

Вячеслав Виноградов

# Модульные компактные НРС-системы и серверы АТСА для телекоммуникаций и промышленности

## Часть 1

В работе представлен анализ модульных систем высокой производительности и серверов стандарта АТСА для телекоммуникаций и компьютерных систем промышленной автоматизации. Особое внимание уделено новым параллельно-конвейерным системам с конвергентной коммутируемой средой связи, обеспечивающей одновременные передачи больших потоков данных с гигабитными скоростями.

Основой дальнейшего развития современных высокопроизводительных вычислительных (НРС) систем и сетей служат модульная структура программно-аппаратных средств, масштабируемая архитектура систем, переход на гигабитные коммутируемые сети связи и новые многоядерные процессоры в виде однокристалльных нанoeлектронных СБИС [1-3]. Традиционные магистральные системы сопряжения и связи процессоров достигли пределов скорости передачи данных и требуют революционных подходов к созданию новых методов последовательной передачи сигналов для быстрой связи и взаимодействия процессорных узлов в вычислительных комплексах. Для сокращения потребляемой энергии и повышения быстродействия микропроцессоров и средств связи их переводят на более низкие уровни напряжения.

Основными требованиями, предъявляемыми к компактным встраиваемым компьютерным системам промышленного назначения и особенно к миниатюрным терминалам новых систем телекоммуникаций, являются габариты и малое потребление электроэнергии в

кристаллах при отсутствии охлаждающих вентиляторов, что приводит к необходимости снижения рабочей частоты процессоров. Между тем, в новых масштабируемых серверах с более высокой концентрацией вычислительных мощностей в единице объема основной задачей является максимальная производительность систем, при этом концентрация вычислительных ресурсов на единицу объема в кристалле процессора и в модулях приводит к

росту потребляемой энергии и огромному выделению тепла. Проблемы роста потребляемой энергии, а также охлаждения серверных кластеров и НРС-систем в центрах данных (ЦД) резко обостряются и становятся приоритетными для будущих систем, требуя новых архитектурных и конструктивных решений для новых модульных серверов и НРС-систем.

Развитие модульных систем первого поколения с программно-управляемой

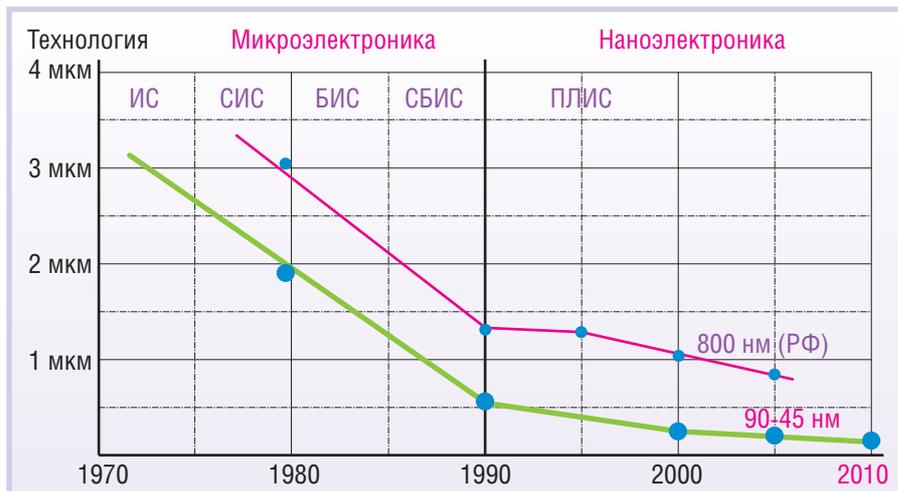


Рис. 1. Развитие электронных технологий

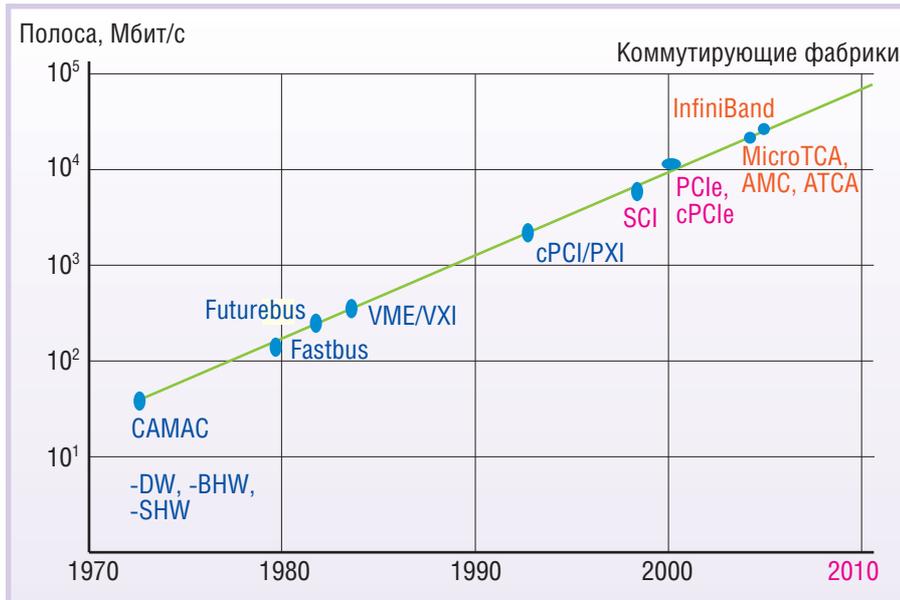


Рис. 2. Расширение полосы пропускания магистральных и коммутируемых компьютерных систем

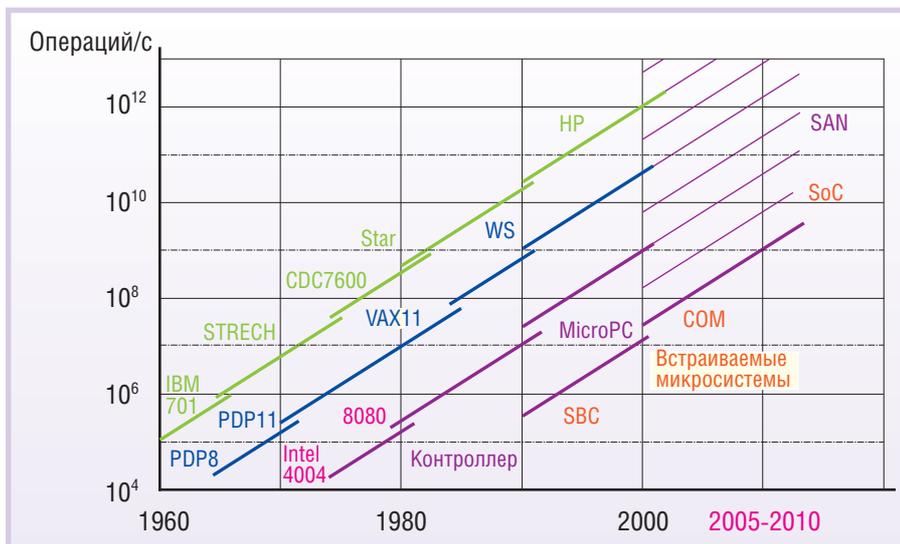


Рис. 3. Рост производительности компьютерных систем

модульной структурой для автоматизации измерений и управления шло на базе технологии ТТЛ силами научных центров разных стран. Второе поколение модульных процессорных систем высокой производительности создавали на базе технологий ТТЛШ (Futurebus) и ЭСЛ (Fastbus) в научных центрах и корпорациях. Большие научно-производственные и финансовые затраты затрудняли научным центрам дальнейшую конкуренцию с промышленными корпорациями, поэтому разработки модульных НРС-систем третьего поколения на основе технологии КМОП (PCI, VME) уже были выполнены компаниями Intel и Motorola. Четвёртое поколение модульных систем с коммутируемой средой (VPX, ATCA) формируется на основе новых нанoeлектронных СБИС международными объеди-

нениями (альянсами) крупных корпораций.

Развитие поколений модульных систем иллюстрируют рис. 1, 2 и 3.

На рис. 4 представлены функции производства и разработки модульных систем, учитывающие их развитие в четырёх описанных поколениях. Идеальная функция производства модулей как основы модульных систем предполагает периодический переход к следующему поколению модульных систем (рис. 4 а). Процесс новых разработок модулей каждого поколения характеризуется производной этой функции (рис. 4 б). В реальной ситуации модули предыдущих поколений часто применяются в новых системах на протяжении 1...2 поколений, а иногда и более. Причинами этого чаще всего выступают их достаточные надёжность и функциональность, а также экономический фактор более высокой стоимости нового оборудования. В силу таких реалий происходит процесс накопления модулей разных поколений (рис. 4 в), а на уровне модульных систем приходится решать дополнительные проблемы по их совместимости.

Первое поколение модульных систем обеспечивало полосу частот в несколько Мбайт/с (десятки Мбит/с). Полоса частот систем второго поколения была расширена на порядок. Это было достигнуто ценой более высокого энергопотребления и тепловыделения, и именно в системах второго поколения впервые пришли к необходимости применения водяного охлаждения электронного оборудования [1]. Более эффективная технология КМОП поз-

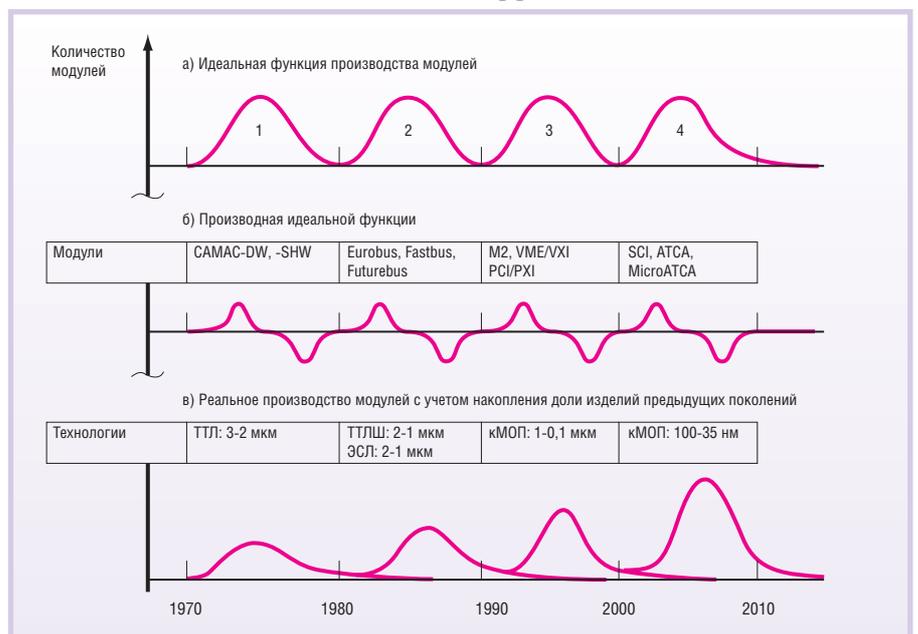


Рис. 4. Функции производства и разработки модульных систем

волила перейти к третьему поколению модульных систем, отличающихся меньшим энергопотреблением и высокими частотными параметрами. Среди представителей этого поколения наибольшее распространение получили модульные системы VME и CompactPCI (сPCI), которые в разных модификациях успешно применяются до настоящего времени. Однако в силу признанной невозможности эффективного роста частоты однокристальных процессоров, сложности создания компактных модулей и кристаллов с быстрыми параллельными связями, а также физических ограничений роста полосы частот взаимодействия процессорных узлов в традиционных магистральных архитектурах возникла задача создания и развития четвертого поколения модульных систем высокой производительности. Решение данной задачи было найдено в сетевых архитектурах на основе коммутируемых вычислительных сред, построенных на базе СБИС (система/сеть на кристалле). Четвертое поколение модульных систем определяет перспективное направление создания и развития будущих компактных модульных микросистем и масштабируемых серверных кластеров высокой производительности, которые и рассматриваются в данной статье.

Проблема эффективности новых серверов и НРС-систем по критериям производительности и энергопотребления решается использованием многоядерных модулей, а также созданием широкополосной коммутируемой среды связи узлов. При этом следует различать производительность подсистемы связи процессоров обработки, ввода-вывода и памяти.

Одним из первых прогрессивных решений для НРС-систем явилась мультипроцессорная архитектура SCI (Scalable Coherent Interconnect) на основе масштабируемого когерентного интерфейса, разработанная в Стэнфордском университете (США) при участии ряда фирм. Первая мощная НРС-система с архитектурой SCI была создана фирмой Sequent, отделившейся от Intel, но вскоре поглощённой IBM [4-8]. Особенностью этой системы стало и то, что здесь впервые применили низковольтные сигналы [9].

Решения комплексных проблем развития технологий компьютерной автоматизации в промышленности и телекоммуникациях реализуются на основе применения компактных встраиваемых

систем контроля и управления, масштабируемых серверов и систем высокой производительности, которые благодаря развитию новых нанoeлектронных технологий СБИС начинают создаваться также в виде модульных blade-серверов («лезвий»). Переход на новые конвергентные структуры модульных систем и сетей позволяет создавать широкополосные коммутируемые вычислительные среды высокой производительности. Новые НРС-системы требуют высокопроизводительных многоядерных процессоров и быстрых компактных средств сетевой связи, выполняемых в виде отдельных кристаллов СБИС.

### **ПРОБЛЕМЫ И МЕТОДЫ ВЗАИМОДЕЙСТВИЯ ПРОЦЕССОРНЫХ УЗЛОВ В МОДУЛЬНЫХ СЕРВЕРАХ**

Новые методы построения НРС-систем и модульных серверов, как и новых встраиваемых модульных систем, ориентированы на гигабитные коммутируемые среды. Параллельные и конвейерные методы обработки потоков данных предусматриваются на всех уровнях НРС-систем – от одновременного исполнения команд в процессорах до распараллеливания процессов в многопроцессорных системах и многоаппаратных комплексах (кластерах).

### **Одновременное исполнение команд в микропроцессорах**

Микропроцессоры развивались как параллельные устройства для выполнения нескольких команд одновременно в каждом такте, при этом важную роль играет выбранный метод реализации функций – аппаратный или программный. Преодоление ограничений традиционной суперскалярной архитектуры компьютерных систем произошло в IA-64 (Intel) и Crusoe (Transmeta). В первом случае процессор выполняет команды по очереди, используя логику группирования для одновременного выполнения команд из одной группы, а от суперскаляра унаследовано предсказание переходов. Микропроцессор Crusoe, отличающийся относительно малой рассеиваемой мощностью, имеет 4 слога в команде и использует динамическую компиляцию. Аппаратный параллелизм для повышения производительности систем использовали в архитектуре процессора «Эльбрус» [10], при этом задача оптимизации и планирования передаётся с ап-

паратного на программный уровень. В процессоре «Эльбрус» использовали статическую компиляцию, а 16 слогов позволили выполнять одновременно до 23 операций (проект E2K).

В проекте «Эльбрус» реализовали сокращение потерь времени за счёт улучшения времени считывания данных из иерархии памяти. Если в процессоре «Эльбрус» берут двоичный код и выполняют программно планирование ресурсов, то в традиционных суперскалярных компьютерах ПО не выполняет этого, так как ресурсы не могут контролироваться двоичным кодом, поэтому каждая команда должна быть широкой для включения спецификации параллельных ресурсов. Аппаратная защита основана на 2 контрольных разрядах в каждом 32 разрядах ЦП и кэш-памяти в кристалле.

Переупорядочиванием последовательности команд сокращают число обменов с внешними устройствами ввода-вывода. Введение конвейерной обработки, сокращение времени ввода-вывода и числа обращений к памяти становятся основными методами дальнейшего увеличения быстродействия систем. В некоторых вычислительных системах (Cray Res) с масштабируемой векторной архитектурой на основе однокристальных процессоров достигли пиковой производительности 1 TFLOPS, обеспечив малую задержку и большую полосу частот для подсистем памяти.

### **Многопроцессорные структуры и коммутируемые сети**

Серверы и НРС-системы создают в виде многопроцессорных и массивно-параллельных систем. Размещение схем (процессора, коммутатора и памяти) в одном кристалле позволит значительно улучшить эти параметры и создать модульные серверы и НРС.

Различают слабые и сильные взаимодействия процессорных узлов в системах. Слабые взаимодействия ядер процессоров основаны на относительно медленных каналах с общей средой связи, допускающей передачу данных только одним узлом в каждый момент времени, что ограничивает скорость взаимодействия всех узлов в комплексе. Сильные взаимодействия ядер процессоров основаны на общей памяти или новых методах коммутируемой связи, допускающей одновременно параллельно-конвейерные потоки данных между несколькими узлами систе-

мы. Под сильными взаимодействиями будем понимать одновременные транзакции процессорных узлов (ядер) между собой, а также с подсистемами ввода-вывода и памяти через широкополосную среду коммутируемой связи. Одновременное взаимодействие многих узлов может быть достигнуто только специальными методами конвейерной связи и коммутируемой средой нового поколения.

Проблемы параллельной обработки в модульных серверах и НРС-системах решают на основе векторной или скалярной формы обработки данных в процессорах, а также распараллеливанием алгоритмов в многоядерных и многопроцессорных структурах и конвейерных сетях. Существует хорошо структурированный параллелизм задач, однако в части программ векторизация затруднена или иногда даже невозможна (эффективность зависит от этой части программ). Требуется оптимизация для увеличения скорости скалярной обработки, что ещё заметнее для систем массового параллелизма. В таких случаях создают неоднородные вычислительные подсистемы, состоящие из сильно связанных модулей, содержащих масштабируемый вектор-

ный унипроцессор и масштабируемый скалярный мультипроцессор, связанные общей памятью. Мониторная подсистема выполняет контроль и анализ задач для их распараллеливания. Различают многопроцессорные системы с симметричным и несимметричным доступом к распределённой памяти и серверные кластеры (многомашинные комплексы) с фиксированной памятью, объединяемые гигабитной коммутируемой сетью.

**Многопроцессорные методы** применяются в системах с адресным пространством 16/32/64 разрядов, при создании многопроцессорных систем обработки сигналов в одном модуле. Многопроцессорные системы имеют симметричную память, несимметричную коммутируемую память, а также общедоступную сетевую память. Многопроцессорные магистральные системы с симметричным доступом к памяти имеют, как правило, физическое ограничение по числу процессоров (обычно до 8...16) и используются при обработке данных в модулях обработки сигналов на основе DSP. Системы с несимметричным доступом к памяти (NUMA) использовали для создания НРС-систем и серверов на основе 32-разрядных процессоров.

Большое адресное пространство современных 64-разрядных процессоров компьютерных систем с общей разделяемой памятью обеспечивает новые возможности прямого доступа к любой области памяти для обработки и быстрого обмена данными многими процессорами. Производительность компонентов серверов и НРС-систем не характеризует их полностью, так как при передаче больших массивов глобальных данных производительность комплекса будет падать из-за взаимодействия процессоров и обмена данными, то есть реальная производительность всего комплекса может быть ниже ожидаемой. Причина этого – в недогрузке процессоров из-за запаздывания данных при передаче или из-за задержки данных, ожидаемых от других процессоров. Реальная производительность комплекса из однородных процессоров пропорциональна числу и производительности базовых процессоров и пропускной способности коммутируемой среды и каналов связи в подсистемах памяти и обратно пропорциональна частоте связи и числу процессоров, участвующих во взаимодействиях.

Система с распределённой памятью имеет ряд особенностей. Влияние за-

держек в коммуникационных средствах проявляется здесь сильнее. При числе узлов более 16 многопроцессорный комплекс становится малоэффективным и для его работы требуется сверхоперативная статистическая кэш-память у каждого процессора. Однако работа с глобальными данными требует когерентности кэш-памяти (например, SCI), а это приводит к дополнительным потерям времени, что критично для обеспечения режима реального времени.

**Вычислительные системы массового параллелизма** представляют другое направление систем высокой производительности. Реализация методов массового параллелизма вычислительных процессов связана с большим увеличением производительности систем. Такие системы стремились создавать и ранее (например, транспьютеры), однако они были ограничены возможностями тогдашних процессорных БИС и скоростью связей узлов (Links — линков). Потом создавали специально сетевые процессоры, которые обеспечивали развитые функции управления связью, но были ограничены по функциям обработки. В настоящее время есть опыт создания мощных многоядерных процессоров и гигабитных коммутирующих сетей в одном кристалле, поэтому дальнейшее развитие систем и сетей идёт по пути создания новых конвергентных коммутируемых вычислительных сред.

Другой проблемой является синхронизация данных параллельных процессов. Естественным способом синхронизации процессов по данным служит использование самих процессоров для слежения за появлением таких процессов в определённых вычислениях, что требует дополнительного времени и приводит к снижению эффективности. Между тем, при синхронизации с использованием ячеек памяти велика вероятность использования устаревших данных.

Эти проблемы были решены в таких суперЭВМ, как ASCI BLUE Mountain и BLUE Pacific System, на основе коммутатора и симметричного доступа к памяти многих процессоров. На втором уровне комплекса используют локальную связь с поддержкой когерентности кэш, а на третьем уровне — разделяемую память и пакетный режим обмена между двухступенчатыми комплексами. Повышение производительности достигается быстрыми коммутаторами

и сокращением длины связей между узлами за счёт новых СБИС, в кристаллах которых более плотно размещено большее количество процессоров.

**Конвергентная коммутируемая среда вычислительных систем** может быть построена по методу распределённых модульных узлов связи или коммутируемых структур и маршрутизаторов в кристалле. В первом случае модули связи устанавливаются в компьютерные модульные системы или в ПК для объединения в кластеры. Во втором случае они встраиваются в отдельные модули и коммутационные задние панели. Более перспективные методы предусматривают создание коммутируемой среды связи модулей в виде многопортовых коммутаторов или встраиваемых коммутаторов в кристалле СБИС. При этом чем больше портов будет включено в коммутатор, тем выше будет концентрация мощностей и компакнее организация больших центров обработки данных.

**Концепция** нового поколения эффективных модульных микросистем, серверов и НРС-систем предлагает их создание на основе единого нового семейства компактных модулей для микросистем и масштабируемых НРС-систем. Масштабируемые серверные системы создают из набора укрупнённых несущих («крейсерских») модулей, в каждом из которых устанавливается до 4...8 компактных модулей (многоядерные процессоры, устройства ввода-вывода и памяти). Такие укрупнённые несущие модули являются модульными серверами (blade-серверами) нового поколения. Они содержат требуемый набор компактных модулей процессоров и выполняются в виде макромодулей с коммутируемой сетевой средой связи для организации подсистем обработки потоков данных, ввода-вывода и хранения больших массивов данных.

Макромодули могут служить базовой платформой для создания масштабируемых модульных серверов и НРС-систем. Набор несущих модулей blade-серверов может устанавливаться в секции или стойки для создания более мощных модульных НРС-систем нового поколения с очень высокой концентрацией вычислительных ресурсов для центров данных.

Другое направление — развитие микросистем из такого же набора компактных модулей. Их устанавливают непосредственно в специальные секции, не

используя макромодули. Это позволяет создавать новые эффективные компактные системы с малым потреблением энергии для промышленной автоматизации, биомедицины, мобильных приложений и других применений.

## Коммутируемые широкополосные среды

С развитием СБИС многоядерных процессоров высокой производительности и конвергентных гигабитных коммутируемых сред сетевая архитектура модульных систем стала основополагающей. Её основу составляют модульные серверные кластеры (наборы blade-серверов), охваченные широкополосными сетями, например, гигабитными Ethernet (GE, 10GE, 100GE) или InfiniBand. При этом следует различать коммутируемую среду кластеров обработки данных, организации подсистем памяти и ввода-вывода. Для организации подсистем ввода-вывода используют коммутируемые среды с последовательными каналами PCIe, Serial RapidIO (SRIO) или другие типы сопряжения, характерные для новых компактных модульных систем, а для подсистем памяти традиционно применяют волоконно-оптические каналы (Fibre Channel — FC).

Среди методов создания модульных структур для серверов и НРС-систем выделяют два основных подхода:

- многопроцессорные модульные системы с распределяемой памятью на основе конвейерно-параллельных каналов коммутируемой среды (типа SCI);
- гигабитные сети типа GE, 10GE, 100GE или InfiniBand.

Многопроцессорные структуры способны объединить в комплекс десятки тысяч процессоров, позволяют создавать кольцевые и многокольцевые топологии, а также распределённые кольцевые структуры для обработки двумерных и трёхмерных данных на основе распределённых коммутируемых конвейерных методов связи (типа SCI). Правда, с ростом числа процессоров возникают проблемы масштабирования и когерентности данных.

Использование для связи процессоров централизованных коммутирующих устройств с гигабитной скоростью улучшает масштабируемость систем. Коммутируемая гигабитная среда может быть создана разными способами, и в первую очередь, на основе хорошо отработанных средств связи и маршру-

тизации Ethernet. Среди других способов выделяется коммутируемая среда с неограниченно масштабируемой полосой частот InfiniBand, первые образцы реализации которой в серверах уже появились на рынке. Архитектура InfiniBand позволила объединить подсистемы ввода-вывода Future I/O и NGIO ([www.infinibandta.org](http://www.infinibandta.org)).

Среда связи InfiniBand была выделена благодаря ряду преимуществ.

1. Высокая производительность характеризуется скоростями связи между узлами — до 20 Гбит/с, связи между коммутаторами — до 60 Гбит/с, общей маршрутизации — до 1220 Гбит/с.
2. Обеспечиваются консолидация многих устройств ввода-вывода с помощью среды связи на отдельной модульной панели или кабеле, а также консолидация кластеров серверов, памяти и управления разными типами данных через одно соединение, в отличие от других типов соединений, не обеспечивающих разные типы трафиков.
3. Эффективная масштабируемость разных типов систем приводит к аппаратной реализации многих функций, что освобождает процессор от целого ряда задач; эффективное ис-

пользование ресурсов каждого узла, добавляемого к кластерам, а также технологии прямого удалённого доступа к памяти (RDMA) с оптимизированным протоколом передачи данных освобождают серверы для работы с приложениями.

4. Обеспечиваются надёжное взаимодействие конечных узлов, виртуализация соединений одновременно для многих приложений, работающих с одними и теми же частями отдельного специализированного приложения.
5. Избыточность в коммутируемых соединениях обеспечивает надёжность связи подсистем.

Сигнальная скорость InfiniBand достигает 2,5 Гбит/с, а при дублировании или квадратировании — 5 или 10 Гбит/с соответственно.

### **Пример массивно-параллельной обработки изображений на новых СБИС**

Систему массивно-параллельной обработки данных на основе ячеечных (клеточных) процессоров рассмотрим на примере комплекса обработки медицинских изображений, который получает данные от таких устройств, как компьютерный томограф (КТ) и маг-

нитно-резонансный томограф. КТ применяют в большом числе клиник и постоянно создают новые алгоритмы обработки медицинских изображений для улучшения их качества и снижения дозы облучения пациентов. Компьютерные системы позволяют реконструировать изображения за короткий интервал времени, используя алгоритмы аппроксимации. Для быстрой реконструкции образов применяют специальные решения.

Фирмы IBM, Sony и Toshiba решили применить технологию создания ячеечной структуры процессора на наборе процессорных элементов (ПЭ), работающих с частотой 3 ГГц, чтобы достичь высокой производительности (200 GFLOPS). Задачи подразделили на конечное число подзадач, выполняемых на ПЭ. Система построена как симметричный мультипроцессор, что позволяет распределённому приложению работать на любом конечном числе ПЭ с параллельно-конвейерной передачей данных.

Реконструкция изображений включает в себя взвешивание данных, конволюцию и реконструкцию обратных проекций. Процессы обработки изображений выполняются параллельно.

Программы обработки проекций (2D, 3D) предварительно загружают во все ПЭ. Ограничением является размер локальной памяти 256 кбайт. По каналу прямого доступа передаются только необходимые части данных. Векторизация достигается «тренировкой» процессора с учётом пикселей соседних узлов изображения как 4 векторов чисел с плавающей запятой. Процессор может обрабатывать по 165 фрагментов из 512 проекций с числом пикселей 512 на каждый фрагмент.

### **Пример отечественной универсальной мультипроцессорной системы**

Компания ПРОСОФТ совместно с ОАО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» (НИИВК) и компанией FASTWEL создала высокопроизводительную универсальную мультипроцессорную систему (УМС), предназначенную для решения прикладных задач, требующих параллельных вычислений. Система построена на базе 24 новейших двухъядерных процессоров Intel Core, объединённых высокоскоростной коммутируемой сетевой средой связи InfiniBand.

В качестве узлов системы использованы серверы Intel с двумя двухъядерными процессорами Intel Xeon серии 5100, работающими при пониженном напряжении питания и, соответственно, с пониженным тепловыделением. Всё оборудование УМС (серверы, коммутаторы Gigabit Ethernet и InfiniBand, система бесперебойного питания) размещено в стандартном 19" шкафу VARISTAR (производитель — Schroff) со степенью защиты IP54.

Использование сети InfiniBand позволило в 10 раз увеличить скорость обмена данными между узлами системы и в 20–30 раз уменьшить задержки.

Вычислительная мощность текущей конфигурации УМС составляет 447 GFLOPS при энергопотреблении 4,5 кВт. Оптимизация архитектуры кластера и применение процессоров с низковольтным питанием обеспечили высокую вычислительную плотность 75 GFLOPS/кВт.

Более подробно УМС описана в «СТА» № 3 за 2008 год (стр. 98).

### **АРХИТЕКТУРА МОДУЛЬНЫХ СИСТЕМ НА ОСНОВЕ SCI**

Структура SCI — одна из первых реализаций модульных систем с быстрой

связью в виде многопроцессорных комплексов (мультипроцессоров). Впервые такое соединение было реализовано фирмой Sequent [6] в НПС-системах на основе низковольтных каналов конвейерной связи с большой полосой частот [9]. Спецификация SCI как международный стандарт для систем с архитектурой неунифицированного доступа к распределённой памяти (NUMA) включает в себя документы, отражающие инновации в её развитии, например:

- IEEE 1596.3 — дифференциальные сигналы низкого уровня;
- IEEE 1596.4 — организация связи быстрой памяти по кольцевой схеме типа SCI;
- IEEE 1596.5 — форматы обмена данными.

Различают совместимые распределённые кольцевые модульные средства связи и коммутируемые среды. Методика кольцевой связи обычно используется для соединения готовых ПК в кластеры или для создания новых модульных систем. Особенностью SCI является возможность создания коммутируемой среды с помощью специального коммутатора, а также объединения обычных ПК в кластеры с по-

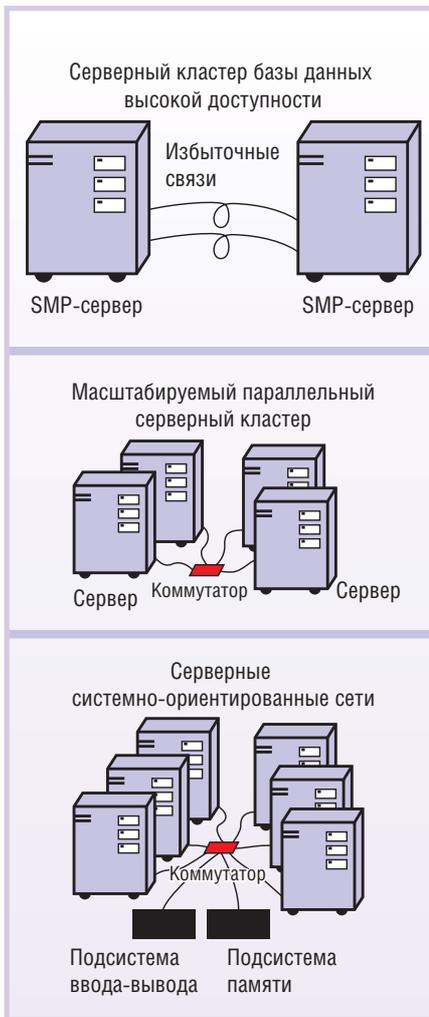


Рис. 5. Связь серверов, подсистем ввода-вывода и памяти с помощью модульных коммутаторов SCI

мощью специальных линков, отличающихся высокой частотой и сверхмалыми задержками на прикладном уровне (рис. 5). Например, для создания надёжной системы с резервированием либо сервера данных достаточно нескольких ПК, объединяемых в сеть, имеющую любую топологию («кольцо», «многокольцевая матрица», «куб» и т.п.) в виде симметричной многопроцессорной структуры (SMP), или связи нескольких серверов с помощью коммутатора. Таким же методом соединяют подсистемы ввода-вывода и памяти в системы высокой производительности. Гибкость построения различных систем на обычных ПК привлекла внимание многих разработчиков, однако высокая стоимость оборудования SCI ограничивала их более широкое применение. Например, несмотря на рекордно малую задержку сообщений, в подсистемах памяти продолжали по экономическим соображениям использовать традиционные волоконно-оптические каналы.

Системно-ориентированные сетевые архитектуры могут быть созданы как на основе типовых персональных компьютеров, соединяемых в различные сетевые топологии с помощью быстродействующих адаптеров связи, так и в виде законченных модульных структур, включающих в себя модули процессоров, ввода-вывода и связи, устанавливаемых в специальные секции (крейты).

Модули SCI выпускаются в разных форматах, с разными интерфейсами (PCI, cPCI, cPCIe, PMC и др.), для работы в разных операционных системах (ОС). Адаптерные модули связи и коммутаторы обеспечивают возможность построения кольцевых структур, матричных систем, двумерных торсидальных структур, а также прямых коммутируемых системных связей для сетей разных топологий.

Рассмотрим разные по типу интерфейса модули SCI, используемые для построения мультипроцессорных и модульных серверов.

**Адаптерные модули связи SCI-PCI** обеспечивают конвейерную передачу сообщений, одновременную работу всех узлов в общей сети и прямой доступ процессоров к распределённой памяти.

**Адаптерные модули связи SCI-cPCI** обеспечивают возможность интеграции коммутируемых соединений для создания кластеров и HPC-систем, а также встраиваемых компьютерных систем реального времени. При этом достигаются предельно малые значения времени задержки передачи сообщений. Фирма Dolphin сообщила о разработке систем высокой производительности на основе 64-разрядных процессоров AMD Opteron и SCI. Были созданы специальная программная инфраструктура SISC (Software Infrastructure SCI) и библиотека API для применения кластеров высокой производительности, рефлексивной памяти и приложений. Применение SISC сделало возможным дистанционный доступ к разделяемой памяти с очень малой задержкой, а также передачи по каналам прямого доступа типа память-память.

**Адаптерные модули связи SCI-cPCIe** применяются для создания мощных рабочих станций, серверных систем, HPC-кластеров, а также используются в различных встраиваемых приложениях. Эти модули обеспечивают большую полосу частот при малой задерж-

ке. Например, скорость двунаправленной передачи при межсистемном обмене большими массивами данных может достигать 10 Гбит/с, а при двух каналах — 20 Гбит/с при задержке всего 1,4 мкс. Подобные решения эффективны для систем реального времени и управляющих комплексов с большим количеством узловых соединений. Технологии конфигурирования и применение коммутаторов позволяют получать избыточные системы высокой надёжности. Адаптерные модули связи SCI-cPCIe предназначены для создания «мостов», кластерных конфигураций, компактных модульных систем и системно-ориентированных сетей (SAN) высокой производительности. ●

*Окончание следует*

## ЛИТЕРАТУРА

1. Виноградов В.И. Информационно-вычислительные системы. — М.: Энергоатомиздат, 1986. — 336 с.
2. Виноградов В.И. Системное развитие быстродействующих коммутируемых сетей Ethernet // Современные технологии автоматизации. 2008. № 1. С. 6-12.
3. Виноградов В.И. Структура современных встраиваемых модульных систем с сетевой архитектурой // Современные технологии автоматизации. 2008. № 2. С. 6-18.
4. Виноградов В.И. Распределённые модульные системы реального времени // Международный симпозиум по проблемам модульных систем и сетей ICSNET'2001, Москва: [сб. трудов].
5. Митрофанов В.В., Слущкий А.И., Ларионов К.А., Эйсымонт Л.К. Результаты испытаний новых образцов SCI-адаптеров в вычислительных кластерах серии ТКС // Международный симпозиум по проблемам модульных систем и сетей ICSNET'2001, Москва: [сб. трудов].
6. Gustavson D.B., Vinogradov V.I. SCI-based Modular Microprocessor Systems // Proceedings of ICSNET'97.
7. Gustavson D.B., Vinogradov V.I. New Technology Based Scalable Modular System Architecture // Proceedings of ICSNET'99.
8. Vinogradov V.I., Lutchev A.N. Scalable SCI-based SMP System Applications // Proceedings of the International School-seminar ACS'98.
9. Kempainen S. Low Voltage Differential signals. IEEE P1596.3 Working Group // Proceedings of ICSNET'93.
10. Бабаян Б.А. Микропроцессор нового поколения для суперкомпьютеров // Международный симпозиум по проблемам модульных систем и сетей ICSNET'2001, Москва: [сб. трудов].