УДК 004.725.5

ПЕРВОЕ ПОКОЛЕНИЕ ВЫСОКОСКОРОСТНОЙ КОММУНИКАЦИОННОЙ СЕТИ «АНГАРА»

И.А. Жабин – начальник сектора, ОАО «НИЦЭВТ»
E-mail: zhabin@nicevt.ru
Д.В. Макагон – начальник отдела, ОАО «НИЦЭВТ»
E-mail: makagond@nicevt.ru
Д.А. Поляков – инженер-программист, ОАО «НИЦЭВТ»
E-mail: polyakov@nicevt.ru
А.С. Симонов – к. т. н., начальник отделения, ОАО «НИЦЭВТ»
E-mail: simonov@nicevt.ru
Е.Л. Сыромятников – ст. науч. сотрудник, ОАО «НИЦЭВТ»
E-mail: syromiatnikov@nicevt.ru
А.Н. Щербак – вед. инженер-электроник, ОАО «НИЦЭВТ»
E-mail: andrey.shcherbak@nicevt.ru

Рассмотрены существующие коммуникационные сети для суперкомпьютеров, основные архитектурные особенности, функциональные возможности и технические характеристики первого поколения отечественной высокоскоростной коммуникационной сети «Ангара» на базе СБИС ЕС8430, разработанной в ОАО «НИЦЭВТ».

Ключевые слова: *суперкомпьютеры, многопроцессорные вычислительные системы, коммуникационная сеть, СБИС, ЕС8430, RDMA, топология «многомерный тор».*

This paper discusses the contemporary interconnection networks for supercomputers. We focus on the architecture, functionality and main features of the first generation of Angara Interconnect – a Russian high-speed interconnection network based on EC8430 ASIC developed by JSC «NICEVT».

Keywords: *supercomputers, HPC, multiprocessor computer systems, interconnection networks, ASIC, EC8430, RDMA, multidimensional torus topology.*

Современные суперкомпьютеры состоят из десятков и даже сотен тысяч узлов, объединяемых высокоскоростной коммуникационной сетью. Поэтому эффективность работы всей системы в целом во многом определяется производительностью не столько отдельных узлов, сколько коммуникационной сетью. Ведь именно от того, как быстро и эффективно сможет взаимодействовать столь огромное число узлов при решении задачи, и зависит, насколько большой выигрыш во времени можно получить по сравнению с последовательным решением этой задачи на одном узле.

В настоящее время существует множество разновидностей суперкомпьютеров различного уровня производительности, отличающихся по архитектуре, вычислительной мощности, энергопотреблению, надёжности, специализированности под определённые классы задач. Особый интерес среди них представляют наиболее производительные суперкомпьютеры, поскольку именно они позволяют получить уникальные результаты, недоступные иными способами. Разработка подобных систем – крайне сложная задача, требующая слаженного решения множества уникальных научных и инженерных задач.

Наиболее мощными суперкомпьютерами на текущий момент (по списку Top500, ноябрь 2013) являются китайские системы Tianhe-2 и Tianhe-1A, японский K Computer, американские Cray Titan, IBM Blue Gene/Q. Все эти суперкомпьютеры используют собственные уникальные («заказные») коммуникационные сети, разрабатываемые в составе этих вычислительных систем и доступные только совместно с ними. Приобретение подобных машин в России в ряде случаев затруднено, а часто является фактически невозможным. В то же время коммерчески доступные сети InfiniBand и Ethernet не подходят для эффективной реализации систем со столь высокими требованиями по масштабируемости, надёжности и производительности. В связи с этим крайне актуальным является вопрос разработки отечественной высокоскоростной сети, сравнимой с западными «заказными» аналогами. Разрабатываемая в ОАО «НИЦЭВТ» высокоскоростная коммуникационная сеть «Ангара» должна занять именно эту нишу.

Для того, чтобы лучше понять требования, предъявляемые к разрабатываемой сети «Ангара», и иметь возможность сравнения этой сети с существующими решениями, в первую очередь необходимо более детально рассмотреть существующие сети.

Сеть InfiniBand широко используется для построения кластерных систем и суперкомпьютеров. Основным производителем адаптеров и коммутаторов InfiniBand является компания Mellanox. Последнее на данный момент поколение сети InfiniBand – InfiniBand FDR – было представлено в июне 2011 г. Основным количественным улучшением нового поколения является увеличенная пропускная способность линков – до 14 Гбит/с.

Архитектура сети InifiniBand оптимизирована под топологию fat tree, однако новейшие коммутаторы и маршрутизаторы поддерживают топологию «многомерный тор», а также гибридную топологию из fat tree и трёхмерного тора.

Популярность сети InfiniBand определяется наличием развитой экосистемы программного обеспечения, относительно невысокой стоимостью и довольно хорошей поддержкой стандарта MPI. Данная сеть, однако, имеет и свои недостатки: относительно большую задержку (более 1,5 мкс на передачу узел-узел и дополнительно 0,1–0,5 мкс на каждый транзитный узел), ограничения по масштабируемости (порядка 48 тысяч узлов на L2-сегмент как ограничение протокола; для больших масштабов требуются маршрутизаторы, существенно ограничивающие пропускную способность между сегментами и усложняющие обслуживание сети). В целом можно сказать, что InfiniBand – это продукт для массового пользователя, при разработке которого была сделана ставка на универсальность (иногда – в ущерб эффективности).

Китайский суперкомпьютер Tianhe-1A, разработанный NUDT (National University of Defense Technology of China) в 2010 г., состоит из 7168 вычислительных узлов, объединенных сетью Arch собственной разработки с топологией fat tree. Сеть строится из 16-портовых маршрутизаторов, односторонняя пропускная способность линка – 8 Гбайт/с, задержка – 1,57 мкс. Поддержаны операции RDMA и оптимизированы коллективные операции. В суперкомпьютере Tianhe-2, появившемся в 2013 г., также используется коммуникационная сеть собственной разработки – TH Express-2 с топологией fat tree. На верхнем уровне сети используются 13 576-портовых коммутаторов на базе специально разработанного чипа NRC (технология 90 нм, 17,16x17,16 мм, 2577 пинов), агрегатная пропускная способность которого составляет 2,56 Тбит/с. Сетевой адаптер использует интерфейс PCI Express 2.0 x16. Коммуникационная задержка для этой сети, измеренная на сообщениях размером 1 КБ на 12000 узлах, равна 9 мкс [1].

Системы серии IBM Blue Gene являются классическими представителями суперкомпьютеров, использующих тороидальную топологию объединения вычислительных узлов. В первых двух поколениях этих систем – Blue Gene/L (2004) и Blue Gene/P (2007) – использовалась топология 3D-тор, дополненная рядом специализированных сетей для синхронизации и коллективных операций, в Blue Gene/Q реализована топология 5D-тор без дополнительных сетей. Пропускная способность линка в Blue Gene/Q составляет 2 ГБ/с, что, с одной стороны, существенно больше 0,425 ГБ/с, предоставляемых в предыдущем поколении, но с другой, – на порядок меньше пропускной способности, предоставляемой, например, InfiniBand или Cray Gemini.

В *суперкомпьютерах фирмы Cray* также использовалась топология «многомерный тор», но в отличие от подхода IBM, которая старалась сделать большую размерность тора, пусть даже за счет более слабых линков, Cray последовательно использовала топологию 3D-тор в сетях SeaStar, SeaStar2, SeaStar2+ и потом Gemini, наращивая пропускную способность линков вплоть до 9,375 ГБ/с. Важно отметить, что один маршрутизатор Gemini соответствует двум маршрутизаторам SeaStar2+, т.е. фактически двум узлам сети, поэтому в Gemini вместо 6 линков для соединения с соседними узлами используется 10 линков (4 линка для направлений X, Z и 2 линка для Y). Новейшая сеть от Cray – Cray Aries использует новую топологию dragonfly.

Сеть Tofu (от TOrus Fusion), которая используется в японском суперкомпьютере K Computer изначально была спроектирована так, чтобы позволить при выборе подмножества узлов сохранить топологию 3D-тор. Данная сеть имеет два уровня иерархии. Верхний уровнь представляет собой масштабируемый 3D-тор, в узлах которого находятся группы (Tofu units) по 12 узлов. Узлы каждой группы соединены между собой 3D-тором со сторонами 2×3×2 без дублирующих связей, что эквивалентно 2D-тору со сторонами 3×4. Таким образом, узел сети Tofu имеет 10 линков с пропускной способностью в 40 Гбит/с

каждый. Tofu имеет аппаратную поддержку синхронизации узлов и редукции (целочисленная и с плавающей запятой).

Ряд отечественных организаций также ведут разработку коммуникационных сетей для использования в суперкомпьютерах, в том числе РФЯЦ ВНИИЭФ (о данных разработках имеется мало информации в открытых источниках); Институт программных систем РАН и РСК «СКИФ»; ИПМ РАН и НИИ «Квант» (сеть «МВС-Экспресс»).

В 2010 г. в Южно-Уральском государственном университете (ЮУрГУ) был представлен российско-итальянский. суперкомпьютер «СКИФ-Аврора», помимо прочего использующий сеть отечественной разработки, полностью построенную на базе ПЛИС Altera Stratix IV. Пропускная способность одного линка в этой сети – 1,25 ГБ/с.

В сети «МВС-Экспресс» для объединения вычислительных узлов используется PCI Express 2.0, при этом узлы объединяются через 24-портовые коммутаторы. Сеть имеет топологию, близкую к fat tree. Сетевой адаптер в вычислительном узле имеет один порт шириной 4 лейна, односторонняя пиковая пропускная способность на линк составляет 20 Гбит/с без учёта накладных расходов на кодирование. Преимуществом применения PCI Express в «МВС-Экспресс» является эффективная поддержка общей памяти с возможностью односторонних коммуникаций. Как следствие, сеть удобна для реализации библиотеки Shmem и PGAS-языков (UPC, CAF).

В ОАО «НИЦЭВТ» с 2006 г. ведётся разработка коммуникационной сети «Ангара» – отечественной высокоскоростной коммуникационной сети с топологией 4D-тор, которая сможет стать основой для создания отечественных суперкомпьютеров.

В 2013 г. появилось на свет первое поколение маршрутизаторов сети «Ангара» на базе СБИС ЕС8430. Этому событию предшествовала длительная подготовительная работа, включавшая несколько отдельных этапов: сначала проводилось общее имитационное моделирование различных вариантов сети и принимались основные решения по топологии, архитектуре маршрутизатора, алгоритмам маршрутизации и арбитража. Затем была выбрана топология «многомерный тор» [2], подобраны различные количественные характеристики сети, включая оптимальные размеры буферов, число виртуальных каналов, проанализированы потенциальные узкие места. При разработке принципов работы сети в качестве руководства использовались [3] и [4], некоторые идеи были также в том или ином виде взяты из описаний архитектур IBM Blue Gene и Cray SeaStar.

В 2007 г. начались работы по макетированию сети с помощью маршрутизаторов на базе ПЛИС (FPGA). В 2008 г. появились первые полнофункциональные прототипы (М2) маршрутизатора на ПЛИС Xilinx Virtex4, с использованием которых был собран макет сети из шести узлов, соединенных в тор 3×2. Данный макет использовался для отладки базовой функциональности маршрутизатора, отработки отказоустойчивой передачи данных. Параллельно были написаны и отлажены начальные варианты драйвера и библиотеки нижнего уровня, портирована библиотека Cray Shmem и обеспечена поддержка MPI [2]. В сентябре 2010 г. был запущен макет с прототипами маршрутизатора третьего поколения (М3), состоящий из девяти узлов, соединенных в двухмерный тор 3×3. В 2012 г. был создан двухузловой макет для отладки высокоскоростных каналов передачи данных с пропускной способностью 12×6,25 Гбит/с. В 2013 г. появилось первое поколение маршрутизаторов сети «Ангара» (рис. 1) на базе СБИС (рис. 2). В настоящий момент продолжается наладка и тестирование этих маршрутизаторов.

Сравнение характеристик сети «Ангара» с зарубежными решениями приведено в таблице.

Сравнительные характеристики сети «Ангара» и зарубежных решений

Характеристика	Сети				
	Ангара М3 (ПЛИС)	Ангара (СБИС)	InfiniBand FDR 4x	IBM BG/Q	Cray XK7
Топология сети	2D-тор	4D-тор	fat tree	5D-тор	3D-тор
ПС с процессором, ГБ/с	2	8	8	~ 20	9,6
ПС линка, ГБ/с	0,625	7,5	6,8	2	9,375
Агрегатная ПС линков, ГБ/с	5	120	–	40	186
Задержка между соседними узлами, мкс	2,5	1,0	1,0	< 1,0	1,4

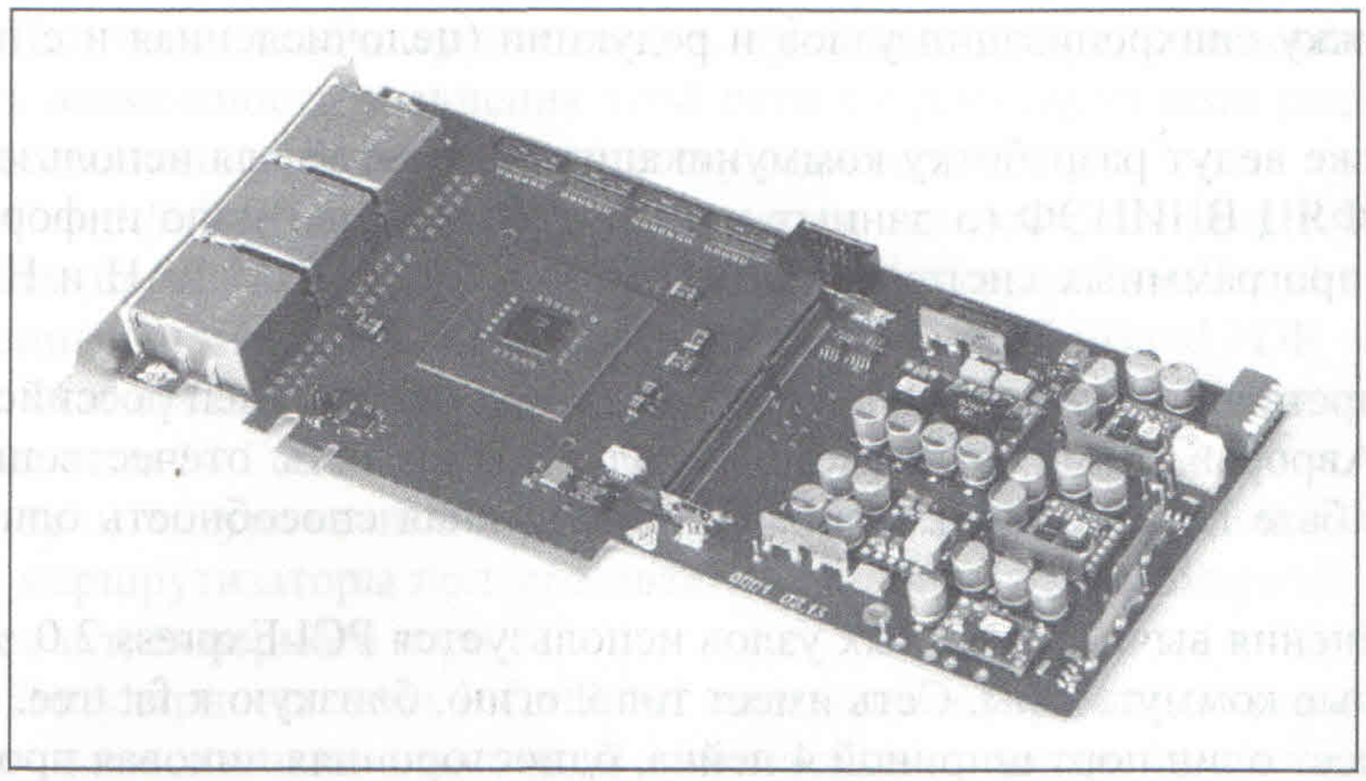

Рис. 1. Плата адаптера PCI Express

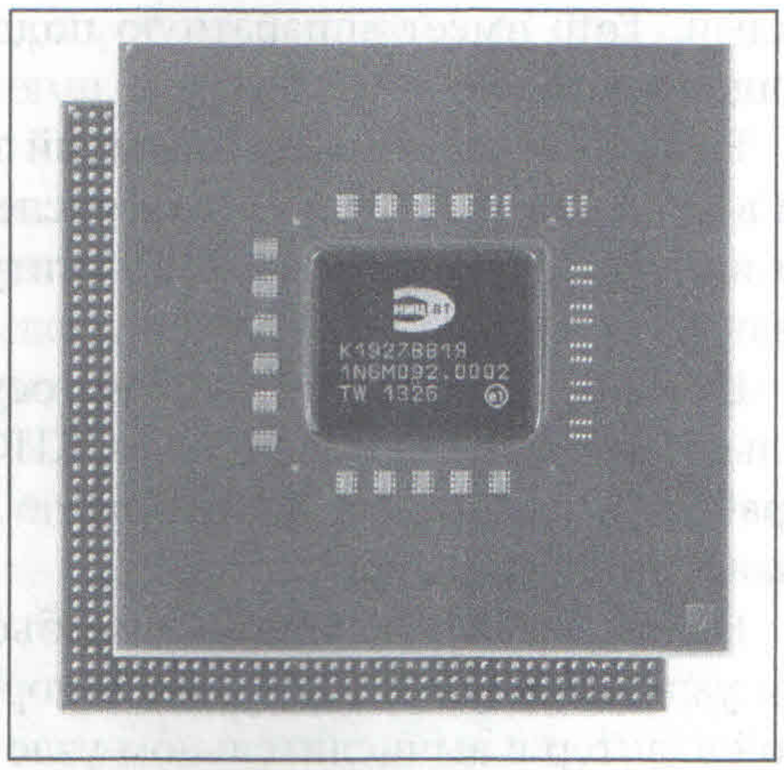

Рис. 2. СБИС «Ангара» (ЕС8430)

Маршрутизатор сети «Ангара» поддерживает топологию 4D-тор. Детерминированная маршрутизация выполняется согласно правилу пузырька (bubble rule) с сохранением порядка направлений +X, +Y, +Z, +W, −X, −Y, −Z, −W. Используется пять виртуальных каналов: два канала для детерминированной маршрутизации – блокируемый канал запросов и неблокируемый канал ответов; отдельный виртуальный канал используется для адаптивной маршрутизации (с возможностью перехода в неблокируемый детерминированный канал в случае потенциального дедлока); ещё два виртуальных канала используются для передачи сообщений по виртуальной подсети для коллективных операций. Детерминированная маршрутизация сохраняет порядок передачи пакетов и предотвращает появление дедлоков (deadlock); адаптивная маршрутизация использует для доставки пакетов один из возможных минимальных маршрутов, игнорируя порядок направлений, что позволяет обходить перегруженные и вышедшие из строя участки сети. Поддержка коллективных операций – широковещательной рассылки и редукции – реализована с помощью виртуальной подсети, имеющей топологию дерева, наложенного на многомерный тор [5].

Адаптер сети на аппаратном уровне поддерживает удаленные (RDMA) записи, чтения и атомарные операции. В текущей версии доступны атомарные операции двух типов – сложение и исключающее ИЛИ. В отдельном блоке реализована логика по агрегации пришедших из сети сообщений с целью повышения доли полезных данных на транзакцию при передаче через интерфейс с хост-системой.

На канальном уровне поддерживается отказоустойчивая передача пакетов с помощью нумерации пакетов, подсчёта для каждого контрольных сумм и повторной передачи в случае, если контрольная сумма, записанная в последнем флите пакета, не совпадает с вычисленной после передачи. Существует также механизм обхода отказавших каналов связи и узлов с помощью перестройки таблиц маршрутизации и использования нестандартных первого/последнего шагов маршрута пакета. Для выполнения различных сервисных операций, включая настройку/перестройку таблиц маршрутизации, и некоторых расчетов может использоваться сервисный процессор, взаимодействующий с адаптером по интерфейсу ELB. В качестве хост-интерфейса используется PCI Express 2.0.

Маршрутизатор имеет следующие основные блоки (рис. 3):

интерфейс с хост-системой, отвечающий за прием и отправку пакетов по хост-интерфейсу;

блок инжекции и эжекции, формирующий пакеты на посылку в сеть и разбирающий заголовки пакетов, пришедших из сети;

блок обработки запросов, обрабатывающий пакеты, требующие информации из памяти хост-системы (например, чтения или атомарные операции);

блок сети коллективных операций, обрабатывающий пакеты, связанные с коллективными операциями, в частности, с выполнением редукционных операций, порождением пакетов широковещательных запросов;

блок служебных операций, обрабатывающий пакеты, идущие в служебный сопроцессор и из него;

кроссбар, соединяющий входы с различных виртуальных каналов и входы с инжекторов с выходами на различные направления и эжекторы;

каналы связи для передачи и приема данных по определенному направлению;

блок передачи данных для отправки пакетов по данному направлению и блок приема и маршрутизации для приема пакетов и принятия решения о дальнейшей их судьбе.

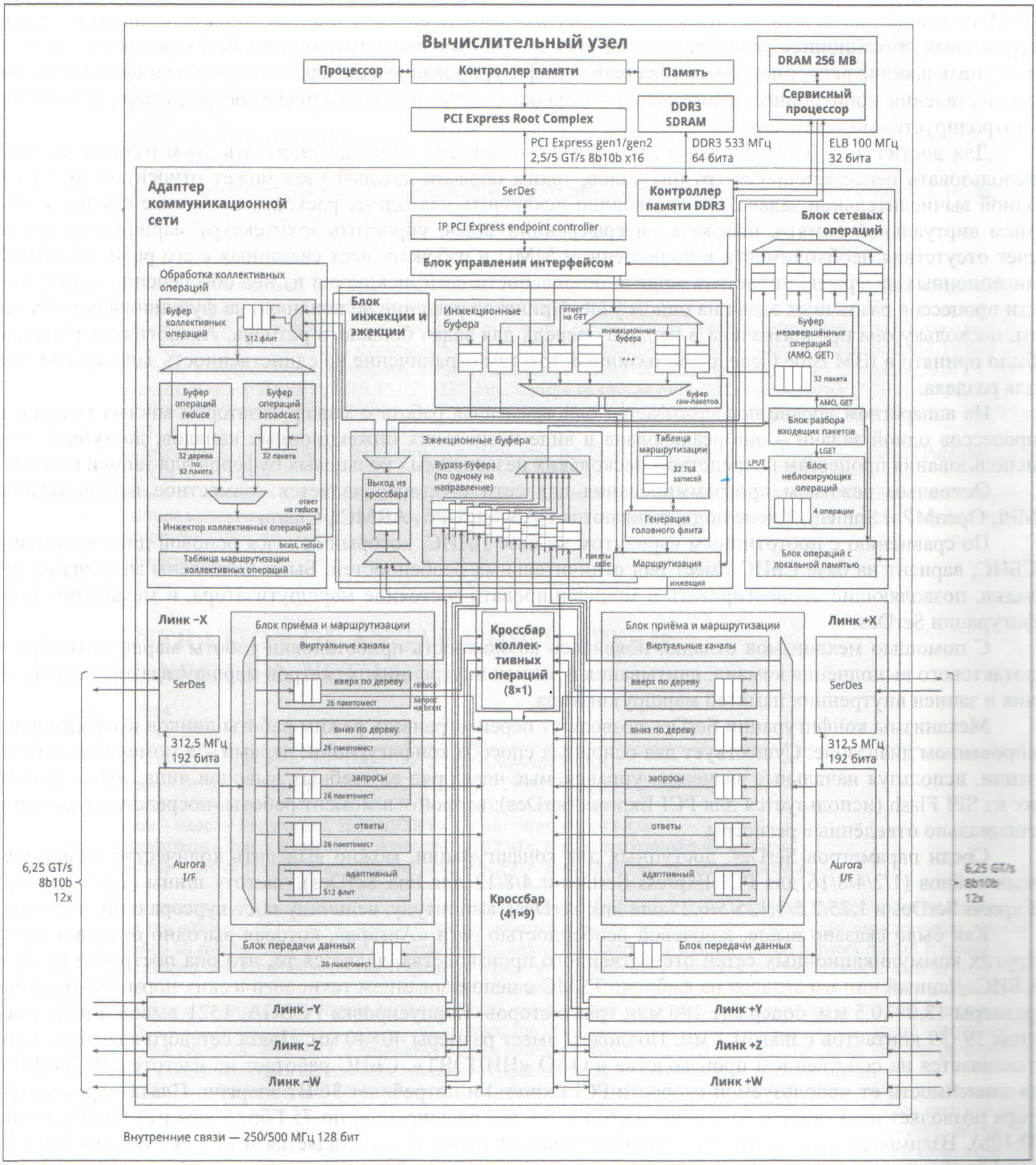

Рис. 3. Структура вычислительного узла с сетевым адаптером/маршрутизатором «Ангара»

Взаимодействие вычислительного узла, т.е. кода, исполняемого на центральном процессоре, с маршрутизатором осуществляется путем записи данных по адресам памяти, которые отображены на адреса ресурсных регионов маршрутизатора (memory-mapped input/output). Это позволяет приложению взаимодействовать с маршрутизатором без участия ядра ОС, что снижает накладные расходы при отправке пакетов, поскольку переключение в контекст ядра и обратно занимает существенное время, в сравнении с временем отправки пакета. Для отправки пакетов используется один из регионов памяти маршрутизатора, рассматриваемый как кольцевой буфер. Также имеется отдельный регион для выполнения

DMA-операций, когда данные читаются из памяти и записываются в удалённую память напрямую адаптером коммуникационной сети без участия процессора вычислительного узла. Ещё один регион содержит управляющие регистры, обеспечивающие конфигурирование адаптера и получение информации для осуществления мониторинга, диагностики и отладки. Доступ к тем или иным ресурсам маршрутизатора контролируется модулем ядра.

Для достижения большей эффективности было принято решение запретить двум разным задачам использовать пересекающиеся группы узлов, таким образом каждый узел может относиться только к одной вычислительной задаче. Это позволило исключить накладные расходы, связанные с использованием виртуальной памяти, избежать интерференции задач, упростить архитектуру маршрутизатора за счет отсутствия необходимости в полноценном MMU и избежать всех связанных с его работой коммуникационных задержек, упростить модель безопасности сети, исключив из нее обеспечение безопасности процессов различных задач на одном узле. Принятое решение не повлияло на функциональность сети, поскольку она предназначена в первую очередь для задач большого размера. Аналогичное решение было принято в IBM Blue Gene, с той разницей, что там ограничение на единственность задачи вводится для раздела.

На аппаратном уровне поддерживается одновременная работа с маршрутизатором многих потоков и процессов одной задачи – она реализована в виде нескольких инжекционных каналов, доступных для использования процессам посредством нескольких независимых кольцевых буферов для записи пакетов.

Основным режимом программирования для сети «Ангара» является совместное использование MPI, OpenMP и Shmem. Также поддерживаются GASNet, UPC, ARMCI, Charm++.

По сравнению с прототипным вариантом на базе ПЛИС, который являлся основой для подготовки СБИС, вариант на базе СБИС имеет ряд отличительных особенностей. Были добавлены механизмы отладки, позволяющие инспектировать и модифицировать состояние маршрутизатора, и механизмы конфигурации SerDes.

С помощью механизмов отладки появилась возможность приостановки работы маршрутизатора и потактового выполнения команд; считывания и записи внутренних регистров маршрутизатора; считывания и записи внутренних памятей маршрутзиатора.

Механизмы конфигурации SerDes позволяют перенастраивать режим работы линков в заранее фиксированном диапазоне. Существует два основных способа конфигурации: первый – в момент инициализации, используя начальные значения, управляемые через ряд служебных выводов чипа, или загружая их из SPI Flash (используется для PCI Express SerDes); второй – в момент работы, посредством записи в специально отведённые регистры.

Среди параметров SerDes, доступных для конфигурации, можно выделить количество используемых лейнов (1/2/4/8/16 для PCI Express SerDes и 4/8/12 для link SerDes), частоту шины (2,5/5 для PCI Express SerDes и 1,25/2,5/3,125/5/6,25 для link SerDes), амплитуду, величину пост-курсора и пре-курсора.

Как было сказано выше, ключевой особенностью сети «Ангара», которая выгодно отличает ее от других коммуникационных сетей отечественного производства, является то, что она построена на базе СБИС. Данный чип изготовлен на фабрике TSMC с использованием технологических норм 65 нм, имеет размеры 13,0×10,5 мм, содержит 180 млн транзисторов. Корпусировка FCBGA, 1521 вывод в виде массива 39×39 контактов с шагом 1 мм. Подложка имеет размеры 40×40 мм. Плата сетевого адаптера изготавливается на собственном производстве в ОАО «НИЦЭВТ». СБИС работает на частоте 250/500 МГц (в зависимости от используемой скорости PCI Express) и потребляет 36 Вт энергии. Плата маршрутизатора позволяет подключить до 6 линков (до 8 с платой расширения) по 75 Гбит/с каждый (кодирование 8b10b). Взаимодействие адаптера с вычислительным узлом осуществляется через PCI Express 2.0 x16 (80 Гбит/с, кодирование 8b10b).

Продвижение сети «Ангара» на рынок планируется осуществлять в двух вариантах: 1) как отдельную коммерческую сеть в виде плат PCI Express (см. рис. 1) для кластерных систем со стандартными процессорами и чипсетами; 2) как интегрированный компонент в составе разрабатываемой в ОАО «НИЦЭВТ» в рамках проекта «Ангара» вычислительной платформы, что позволит объединить до 32 тысяч узлов в составе суперкомпьютера транспетафлопсного уровня производительности.

Параллельно с выпуском СБИС первого поколения продолжается дальнейшая разработка и оптимизация архитектуры сети «Ангара», готовится макет М4 (на базе ПЛИС Virtex 7). Опыт эксплуатации предыдущих макетов и кластера с маршрутизаторами на базе СБИС является основой для разработки

принципов работы сети «Ангара» второго поколения. Основные доработки будут направлены на поддержку большего числа топологий, повышение безопасности выполнения прикладных задач на узлах, добавление аппаратной поддержки атомарных операций с возвратом значений, поддержки GPU Direct, оптимизацию RDMA-операций и поддержки большого числа тредов на узле.

- В целом создание отечественной СБИС маршрутизатора высокоскоростной коммуникационной сети – это реальный шаг на пути к созданию полностью отечественного суперкомпьютера субэкзафлопсной производительности, значительно приближающий Россию к ведущим мировым державам в области высокопроизводительных вычислений.

🕮 Литература

1. *Dongarra J.* Visit to the National University for Defense Technology Changsha, China. Technical report, Oak Ridge National Laboratory, USA, 2013. http://www.netlib.org/utk/people/JackDongarra/PAPERS/tianhe-2-dongarra-report.pdf
2. *Корж А.А., Макагон Д.В., Жабин И.А., Сыромятников Е.Л.* и др. Отечественная коммуникационная сеть 3D-тор с поддержкой глобально адресуемой памяти для суперкомпьютеров транспетафлопсного уровня производительности // Труды междунар. науч. конф. «Параллельные вычислительные технологии» (ПаВТ'2010). (Уфа, 29 марта – 2 апреля 2010 г.). Челябинск: Издательский центр ЮурГУ. 2010. С. 227–237. http://omega.sp.susu.ac.ru/books/conference/PaVT2010/full/134.pdf
3. *Dally W. and Towles B.* Principles and Practices of Interconnection Networks. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
4. *Duato J., Yalamanchili S. and Lionel N.* Interconnection Networks: An Engineering Approach. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
5. *Сыромятников Е.Л., Макагон Д.В., Парута С.И., Румянцев А.А.* Реализация аппаратной поддержки коллективных операций в маршрутизаторе высокоскоростной коммуникационной сети с топологией «многомерный тор» // Успехи современной радиоэлектроники. 2012. № 1. С. 11–15.

Поступила 1 ноября 2013 г.

FIRST GENERATION OF ANGARA HIGH-SPEED INTERCONNECTION NETWORK

I.A. Zhabin – Head RTL Designer, JSC «NICEVT». E-mail: zhabin@nicevt.ru
D.V. Makagon – Senior Sostware Engineer, JSC «NICEVT». E-mail: makagond@nicevt.ru
D.A. Polyakov – Sostware Engineer, JSC «NICEVT». E-mail: polyakov@nicevt.ru
A.S. Simonov – Head of Department, JSC «NICEVT». E-mail: simonov@nicevt.ru
E.L. Syromyatnikov – Senior Research Engineer, JSC «NICEVT». E-mail: syromiatnikov@nicevt.ru
A.N. Shcherbak – Senior Digital Design Engineer, JSC «NICEVT». E-mail: andrey.shcherbak@nicevt.ru

In our days supercomputers are built with tens and even hundreds of thousands of nodes connected by high-speed interconnection networks, so the efficiency of the whole system is largely determined by the interconnection network. All interconnection networks can be classified as commercially available and proprietary/custom. The commodity interconnection network market is mostly dominated by InfiniBand and Ethernet networks. Typical examples of the proprietary networks are interconnects used in Blue Gene, Tianhe-2 and Tianhe-1A, K Computer and Cray Titan. The commercially available networks being universal are hardly suitable for the high-end systems with tough requirements on the scalability, reliability and performance. The purchase of the top proprietary networks in Russia is practically impossible. This situation makes the development of a competitive Russian high-speed interconnect an extremely topical issue. JSC «NICEVT» is developing Angara interconnection network for precisely this niche. The first generation of network routers based on Angara EC8430 ASIC has already appeared in 2013. This is a unique result for Russia.
The EC8430 ASIC was fabricated with TSMC 90 nm process, die size is 13.0x10.5 mm with over 180 million transistors. The network adapter PCB is manufactured by JSC "NICEVT". Adapter has up to 8 high-speed connectors operating up to 75 Gbps each and a PCI Express 2.0 x16 interface with a host system. Angara interconnection network is planned to be supplied in two variants: as standard commercially available PCI Express network adapters for cluster systems with standard processors and chipsets, and as an integrated part of Angara Computing Platform server boards.

References

1. Dongarra J. Visit to the National University for Defense Technology Changsha, China. Technical report, Oak Ridge National Laboratory, USA, 2013. http://www.netlib.org/utk/people/JackDongarra/PAPERS/tianhe-2-dongarra-report.pdf
2. Korzh A.A., Makagon D.V., Zhabin I.A., Sy'romyatnikov E.L. i dr. Otechestvennaya kommunikaczionnaya set' 3D-tor s podderzhkoj global'no adresuemoj pamyati dlya superkomp'yuterov transpetaflopsnogo urovnya proizvoditel'nosti // Trudy' mezhdunar. nauch. konf. «Parallel'ny'e vy'chislitel'ny'e texnologii» (PaVT'2010). (Ufa, 29 marta – 2 aprelya 2010 g.). Chelyabinsk: Izdatel'skij czentr JuurGU. 2010. S. 227–237. http://omega.sp.susu.ac.ru/books/ conference/PaVT2010/full/134.pdf
3. Dally W. and Towles B. Principles and Practices of Interconnection Networks. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
4. Duato J., Yalamanchili S. and Lionel N. Interconnection Networks: An Engineering Approach. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
5. Sy'romyatnikov E.L., Makagon D.V., Paruta S.I., Rumyanczev A.A. Realizacziya apparatnoj podderzhki kollektivny'x operaczij v marshrutizatore vy'sokoskorostnoj kommunikaczionnoj seti s topologiej «mnogomerny'j tor» // Uspexi sovremennoj radioe'lektroniki. 2012. № 1. S. 11–15.