

РАЗВИТИЕ АППАРАТНОЙ ПЛАТФОРМЫ РЕКОНФИГУРИРУЕМЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Н.Н. Дмитренко, И.А. Каляев, И.И. Левин, Е.А. Семерников

В последние годы наметилась устойчивая тенденция использования ПЛИС для высокопроизводительных вычислений. В первой строчке списка TOP-500 значится суперЭВМ Jaguar - Cray XT5-HE, произведенная фирмой Cray Inc., с пиковой производительностью 2331.00 Тфлопс, в составе которой в качестве сопроцессоров используются ПЛИС большой интеграции. В целом ряде разработок высокопроизводительной вычислительной техники последних лет также применяются ПЛИС, позволяющие существенно увеличить их производительность на многих задачах. В этих системах ПЛИС, так же как и в Jaguar - Cray XT5-HE, используются как дополнение к микропроцессорам, позволяющее выполнять те фрагменты вычислений, которые трудно или неэффективно реализуются на универсальных микропроцессорах.

Однако, как это показано в [1,2], ПЛИС обладают значительно большим вычислительным потенциалом, чем тот, который раскрывается при их использовании в качестве сопроцессоров. В полной мере вычислительный потенциал ПЛИС может быть реализован в системах, где на ПЛИС возлагается основная вычислительная нагрузка и где они используются в составе больших вычислительных полей.

Ниже будет рассмотрен ряд созданных и перспективных вычислительных систем с большими вычислительными полями из ПЛИС, обладающих производительностью в несколько десятков терафлопс, которые создаются в рамках концепции многопроцессорных вычислительных систем с программируемой архитектурой [1, 2]. Эта концепция была впервые сформулирована академиком А.В. Каляевым в 70–80-х годах прошлого века и развивается в НИИ МВС ЮФУ (г. Таганрог) уже более 30 лет.

В отличие от кластерных суперЭВМ, обладающих «жесткой» архитектурой, архитектура реконфигурируемых вычислительных систем (РВС) может динамически изменяться в процессе функционирования. В результате у пользователя появляется возможность адаптации архитектуры вычислительной системы под структуру решаемой задачи. В качестве элементной базы для построения РВС используются ПЛИС большой интеграции, соединенные в вычислительные поля. Вычислительные структуры, реализуемые в доступном пользователю ресурсе ПЛИС, обеспечивают высокую реальную производительность и пропорциональный рост производительности при увеличении задействованного оборудования.

Рассмотрим особенности построения и развития аппаратной платформы высокопроизводительных РВС на примере разработок НИИ МВС ЮФУ последних лет. Для сравнения будем рассматривать три проекта, направленных на создание РВС высокой производительности.

В 2007-2009 годах в рамках проекта по госконтракту № 02.524.12.4002 (шифр «Большая медведица») было создано семейство программно-совместимых реконфигурируемых высокопроизводительных вычислительных систем производительностью от 0,025 Тфлопс до 6 Тфлопс, в состав которого входят: РВС-5 – высокопроизводительная система производительностью 6 Тфлопс; РВС-1Р и РВС-1К – системы производительностью более 1 Тфлопс; РВС-0.2-РС – рабочая станция производительностью 300 Гфлопс; РУПК-50 и РУПК-25 – ускорители персональных компьютеров производительностью 50 и 25 Гфлопс. Старшие представители семейства РВС-5, РВС-1Р и РВС-0.2-РС выполнены на принципах модульной наращиваемости и обладают почти линейным ростом реальной производительности в зависимости от увеличения аппаратного ресурса [2, 3]. Реальная производительность всех представителей семейства РВС на задачах различных классов составляет более 50% от указанной пиковой производительности. Старшие представители семейства выполнены на основе единственного базового модуля 16V5-75. Модели РВС-5 и РВС-1Р содержат соответственно пять и одну 19' стойку СТ-1Р. Каждая стойка включает четыре блока РВС-0.2-ВБ 6U, имеющих в своем составе до четырех базовых модулей 16V5-75 с пиковой производительностью до 200 Гфлопс каждый.

ОКР «Орион», выполнение которой планируется завершить в 2010 году, предусматривает создание модульно наращиваемой системы с производительностью до 20 Тфлопс в 19' стойке. Стойка содержит до 24 блоков «Орион» 1U на основе четырех базовых модулей 16V5-125OP с производительностью до 250 Гфлопс каждый.

ОКТР «Орион-6», выполнение которой также планируется завершить в 2010 году, предусматривает создание опытных образцов базовых модулей и блоков модульно наращиваемой системы с производительностью до 30 Тфлопс в 19' стойке. Стойка содержит до 24-х блоков «Орион-6» 1U на основе четырех базовых модулей 8V6-200OP с производительностью до 400 Гфлопс каждый на операциях с одинарной точностью.

Общие принципы создания РВС подробно рассмотрены в [1, 2].

В таблице 1 приведены пиковые производительности узлов высокопроизводительных РВС. Производительность соответствует обработке данных с одинарной ($P_{i_{32}}$) и двойной ($P_{i_{64}}$) точностью в соответствии со стандартом IEEE-754 для базовых модулей, блоков и стоек описанных изделий.

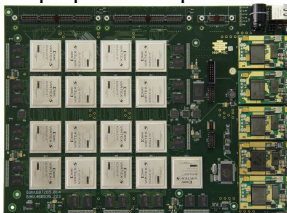
Таблица 1. Производительность базовых модулей, блоков и стоек

| Изделие | Базовый модуль Pi_{32}/Pi_{64} (Гфлопс) | Блок Pi_{32}/Pi_{64} (Гфлопс) | Стойка Pi_{32}/Pi_{64} (Гфлопс) |
|---------------------------------|--|------------------------------------|--------------------------------------|
| Госконтракт № 02.524.12.4002 | 16V5-75 200/75 | PBC-0.2-ВБ 800/300 | СТ-1P 2400/1200 |
| Проект «Орион» | 16V5-125OP 250/85 | Блок «Орион» 1000/340 | Стойка «Орион» 19200/8160 |
| Проект «Орион-6» | 16V6-200OP 400/125 | Блок «Орион-6» 1600/500 | Стойка «Орион-6» 30000/12500 |

Базовые модули всех описываемых систем имеют похожие структуры и содержат одинаковые типовые узлы. Основу вычислительной части базовых модулей составляет вычислительное поле из ПЛИС, соединенных между собой ортогональной системой связей по принципу близкодействия. К ПЛИС вычислительного поля подключены микросхемы памяти и разъемы для объединения вычислительных полей базовых модулей в вычислительные поля блоков. На базовых модулях имеется также ряд вспомогательных подсистем, предназначенных для обеспечения их работоспособности. Для связи с внешними устройствами, настройки и управления ресурсами предусмотрена дополнительная ПЛИС контроллера базового модуля, имеющая соответствующие внешние интерфейсы.

На рис. 1. *а, б и в* показаны печатные платы базовых модулей 16V5-75, 16V5-125OP и компоновка 8V6-200OP соответственно.

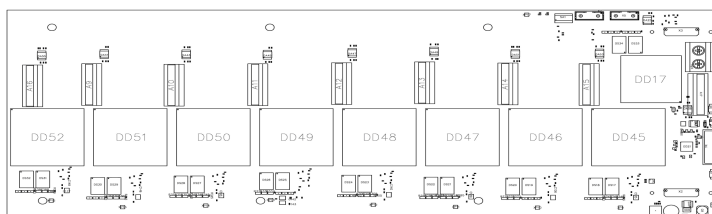
Таким образом, базовые модули представляют собой мощные вычислительные узлы производительностью до 400/125) Гфлопс, на основе которых строятся вычислительные блоки, содержащие до четырех базовых модулей производительностью от 800 до 1600 Гфлопс. В то же время базовые модули обладают достаточной автономностью и могут легко комплексоваться с персональным компьютером типа IBM PC в качестве ускорителей и использоваться при решении различных задач.



а



б



в

Рис. 1. Печатные платы базовых модулей

В таблице 2 приведены производительности базовых модулей на задачах криптологии, использующих преобразования символьных данных, и задачах на основе арифметики с плавающей запятой одинарной точности. Увеличение производительности базового модуля 16V5-125OP по сравнению с 16V5-75 для задач, использующих представление данных с плавающей запятой, при наличии одинакового аппаратного ресурса

вычислительного поля объясняется наличием большего количества каналов распределенной памяти – 32 и 20 соответственно.

| Базовый модуль | Криптология (Ключей в секунду) | Плавающая арифметика (Тфлопс) |
|----------------|--------------------------------|-------------------------------|
| 16V5-75 | 128,0*10 ⁹ | 200/75 |
| 16V5-125OP | 128,0*10 ⁹ | 250/85 |
| 8V6-200OP | 215,0*10 ⁹ | 400/125 |

В таблице 3 приведены суммарные скорости передачи данных между вычислительным полем и блоками распределенной памяти, между вычислительными полями других базовых модулей и внутри вычислительного поля для базовых модулей 16V5-75, 16V5-125OP и 8V6-200OP.

| Базовый модуль | С блоками распределенной памяти (Тбит/с) | С другими базовыми модулями (Тбит/с) | Между ПЛИС вычислительного поля (Тбит/с) |
|----------------|--|--------------------------------------|--|
| 16V5-75 | 0,205 | 0,268 | 3,225 |
| 16V5-125OP | 0,307 | 0,499 | 2,533 |
| 8V6-200OP | 0,610 | 0,638 | 2,100 |

Снижение скорости обмена между ПЛИС вычислительного поля объясняется тем, что вычислительное поле 8V6-200OP содержит всего восемь ПЛИС XC6VLX240T-1FFG1759 вместо шестнадцати XC5VLX110-2FF1153 фирмы Xilinx для базовых модулей 16V5-75, 16V5-125OP.

Первый этап наращивания ресурса вычислительного поля на основе вычислительных полей базовых модулей выполняется при создании блоков РВС-0.2-ВБ, блока «Орион» и блока «Орион-6». Основу этих изделий составляют объединенные вычислительные поля, включающие в себя вычислительные поля четырех базовых модулей, соединенные между собой в единый вычислительный ресурс быстрыми LVDS-каналами (см. рис.2 и рис. 3).

На рис. 2 прямыми двунаправленными стрелками показаны каналы между ПЛИС вычислительных полей базовых модулей 16V5-75. Изогнутыми стрелками показаны LVDS-каналы, соединяющие вычислительные поля отдельных базовых модулей в вычислительные поля блоков. Следует отметить, что связи внутри вычислительного поля 16V5-75 также выполнены по стандарту LVDS и обеспечивают суммарную скорость передачи данных свыше 3 Тбит в секунду на частоте 1200 МГц.

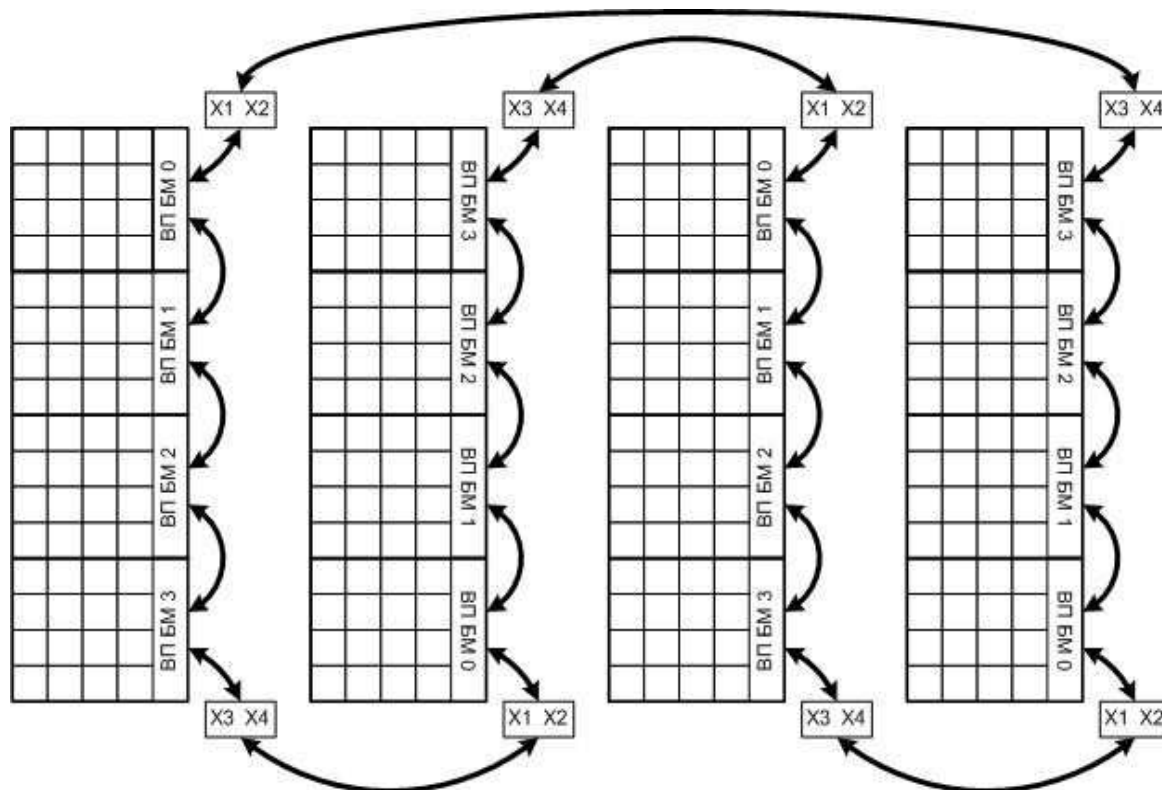


Рис.2. Объединение вычислительных полей базовых модулей в вычислительные поля блока РВС-0.2-ВБ

На рис 3. показано вычислительное поле блока «Орион». В отличие от базового модуля 16V5-75, где связи для объединения имеются только у четырех ПЛИС из шестнадцати, в базовом модуле 16V5-125OP (как и

в 8V6-2000P) все ПЛИС вычислительного поля имеют LVDS-каналы для наращивания вычислительного поля. Это обеспечивает большую скорость обмена информацией между базовыми модулями в составе блока «Орион» и «Орион-6».

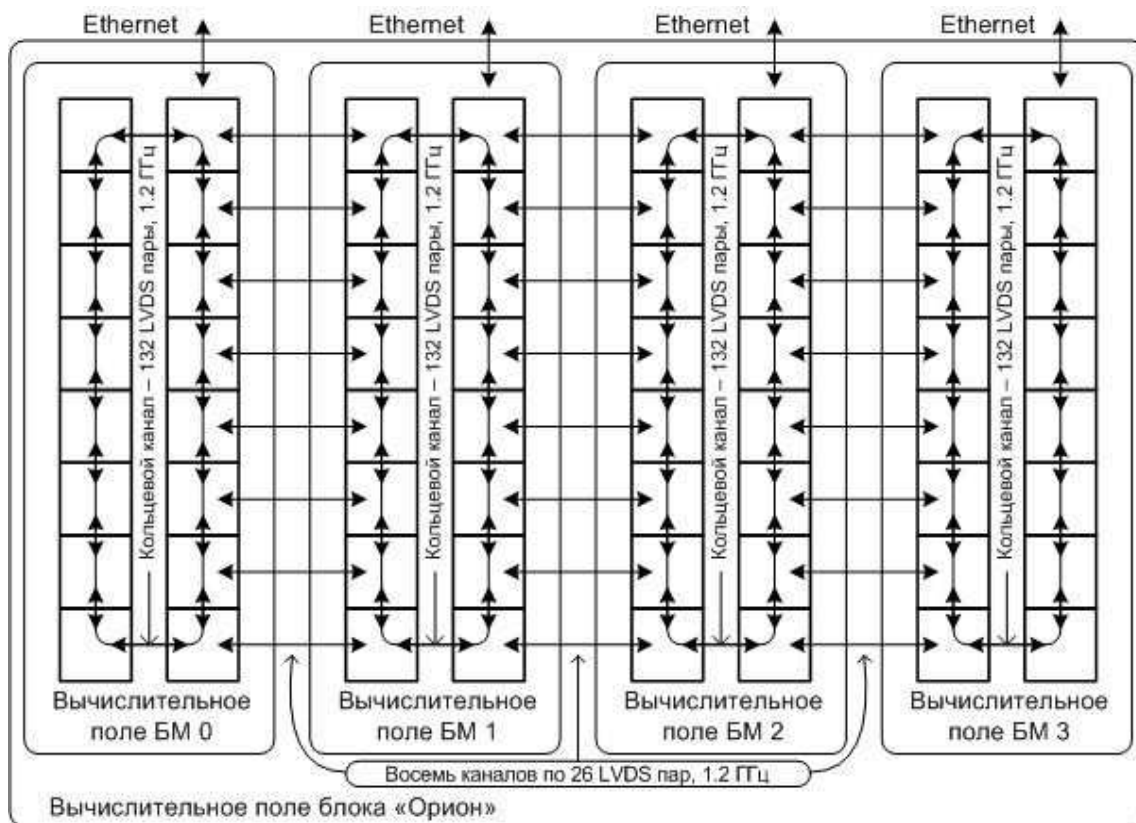
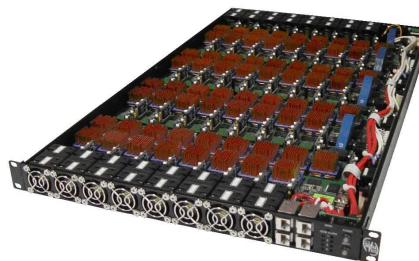


Рис.3. Объединение вычислительных полей базовых модулей в вычислительное поле блока «Орион»

На рис. 4 показаны блоки PBC-0.2-ВБ и блок «Орион» со снятыми верхними крышками. Внешний вид, компоновка и размеры блока «Орион-6» соответствуют параметрам блока «Орион».



а

б

Рис. 4. Печатные платы базовых модулей

На рис. 4а видны кабели синего цвета, которые соединяют вычислительные поля отдельных базовых модулей в вычислительное поле PBC-0.2-ВБ. Объединение базовых модулей 16V5-1250P в блоке «Орион» (рис. 4б) и объединение вычислительных полей 8V6-2000P в блоке «Орион-6» выполнено не кабелями, а специальными жесткими соединителями в виде интерфейсных плат с двумя LVDS-разъемами.

В таблице 4 приведены значения производительности блоков PBC-0.2-ВБ, блока «Орион» и оценки производительности блока «Орион-6» на задачах различных классов: криптологии, цифровой обработки сигналов, линейной алгебры и математической физики.

| Таблица 4. Производительность блоков PBC-0.2-ВБ, «Орион» и блока «Орион-6» | | | | |
|--|------------------------------|------------------|-------------------------------|-------------------------|
| Задачи | Криптология, (Ключей/сек) | ЦОС, (Гфлопс) | Линейная алгебра, (Гфлопс) | Мат. Физика (Гфлопс) |
| PBC-0.2-ВБ | $512,7 \cdot 10^9$ | 647,7 | 423,2 | 535,2 |

| | | | | |
|-----------|-----------------------|--------|-------|--------|
| «Орион» | 512,7*10 ⁹ | 809,6 | 528,7 | 669,0 |
| «Орион-6» | 860,0*10 ⁹ | 1380,3 | 895,7 | 1140,3 |

Из таблицы 4 видно, что вычислительные блоки PBC-0.2-ВБ и «Орион», имея практически одинаковые ресурсы вычислительных полей, незначительно различаются по производительности. Увеличение производительности блока «Орион» объясняется двумя факторами. Первый из них – это большая пропускная способность каналов, объединяющих вычислительные поля базовых модулей (см. таблицу 2). Второй фактор – большее количество каналов распределенной памяти, подключенных к вычислительному полю: 80 каналов в блоке PBC-0.2-ВБ и 128 каналов в блоке «Орион».

Однако следует отметить, что величины производительностей этих блоков, приходящиеся на единицу объема, существенно различаются в пользу блока «Орион». Это связано с более рациональной компоновкой блока, объем которого составляет примерно 18% от общего объема блока PBC-0.2-ВБ или 32% от объема его вычислительной части.

В таблице 5 приведены значения удельной производительности блоков PBC-0.2-ВБ, блока «Орион» и блока «Орион-6» на задачах криптологии и удельная пиковая производительность на задачах, использующих для представления данных плавающую запятую одинарной точности.

| Производительность | (Ключей в сек/дм ³) | (Гфлопс/дм ³) |
|--------------------|---------------------------------|---------------------------|
| PBC-0.2-ВБ | 11*10 ⁹ | 16,7 |
| «Орион» | 33,2*10 ⁹ | 64,9 |
| «Орион-6» | 56,4*10 ⁹ | 110,3 |

Следующий этап наращивания ресурса вычислительных полей на основе вычислительных полей блоков PBC-0.2-ВБ может быть выполнен на уровне стоек. Ниже приводится пример объединения вычислительных полей четырех блоков PBC-0.2-ВБ в составе стойки СТ-1Р, предназначенной для комплектования представителей семейства PBC – PBC-1Р производительностью свыше 1,2 Тфлопс и PBC-5 производительностью свыше 6 Тфлопс.

Вычислительные поля четырех вычислительных блоков PBC-0.2-ВБ объединяются с помощью LVDS-каналов в единый вычислительный ресурс, содержащий до 16-ти базовых модулей 16V5-75 с общей пиковой производительностью 1,2 Тфлопс, как это показано на рис. 5.

Межблочные связи являются продолжением межмодульных связей и, в свою очередь, продолжением связей между ПЛИС вычислительных полей базовых модулей. В целом подобная организация быстрых связей реализуется в составе стойки СТ-1Р глобальный LVDS-канал передачи данных с единым темпом продвижения информации в объединенном вычислительном поле стойки, содержащем 256 ПЛИС, или, с учетом их интеграции, свыше 28 миллиардов эквивалентных вентилялей.

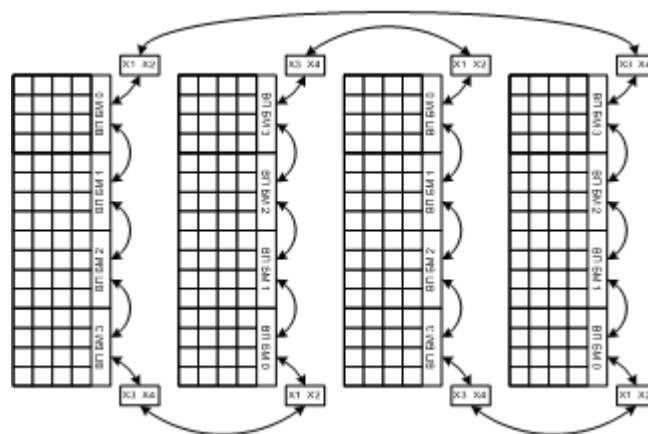


Рис. 5. Вычислительное поле стойки СТ-1Р

Блоки «Орион» и «Орион-6», как и стойка СТ-1Р, не имеют внешних быстрых LVDS-каналов для дальнейшего наращивания ресурса вычислительных полей. Тем не менее, вычислительное поле одного блока «Орион-6», включающего 32 ПЛИС XC6VLX240T-1FFG1759, составляет около 7,7 миллиардов вентилялей. Наращивание производительности систем, составленных из блоков «Орион», «Орион-6» и стоек СТ-1Р, выполняется с использованием сетевых технологий посредством каналов Ethernet, как это реализовано, например, в PBC-5 [3, 4].

Программирование PBC отличается от программирования суперЭВМ традиционной архитектуры, поскольку включает программирование структуры вычислительной системы в поле логических ячеек ПЛИС. Программирование структуры системы вызывает у пользователей наибольшие трудности, так как традиционно они привыкли программировать только организацию вычислительного процесса, в то время как для программирования вычислительных структур требуется специалист с квалификацией схемотехника. В

настоящее время в НИИ МВС ЮФУ ведутся работы по разработке и совершенствованию системного программного обеспечения РВС и программных средств, призванных облегчить создание прикладных параллельных программ. Созданные и уже поставляемые с изделиями системное программное обеспечение и средства разработки, а также ресурсонезависимый язык высокого уровня COLAMO с неявным описанием параллелизма обеспечивают потенциальным пользователям возможности создания прикладных программ для РВС без привлечения специалистов в области схемотехники ПЛИС и по сложности приближены к обычному программированию для традиционных многопроцессорных ЭВМ.

В заключение следует отметить, что переход на новую компоновку блоков, позволившую сосредоточить в пределах блоков «Орион» и «Орион-6» высотой 1U мощный вычислительный ресурс на основе ПЛИС, обеспечивает удельную производительность РВС на уровне лучших мировых показателей для суперЭВМ с кластерной архитектурой. Однако, учитывая тот факт, что РВС обладают реальной производительностью на широком классе задач до 10 раз выше, чем кластерные суперЭВМ, можно сделать вывод о том, что РВС значительно превосходят их по вычислительной эффективности.

Следует также подчеркнуть, что РВС являются принципиально новым направлением развития высокопроизводительной вычислительной техники, которые, в отличие от кластерных суперЭВМ, предоставляют пользователю возможность создавать в базовой архитектуре виртуальные специализированные вычислители, структура которых адекватна структуре решаемой задачи. Это, в свою очередь, обеспечивает высокую эффективность вычислений и близкий к линейному рост производительности при наращивании вычислительного ресурса.

РВС предназначены для решения вычислительно трудоемких задач, решение которых на суперЭВМ традиционной архитектуры является затруднительным. В качестве примера можно привести задачи, требующие при выполнении на суперЭВМ традиционных архитектур недопустимо больших временных затрат, а также задачи, требующие обработки информации с использованием нестандартных представлений данных.

ЛИТЕРАТУРА:

1. Каляев А.В., Левин И.И. Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. - М.: Янус-К, 2003. – 380 с.
2. Каляев И.А., Левин И.И., Семерников Е.А., Шмойлов В.И. Реконфигурируемые мультимедийные вычислительные структуры /Изд. 2-е, перераб. и доп. / Под общ. Ред. И.А. Каляева. - Ростов-на-Дону: Изд-во ЮНЦ РАН, 2009. – 344 с.
3. Дмитренко Н.Н., Каляев И.А., Левин И.И., Семерников Е.А. Семейство многопроцессорных вычислительных систем с динамически перестраиваемой архитектурой. Вестник компьютерных и информационных технологий. - М.: Изд-во Машиностроение, 2009. – Ч. 1. - №6. – С.2-8. –Ч. 2.- № 7. – С.2-10.
4. Каляев И.А., Левин И.И., Семерников Е.А. Семейство вычислительных систем с высокой реальной производительностью на основе ПЛИС // Параллельные вычислительные технологии (ПаВТ'2010): Труды международной научной конференции (Уфа, 29 марта – 2 апреля 2010 г.) [Электронный ресурс] – Челябинск: Издательский центр ЮУрГУ, 2010. – 723 с.С. 199 – 210.