На правах рукописи

Шабанов Борис Михайлович

МЕТОДЫ И СПОСОБЫ ПОСТРОЕНИЯ, ВЫБОРА И ПРИМЕНЕНИЯ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ДЛЯ ВЫПОЛНЕНИЯ НАУЧНЫХ И ТЕХНИЧЕСКИХ ЗАДАЧ

05.13.15 — вычислительные машины, комплексы и компьютерные сети

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора технических наук

Москва — 2019

Работа выполнена в Межведомственном суперкомпьютерном центре Российской академии наук – филиале Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»

Официальные оппоненты:

Фельдман Владимир Марткович,
доктор технических наук, старший научный сотрудник, заместитель генерального директора ОАО «ИНЭУМ им. И.С. Брука»

Воеводин Владимир Валентинович,
член-корреспондент РАН, доктор физико-математических наук, профессор, заместитель директора Научно-исследовательского вычислительного центра МГУ имени М.В. Ломоносова

Ильин Вячеслав Анатольевич,
доктор физико-математических наук, главный научный сотрудник Курчатовского комплекса НБИКС-природоподобных технологий НИЦ «Курчатовский институт»

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук

Защита диссертации состоится «30» октября 2019 г. в 15 ч. 00 мин. на заседании диссертационного совета Д 002.073.02 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук по адресу: 119333, г. Москва, ул. Вавилова, д.44, кор.2.

С диссертацией можно ознакомиться в библиотеке Федерального исследовательского центра «Информатика и управление» Российской академии наук по адресу: г. Москва, ул. Вавилова, д.44, кор. 2 и на сайте www.frccsc.ru.

Автореферат разослан «_____» _____ 2019 г.

Ученый секретарь
диссертационного совета  Р.В. Разумчик

Общая характеристика работы

Актуальность. В мире на протяжении десятилетий реализуются национальные и наднациональные проекты по созданию и использованию высокопроизводительных вычислительных систем, уровень развития которых является фактором стратегического значения. Масштабные программы в данном направлении, требующие огромных инвестиций, представлены в США, Европейском Союзе, Китае, Японии, России. Очевидно, что только индустриально развитые страны в состоянии поддерживать исследования и проводить разработки в области суперкомпьютерных технологий, а также эффективно использовать высокопроизводительные вычисления.

Постоянное усложнение фундаментальных и прикладных задач требует систематических исследований в направлении повышения производительности суперкомпьютеров и организации вычислительного процесса. Актуальность подобных исследований в России определяется приоритетами Стратегии научно-технологического развития Российской Федерации в части перехода к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создания систем обработки больших объемов данных, машинного обучения и искусственного интеллекта. Разнообразие, масштаб и сложность возникающих при этом задач требуют создания и развития специальных технологий, вычислительной инфраструктуры и организационных решений. Суперкомпьютерные вычисления являются неотъемлемой частью новых производственных технологий, определенных в национальной программе «Цифровая экономика Российской Федерации» в качестве «сквозных» цифровых технологий.

Суперкомпьютеры являются сложными, дорогостоящими вычислительными системами с коротким жизненным циклом, поэтому важно обеспечить оперативность проектирования, экономичность эксплуатации и результативность их применения. В настоящее время доминирует тенденция увеличения производительности суперкомпьютерных вычислительных систем в основном за счет роста количества процессорных ядер, число которых в наиболее производительных современных системах исчисляется миллионами. Однако, эффективное использование столь большого числа процессоров в настоящее время возможно для выполнения весьма ограниченного набора приложений.

Можно выделить два основных направления оснащения суперкомпьютерных центров новыми системами. В первом случае используются системы, основная доля производительности которых обеспечивается ускорителями вычислений. Достижение максимальной эффективности вычислительного процесса в этом случае возможно при использовании пакетов программ, оптимизированных для конкретных ускорителей. Во втором случае используются универсальные системы общего назначения на традиционных процессорах с меньшей пиковой производительностью, которые, как показывает практика, наиболее востребованы пользователями суперкомпьютерных центров в сфере науки и образования. Универсальные системы могут эффективно использоваться как для стандартных оптимизированных пакетов, так и для оригинальных пользовательских программ, особенно при переносе с систем предыдущих поколений.

Повышение производительности вычислительных систем, в том числе гетерогенных, может быть достигнуто как за счет методов и средств балансировки вычислительной нагрузки между узлами (процессорными модулями) системы с учетом их характеристик и конфигурации коммуникационной сети, так и за счет организации параллельных вычислений внутри узла. Важным инструментом, позволяющим повысить производительность выполнения приложений, остается векторизация программного кода.

Помимо создания универсальных суперкомпьютерных систем, построенных на основе процессоров традиционной архитектуры, ведутся активные исследования, связанные с разработкой и поиском новых решений для повышения их реальной производительности, с учетом того, что производительность таких процессоров падает с увеличением количества обращений к медленным уровням иерархии памяти. Это отчетливо проявляется на программах с мелкоструктурным параллелизмом. Одним из таких подходов является переход на системы с архитектурой управления потоком данных, высокая производительность которых достигается за счет использования естественного параллелизма выполнения команд.

Важным направлением повышения эффективности вычислений является создание сервисно-ориентированных сетей вычислительных центров. Это обусловлено тем, что вычислительные центры могут специализироваться на решении определенных классов задач с применением проблемно-ориентированных вычислительных систем и пакетов программ. Сервисно-ориентированная сеть позволяет достигать наибольшего эффекта при решении вычислительных задач или их частей на наиболее подходящих вычислительных ресурсах независимо от их принадлежности конкретному вычислительному центру. Реализация данного подхода требует проведения исследований теоретических аспектов и поиска практических решений по созданию и функционированию сервисно-ориентированных сетей вычислительных центров, формированию профилей субъектов интеграции и оценки качества их деятельности при различных типах информационного и сетевого взаимодействия между центрами с учетом многоаспектности вычислительных потребностей.

Изложенное обуславливает своевременность и актуальность исследования, направленного на решение крупной научно-технической проблемы – поиска новых технических, технологических и организационных решений для обеспечения возрастающих потребностей в высокопроизводительных вычислениях за счет разработки и освоения новых суперкомпьютерных систем, их интеграции в единую вычислительную среду средствами сетей сервисно-ориентированных вычислительных центров, повышения производительности суперкомпьютеров за счет разработки новых архитектур процессоров.

Существенное влияние на развитие суперкомпьютерных технологий и их применение, включая разработку элементной базы, суперкомпьютерных систем, сетевой и коммуникационной инфраструктуры, построение суперкомпьютерных центров общего и специального назначения, создание математических методов, комплексов прикладных программ и организацию высокопроизводительных вычислений, решение вопросов стандартизации и повышения надежности вычислительных систем, подготовку разработчиков и обучение пользователей оказали работы В.И. Бердышева, В.Б. Бетелина, В.К. Левина, Г.И. Савина, И.А. Соколова, Б.Н. Четверушкина, Б.А. Бабаяна, Вл.В. Воеводина, И.А. Каляева, С.М. Абрамова, Г.С. Елизарова, А.А. Зацаринного, В.В. Корнеева, Р.М. Шагалиева, А.Н. Томилина, А.О. Лациса.

Цель диссертационной работы состоит в исследовании и разработке технических, технологических и организационных решений для анализа, построения и применения высокопроизводительных вычислительных систем как основы суперкомпьютерной инфраструктуры для науки, образования и инновационной деятельности.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

- Разработать метод построения суперкомпьютерных систем на основе выделения и классификации факторов, определяющих результативность применения вычислительных систем.
- Разработать комплекс решений по тестированию и анализу вычислительных кластерных систем с многоядерной архитектурой с целью определения влияния параметров на время выполнения программ.

- Создать суперкомпьютерные системы с различным набором характеристик, воплощающие разработанные архитектурные, сетевые и программные решения, сформировать на их основе интегрированную вычислительную информационно-коммуникационную среду проведения научных исследований и решения прикладных задач.
- Разработать решения по повышению производительности процессоров и вычислительных систем за счет развития архитектуры векторно-потокового процессора.
- Разработать основы, архитектурные и технические решения для создания сервисно-ориентированных программно-определяемых центров обработки данных межведомственного уровня как современной среды проведения научных исследований и инновационных разработок, формирования профессиональных компетенций специалистов в области суперкомпьютерных технологий.

Научная новизна исследования содержится в комплексе выполненных автором работ по всему технологическому циклу от проектирования и оптимизации суперкомпьютеров до создания распределенных комплексов и инфраструктур высокопроизводительных вычислительных систем, включая организацию эффективного вычислительного процесса, а именно:

- Разработан метод построения суперкомпьютерных систем на основе выделения и классификации факторов, определяющих результативность применения вычислительных систем для исследования актуальных вычислительно сложных научных проблем.
- Разработан и реализован комплекс решений по тестированию и анализу вычислительных систем с целью определения влияния параметров на время выполнения программ в вычислительных кластерах с многоядерной архитектурой.
- На основе разработанных решений создана уникальная серия суперкомпьютеров с различным набором характеристик, оригинальных архитектурных, сетевых и программных решений, сформирована интегрированная высокопроизводительная вычислительная информационно-коммуникационная среда проведения научных исследований и решения прикладных задач.
- Разработана оригинальная архитектура векторно-потоковой вычислительной системы, поддерживающая аппаратное распределение памяти, позволяющая увеличивать производительность процессора в 7–10 раз, и обеспечивающая устойчивость к изменениям латентности памяти и межпроцессорного обмена в диапазоне изменения задержек, на порядок более широком по сравнению с системами традиционной архитектуры.
- Разработаны базовые принципы, архитектурные и технические решения для создания сервисно-ориентированных программно-определяемых центров обработки данных межведомственного уровня как современной среды проведения научных исследований и инновационных разработок, формирования профессиональных компетенций специалистов в области суперкомпьютерных технологий.

Практическая ценность и достоверность полученных результатов подтверждаются положительным опытом создания и практического использования высокопроизводительных систем МВС-1000М, МВС-15000ВМ, МВС-6000IM, МВС-100К, МВС-10П, в основу которых легли научно обоснованные архитектурные, технические и технологические решения по применению суперкомпьютерных технологий в сфере науки и образования.

Разработанный и реализованный комплекс решений по созданию, развитию и интеграции высокопроизводительных вычислительных ресурсов обеспечил качественно новый уровень проведения фундаментальных исследований, реализации наукоемких проектов, что подтверждается многочисленными научными публикациями пользователей Межведомственного суперкомпьютерного центра РАН (МСЦ РАН) в ведущих зарубежных и российских научных изданиях.

Теоретические положения и накопленный практический опыт, представленные в настоящей работе, могут служить основой дальнейшего развития суперкомпьютерных технологий, создания перспективных высокопроизводительных вычислительных систем и организации на их основе распределенных вычислительных центров.

Методология и методы исследования. Результаты диссертации были получены с привлечением моделей и методов, используемых при тестировании и анализе производительности вычислительных систем, поиске архитектурных и системотехнических решений. Математическую основу исследования составляют системный анализ, теория алгоритмов, математическая логика, теория графов.

Основные положения, которые выносятся на защиту.

- Разработанный метод выбора вычислительных систем на основе выделения и классификации факторов, определяющих результативность применения вычислительных систем, позволяет создавать суперкомпьютерные системы для решения актуальных вычислительно сложных научных проблем и делает возможным выполнение оптимизации состава вычислительного центра.
- Разработанный и реализованный комплекс решений по тестированию и анализу вычислительных систем позволяет определять влияние параметров на время выполнения программ в вычислительных кластерах с многоядерной архитектурой.
- Разработанные и реализованные суперкомпьютерные системы МВС-1000М, МВС-15000ВМ, МВС-6000IM, МВС-100К и МВС-10П с различным набором характеристик, оригинальные архитектурные, сетевые и программные решения составляют основу интегрированной высокопроизводительной вычислительной информационно-коммуникационной среды проведения научных исследований и решения прикладных задач, позволившей поднять фундаментальные научные исследования в России на качественно новый уровень.
- Разработанная оригинальная архитектура векторно-потоковой вычислительной системы обеспечивает аппаратное распределение памяти, позволяет повышать производительность процессора в 7–10 раз, обеспечивает устойчивость производительности к десятикратному увеличению задержек памяти и межпроцессорного обмена.
- Разработанные архитектурные, технические и технологические решения составляют базовые основы создания сервисно-ориентированных программно-определяемых центров обработки данных межведомственного уровня как современной среды проведения научных исследований и инновационных разработок, формирования профессиональных компетенций специалистов в области суперкомпьютерных технологий.

Реализация результатов работы. Полученные результаты нашли применение при создании и использовании суперкомпьютеров МВС-1000М, МВС-15000ВМ, МВС-6000IM, МВС-100К, МВС-10П в МСЦ РАН, а также оказали существенное влияние на формирование и развитие инфраструктуры суперкомпьютерных вычислений науки и образования в стране. Результаты диссертации были получены соискателем лично или при его руководстве в научно-исследовательских работах по Программам фундаментальных научных исследований

государственных академий наук, Программам фундаментальных исследований президиума РАН, проектам РФФИ и Минобрнауки России, международным проектам DEISA-2, ExaHYPE.

Апробация. Материалы диссертации докладывались на следующих международных и всероссийских конференциях: Federated Conference on Computer Science and Information Systems, Szczecin, Poland, 18-21 September, 2011; 4-я Международная научно-техническая конференция «Распределенные вычисления и Грид-технологии в науке и образовании (GRID'2010)» Дубна 28 июня – 3 июля 2010 г.; Международная конференция «Суперкомпьютерные дни в России: Труды международной конференции», Москва, 25 – 26 сентября 2017 г.; Национальный Суперкомпьютерный Форум (НСКФ-2018), Переславль-Залесский, 27 – 30 ноября 2018 г.; Национальный Суперкомпьютерный Форум (НСКФ-2017), Переславль-Залесский, 28 ноября – 01 декабря 2017 г.; Национальный Суперкомпьютерный Форум (НСКФ-2015), Переславль-Залесский, 24-27 ноября 2015 г.; Национальный Суперкомпьютерный Форум (НСКФ-2014), Переславль-Залесский, 25 – 27 ноября 2014 г.; Конференция «Суперкомпьютерные вычисления для развития российской науки» Москва, 26 апреля 2017 г.; International conference Engineering & Telecommunications – En&T 2014 Moscow. 26 – 28 November, 2014 г.; Международная конференция «Информационные технологии в науке, социологии и бизнесе (осенняя сессия)», Ялта-Гурзуф, 1 – 10 октября 2011г.; Суперкомпьютерные технологии: разработка, программирование, применение СКТ-2010. Международная научно-техническая конференция. Таганрог, 2010 г.; Научный совет РАН «Высокопроизводительные вычислительные системы, научные телекоммуникации и информационная инфраструктура», май 2009 г., апрель 2013 г., февраль 2016 г.; II Всероссийская конференция «Центры коллективного пользования и уникальные научные установки организаций, подведомственных ФАНО России», Москва, 25 – 27 октября 2017 г.; 2018 IEEE Conference EIConRus, Москва, 29 января – 1 февраля 2018г.; 6-я Всероссийская научно-техническая конференция «Суперкомпьютерные технологии СКТ-2018» Дивноморское, Геленджик, 17 – 22 сентября 2018 г.; 4-я Всероссийская научно-техническая конференция «Суперкомпьютерные технологии СКТ-2016» Дивноморское, Геленджик, 19 – 24 сентября 2016 г.; 3-я Всероссийская научно-техническая конференция «Суперкомпьютерные технологии СКТ-2014», Дивноморское, Геленджик, 29 сентября – 4 октября 2014 г.; 2-я Всероссийская научно-техническая конференция «Суперкомпьютерные технологии СКТ-2012», Дивноморское, Геленджик, 24 – 29 сентября 2012 г.

Публикации. По теме диссертации автором опубликовано 83 печатные работы, из них 24 работы опубликованы в изданиях, входящих в Перечень рецензируемых научных изданий, рекомендованных ВАК.

Структура и объем работы. Диссертация состоит из введения, четырех глав и заключения. Содержание работы изложено на 264 страницах машинописного текста. Список использованных источников составляет 172 наименования.

Краткое содержание работы

Во введении обоснована актуальность работы, сформулированы цель и задачи работы, приведены научная новизна, практическая значимость полученных результатов и защищаемые положения, рассмотрена структура диссертации.

В первой главе представлены анализ тенденций развития суперкомпьютеров, требования к их производительности и применению.

В разделе 1.1 приведен обзор развития суперкомпьютерных вычислений. Мировой опыт показывает, что крупные научно-исследовательские организации часто используют собственные суперкомпьютерные центры, ориентированные на решение определенных

классов научных и технических задач. Наиболее производительные суперкомпьютеры находятся в США и Китае. Рассмотрены крупные зарубежные центры: Окриджская национальная лаборатория (ORNL), Ливерморская национальная лаборатория им. Лоуренса (LLNL), Лос-Аламосская национальная лаборатория (LANL), Национальный суперкомпьютерный центр Уси (NSCC-Wuxi), Национальный суперкомпьютерный центр в Гуанчжоу, Национальный центр суперкомпьютерных приложений (NCSA) при Иллинойском университете, Суперкомпьютерный центр Сан-Диего (SDSC), Суперкомпьютерный центр имени Лейбница Баварской академии наук (LRZ), Юлихский исследовательский центр (JSC), Центр высокопроизводительных вычислений (HLRS) при Университете Штутгарта, Суперкомпьютерный и дата центр общества Макса Планка (MPCDF), Техасский вычислительный центр (TACC) при Техасском университете в Остине, Исследовательское вычислительное подразделение Университета Пердью, консорциум CINECA, Барселонский суперкомпьютерный центр (BSC), Объединенный центр высокопроизводительных вычислений университета Цукуба и Токийского университета (JCAHPC), Суперкомпьютерный центр Китайской академии наук (SCCAS).

В США, Европе и Азии выполняются проекты, направленные на объединение суперкомпьютерных центров в рамках единой цифровой инфраструктуры для науки и образования. Рассмотрены проекты объединения суперкомпьютерных центров XSEDENet, PRACE и InfiniCortex. Отмечена устойчивая тенденция консолидации ресурсов вычислительных центров за счет объединения в сети с целью оптимизации использования ресурсов.

В разделе 1.2 рассмотрены суперкомпьютерные центры в России, в том числе в Российской академии наук (РАН). Суперкомпьютерные центры являются неотъемлемой частью информационно-телекоммуникационной инфраструктуры науки и образования России. Лидером рейтинга Top-50 самых мощных суперкомпьютеров СНГ является суперкомпьютер «Ломоносов-2» (пиковая производительность 4,95 ПФлопс), установленный в Научно-исследовательском вычислительном центре Московского государственного университета имени М.В. Ломоносова (НИВЦ МГУ). Высокопроизводительные системы многие годы развиваются в НИЦ «Курчатовский институт» и Объединенном институте ядерных исследований (ОИЯИ), ряде университетов, например, Санкт-Петербургском политехническом университете Петра Великого (СПбПУ), Национальном исследовательском Нижегородском государственном университете им. Н.И. Лобачевского (ННГУ), Томском государственном университете(ТГУ). В 2018 г. новые суперкомпьютеры были установлены в Сколковском институте науки и технологий (Сколтех) и Национальном исследовательском университете «Высшая школа экономики» (НИУ ВШЭ). На рисунке 1 приведены организации науки и образования, в которых имеются вычислительные системы по данным 30-го рейтинга Top-50, составленного 2 апреля 2019 г.

В институтах РАН суперкомпьютеры используются в основном для решения фундаментальных научных задач. В рейтинге Top-50 находятся системы из 9 организаций РАН.

С момента создания в 1996 г. по настоящее время в МСЦ РАН устанавливались и эксплуатировались вычислительные системы мирового уровня производительности, входившие в первую сотню самых мощных суперкомпьютеров мира. За это время доля предоставляемых центром ресурсов в общем объеме суперкомпьютерных ресурсов для академии наук не опускалась ниже 55% и доходила до 85%.

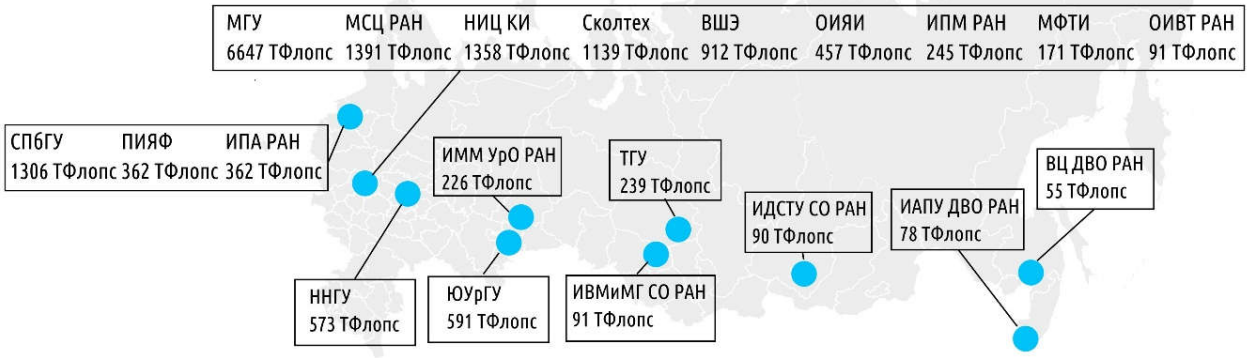

Рисунок 1. Суперкомпьютерные центры науки и образования в России

В разделе 1.3 определен и рассмотрен состав ключевых компонентов суперкомпьютерного центра.

Главным компонентом суперкомпьютерного центра является вычислительная инфраструктура, которая может включать суперкомпьютеры различной архитектуры. Рассмотрены основные архитектуры, используемые для построения суперкомпьютеров: MPP-системы (массово-параллельные системы) и кластеры. Приведен обзор процессоров для высокопроизводительных вычислений, рассмотрены свойства коммуникационных сетей суперкомпьютеров Intel Omni-Path, InfiniBand, Ангара, Ethernet.

Особенность современного состояния и развития элементной базы вычислительной техники – прекращение роста тактовой частоты универсальных микропроцессоров. Дальнейший рост производительности вычислительных систем достигается за счет увеличения числа процессорных ядер, применения ускорителей и поиска новых архитектур. При этом число ядер в процессорах растет быстрее, чем пропускная способность взаимодействия процессора с памятью.

Система хранения данных (СХД) используется для хранения программных модулей и проектов, исходных данных, промежуточных и конечных результатов вычислений. Рассмотрены подходы к построению крупных систем хранения данных, предполагающие использование таких параллельных файловых систем, как Lustre, GPFS, OrangeFS, Ceph.

Для доступа пользователей, загрузки/выгрузки данных необходима сетевая инфраструктура. Рассмотрено ее функциональное деление на коммуникационную, транспортную, управляющую и служебную сети, определены требования к их характеристикам и возможности совмещения функций одной сетью. Показаны способы организации файлового обмена между вычислителями и СХД при использовании коммуникационных и транспортных сетей на основе разных сетевых технологий InfiniBand, Intel Omni-Path, Ethernet.

Для функционирования оборудования необходима инженерная инфраструктура, включающая, как правило, подсистемы охлаждения, электроснабжения, охранно-пожарной сигнализации, пожаротушения и контроля доступа.

К числу традиционных проблем создания суперкомпьютерных систем относят обеспечение их энергоэффективности и надежности. Сегодня технически возможно построить суперкомпьютер производительностью более 1 экзафлопса, но потребляемая им мощность будет слишком высока (около 50 МВт). Поэтому новые системы должны быть энергоэффективны, а программное обеспечение должно учитывать те возможности

архитектуры и аппаратных средств, которые повышают энергоэффективность. Кроме того, необходимо использование методов динамического распределения вычислительных операций на миллионы процессорных ядер с учетом того, что некоторые из них могут выйти из строя во время вычислений.

Показателем качества суперкомпьютерной системы и соответствующей инженерной инфраструктуры является коэффициент энергоэффективности (PUE), используемый для оценки потерь энергии во время эксплуатации. Помимо повсеместно используемого воздушного охлаждения для снижения энергопотребления применяется жидкостное охлаждение, в частности, «холодной водой» (температура 18°–22°C) и «горячей водой» (температура 40°–60°C). Возможно также использование иммерсионного охлаждения с погружением вычислительных модулей в диэлектрическую жидкость.

Представлено используемое в суперкомпьютерных системах программное обеспечение, приведен анализ его применения пользователями, выделены наиболее употребляемые пакеты программ. Рассмотрены стеки (согласованные многоуровневые наборы) базового программного обеспечения, используемого в крупных зарубежных и российских суперкомпьютерных центрах.

Проанализировано управление вычислительными ресурсами и пользовательскими заданиями в суперкомпьютерах, которое осуществляется системами управления заданиями (СУЗ). Приведено описание СУЗ: Moab HPC Suite, SLURM, а также отечественной системы управления прохождением параллельных заданий (СУППЗ). Данные экспериментального сравнения планирования СУППЗ и SLURM показывают паритет по качеству планирования заданий. Преимуществами СУППЗ, благодаря которым система эксплуатируется в МСЦ РАН в течение последних двух десятилетий, являются соответствие сложившемуся порядку работы и практическим потребностям пользователей и системных администраторов МСЦ РАН, развитая подсистема сбора и обработки статистики, а также предусмотренные в системе возможности комплексирования суперкомпьютерных центров в единую систему коллективного доступа.

Отмечено, что процесс создания новых вычислительных систем для оснащения суперкомпьютерных центров требует решения сложных научных, организационных и технических задач, среди которых выделяются выбор вычислительной инфраструктуры и коммуникационной среды с учетом требований инженерной и сетевой инфраструктуры, программного обеспечения и систем хранения данных.

Во второй главе рассматриваются вопросы создания высокопроизводительных вычислительных систем.

В разделе 2.1 рассмотрены архитектуры суперкомпьютерных систем и особенности их программирования. Основным решением в последние десятилетия, обеспечивающим рост производительности, является развитие способов параллельной обработки данных. Рассмотрены особенности вычислительных систем, построенных на основе векторных процессоров и ускорителей вычислений, многопроцессорных систем с общей памятью, кластеров и MPP-систем.

Рассмотрены уровни выполнения программ на высокопроизводительных системах – от уровня функциональных устройств процессора до уровня метасети. Приведено сопоставление организации программ с уровнями аппаратной поддержки параллельных вычислений.

В разделе 2.2 приводится разработанный на основе опыта создания, эксплуатации и тестирования вычислительных систем метод выбора вычислительных систем по набору факторов, определяющих результативность применения вычислительных систем и функционирование суперкомпьютерного центра.

Выбор вычислительной системы для суперкомпьютерного центра определяется набором факторов, наиболее значимые из которых следующие.

1. Производительность, которая является ключевым параметром при выборе вычислительных систем. При этом следует учитывать специфику и особенности задач вычислительного центра. Для оценки производительности предложен комплекс решений по тестированию и анализу вычислительных систем.

Для оценки влияния на производительность одновременной работы множества ядер введены коэффициент конфликтов по использованию памяти (*CR*) и коэффициент конфликтов по сетевой инфраструктуре (*CL*). Коэффициент *CR* определяется как отношение времени выполнения фрагмента программного кода с интенсивной выборкой из памяти при одновременной работе нескольких ядер ко времени выполнения того же кода при работе одного ядра. Коэффициент *CL* определяется как отношение времени выполнения фрагмента программного кода с интенсивной передачей данных по сети при одновременной работе нескольких пар ядер ко времени выполнения того же кода при работе одной пары ядер.

Значения коэффициентов *CR* и *CL* используются в качестве параметров при априорной оценке времени выполнения программ. Такой способ позволяет проводить оценку масштабируемости программ без использования ресурсоемкого тестирования.

2. Стоимость и доступность. Создание суперкомпьютерных систем связано с освоением и использованием широкого спектра новейших технологий, при этом существуют естественные и искусственные ограничения на использование конкретных типов модулей и узлов. Не все возможные конфигурации оборудования (например, количество процессоров, ускорителей, объем и скорость работы памяти и другие) могут быть доступны для приобретения в определенный момент времени. При этом компоненты системы должны иметь приемлемую стоимость.

3. Соответствие экосистеме суперкомпьютерного центра. Динамичная совокупность научного оборудования центра, его персонала, пользователей и отношений между ними образуют экосистему суперкомпьютерного центра. Нарушение сложившегося в экосистеме порядка работы пользователей приводит к задержкам освоения новых систем, снижению активности и даже оттоку пользователей, что сказывается на эффективности работы центра.

4. Возможность масштабирования, востребованность, защита инвестиций. Как правило, создание и развитие вычислительной системы происходит в несколько этапов, связанных с наращиванием требуемых вычислительных мощностей. Поэтому возможность масштабирования является принципиальной, что предъявляет определенные требования к структуре коммуникационных, транспортных и служебных сетевых соединений, а также к производительности СХД. При создании новой системы крайне важно использовать новейшие поколения процессоров, а также модули с возможностью модернизации заменой части оборудования на более современное.

5. Демасштабирование, продление жизненного цикла системы. Помимо наращивания вычислительной мощности важное значение имеет возможность использования системы после ее замены на более производительную. Необходимо обеспечить возможность разделения системы на части, которые смогут использоваться в других вычислительных центрах. Кроме того, отдельные узлы системы могут использоваться в дальнейшем как серверы доступа, серверы управления сетью, файловые серверы, веб-серверы и др. В этом случае для обеспечения указанных требований желательна возможность изменения конфигурации узлов суперкомпьютера, например, увеличения объема памяти, добавления сетевых интерфейсов.

6. Стоимость и трудоемкость эксплуатации. Эксплуатация системы требует затрат на квалифицированный инженерный персонал, программистов и администраторов. В вычислительных системах используются сотни и тысячи вычислительных узлов, что приводит к частым отказам и требует сокращения времени восстановления работоспособности. Поэтому обслуживание системы должно быть технологичным и минимально трудоемким, что является сложной технической и организационной задачей,

предъявляющей жесткие требования к системам администрирования и мониторинга. В связи с постоянным ростом энергопотребления современных вычислительных систем стоимость потребляемой электроэнергии становится основной частью затрат на эксплуатацию, что оказывает существенное влияние на выбор решений для вычислительного центра.

7. Требования к созданию и развитию инфраструктуры. Помещения, в которых размещаются суперкомпьютеры, как правило, ограничены по площади и объему, поэтому для обеспечения масштабируемости систем и снижения латентности в коммуникационной среде важна компактность решений. Охлаждение вычислительных систем требует дополнительных затрат как на создание специальной инфраструктуры, так и на ее эксплуатацию. Кроме того, мощность подводящих линий электропитания, как правило, ограничена, что ставит задачу оптимизации энергопотребления. Центр должен иметь каналы связи с высокой пропускной способностью с основными пользователями, так как для выполнения задач пользователями часто требуется загрузка и выгрузка значительного объема данных, в том числе конфиденциальных.

8. Возможность распределенной обработки данных. Срок эксплуатации суперкомпьютерной системы (5–10 лет) значительно превышает срок появления новых поколений вычислительных систем (3–4 года), поэтому суперкомпьютеры нового поколения эксплуатируются совместно с системами предыдущих поколений. Устойчивой тенденцией развития высокопроизводительных вычислений является консолидация суперкомпьютерных ресурсов разных центров в распределенные сети, при этом объединяться в них могут системы разных поколений. Вводимые в эксплуатацию новые суперкомпьютеры должны быть совместимы как с системами предыдущих поколений, так и с другими суперкомпьютерами из состава распределенной сети для высокопроизводительных вычислений.

На основе опыта создания, использования и тестирования широкого спектра высокопроизводительных вычислительных систем разработан метод выбора таких систем по набору перечисленных факторов. В отношении каждого из возможных технических, технологических и организационных решений реализации суперкомпьютеров выбираются альтернативные варианты, обеспечивающие учет факторов 2–8. Производительность выбранных вариантов определяется с помощью специально разработанного комплекса тестирования и анализа. По результатам тестирования производится выбор варианта.

В разделе 2.3 рассмотрен комплекс решений по тестированию и анализу вычислительных систем, а также приведены характеристики оборудования и результаты работы тестов.

В разработанном комплексе решений используются 3 вида тестов (таблица 1): стандартные тесты, пользовательские программы и синтетические тесты.

Стандартные тесты SPEC, NAS NPB и HPL (Linpack) широко используются для тестирования вычислительных систем. В комплекс помимо стандартных тестов включен ряд программ, применяемых пользователями (газовая динамика, наноэлектроника, термодинамика атмосферы, молекулярная динамика). Синтетические тесты включают блочное умножение матриц (Matmul), а также работу с памятью и коммуникационной сетью (Laptest).

С помощью рассмотренных выше тестов была проведена оценка производительности прототипов суперкомпьютеров для использования в вычислительных системах в МСЦ РАН. На рисунке 2 показано применение одной из программ пользователей, широко применяемой для расчетов молекулярной динамики, для тестирования решений (IBM JS20 и Regatta p655) при выборе вычислительной системы МВС-15000ВМ. Также приведено сравнение с системой МВС-1000М. Как видно из графика, решение JS20 показывает лучшую производительность и масштабируемость.

Таблица 1. Состав тестов для комплекса решений

Уровень параллелизма	Объект тестирования	Тесты
Операции над регулярными структурами данных	Процессорное ядро	SPEC Laptest
Нити	Узел системы	SPEC NAS Matmul
Процессы	Система с распределенной памятью	NAS HPL Программы пользователей Matmul Laptest

Тестирование и эксплуатация систем показали, что, во-первых, стандартные тесты и программы пользователей не в полной мере отражают специфику выполнения пользовательских программ, во-вторых, особенности поведения программ на новом оборудовании (размерность, точность) не всегда очевидны, в-третьих, требуется оптимизация процесса тестирования, так как проведение всеобъемлющего тестирования является затратным и часто невозможным.

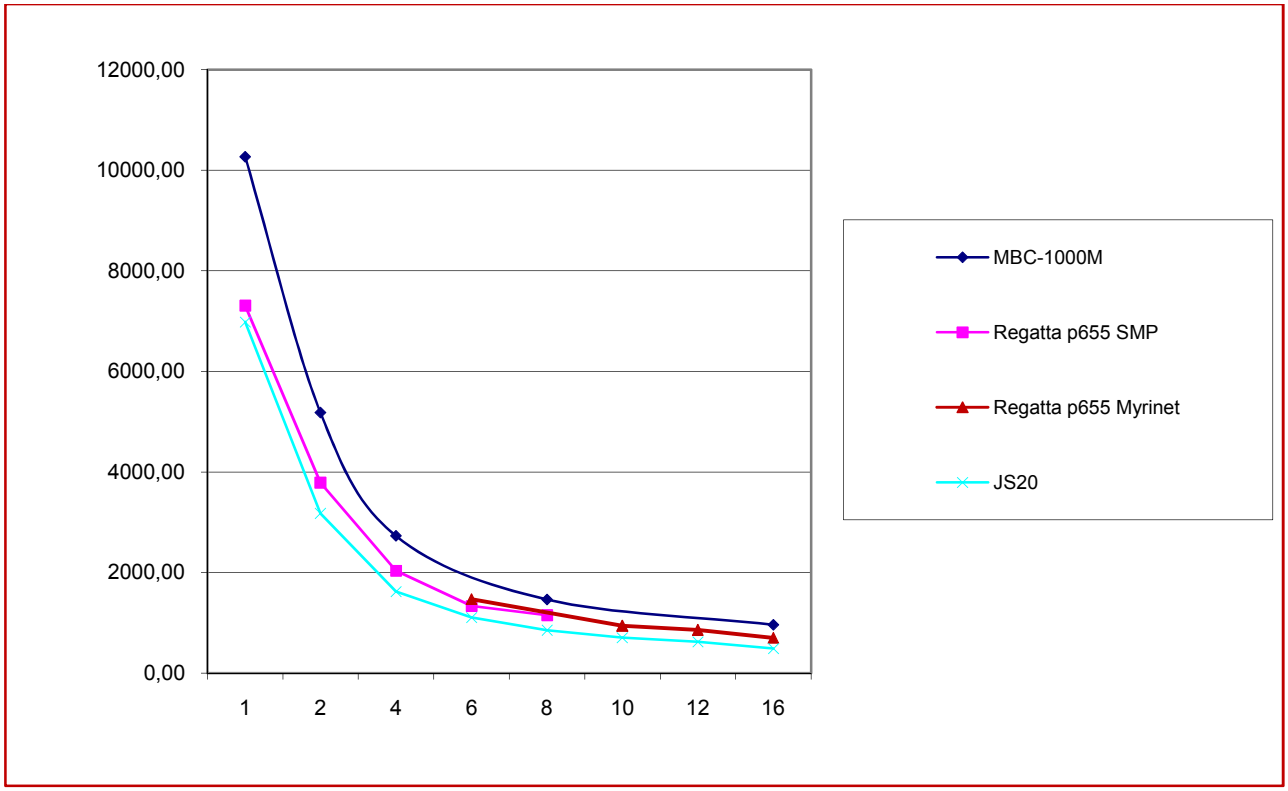

Рисунок 2. Молекулярная динамика, GROMACS, время (секунды) в зависимости от числа ядер

Для определения параметров вычислительной системы под нагрузкой и предсказания поведения системы при выполнении пользовательских программ были разработаны синтетические тесты Matmul (блочное умножение матриц) и Laptest (определение времени выполнения фрагментов программного кода при работе с памятью и передачах по коммуникационной сети). При подготовке синтетических тестов были определены типовые элементы параллельных программ, моделирование которых позволило определить

численные параметры оценки выполнения программ. Эти тесты определяют производительность ядер при выполнении операций с плавающей запятой, пропускную способность оперативной памяти при выполнении операций чтения-записи, коммуникационной среды при выполнении вызовов MPI, подсистемы ввода-вывода при выполнении файловых операций с учетом наличия конкуренции ядер по доступу к общей памяти и сетевым соединениям вычислительного узла, локальности выполняемых программ по данным и используемым моделям программирования.

В качестве характерного примера тестирования рассмотрено сравнение процессоров, которое было проведено перед расширением вычислительной системы МВС-100К. Проводилось сравнение 4-ядерного процессора Intel Xeon 5450 (Harpertown) с тактовой частотой 3 ГГц и 6-ядерного Intel Xeon X5670 (Westmere) с частотой 2,93 ГГц на программе Matmul. Результаты сравнения представлены на рисунке 3. Как видно из графиков, влияние переполнения кэш-памяти, которое происходит при размере блока от 256 до 1024, для процессора Westmere выражено слабее. На больших размерах блоков процессор Westmere показывает лучшие характеристики, что предъявляет менее жесткие требования к локализации данных в программах и тем самым расширяет круг эффективно решаемых задач. Таким образом, несмотря на более низкую тактовую частоту процессоров Westmere, их производительность значительно выше при больших объемах обрабатываемых данных, а при небольших сопоставима. В связи с этим расширение системы производилось с использованием процессоров Westmere.

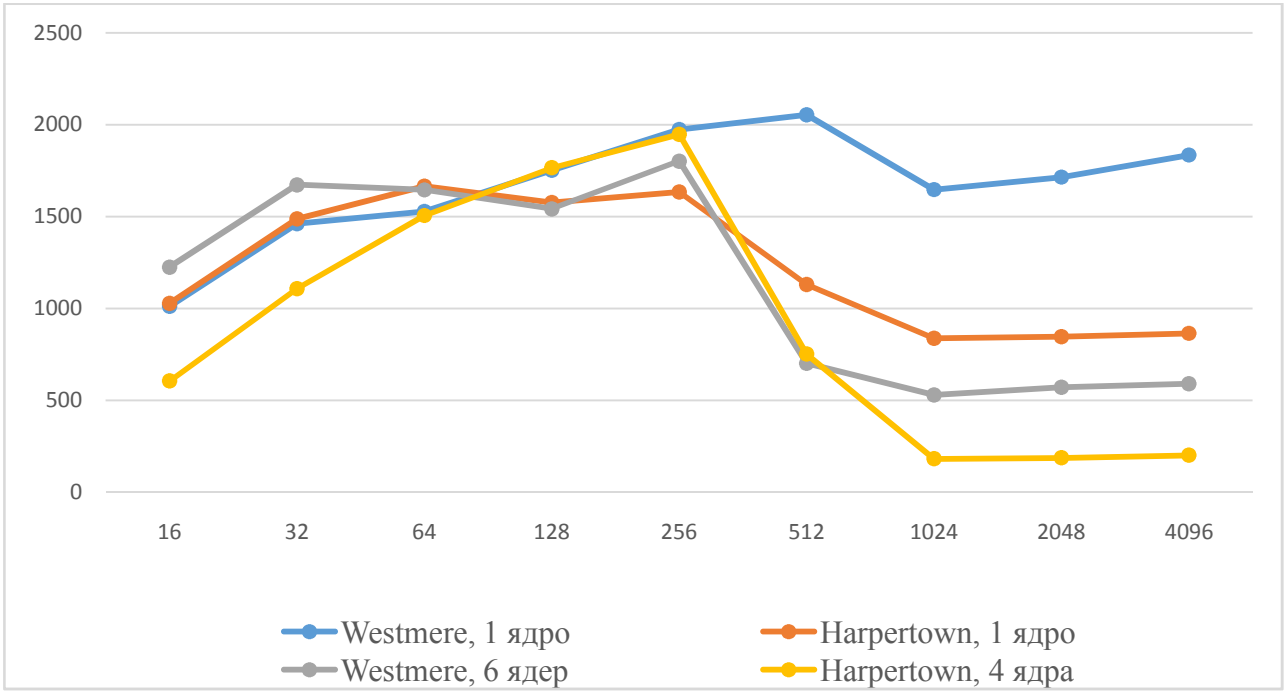

Рисунок 3. Блочное умножение матриц 4096×4096. Зависимость производительности (МФлопс) одного ядра от размера блока и числа одновременно работающих ядер

С помощью синтетических тестов определяются значения коэффициентов конфликтов по использованию памяти *CR* и по сетевой инфраструктуре *CL*. Для определения значения коэффициента *CR* выполняется скалярное произведение векторов с интенсивной выборкой из памяти в одном узле при различном числе работающих ядер. Для определения значения коэффициента *CL* производится интенсивная передача данных между узлами по коммуникационной сети при различном числе работающих ядер. Замедление, вызванное конкурентными обращениями к памяти (равное значению коэффициента *CR*), оказывает значительное влияние на производительность вычислительных узлов.

Тестирование проводилось для всех сегментов вычислительной системы МВС-10П. Согласно результатам тестирования универсальные процессоры (Intel Xeon) показывают большую эффективность (отношение реальной производительности к пиковой), однако процессоры Intel Xeon Phi обеспечивают лучшую масштабируемость.

В качестве примера на рисунке 4 приведены значения коэффициентов *CR* и *CL* для узлов МВС-10П ОП, построенных на двух универсальных 16-ядерных процессорах Intel Xeon E5-2697Av4 (Broadwell), и МВС-10П МП2, построенных на 72-ядерных процессорах Intel Xeon Phi 7290 (KNL). Тестирование проводилось при последовательном чтении из памяти или передаче по коммуникационной сети данных объемом 1ГБ при разном числе одновременно работающих ядер на каждом узле. Дополнительно приведено значение коэффициента *CR* при задействованном режиме одновременной многопоточности (hyper-threading) на 64 виртуальных ядрах процессора Broadwell.

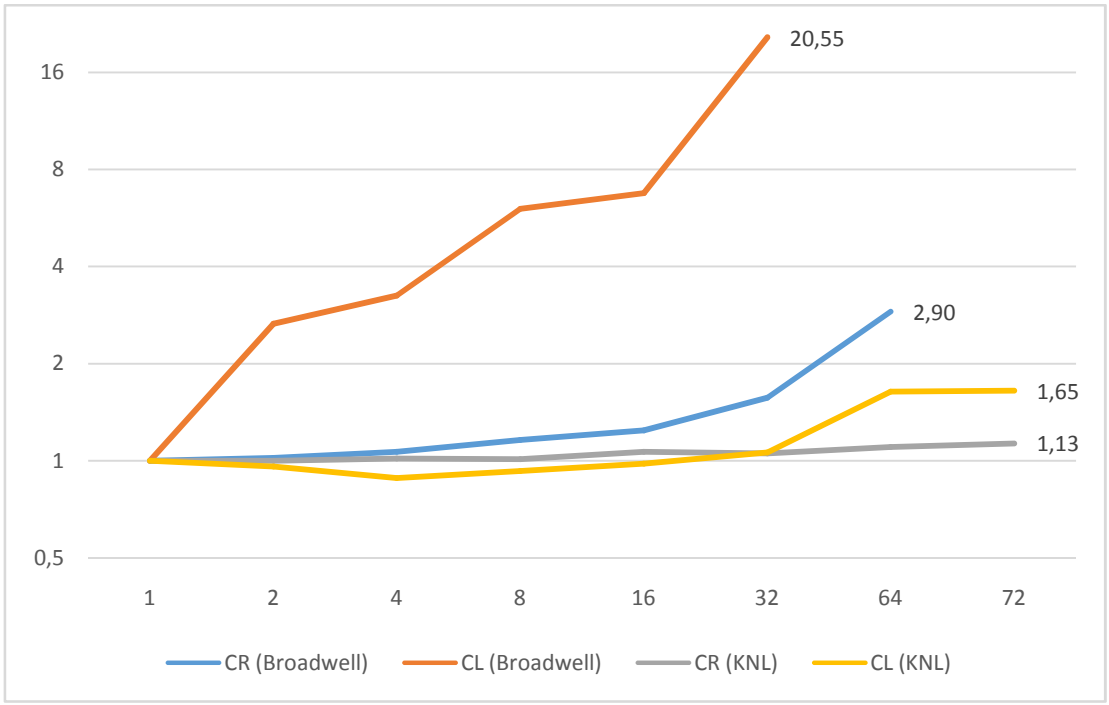

Рисунок 4. Зависимость значений коэффициентов *CR* и *CL*от числа ядер в узле для сегмента МВС-10П ОП

Помимо оценки производительности для системы МВС-10П проведено тестирование энергоэффективности, в результате которого определена зависимость энергопотребления от тактовой частоты при выполнении программы численного решения гиперболических уравнений. Тестирование производилось на 128 узлах суперкомпьютера МВС-10П ОП (Broadwell). Наименьшее потребление энергии достигнуто при тактовой частоте 2,2 ГГц. Снижение тактовой частоты ниже номинальной (2,6 ГГц) приводит к значительному сокращению энергопотребления (рисунок 5). Полученный результат позволяет производить управление бюджетом энергопотребления при выполнении пользовательских задач.

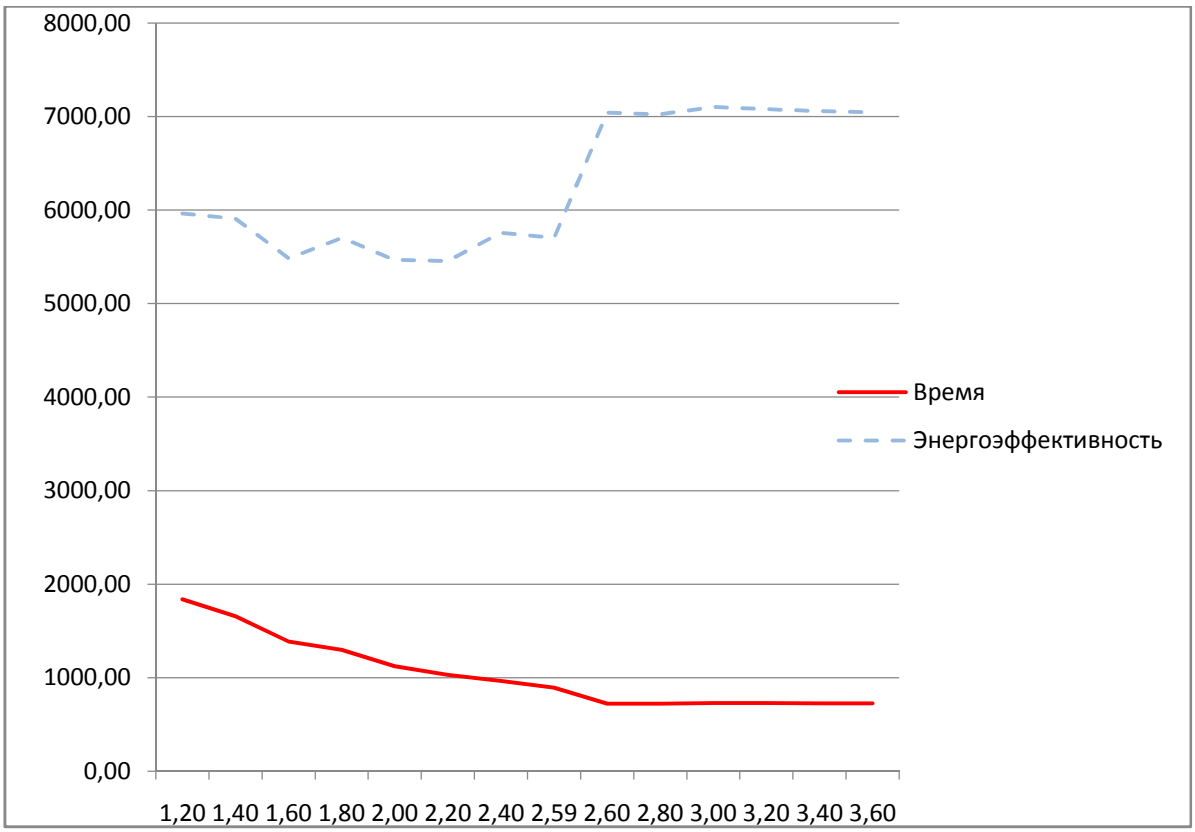

Рисунок 5. Зависимость времени выполнения (секунды) и энергопотребления (Вт·ч) от тактовой частоты процессора при выполнении программы решения гиперболических уравнений

В разделе 2.4 рассмотрена реальная производительность вычислительных систем.

При создании гетерогенных систем возникает вопрос достижения оптимального соотношения мощных универсальных процессоров и ускорителей. В случае, когда прикладная программа не оптимизирована для использования ускорителей, пользователи, как показывает опыт, будут запускать программу только на универсальных процессорах, и ускорители будут простаивать. Если программа оптимизирована для использования ускорителей, возникает обратная ситуация – основные вычисления производятся именно на ускорителях, и универсальные процессоры остаются недогруженными.

Системы, содержащие ускорители, обеспечивают увеличение производительности при выполнении только части задач. Для решения широкого круга задач требуются и востребованы универсальные процессоры. В настоящее время пользователи предпочитают использовать решения с системой команд x86, так как это обеспечивает переносимость прикладных программ. Что касается использования универсальных процессоров и ускорителей в одном узле, то результаты моделирования и опыт реальной работы пользователей показали, что вместо одновременного использования универсальных процессоров и ускорителей предпочтительно использовать один класс обрабатывающих устройств. Таким образом, целесообразно иметь 2 вида сегментов вычислительной системы: один, состоящий полностью из универсальных процессоров, для задач, которые плохо соответствуют архитектуре ускорителей, и сегмент, узлы которого содержат несколько сопроцессоров и универсальные процессоры относительно небольшой вычислительной мощности для организации вычислений.

Реальная производительность при выполнении задач пользователей существенно ниже пиковой производительности вычислительных систем. Если для теста Linpack эффективность, как правило, находится в диапазоне 65-85%, то для реальных задач это

значение значительно меньше. Это связано с тем, что такие тесты, как Linpack, оптимизированы производителями оборудования, в первую очередь, из рекламных соображений. На рисунке 6 представлены максимальные полученные значения эффективности использования вычислительных систем для различных программ при вычислениях над вещественными числами двойной точности (64 бита) для тестов Linpack, синтетического теста Matmul и пользовательских тестов Gromacs, Jet (газовая динамика), Nano (наноэлектроника) для систем, представленных в таблице 2. С проблемой эффективности выполнения программ сталкиваются все суперкомпьютерные центры.

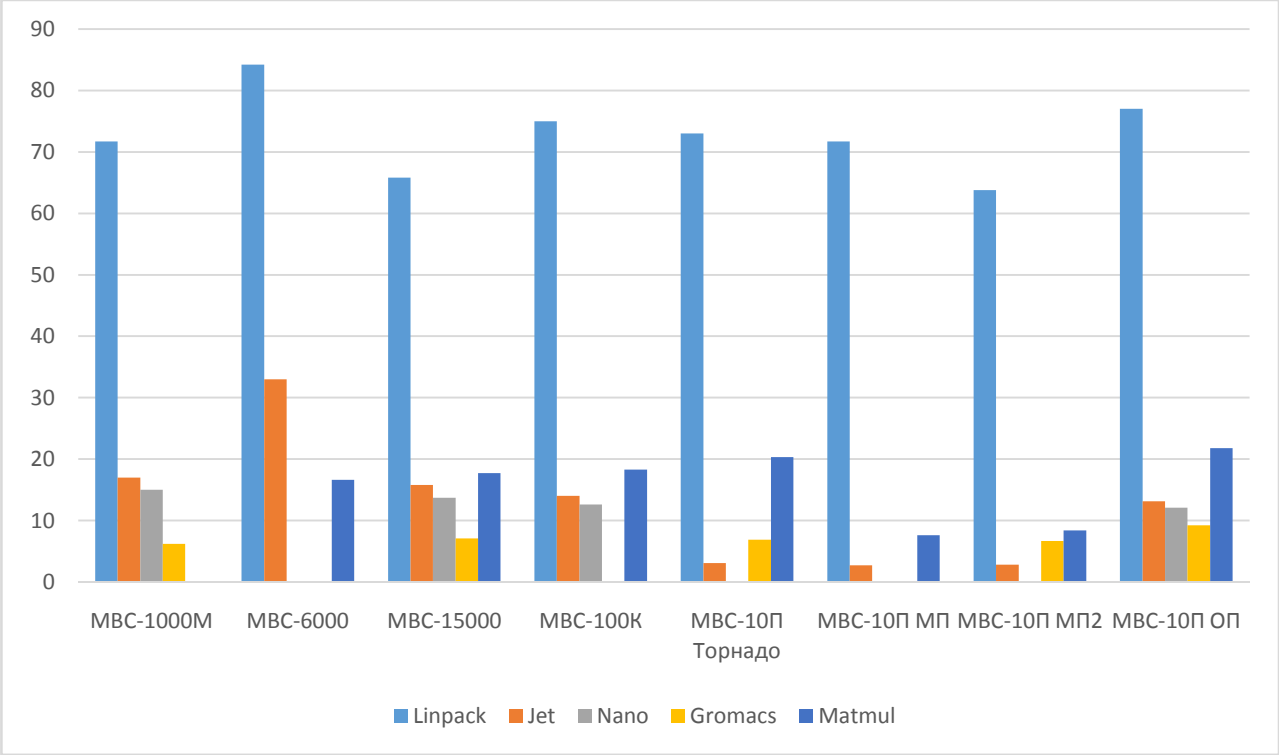

Рисунок 6. Эффективность использования вычислительных систем (%)

Поскольку при выполнении пользовательских программ производительность современных многоядерных платформ значительно ниже пиковой, требуется рассмотрение альтернативных архитектур процессоров.

В третьей главе предложен и исследован один из возможных путей преодоления недостатков вычислительных систем классической архитектуры – переход на использование процессоров с архитектурой управления потоком данных (потоковые процессоры).

В разделе 3.1 рассмотрены основные недостатки процессоров классической архитектуры – снижение производительности из-за недостаточного быстродействия памяти и сложность повышения производительности одного процессорного ядра, что приводит к увеличению накладных расходов на синхронизацию и обмен данных в многопроцессорных системах. Повысить в несколько раз производительность одного процессорного ядра и поддерживать ее на высоком уровне при увеличении времени обращения к памяти можно за счет перехода на архитектуру управления потоком данных.

В отличие от лучших процессоров традиционной архитектуры, которые способны выполнять не более 4–6 команд за такт при поиске команд с готовыми операндами в окне из примерно 200 команд, векторный потоковый процессор (ВПП) способен выполнять до 16 команд за такт. За счет поиска команд по готовности операндов в пределах окна из 20000 команд ВПП может компенсировать большие задержки обращения к памяти и межпроцессорного обмена, что должно обеспечить многопроцессорным системам на его

основе значительно более высокую реальную производительность по сравнению с системами традиционной архитектуры.

В разделе 3.2 описан принцип работы и рассмотрены проблемы создания потокового процессора. Программа в потоковом процессоре представляет собой граф, узлами которого являются команды, а информация по дугам передается в виде токенов, содержащих поле данных (значение операнда) и поле контекста. Этот контекст определяет, куда должен быть отправлен токен, он содержит номер команды приемника в графе программы, а также номера итераций вложенных циклов и запуска процедур, которые также должны совпадать у команд, выдаваемых на выполнение. Это позволяет выполнять различные итерации циклов и запуски процедур на одном и том же графе программы, хотя и усложняет реализацию памяти, которая обеспечивает поиск команд с готовыми операндами в потоковом процессоре, поскольку она должна выполнять ассоциативный поиск. Любая команда в графе выдается на исполнение по прибытию на ее входы последнего из токенов со значениями операндов. После вычисления результата в исполнительном устройстве (ИУ) формируются новые токены со значением результата, которые отправляются на входы последующих команд согласно графу программы, а использованные токены операндов уничтожаются. Тем самым в потоковом процессоре работает принцип единственного присваивания, при котором выдача команды на выполнение определяется лишь наличием операндов на ее входах.

Приводится перечень недостатков, препятствовавших успешной реализации потоковых процессоров. Главными из них являются сложность реализации ассоциативной памяти (АП) большой емкости, осуществляющей поиск команд с готовыми операндами в потоковом процессоре, и в 2-3 раза большее число выполняемых команд при обработке массивов данных. Кроме того, потоковый процессор имеет в 2-3 раза меньшую производительность на чисто последовательном коде и может реализовать свое преимущество в производительности лишь при наличии в программе достаточного уровня параллелизма.

В ВПП так же, как и в Манчестерском проекте потокового процессора (MDFM), для хранения массивов данных используется память структур данных на основе обычной линейно адресуемой памяти. В MDFM при активации нового программного блока «глобальный распределитель» выделяет место в памяти структур данных для записи данных из этого блока. Произвольная длина создаваемых массивов обусловила сложность алгоритма работы «глобального распределителя» в MDFM и его медленную работу, что, в свою очередь, привело к большому размеру активируемых им параллельных блоков.

В разделе 3.3 приводятся особенности реализации ВПП, его структурная схема (рисунок 7) и результаты моделирования.

Основное отличие ВПП от MDFM – в устройстве распределения памяти структур данных, которое выделяет память для записи результата каждой векторной команды фрагментами постоянной длины, равной аппаратной длине вектора VL_{max}=256 слов. Это позволяет реализовать быстрое аппаратное распределение памяти на основе ведения списков свободных векторов и значительно снизить инерционность системы авторегулирования уровня нагрузки в процессоре. Так на программе умножения матриц размером 128×128 суммарное число команд в буферах на входе ИУ в ВПП не превышает 25 команд, ожидающих в АП готовности операндов, – 550, что на 2 порядка меньше по сравнению с MDFM. Память структур данных в ВПП имеет два уровня – память векторов (ПВ) большой емкости, реализованную на микросхемах динамической памяти, и быструю локальную память векторов (ЛПВ) значительно меньшей емкости на процессорном кристалле.

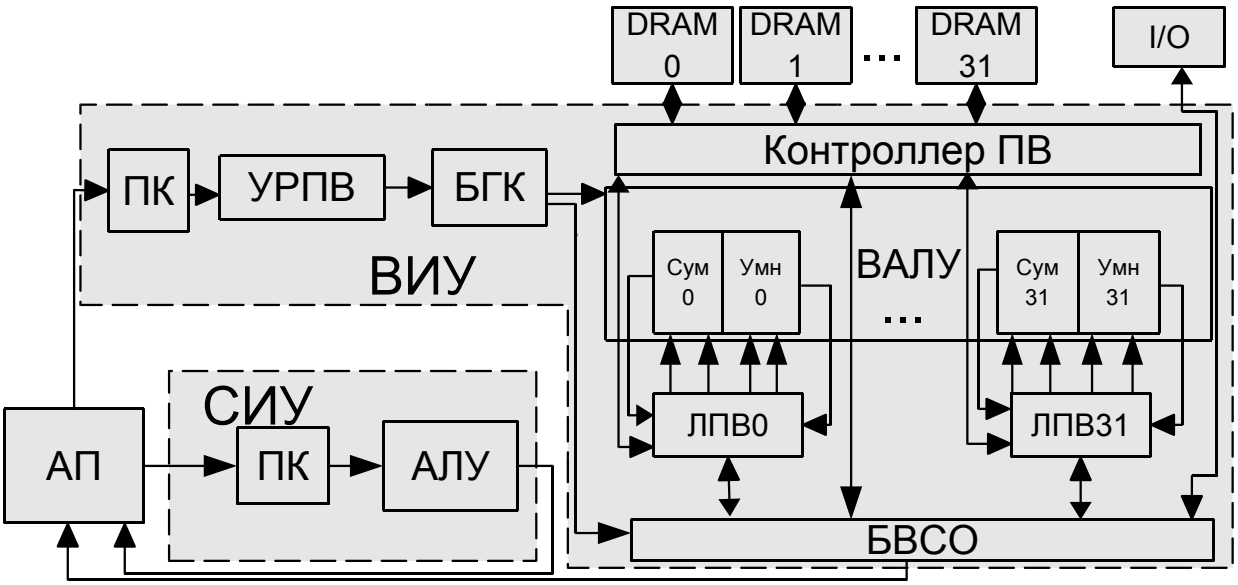

Рисунок 7. Структурная схема ВПП

Блок обработки векторных команд, обозначенный на рисунке 7 как векторное исполнительное устройство (ВИУ), содержит два ИУ: векторное АЛУ (ВАЛУ) для выполнения векторных арифметических команд и блок выполнения специальных операций (БВСО). БВСО может работать одновременно с ВАЛУ, выполняя команды чтения и записи одиночных элементов вектора, команды редукции, среди которых вычисление суммы элементов вектора и поиска максимального (минимального) элемента, а также команды сбора и распределения элементов вектора. Наконец, как векторное АЛУ, так и БВСО могут читать и записывать элементы вектора не только в быстродействующей ЛПВ, но и в ПВ большой емкости, реализованной на микросхемах динамической памяти, через находящийся на процессоре контроллер ПВ.

Входящие в состав ВИУ ВАЛУ и БВСО так же, как и скалярное ИУ (СИУ), получают готовые для выполнения команды (пары токенов операндов) из АП и туда же пересылают токены с результатами выполненных команд. В ВПП вся необходимая для выполнения команды и формирования токенов результата информация выбирается из памяти команд (ПК). В отличие от выполнения скалярных команд, векторные команды сначала получают адрес для записи вектора результата в ПВ или ЛПВ из устройства распределения памяти векторов (УРПВ), которое аппаратно распределяет адресное пространство ПВ и ЛПВ, и лишь затем поступают для выполнения в ВАЛУ или БВСО через буфер готовых команд (БГК).

При аппаратном распределении памяти в ВПП большие массивы хранятся в виде векторов указателей, то есть векторов, элементами которых являются указатели векторов подмассивов. Это увеличивает время доступа к произвольному элементу массива, однако дает возможность выполнять одинаковые операции над всеми элементами вектора указателя, например, над строками, содержащимися в векторе указателе матрицы. Тем самым удается значительно сократить адресные и другие скалярные вычисления и, следовательно, повысить степень векторизации и производительность векторной обработки в ВПП. Для иллюстрации рассмотрим программу умножения матриц A и B, в которой основной объем вычислений приходится на подпрограмму SGEMV по вычислению элементов j-го столбца матрицы результата C:

```
DO k=1, N
C(1:N,j) = C(1:N,j) + A(1:N,k) * B(k,j)
ENDDO
```

Граф рассматриваемой подпрограммы для ВПП при использовании метода векторов указателей показан на рисунке 8 в предположении, что размер матрицы $N<VL_{max}$.

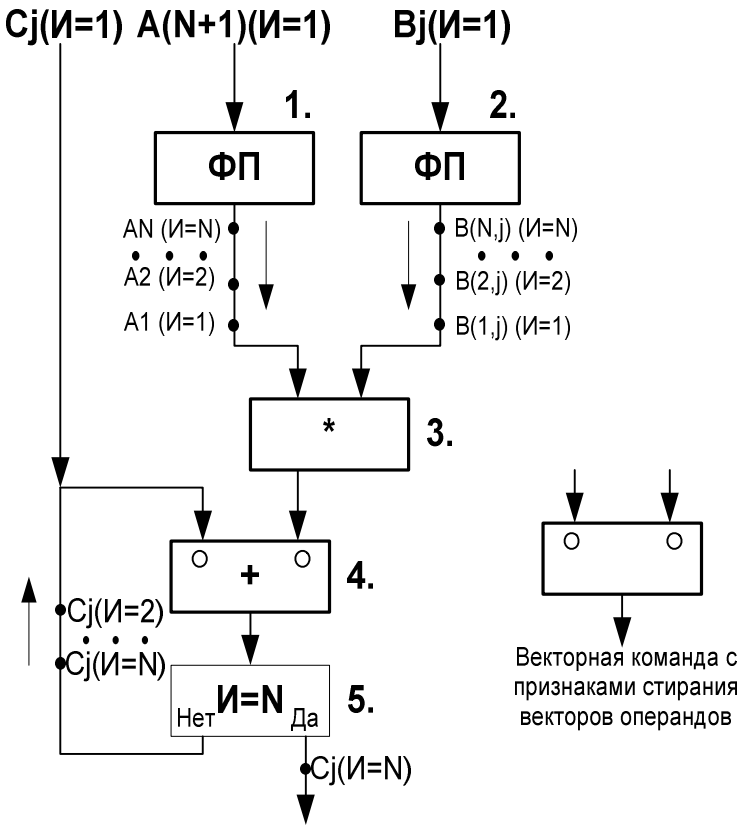

Рисунок8. Граф подпрограммы SGEMV для выполнения в ВПП

На вход графа поступают подлежащий вычислению указатель вектора столбца Cj с нулевыми значениями элементов, указатель матрицы A – вектор A(N+1), содержащий указатели векторов столбцов A1,A2,... AN, и указатель j-го столбца матрицы B – Bj. Для возможности параллельного выполнения команд из разных итераций цикла контекст токена в ВПП имеет поля индекса (И), итерации (Т) и подпрограммы (П). Поле И содержит номер итерации k, поэтому токены указателей, приходящие на вход графа, имеют поле И=1. Команды 1 и 2 «Формирование Потока» (ФП) в графе на рисунке 8 читают из памяти элементы вектора указателя A(N+1) матрицы A и вектора столбца Bj соответственно, формируя последовательность токенов со значениями элементов A1,A2,... AN и B(1,j),B(2,j),... B(N,j) для всех итераций цикла, выполняемого подпрограммой. Поэтому в каждой из N итераций цикла в ВПП выполняются лишь две векторные команды с плавающей запятой (команды 3 и 4 на рисунке 8) и одна скалярная команда (команда 5). Команда 5 сравнивает номер текущей итерации k (поле И) в токене вектора Cj с длиной вектора N и осуществляет выход из цикла при k=N, а при невыполнении этого условия увеличивает на 1 значение И в токене результата для передачи Cj на вход команды сложения 4 в следующей итерации цикла. Признаки стирания на входах команды 4 указывают, что оба вектора используются в качестве операнда последний (единственный) раз и после выполнения команды должны быть отправлены в список свободных векторов.

При выполнении подпрограммы SGEMV в ВПП и в векторном процессоре традиционной архитектуры каждая итерация цикла состоит из 3 и 8 команд соответственно. Использование команд ФП позволило исключить в ВПП одиночные команды обращения к памяти по чтению векторов Ak и элементов B(k,j), команды вычисления адресов чтения векторов Ak, а также увеличения индекса, необходимые в традиционном процессоре. Тем

самым в ВПП удается уменьшить число выполняемых команд в 2,7 раза и фактически векторизовать не только внутренний, но и следующий из вложенных циклов в программе.

При моделировании ВПП считалось, что время выборки из ПВ и ЛПВ равно 100 и 10 тактам соответственно. Показано, что производительность одного ядра ВПП может быть повышена до 256–512 флоп в такт. При пиковой производительности 256 и 512 флоп в такт реальная производительность ВПП на блочном варианте программы умножения матриц достигает 95% и 83% от пиковой производительности соответственно. Заметим, что производительность одного ядра процессора Intel Skylake составляет 32 флоп в такт, и у векторного процессора NEC SX-ACE, который так же, как и ВПП, обрабатывает векторы длиной 256 слов – 64 флоп в такт. При этом ВПП имеет еще одно важное преимущество перед процессорами традиционной архитектуры – способность сохранять значительно более высокую производительность при уменьшении размера обрабатываемых матриц. При росте пиковой производительности ВПП или системы на традиционных процессорах происходит увеличение размера матрицы N1/2, на котором достигается производительность, равная половине от пиковой производительности. Однако для ВПП с пиковой производительностью 128 и 256 флоп за такт N1/2 составляет примерно 60 и 100, а для систем на Intel Xeon E5 с пиковой производительностью 64 и 128 флоп в такт N1/2 составляет 1500 и 2500 соответственно. Такая разница объясняется эффектом «холодного кэша» у процессора традиционной архитектуры, когда при уменьшении объема вычислений в выполняемой программе процессор начинает резко терять производительность, в первую очередь, из-за недостаточного накопления данных в кэш-памяти, что приводит к частым промахам и падению производительности.

Показана более высокая производительность ВПП на программах сортировки и решения систем дифференциальных уравнений 2D Stencil по сравнению с процессорами Intel Xeon. На параллельных алгоритмах сортировки, таких как битонная сортировка и сортировка с использованием команд подсчета совокупностей, производительность ВПП оказалась в 3 – 12 раз выше. В то же время на чисто последовательных алгоритмах, таких как сортировка слиянием, ВПП уступает в производительности современным процессорам до 3 раз. Здесь в ВПП проявляется известный недостаток потоковой архитектуры, заключающийся в низкой производительности выполнения последовательных команд. В программе битонной сортировки параллельная сортировка 16 векторов по 256 элементов в ВПП приводит к увеличению времени прохождения одной ступени в сети, однако время сортировки одного вектора из 256 элементов уменьшается почти в 9 раз, и производительность оказывается выше, чем в Intel KNL в 3,2 раза. Показано, что при числе сортируемых элементов более 256 увеличение времени выборки из памяти с 10 до 100 тактов практически не приводит к снижению производительности ВПП.

Показано на тестовых программах, что одно ядро ВПП обеспечивает значительно более высокую производительность по сравнению с процессорным ядром традиционной архитектуры на программах с высокой долей вычислений, выполняемых с помощью векторных команд. Из числа рассмотренных программ это умножение матриц, битонная сортировка и 2D Stencil. Причем во всех этих программах известный недостаток потоковых процессоров, заключающийся в низкой производительности при выполнении цепочек из последовательных команд, компенсируется возможностью параллельного выполнения мелких блоков кода, таких как итерации вложенных циклов.

В программе умножения матриц такое распараллеливание итераций вложенных циклов производится аппаратно без переделки программного кода. В программе битонной сортировки необходимо введение в граф программы дополнительных команд, чтобы выполнять параллельно слияние не одной, а нескольких пар векторов. Кроме того, в ВПП за счет возможности одновременного выполнения мелких блоков можно распараллелить программу с чисто скалярной обработкой, что сложно осуществить в традиционном

процессоре. Помимо программ с высокой степенью векторизации ВПП также имеет преимущество перед современными процессорами на программах с мелкоструктурным и нерегулярным параллелизмом.

Еще одним классом программ, позволяющих повысить реальную производительность ядра ВПП без увеличения скорости работы памяти, являются программы, в которых элементы массива, читаемые из памяти, обрабатываются цепочкой из нескольких ИУ до получения результата, который записывается в память. Чем длиннее оказывается цепочка из ИУ в построенном графе, тем большее число арифметических операций выполняется по отношению к числу слов читаемых и записываемых в память на начальных и конечных вершинах графа.

В разделе 3.4 приводятся результаты исследования системы из нескольких ядер ВПП с общей памятью и рассматриваются перспективы ее практического применения. Согласно предварительной оценке при 22 нм технологии изготовления на кристалле СБИС площадью 450 мм^2 возможно разместить многопроцессорную систему из 4 ядер ВПП с суммарной производительностью 1024 флоп в такт. Модель ядра ВПП была дополнена схемой разрешения конфликтов при обращении к общим ресурсам системы, таким как контроллеры динамической памяти процессорного кристалла для доступа к ПВ и к устройству распределения адресов этой памяти. Кроме того, для распределения задач по процессорам был разработан граф управляющей программы, которая ведет список свободных ядер в многопроцессорной системе и формирует токены с исходными данными для запуска очередного процесса на выполнение. Результаты моделирования многопроцессорной системы из нескольких ядер ВПП подтвердили, что при увеличении числа ядер в системе можно получить близкое к линейному ускорение времени выполнения программы перемножения матриц, поскольку на этой программе пропускная способность к ПВ не ограничивает рост производительности процессора. Моделирование также показало, что несколько программ, такие как блочное умножение матриц и управляющая программа, могут выполняться на одном ядре ВПП с максимальной производительностью, так как они задействуют разные ИУ.

ВПП является универсальным процессором, но ближайшей перспективой его практического применения следует рассматривать использование в качестве ускорителя к процессору традиционной архитектуры, по аналогии с графическими ускорителями.

В четвертой главе представлены созданные для МСЦ РАН суперкомпьютеры различной архитектуры, а также определены направления развития научного вычислительного центра.

В разделе 4.1 представлены суперкомпьютеры МВС-1000М, МВС-15000ВМ, МВС-6000IM, МВС-100К и МВС-10П и описано применение метода выбора вычислительных систем.

МВС-1000М. Создание системы МВС-1000М преследовало цель выхода на уровень производительности в 1 ТФлопс. Для этого были предложены следующие принципиально новые технологические решения:

- переход с массово-параллельной на кластерную архитектуру суперкомпьютера;
- применение новейших высокопроизводительных процессоров Alpha-21264;
- применение в качестве коммуникационной среды низколатентной сетевой технологии Myrinet;
- использование полнофункциональной ОС Linux и развитой среды разработки параллельных программ;
- разработка и применение развитой подсистемы коллективного доступа пользователей, основанной на СУППЗ.

Следует особо отметить, что предшествовавшие массово-параллельные системы предлагали достаточно ограниченные возможности по разработке, отладке и выполнению

параллельных приложений в POSIX-совместимой операционной среде, что существенно ограничивало круг потенциальных пользователей.

МВС-1000М стал первым российским научным суперкомпьютером, преодолевшим терафлопсный рубеж производительности и вошедшим в первую сотню списка TOP500 (64-е место). Система МВС-1000М содержала 384 двухпроцессорных узла на базе процессоров Alpha-21264 и имела производительность 1,024 ТФлопс.

МВС-1000М применяли в исследованиях 482 пользователя из 81 организации, выполнившие 404 научных проекта. В системе МВС-1000М было обработано свыше 300 тыс. пользовательских заданий. При замене на систему следующего поколения в целях продления жизненного цикла суперкомпьютер МВС-1000М был демасштабирован и разбит на сегменты, которые были переданы в филиал МСЦ РАН в г. Казани, а также в вычислительные центры ИММ УрО РАН (г. Екатеринбург) и ИВМиМГ СО РАН (г. Новосибирск).

МВС-15000ВМ. Создание системы МВС-15000ВМ преследовало цель выхода на уровень производительности свыше 10 ТФлопс. Для этого были предложены и использованы следующие принципиально новые технологические решения:

- конструктив на базе блейд-серверов, обеспечивший компактность системы и возможность концентрации большой вычислительной мощности на ограниченной площади помещения;
- применение высокопроизводительных процессоров IBM PowerPC 970FX;
- использование параллельной файловой системы GPFS, обеспечившей сверхоперативный доступ пользователей к данным.

Впервые в Европе для суперкомпьютеров было предложено решение на блейд-серверах IBM JS20, каждый из которых содержал два двухъядерных процессора упрощенной архитектуры IBM Power, что обеспечило компактность и низкое энергопотребление. В качестве коммуникационной среды использовалась низколатентная сеть Myrinet 2000. Кластер МВС-15000ВМ с пиковой производительностью 10,1 ТФлопс в 2005 году занял 56-е место в списке TOP500. В его состав входили 574 двухпроцессорных узла на базе процессоров IBM PowerPC 970FX. Заложенные технологические решения обеспечивали возможность масштабирования системы до уровня 30 ТФлопс.

При создании суперкомпьютера МВС-15000ВМ использовался метод выбора вычислительных систем по набору факторов. Выбор системы осуществлялся из следующих вариантов: IBM JS-20, IBM Regatta p655, IBM BlueGene/L, Dell Power Edge, серверы Supermicro. Последующее тестирование вычислительных систем показало преимущество блейд-серверов IBM JS-20.

МВС-15000ВМ применяли в исследованиях 639 пользователей из 107 организаций, выполнившие 481 научный проект. В системе было обработано свыше 500 тыс. пользовательских заданий. При замене на систему следующего поколения в целях продления жизненного цикла суперкомпьютер МВС-15000ВМ был демасштабирован и разбит на сегменты, которые были переданы в филиалы МСЦ РАН в гг. Санкт-Петербурге и Казани, а также в ИАПУ ДВО РАН (г. Владивосток) и ИПХФ РАН (г. Черноголовка).

МВС-6000IM. Система МВС-6000IM представляла диапазон производительности до 10 ТФлопс. В этой системе были опробованы следующие принципиально новые технологические решения:

- применение функционально сложных процессоров Intel Itanium-2, построенных на базе архитектуры с явным параллелизмом команд (EPIC);
- использование управляющих серверов суперкомпьютера с архитектурой, отличной от архитектуры вычислительных модулей, что потребовало организации процесса кросс-компиляции на сервере доступа.

Пиковая производительность суперкомпьютера МВС-6000IM составила 1,54 ТФлопс. Система заняла 412-е место в рейтинге TOP500. Заложенные технологические решения обеспечивали возможность масштабирования системы до уровня 10 ТФлопс.

При создании суперкомпьютера МВС-6000IM рассматривались следующие варианты: HP Integrity, IBM Regatta, Alpha Server. Тестирование показало преимущество HP Integrity.

Суперкомпьютер применяли в исследованиях 431 пользователь из 72 организаций, выполнившие 129 научных проектов. В системе было обработано свыше 100 тыс. пользовательских заданий.

МВС-100К. Создание системы МВС-100К преследовало цель выхода на уровень производительности диапазона 0,1 – 1 ПФлопс. Для этого были предложены и использованы следующие принципиально новые технологические решения:

- конструктив на базе блейд-серверов, обеспечивший эффективное воздушное охлаждение системы и возможность подведения мощности в 40 кВт на стойку;
- применение высокопроизводительных процессоров IntelXeon новой микроархитектуры;
- использование высокоскоростной низколатентной сети InfiniBand;
- двухуровневая топология сети, обеспечившая высокую масштабируемость системы;
- использование графических ускорителей в составе части вычислительных модулей;
- применение высокопроизводительной и высоконадежной кластерной СХД NetApp FAS3140;
- разделение системы на связанные сегменты и постепенное наращивание ее производительности за счет добавления новых сегментов.

Для кластерной вычислительной системы МВС-100К были предложены и использованы высокоэффективные процессоры Intel Xeon семейств Harpertown, Clovertown, Westmere на базе блейд-серверов HP ProLiant и низколатентная сетевая технология InfiniBand DDR. Система МВС-100К заняла 33-е место в списке TOP500. В дальнейшем система была расширена новыми узлами, в том числе с ускорителями Nvidia Tesla M2090, с использованием сетевой технологии InfiniBand QDR. Заложенные технологические решения обеспечивали возможность масштабирования системы до уровня 800 ТФлопс.

При выборе оборудования учитывалось широкое распространение и востребованность у пользователей x86-совместимой архитектуры. Были рассмотрены следующие альтернативы: HP ProLiant BL460, IBM x3550, IBM BladeCenter LS21, в результате было принято решение о выборе платформы HP ProLiant BL460, которая получила развитие с удвоением плотности компоновки.

МВС-100К применяли в исследованиях 1078 пользователей из 141 организации, выполнившие 1146 научных проектов. В системе МВС-100К было обработано свыше 1,8 млн. пользовательских заданий. Для продления жизненного цикла в рамках демасштабирования отдельные сегменты суперкомпьютера МВС-100К были переданы в филиалы МСЦ РАН в гг. Санкт-Петербурге и Казани, в ИПХФ РАН (г. Черноголовка). Система МВС-100К оказалась настолько востребована пользователями, что полученные в процессе демасштабирования сегменты используются по настоящее время.

МВС-10П. Создание и развитие системы МВС-10П преследовало цель выхода на уровень производительности диапазона 1–10 ПФлопс. Для этого были предложены и использованы следующие принципиально новые технологические решения:

- конструктив, обеспечивший водяное охлаждение по инновационной отечественной технологии;
- переход с охлаждения «холодной водой» на охлаждение «горячей водой», обеспечивавший существенное повышение энергоэффективности;

- применение новейших высокопроизводительных процессоров Intel Xeon и x86-совместимых ускорителей Intel Xeon Phi;
- использование высокоскоростной низколатентной сети Intel Omni-Path;
- совместное применение систем управления заданиями SLURM и СУППЗ.

Гетерогенная вычислительная система МВС-10П, одна из первых в мире с водяным охлаждением, содержала помимо обычных процессоров сопроцессоры с большим числом ядер Intel Xeon Phi. Суперкомпьютер МВС-10П занимал 59-е место в списке TOP500. С 2016 г. в этой системе впервые в России стали применяться сетевая технология Intel Omni-Path и охлаждение «горячей водой». Последнее позволило достичь значения коэффициента энергоэффективности PUE 1,06, в то время как среднее значение этого коэффициента для центров обработки данных превышает 1,5.

С 2017 г. по настоящее время система МВС-10П развивается в виде отдельных сегментов с использованием двух архитектур: с экстрамассивным (manycore) параллелизмом для выполнения специально оптимизированных пакетов программ и с массивным (multicore) параллелизмом для выполнения широкого спектра программ. Заложенные в системе МВС-10П технологические решения обеспечивают возможность масштабирования системы до уровня 10 ПФлопс, а с использованием ускорителей – до 20 ПФлопс.

При выборе системы рассматривались следующие альтернативные предложения отечественных и зарубежных производителей: РСК Торнадо, IBM iDataPlex DX360M3, IBM X360M4, HPE SL250s. С ростом числа ядер в микропроцессорах важным аспектом стало увеличение эффективности теплоотвода в узле. Для системы МВС-10П после рассмотрения предложенных решений была определена необходимость использования водяного охлаждения. В результате было принято решение о выборе блейд-серверов РСК Торнадо.

В настоящее время суперкомпьютер МВС-10П содержит пять сегментов, ориентированных на решение разных классов задач с использованием как массивного, так и экстрамассивного параллелизма. Структурная схема системы приведена на рисунке 9.

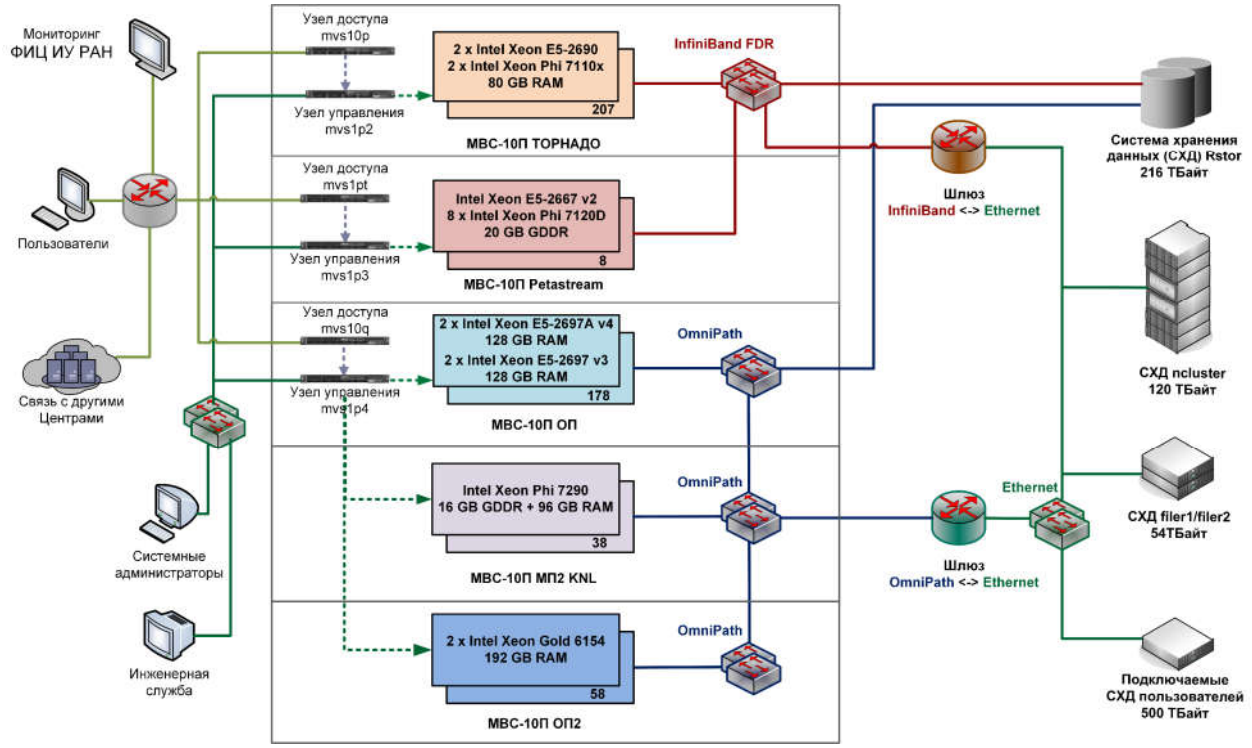

Рисунок 9. Структурная схема системы МВС-10П

Систему используют 692 пользователя из 129 организаций, за время эксплуатации реализовано 714 научных проектов. В системе МВС-10П выполнено свыше 500 тыс. пользовательских заданий. Объем заявок на предоставление ресурсов системы составил свыше 487 млн. ядро-часов в год.

Общие данные по разработанным вычислительным системам приведены в таблице 2.

Таблица 2. Характеристики суперкомпьютеров серии МВС

Система	Процессоры, сопроцессоры, ускорители	Коммуникационная среда	Количество узлов / ядер	Производительность пик / Linpack, ТФлопс
МВС-1000М, (2001 г.)	HP Alpha 21264A	Myrinet 2000, 2 Гбит/с	384 / 768	1,176 / 0,734
МВС-15000ВМ (2003 г.)	IBM PowerPC 970	Myrinet 2000, 2 Гбит/с	576 / 2296	10,102 / 6,680
МВС-6000IM (2005 г.)	Intel Itanium 2	Myrinet 2000, 2 Гбит/с	128 / 256	1,536 / 1,293
МВС-100К (2009 г.)	Intel Xeon E5450, Intel Xeon X5365, Intel Xeon X5670, Intel Xeon X5675, ускорители: NVIDIA Tesla M2090	InfiniBand DDR/QDR, 20/40 Гбит/с	1275 / 13004	227,9 / 119,9
МВС-10П Торнадо (2013 г.)	Intel Xeon E5-2690, сопроцессоры: Intel Xeon Phi 7110X	InfiniBand FDR, 56 Гбит/с	208 / 28704	523,83 / 383,21
МВС-10П МП (2014 г.)	Intel Xeon E5-2667v2, сопроцессоры: Intel Xeon Phi 7120D	InfiniBand FDR, 56 Гбит/с	64 / 3904	77,33 / 53,51
МВС-10П ОП (2016 г.)	Intel Xeon E5-2697Av4, Intel Xeon E5-2697v3	Intel Omni-Path, 100 Гбит/с	178 / 5528	229,96 / 171,89
МВС-10П МП2 (2017 г.)	Intel Xeon Phi 7290	Intel Omni-Path, 100 Гбит/с	38 / 2736	131,33 / 83,91
МВС-10П ОП2 (2018 г.)	Intel Xeon Gold 6154	Intel Omni-Path, 100 Гбит/с	58/2088	200,45 / 132,91

В перечень основных научных достижений Российской академии наук в соответствующие годы вошли созданные вычислительные системы и полученные с их использованием результаты фундаментальных научных исследований.

В разделе 4.2 представлена грид-инфраструктура для суперкомпьютерных приложений (РИСП), созданная в Российской академии наук на базе кластерных систем Межведомственного суперкомпьютерного центра. В грид-инфраструктуре обеспечены динамическое выделение ресурсов по принципу наименьшей загруженности, общая очередь заданий, защита от несанкционированного доступа, единая точка доступа к ресурсам грид. РИСП объединила вычислительные системы МВС-10П, а также отдельные сегменты

вычислительных систем МВС-15000ВМ и МВС-100К, размещенные в МСЦ РАН (г. Москва) и его филиалах в Санкт-Петербурге и Казани.

Развитием РИСП является проект распределенной сети суперкомпьютерных центров коллективного пользования. Проект предполагает объединение ресурсов разных суперкомпьютерных центров в соответствии с принципами федеративности и децентрализации управления, унификации доступа пользователей к суперкомпьютерным ресурсам, унификации представления информации в системе мониторинга.

В разделе 4.3 приведен анализ использования суперкомпьютерных ресурсов.

Загрузка суперкомпьютеров МСЦ РАН находится на уровне 90% и выше. Анализ эффективности использования суперкомпьютерных ресурсов отразил следующие аспекты:

а) Использование ресурсов активными пользователями, потребляющими более 2% доступных ресурсов суперкомпьютера на каждый проект. Активные пользователи расходуют от 67% до 75% всех ресурсов.

б) Использование ресурсов пользователями, на каждый проект которых расходуется менее 2% доступных ресурсов суперкомпьютера. Несмотря на то, что таких пользователей более 75%, суммарный расход ими ресурсов не превышает 15%.

в) Использование ресурсов сотрудниками МСЦ РАН для разработок, тестирования, проведения работ по модернизации и профилактике оборудования – около 11%.

г) Потери, складывающиеся из следующих составляющих:

- потери по причине аварий систем электропитания и охлаждения;
- потери, связанные с переводом управления очереди заданий в ручной режим (как правило, по причине устранений последствий аварий и отказов инфраструктуры);
- потери, связанные с ремонтом вычислительных модулей.

Анализ статистики позволяет сделать вывод, что переход на более эффективное охлаждение «горячей водой» на суперкомпьютере МВС-10П ОП позволил более чем в 2 раза снизить потери, связанные с авариями и ремонтом оборудования. Полученный эффект был достигнут в том числе за счет автономизации основных инфраструктурных систем.

Анализ статистики работы суперкомпьютера МВС-10П также показал, что доля заданий с низкой степенью параллелизма, использующих 128 процессоров и менее, в сегменте МВС-10П Торнадо составляет более 42%, а на МВС-10П ОП – менее 24%. При этом доля заданий с высокой степенью параллелизма (использующих более 512 процессоров) на суперкомпьютере МВС-10П ОП (40,4%) существенно выше, чем на Торнадо (23,2%). Это объясняется большим количеством процессорных ядер в МВС-10П ОП на одном вычислительном модуле и современной средой разработки параллельных программ, позволяющей пользователям лучше масштабировать свои приложения.

Следует отметить, что режим эксплуатации суперкомпьютеров предусматривает еженедельную профилактику оборудования, во время которой вычислительные системы проходят всестороннее тестирование и обновление программного обеспечения. По окончании профилактики пользователям предоставляется возможность выполнения заданий с высокой степенью параллелизма, задействующих весь доступный объем суперкомпьютерных ресурсов.

В разделе 4.4 приведены направления развития научного суперкомпьютерного центра.

Спрос на вычислительные ресурсы МСЦ РАН постоянно возрастает, о чем свидетельствуют заявки пользователей, согласно которым суммарный запрашиваемый объем ресурсов более чем в 16 раз превосходит объем имеющихся в центре ресурсов. Среднее время ожидания задания в очереди находится на уровне 15 часов.

Среди основных направлений развития научного суперкомпьютерного центра отмечены:

- увеличение объемов и номенклатуры предоставляемых услуг по высокопроизводительным вычислениям за счет повышения производительности и разнообразия суперкомпьютерного оборудования;
- сокращение максимального времени ожидания результата выполнения задач;
- расширение круга исследователей, имеющих доступ к суперкомпьютерным ресурсам вычислительного центра;
- распространение передового опыта использования суперкомпьютерных технологий в исследованиях;
- привлечение молодых специалистов за счет возможности работы с новейшими суперкомпьютерными архитектурами;
- развитие международного сотрудничества в таких проектах, как DEISA и PRACE.

В разделе 4.5 рассмотрено построение межведомственного центра коллективного пользования в модели программно-определяемого центра обработки данных (ЦОД). ЦОД является прогрессивной формой предоставления вычислительных ресурсов, когда необходимо обеспечить обслуживание широкого круга пользователей. Рассматривается один из подходов к созданию ЦОД – концепция программно-определяемой инфраструктуры. Программно-определяемой является такая инфраструктура ЦОД, в которой ее ключевые элементы – вычислительные ресурсы, сеть, системы хранения данных виртуализованы и предоставляются пользователям как сервисы с заданными характеристиками.

ЦОД межведомственного уровня обеспечивает:

- выполнение запросов на исполнение потока задач, относящихся к разным классам;
- выполнение запросов пользователей разного уровня подготовки.

Современным подходом к созданию вычислительной инфраструктуры ЦКП, который позволяет удовлетворить представленным требованиям, является использование концепции программно-определяемого центра обработки данных.

Программно-определяемым является центр обработки данных, в котором все элементы вычислительной инфраструктуры (сеть, системы хранения данных, вычислительные ресурсы, приложения) виртуализованы и предоставляются как сервисы с заданными характеристиками.

Архитектуру программно-определяемого ЦОД можно разделить на три логических уровня: аппаратуры, виртуализации, управления.

Рассмотрены и конкретизированы общие требования к организации межведомственного ЦОД:

- развертывание вычислительных платформ межведомственного ЦОД, обеспечивающее максимальное использование аппаратуры;
- обеспечение выполнения программ из заданных классов задач на используемом оборудовании;
- предоставление пользователю выбора системного программного обеспечения.

Показано, что реализация предлагаемой инфраструктуры позволяет обеспечить возможность каждому пользователю продуктивно решать задачи за приемлемое время с приемлемым уровнем затрат. Описанный подход был успешно внедрен при создании вычислительной инфраструктуры программы «Университетский кластер», испытательного стенда международного проекта OpenCirrus и проблемно-ориентированной web-лаборатории решения задач механики сплошных сред UniCFD.

В заключении представлены основные результаты и выводы диссертации:

1. На основе разработанных в диссертации метода, технических и технологических решений создана серия суперкомпьютеров с различным набором характеристик и оригинальными архитектурными, сетевыми и программными решениями. Суперкомпьютеры обеспечили соответствующий мировому уровень развития отечественных высокопроизводительных вычислительных средств для науки и образования и составили основу интегрированной высокопроизводительной вычислительной среды проведения научных исследований и решения прикладных задач, позволившей поднять фундаментальные научные исследования в России на качественно новый уровень. Созданные и установленные в МСЦ РАН вычислительные системы наряду с предоставлением вычислительных услуг обеспечивают полнофункциональную среду проведения исследований и разработок в области суперкомпьютерных технологий, отработки архитектурных, технических, технологических решений по созданию и использованию высокопроизводительных вычислительных систем. Сформированная инфраструктура высокопроизводительных вычислений обеспечила доступ к суперкомпьютерам более чем 1400 исследователям, результаты работы которых опубликованы в более чем 5900 научных статьях (из них только в 2017-2018 г. 189 статей в 1-2 квартилях Web of Science и Scopus).

2. Разработанный метод построения вычислительных платформ для задач, решаемых вычислительным центром, связывает в единый комплекс такие факторы, как производительность, стоимость и доступность оборудования, возможность масштабирования и демасштабирования, востребованность у пользователей, продление жизненного цикла системы, стоимость и трудоемкость эксплуатации, возможности развития инфраструктуры, обеспечения глобально распределенной обработки данных. Разработанный метод позволяет создавать суперкомпьютерные системы для решения актуальных вычислительно сложных научных проблем и делает возможным выполнение оптимизации состава вычислительного центра.

3. Разработанный и реализованный комплекс решений по тестированию и анализу вычислительных систем позволяет определять влияние параметров на время выполнения программ в вычислительных кластерах с многоядерной архитектурой и используется для определения характеристик оборудования, необходимого при создании вычислительных систем для решения научных и технических задач.

4. Исследована и разработана оригинальная архитектура векторно-потокового процессора, которая обеспечивает увеличение производительности одного процессора до 10 раз по сравнению с процессорами традиционной архитектуры. Показано, что производительность векторно-потокового процессора устойчива к изменениям латентности памяти и межпроцессорного обмена в диапазоне изменения задержек, на порядок более широком по сравнению с системами традиционной архитектуры.

5. Разработанные архитектурные, технические и технологические решения составляют базовые основы построения и определяют направления развития сервисно-ориентированных программно-определяемых центров обработки данных межведомственного уровня как среды проведения научных исследований и инновационных разработок, формирования профессиональных компетенций специалистов в области суперкомпьютерных технологий.

Основные результаты, выводы и рекомендации, изложенные в диссертации, получены и использовались при реализации следующих национальных и международных проектов: программ фундаментальных научных исследований государственных академий наук на 2008-2012 годы и на 2013-2020 годы; программ фундаментальных исследований Президиума и Отделения математических наук РАН; проектов РФФИ и Минобрнауки России; научно-технической программы Союзного государства «Исследования и разработка

высокопроизводительных информационно-вычислительных технологий для увеличения и эффективного использования ресурсного потенциала углеводородного сырья Союзного государства» («СКИФ-НЕДРА»); международного проекта распределенной европейской инфраструктуры для суперкомпьютерных приложений DEISA-2; международного проекта создания решателей гиперболических уравнений для вычислительных систем экзафлопсного диапазона ExaHYPE; международного проекта облачных вычислений OpenCirrus; вычислительной инфраструктуры программы «Университетский кластер»; проблемно-ориентированной web-лаборатории решения задач механики сплошных сред UniCFD.

Результаты диссертационной работы могут быть использованы для исследований и разработки суперкомпьютерных систем, организации на их основе высокопроизводительных вычислений, создания и развития вычислительных центров и сетей вычислительных центров.

По материалам диссертации подготовлены и читаются лекционные курсы и проводятся лабораторные практикумы для студентов РТУ МИРЭА, НИУ МИЭТ.

Основные публикации по теме диссертации

1. Savin G.I., Vdovikin O.I., Shabanov B.M., Chetverushkin B.N., Gorobets A.V., Kozubskaya T.K., Sukov S.A. Gasdynamic and aeroacoustic simulations on the MBC-100M supercomputer // Doklady Mathematics. 2008, т. 78, № 3, pp. 932-935. (Scopus, WOS:000261967100034).
2. Savin G. I., Korneyev V. V., Shabanov B.M., Telegin, P.N., Semonov, D. V., Kiselev, A.V., Baranov, A.V., Vdovikin, O. I., Aladyshev, O.S., Ovsyanikov A.P. Grid-infrastructure JSCC RAS for supercomputing applications // Distributed Computing and Grid-Technologies in Science and Education: Proceedings of the 4th Intern. Conf. (Dubna, June 28–July 3, 2010) JINR Dubna, 2010, с. 406-410 (WOS:000393794200066)
3. Reznikov G.V., Smirnov G.F., Shabanov B.M. Thermophisical Problems with Freon Evaporative Systems and Heat Pipes Employed in Cooling Electronic Equipment. Heat Transfer in Electronic and Microelectronic Equipment // Proceedings of the International Centre for Heat and Mass Transfer. V 29, 1990, pp.841-859 (WOS:A1990BQ69J00053)
4. Boyarinov I.M., Davydov A.A., Shabanov B.M. Error correction in main memory of high-capacity computer // Automation and Remote Control. т. 48, выпуск 7, pp. 956-965, часть 2, Опубликовано JUL 1987 (WOS:A1987L826700007)
5. Benderskii L. A., Lyubimov D. A., Chestnykh A. O., Shabanov B. M., Rybakov A. A. The use of the RANS/ILES method to study the influence of coflow wind on the flow in a hot, nonisobaric, supersonic airdrome jet during its interaction with the jet blast deflector // Hight temperature, 2018, Vol. 56, No. 2, pp. 247-254. (ВАК, Scopus)
6. Victor Korneev, Dmitry Semenov, Andrey Kiselev, Boris Shabanov, Pavel Telegin Multiagent distributed grid scheduler // Proceedings of the Federated Conference on Computer Science and Information Systems. Szczecin, Poland, 18-21 September, 2011, pp. 577-580 (Scopus)
7. Igor V. Polyakov, Maria G. Khrenova, Alexander A. Moskovsky, Boris M. Shabanov, Alexander V. Nemukhin. Towards first-principles calculation of electronic excitations in the ring of the protein-bound bacteriochlorophylls // Chemical Physics, Vol. 505, 13 April 2018, pp 34-39. (WOS:000429471000006)
8. Boris M. Shabanov, Pavel Telegin, Oleg S. Aladyshev, Anton V. Baranov, Artem Tikhomirov. Comparison of priority-based and first price sealed-bid auction algorithms of job scheduling in a geographically-distributed computing system // Proceedings of the 2018 IEEE Conference ElConRus. 2018, pp. 1557-1562. (WOS: 000450337100367, Scopus)
9. Калинов А.Я., Климов С.А., Посыпкин М.А., Савин Г.И., Устюгов С.Д., Чечеткин В.М., Шабанов Б.М. Математическое моделирование задач о взрыве сверхновой на параллельном компьютере // Ж. вычисл. матем. и матем. физ., 44:5 (2004), 953–960; Comput. Math. Math. Phys., 44:5 (2004), с. 903-–910 (Scopus, ВАК)

10. Aladyshev O.S., Baranov A.V., Ionin R.P., Kiselev E.A., Shabanov B.M. Variants of deployment the high performance computing in clouds // Proceedings of the 2018 IEEE Conference ElConRus. 2018, pp. 1453-1457. (WOS :000450337100342, Scopus)

11. Мельников В.А., Шабанов Б.М. Применение высокопроизводительных вычислений в машиностроении и электронике // Проблемы машиностроения и надежности машин. 1993 г., №5, с. 3-9. (ВАК).

12. Борисов Ю.И., Шабанов Б.М. Одно из направлений развития САПР для создания сложных технических систем // Информационные технологии. 2003 г., №10, с. 2-9. (ВАК)

13. Клинов М.С., Лапшина С.Ю., Телегин П.Н., Шабанов Б.М. Особенности использования многоядерных процессоров в научных вычислениях // Вестник УГАТУ. – 2012. – т.16. – № 6 (51). – с. 25-31. (ВАК)

14. Аладышев О. С., Киселев Е. А., Савин Г. И., Телегин П. Н., Шабанов Б. М. Влияние характеристик внешней памяти суперкомпьютерных комплексов на выполнение параллельных программ // Системы и средства информатики, 2014, том 24, выпуск 4, с. 111-123. (ВАК)

15. Дикарев Н.И., Шабанов Б.М. Архитектура процессора и ее влияние на производительность суперЭВМ // Программные продукты и системы. 2007, №2, с.2-5. (ВАК)

16. Дикарев Н.И., Шабанов Б.М., Шмелев А.С. Использование «сдвоенного» умножителя и сумматора в векторном процессоре с архитектурой управления потоком данных // Программные системы: теория и приложения, 2015, 6:4(27), с. 227-241. (ВАК)

17. Дикарев Н.И., Шабанов Б.М., Шмелев А.С. Выполнение задач сортировки на векторном процессоре с архитектурой управления потоком данных // Программные системы: теория и приложения, 2017, 8:4(35), с. 305–317. (ВАК)

18. Дикарев Н.И., Шабанов Б.М., Шмелев А.С. Моделирование параллельной работы ядер векторного потокового процессора с общей памятью // Программные системы: теория и приложения, 2018, 9:1(36), с. 37-52. (ВАК)

19. Корнеев В.В., Семенов Д.В., Телегин П.Н., Шабанов Б.М. Отказоустойчивое децентрализованное управление ресурсами грид // Известия высших учебных заведений. Электроника. 2015, №1(111), с. 83-90. (ВАК)

20. Савин Г.И., Шабанов Б.М., Телегин П.Н., Вдовикин О.И., Козырев И.А., Корнеев В.В., Семенов Д.В., Киселев А.В., Кузнецов А.В., Овсянников А.П. Инфраструктура ГРИД для суперкомпьютерных приложений // Известия вузов. Электроника, 2011, №1(87), с. 51-55. (ВАК)

21. Савин Г.И., Шабанов Б.М., Корнеев В.В., Телегин П.Н., Семенов Д.В., Киселев А.В., Кузнецов А.В., Вдовикин О.И. Аладышев О.С., Овсянников А.П. Создание распределенной инфраструктуры для суперкомпьютерных приложений. // Программные продукты и системы. 2008, №2, с.2-7. (ВАК)

22. Овсянников А.П., Шабанов Б.М., Аладышев О.С., Опалев В.М., Вдовикин О.И., Захарченко А.В. Вычислительная сеть Межведомственного суперкомпьютерного центра. // Известия ВУЗов. «Электроника». 2004, №1, с.18-21. (ВАК)

23. Фортов В. Е., Савин Г.И., Левин В.К., Забродин А.В., Шабанов Б.М. Создание и применение системы высокопроизводительных вычислений на базе высокопроизводительных сетевых технологий // Информационные технологии и вычислительные системы. 2002, № 1, с. 3-10. (ВАК)

24. Аладышев О.С., Шабанов Б.М. Параллельная кластерная файловая система для высокопроизводительных вычислительных систем // Программные продукты и системы. – 2007. №2. (ВАК)

25. Шабанов Б.М. Выбор вычислительной системы для решения научных задач // Программные продукты и системы. 2012, № 4 (100), с.7-10 (ВАК)

26. Аладышев О.С., Биктимиров М.Р., Жижченко М.А., Овсянников А.П., Опалев В.М., Шабанов Б.М., Шульга Н.Ю. Особенности построения объединенной среды суперкомпьютерного центра // Программные продукты и системы, 2008, №2, с. 9-11. (ВАК)

27. Шабанов Б.М., Телегин П.Н., Аладышев О.С. Особенности использования многоядерных процессоров // Программные продукты и системы. 2008. №2. с. 7-9. (ВАК)

28. Телегин П.Н., Телегина Е.В., Шабанов Б.М. Влияние архитектуры на модели программирования параллельных вычислительных систем // Известия высших учебных заведений. Электроника. 2011. №2 (88), с. 60-65. (ВАК)

29. Аладышев О. С., Дикарев Н. И., Овсянников А. П., Телегин П. Н., Шабанов Б. М. СуперЭВМ: области применения и требования к производительности // Известия высших учебных заведений. Электроника. 2004. № 1. с. 13-17. (ВАК)

30. Савин Г.И., Телегин П.Н., Шабанов Б.М. Кластеры Беовульф // Известия высших учебных заведений. Электроника. 2004, №1, с. 7-12. (ВАК)

31. Шабанов Б. М., Овсянников А. П., Баранов А. В., Лещев С. А., Долгов Б. В., Дербышев Д. Ю. // Проект распределенной сети суперкомпьютерных центров коллективного пользования. // Программные системы: теория и приложения, 2017, том 8, выпуск 4, страницы 245-262. (ВАК)

32. Аладышев О.С., Телегин П.Н., Шабанов Б.М., Овсянников А.П., Опалев В.М., Вдовикин О.И. Аспекты разработки и создания кластерных вычислительных систем. // Известия высших учебных заведений. Электроника. 2004, №1, с.36-42. (ВАК)

33. Шабанов Б.М., Самоваров О.И. Принципы построения межведомственного центра коллективного пользования общего назначения в модели программно-определяемого ЦОД. Труды ИСП РАН. 2018, том 30, выпуск 6, с. 7-24. (ВАК)

34. Дикарев Н.И., Шабанов Б.М., Шмелев А.С. Векторный потоковый процессор: оценка производительности // Известия ЮФУ. Технические науки. Тематический выпуск: Суперкомпьютерные технологии. – Таганрог: Изд-во ТРТУ, 2014. №12 (161), с. 36-46. (ВАК)

35. Дикарев Н.И., Шабанов Б.М. Архитектура высокопроизводительных вычислительных систем. М.: ФАЗИС, 2015. 108 с.