УДК 004.3'12

К.т.н. С.В. Юрлин, д.т.н. В.М. Фельдман
(АО «МЦСТ», ПАО «ИНЭУМ им. И.С. Брука»)

S. Yurlin, V. Feldman

ПРОБЛЕМЫ РЕАЛИЗАЦИИ МОДУЛЕЙ НА ОСНОВЕ МИКРОПРОЦЕССОРА АО «МЦСТ» НОВОГО ПОКОЛЕНИЯ

DEVELOPMENT ISSUES OF MODULES BASED ON AO «MCST» NEW GENERATION MICROPROCESSOR

Статья иллюстрирует возможность совместить процесс проектирования нового микропроцессора с решением базовых задач, определяющих конструкторско-технологические принципы построения вычислительных средств на его основе, включая компоновку оперативной памяти и реализацию выводов корпуса микросхемы.

This article illustrates the possibility of combination new microprocessor designing with solving basic tasks of it integration in different modules due to design and technology principles, including RAM and pinout table.

Ключевые слова: микропроцессор, процессорный модуль, многопроцессорный вычислительный комплекс, оперативная память, матрица выводов.

Keywords: microprocessor, processor module, multiprocessor computer system, RAM, pinout table.

Введение

Специалистами АО «МЦСТ» в сжатые сроки выполняется большой объем проектирования отечественных высокопроизводительных вычислительных средств на основе микропроцессоров собственной разработки. Это обусловило необходимость совмещать и взаимно учитывать решение базовых проблем, последовательный и всесторонний анализ которых был бы весьма трудоемок и продолжителен по времени. Такой подход освоен в

проектной практике компании и при профессиональной, глубоко продуманной реализации в целом дает ожидаемые результаты [1]. В данной статье описание этого подхода приводится по отношению к исследованию, проведенному для оценки и выбора конструкторско-технологических основ, обеспечивающих полномасштабное применение проектируемого микропроцессора нового поколения с архитектурой SPARC v9 для создания многопроцессорных многомашинных вычислительных систем с неоднородным доступом к оперативной памяти (NUMA).

1. Исходные положения и цель исследования

Выполненный анализ исходил из ряда положений, фиксирующих предыдущий опыт и возможности компании. Определяющее значение имели следующие факторы:

А. Функциональные возможности проектируемого микропроцессора в составе вычислительных систем реализуются путем использования построенных на его базе вычислительных модулей, выпускаемых в двух вариантах:

1) *одномодульное исполнение* [2]. Конструктивной основой модуля является панель, или материнская плата, содержащая до четырех микропроцессоров, до двух контроллеров периферийных интерфейсов КПИ-2 и сопутствующее оборудование. Как правило, такой модуль может комплектоваться картами расширения. Объединение входящих в модуль микропроцессоров высокоскоростными каналами позволяет создать рассчитанный на независимую работу вычислительный комплекс (ВК);

2) *многомодульное исполнение* [3]. Каждый модуль содержит набор элементов, необходимых для его функционирования в качестве однопроцессорной вычислительной машины. Объединение модулей для создания четырехпроцессорного ВК и подача на них электропитания осуществляются посредством коммутационной панели. Периферийное оборудование может быть подключено к вычислительному модулю или размещено на отдельных модулях, соединение с которыми также осуществляется через панель.

Эти варианты отличаются друг от друга взаимным расположением и типом приме-

няемых компонентов, трассами прокладки высокоскоростных каналов. Данные аспекты должны учитываться при разработке нового микропроцессора.

Б. Соответственно специфике применения в микропроцессоре реализуется большое число каналов:

- два канала оперативной памяти;
- три канала межпроцессорного взаимодействия шириной x16 (IPLink A, B, C), предназначенных для организации связей внутри одного ВК;
- один канал межмашинного взаимодействия шириной x8 (RDMA), предназначенный для организации связи между ВК в рамках одного вычислительного кластера;
- один канал для взаимодействия с контроллером периферийных интерфейсов шириной x16 (WLink).

Каналы оперативной памяти имеют встроенную опцию отключения питания микропроцессора с сохранением данных в модулях памяти.

Высокоскоростные каналы IPLink, WLink и RDMA обеспечивают частоту передачи данных до 6 Гб/с и реализуются на одинаковых блоках CEI 6G SR x4 (Common Electrical I/O), предназначенных для близкого взаимодействия. Максимальная длина связей от передатчика до приёмника составляет 20 см [4]. В данном микропроцессоре требуется 18 таких блоков.

В. Широкий диапазон производительности микропроцессора обеспечивается регулировкой тактовой частоты и уровней напряжения питания, что также дает экономию мощности при его низкой загруженности. Предусмотренное число режимов обеспечивается выделением нескольких независимых доменов процессорной логики, в которых частота и уровень напряжения регулируются независимо. Основных доменов три: два кластера процессорных ядер (по четыре ядра в каждом) и домен северного моста. Дополнительно присутствует большое количество доменов питания периферийных блоков.

Приведенное в статье исследование ставило целью на базе этих и других установок

еще в процессе проектирования нового микропроцессора сформулировать основные конструктивно-технологические решения в части обоих вариантов исполнения вычислительных модулей и в контексте этой работы определиться с такими важными проблемами, как компоновка оперативной памяти и реализация матрицы выводов корпуса микросхемы.

2. Одномодульное исполнение

В одномодульном исполнении вычислительный комплекс реализуется путём размещения на одной панели формата WTX (или большего формата) четырёх микропроцессоров, объединённых через каналы IPLink, до двух микросхем КПИ-2, модулей памяти и набора периферии, в которую в т.ч. входят слоты расширения.

Подобный вариант компоновки модуля приведён на рис. 1.

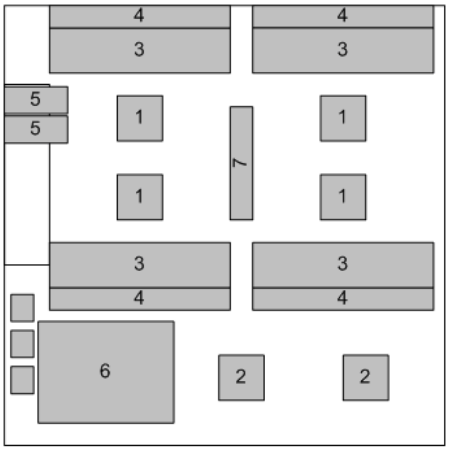

Рисунок 1. Размещение компонентов на четырёхпроцессорной панели:
1 – микропроцессор; 2 – КПИ-2; 3 – зона сокетов оперативной памяти; 4 – зона преобразователей напряжения DC-DC; 5 – разъёмы QSFP для соединения модулей через RDMA; 6 – зона периферийных слотов; 7 – микропереключатели для задания режимов

Данная компоновка обусловлена следующим. Сложность проектирования столь больших изделий уменьшается в результате размещения на панели четырех экземпляров унифицированного процессорного блока. Плоскости блоков разворачиваются так, что расположение их сторон относительно друг друга обеспечивает длину проводников каналов IPLink, не превышающую 20 см, т.е. трассировка межпроцессорных каналов осу-

ществляется в центре панели между процессорными блоками. Для равномерного распределения трассировки многослойной печатной платы (МПП) сокеты модулей оперативной памяти должны располагаться с другой стороны – по внешней границе зоны микропроцессоров.

Размещение преобразователей напряжения, которые обеспечивают различные номиналы электропитания, связано с определенными проблемами. Если установить преобразователи с одной из сторон корпуса микросхемы, не ассоциированной с каналами оперативной памяти или высокоскоростными каналами, это приведет к нерациональному использованию площади МПП. Расположение преобразователей с двух сторон ограничивает возможность поворота унифицированных блоков на панели, что препятствует эффективной трассировке. Размещение преобразователей в зоне высокоскоростных каналов запрещено, т.к. это нарушит целостность сигналов.

Другим вариантом размещения преобразователей напряжения является их установка за модулями памяти. В этом случае освобождается пространство рядом с микропроцессорами. Его можно использовать для размещения элементов системы синхронизентов управления и настройки работы панели. Негативным эффектом такого решения становится ухудшение качества полигонов питания. Однако, учитывая сравнительно небольшую мощность микропроцессора при достаточной ширине полигонов или их дублировании, этим можно пренебречь.

В одномодульном исполнении четырёхпроцессорного ВК вывод канала RDMA возможен через слот расширения с подключением периферийной карты или прямым подключением контактов микропроцессора к контактам QSFP разъёма, располагаемого в зоне ввода-вывода, предусмотренной спецификацией ATX. В обоих вариантах длины трасс от микропроцессоров, расположенных вдали от указанных мест, до разъёма будут больше 20 см, поэтому необходимо использовать повторители.

3. Многомодульное исполнение

При многомодульном исполнении четырёхпроцессорный ВК реализуется путём объединения через соединительную панель четырёх одинаковых модулей, каждый из которых содержит один микропроцессор, один контроллер периферийных интерфейсов КПИ-2 и набор интерфейсных разъёмов. Примерное расположение компонентов в этом варианте, которое обусловлено приведенными ниже ограничениями, представлено на рис. 2.

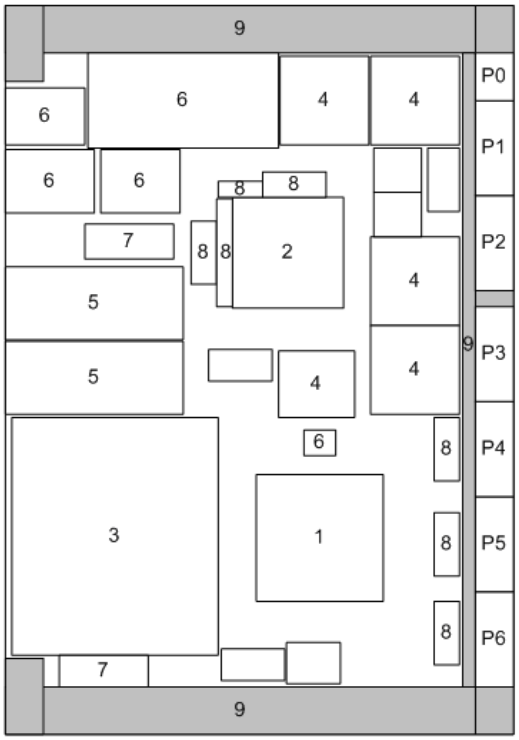

Рисунок 2. Размещение компонентов модуля в конструктиве VPX:
1 – микропроцессор; 2 – КПИ-2; 3 – зона сокетов оперативной памяти; 4 – зона преобразователей напряжения DC-DC; 5 – разъёмы QSFP для соединения модулей через RDMA; 6 – зона периферийных интерфейсов; 7 – микропереключатели для задания режимов; 8 – последовательные конденсаторы высокоскоростных линков; 9 – зоны высвобождения от компонентов

При реализации ВК в многомодульном исполнении принципиальное значение имеют следующие условия:

- назначение сигналов в разъемах для подключения к соединительной панели, обозначенных с учетом спецификаций VPX, предполагает ввод питания +12 В через разъём

P0, размещение интерфейсов PCIe, PCI и SATA в P1, P2 и P3 соответственно, а вывод трех межпроцессорных каналов IPLink – через P3, P4, P5;

- высота вычислительных модулей ограничивается размером 4HP, а их размещение осуществляется непосредственно рядом друг с другом;
- тракт каналов IPLink содержит последовательные конденсаторы, а его длина должна оставаться в рамках допустимой величины для блоков CEI 6G SR x4 и не требовать установки повторителей;
- микропроцессор размещен максимально близко к разъёмам P3, P4, P5 так, чтобы сторона микросхемы с шариковыми выводами, ассоциированными с межпроцессорными каналами, была направлена в сторону разъёмов.

В приведённом на рис. 2 модуле вывод канала RDMA из микропроцессора возможен через соединительную панель на периферийный модуль или через разъём на лицевой панели. Свойства RDMA канала позволяют осуществлять прямое подключение контактов микропроцессора к контактам QSFP разъёма, обеспечивающего применение как медного, так и оптического кабелей. Кабели для данного разъёма имеют встроенный повторитель, поэтому при сохранении длины трассы от микропроцессора до разъёма в пределах 20 см обеспечивается возможность подключения разных, достаточно удалённых друг от друга ВК.

В случае четырёхпроцессорной конфигурации прямое соединение нескольких ВК через RDMA каналы позволяет объединять в кластер только пять узлов по схеме «каждый с каждым». Большей гибкости можно достичь путём преобразования интерфейса RDMA канала к стандартному интерфейсу, например, 10 G или 40 G Ethernet. Для этого необходимо применение ПЛИС. Размещение ПЛИС на модуле потребует значительной площади и дополнительных источников питания. В то же время, размещение преобразователя интерфейса на соседнем периферийном модуле позволит уменьшить сложность вычислительного модуля и количество применяемых в нём дорогих компонентов, а также увели-

чит модульность ВК.

Равномерное заполнение матрицы выводов микропроцессора требует расположить выводы каналов оперативной памяти и каналов ввода-вывода на разных сторонах корпуса. Ввиду ограничений, вносимых каналами IPLink на размещение микропроцессора, положение контроллера периферийных интерфейсов и разъёма RDMA-канала, применение модулей оперативной памяти DDR4 в формате DIMM оказывается невозможным. Максимальная высота модуля 4HP не позволяет использовать сокеты для вертикального размещения модулей памяти, поэтому в данном случае необходимо применение сокетов для углового размещения оперативной памяти формата SO-DIMM или использование распаянной памяти.

4. Оперативная память

Оперативная память в изделиях на основе проектируемого микропроцессора может быть реализована с применением сокетов форматов DIMM и SODIMM, а также в распаянном исполнении. Каждый из вариантов имеет свои ограничения при реализации топологии модулей.

Расположение ячеек ввода-вывода каналов оперативной памяти в кристалле предполагает назначение сигналов разных каналов на разных углах одной стороны матрицы выводов. Однако в случае такого расположения при использовании модулей памяти формата SODIMM трассировка на МПП оказывается затруднительной и приводит к увеличению числа слоёв платы по следующим причинам:

- из-за отсутствия свободного места для прокладки проводников между микропроцессором и сокетом. Исходя из норм проектирования и геометрии прокладки проводников, между микропроцессором и сокетом памяти должно быть порядка 20 мм, а в случае компоновки модуля в VPX конструктиве оценочно доступно только 12 мм;
- из-за невозможности обеспечить сдвиг сокетов памяти для обеспечения параллельной трассировки двух каналов в одном слое. Это сократило бы число необходимых

слоёв при использовании сокетов формата DIMM.

Сокет формата SODIMM предназначен для поверхностного монтажа. В отличие от формата DIMM трассировка на внешнем слое позволяет подключить только ближайший к микропроцессору ряд контактов сокета. В этом случае потребуется разделение проводников каждого байта памяти по двум разным слоям, что недопустимо. Поэтому трассировка, обеспечивающая соединение каналов памяти с сокетом формата SODIMM, будет осуществляться только на внутренних слоях.

Ещё одной особенностью использования оперативной памяти DDR4 в данном микропроцессоре является наличие опции отключения питания ячеек ввода-вывода в режиме сна S3, при котором модули памяти вводятся в состояние сохранения данных. Из микропроцессора, находящегося в режиме сна, это состояние поддерживается группой сигналов CKE, reset_n, alert_n, расположенных в каждом канале в отдельной области неотключаемого питания. При этом питание на модулях и в этих областях сохраняется, а питание остальных ячеек ввода-вывода может быть отключено. Поддержание этой опции требует разделения полигона питания номиналом +1,2 В на два независимых и добавления ещё одного источника питания.

Сохранение целостности высокочастотных сигналов памяти требует проведения каждой группы сигналов над своим полигоном земли-питания. Причём, с учётом структуры модулей памяти, желательно провести байты над опорным слоем земли, а адресно-командные сигналы – над питанием.

5. Матрица выводов корпуса микросхемы

В принципе, в качестве автоматизированного средства назначения сигналов на матрице выводов можно использовать инструмент, описанный в [5]. Однако, будучи направлен на максимальное уменьшение размеров матрицы, он не учитывает факторы, приведенные в предыдущих разделах относительно выбора компоновки модулей и особенностей реализации кристалла. Такой вариант приведен на рис. 3. Представленное решение

обусловлено удобством трассировки и минимизацией количества слоев в МПП процессорных модулей.

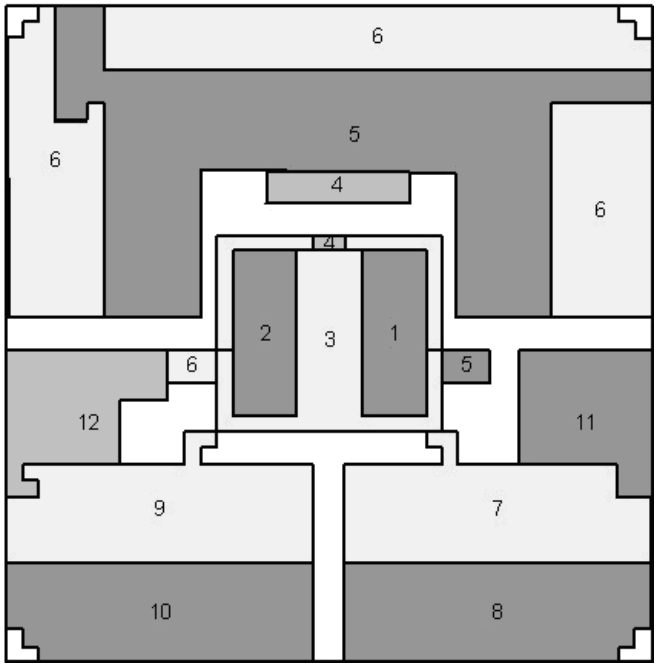

Рисунок 3. Матрица выводов корпуса микропроцессора:
1 и 2 – зоны кластеров процессорных ядер; 3 – зона домена северного моста; 4 – домен синхронизации; 5 и 6 – каналы оперативной памяти; 7, 8, 9, 10 и 11 – высокоскоростные каналы; 12 – зона служебных сигналов

Для корпуса, в качестве экономичного, рассматривался вариант с размером матрицы 38×38, но он не подошел, т.к. увеличивал сложность и стоимость разработки из-за проблем, связанных с формированием методики и средств для отбраковки микропроцессоров, а также с реализацией посадочного места микропроцессора в электронной библиотеке САПР по разработке модулей. Поэтому был принят освоенный при создании микропроцессора Эльбрус-4С размер 40×40, позволяющий использовать полученные опыт и технологии. Также существенно то, что в этом варианте после назначения всех информационных сигналов и выводов земли-питания микропроцессора остаётся достаточно большое количество свободных мест.

Выводы каналов оперативной памяти и каналов ввода-вывода должны быть распо-

ложены на разных сторонах микропроцессора. Это обеспечивает равномерное распределение проводников при топологическом проектировании МПП.

Сигналы одного канала оперативной памяти назначаются на контакты у внешней границы матрицы выводов, а второго – за первым, на внутренней стороне матрицы. Зона каждого канала делится на четыре части, расположенные последовательно одна за другой вдоль границы корпуса: группа из пяти байтов, группа сигналов из области неотключаемого питания, адресно-командная группа, группа из четырёх байтов. Каждый канал должен выводиться из-под матрицы в двух слоях МПП. Это необходимо, чтобы каналы выводились без нарушений норм проектирования и требований к интерфейсу памяти [6], обеспечивающих высокий уровень качества сигналов.

Контакты корпуса, не имеющие назначенного сигнала микропроцессора, не требуют размещения переходных отверстий. Это позволяет на обратной стороне МПП размещать в зоне матрицы микропроцессора LC-фильтры аналогового питания, что освобождает пространство вокруг микросхемы и уменьшает количество полигонов, проходящих через переходные отверстия матрицы.

Контакты питания процессорной логики в матрице выводов размещены под проекцией соответствующих контактов кристалла. Полигоны разных доменов реализуются отдельно, чтобы обеспечить возможность независимой регулировки номиналов питания. Это приводит к наличию у микропроцессора четырёх разных полигонов питания одного номинала, питающих домен северного моста, два домена кластеров процессорных ядер и полигон цифрового и аналогового питания логики периферийных блоков. При такой структуре усложняется проектирование МПП, но уменьшаются абсолютные значения текущих по платам токов. В то же время, снижается негативное взаимное влияние доменов друг на друга, описанное в [7]. Из-за различных исполнений модулей размещение выводов питания сделано с некоторой симметрией, что обеспечивает возможность размещения преобразователей DC-DC с любой стороны микропроцессора без ухудшения качества по-

лигонов питания.

Заключение

Проведенное авторами этой статьи исследование показало возможность развертывания работы по созданию конструкторско-технологической базы использования микропроцессора нового поколения в вычислительных средствах уже в процессе его проектирования. Описанные в статье решения имеют достаточно универсальный характер и могут с большим основанием непосредственно применяться или учитываться в следующих проектах вследстаие того, что модульное исполнение сейчас стало основной или предпочтительной формой реализации вычислительных средств. С учетом этой перспективы в основном качественный на данный момент характер исследования будет дополняться аналитическим материалом.

Литература

1. Бычков И.Н., Воробьев А.С., Рябцев Ю.С. Разработка таблицы выводов серверного процессора // Вопросы радиоэлектроники. – 2015. – Сер. ЭВТ. – Вып. 1. – С. 117-129.

2. Бычков И.Н., Воробьев А.С., Рябцев Ю.С. И Стенд тестирования и разбраковки многоядерных процессоров // Приборы. – 2015. – № 2(176). – С. 16-22.

3. Бычков И.Н., Халиуллин Ю.Х. Масштабируемые по количеству ядер и периферийных интерфейсов промышленные вычислительные системы: 57-я научная конференция МФТИ с международным участием, посвященная 120-летию со дня рождения П.Л. Капицы: Тез. докл. – М., 2014.

4. OIF. Common Electrical I/O (CEI) – Electrical and Jitter Interoperability agreements for 6G+bps and 11G+bps. I/O IA # OIF-CEI-02.0, 28th February 2005.

5. R.J. Lee, M.F. Lai, and H.M. Chen. Fast flipchip pinout designation respin by pinblock design and floorplanning for packageboard codesign // Design Automation Conference, 2007 /ASPDAC’07, Jan. 2007, pp. 804-809.

6. Synopsys. Guidelines for Implementing Signaling Environments for DDRn Interfaces: PCB, Package, Power, and Timing Budgets. Signal Integrity, June 26, 2015 Revision 6.3.

7. Ramon Bertran, Alper Buyuktosunoglu, Pradip Bose, Timothy J. Slegel, Gerard Salem, Sean Carey, Richard F. Rizzolo, Thomas Strach. Voltage Noise in Multi-core Processors: Empirical Characterization and Optimization Opportunities // 47th Annual IEEE/ACM International Symposium on Microarchitecture, 2014.