов обследования пациентов;

– создание и ведение единого архива медицинских изображений микроциркуляторного русла (статических и динамических);

– различная обработка медицинских изображений, в том числе: изменение контрастности, изменение масштаба (увеличение и уменьшение), линейные измерения, межкадровая обработка, фильтрация и др.;

– количественная оценка изменений микроциркуляторного русла;

– документирование результатов исследований.

В состав АПКК входят:

– щелевая лампа (серии ЩЛ-3Г);

– суперкомпьютерная конфигурация «СКИФ» (кластер из 3-х двухпроцессорных вычислительных узлов и управляющая машина);

– инструментальные средства ввода изображений микроциркуляторного русла;

– специализированное программное обеспечение.

Аппаратная часть АПКК базируется на технических решениях, принятых при создании суперкомпьютеров семейства «СКИФ».

Технические характеристики:

– пиковая производительность – 9.4 GigaFlops;

– число вычислительных узлов – 3 и один управляющий узел;

– оперативная память – 3x512 + 1024Mb;

– дисковая память – 3x18Gb + 18Gb;

– тип системной и вспомогательной сетей – Ehternet 100Mbit;

– конструктив узла (форм-фактор) – 1U.

Рис. 5.5. Аппаратно-программный кардиологический комплекс

Каждый вычислительный узел содержит 2-х процессорную системную плату, два микропроцессора Pentium III 1,2 Ггц, оперативную память объемом 512 Мбайт, жесткий диск объемом 18 Гбайт, 2 встроенных сетевых адаптера Fast Ethernet.

Управляющий узел на Pentium IV-1800MHz содержит оперативную память объемом 1024 Мбайт, жесткий диск SCSI объемом 18 Гбайт, 2 сетевых адаптера Fast Ethernet, плату видеотюнера, видеокарту AGP, флоппи-диск, CD-ROM, монитор 17”, клавиатуру, манипулятор «мышь».

Управляющий узел обеспечивает управление вычислительными узлами, ввод изображений микроциркуляторного русла, хранение изображений в базе данных и интерфейс пользователя для работы с базой данных. Вычислительные узлы обеспечивают обработку изображения поступающего со щелевой лампы в режиме реального времени и обработку изображения для выделения и классификации сосудов.

5.1.3. Описание программного комплекса опытного образца АПКК

Программное обеспечение для обеспечения работы АПКК:

1) Операционная система Linux (RedHat 7.3 с ядром 2.4.18-3smp). В качестве среды выполнения параллельных приложений используется программная среда LAM (Local Area Multicomputer)

2) Библиотека программ MPI, реализующих прикладной интер-

фейс передачи сообщений высокого уровня по стандарту MPI 1.2 для языков программирования Фортран и Си.

3) Сервер баз данных InterBase 6.5.
4) Специализированное программный комплекс (ПРК).

Структурная схема ПРК показана на рис. 5.6. Стрелками обозначено движение данных между модулями. Взаимодействие сервера БД с клиентом происходит по протоколу TCP/IP.

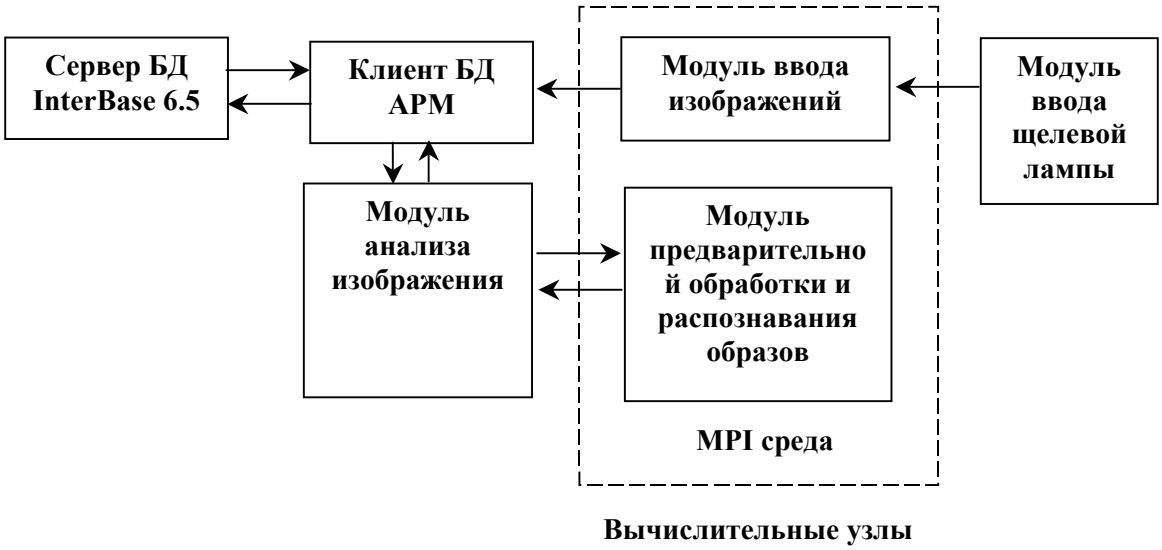

Рис. 5.6. Структурная схема ПРК.

Сервер баз данных (БД) InterBase 6.5 обеспечивает хранение медицинских карт пациентов.

Клиентская часть БД обеспечивает интерфейс взаимодействия ПРК с пользователем и вызов модулей обработки изображения.

Модуль ввода изображения обеспечивает ввод изображения с видеотюнера, подключенного к щелевой лампе. Обрабатывает изображение в режиме реального времени с целью улучшения их качества. Модуль работает в среде MPI на вычислительных узлах суперкомпьютера.

Модуль предварительной обработки и распознавания образов выполняет обработку изображения алгоритмами большой вычислительной мощности. Модуль работает в среде MPI на вычислительных узлах суперкомпьютера.

Модуль анализа изображения обеспечивает интерфейс с пользователем для устранения неточностей распознавания, классификации сосудов и вычисления количественной оценки состояния микроциркуляторного звена сердечно-сосудистой системы.

Модуль ввода изображения обеспечивает ввод изображения полученного при помощи щелевой лампы. Изображение поступает на аналоговый вход платы видеотюнера управляющего узла. На управляющем узле изображение делится на равные куски, которые рассылаются на вычислительные узлы суперкомпьютера. На вычислительных узлах оно обрабатывается одним из алгоритмов, затем отсылается обратно на управляющий узел и записывается в базу данных.

Основной причиной сложности поиска кровеносных сосудов на изображениях и измерения их характеристик является то, что реальные изображения характеризуются значительным динамическим диапазоном изменения яркости и палитры, зашумленностью и существенно зависят от условий наблюдений. Это требует больших усилий по приведению исходных изображений к определенному нормализованному виду необходимому для обеспечения надежной работы системы. Примеры работы алгоритмов предварительной обработки и распознавания изображений сосудов микроциркуляторного русла приведены на рис.5.7-5.18.

Модуль предварительной обработки и распознавания образов, получает изображение из базы данных, проводит над ним предварительную обработку специальным банком фильтров, распознает объекты и передает информацию в модуль анализа изображений.

Общая архитектура параллельной обработки модуля аналогична модулю ввода изображения, за исключением специализированных фильтров.

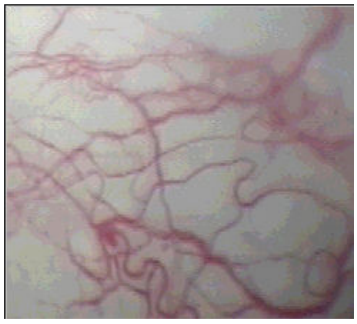

Рис. 5.7. Исходное изображение

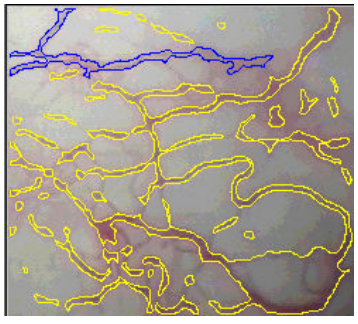

Рис. 5.8. Изображение после применения универсального алгоритма предварительной обработки и поиска сосудов по контурам

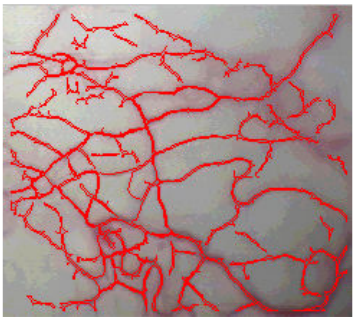

Рис. 5.9. Изображение после применения универсального алгоритма предварительной обработки и поиска сосудов методом скелетизации

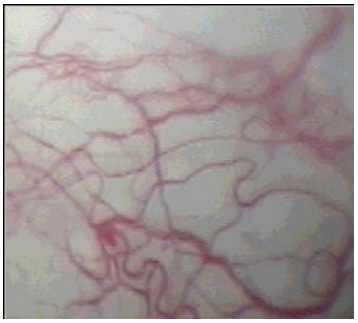

Рис. 5.10. Исходное изображение

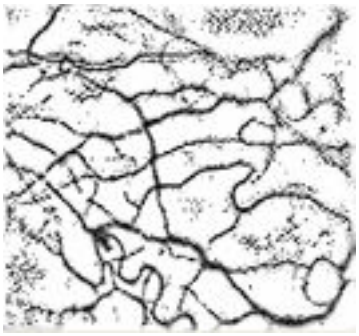

Рис. 5.11. Изображение после применения точного алгоритма предварительной обработки

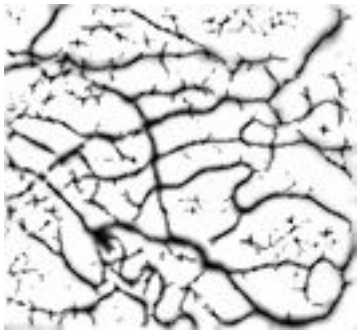

Рис. 5.12. Изображение после применения размытия с $\sigma = 0.5$ и точного алгоритма предварительной обработки

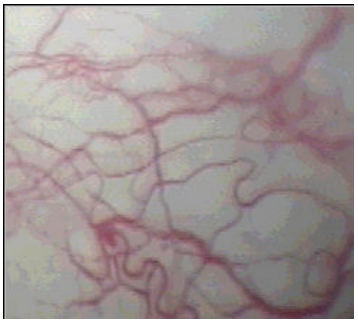

Рис. 5.13. Исходное изображение

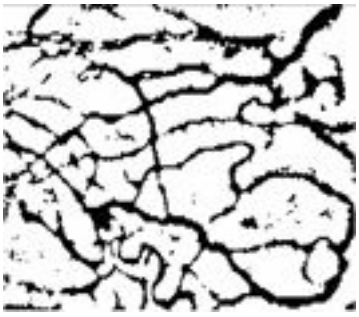

Рис. 5.14. Изображение после применения универсального алгоритма предварительной обработки и бинаризации

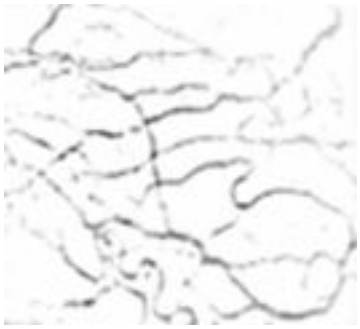

Рис. 5.15. Изображение после применения размытия с $\sigma = 2.5$ и точного алгоритма предварительной обработки

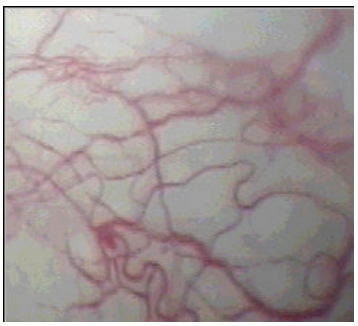

Рис. 5.16. Исходное изображение

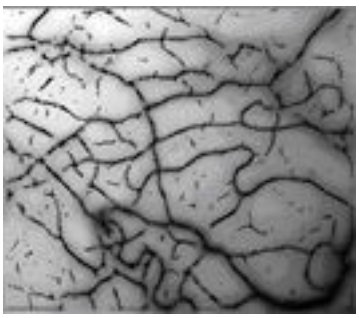

Рис. 5.17. Изображение после применения размытия с $\sigma = 1.0$ и алгоритма скелетизации

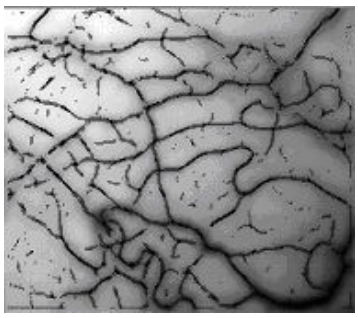

Рис. 5.18. Изображение после применения размытия с $\sigma = 0.5$ и алгоритма скелетизации

Сравнительные характеристики последовательных и параллельных алгоритмов, выполняемых на 1 и 7 процессорах соответственно представлены в таблице 5.1 и на рис.5.19.

Таблица 5.1

Время выполнения алгоритмов обработки изображения

Тип фильтра	1 процессор	7 процессоров
1. Сглаживающий	99500	18633
2. Медианный	114624	21465
3. ВЧ	109567	20518
4. Дифференцирующий	105730	19800
5. Комбинированный	199265	37316

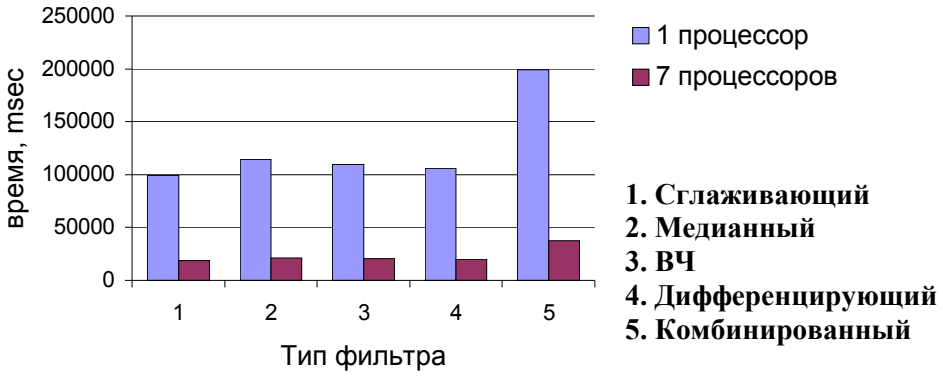

Рис. 5.19. Время выполнения алгоритмов фильтрации на 1 и 7 процессорах

Как видно из таблицы 5.1, на 7 процессорах комбинированный алгоритм выполняется в 5.36 раз быстрее, это значит, что используется 76.29% пиковой производительности кластера.

5.1.4. Разработка телекоммуникационной кардиодиагностической сети

В настоящее время для медицинской практики все большее значение приобретает перспектива развития такого направления цифровой технологии, как телемедицина.

Использование Интернет расширяет возможность свободного обмена разнообразной информацией между клиниками различных географических регионов путем передачи диагностического изображения, текстовых и графических файлов с унифицированной тактикой диагностики

и лечения с целью научных исследований, телеконференций и медицинского образования.

Современная телемедицина обеспечивает создание, передачу, хранение и отображение информационного продукта (данных, знаний) с целью проведения лечебно-диагностических мероприятий на расстоянии в заданный интервал времени. Одним из наиболее распространенных направлений телемедицины является организация удаленных консультаций врачей-специалистов. Традиционно приоритетным разделом этой области считается кардиология.

Разработка и внедрение АПКК позволяет телемедицинским технологиям в области кардиологии охватить обширную область – от телемониторинга состояния больных и экспресс-диагностики на расстоянии до имитационного моделирования и обучения.

Внедрение АПКК в лечебно-профилактические учреждения позволит создать технологию мониторинга и формирование алгоритма первичного обследования для выявления лиц с повышенным риском развития ишемической болезни сердца и артериальной гипертензии. Общая схема телемедицинской системы представлено на рис. 5.20.

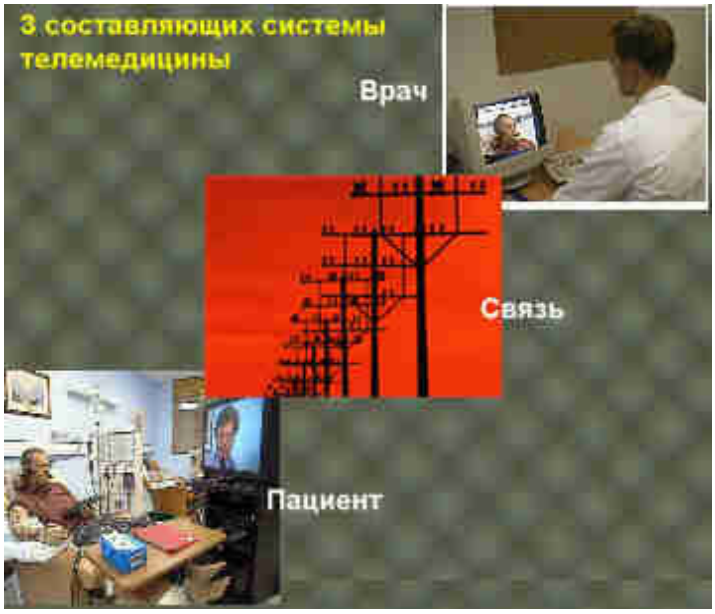

Рис. 5.20. Общая схема телемедицинской системы представлено

Структурная схема 1-й очереди ТККДС приведена на рис. 5.21.

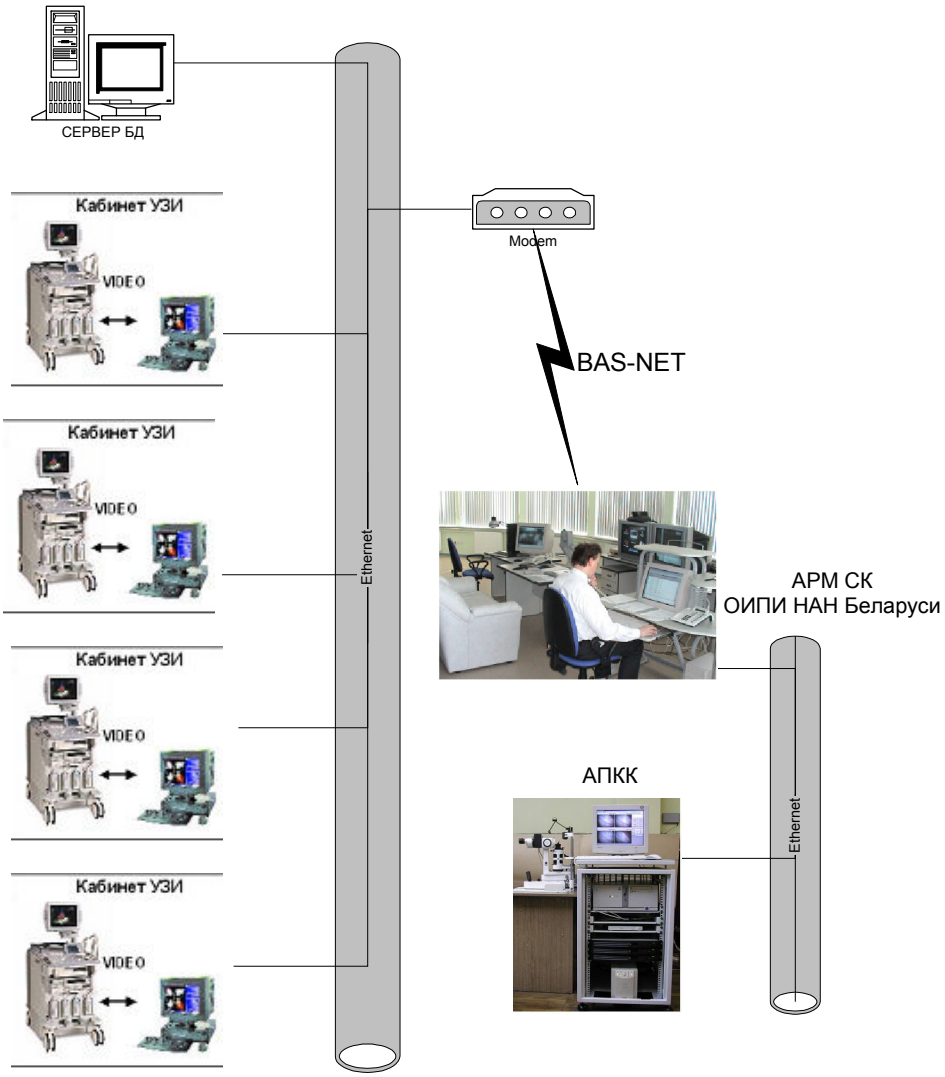

Рис. 5.21. Структурная схема 1-й очереди ТККДС

Телекоммуникационная кардиодиагностическая сеть включает автоматизированное рабочее место специалиста-консультанта (АРМ СК) телемедицинского центра на базе ОИПИ НАН Беларуси и кардиодиагностическую сеть на базе РНПЦ «Кардиология».

Кардиодиагностическая сеть на базе РНПЦ «Кардиология» включает консультативно-экспертную систему на базе опытного образца АПКК и диагностическую информационную систему в составе выделенного сервера базы данных и 5 автоматизированных рабочих мест (АРМ) врачей-диагностов.

Опытный образец АПКК может также размещаться на базе ОИПИ НАН Беларуси и быть объединенным в рамках локальной сети с АРМ

СК.

Программное обеспечение ТККДС обеспечивает проведение телемедицинских консультаций в отложенном режиме (по электронной почте), а также – в реальном режиме времени (с использованием видеоконференцсвязи).

Все консультативные сеансы, проводимые в ТККДС, протоколируются с помощью АРМ СК. Результаты протоколирования подвергаются обработке, компрессированию и последующей каталогизации в рамках единой базы данных ТККДС. Каждый пользователь ТККДС имеет свои уникальные код и пароль. Удаленный доступ к опытному образцу АПКК, установленному в ОИПИ НАН Беларуси, осуществляется посредством криптованного ssh-соединения.

5.2. Программно-аппаратный комплекс для численного моделирования процессов в задачах радиационной газодинамики

Работы по разработке программно-аппаратного комплекса для численного моделирования процессов в задачах радиационной газодинамики на суперкомпьютере «СКИФ К-500» выполнены в ГНУ «ИТМО им. А.В.Лыкова» НАН Беларуси в рамках программы «СКИФ». Руководитель темы – Романов Г.С., основные исполнители работы – Станкевич Ю.А., Сметанников А.С., Степанов К.Л., Окунев В.Е., Зенькевич С.М.

Введение в последние годы в практику научных расчетов векторных, матричных и многопроцессорных вычислительных систем способствовало появлению совершенно новых возможностей при решении многих научных и технических задач методами численного моделирования и вычислительного эксперимента. Одним из таких направлений является высокотемпературная гидродинамика, описывающая поведение вещества и протекающие в нем физические процессы при интенсивных воздействиях, сопровождающихся большим выделением энергии. Объектом высокотемпературной газодинамики являются горячие газы, плазма, а также вещество в твердом и жидком состоянии, в которых под действием мощных потоков энергии различной физической природы протекают разнообразные процессы. В силу своей сложности, нелинейности, многообразия и взаимосвязи они могут быть эффективно проанализированы и исследованы только путем построения численных моделей, описывающих поведение физических систем, и выполнения с их помощью вычислительного эксперимента.

В процессе выполнения работы были выполнены теоретические исследования и численное моделирование на суперкомпьютере «СКИФ К-500» физических и радиационно-газодинамических процессов в нескольких задачах динамики излучающего газа:

1) Взаимодействия лазерного излучения с мишенью, динамика и излучение эрозионного факела.

2) Динамика высокоскоростного астероидного удара по поверхности Земли.

3) Сильный высокотемпературный взрыв.

4) Динамика входа космического тела в атмосферу Земли.

Первый из этих комплексов описывает процессы нагрева и испарения материала мишени под действием лазерного излучения, образование эрозионного плазменного факела, гидродинамику его разлета в окружающую среду, поглощение плазмой лазерного излучения и возникновение экранировки поверхности, собственное тепловое излучение плазмы эрозионного факела. Комплекс позволяет моделировать пространственно-временную картину действия лазерного излучения на непрозрачную мишень при различных режимах воздействия (длительность и форма лазерного импульса, мощность излучения и т.д.) и в зависимости от внешних условий (например, давление окружающего газа).

В таблице 5.2 представлены результаты тестирования эффективности многопроцессорных вычислений задачи на суперкомпьютере «СКИФ-500». Результаты определения эффективности приведены для расчета поля излучения в одной пространственной точке.

Таблица 5.2.

Результаты тестирования эффективности расчета

Число узлов кластера, N	1	2	4	8	16
Процессорное время t_N, сек	17.67	8.91	4.61	2.45	1.44
Эффективность расчета $100t_1/(Nt_N)$, %		99.16	95.82	90.15	76.69

t_1 – время решения задачи на одном процессоре; t_N – время расчета на N процессорах.

Большая эффективность распараллеливания данной задачи объясняется достаточно высокой однородностью загрузки процессоров и сравнительно малым объемом информации, транспонируемой между процессорами. Поэтому затраты времени на обмен данными и ожидание завершения параллельных вычислений составляют незначительную часть полного расчетного времени. Расчет на пространственной сетке 30x100 на машине «Pentium-150» требует 24 часов, на одном узле «СКИФ-500» расчетное время составляет 8 часов, на 16 узлах кластера «СКИФ-500» расчетное время – около получаса.

Второй программный комплекс осуществляет численное модели-

рование динамики процессов, происходящих при высокоскоростном ударе. Эта задача представляет интерес для многих проблем астрофизики и космической физики – создания систем противометеоритной защиты космических аппаратов, изучения метеоритных кратеров, происхождения планетных атмосфер, возможных последствий падения крупных космических объектов на Землю и др.

Сравнение эффективности решения этой задачи на многопроцессорной системе в зависимости от числа процессоров представлено в таблице 5.3.

Таблица 5.3

Тестирование расчета на многих процессорах

Код\ N_PROC	1	2	3	5	10
IMPACT efficienty %	100	61	57	39	23

Как показывает сравнение, мультипроцессорный расчет заметно сокращает время выполнения задачи. Вместе с тем, нельзя сказать, что отношение времени счета на одном процессоре к времени счета на большем числе процессоров есть увеличение производительности. Вызвано это следующими обстоятельствами. Во-первых, распараллелена лишь часть расчетного цикла, которая отвечает решению уравнений радиационной диффузии. Во-вторых, на начальном этапе проникновения астероида в грунт и образования кратера (пока еще ударная волна не прошла по телу астероида и не испарила его) расчет переноса излучения не ведется, поскольку это лишь удлинит время расчета, а выноса энергии излучением из прикратерной области практически не будет. Т.е. на этом этапе расчета никакого расщепления расчета нет Кроме того, если число групп не кратно числу процессоров (например, когда расчет ведется на трех или четырех процессорах), распараллеливание оказывается не максимально эффективным.

С помощью третьего пакета программ исследуется одна из классических задач динамики излучающего газа – сильный взрыв с учетом излучения.

В таблице 5.4 показана эффективность распараллеливания вычислительного алгоритма на стадии расчета процессов радиационного переноса. Производительность параллельных вычислений рассчитывалась путем сравнения времени решения задачи на различном числе процессоров кластера «СКИФ-500».

Таблица 5.4

Протокол тестирования задачи на различном числе процессоров

Код \ N_PROC	1	3	5	7	9	11	13	15
EXPLOSION efficienty %	100	97.4	95	89.1	91.5	82.0	86.3	85.0

Следует отметить, что эффективность распараллеливания вычислений в данном элементе вычислительного алгоритма целиком зависит от соотношения времени расчета газодинамической части программы и радиационной ее части.

Четвертый комплекс программ позволяет выполнять численное моделирование явлений, сопровождающих движение космического тела в атмосфере.

Численное моделирование позволило получить картину нестационарного гиперзвукового течения вокруг метеороида. Получены подробные двумерные поля газодинамических параметров газа при обтекании метеороида. Проведенные исследования свидетельствуют о соответствии выбранной физической и математической модели реальным процессам, протекающим при обтекании космических тел гиперзвуковыми потоками газа при их вхождении в атмосферу. Данные об эффективности распараллеливания вычислительного алгоритма в данной задаче в зависимости от числа процессоров приведены в таблице 5.5.

Таблица 5.5

Тестирование расчета на многих процессорах

Код \ N_PROC	1	2	5	10
STREAMLINE efficienty %	100	89.6	61.2	54.2

Отметим, что эффективность распараллеливания существенным образом зависит от соотношения времен, затрачиваемых газодинамическим и радиационным блоками программного кода.

Несмотря на кажущуюся разнородность этих задач, они имеют целый ряд общих характерных черт:

– интенсивное энерговыделение (в первой задаче – поглощение падающего на мишень лазерного излучения, во второй и четвертой – переход кинетической энергии движения в тепловую энергию, в третьей – энерговыделение в результате химической, ядерной или иной реакции);

– гидродинамическое течение, теплообмен, фазовые переходы и радиационный перенос энергии. Это также широкий диапазон параметров состояния, в котором находится вещество на разных стадиях процессов и различных пространственных областях.

Численное решение уравнений радиационной газодинамики, лежа-

щих в основе адекватного описания рассматриваемых процессов, весьма удачно соответствует архитектуре многопроцессорных систем. В частности, достаточно проста векторизация расчетов гидродинамики по явным разностным схемам, что объясняется их локальностью. Также естественно распараллеливается расчет радиационного переноса энергии при использовании многогруппового по спектру приближения. Таким образом, очевидно, что высокопроизводительные мультипроцессорные вычислительные системы с распараллеливанием вычислений, большим объемом оперативной памяти и высоким быстродействием в полной степени соответствуют требованиям для эффективного решения сформулированных задач. Для осуществления такого распараллеливания были использован программный инструментарий MPI (message passing interface), который позволяет расщепить алгоритм расчета и осуществить связь между различными ветвями параллельного алгоритма.

Отметим некоторые особенности программно-аппаратных комплексов для их решения. Прежде всего, все они нестационарные. За исключением третьей задачи, которая сформулирована в одномерной (плоской, сферической или цилиндрической) постановке, все они являются двухмерными и осесимметричными. Это приближение справедливо, если лазерное излучение и скорость астероида направлены нормально к поверхности, на которую они воздействуют, и если космическое тело имеет ось симметрии, совпадающую с направлением его движения. Во всех задачах используются реальные оптико-физические характеристики вещества. К ним относятся уравнения состояния, замыкающие систему уравнений газовой динамики, и коэффициенты поглощения среды, используемые для расчета радиационного переноса энергии. Все описанные программные комплексы работают с базой данных по физическим характеристикам вещества, созданной в коллективе авторов.

В процессе выполнения работы показано, что сложные нелинейные, нестационарные и многомерные задачи динамики излучающего газа, имеющие важное научное и научно-техническое значение, и требующие для своего анализа использования мощных компьютеров с большим объемом оперативной памяти и высоким быстродействием, могут быть успешно решены на суперкомпьютерах семейства «СКИФ».

5.3. Эксплуатация инженерных пакетов на суперкомпьютерах семейства «СКИФ»

Данный раздел содержит описание технологий доступа и эксплуатации инженерных пакетов на суперкомпьютерных конфигурациях «СКИФ» на примере выполнения работ по проектированию моделей турбокомпрессоров.

Работа выполнялась коллективом ОИПИ НАН Беларуси под руко-

водством Чижа О.П. в составе Мурашко В.В., Кулешевой М.Е.

Объектом исследования являлись особенности технологии информационного взаимодействия процессов и этапов компьютерной технологии проектирования турбокомпрессора, удаленного доступа к кластерным конфигурациям «СКИФ» для моделирования процессов газодинамики, механики и теплообмена турбокомпрессора на пакете конечно-элементного анализа (КЭА) LS-DYNA и визуализации результатов проектирования и моделирования турбокомпрессора.

Цель работы – исследование и анализ методов обеспечения информационного взаимодействия процессов и этапов компьютерной технологии проектирования турбокомпрессора, методов эксплуатации пакета КЭА LS-DYNA на кластерных конфигурациях «СКИФ» и визуализации результатов проектирования и моделирования при удаленном доступе на коммуникационных каналах с высокой и с низкой пропускной способностью, методов создания файлов сеток и входных файлов для пакета КЭА LS-DYNA.

Для решения поставленной задачи были разработаны методы обеспечения взаимодействия процессов и этапов компьютерной технологии проектирования турбокомпрессора, методы удаленной работы с пакетом КЭА LS-DYNA на суперкомпьютерной системе, разработаны методы получения в командном режиме видео-файлов (avi и mpeg) из программы пре-постпроцессора LS-PREPOST, разработаны командные файлы, позволяющие автоматизировать процессы запуска задачи пакета LS-DYNA на кластере и получения результатов при удаленном доступе на каналах с низкой пропускной способностью. Разработаны командные файлы получения файлов сеток и входных файлов для пакета КЭА LS-DYNA.

В результате проведенных работ были разработаны руководство пользователя по эксплуатации пакета КЭА LS-DYNA при удаленном доступе на коммуникационных каналах с высокой и с низкой пропускной способностью и методика создания файлов сеток и входных файлов для пакета КЭА LS-DYNA с помощью пакета ICEM CFD.

Основными этапами компьютерной технологии проектирования турбокомпрессора (ТКР) являются:

– построение электронных 3-D моделей деталей и сборок ТКР на основе технической документации предоставляемой РУП «БЗА»;

– компьютерный анализ и моделирование газодинамического состояния конструкций ТКР;

– компьютерный анализ и моделирование динамических прочностных характеристик конструкций ТКР;

– корректировка электронных 3-D моделей по результатам анализа газодинамических и прочностных расчетов;

– разработка по электронным 3-D моделям деталей и сборок ТКР электронных моделей ассоциативных чертежей.

Проектирование турбокомпрессора проводилось специалистами предприятия РУП «БЗА», ОИПИ НАН Беларуси, НИЧ БНТУ как на ПЭВМ участников проекта, так и на высокопроизводительных кластерных конфигурациях семейства «СКИФ». На рис. 5.22 представлена блок-схема взаимодействия участников проекта.

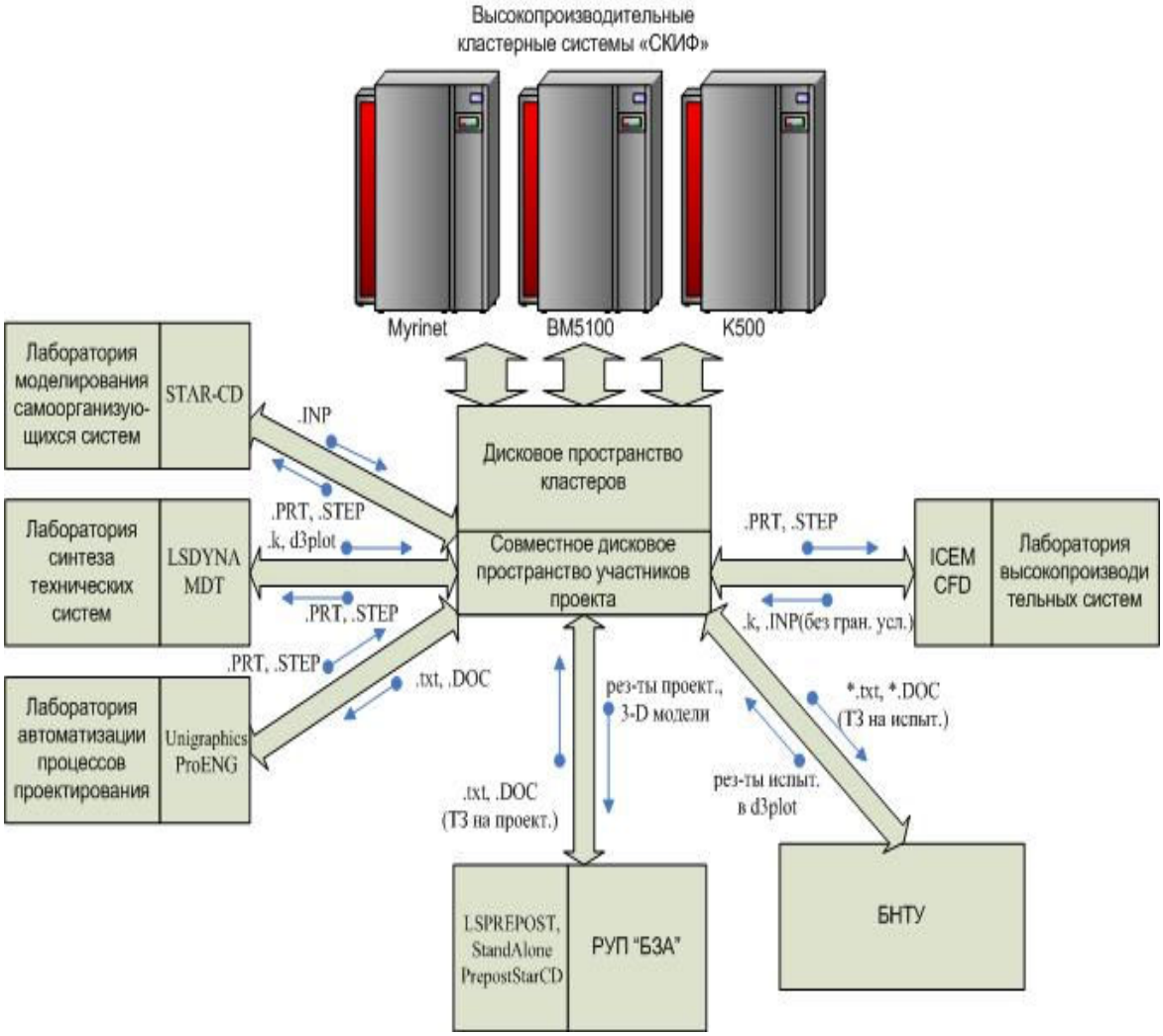

Рис. 5.22. Блок-схема взаимодействия участников проекта по разработке турбокомпрессора

Для функционирования компьютерной технологии проектирования деталей и сборок ТКР необходимо обеспечить доступ к кластерным конфигурациям «СКИФ», а также оперативный обмен данными между участниками проектирования.

Создание общего дискового пространства. Для обмена электронными данными на дисковом пространстве кластера BM5100 была создана структура каталогов (рис. 5.23), обеспечивающая оперативный обмен и хранение данных (файлов) между участниками проектирования. Были настроены права доступа к общим каталогам и подкаталогам.

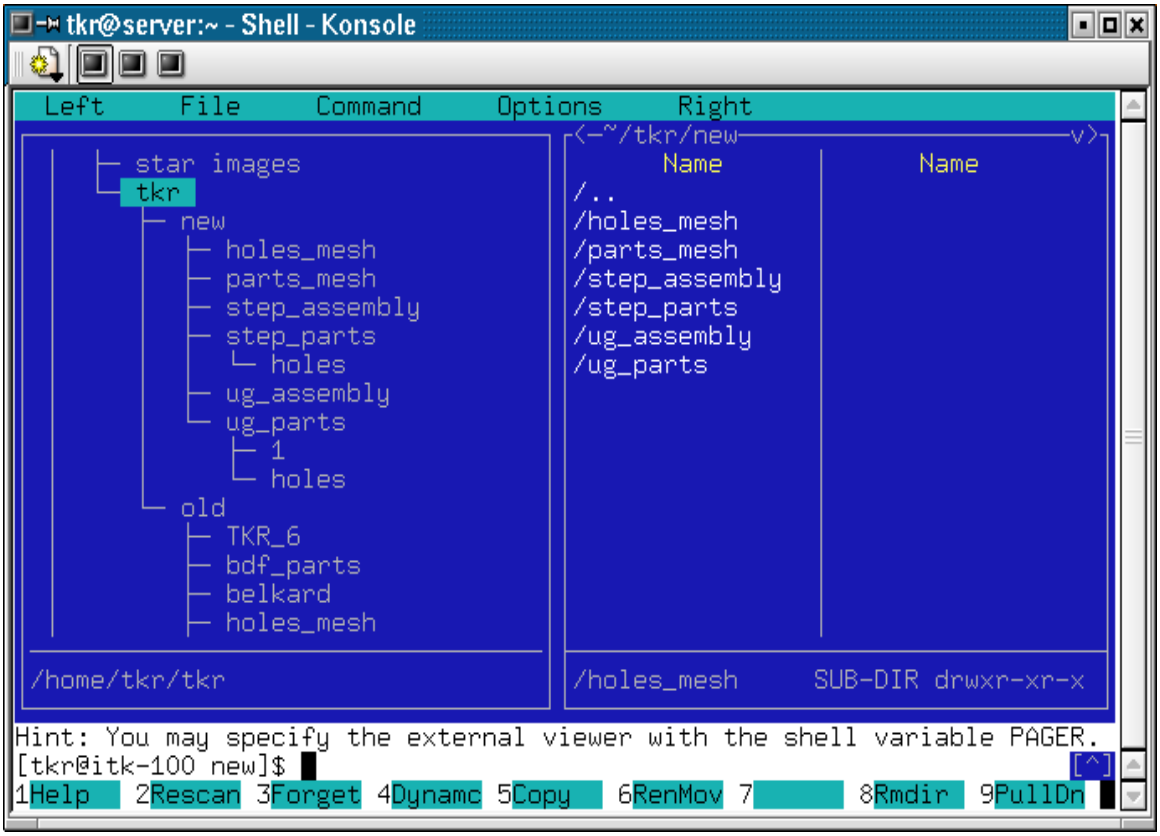

Рис. 5.23. Структура каталогов

Установка прикладного программного обеспечения для удаленного доступа к кластерам «СКИФ». Для обеспечения удаленного доступа к кластерам «СКИФ», возможности обмена данными между лабораториями ОИПИ НАН Беларуси, РУП «БЗА» и для визуализации результатов проектирования на персональных компьютерах участников проекта было установлено и настроено следующее прикладное программное обеспечение:

– утилита для защищенного терминального доступа на UNIX-хосты **putty** и клиент безопасного копирования **pscp**;

– менеджер файлов Total Commander с **sftp** плагином для удаленного копирования файлов;

– NFS-клиент для ОС Windows, для обеспечения возможности работы по протоколу NFS с пользовательскими директориями, расположенными на управляющих машинах кластеров (только в лабораториях ОИПИ НАН Беларуси);

– LSPREPOST, программа пре-постпроцессинга пакета LS-DYNA.

Пользователь работал с пакетом КЭА LS-DYNA, установленном на кластере BM5100, при удаленном доступе на коммуникационных каналах с высокой (Fast Ethernet/Gigabit Ethernet, более 100 Мбит/с) и низкой (до128 Кбит/с) пропускной способностью. Осуществлялась визуализацию полученных результатов.

В качестве примера использована модель ротора ТКР.

На рис. 5.24 показано диалоговое окно открытия ssh-сессии с кластером BM5100.

На рис. 5.25 представлено диалоговое окно программы LSPREPOST после загрузки результатов вычислений.

На рис. 5.26 показан просмотр динамических результатов вычислений для ротора решателя пакета LS-DYNA.

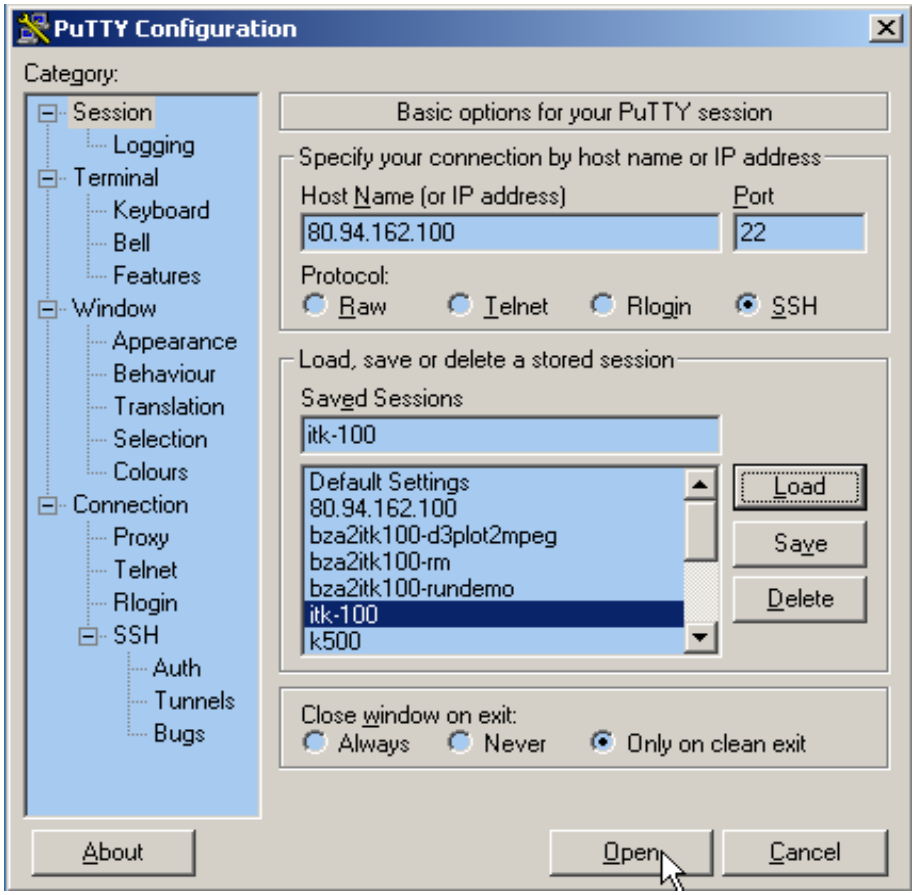

Рис. 5.24. Диалоговое окно открытия ssh-сессии с кластером BM5100

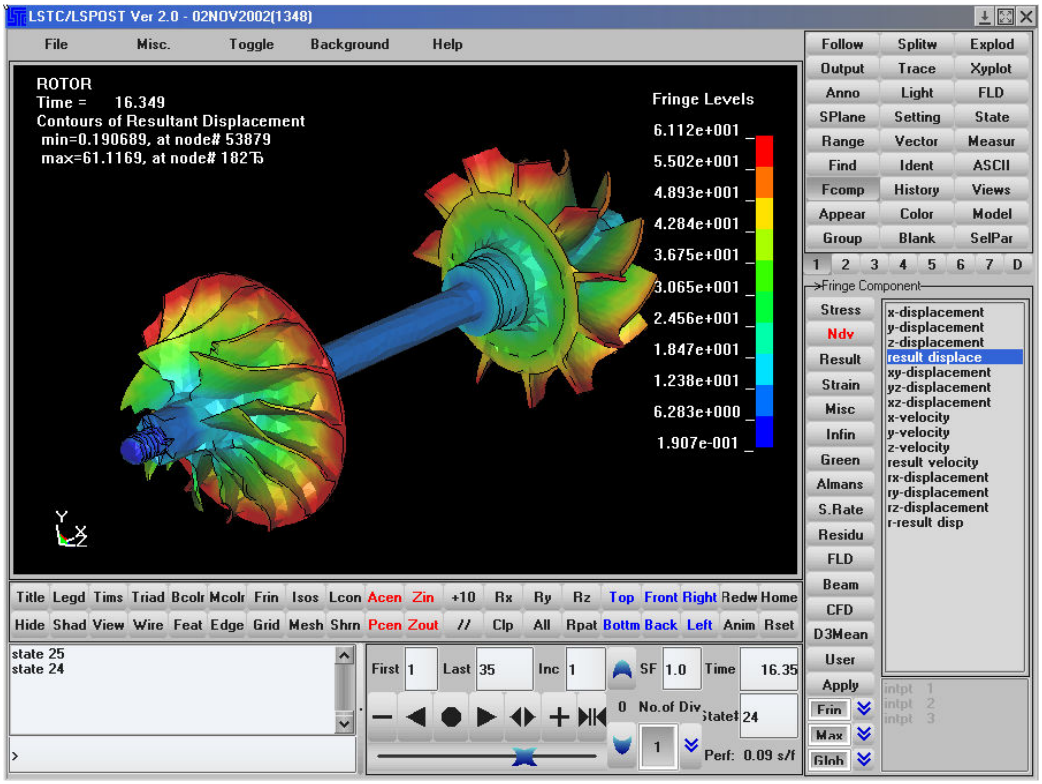

Рис. 5.25. Диалоговое окно программы LSPREPOST после загрузки результатов вычислений

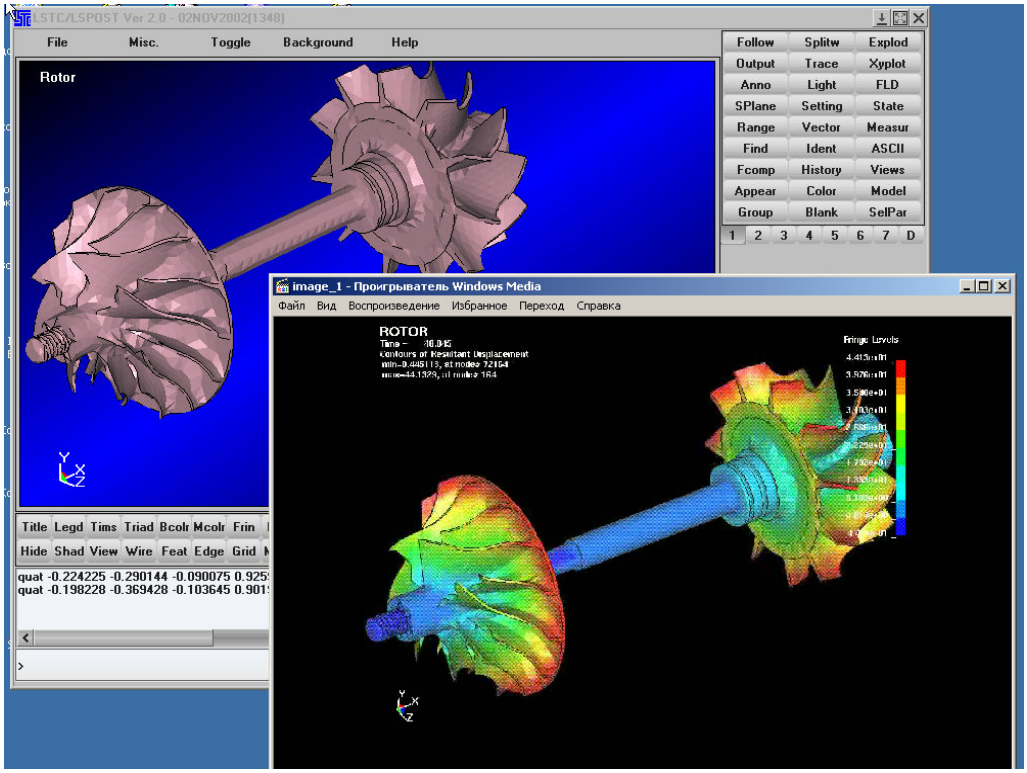

Рис. 5.26. Динамические результаты вычислений решателя пакета LS-DYNA

5.4. Конечно-элементный анализ машиностоительных конструкций на суперкомпьютерах семейства «СКИФ»

Работы в этом направлении выполнялись в ОИПИ НАН Беларуси Медведевым С.В., Чижом О.П., Петрушиной М.В.

По независимым экспертным оценкам до 75% всех дефектов машиностроительных конструкций закладываются на этапе конструирования и, в меньшей степени, технологической подготовки.

При инженерных расчетах конструкций на прочность и жесткость в последнее время широко используются методы и соответствующие пакеты конечно-элементного анализа (Finite Element Analysis – FEA). Однако в реальных производственных условиях расчеты сложных в структурном отношении моделей конструкций требуют десятки, сотни и более часов процессорного времени как на персональных вычислительных машинах, так и на графических станциях. Некоторые расчетные задачи (например, взаимодействие газовых потоков с объектами сложных геометрических форм и т.д.) принципиально не могут быть решены на традиционных средствах вычислительной техники за приемлемое процессорное время; другие задачи неразрешимы в вычислительном отношении из-за слишком большого объема исходных данных.

Производительность, объемы памяти, удаленный доступ к суперкомпьютерам семейства «СКИФ» обеспечивают возможности построения систем проектирования конкурентоспособных конструкций, в которых процессы поиска конструктивно-технологических вариантов совмещены по времени с процессами расчетов динамической прочности, ресурса конструкций с учетом особенностей технологических процессов их изготовления.

Заслуживает внимания в связи с этим создание региональных суперкомпьютерных центров коллективного пользования, располагающих как высокопроизводительными кластерными системами, так и программным обеспечением мирового уровня.

На данный момент сложилось несколько направлений достаточно эффективного использования суперкомпьютеров «СКИФ» для конечно-элементного анализа:

1) Конструктивно-технологическое проектирование и анализ остаточных деформаций и напряжений сварных конструкций общемашиностроительного применения.

2) Моделирование столкновений транспортных средств с неподвижными препятствиями, оценка пассивной безопасности водителя и пассажиров.

3) Компьютерное моделирование взаимодействия лазерных излу-

чений с веществами различной природы.

4) Моделирование динамических прочностных явлений, возникающих при проектировании и эксплуатации карданных валов транспортных средств.

5) Моделирование в связанной постановке явлений взаимодействия газовой среды с твердыми телами.

Рассмотрим обозначенные направления более подробно.

Одним из направлений является моделирование сварочных явлений в среде LS-DYNA. Для этого используется термомеханическая постановка задачи, реализуемая, к сожалению, только в однопроцессорной (SMP) версии пакета. Параллельно разрабатывается подход, который позволяет использовать механическую постановку задачи и многопроцессорную (MPP) версию LS-DYNA.

По этой методике расчету подвергаются такие объекты, как лонжероны и рамы транспортных средств высокой грузоподъемности, элементы шахтных крепей и другие объекты (рис. 5.27-5.29).

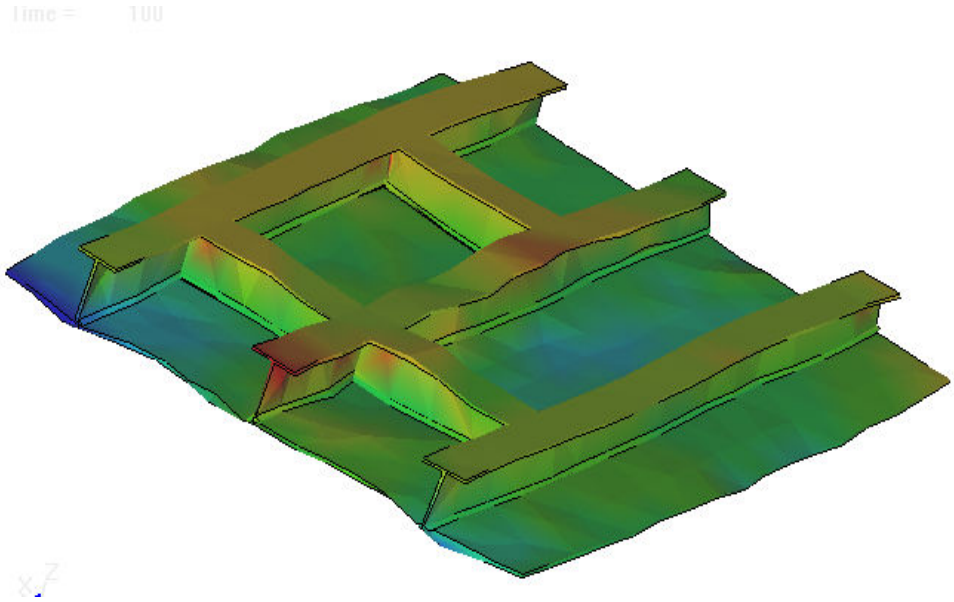

Рис. 5.27. Пример распределение остаточных деформаций плоской сварной конструкции как результат решения термомеханической задачи

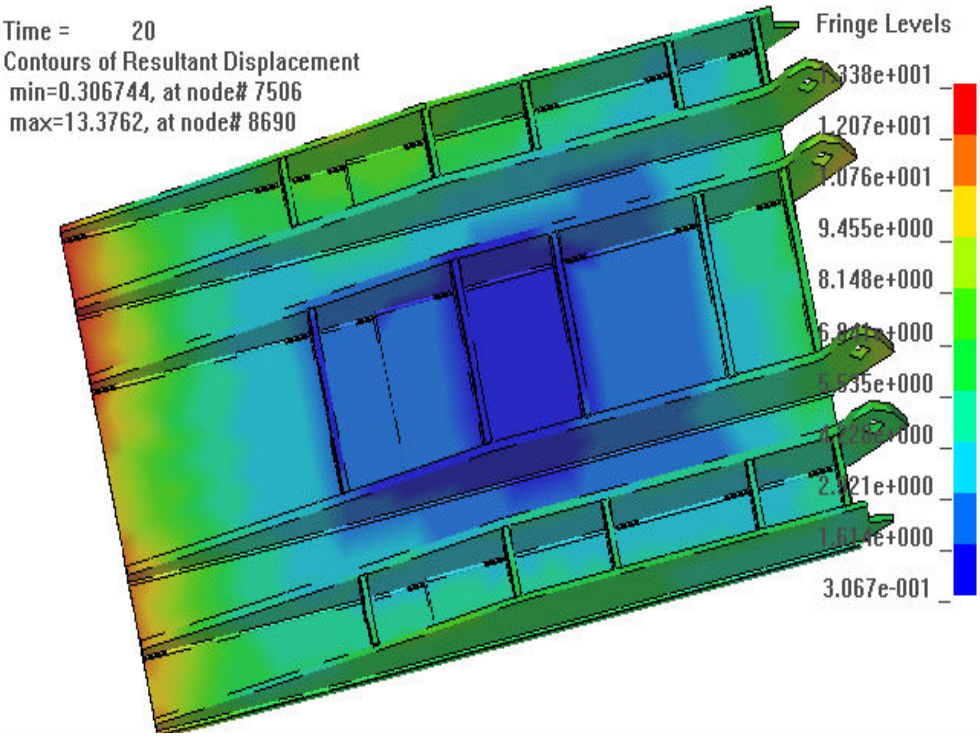

Рис. 5.28. Характер распределения остаточных деформаций при сварке деталей шахтной крепи

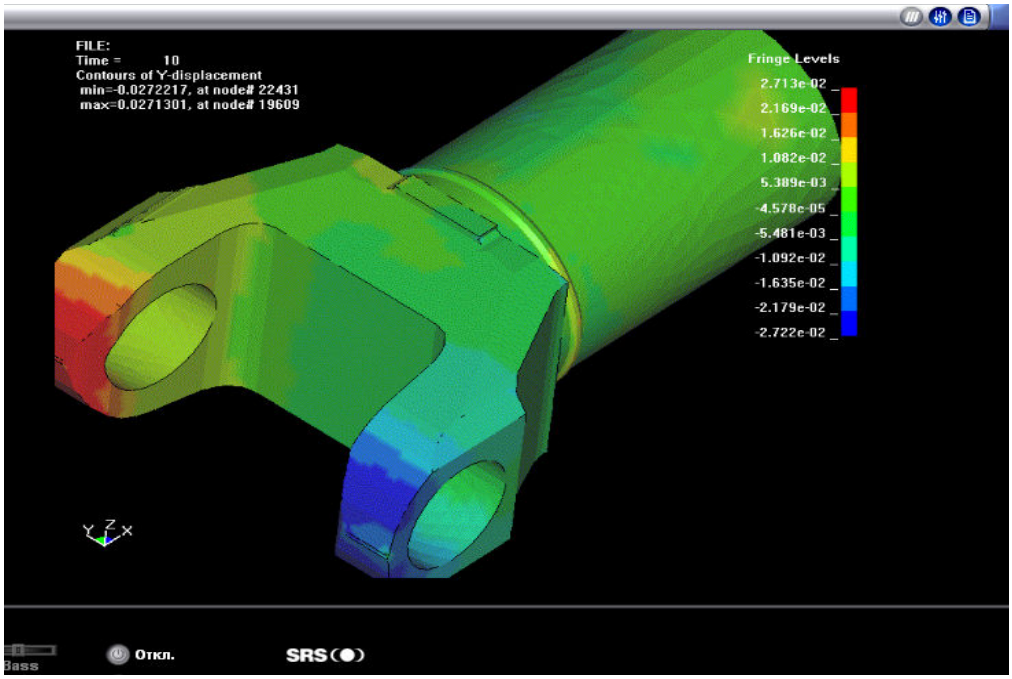

Рис. 5.29. Моделирование остаточных перемещений карданной вилки, вызываемых сваркой кольцевого шва

На рис.5.30 и 5.31 представлены достаточно сложные в структурном отношении объекты, исходные данные о которых представлены моделями не единых твердых тел, а совокупностью моделей отдельных деталей, между которыми автоматически определяется наличие контактов. Модель почвообрабатывающего агрегата в сборе занимает более 450 Мб дискового пространства, имеет в своем составе около 3 млн. пространственных конечных элементов.

Проведены исследовательские работы по частичной замене натурных испытаний транспортных средств виртуальными путем решения задачи моделирования деформаций и оценки жизненного пространства кабины автомобиля при ударных воздействиях в среде LS-DYNA на суперкомпьютерах семейства «СКИФ» (рис. 5.32, 5.33).

Исследованы температурные условия и режим работы подшипникового узла, обуславливающие потерю жёсткости в подшипниках скольжения турбокомпрессора ТКР 6.1 Борисовского завода агрегатов. Изучены модели зоны нагрева ротора в области контакта с подшипниками скольжения с целью получения картин температурных полей и полей деформаций в зоне контакта.

Проведен статический модальный анализ пар «ротор-подшипник» и турбокомпрессора в сборе для получения спектра частот, обеспечивающих заданные эксплуатационные характеристики турбокомпрессора (рис. 5.34).

Исследована газодинамическая задача продува турбокомпрессора и взаимодействия газовой среды с лопаточными колесами турбины и компрессора. Использовалась Arbitrary Lagrangian Eulerial (ALE) постановка (рис. 5.35, 5.36), изначально требующая очень серьезных вычислительных ресурсов.

В программной системе LS-DYNA, развернутой на суперкомпьютере «СКИФ К-500», выполнен ряд расчётов для Института молекулярной и атомной физики НАН Беларуси, связанных с взаимодействием лазерного излучения с веществом. В рамках задачи селективного лазерного спекания титанового порошка выполнялся расчёт температурного поля, создаваемого импульсным лазерным излучением (рис. 5.37).

Проведен анализ динамических прочностных характеристик карданных валов автомобилей с целью оптимизации конструкции, улучшения прочностных свойств и эксплуатационных качеств (рис. 5.38, 5.39).

На суперкомпьютерной конфигурации «СКИФ К-500» с помощью пакета LS-DYNA был проведен расчет задачи столкновения трех автомобилей (3 Vehicle Collision**).** Конечно-элементная модель трех автомобилей включает конечные элементы разных типов и содержит порядка 1,5 млн. степеней свободы (рис. 5.40).

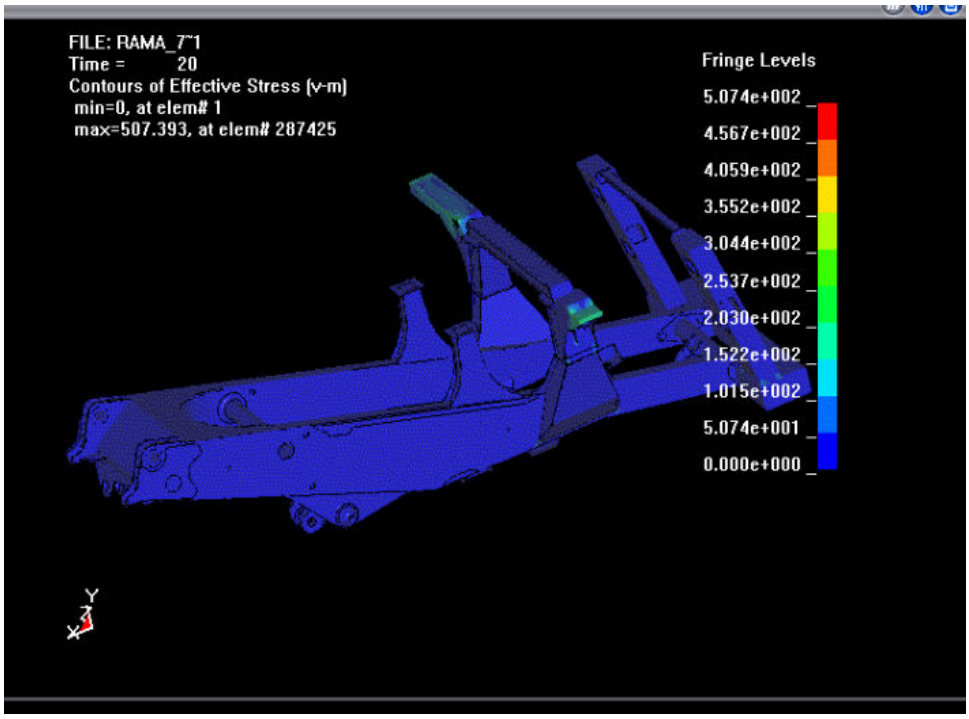

Рис. 5.30. Напряженное состояние рамы транспортного средства с учетом нерелаксированных остаточных сварочных напряжений

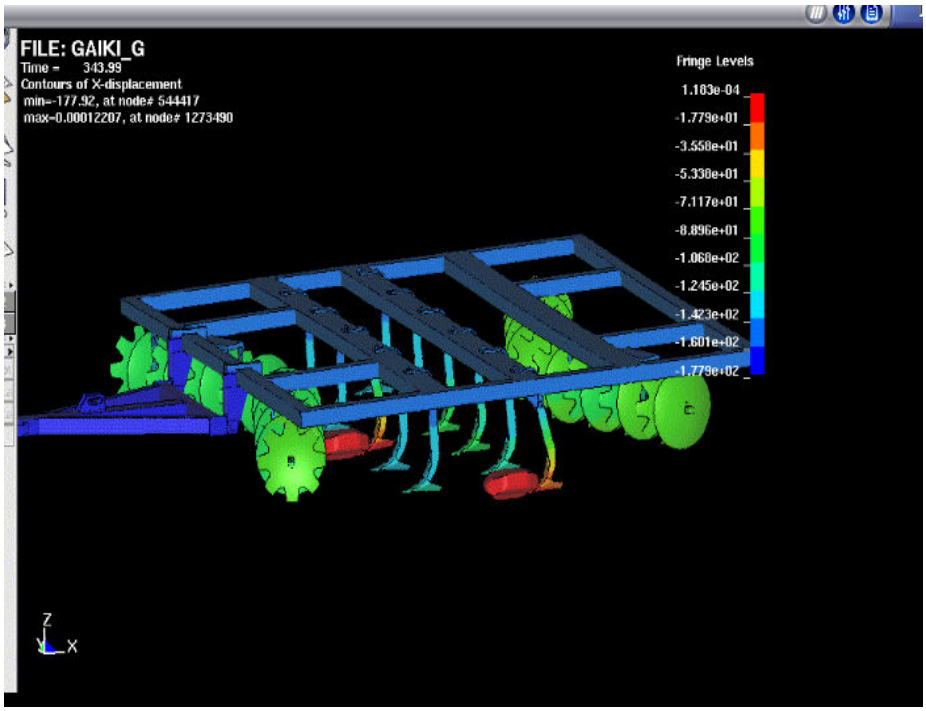

Рис. 5.31. Компьютерная модель почвообрабатывающего агрегата, взаимодействующая с неподвижными препятствиями

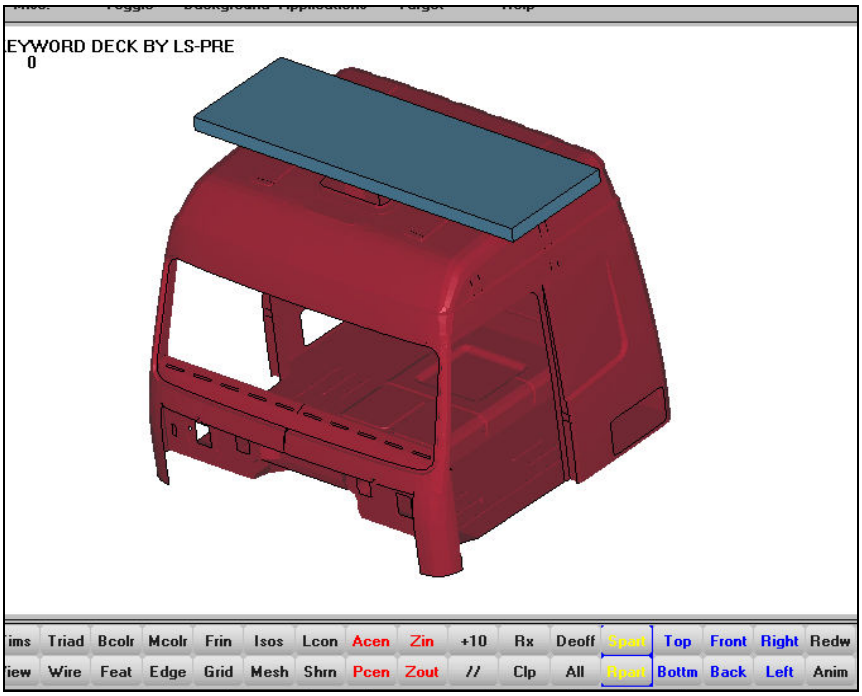

Рис. 5.32. Исходное состояние модели при моделировании процессов смятия кабины неподвижными препятствиями

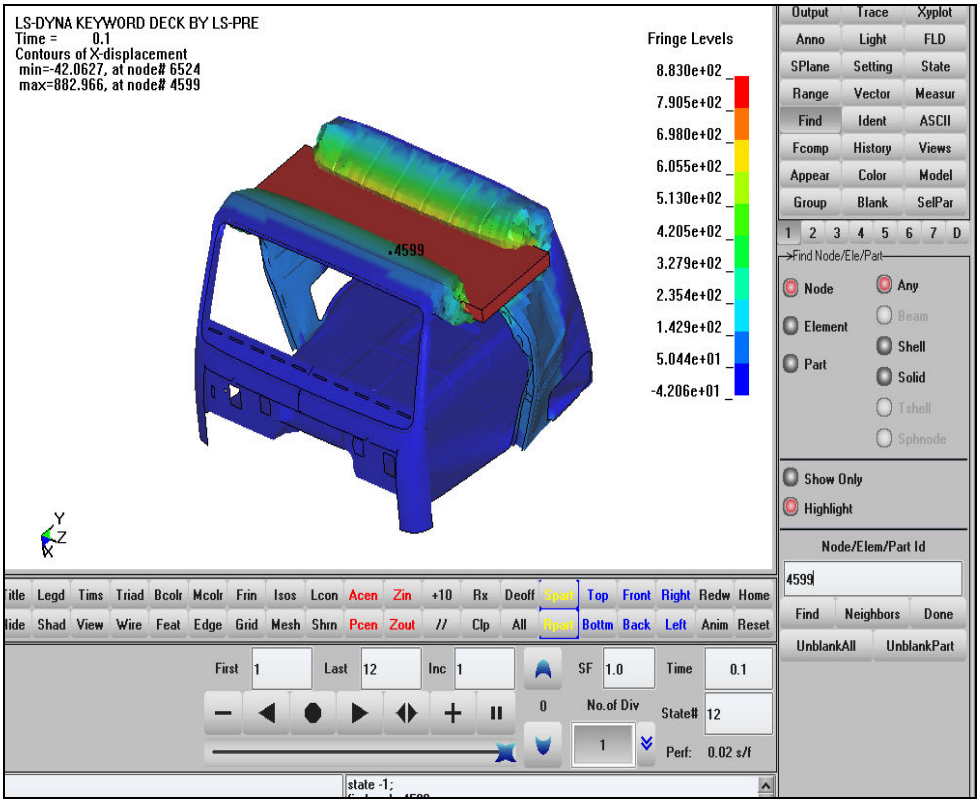

Рис. 5.33. Моделирование процессов смятия кабины неподвижными препятствиями, результаты расчета

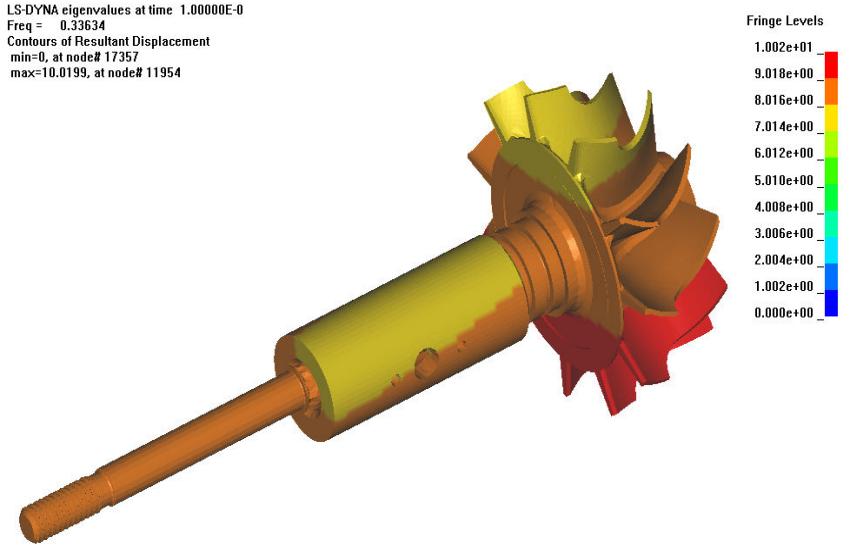

Рис. 5.34. Результаты модального анализа системы подшипник-ротор турбокомпрессора

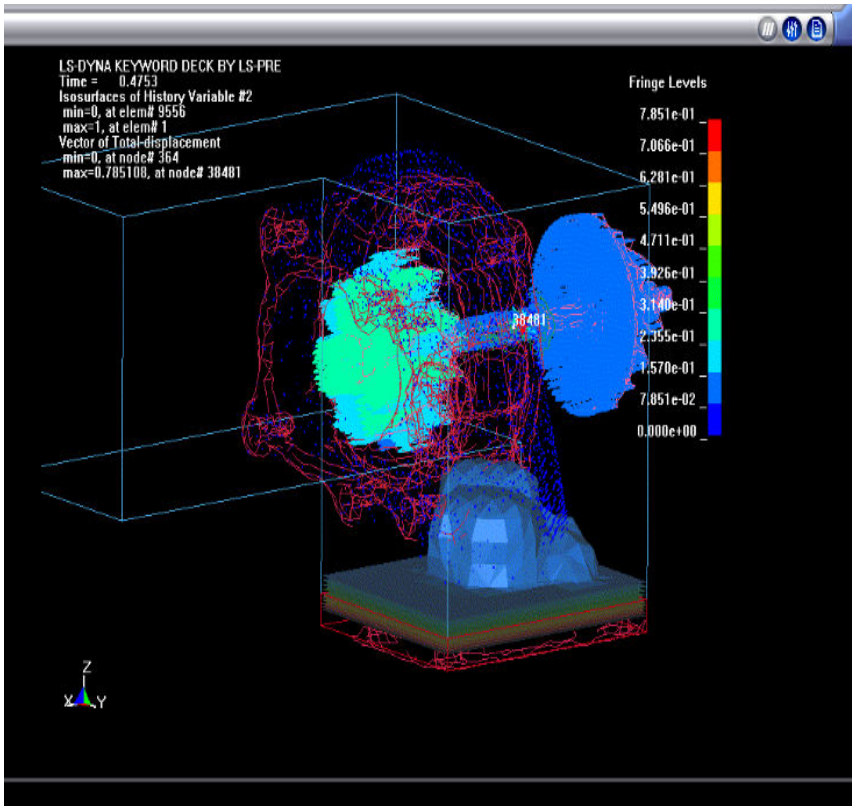

Рис. 5.35. Моделирование взаимодействий газового потока с колесом турбины турбокомпрессора

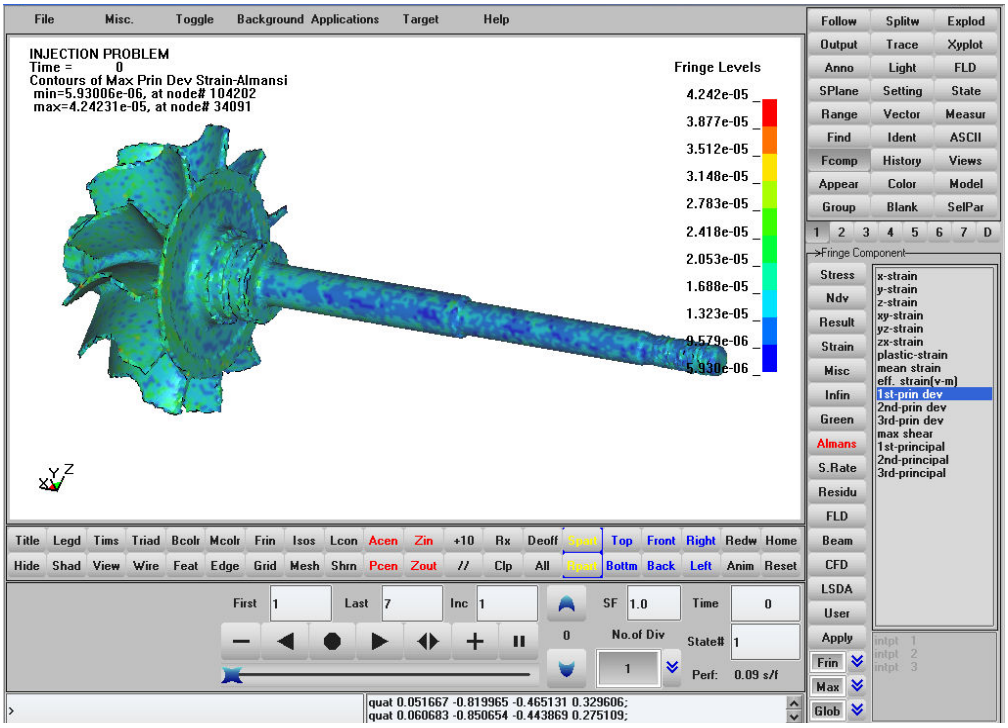

Рис. 5.36. Напряженное состояние ротора, обусловленное влиянием газового потока

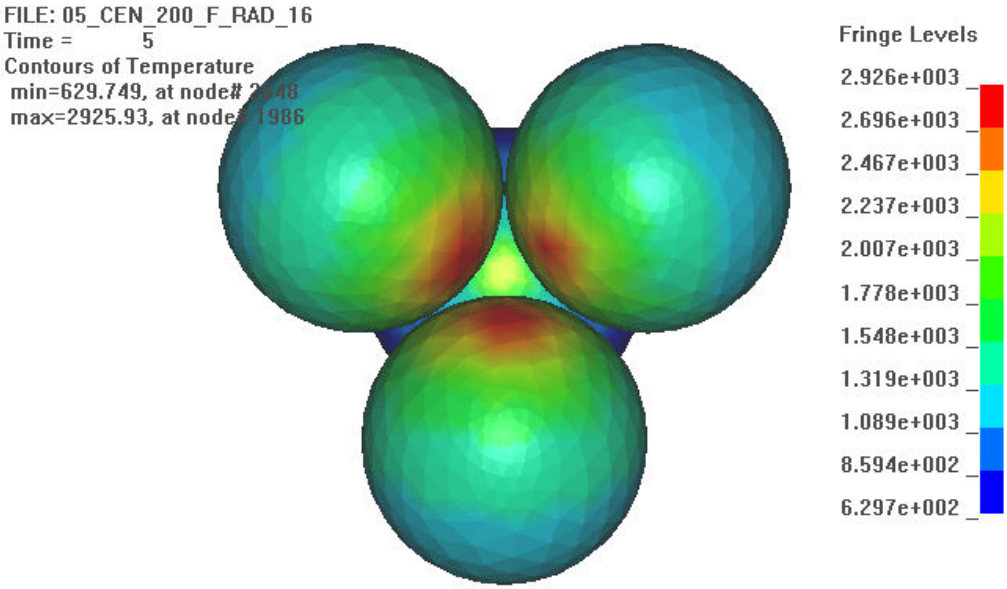

Рис. 5.37. Распределение полей температур в слоях титанового порошка

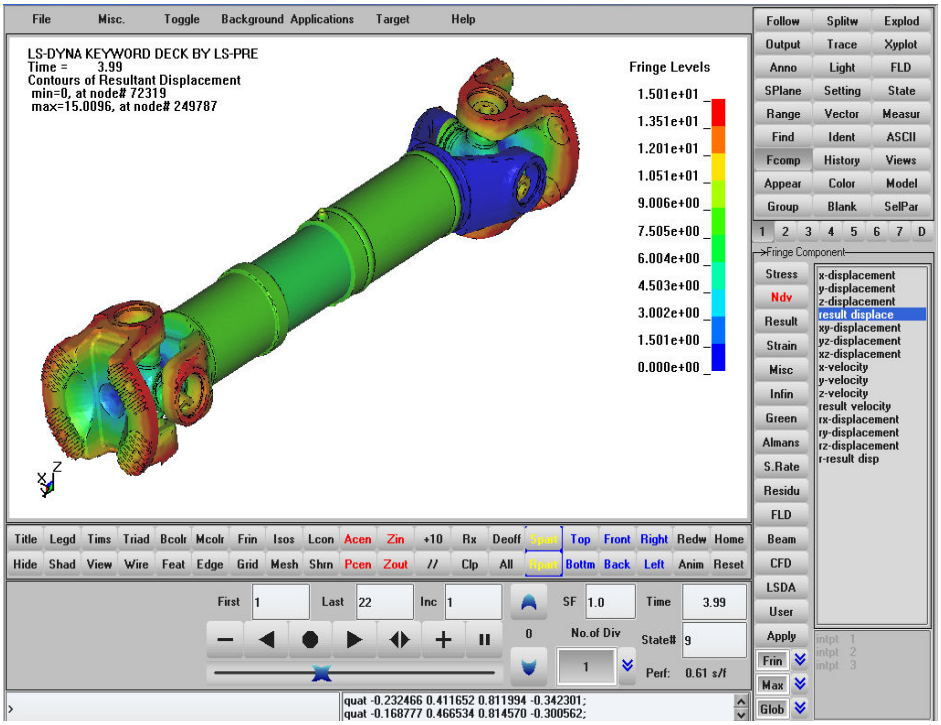

Рис. 5.38. Распределение напряжений в деталях карданного вала при заданном режиме работы

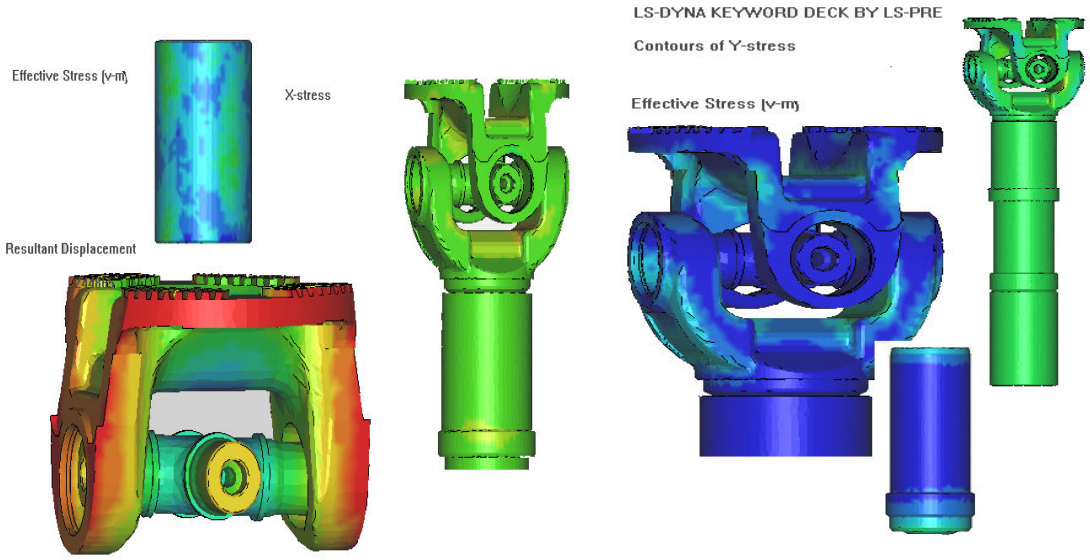

Рис. 5.39. Распределение эквивалентных напряжений в деталях карданного вала

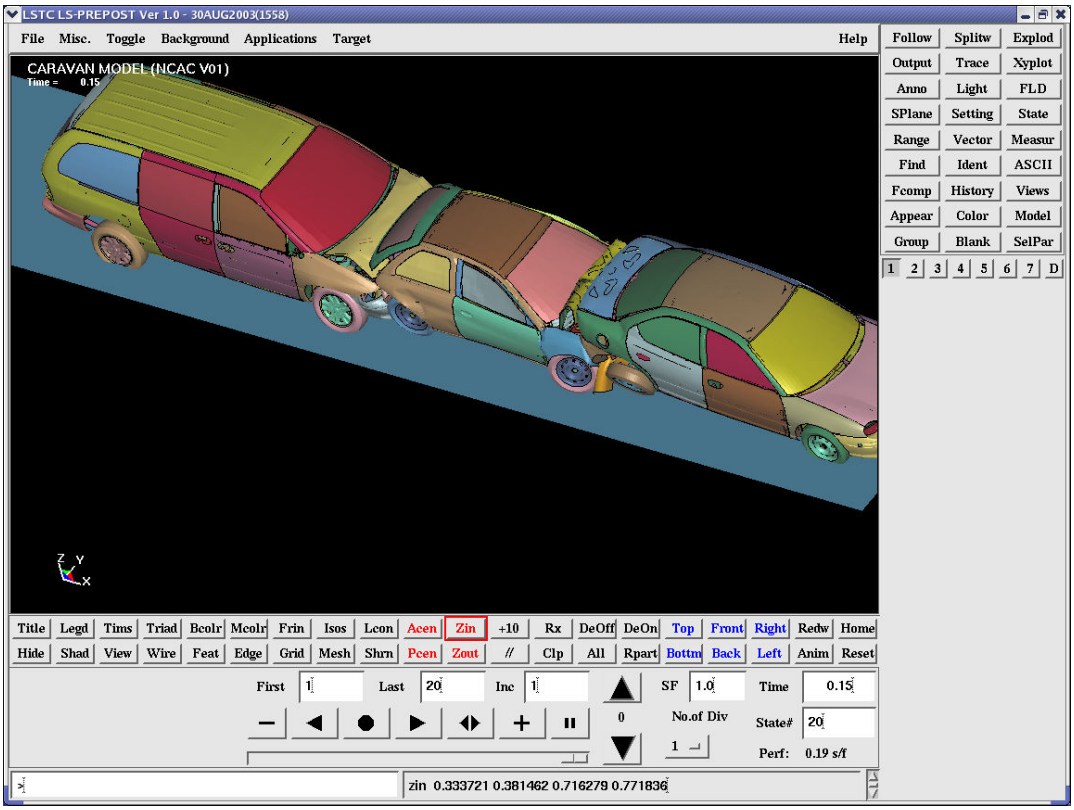

Рис. 5.40. Моделирование столкновения трех автомобилей

Время прогона задачи столкновения трех автомобилей (3 Vehicle Collision, конечно-элементная модель трех автомобилей включает конечные элементы разных типов и содержит порядка 1,5 млн. степеней свободы) на 35 процессорах К-1000 равно 6958 сек.

Время прогона задачи на всех 128 процессорах конфигурации «СКИФ К-500» равно 9989 секунд. Данный результат был принят к опубликованию на сайте: http://www.topcrunch.org/benchmark_results.sfe, содержащем информацию о производительности пакета LS-DYNA на различных компьютерных конфигурациях. Результат прогона задачи позволил суперкомпьютеру «СКИФ К-500» занять почетное общее четвертое место и первое место среди установок, основанных на процессорах имеющих разрядность 32 бита. Более высокие места заняли кластера имеющие разрядность 64 бита. Так время прогона задачи столкновения трех автомобилей на 35 процессорах суперкомпьютера «СКИФ К-1000» равно 6958 сек. Это позволило в декабре 2004 года кластеру «СКИФ К-1000» занять первое место на данной задаче. При этом было задействовано только 12% от общего количества узлов «СКИФ К-1000», что подтверждает хорошую адекватность выбранного технологического решения для задач конечно-элементного анализа.

Проведенные вычислительные эксперименты в среде LS-DYNA на

суперкомпьютерах СКИФ показали приемлемую для широкого практического применения надежность вычислительных комплексов, а также высокую производительность вычислений, на большинстве проведенных расчетов линейно возрастающую пропорционально числу используемых процессоров. Апробированы прямые интерфейсы пакетов ANSYS и NASTRAN в среду LS-DYNA.

Таким образом, технические характеристики суперкомпьютеров «СКИФ» позволяют на качественно новом уровне ставить и решать актуальные задачи компьютерного моделирования машиностроительных конструкций, направленные на повышение их конкурентоспособности и ресурса.

Направления дальнейших исследовательских и прикладных работ – совершенствование имеющихся методик расчета сварных конструкций, увязанных с процессами компьютерного проектирования сборочно-сварочной оснастки, освоение интерфейсов ADAMS-LS-DYNA, создание методик построения виртуальных испытательных стендов для объектов автотракторного и сельскохозяйственного назначения.

5.5. Прикладные комплексы обработки космической информации

5.5.1. Программная система синтеза фокусированных радиолокационных изображений

Основными разработчиками программной системы синтеза фокусированных радиолокационных изображений являются Научно-исследовательский институт космических систем (НИИ КС) и ИПС РАН.

Программная система предназначена для формирования фокусированных радиолокационных изображений из голограмм радиолокационной станции космического базирования «Алмаз» и состоит из следующих компонентов:

– графический интерфейс пользователя (rliX-niiks);

– терминальное приложение обработки голограммы и формирования радиолокационного изображения (rli-niiks).

Формируемое изображение предназначено для решения задач разведки (обнаружение объектов, распознавание объектов), картографирования местности, экологического мониторинга, оценки топографического рельефа местности и т.д.

При реализации программы синтеза используются методы сжатия сигналов в радиолокатор с синтезированной апертурой по азимуту, в частности, реализован метод быстрой свертки, основанный на операции быстрого преобразования Фурье. Автофокусировка выполняется по методу минимизации энтропии.

Особенности программной реализации:

– возможность обработки в режиме автоматической фокусировки изображения;

– возможность автоматического сохранения полученных результатов с целью ускорения обработки поступающих данных;

– отображение статистики в нижней части окна;

– вывод оператору всевозможных сообщений об ошибках.

Программа синтеза реализована на языке программирования T++. Графический интерфейс пользователя выполнен на языке программирования C++.

Управляющий компьютер кластера должен имень следующие характеристики:

– CPU Pentium-533, 128 МБ RAM (желательно 256), жесткий диск не менее 1000 МБ, net card 100Мбит/с.

Программное обеспечение управляющего компьютера:

– OS Linux, ядро не ниже 2.4, XFree86, библиотеки libc, libstdc++, gtk+-2.0;

– OpenTS-1.999, библиотеки libdmpi_lam, libdmpi_scali.

Узлы кластера должны иметь следующие характеристики:

– CPU Pentium-1400, 256 МБ RAM (желательно 512), жесткий диск не менее 300 МБ, net card 100Мбит/с.

Программное обеспечение узлов кластера:

– OS Linux, ядро не ниже 2.4, библиотеки libc, libstdc++;

– OpenTS-1.999, библиотеки libdmpi_lam, libdmpi_scali.

Тестирование программы синтеза радиолокационных изображений проводилось на компьютере со следующими основными характеристиками:

Оборудование:

– CPU AMD Athlon 2500+;

– MB ASUS A7V600 (VIA KT600);

– SDRAM DDR2700 512МБ.

Программное обеспечение:

– Linux Red Hat 9, ядро 2.4.8-20;

– OpenTS-1.999.

Для тестирования применялись голограммы размером 8,4МБ и 22,5МБ.

Результаты тестирования сведены в таблицу 5.6.

Таблица 5.6.

Размер голограммы, МБ	Время выполнения для фиксированного значения наклонной дальности, с	Время выполнения для режима автофокусировки, с
8,4	6,298	14,112
22,5	15,582	35,4

Проводилось удаленное тестирование программы rli-niiks на кластере «СКИФ», размер голограммы – 22,5МБ.

Результаты тестирования сведены в таблицу 5.7.

Таблица 5.7

Количество процессоров	Время выполнения для фиксированного значения наклонной дальности, с	Время выполнения для режима автофокусировки, с
1	19.970	38.655
4	14.748	25.152
8	10.239	16.366
12	8.345	12.117
16	7.596	9.744
20	7.015	8.415
24	6.162	8.148
26	5.948	7.467

5.5.2. Программный комплекс моделирования широкополосных пространственно-временных радиолокационных сигналов

Программный комплекс моделирования широкополосных пространственно-временных радиолокационных сигналов разработан НИИ КС и ИПС РАН.

Программный комплекс предназначен для моделирования широкополосных пространственно-временных радиолокационных сигналов и состоит из следующих компонентов:

– графический интерфейс пользователя (modX-niiks);

– терминальное приложение моделирования и записи результатов в файл голограммы (mod-niiks);

– терминальное приложение синтеза радиолокационного изображения из голограммы (mod-synth-niiks).

Программа позволяет смоделировать сигнал с одно- и много- то-

чечной фоно-целевой обстановкой (кадр) на земной поверхности для вводимых пользователем параметров сигнала и съемки.

Радиолокационные изображения предназначены для решения задач разведки (обнаружение объектов, распознавание объектов), картографирование местности, экологического мониторинга, оценки топографического рельефа местности и т.д.

При реализации программы синтеза использовались методы сжатия сигналов в радиолокаторах с синтезированной апертурой по азимуту, в частности, реализован метод быстрой свертки, основанный на операции быстрого преобразования Фурье.

Программа работает в интерактивном режиме: требует ввода пользователем исходных данных.

Особенности программной реализации:

– возможность формирования в результате моделирования голограммы и последующая ее обработка с целью получения радиолокационного изображения;

– возможность сохранения синтезированной голограммы и синтезированного изображения;

– отображение статистики в нижней части окна;

– вывод оператору всевозможных сообщений об ошибках.

Программа моделирования реализована на языке программирования T++. Графический интерфейс пользователя и программа синтеза изображения реализованы на языке программирования C++

Программный комплекс предназначен для работы на кластере со следующими минимальными требованиями к его составляющим.

Управляющий компьютер кластера должен иметь следующие характеристики:

– CPU Pentium-533, 256 МБ RAM (желательно 512), жесткий диск не менее 1000 МБ, net card 100Мбит/с.

Программное обеспечение управляющего компьютера:

– OS Linux, ядро не ниже 2.4, XFree86, библиотеки libc, libstdc++, gtk+-2.0.

– OpenTS-1.999, библиотеки libdmpi_lam, libdmpi_scali.

Узлы кластера должны иметь следующие характеристики:

– CPU Pentium-1400, 256 МБ RAM (желательно 512), жесткий диск не менее 300 МБ, net card 100Мбит/с.

Программное обеспечение узлов кластера:

– OS Linux, ядро не ниже 2.4, библиотеки libc, libstdc++.

– OpenTS-1.999, библиотеки libdmpi_lam, libdmpi_scali.

Тестирование программы моделирования проводилось на компьютере со следующими основными характеристиками:

Оборудование:

– CPU AMD Athlon 2500+;
– MB ASUS A7V600 (VIA KT600);
– SDRAM DDR2700 512МБ.

Программное обеспечение:

– Linux Red Hat 9, ядро 2.4.8-20;
– OpenTS-1.999.

Для тестирования были использованы модели местности размером M на N точек с расстоянием между точками 5 м. Время выполнения программы моделирования (синтез голограммы) для модели 20 на 20 точек – 25,882 с и для модели 50 на 50 точек – 174,265 с. Остальные параметры моделирования: наклонная дальность – 300 км; высота полета – 500 км; длинна волны – 0,05 м; скорость полета – 7500 м/с; ширина ДНА – 0,5°; длительность импульса – 1 мкс; ширина спектра – 100 МГц; частота АЦП – 150 МГц.

Проводилось удаленное тестирование программы mod-niiks на кластере «СКИФ», размер моделируемой обстановки – 20 х 20 и 50 х 50 точек. Остальные параметры моделирования: расстояния “x” и “y” между точками – 10 м; наклонная дальность – 300 км; высота полета – 500 км; длина волны – 0,05 м; скорость полета – 7500 м/с; ширина ДНА – 0,5°; длительность импульса – 1 мкс; ширина спектра – 100 МГц; частота АЦП – 150 МГц.

Результаты тестирования сведены в таблицу 5.8.

Таблица 5.8

Количество процессоров	Время выполнения при размере моделируемой обстановки 20 х 20 точек	Время выполнения при размере моделируемой обстановки 50 х 50 точек
1	31.715	254.980
4	16.589	133.280
8	10.624	84.245
12	7.226	56.550
16	5.445	42.217
20	4.526	34.264
24	3.888	29.633
26	3.500	25.865

5.5.3. Программный комплекс поточечной обработки цветных и полутоновых видеоданных космических систем дистанционного зондирования

Основными разработчиками программного комплекса поточечной обработки цветных и полутоновых видеоданных космических систем дистанционного зондирования являются НИИ КС и ИПС РАН.

Программный комплекс предназначен для выполнения операций поточечного преобразования полутоновых и цветных изображений с целью изменения их статистических характеристик и повышения дешифровочных свойств.

Форматы исходного и результирующего изображений – *Bitmap Windows* с разрядностью 8 бит/пиксел (для полутоновых изображений) и 24 бит/пиксел (для цветных изображений). При этом данные палитры (карты цветов), которые могут содержаться в файле с 8-битным изображением, при обработке игнорируются.

Для нормального функционирования программного комплекса поточечной обработки цветных и полутоновых изображений необходимы соответствующие конфигурация аппаратных средств, программное обеспечение и исходная информация.

Управляющий компьютер кластера:

– процессор Intel Pentium или аналогичный с тактовой частотой не менее 700 МГц;

– оперативная память – не менее 128 Мбайт;

– видеоадаптер – SuperVGA с разрешением не менее 1024x768x8 при кадровой частоте не менее 75 Гц. При работе с цветными и спектрозональными изображениями при данном разрешении должна обеспечиваться передача 16 млн.цветов (режим True Color);

– монитор – с диагональю не менее 19’’, зерном 0.25 и частотными характеристиками, соответствующими параметрам видеоадаптера;

– свободный объем дискового пространства не менее 2 Гбайт;

– сетевая карта типа Ethernet с пропускной способностью не менее 100 Мбит/с.

Узлы кластера:

– процессор Intel Pentium или аналогичный с тактовой частотой не менее 1 Гц;

– оперативная память – не менее 256 Мбайт;

– свободный объем дискового пространства не менее 500 Мбайт;

– сетевая карта типа Ethernet с пропускной способностью не менее 100 Мбит/с.

Программное обеспечение:

– операционная система Linux Red Hat 9;

– LAM MPI 7.0.2.;

– OpenTS-1.999-1.

Исходная информация:

– файлы полутоновых и цветных изображений в формате Windows Bitmap разрядностью 8 и 24 бит/пиксел.

Специальные требования и условия организационного и технологического характера не предъявляются.

Поточечное преобразование изображения (наряду с другими видами обработки – сверткой, унитарными преобразованиями и т.д.) является одним из типовых алгоритмов обработки видеоданных.

Данный алгоритм используется для решения следующих задач:

– преобразования вида негатив/позитив;

– расширения динамического диапазона изображения;

– преобразования яркости и контраста изображения;

– нормализации статистических характеристик изображения – линейного преобразования параметров распределения сигналов изображения к заданным значениям;

– видоизменения гистограмм, в результате которого функция распределения результирующего изображения принимает заданный вид с требуемыми значениями параметров распределения;

– псевдоцветового кодирования, в результате которого полутоновое изображение преобразуется в цветное;

– преобразования цветного (спектрозонального) изображения в полутоновое путем линейной комбинации сигналов спектральных каналов (сигналов цветности) с целью получения яркостных (цветовых) сечений и формирования канала яркости цветного (спектрозонального) изображения;

– совместной обработки спектральных каналов цветных изображений с учетом взаимосвязей с целью повышения дешифровочных свойств цветных (спектрозональных) изображений и декорреляции спектральных компонент; заключается в выполнении различных линейных комбинаций по каждому из спектральных каналов результирующего изображения.

Алгоритм поточечного преобразования значений кодов яркости элементов изображения изначально ориентирован на параллельную обработку, поскольку обработка каждого элемента изображения производится независимо от других.

Для снижения вычислительной сложности алгоритма поточечного преобразования изображения целесообразно использовать метод про-

смотровых таблиц, который заключается в следующем. На первом этапе формируется таблица соответствия между значениями кодов исходного и результирующего изображений. Размерность таблицы определяется разрядностью кодовых значений исходного изображения.

Выходное значение данного преобразования может быть как скаляром (в операциях преобразования гистограмм полутоновых изображений), так и трехкомпонентным вектором (при выполнении операций псевдоцветового или ложноцветового кодирования).

В разрабатываемом программно-алгоритмическом комплексе процедура формирования просмотровой таблицы реализована в головном модуле, поскольку не подлежит распараллеливанию.

На втором этапе с использованием таблицы соответствия (просмотровой таблицы) происходит присвоение значений элементам результирующего изображения на основе кодов исходного. Данная процедура реализована в модуле преобразования изображения, написанном на алгоритмическом языке T++ и исполняется в параллельном режиме.

Для цветного изображения разрядностью 24 бит/пиксел использование данного алгоритма целесообразно при размерах изображения свыше 48 Мбайт, а для полутонового разрядностью 8 бит/пиксел – более 256 байт.

Результаты тестирования программного комплекса.

При тестировании программного комплекса поточечной обработки изображений была реализована функция преобразования вида негатив/позитив. Программа выполнялась на управляющей ПЭВМ с центральным процессором Athlon – 2600+, 1Гб ОЗУ.

В результате выполнения программы по обработке двух тестовых изображений (размерами 2075×1905 и 1000×981) сформированы негативы изображений соответствующих размеров.

Фрагменты исходного и результирующего изображений размером 1024×*760 представлены на рис. 5.41 и рис. 5.42 соответственно.

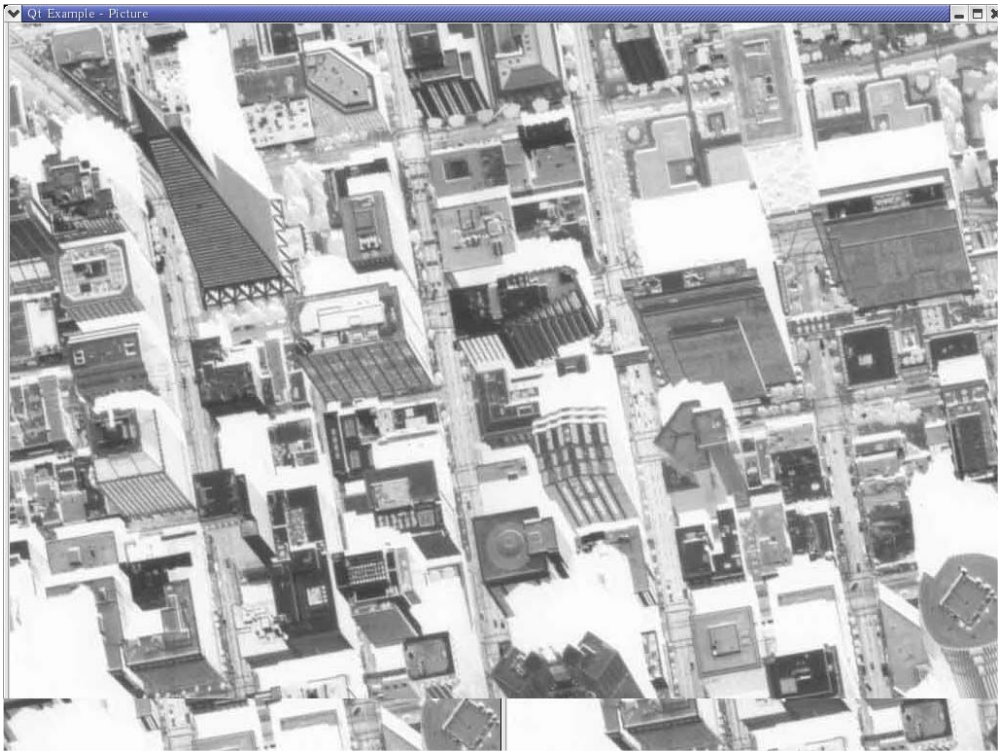

Рис. 5.41. Фрагмент исходного изображения

Рис. 5.42. Фрагмент результирующего изображения

5.6. Интеллектуальные прикладные системы

5.6.1. Программная система извлечения информации из текстов (ПС INEX)

Программная система извлечения информации из текстов (ПС INEX) разработана ИПС РАН.

ПС INEX осуществляет извлечение данных из текстов на основе морфологических, синтаксических, лексических и семантических особенностей обрабатываемого текста.

В состав ПС INEX входят пять тестовых приложений, демонстрирующих возможности различных уровней анализа документа:

– программа, демонстрирующая возможности модулей графематического анализа;

– программа, демонстрирующая возможности морфологического анализа;

– программа, демонстрирующая возможности модуля аннотирования по онтологии;

– программа, демонстрирующая применение правил на языке CPSL;

– программа, демонстрирующая работу всех модулей системы, включая порождение и заполнение фрейма для извлекаемой информации.

Описание технических требований относится ко всем пяти приложениям.

Требования к управляющему компьютеру кластера.

Требования к аппаратному обеспечению:

– процессор Intel Pentium III или AMD Duron,

– не менее 128 МБ оперативной памяти,

– не менее 500 МБ свободного пространства на жестком диске.

Требования к программному обеспечению:

– операционная система Linux с ядром версии не ниже 2.4,

– компилятор gcc версии 3.2.2,

– Scali SSP версии не ниже 2.1.0.

Требования к узлам кластера.

Требования к аппаратному обеспечению:

– процессор Intel Pentium III или AMD Duron,

– не менее 128 МБ оперативной памяти (желательно 256 Мб),

– не менее 100 МБ свободного пространства на жестком диске.

Требования к программному обеспечению:

– операционная система Linux с ядром версии не ниже 2.4,
– Scali SSP версии не ниже 2.1.0.

5.6.2. Программный комплекс рубрикации, кластеризации и поиска полуструктурированных данных

Основными разработчиками программного комплекса рубрикации, кластеризации и поиска полуструктурированных данных являются НИИ механики МГУ им. М.В. Ломоносова и ИПС РАН.

Программа рубрикации текстов осуществляет автоматическое рубрицирование тестовых массивов согласно статистическим портретам рубрик, заданных на этапе обучения. Высокая производительность рубрикации обеспечивается за счет эффективного динамического распараллеливания с использованием «Т-системы».

Программа кластеризации статистических портретов осуществляет построение дерева статистических портретов для повышения качества рубрикации текстов.

Перед установкой программы необходимо убедиться, что в системе корректно установлен пакет «Т-система».

Для установки программы следует установить RPM-пакет программы средствами системы. Дополнительная настройка программы не требуется.

Программа поиска в полуструктурированных базах данных осуществляет вычисление конъюктивного регулярного путевого запроса к полуструктурированным данным в OEM-модели на кластере с использованием «Т-системы» как средства автоматического распараллеливания.

Программа предназначена для работы в параллельном режиме на кластере однако может выполняться на имеющейся в составе «Т-системы» эмуляции кластера. Эффективность работы программы может существенно зависеть от выбора режимов поиска.

5.6.3. Программная система инструментальных средств проектирования интеллектуальных систем

Программная система разработана ИПС РАН.

Программная система инструментальных средств проектирования интеллектуальных систем (ПС «Miracle») предназначена для создания динамических систем, основанных на знаниях.

С использованием инструментального комплекса ПС «Miracle» можно решать следующие классы задач:

– задачи мониторинга, диагностики и прогнозирования состояния сложных технических систем;

– задачи планирования целенаправленного поведения интеллектуальных систем, – выбор последовательности действий по достижению

поставленной цели. Обычно задача планирования предполагает решение задачи прогнозирования. Например, выработка последовательности действий по осуществлению стыковки космического корабля со станцией;

– задачи интеллектуального управления в автоматических системах, например, автоматическое управление роботом, управление сложными движениями летательных аппаратов;

– задачи проектирования – определение конфигурации объектов с точки зрения достижения заданных критериев эффективности и ограничений, например, проектирование бюджета предприятия;

– задачи диспетчирования – распределение работ по времени, составление расписаний.

В состав комплекса ПС «Miracle» включены:

– интерфейс разработчика динамической интеллектуальной системы;

– средства для создания и управления базами знаний, созданных на основе гибридного метода представления знаний;

– средства анализа текущего состояния предметной области, динамики её изменения, и управления целенаправленным поведением динамической системы;

– средства моделирования динамики системы и прогнозирования развития при выборе различных вариантов управления;

– средства планирования поведения системы на основе результатов прогнозирования.

В инструментальном комплексе используется гибридный метод представления знаний на основе семантических сетей, правил и фреймов. Гибридный метод сочетает декларативный и процедурный подходы к представлению знаний.

В методе имеются:

– средства корректного описания структуры и динамики предметных областей;

– средства динамического целеуказания, которые позволяют задавать целевые состояния динамической системы;

– средства выбора управления в зависимости от имеющихся целей.

В инструментальном комплексе имеются средства моделирования динамики и целенаправленного поведения динамических интеллектуальных систем.

Для функционирования ПС «Miracle» необходимо следующее оборудование и ресурсы:

– управляющий компьютер кластера CPU AMD Athlon 1000, не менее 128 МБ RAM, свободное место на жестком диске не менее 20 МБ;

– узлы кластера CPU AMD Athlon 2000, не менее 256 МБ RAM, свободное место на жестком диске не менее 128 МБ.

Необходимое программное обеспечение:

– ОС RedHat Linux версии не ниже 6.2;

– Active Tcl/Tk версии не ниже 8.4.1;

– средство автоматического распараллеливания программ T-система.

5.7. Программный комплекс «Аэромеханика подвижных плохообтекаемых тел»

Основными разработчиками программного комплекса «Аэромеханика подвижных плохообтекаемых тел» являются НИИ механики МГУ им. М.В. Ломоносова и ИПС РАН.

Программный комплекс предназначен для моделирования обтекания набегающим потоком тел, имеющих одну либо три степени свободы. Более конкретно, речь идет о моделировании режимов автоколебаний и авторотации плохообтекаемых тел в сплошной среде при наличии вращательной степени свободы, с учетом динамики дискретных вихревых структур. Возможен как фиксированный закон перемещения точки закрепления тела (центра вращения), так и добавление двух поступательных степеней свободы. Данная задача имеет значение для целого ряда областей, в которых необходимо рассматривать нестационарные дозвуковые течения около вращающихся или колеблющихся плохообтекаемых тел и экранов (парашютные системы, антенные устройства, флюгера, различные оперенные тела при поперечном обтекании, ветроэнергетические установки).

Используется новый, разработанный авторами комплекса, вычислительный вихревой метод, позволяющий впервые в практике применения вихревых методов корректно учесть влияние вязкости. Однако при этом требуется значительно большее число дискретных вихревых элементов по сравнению с традиционным методом дискретных вихрей. В связи с этим остро встает проблема повышения производительности разрабатываемых численных алгоритмов, которую невозможно решить без использования современной многопроцессорной техники. Поэтому настоящий программный продукт направлен на использование вычислительных кластеров семейства «СКИФ» с T-системой автоматического динамического распараллеливания (OpenTS).

Комплекс работает в последовательном и параллельном режимах на вычислительном кластере под управлением ОС Linux с установленными пакетами MPI, OpenTS, Qt и графической подсистемой.

Комплекс предназначен для использования на вычислительном кластере под управлением ОС Linux с установленными пакетами MPI и

OpenTS. Возможен также вариант последовательного запуска на компьютере под управлением ОС Linux. Визуализационная часть может быть запущена на локальной машине и получать данные от вычислительного модуля, работающего на удаленном кластере, посредством механизма сокетов. Для работоспособности модуля, осуществляющего визуализацию, необходимо наличие библиотеки Qt.

5.8. Программная система мультиконформационного моделирования (ПС MULTIGEN)

Программная система мультиконформационного моделирования (ПС MultiGen) разработана Челябинским государственным университетом и ИПС РАН.

ПС MultiGen предназначена для поиска конформеров молекул в заданном энергетическом интервале с последующей их ориентацией в поле рецептора или самосогласованном реакционном поле.

ПС MultiGen использует преимущества параллельной архитектуры, в частности:

– обеспечивает возможность распределения вычислительной нагрузки на вычислительные узлы кластерного уровня в процессе конформационного анализа;

– обеспечивает ускорение процесса конформационного поиска при увеличении числа используемых вычислительных узлов.

Поиск конформеров сводится к математической задаче определения локальных минимумов потенциальной энергии в многомерной системе путем перемещения атомов вдоль мод гессиана с последующей квазиньютоновской оптимизацией. Перемещения вдоль 3N мод (N – число атомов) являются параллельными процессами и могут производиться в параллельном режиме.

Для поиска наиболее вероятных конформационных состояний молекул может быть использована следующая методология:

1) Рассчитывается гессиан.

2) Далее осуществляется перемещение атомов молекулы вдоль каждой из колебательных мод, как в положительном, так и в отрицательном направлении до достижения ближайшего максимума.

3) После преодоления максимума вновь производится минимизация потенциальной энергии молекулы по координатам атомов. При этом молекула попадает в иной минимум потенциальной энергии.

4) Производится сравнение энергии вновь полученного конформера и наиболее выгодной конформации. Если разница энергий не превышает некоторого предела (ΔElim) и вновь найденный конформер не был уже ранее обнаружен, то геометрия полученного конформера запоминается.

5) Для каждого из вновь найденных конформеров рассчитывается гессиан и вновь производятся операции 2 – 5.

Очевидно, что начало решения данной задачи представляет собой разветвленный алгоритм с параллельными операциями 2 и 3, представленный на рис. 5.43. Из исходного конформера К0 определяются n конформеров (конформеры первого уровня К1i). Каждый из них дает начало новому поиску. Из них отыскиваются m новых конформеров (конформеры второго уровня К2i). Каждый из конформеров второго уровня также дает начало новому поиску и т.д. Однако, впоследствии, на более высоких уровнях находятся конформеры, ранее уже обнаруженные (операция 4) и число ветвей поиска сокращается.

Поиск прекращается, если на новом уровне не находится ни одного ранее не найденного конформера. Каждая стрелка на рис. 5.43 включает в себя операции 1 – 4.

Разработанная система состоит из четырех основных функциональных компонентов: система перемещения атомов молекулы, система оптимизации геометрических характеристик конформеров, блок сопоставления геометрических характеристик конформеров и система ориентации конформеров.

Необходимость системы перемещения атомов молекулы вдоль мод гессиана обусловлена тем, что конформационные переходы в молекуле осуществляются вследствие колебательного движения атомов. Направление колебательного движения определяется как собственные векторы матрицы силовых постоянных (то есть вторых производных энергии по колебательным координатам), которая является гессианом системы. Таким образом, для имитации конформационного перехода производится вычисление гессиана системы, определение собственных векторов гессиана с последующим перемещением атомов вдоль данных мод в прямом и противоположном направлениях до преодоления ближайшего максимума потенциальной энергии.

Необходимость системы перемещения атомов молекулы вдоль мод гессиана обусловлена тем, что конформационные переходы в молекуле осуществляются вследствие колебательного движения атомов. Направление колебательного движения определяется как собственные векторы матрицы силовых постоянных (то есть вторых производных энергии по колебательным координатам), которая является гессианом системы. Таким образом, для имитации конформационного перехода производится вычисление гессиана системы, определение собственных векторов гессиана с последующим перемещением атомов вдоль данных мод в прямом и противоположном направлениях до преодоления ближайшего максимума потенциальной энергии.

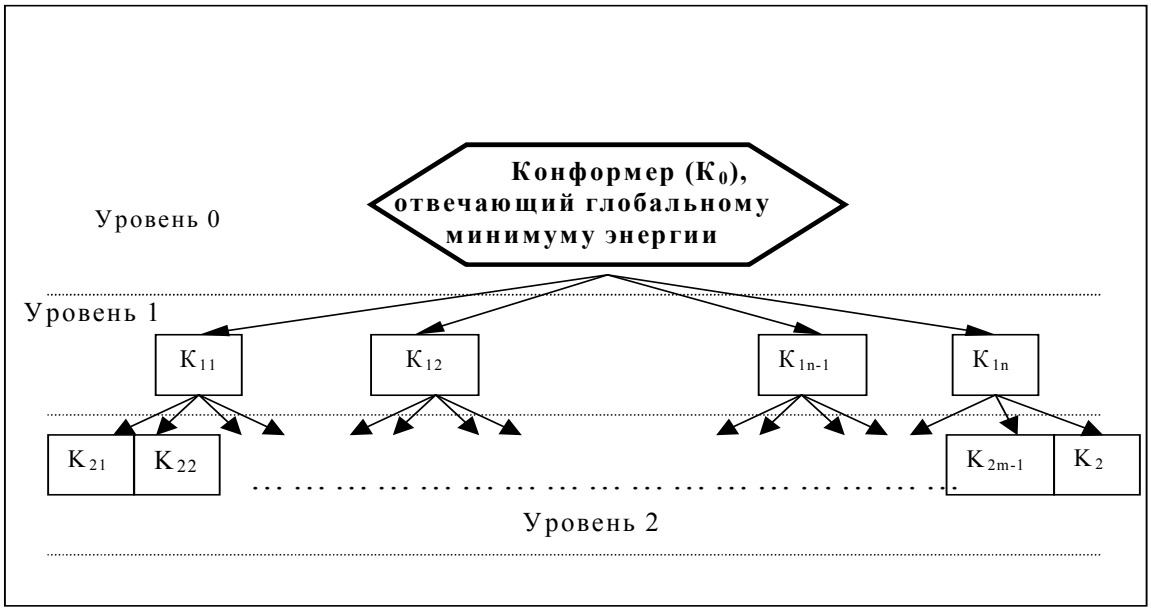

Рис. 5.43. Разветвленный алгоритм поиска вероятных конформеров

После преодоления ближайшего максимума потенциальной энергии требуется привести молекулу к минимуму потенциальной энергии, соответствующему вновь найденному конформеру. Для этого используется система оптимизации геометрических характеристик конформеров, реализующая квази-ньютоновский алгоритм приведения системы к минимуму потенциальной энергии.

Блок сопоставления геометрических характеристик конформеров производит сопоставление вновь обнаруженного конформера с ранее найденными, для установления уникальности полученной структуры.

Система ориентации суперпозиции конформеров в поле производит оптимизацию ориентации конформеров в поле рецептора или самосогласованном реакционном поле для последующей мультиконформационной оценки биологической активности или реакционной способности соединения.

ПС MultiGen предназначена для использования в среде версии не ниже 2.4 дистрибутива RedHat ОС Linux совместно с программным комплексом Scali SSP версии не ниже 2.1.0. Кроме того, требуется предустановка Т-системы с открытой архитектурой.

Компоненты ПС MultiGen реализованы на языке программирования T++.

5.9. Кардиологическая экспертная система реального времени «ADEPT-C»

Кардиологическая экспертная система реального времени «ADEPT-C» разрабатана АНО ИВВиИС и ИПС РАН.

Система должна быть введена в опытную эксплуатацию в лечебных учреждениях Ленинградской области. При эксплуатации кардиологического комплекса будут использоваться аппаратные средства (вычислительная и медицинская аппаратура) отечественного производства. В качестве аппаратной реализации вычислительной техники будут использованы кластеры, созданные в процессе выполнения программы «СКИФ». В качестве периферийной медицинской аппаратуры, по согласованию с привлекаемыми для выполнения данной программы врачами Ленинградского областного кардиологического диспансера, будет использован компьютерный анализатор ЭКГ ЗАО «Микард-Лана», Санкт-Петербург.

Функционально программный комплекс ADEPT-C должен реализовать клиент-серверную схему взаимодействия между компонентами.

Он состоит из трех независимых компонентов:

– ядро комплекса ADEPT-C, функционирующее на кластере семейства «СКИФ» (сервер);

– АРМ лечащего врача (клиент), реализующий функции сбора и пересылки медицинской информации (как инструментальных данных – записи и параметров ЭКГ, так и результатов врачебного расспроса и осмотра);

– АРМ администратора (клиент).

Программный комплекс ADEPT-C обеспечивает возможность выполнения автоматизированных телемедицинских консультаций в области кардиологической диагностики в реальном режиме времени. Это подразумевает следующие функциональности:

– представление пользователю (с учетом его прав и ролей) доступа к единой базе данных, содержащей электронную информацию о пациентах;

– выдача заключения экспертной системы на основе правил (rules-based reasoning);

– выполнение процедуры имитационного моделирования изменчивости электрической активности сердца на основе данных ЭКГ-обследования;

– просмотр, редактирование и пополнение базы знаний экспертной системы.

Программный комплекс ADEPT-C обеспечивает возможность своей установки (инсталляции) в качестве компоненты ПО КУ «СКИФ» на кластере.

Реакция системы на запросы пользователя отвечает требованиям, предъявляемым к системам реального времени. В зависимости от мощности используемого аппаратного обеспечения открытие страницы не занимает более 3-5 сек., обработка запроса на моделирование кардиограммы и выдачи результатов сопоставления не более 1 минуты (при параметрах системы, предусмотренных по умолчанию).

С точки зрения обеспечения надежного функционирования системы выполняются следующие условия:

1) Серверная часть комплекса автоматически восстанавливает свою работу в случае возникновения аппаратных сбоев на узлах кластера.

2) Серверная часть комплекса формирует log-файлы, содержащие информацию о текущем состоянии и возможных сбоях.

3) Серверная часть комплекса увеличивает скорость обработки одного клиентского запроса в случае увеличения числа задействованных вычислительных узлов кластера «СКИФ».

4) Клиентские части комплекса («АРМ лечащего врача» и «АРМ администратора») обеспечивают проверку корректного ввода исходных данных и обрабатывают соответствующие ошибки.

5) Клиентские части комплекса («АРМ лечащего врача» и «АРМ администратора») корректно обрабатывают сбои, возникающие при соединении с серверной частью, в случае сбоя коммуникационной сети или выключения серверной вычислительной системы.

Функционирование кардиологической экспертной системе реального времени «ADEPT-C» обеспечивается с использованием двух классов технических вычислительных средств: «сервер» и «клиент», к параметрам которых предъявляются следующие технические требования.

Сервер:

аппаратная платформа кластерного уровня суперкомпьютеров «СКИФ», включающая:

– управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура);

– вычислительные узлы кластерного уровня (базовый вычислительный модули кластерного уровня, БВМ КУ) со следующими параметрами: класс процессоров – не менее Pentium III; количество оперативной памяти на узле – не менее 128 Мб;

– системную сеть кластера (SCI), объединяющую вычислительные узлы;

– вспомогательную сеть 100 Mbit Ethernet, с поддержкой TCP/IP, объединяющую управляющую ЭВМ и вычислительные узлы;

– внешний сетевой интерфейс с пропускной способностью не менее 10 Мб/с.

Клиент:

– персональный компьютер на базе процессора – не менее Pentium III;

– свободное дисковое пространство – не менее 10 Мб;

– аппаратура для соединения по TCP/IP протоколу.

Серверная часть ADEPT-C функционирует на кластере «СКИФ», в среде ОС Linux (дистрибутив RedHat 7.3 и выше), с предустановленным программным комплексом Scali SSP. Кроме того, она требует корректной предустановки на кластере следующих программных средств и библиотек:

– коммуникационная библиотека Scali MPI в составе программного комплекса Scali SSP;

– программная библиотека работы с XML-документами Xerces,

– СУБД PostgreSQL (версия не ниже 7.3), с предустановленной библиотекой libpq.

Клиентские части ADEPT-C предназначены для функционирования на персональном компьютере под управлением Windows 2000/XP, с предустановленным программным комплексом Java 2 SDK, Standard Edition, версии 1.5.0 или выше. Для обеспечения сбора данных ЭКГ-обследований требуется предустановка специализированных программных средств управления кардиомонитором «Кардиометр-МТ» (драйверы сбора и обработки кардиологической информации, интерфейс взаимодействия).

6. Развитие суперкомпьютерного направления «СКИФ»

6.1. Практическое использование суперкомпьютерной техники в Беларуси и в России

Создание суперкомпьютерных конфигураций «СКИФ К-500» и «СКИФ К-1000» позволило уже в 2003-2004 годах развернуть в ОИПИ НАН Беларуси фронт работ по практическому использованию суперкомпьютерных технологий.

Развернуты работы по созданию сквозной компьютерной технологии проектирования, испытаний и технологической подготовки турбокомпрессоров для наддува дизельных двигателей Минского моторного завода. Заказчик – Борисовский завод агрегатов Министерства промышленности Республики Беларусь. Компьютерная технология базируется на платформе СКИФ и программных системах LS-DYNA и Star-CD, адаптированных для работы на многопроцессорных системах.

Проведены работы по использованию суперкомпьютеров «СКИФ» для расчетов и моделирования остовов перспективных универсальных тракторов «Беларусь», которые принципиально не могут быть рассчитаны на традиционных средствах вычислительной техники. Получены положительные решения, ведутся дальнейшие работа по совершенствованию методик.

В рамках отраслевой программы «Компьютерные технологии проектирования новых изделий» Министерства промышленности Республики Беларусь проводятся работы по расчету динамических характеристик почвообрабатывающих агрегатов с использованием программного обеспечения конечно-элементных расчетов, развернутого на семействе кластеров «СКИФ».

Поставлены вычислительные эксперименты по расчетам на «СКИФ» несущих конструкций карьерных самосвалов БелАЗ и шахтных крепей.

Проведены работы по моделированию на суперкомпьютерах «СКИФ» процессов лазерного спекания порошковых материалов для технологий быстрого прототипирования и изготовления медицинских изделий.

В интересах МАЗа совместно с НИРУП «Белавтотракторостроение» НАН Беларуси проводятся работы по суперкомпьютерному моделированию столкновений транспортных средств с неподвижными препятствиями.

В рамках договора с Комитетом государственной безопасности суперкомпьютеры «СКИФ» задействованы в задачах криптоанализа и от-

работки технологий решения задач перебора большой размерности.

Завершена клиническая апробация аппаратно-программного кардиологического комплекса на основе суперкомпьютерных вычислительных модулей для исследования микроциркуляторного звена сердечно-сосудистой системы патентуемым методом биомикроскопии (исполнители работ – ОИПИ НАН Беларуси, РНПЦ «Кардиология», УП «НИИЭВМ»).

Совместно с Республиканским Гидрометеорологическим центром Беларуси проведено удаленное тестирование счета модели регионального прогноза погоды на 48 часов. В результате счета задачи на 32 процессорной установке «СКИФ» прогностические значения приземного давления, составляющих скорости ветра и осадков получены в 12 раз быстрее, чем при расчетах на обычных вычислительных средствах Гидрометеоцентра.

Для практического внедрения результатов Программы «СКИФ» организован режим удаленного доступа к вычислительным ресурсам суперкомпьютерных конфигураций ОИПИ НАН Беларуси в рамках заключенных договоров (соглашений) о научно-техническом сотрудничестве с:

– Институтом механики металлополимерных систем им. В.А. Белого НАН Беларуси (численные модели в области механики материалов и расчет изделий из полимерных композитов);

– Институтом молекулярной и атомной физики НАН Беларуси (исследовательские работы по моделированию на суперкомпьютерных установках «СКИФ» в среде LS-DYNA процессов лазерного спекания порошковых материалов для технологий быстрого прототипирования);

– Институтом математики НАН Беларуси (многомерные задачи математической физики, моделирование физико-химических процессов, параллельные алгоритмы вычислительной математики);

– Институтом тепло- и массообмена им А.В. Лыкова НАН Беларуси;

– ООО «Информационные системы» (банковские информационные технологии);

– Белорусским национальным техническим университетом (подготовка кадров, CAD/CAE/CAM/PDM – технологии);

– Государственным экспертно-криминалистическим центром Министерства внутренних дел (ГЭКЦ МВД) Республики Беларусь (специализированные программно-технические средства для борьбы с телефонным терроризмом);

– Белорусским государственным университетом информатики и радиоэлектроники (инструментальные средства проектирования интел-

лектуальных систем);

– ООО «Софтклуб» (банковские информационные технологии).

Создание в рамках совместной белорусско-российской программы «СКИФ» суперкомпьютеров «СКИФ К-500» и «СКИФ К-1000» позволило создать в ОИПИ НАН Беларуси Республиканский суперкомпьютерный центр коллективного пользования для проектирования и анализа объектов новой техники.

Суперкомпьютеры триллионного диапазона производительности из-за высокой стоимости не предназначены для выпуска большими партиями, так как закупка таких конфигураций зачастую не по карману даже крупным предприятиям. Так например, стоимость суперкомпьютера «СКИФ К-1000» составила примерно 2000 тыс. долларов США, что в несколько раз меньше, чем аналогичные разработки зарубежных производителей.

Учитывая высокую стоимость суперкомпьютерных конфигураций, территориальную компактность Республики и развитую телекоммуникационную инфраструктуру, был выбран путь создания в рамках Национальной академии наук Беларуси Республиканского суперкомпьютерного центра коллективного пользования для проектирования и анализа объектов новой техники. Проблема создания в системе НАН Беларуси суперкомпьютерного центра коллективного пользования возникла в связи с отставанием от ведущих мировых держав в развитии и применении новейших наукоёмких информационных технологий, нацеленных на решение сложных задач машиностроения, геологоразведки, контроля окружающей среды, транспорта и связи, государственных, коммерческих, военных и других приложений.

Создание суперкомпьютерного центра в ОИПИ НАН Беларуси для развития и внедрения в НАН Беларуси наукоемких информационных технологий обеспечило возможность предоставления услуг для решения наукоёмких задач, возникающих в промышленности и в других областях народного хозяйства страны, требующих компьютерных и информационных ресурсов, владение которыми недоступно или экономически нецелесообразно для отдельных организаций.

6.2. Создание телекоммуникационной сети, объединяющей участников совместной Программы Беларуси и России, с выделенным высокоскоростным каналом связи

Работы по созданию телекоммуникационной сети, объединяющей участников совместной Программы Беларуси и России, выполнены в НИРУП «Национальный центр информационных ресурсов и технологий» НАН Беларуси в рамках программы «СКИФ». Основные исполнители работы – С.А. Анейчик (ответственный исполнитель), Ю.В. Костю-

кевич, О.А. Носиловский, Л.В. Гарустович, В.М. Горулько, О.К. Мойсейчук.

Формирование телекоммуникационной инфраструктуры распределенной суперкомпьютерной системы выполнялось в рамках задания программы Союзного государства «СКИФ» – «Разработать и реализовать программно-технические решения с использованием суперкомпьютеров и создать на их основе телекоммуникационную сеть, объединяющую участников совместной программы Беларуси и России, с выделенным высокоскоростным каналом связи». Работы проводились по двум основным направлениям:

1) Организация высокоскоростного канала для обмена информацией научных сетей Беларуси и России в рамках программы «СКИФ».

2) Разработка и создание телекоммуникационной инфраструктуры, обеспечивающей белорусским потребителям режим удаленного доступа к вычислительным ресурсам суперкомпьютерного комплекса.

При выполнении указанных работ были решены следующие задачи:

– проведено исследование существующей телекоммуникационной инфраструктуры научных сетей Европы и сети GEANT с точки зрения возможности ее использования для решения задач программы «СКИФ»;

– проведено исследование возможностей подключения к узлам обмена научным трафиком сети GEANT и выбрана оптимальная точка подключения для решения задач программы «СКИФ»;

– проведен анализ и выбраны программно-аппаратные средства для обеспечения коннективности белорусских и российских участников программы «СКИФ»;

– проведен анализ и выбрана оптимальная начальная пропускная способность канала, предусмотрена возможность ее дальнейшего расширения;

– проведена модернизация узла управления высокоскоростным каналом на производственных площадях РО «Белтелеком» для обеспечения функционирования международного канала на скорости не менее 34 Мбит/сек;

– проведена опытная эксплуатация канала обмена информацией научных сетей Беларуси и России через сеть GEANT на скорости 2 Мбит/сек;

– по результатам опытной эксплуатации выполнена настройка конфигурационных параметров канала на особенности использования распределенных суперкомпьютерных ресурсов «СКИФ»;

– выполнены работы по обеспечению удаленного доступа к суперкомпьютерным ресурсам «СКИФ» для ГГУ, КГБ, НИИ ЭВМ, РНПЦ «Кардиология», БГУИР, МТЗ и НИИ цифрового телевидения;

– разработаны и апробированы технические решения по организации удаленного доступа к суперкомпьютерным ресурсам «СКИФ» с использованием радиотехнологий для Гомельского кардиодиспансера и Военной академии.

В качестве базовой при создании телекоммуникационной инфраструктуры суперкомпьютерного комплекса «СКИФ» была использована существующая телекоммуникационная инфраструктура научно-исследовательской компьютерной сети НАН Беларуси BASNET как наиболее крупной научной сети Беларуси. BASNET объединяет более 100 научных учреждений республики, в том числе более 50 институтов Национальной академии наук, являющихся потенциальными потребителями вычислительных ресурсов суперкомпьютерного комплекса. Сеть имеет развитую распределенную инфраструктуру, построенную на базе современных сетевых технологий, располагает узлами в областных центрах и собственной приемо-передающей спутниковой системой. Данное решение позволило уже на начальных этапах с одной стороны организовать сетевое взаимодействие большинства белорусских соисполнителей программы «СКИФ», а с другой стороны – обеспечить коннективность с российскими соисполнителями программы по существующим спутниковым каналам связи.

Коннективность с российскими участниками программы «СКИФ» была обеспечена через паневропейскую научную сеть GEANT путем резервирования гарантированного сегмента. Подключение к GEANT осуществляется по каналу 34 Мбит/сек. через центральный узел польской научной сети PIONIER, расположенный в Познанском суперкомпьютерном и сетевом центре Польской академии наук.

Российские научные сети подключены к GEANT через узел в Стокгольме. Подключение организовано на базе магистральной цифровой волоконно-оптической линии связи на скорости 155 Мбит/сек. Координатором этих работ являлся Межведомственный суперкомпьютерный Центр в Москве.

Структурная схема подключения научно-информационной компьютерной сети Республики Беларусь (НИКС) к единой Европейской научной сети GEANT приведена на рис. 6.1.

Наличие высокоскоростных каналов подключения России и Беларуси к GEANT позволило не только обеспечить резервирование гарантированных сегментов для обмена данными между суперкомпьютерными комплексами, но и реализовать виртуальный шифрованный канал Минск-Москва-Переславль по VPN-технологии.

Схема организации взаимодействия белорусских и российских участников программы «СКИФ» приведена на рис. 6.2.

Организация сетевого взаимодействия российских и белорусских

соисполнителей программы «СКИФ» через сеть GEANT открыла перспективы для сотрудничества в сфере суперкомпьютерных технологий с европейскими странами.

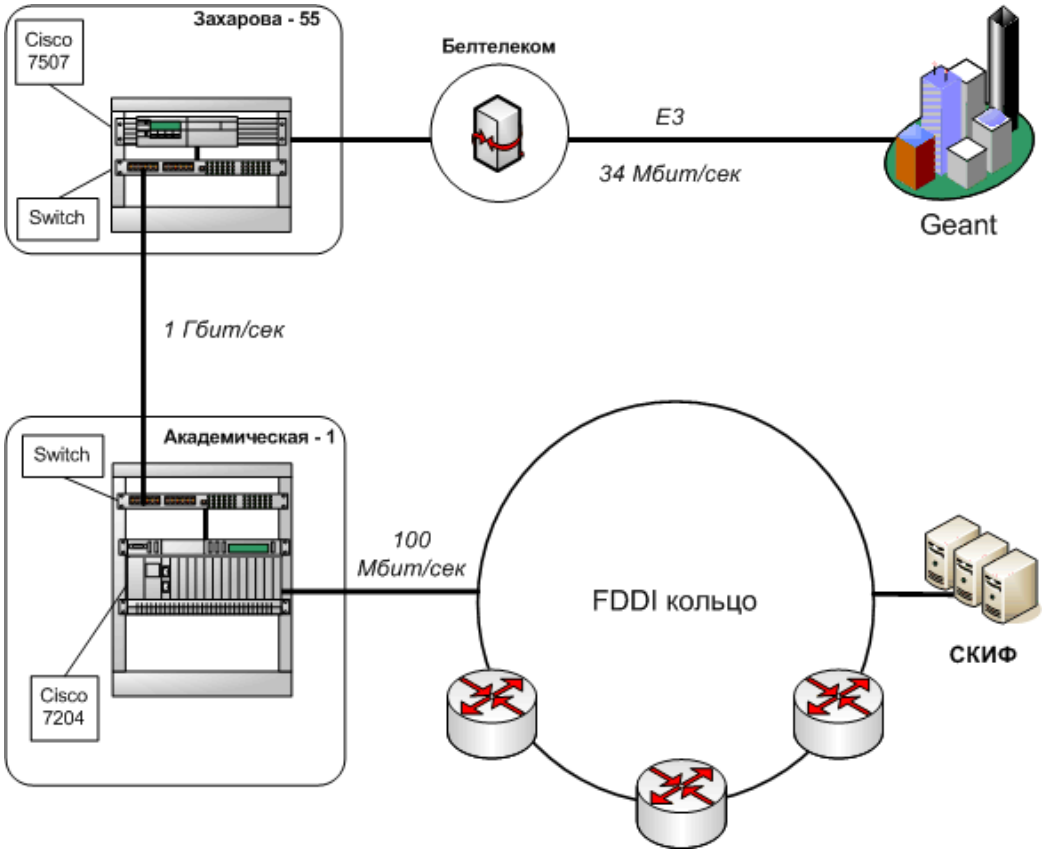

Рис. 6.1. Структурная схема подключения НИКС Республики Беларусь к единой Европейской научной сети GEANT

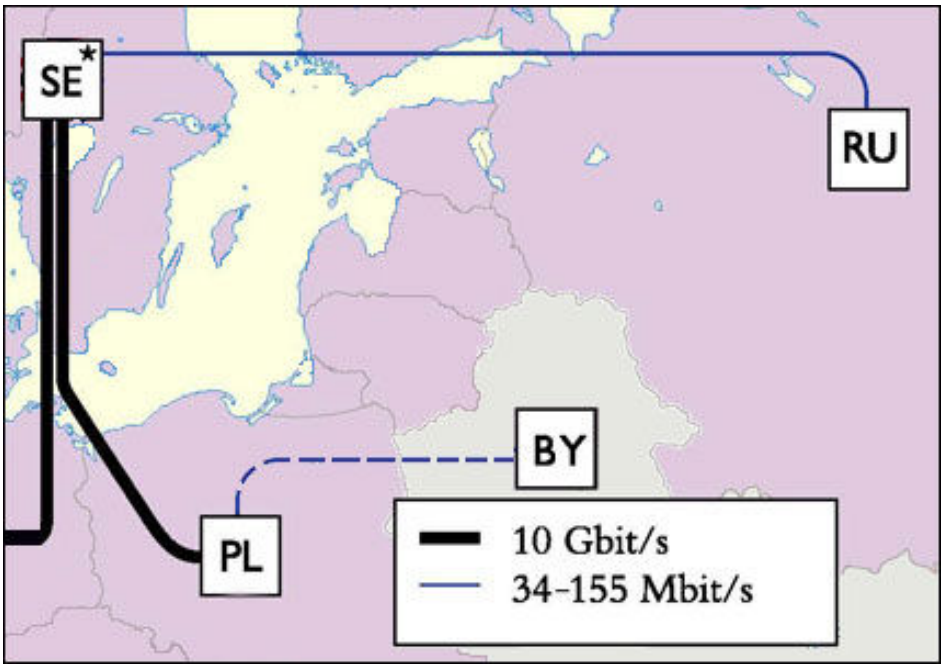

Рис. 6.2. Схема организации взаимодействия белорусских и российских участников программы «СКИФ»

При организации удаленного доступа к ресурсам суперкомпьютерного комплекса «СКИФ», с целью минимизации затрат, были максимально учтены возможности существующей телекоммуникационной инфраструктуры сети BASNET, поскольку суперкомпьютерный комплекс «СКИФ» уже подключен к сети BASNET на скорости 100 Мбит/сек, а сама сеть имеет развитую телекоммуникационную инфраструктуру как в пределах города Минска, так и в регионах.

BASNET основывается на восьми базовых сетевых узлах, шесть из которых связаны высокоскоростными оптоволоконными каналами общей длиной более 10 км, обеспечивающими передачу данных по сети со скоростью 100-1000 Мбит/сек. Несколько крупных организаций подключены радио-Ethernet и радиорелейными каналами на скоростях 2-11 Мбит/сек. В институтах, подключенных к BASNET, созданы современные локальные сети, объединяющие более 5000 компьютеров.

Региональные узлы сети BASNET функционируют в Гомеле, Бресте, Витебске и Могилеве.

Для подключения организаций были использованы три технологии – Ethernet с использованием ВОЛС, DSL и радио-Ethernet.

Наиболее крупные организации, расположенные на небольшом удалении от коммуникационных узлов BASNET, были подключены по технологии Ethernet с использованием волоконно-оптических линий связи (ВОЛС). В ВОЛС среда передачи – стеклянная, а информация пере-

носится модулированным световым потоком, генерируемым светодиодами или лазерами.

Достоинства ВОЛС:

– высокая пропускная способность;

– отсутствие электромагнитного излучения, что исключает утечку информации;

– высокая помехоустойчивость;

– большое расстояние передачи (не менее 2 км без повторителей);

– малый вес и диаметр кабеля;

– высокое электрическое сопротивление, обеспечивающее гальваническую развязку соединяемых устройств.

Менее крупные организации, расположенные на удалении до 10-12 км. были подключены по технологии DSL. Технологии DSL обеспечивают высокую скорость передачи данных. Различные варианты технологий DSL обеспечивают различную скорость передачи данных (до 8 Мбит/сек.).

И, наконец, подключение абонентов, находящихся на значительном удалении, осуществлялось с использованием радио-Ethernet технологии, которая позволяет передавать данные на скорости до 11 Мбит/сек.

Для подключения организаций г. Гомеля был создан международный цифровой канал связи 2 Мбит/сек. с использованием сети SDH, являющейся основной транспортной средой общей сети РО «Белтелеком», между Гомельским региональным узлом сети BASNET и коммуникационным узлом на ул. Захарова-55.

Общая схема организации удаленного доступа приведена на рис. 6.3.

Организации НАН Беларуси – Институт физики, Институт прикладной физики, ИТМО, Институт математики, УП «Автотракторостроение» и др. были подключены к сети BASNET на скоростях 10-100 Мбит/сек. вне рамок программы «СКИФ» и имеют скоростной доступ к ресурсам суперкомпьютерного комплекса «СКИФ» с использованием существующей телекоммуникационной инфраструктуры BASNET. Также вне рамок программы «СКИФ» были подключены к сети BASNET через Научно-информационную компьютерную сеть Республики Беларусь на скорости 100 Мбит/сек по Ethernet-технологии Белорусский государственный университет (БГУ) и Белорусский национальный технический университет (БНТУ). Для подключения других организаций в рамках программы «СКИФ» были предприняты соответствующие меры.

Белорусский государственный университет информатики и радиоэлектроники (БГУИР) подключен к Fddi-кольцу сети BASNET по волоконно-оптической линии связи радиальным сегментом по Ethernet тех-

нологии и имеет возможность использования ресурсов «СКИФ». Для обеспечения надежности подключения, на промежуточном узле сети BASNET в Институте порошковой металлургии был установлен дополнительный коммутатор.

УП «НИИЭВМ» было подключено к сети BASNET по волоконно-оптической линии связи на скорости 100 Мбит/сек с использованием технологии Fast Ethernet к высокопроизводительному коммутатору Cabletron ELS100-24TXG.

Республиканский научно-практический центр «Кардиология» был подключен к сети BASNET по волоконно-оптической линии связи на скорости 100 Мбит/сек с использованием технологии Fast Ethernet к высокопроизводительному коммутатору Cabletron ELS100-24TXG.

Комитет государственной безопасности Республики Беларусь был подключен к центральному узлу сети BASNET (ул. Академическая-1) по выделенной двухпроводной линии связи по технологии ADSL с использованием существующего ADSL-концентратора Zyxel IES1000 (со стороны сети BASNET) и модема-маршрутизатора Corecess 3113 со скоростью доступа равной 2 Мбит/сек.

Республиканский Гидрометеорологический центр был подключен к центральному узлу сети BASNET (ул. Академическая-1) по волоконно-оптической линии связи на скорости 100 Мбит/сек с использованием технологии Fast Ethernet.

Минский тракторный завод был подключен к сети BASNET по волоконно-оптической линии связи на скорости 100 Мбит/сек с использованием технологии Fast Ethernet. Данный узел связан с центральным узлом сети BASNET на скорости 1 Гбит/сек.

Производственное объединение «Горизонт» (ПО «Горизонт»). Основное проектное предприятие ПО «Горизонт» – НИИ цифрового телевидения было подключено к сети BASNET по волоконно-оптической линии связи на скорости 100 Мбит/сек с использованием технологии Fast Ethernet. Данный узел связан с центральным узлом сети BASNET на скорости 1 Гбит/сек.

Военная академия. Военная Академия находится за кольцевой дорогой г. Минска, в районе, в котором нет телекоммуникаций сети BASNET и на значительном удалении от возможной точки подключения к сети BASNET. Поэтому для ее подключения была использована технология беспроводной связи RadioEthernet по протоколу 802.11b. Стандарт IEEE 802.11b позволяет передавать данные на скорости до 11 Мбит/с.

Схема радиоканала приведена на рис. 6.4.

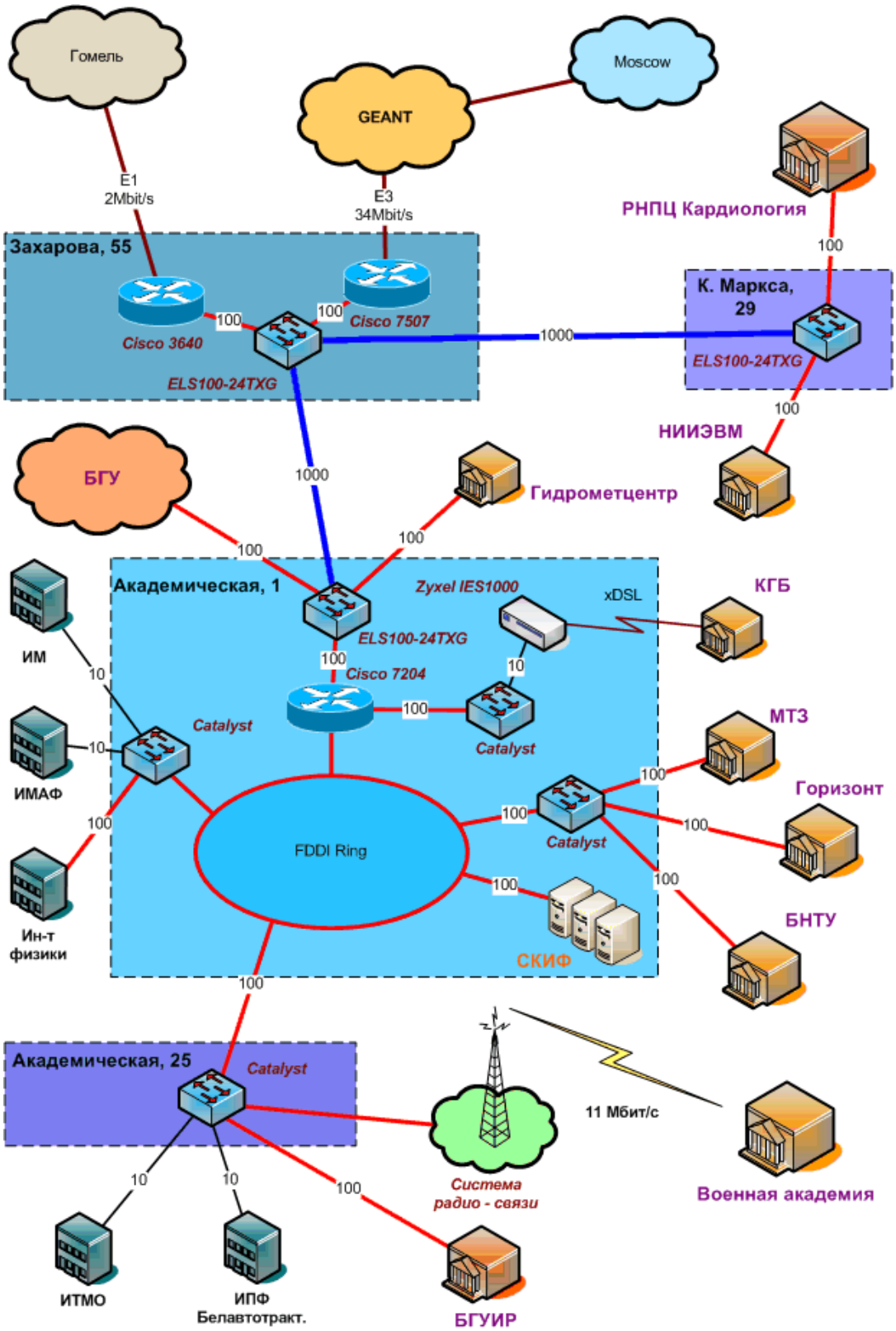

Рис. 6.3. Общая схема организации удаленного доступа

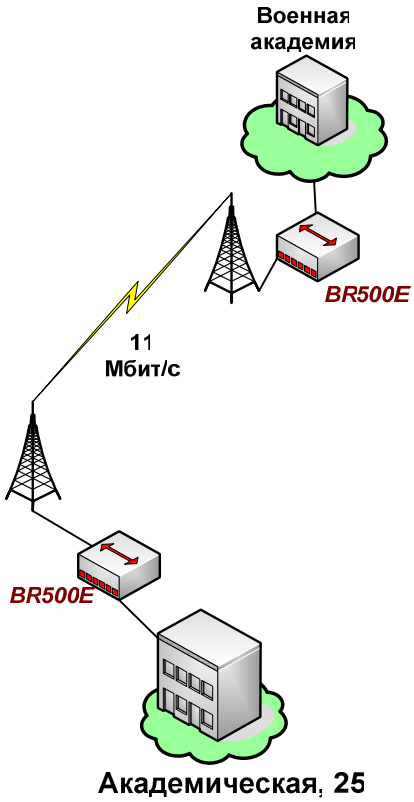

Рис. 6.4. Схема радиоканала

В рамках программы «СКИФ» было также осуществлено подключение к ресурсам суперкомпьютерного комплекса «СКИФ» гомельских организаций.

Подключение организаций к ресурсам «СКИФ» через Гомельский региональный узел осуществлялось по арендованным двухпроводным линиям связи с использованием технологии SHDSL. Скорость подключения составила 1-2 Мбит/сек (в зависимости от длины и качества арендованной телефонной линии). Для этих целей в каждой из подключаемых организаций были установлены модемы-маршрутизаторы SHDSL Corecess 3311.

Схема подключения Гомельских организаций к ресурсам «СКИФ». приведена на рис. 6.5.

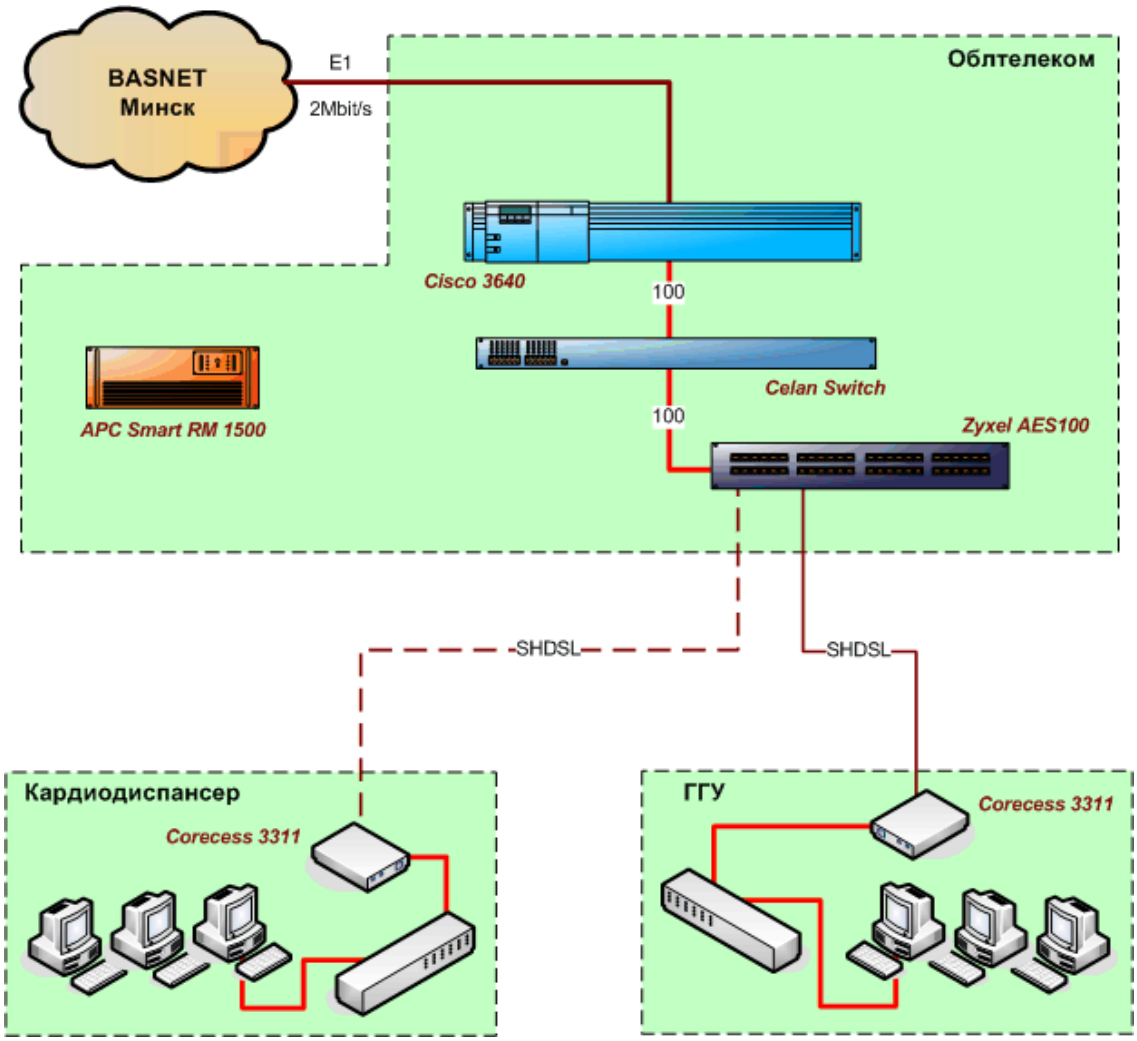

Рис. 6.5. Схема подключения гомельских организаций к ресурсам «СКИФ»

6.3. Формирование новых программ Союзного государства по развитию направления «СКИФ»

Востребованность работ по развитию в странах Союзного государства направления суперкомпьютерных технологий определяется отставанием этих стран от ведущих мировых держав в развитии и применении новейших наукоёмких информационных технологий, нацеленных на решение сложных задач машиностроения, биотехнологии, геологоразведки, контроля окружающей среды, транспорта и связи, государственных, коммерческих, военных и других приложений. Применение суперкомпьютерных систем и наукоёмких программных продуктов актуально также и в таких приложениях, как банки данных, информационно-аналитические системы, ситуационные центры управления, системы управления в реальном масштабе времени, боевые информационно-управляющие системы и др.

Для эффективного решения задач на базе наукоемких технологий для различных областей применения необходимо использование принципиально новых суперкомпьютерных технологий обработки данных в широком диапазоне производительности – от сотен миллиардов операций в секунду до вычислительных систем с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду). Закупка импортных суперкомпьютерных конфигураций связана с систематическими политическими ограничениями со стороны потенциальных экспортеров и с утечкой значительных финансовых средств за рубеж. Проблема усугубляется также и крайней сложностью освоения современных наукоёмких программных систем мирового уровня. Необходима инфраструктура, предоставляющая комплексные услуги по внедрению основных компонентов наукоёмких информационных технологий, в т. ч. предоставление доступа к суперкомпьютерным вычислительным ресурсам, консультации по выбору и применению технологий, сопровождение компьютерных средств и программного обеспечения, обучение и др. Решению этих проблем может способствовать создание в Республике Беларусь и Российской Федерации суперкомпьютерных центров для развития и внедрения наукоемких информационных технологий. Наличие таких центров обеспечит возможность предоставления услуг для решения наукоёмких задач, возникающих в промышленности и в других областях народного хозяйства, требующих компьютерных и информационных ресурсов, владение которыми недоступно или экономически нецелесообразно для отдельных организаций.

В рамках программы «СКИФ» создан существенный научно-технический задел, позволяющий создавать суперкомпьютерные конфигурации в широком диапазоне производительности, вплоть до терафлопового (триллионы операций в секунду). Создание суперкомпьютеров «СКИФ К-500» и «СКИФ К-1000» позволяет утверждать, что научный и технический уровень принятых решений соответствует современному мировому уровню и имевшееся отставание в этой части, фактически, ликвидировано.

Результаты комплексной реализации программы «СКИФ» являются существенным научно-техническим и организационным заделом для дальнейшего развития суперкомпьютерного направления, в том числе, для формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ» (НП «СКИФ»).

6.3.1. Стратегия формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ»

Стратегия формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ» базируется на

трех концептуальных уровнях (этапах) проектирования – технологическом, системном и прикладном (техническая реализация проекта). Отмеченные концептуальные уровни являются естественными компонентами процесса проектирования любой системы, тем более, сложнейшей научно-технической программы.

На технологическом (предварительном или организационном) уровне анализируются существующие организационные и научно-технические предпосылки формирования новой программы (сложившиеся коллективы предприятий и связи между ними, научно-технический задел и перспективы его развития и т. д.), адекватность соответствия предпосылок стратегическим целям и задачам формируемой программы, выявляются узкие места и намечаются перспективы по их устранению.

Отправной точкой анализа организационных и научно-технических предпосылок формирования НП «СКИФ» является совокупность следующих реперных моментов:

– программа Союзного государства «СКИФ» – базовая для НП «СКИФ»;

– ход реализации базовой программы «СКИФ»;

– пилотные демонстрационные и прикладные пакеты «СКИФ»;

– перспективные области применения суперкомпьютеров «СКИФ»;

– перспективы внедрения и развития суперкомпьютерных конфигураций «СКИФ»;

– перспективы реализации и формирования программ Союзного государства, связанных с использованием суперкомпьютерных технологий;

– основные цели и задачи программы НП «СКИФ»;

– главная цель программы НП «СКИФ»;

– перспективные области применения программы НП «СКИФ»;

– стратегическая область применения программы НП «СКИФ»;

– коллектив предприятий-исполнителей программы НП «СКИФ»;

– предложения по участию в программе «СКИФ»;

– предложения по участию в программе НП «СКИФ»;

– предложения головных исполнителей по срокам формирования и реализации программы НП «СКИФ».

На технологическом уровне становится актуальной проблема анализа и оптимизации взаимодействия потенциальных предприятий-соисполнителей новой программы, а также анализа многообразия конкретных форм проявления деятельности этих предприятий. Следует учитывать также, что специфику факторов, влияющих на успешное формирование новой программы, невозможно оценить на основании

анализа только ее элементов (предприятия-исполнители, программные мероприятия, их содержание и сроки реализации, объемы затрат и т. п.) и связей между ними. Возникает противоречие частного и общего. С одной стороны, сложную систему (программу) необходимо декомпозировать (разделить на функциональные компоненты, например, проекты), с другой стороны, найти критерии выделения параметров системы, характеризирующие ее (программу) как единое целое.

Проблемы подобного уровня преодолеваются на системном уровне формирования программы. На базе представления формируемой программы на технологическом уровне на системном уровне (этап подготовки предложений по формированию новой программы) фактически определяется концепция формирования новой программы, позволяющая перейти к техническому проектированию (т.е. к формированию) новой программы с учетом предъявляемых требований и конкретных ограничений (например, требований «Порядка разработки и реализации программ Союзного государства» и др.).

В общем виде на технологическом концептуальном уровне информационную среду для формирования новой программы можно представить в виде некоторой технологической структуры (модели) кластерного типа, состоящей из множества информационных узлов и связей между ними. Каждый из таких информационных узлов отображает на модели определенный информационный аспект (или совокупность аспектов) в соответствии с той или иной спецификой отображаемого конкретным узлом реперного момента, используемого для анализа предпосылок формирования новой программы. Укрупненная методика формирования новой программы Союзного государства НП «СКИФ» приведена на рис. 6.6.

Для форсирования начала работ по формированию НП «СКИФ» ключевое значение приобретают работы на технологическом и системном концептуальных уровнях, на которых, по сути, используется метод декомпозиции стратегических целей. Этапность реализации этого метода может быть представлена следующим образом:

1 этап. Анализ и формулировка тенденций и перспективных направлений развития суперкомпьютерной отрасли.

2 этап. Формулировка стратегических целей программы НП «СКИФ» в рамках перспективных направлений развития суперкомпьютерной отрасли (каждому направлению может соответствовать несколько стратегических целей).

3 этап. Формулировка задач перспективных разделов программы НП «СКИФ» в соответствии с ее стратегическими целями.

4 этап. Составление предварительного укрупненного перечня проектов по разделам программы, формулировка целей для каждого проекта (группы проектов).

5 этап. Отработка концепции формирования программы и предварительного перечня проектов по разделам программы, уточнение целей для каждого проекта.

Последний этап позволяет перейти на уровень конкретных проектов (программных мероприятий), определить показатели для оценки их эффективности и, следуя в обратном направлении по указанным шагам, прийти к формированию (композиции) новой программы Союзного государства на основе конкурентоспособных и экономически эффективных проектов.

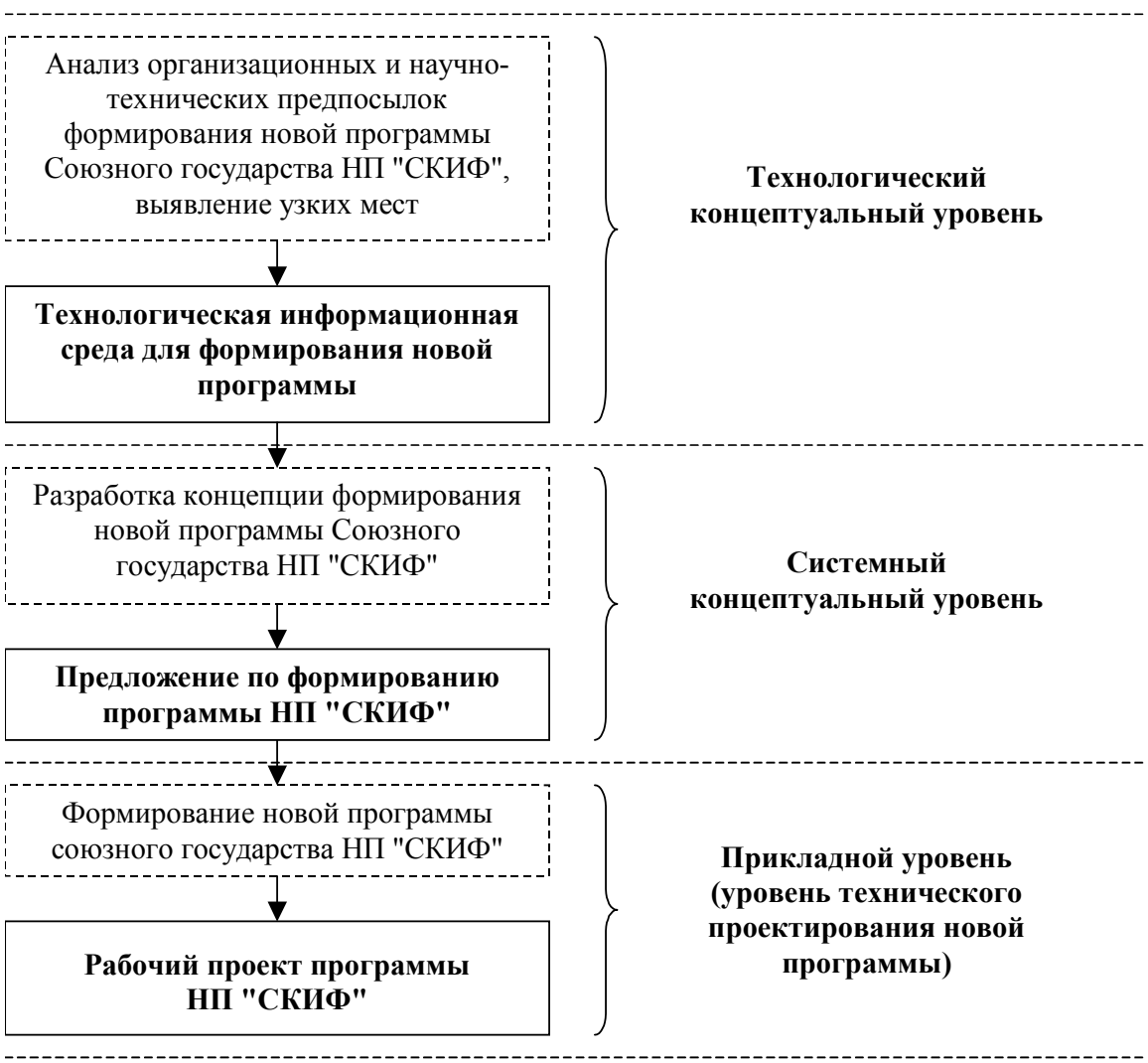

Рис. 6.6. Методика формирования НП «СКИФ»

С учетом созданного в рамках программы «СКИФ» научно-технического задела главной целью формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ» может быть освоение и адаптация передовых зарубежных и отечественных наукоемких технологий на отечественных суперкомпьютерных конфигурациях семейства «СКИФ», внедрение этих технологий в основных отраслях гражданской и военной промышленности и социально-экономической сфере Союзного государства, оптимизация отечественных суперкомпьютерных конфигураций семейства «СКИФ» с учетом требований современных наукоемких технологий и специфики их приложений.

Создаваемые суперкомпьютеры семейства «СКИФ» являются базой для реализации формируемой Министерством промышленности, науки и технологий Российской Федерации и Национальной академии наук Беларуси научно-технической программы Союзного государства «Развитие и внедрение в государствах-участниках Союзного государства наукоёмких компьютерных технологий на базе мультипроцессорных вычислительных систем» (шифр **«Триада»**).

На базе результатов программы «СКИФ» могут быть сформированы и другие (наряду с программой «Триада») направления – защита информации, специальные условия эксплуатации и т.п. Однако, все эти направления перспективны лишь при развитии стратегического направления – оптимизации отечественных суперкомпьютерных конфигураций семейства «СКИФ» с учетом специфики требований современных наукоемких технологий и других суперкомпьютерных приложений. Развитие этого стратегического направления предполагается в рамках новой программы Союзного государства (базовой для развития суперкомпьютерного направления «СКИФ») – «Разработка и использование программно-аппаратных средств ГРИД – технологий и перспективных высокопроизводительных (суперкомпьютерных) вычислительных систем семейства 'СКИФ'» (шифр **«СКИФ-ГРИД»**).

6.3.2. Программа Союзного государства «СКИФ-ГРИД»

Государственными заказчиками программы «СКИФ-ГРИД» являются: Национальная академия наук Беларуси и Федеральное агентство по науке и инновациям Российской Федерации.

Головные исполнители программы:

– от Республики Беларусь – ОИПИ НАН Беларуси.

– от Российской Федерации – ИПС РАН.

Магистральным путём развития современных суперкомпьютерных технологий является построение распределённых вычислительных систем с массовым параллелизмом. Как в научной среде, так и для нужд

промышленности существует необходимость создания высокопроизводительной высоконадёжной среды для создания мобильных и масштабируемых приложений. Поэтому в качестве стратегической линии формирования программы «СКИФ-ГРИД» планируется развитие суперкомпьютерного направления «СКИФ» на базе технологий Grid Computing.

Основные идеи ГРИД-технологий заимствованы из практики работы национальных электрических сетей (power grid), которые включают в себя распределенные по огромным территориям потребителей электроэнергии, линии передачи и генерирующие мощности различных электростанций. Потребитель электроэнергии обслуживается этой гигантской системой и ему не важно знать, какая электростанция в настоящее время питает его оборудование, какого типа эта электростанция (тепловая, атомная, гидростанция или другая), какие линии передачи электроэнергии задействованы для этого. Электрическая сеть поддерживает разные аспекты такого обслуживания: эффективное использование в национальном масштабе имеющихся генерирующих мощностей, переброса избытка мощности из одного региона в другой, использование резервных линий для нейтрализации последствий аварий на линиях передачи электроэнергии или на электростанциях и т.п.

Сегодня цели и планы разработки ГРИД-технологий и ГРИД-систем не менее важны и грандиозны, чем цели и планы национальных электрических систем, создаваемых в первой половине двадцатого века. Необходимо объединить в единую систему различные по технической реализации и различные по типам компьютерные ресурсы (вычислительные ресурсы, ресурсы хранения и передачи информации) и донести совокупный ресурс до потребителя. Потребитель должен получать услуги от системы в целом, ему не важно знать, где и какая установка хранит или обрабатывает его информацию, какого типа данная установка, какие линии передачи информации при этом задействованы и т.п. Также как и в электрических сетях, высокопроизводительная вычислительная ГРИД-система должна обеспечивать эффективное использование всей совокупности ресурсов (вычислительных и ресурсов хранения информации), нейтрализацию последствий аварий на линиях передачи, в устройствах хранения или обработки информации. Тем самым, создание подобных систем должно кардинально улучшить эффективность использования совокупных компьютерных ресурсов страны.

Таким образом, ГРИД – это технология создания эффективных территориально-распределенных гетерогенных (объединяющих компьютеры с самыми различными аппаратными и программными системами) сетей. Основная задача ГРИД – реализация гибкого, защищенного, скоординированного вычислительного пространства для совместного использования ресурсов между динамически меняющимися сообществами

пользователей. Данная технология призвана осуществлять хранение информации и высокопроизводительные параллельные вычисления на сети глобально распределенных вычислительных средств и других ресурсов: суперкомпьютеров, отдельных серверов, мэйнфреймов, систем хранения и поддержки сетевых ресурсов, баз данных (с разной реализацией этих баз).

На основе ГРИД-технологий можно создавать супермощные информационно-вычислительные среды, обладающие также уникально низким соотношением «эксплуатационные расходы/производительность», за счет повышения коэффициента использования всей совокупности ресурсов, включенных в ГРИД-сеть.

Предпосылкой и базисом для развития в программе «СКИФ-ГРИД» ГРИД-технологий являются созданные в рамках программы «СКИФ» кластерные и мета-кластерные конфигурации широкого диапазона производительности, а также новые технологии динамического распараллеливания вычислений, мониторинга и управления кластерными и мета-кластерными конфигурациями.

В настоящее время практически в каждой развитой стране развернуты национальные ГРИД-проекты, имеющие целью создание соответствующей инфраструктуры и развитие технологий, обеспечивающих удаленный доступ к разнообразным вычислительным ресурсам независимо от места расположения потребителя.

Главной целью программы «СКИФ-ГРИД» является освоение и адаптация передовых наукоемких технологий на перспективных суперкомпьютерных платформах, оптимизация суперкомпьютерных конфигураций семейства «СКИФ», ориентированных на построение на их основе ГРИД-компьютерных сетей, т.е. создание новой технологической базы для обеспечения динамики роста экономического и оборонного потенциала России и Беларуси, укреплении безопасности Союзного государства.

Проведение целевой научно-технической программы «СКИФ-ГРИД» в области высокопроизводительных вычислений, безусловно, станет также катализатором роста в отрасли наукоёмких программных продуктов и услуг для суперкомпьютерных приложений, в том числе и условием выхода с ними на мировой рынок.

Высокопроизводительные вычисления являются одной из ключевых технологий, необходимых для развития наукоемких технологий. Помимо создания собственно высокопроизводительных вычислительных средств (суперкомпьютеров) необходимо также развитие целого ряда смежных областей науки и техники, связанных с решением конкретных научно-технических проблем, таких как ядерная и водородная энергетика, генная инженерия, перспективное материаловедение, нанотехно-

логии, гидрометеорология, биоинформатика, молекулярное моделирование, геоинформатика и другие.

Концептуально Программа представляется четырьмя взаимоувязанными стратегическими направлениями.

Развитие высокопроизводительных вычислений на основе ГРИД-технологий, ориентированных на создание гетерогенных, территориально-распределённых вычислительных комплексов.

Создание перспективных суперкомпьютерных платформ и конфигураций следующего поколения, ориентированных на использовании в ГРИД-системах и основанных на новых технологиях взаимодействия узлов (интерконнект) и управления узлами и кластерами, на использовании гибридных узлов, различного набора реконфигурируемых и специализированных вычислителей.

Обеспечение информационной безопасности создаваемых распределенных ГРИД-вычислительных сред с учетом взаимодействия их компонент по открытым (общедоступным) коммуникационным сетям (идентификация и аутентификация, управление доступом, мониторинг и аудит, контроль целостности, защищенные режимы передачи данных по открытым каналам связи), а также создание средств защиты суперкомпьютерных установок от побочных электромагнитных излучений и наводок, обеспечение критических узлов технологическими решениями для поддержки требований безопасности на аппаратном уровне.

Научно-исследовательские работы по перспективным областям применения создаваемых вычислительных установок, в том числе разработка пилотных образцов прикладных систем

Заложенное в программе «СКИФ-ГРИД» развитие распределенных высокопроизводительных вычислений на базе ГРИД-технологий позволит более эффективно использовать вычислительные мощности и ресурсы хранения данных, а также решать социальные и оборонно-технические задачи стратегической значимости и высокой информационной сложности, осуществление которых до этого времени представлялось невозможным.

Работа по программе рассчитана на четыре года и будет выполняться в два этапа:

Этап 1 (первый и второй годы выполнения Программы):

– разработка концепции ГРИД-инструментария, разработка комплектов конструкторской и программной документации (КД и ПД) для создания конфигураций семейства «СКИФ» Ряда 3 и метакластерных систем на их основе;

– создание образцов вычислительных систем конфигураций «СКИФ» Ряда 3 и проведение их испытаний, доработка КД и ПД по результатам испытаний.

Этап 2 (третий и четвертый годы выполнения Программы):

– разработка комплектов конструкторской и программной документации (КД и ПД) для создания конфигураций семейства «СКИФ» Ряда 4 и территориально – распределенных гетерогенных ГРИД-сетей на их основе;

– создание комплекта программного обеспечения ГРИД-инструментария, образцов вычислительных систем конфигураций «СКИФ» Ряда 4 и опытного участка ГРИД-сети, включающего суперкомпьютеры семейства «СКИФ» Ряда 3 и Ряда 4 и отдельные серверы и рабочие станции, установленные в России и Беларуси, проведение их испытаний, доработка КД и ПД по результатам испытаний.

Предложения Национальная академия наук Беларуси и Федерального агентства по науке и инновациям Российской Федерации с белорусской стороны согласованы Минэкономики, Минфином и Министерством иностранных дел Республики Беларусь, с российской стороны Минэкономразвитием и по состоянию на 01.09.2005 находятся на согласовании в Минфине России.

6.3.3. Программа Союзного государства «Триада»

Предложение Министерства образования и науки Российской Федерации и Национальной академии наук Беларуси о разработке программы одобрено Советом Министров Республики Беларусь (Постановление № 680 от 22 мая 2003 г.) и правительством Российской Федерации (Распоряжение от 13.07.2004 № 945-р).

Государственный заказчик-координатор программы: Министерство образования и науки Российской Федерации.

Государственный заказчик программы от Российской Федерации: Федеральное агентство по науке и инновациям.

Государственный заказчик программы от Республики Беларусь: Национальная академия наук Беларуси.

Головной исполнитель программы от Российской Федерации: «Научно-исследовательский центр электронно-вычислительной техники» (ОАО НИЦЭВТ). Научный руководитель программы – В.В. Митрофанов, исполнительный директор программы – А.И. Слуцкин.

Головной исполнитель программы от Республики Беларусь: Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ НАН Беларуси). Научный руководитель программы – В.И. Махнач, исполнительный директор программы – Г.М. Левин.

Проект научно-технической программы Союзного государства «Триада» одобрен Постановлением Совета Министров Республики Беларусь от 4 марта 2005 г. № 246. Предполагается, что в 2005 году проект

программы будет одобрен Правительством Российской Федерации и Советом Министров Союзного государства.

Главная цель программы – разработка и внедрение новых наукоемких технологий в промышленности и социально-экономической сфере, адаптация и освоение передовых зарубежных наукоемких технологий на производимых в государствах-участниках высокопроизводительных мультипроцессорных (суперкомпьютерных) вычислительных системах (ВМВС), внедрение этих технологий в основных отраслях промышленности государств-участников, оптимизация создаваемых в государствах-участниках мультипроцессорных вычислительных систем с учетом требований современных наукоемких технологий.

Данная программа является комплексной, объединяющей приоритетные проекты, нацеленные на решение ключевых проблем использования ВМВС в наиболее важных областях приложений, а также проекты развития ВМВС для удовлетворения требований приложений.

Развитие и внедрение наукоемких технологий на базе ВМВС позволит решить проблемы поднятия уровня промышленности государств-участников. Это сделает реальным выход на мировой рынок интеллектуальных информационных продуктов и услуг, что возможно благодаря сохранившимся еще в государствах-участниках мощным школам по фундаментальным наукам (математика, физика, механика и др.), способным решать уникальные научно-технические задачи.

Применение наукоемких технологий инженерного проектирования крайне актуально. Государства-участники Союзного государства в настоящее время отстают от ведущих мировых держав в применении современных наукоёмких информационных технологий компьютерного проектирования (CAD-технологии) и инженерного анализа (CAE-технологии), планирования и управления производством на основе данных об изделии (CAM и PDM-технологии) и других технологий, объединяемых в комплекс CALS-технологий.

С проблематикой расчетов для машиностроительных отраслей промышленности и оптимизацией проектных решений связаны направления исследований:

– организация параллельных вычислительных процессов в инженерных расчетах, создание высокоточных эффективных методов и на их основе пакетов наукоемких технологий, ориентированных на суперкомпьютерные вычислительные системы;

– проблемы адекватности, точности и масштабируемости, возникающие при решении задач из области CAE - приложений, аэрогидродинамики и электромагнитных расчетов на суперкомпьютерных вычислительных системах с использованием пакетов мирового уровня;

– разработка научно–методических основ применения наукоемких информационных CAE - технологий на базе кластерных высокопроизводительных мультипроцессорных вычислительных систем.

По этим направлениям предполагается выполнение следующих основных работ:

Название работы	**Исполнители от Республики Беларусь**	**Потребители в Республике Беларусь**
Разработка научно-методических основ компьютерного инженерного анализа тракторных конструкций на базе кластерных мультипроцессорных вычислительных систем с использованием отечественных и мировых CAE-систем	ОИПИ НАН Беларуси ПО «МТЗ»	РУП «МАЗ», РУП «Белаз», «Гомсельмаш»
Исследование проблем адекватности, точности и масштабируемости, возникающих при решении задач из области CAE-приложений, аэрогидродинамики и электромагнитных расчетов на кластерных высокопроизводительных мультипроцессорных вычислительных системах с использованием пакетов мирового уровня	ОИПИ НАН Беларуси ОАО «ГОРИЗОНТ»	РУП «Витязь», г. Витебск, ОАО «Электроаппаратура», г. Гомель
Разработка технологий, подходов и методов для выполнения компьютерного моделирования и проведения исследований сложных механических процессов и явлений, имеющих место в узлах, механизмах, конструкциях существующих и проектируемых машин и механизмов на базе кластерных высокопроизводительных мультипроцессорных вычислительных систем с использованием пакетов мирового уровня. Разработка и создание специализированных корпоративных систем на основе разработанных технологий	ОИПИ НАН Беларуси БГУ Институт механики машин НАН Беларуси	МАЗ, МТЗ, БелАЗ

Разработка научно-методических и практических основ применения пакета LS-DYNA 3D на базе суперкомпьютерной системы «СКИФ» при решении задач пассивной безопасности автотракторной техники	ОИПИ НАН Беларуси ИМИНМАШ НАН Беларуси РУП «МАЗ»	РУП «МАЗ»
Разработка на базе суперкомпьютерной системы «СКИФ» типового программно-аппаратного комплекса (ПАК) и базовых методик применения в автомобильной промышленности	ИМИНМАШ НАН Беларуси РУП «МАЗ»	РУП «МАЗ»
Разработка систем управления пространственным развитием процессов пластического формо- и структурообразования, включая наноструктуры, с целью формирования деформационных и механических свойств материалов с применением наукоемких информационных Cae-технологий на базе кластерных высокопроизводительных мультипроцессорных вычислительных систем	ОИПИ НАН Беларуси ФТИ НАН Беларуси	ПО «МТЗ» РУП «ММЗ» ООО «Техстроймаш»
Разработка эффективных методов и алгоритмов параллельной обработки и идентификации изображений фотошаблонов и топологических слоев интегральных схем, а также программных средств восстановления и контроля топологии, ориентированных на кластерные высокопроизводительные мультипроцессорные вычислительные системы	ОИПИ НАН Беларуси КБТЭМ-ОМО концерна «Планар»	ГНПК ТМ «Планар», НПО «Интеграл», ГП «Горизонт», МПО «ВТ им. Орджоникидзе»
Математическое моделирование технологических процессов производства субмикронных элементов сверхбольших интегральных схем	Институт математики НАН Беларуси	НПО «Интеграл» ПО «Горизонт» АО «Минский часовой завод» УП «Минский НИИ радиоматериалов»
Исследование и разработка технологии создания информационно-аналитических систем на базе архитектуры высокопроизводительных мультипроцессорных вычислительных систем для органов государственного управления и промышленных предприятий	ЦНИИТУ	Министерство промышленности

Разработка и внедрение на предприятиях автотракторного сельскохозяйственного и строительно-дорожного машиностроения методики расчета напряженно-деформированного состояния, усталости и ресурса несущих конструкций остова и ходовой системы энергонасыщенных тягово- транспортных машин методами конечных и граничных элементов, создание и внедрение на указанных предприятиях программно-технических комплексов для их реализации	ОИПИ НАН Беларуси РУП МТЗ НИРУП «Белавтотракторо-строение»	РУП МТЗ
Разработка и внедрить на машиностроительных предприятиях по производству тягово-транспортных машин методики расчета и оптимизации параметров многозвенных механических трансмиссий с учетом многорежимного вероятностного характера нагружения и требуемых ресурса и надежности	ОИПИ НАН Беларуси РУП МТЗ НИРУП «Белавтотракторо-строение»	РУП МТЗ

Ожидаемые результаты работ:

1) Разработка и накопление представительного пакета оценочных программ для исследования корректности вычислений (адекватность и точность), масштабируемости производительности при увеличении количества используемых процессоров ВМВС.

2) Проведение исследований пакетов, подготовка научно - технических отчетов с результатами исследований и обобщениями опыта использования.

3) Разработка параллельных версий уже зарекомендовавших себя отечественных пакетов инженерных расчетов для кластерных ВМВС.

4) Обеспечение технической поддержки пакетов, организация обучения, предоставление аренды пакетов и консультаций, решение задач по заказам.

5) Анализ границ применимости пакетов инженерных расчетов, используемых в них алгоритмам, ожидаемых характеристик при решении различных классов задач.

6) Анализ наблюдаемых недостатков зарубежных пакетов инженерных расчетов. Создание новых блоков для зарубежных инженерных пакетов, а также рекомендации и оценки целесообразности создания новых пакетов инженерных расчетов при последующем развитии программы.

Работы, проведенные в рамках программы Союзного государства «СКИФ», создали научно-технические предпосылки для удовлетворения потребности государств-участников в суперкомпьютерных ресурсах. Однако, ориентация на применение различных наукоемких технологий требует проведения дополнительных исследований и разработок. В частности, высокопроизводительные вычислительные системы для промышленных приложений инженерного проектирования и виртуального эксперимента, задачи разведки природных ископаемых и ряд других важных задач должны отвечать специфическим для наукоемких технологий требованиям, например: по организации мощной подсистемы дисковой памяти с большим расслоением; по организации визуализации результатов расчетов и т. п.

Заключение

Суперкомпьютерные технологии в развитых странах применяются во многих отраслях промышленности: в машиностроении (в том числе оборонном); в химической, фармацевтической, аэрокосмической промышленности: в добыче полезных ископаемых, сельском хозяйстве, логистике, транспорте, телекоммуникациях, и многих других. Не менее важны высокопроизводительные вычислительные системы и при решении многих практически важных научных задач, например: прогнозировании погоды, создании новых материалов, поиске новых лекарственных средств, оптимизации управления сложными системами. Помимо непосредственного эффекта от использования высокопроизводительных вычислительных установок при создании суперкомпьютеров часто используются новые технологии и методики, которые затем находят самое широкое применение в других отраслях науки и техники.

Задачи поддержания конкурентоспособности на мировом рынке ставят перед промышленностью и наукой Беларуси и России проблему адаптации к происходящим изменениям, своевременного использования передового мирового опыта, вновь открывающихся возможностей по повышению конкурентоспособности предлагаемых товаров и услуг, проведению новых исследований.

Отставание в использовании суперкомпьютерных средств создаёт также проблему поддержания престижа государств-участников Союзного государства, как одних из ведущих в мире государств в области высоких информационных технологий. На 01.01.2005 года в списке 500 наиболее высокопроизводительных вычислительных систем (Top-500) находилось лишь по одной установке в России и в Беларуси, общей мощностью 0,6% от суммарной мощности суперкомпьютеров этого списка. При этом, в Мексике находится 4 установки (0,8%), в Саудовской Аравии – 9 (1,8%), в Австралии – 6 (1,2%), в Новой Зеландии – 4(0,8%). Следует ожидать, что если не будет предпринято дополнительных мер, в ближайшие год-два в Союзном государстве не останется ни одной установки, входящей в этот список.

Экономический спад в последнее десятилетие XX века и недостаточное финансирование фундаментальных исследований и новых прикладных научно-технических разработок в области разработки и производства элементной базы и вычислительной техники предопределили вынужденный болезненный переход к импортной элементной базе и вычислительным средствам и, как следствие, зависимость от иностранных производителей. При этом образуется «замкнутый круг», когда следст-

вие усиливает причину. Так, например, для создания современной машиностроительной продукции предприятиям необходимо использовать суперкомпьютеры для инженерных расчётов. Отсутствие возможности проводить такие расчёты ухудшает рыночные перспективы новой продукции предприятий, что, в свою очередь, отрицательно сказывается на финансовом положении и ещё более затрудняет приобретение и использование суперкомпьютерной техники. Другой пример – отсутствие высокопроизводительных вычислительных средств не позволяет проводить перспективные исследования во многих областях науки, например, гидрометеорологии. А отсутствие новых разработок, например, более точных моделей для прогноза погоды, не позволяет эффективно использовать новые высокопроизводительные вычислительные средства.

Как уже отмечалось, указанные проблемы начали решаться в программе Союзного государства «СКИФ», завершившейся в 2004 году. Одним из важнейших результатов программы «СКИФ» является формирование совместного (Беларуси и России) высокопрофессионального коллектива с налаженным взаимопониманием и тесными кооперационными связями, с большим научным заделом и опытом практической работы в суперкомпьютерной отрасли, включая опыт разработки и создания высокопроизводительных вычислительных систем с параллельной архитектурой и программных комплексов на их основе. Сформированный в рамках программы «СКИФ» коллектив способен решать в суперкомпьютерной отрасли технические задачи любой сложности. Создание суперкомпьютеров «СКИФ К-500» и «СКИФ К-1000» позволяет утверждать, что сегодняшний научный и технологический уровень созданного коллектива исполнителей заведомо соответствует мировому уровню и имевшееся отставание в этой области на сегодняшний день в значительной степени ликвидировано.

Однако, отрасль информационных технологий развивается стремительными темпами и «вещественные» результаты программы «СКИФ» (конструкторская и программная документация, образцы семейства «СКИФ») будут подвержены моральному старению, и если не будет обеспечено дальнейшее эффективное использование сформированной команды исполнителей, через год-другой отставание в суперкомпьютерных технологиях, с таким трудом ликвидированное сегодня, будет снова иметь место. Поэтому, крайне необходимо, создать условия для сохранения одного из основных результатов программы «СКИФ», условий, в которых коллектив исполнителей мог бы продолжить работу, направленную на удержание достигнутых позиций в части суперкомпьютерных технологий. Любое иное развитие событий сведет на нет усилия и средства, вложенные в программу «СКИФ». Это обстоятельство (наряду с созданным научно-техническим заделом) является важнейшим обосно-

ванием необходимости формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ».

С учетом созданного в рамках программы «СКИФ» научно-технического задела главной целью формирования новых программ Союзного государства по развитию суперкомпьютерного направления «СКИФ» может быть освоение и адаптация передовых зарубежных и отечественных наукоемких технологий на отечественных суперкомпьютерных конфигурациях семейства «СКИФ», внедрение этих технологий в основных отраслях гражданской и военной промышленности и социально-экономической сфере Союзного государства, оптимизация отечественных суперкомпьютерных конфигураций семейства «СКИФ».

Суперкомпьютеры семейства «СКИФ» являются базой для реализации формируемой Национальной академией наук Беларуси и Министерством образования и науки Российской Федерации научно-технической программы Союзного государства «Развитие и внедрение в государствах-участниках Союзного государства наукоёмких компьютерных технологий на базе мультипроцессорных вычислительных систем» (шифр «ТРИАДА»).

Данная программа является комплексной, объединяющей приоритетные проекты, нацеленные на решение ключевых проблем использования суперкомпьютерных технологий в наиболее важных областях приложений.

Развитие и внедрение наукоемких технологий на базе суперкомпьютеров позволит решить проблемы поднятия уровня промышленности в Республике Беларусь и Российской Федерации, сделает реальным выход на мировой рынок интеллектуальных информационных продуктов и услуг, что возможно благодаря сохранившимся еще мощным школам по фундаментальным наукам (математика, физика, механика и др.), способным решать уникальные научно-технические задачи.

На базе результатов программы «СКИФ» могут быть сформированы и другие (наряду с программой «ТРИАДА») направления – защита информации, специальные условия эксплуатации и т.п. Однако все эти направления перспективны лишь при развитии стратегического направления – оптимизации отечественных суперкомпьютерных конфигураций семейства «СКИФ» с учетом специфики требований современных наукоемких технологий и других суперкомпьютерных приложений. Развитие этого стратегического направления предусмотрено в предложении Национальной академии наук Беларуси и Федерального агентства по науке и инновациям Российской Федерации о разработке научно-технической программы Союзного государства «СКИФ-ГРИД».

Актуальность разработки программы «СКИФ-ГРИД» определяется необходимостью своевременного освоения новых технологий высоко-

производительных вычислений, их адаптации к технологическому и организационному укладу государств-участников, решения вопросов информационной безопасности создаваемых средств для специальных приложений. Сегодня цели и планы разработки ГРИД-технологий и ГРИД-систем не менее важны и грандиозны, чем цели и планы национальных электрических систем, создаваемых в первой половине двадцатого века. Необходимо объединить в единую систему различные по технической реализации и различные по типам компьютерные ресурсы (вычислительные ресурсы, ресурсы хранения и передачи информации) и донести совокупный ресурс до потребителя. Таким образом, ГРИД – это технология создания эффективных территориально-распределенных гетерогенных сетей.

Предпосылкой и базисом для развития в программе «СКИФ-ГРИД» ГРИД-технологий являются созданные в рамках программы «СКИФ» кластерные и метакластерные конфигурации широкого диапазона производительности, а также новые технологии динамического распараллеливания вычислений, мониторинга и управления кластерными и метакластерными конфигурациями.

Важнейшим практическим внедрением результатов реализации программы «СКИФ» является создание в ОИПИ НАН Беларуси **суперкомпьютерного центра коллективного пользования** с возможностью удаленного доступа к его вычислительным ресурсам. Создание суперкомпьютерного центра в ОИПИ НАН Беларуси для развития и внедрения в НАН Беларуси наукоемких информационных технологий позволяет предоставлять услуги для решения наукоёмких задач, возникающих в промышленности и в других областях народного хозяйства, требующих компьютерных и информационных ресурсов, владение которыми недоступно или экономически нецелесообразно для отдельных организаций.

Для развития такой стратегически важной области, как суперкомпьютерное направление, необходимо:

1) Поддерживаемая государством инфраструктура комплексного проекта. Эта задача была частично решена в рамках программы «СКИФ» и более полно может быть решена в рамках формируемых программ Союзного государства «ТРИАДА» и «СКИФ-ГРИД».

2) Внедрение и применение современных наукоёмких информационных технологий компьютерного проектирования (CAD-технологии) и инженерного анализа (CAE-технологии), планирования и управления производством на основе данных об изделии (CAM и PDM-технологии) и других технологий, объединяемых в комплекс CALS-технологий в основных отраслях гражданской, военной промышленности и социально-экономической сфере.

3) Организация подготовки и переподготовки кадров в области передовых информационных технологий.

4) Проведение научно-исследовательских работ по использованию созданных высокопроизводительных систем в перспективных областях применения: ядерная и водородная энергетика, генная инженерия, перспективное материаловедение, нанотехнологии, гидрометеорология, биоинформатика, молекулярное моделирование, геоинформатика и другие.

С учетом вышеизложенного следует отметить, что в последнее время наметилась тенденция требовать от науки конкретной экономической отдачи, обуславливать бюджетное финансирование научных исследований возможностью софинансирования: либо из собственных средств исследовательского учреждения, либо из средств потенциальных заказчиков и заинтересованных ведомств. В целом такой подход правильный. Однако в отрасли информационно-телекоммуникационных технологий он может использоваться только для весьма частных прикладных проектов. Крупные фундаментальные исследования и инфраструктурные проекты должны иметь полное, адекватное и стабильное государственное финансирование. Это объясняется тем, что речь идет об исследованиях и разработках, призванных создать новую инфраструктуру государства, в интересах всего государства в целом, т.е. создать «общественное достояние» («public goods»).

Такие проекты – создание инфраструктуры страны в интересах всех («общественное достояние») – должны выполняться за счет средств бюджета. Свои отчисления все ведомства, все предприятия страны, даже физические лица и исполнители уже сделали, в виде отчислений налогов в бюджет, который и предназначен для реализации общегосударственных нужд.

Точно так же, надо быть весьма аккуратными при постановке вопросов об экономической эффективности таких системообразующих инфраструктурных проектов. Например, основной эффект от программы создания семейства высокопроизводительных вычислительных установок (например, суперкомпьютеров «СКИФ») будет получен не за счет продаж суперкомпьютеров или продажи расчетного времени на них, а за счет уже упоминавшихся вторичных эффектов:

– после освоения новых информационных технологий на многих предприятиях (использующих разработанную технику) удастся расширить выпуск конкурентоспособной продукции;

– повысится эффективность предприятия, расширится производство;

– увеличатся налоговые поступления в бюджет.

Вот эти самые расширенные налоговые поступления от эффективно работающих предприятий и вернут в бюджет те деньги, которые были потрачены на создание общегосударственной инфраструктуры – отечественного семейства суперкомпьютеров. Точно такие же обстоятельства связаны и с крупными GRID-проектами. Как аналог можно упомянуть национальные автомагистрали, которые, как правило, строятся за счет бюджетных средств. Эффект же для экономики государства от наличия в стране хороших автомагистралей вполне очевиден. В таких проектах создается общенациональный информационно-коммуникационный фундамент новой экономики и нового государства – экономики и государства, основанных на знаниях. Экономический эффект от этого надо ожидать в структурных изменениях и повышении эффективности всей экономики и государства в целом.

Список использованной литературы

1) ГОСТ 21552-84. Средства вычислительной техники. Общие технические требования, правила приёмки, методы испытаний, маркировка, упаковка, транспортирование и хранение.

2) ГОСТ 16325-88. Машины вычислительные электронные цифровые общего назначения. Общие технические требования.

3) ГОСТ 27.002-89. Надёжность в технике. Основные понятия. Термины и определения

4) ГОСТ 27.003-90. Надёжность в технике. Состав и общие правила задания требований по надёжности.

5) Соловьев А.Д. Оценка надежности восстанавливаемых систем. // Изд. «Знание», Москва, 1987 – С.50.

6) Гнеденко Б.В. , Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности. // Изд. «Наука», Москва, 1966. – С.524.

7) Овчаров Л.А. Прикладные задачи теории массового обслуживания. // Изд. «Машиностроение», Москва,1969. – С.324.

8) Романов Г.С., Станкевич Ю.А. Физика горения и взрыва. - 1981. – Т. 17, №6 – С. 77-82.

9) Станкевич Ю.А., Станчиц Л.К., Степанов К.Л. Publ. Obs. Astron. Belgrade – 2002 –V.74 –P.179-183.

10) Броуд Г. Расчеты взрывов на ЭВМ. // Газодинамика взрывов. М.- Мир – 1976.

11) Romanov G.S., Stankevich Yu.A., Stanchits L.K., Stepanov K.L. Int. J. Heat Mass. Transfer. - 1995. - V. 38 - No.3 -P. 545.

12) Лапицкий В.А., Стежко И.К., Трухан В.А. Количественная оценка статистических и динамических характеристик медицинских изображений // Цифровая обработка изображений. Сб.науч.тр. Вып.4./ Под ред. С.В. Абламейко. -Минск: Ин-т техн. кибернетики НАН Беларуси, 2000. – С.165-179.

13) Левшинский Л.И., Лапицкий В.А., Трухан В.А., Стежко И.К. Аппаратно-программный комплекс ввода, обработки и архивации видеоизображений (АПК-видео) // Цифровая обработка изображений. Сб.науч.тр. Вып.3./ Под ред. С.В. Абламейко. -Минск: Ин-т техн. кибернетики НАН Беларуси, 1999 – С.7-14.

14) Ливенцева М.М., Константинова Е.Э. Дифференцированная терапия больных рефрактерной артериальной гипертонией с учетом состояния микроциркуляции и функциональных свойств эритроцитов // Микроциркуляция и гемореология: Материалы II Международной конференции 29-30 августа 1999 г., Ярославль-Москва. – М, 1999. – С.320.

15) Состояние системы микроциркуляции и показатели гемореологии при острых и хронических формах ИБС/ Константинова Е.Э., Цапаев В.Г., Цапаева Н.Л., Милютин А.А. // Актуальн. вопросы кардиол.: Сб. научн. трудов. Вып.1. – Минск, 1997. – С.127-130.

16) Lapitskii V.A., Levshinskii L.I.,Trukhan V.A., etc. Some quantitative methods for examination of medical multispectral images with the help of a personal computer // Pattern recognition and image analysis, Vol.6 No.3/ Advances in mathematical theory and applications - USA, Birmingnam, 1996 – P.634-640.

17) Курейчик В.М. Генетические алгоритмы // Материалы всероссийской научно-технической конференции с участием зарубежных представителей «Интеллектуальные САПР-97»/Известия ТРТУ – Таганрог, 1998. – С.4-7.

18) Nabhan Tarek M. and Zomaya Albert Y., Toward generating neural network structures for function approximation. Neural Networks, 1994, vol. 7. no 1. pp. 89-99.

19) Chol Chong-Ho and Choi I In Young, Construction of neural networks for piecewise approximation of continous functions. IEEE Int. Conf. Neural Networks, San Francisco. Calif. March 28-Apr.l. 1993, ICNN'93, vol. l.p. 428.

20) Godfrey. K.R.L. and Attikiouiel, Y., Self-organized color image quantization for color image data compression, IEEE Int. Conf. Neuro. Networks. San Francisco. Calif., March 28-Apr.l 1993. ICNN-93, vol. 3. pp. 1622-1626.

Публикации авторов

1) Абламейко С.В., Абрамов С.М., Анищенко В.В. , Парамонов Н.Н., Чиж О.П. Совместная белорусско-российская программа "СКИФ". // Суперкомпьютерные системы и их применение. Доклады Международной научной конференции SSA'2004. – Минск, 26-28 октября 2004. – С. 23-27.

2) Абламейко С.В., Абрамов С.М., Анищенко В.В., Парамонов Н.Н., Чиж О.П. Кластерные конфигурации СКИФ Ряда 1. // Суперкомпьютерные системы и их применение. Доклады Международной научной конференции SSA'2004. – Минск, 26-28 октября 2004. – С. 54-61.

3) Абламейко С.В., Абрамов С.М., Анищенко В.В., Парамонов Н.Н., Чиж О.П. Модели суперкомпьютеров СКИФ Ряда 2. // Суперкомпьютерные системы и их применение. Доклады Международной научной конференции SSA'2004. – Минск, 26-28 октября 2004. – С. 73-76

4) Абламейко С.В., Абрамов С.М., Анищенко В.В. , Парамонов Н.Н. , Чиж О.П. Принципы построения суперкомпьютеров семейства СКИФ. // Суперкомпьютерные системы и их применение. Доклады Международной научной конференции SSA'2004. – Минск, 26-28 октября 2004. – С. 109-115

5) Абрамов С.М., Адамович А.И., Инюхин А.В., Московский А.А., Роганов В.А., Шевчук Ю.В., Шевчук Е.В. 2004. Т-система с открытой архитектурой – // Международная научная конференция "Суперкомпьютерные системы и их применение" SSA' 2004: Доклады конференции (26-28 октября 2004 года, Минск). – Мн.: ОИПИ НАН Беларуси, 2004. – С. 18-22

6) Медведев С.В., Назаренко А.А., Петрушина М.В., Чиж О.П. Конечно-элементный анализ машиностроительных конструкций на суперкомпьютерах семейства СКИФ // Международная научная конференция "Суперкомпьютерные системы и их применение" SSA' 2004: Доклады конференции (26-28 октября 2004 года, Минск). – Мн.: ОИПИ НАН Беларуси, 2004. – С. 198-206.

7) Анищенко В.В., Кульбак Л.И., Фисенко В.К. Методология оценки надежности кластерных суперкомпьютеров // Международная научная конференция "Суперкомпьютерные системы и их применение" SSA' 2004: Доклады конференции (26-28 октября 2004 года, Минск). – Мн.: ОИПИ НАН Беларуси, 2004. – С. 244-249.

8) Абламейко С.В., Абрамов С.М., Анищенко В.В., Парамонов Н.Н. Принципы построения суперкомпьютеров семейства "СКИФ" и

их реализация. // Ежеквартальный научный журнал "Информатика". ОИПИ НАН Беларуси. – Минск, №1, январь-март 2004. – С. 89-106.

9) Анищенко В.В., Парамонов Н.Н. Программа Союзного государства "Разработка и освоение в серийном производстве семейства высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе" (шифр "СКИФ"). // Ежеквартальный научный журнал "Информатика". – ОИПИ НАН Беларуси, Минск, №1, январь-март 2004. – С. 166-167.

10) Анищенко В.В., Кульбак Л.И., Фисенко В.К. Показатели и математическая модель надежности кластерного суперкомпьютера // Информатика.– Минск, № 2, 2004. – С. 5 – 12.

11) Анищенко В.В., Кульбак Л.И., Фисенко В.К. Надежность и отказоустойчивость кластерных вычислительных систем // автоматика и вычислительная техника. № 5 – 2004. – С. 32-42.

12) Абрамов С.М., Анищенко В.В., Парамонов Н.Н. Суперкомпьютерные кластерные конфигурации "СКИФ". // Научный сервис в сети ИНТЕРНЕТ. Труды Всероссийской научной конференции. – Новороссийск, 20-25 сентября 2004. – С. 216-218.

13) Абламейко С.В., Абрамов С.М., Анищенко В.В., Парамонов Н.Н. Суперкомпьютеры семейства "СКИФ". // Программные системы: теория и приложения. Труды международной конференции. – Переславль-Залесский, май 2004. – С. 157-183.

14) Абрамов С.В., Парамонов Н.Н. Дальнейшее развитие суперкомпьютерного направления "СКИФ". // Программные системы: теория и приложения. Труды международной конференции – Переславль-Залесский, май 2004. С. 185-196.

15) Абрамов С.М., Адамович А.И., Коваленко М.Р., Парамонов Н.Н., Слепухин А.Ф. Кластерные системы семейства "СКИФ". // Научный сервис в сети ИНТЕРНЕТ. Труды Всероссийской научной конференции. – Новороссийск, 22-27 сентября 2003. – С. 147-151.

16) Медведев С.В., Чиж О.П. Компьютерное проектирование и испытание сварных конструкций в мультипроцессорной параллельной среде СКИФ // Компьютерные технологии в соединении материалов: Тез. докл. 4-й Всерос. науч.-техн.конф., 8-10 октября 2003 г. – Тула: ТулГУ, 2003. – С. 27-28.

17) Абрамов С.М., Анищенко В.В., Парамонов Н.Н., Чиж О.П. Разработка и опыт эксплуатации суперкомпьютеров семейства "СКИФ". // Информационные системы и технологии IST 2. Материалы 1 Международной конференции. – Минск, 5-8 ноября 2002. – С. 115-117.

18) Парамонов Н.Н., Поденок Л.П., Садыхов Р.Х. Специализированный процессорный модуль MLO (Скиф-У) для высокопроизводи-

тельных вычислений: архитектура и функционирование. // Информационные системы и технологии IST 2. Материалы 1 Международной конференции. – Минск, 5-8 ноября 2002.

19) Отвагин А.В., Садыхов Р.Х., Парамонов Н.Н. Система поддержки проектирования параллельных программ на базе MPI. // Информационные системы и технологии IST 2. Материалы 1 Международной конференции. – Минск, 5-8 ноября 2002. – С. 147-151.

20) Абрамов С.М., Айламазян А.К., Анищенко В.В., Парамонов Н.Н., Танаев В.С. Совместная суперкомпьютерная программа "СКИФ" в аспекте информационной безопасности. Комплексная защита информации. Сборник материалов VI междунар. конф. (26 февраля – 1 марта 2002 г., Суздаль). – Минск, 2002. - С.9-11.

21) Абрамов С.М., Айламазян А.К., Анищенко В.В., Парамонов Н.Н., Танаев В.С., Чиж О.П. Основные принципы создания и применения перспективных моделей семейства суперкомпьютеров "СКИФ". // Вестник связи. – Минск, 2002. №4. – С. 52-55.

22) Абрамов С.М., Айламазян А.К , Анищенко В.В., Парамонов Н.Н., Танаев В.С. Программа Союзного государства "Разработка и освоение в серийном производстве семейства высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе" (шифр "СКИФ"): результаты и перспективы. // Вестник связи. – Минск, 2001, №3. – С 35-37.