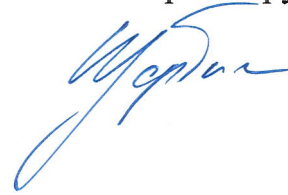


На правах рукописи



Щербина Николай Александрович

**Разработка системы взаимосвязи блоков для
микропроцессоров семейства МЦСТ-R**

05.13.05 — Элементы и устройства вычислительной техники и систем
управления

Автореферат диссертации на соискание ученой степени
кандидата технических наук

Москва - 2014

Работа выполнена в ОАО «ИНЭУМ им. И.С. Брука» и ЗАО «МЦСТ».

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

Официальные оппоненты: **Бобков Сергей Геннадьевич**
доктор технических наук,
старший научный сотрудник,
заместитель директора НИИСИ РАН

Бутузов Александр Валерьевич
кандидат технических наук,
руководитель подразделения по разработке микроархитектуры
ЗАО «Интел А/О»

Ведущая организация: Институт проблем проектирования в микроэлектронике
Российской академии наук (ИППМ РАН)

Защита состоится «12» ноября 2014 г. в 15 ч. 00 мин. на заседании диссертационного совета Д 409.009.01 при ОАО «Институт электронных управляющих машин имени И.С. Брука» по адресу: 119334, г. Москва, ул. Вавилова, д. 24.

С диссертацией можно ознакомиться в библиотеке ОАО «Институт электронных управляющих машин имени И.С. Брука» и на официальном сайте www.ineum.ru.

Автореферат разослан « » _____ 2014 г.

Ученый секретарь
диссертационного совета,
кандидат технических наук, профессор


Красовский В.Е.

Общая характеристика работы

Актуальность работы

Современные микропроцессоры являются системами на кристаллах (СнК), состоящими из вычислительных ядер, иерархии кэш-памятей, контроллеров памяти, контроллеров периферийных интерфейсов и других компонентов. Многие компании-производители в рамках одного поколения микропроцессоров выпускают ряд СнК, различающихся набором перечисленных компонентов, в то время как микроархитектура самих блоков не меняется. Данная стратегия, в силу ее универсального подхода, используется как производителями микропроцессоров для встроенных применений, к которым можно отнести такие компании, как Freescale, Qualcomm, TI, Marvell, Broadcom и др., так и компаниями, специализирующимися на выпуске серверных микропроцессоров, среди которых стоит отметить Intel и IBM.

Даже при наличии разработанных блоков важную задачу составляет их объединение в составе одной СнК. На протяжении последних десяти лет методы объединения блоков СнК, то есть организации их взаимосвязи, активно развиваются. Сравнительно недавно группа ученых во главе с профессором Стэнфордского университета W.J. Dally, предложила использовать принципы организации взаимосвязи блоков, подобные применяемым в компьютерных сетях. Эти принципы нашли отражение во многих реализациях микропроцессоров и в данный момент являются доминирующими в проектировании системы взаимосвязи блоков СнК.

Актуальность диссертационной работы обусловлена необходимостью освоения методов взаимосвязи блоков для отечественных СнК. Эти методы дадут возможность создавать серии СнК для различных сегментов рынка вычислительной техники (ВТ), отличающиеся составом вычислительных блоков, интегрированных на одной и той же микроархитектурной платформе.

Подобные методы включают в себя общие принципы разработки системы взаимосвязи СнК. Конкретные аппаратные решения могут быть получены на основе этих принципов для определенного семейства микропроцессоров. В данной работе речь идет о серии микропроцессоров МЦСТ-R. Это семейство на данный момент включает в себя микропроцессоры: МЦСТ-R150, МЦСТ-R500 и многоядерные СнК: МЦСТ-R500S, МЦСТ-R1000 и МЦСТ-R1000M¹. Все микросхемы реализуют архитектуру SPARC.

В данной работе предложенные автором принципы объединения компонентов СнК применяются при разработке системы взаимосвязи блоков четырехъядерной СнК

¹ В данный момент осуществляется тестирование опытных образцов СнК МЦСТ-R1000M.

МЦСТ-R1000. Заложенные в этой разработке свойства позволяют использовать ее как в указанной СнК, так и в будущих микропроцессорах серии МЦСТ-R.

Автор предлагает разделить функции, реализация которых необходима в СнК для обеспечения эффективного взаимодействия блоков, на три уровня: сетевой, транспортный и сеансовый. Такой шаг упрощает анализ и проектирование системы взаимосвязи, обеспечивает свойство ее модульности за счет прямого соответствия определенных аппаратных решений выделенным функциональным уровням. Приведенный подход разбиения микроархитектуры системы взаимосвязи, являясь научной новизной представленной работы, дополняется некоторыми методами обеспечения взаимодействия блоков в СнК, предложенными группой W.J. Dally. На основе этого формируется теоретическая база исследования. Практические результаты работы, состоящие в разработке системы взаимосвязи СнК МЦСТ-R1000, серийно выпускаемой в настоящее время, подтверждают эффективность решений, основанных на представленных автором теоретических выводах.

Цель диссертационного исследования состоит в

- разработке системы аппаратных средств, обеспечивающей эффективную взаимосвязь блоков в СнК МЦСТ-R1000;
- адаптации микроархитектуры микропроцессоров семейства МЦСТ-R к возможности унифицированного введения новых типов блоков;
- определении и реализации механизмов взаимодействия блоков, приемлемых для перспективных микропроцессоров с многоузловой топологией системы взаимосвязи.

Для достижения поставленной цели решаются следующие задачи:

- формирование иерархии функций, необходимых для взаимодействия блоков СнК;
- разработка аппаратных средств СнК, соответствующих каждому выделенному функциональному уровню иерархии;
- анализ характеристик оборудования каждого уровня и системы взаимосвязи в целом;
- выработка рекомендаций по развитию представленных принципов построения системы взаимосвязи и разработанных микроархитектурных элементов в будущих поколениях микропроцессоров серии МЦСТ-R.

Методы исследования, выбранные для решения поставленных задач, базируются на компьютерном моделировании, а также на теории и технологиях проектирования микропроцессоров.

Научная новизна исследования

К составляющим научную новизну диссертационной работы решениям следует отнести:

- введение *уровневой иерархии* функций системы взаимосвязи блоков в СнК;
- рассмотрение организации *транзакций* обмена данными между блоками и обеспечения необходимого уровня параллелизма в их выполнении, как функций *транспортного уровня* системы взаимосвязи;
- выделение функций *синхронизации* транзакций обмена данными в *сеансовый уровень* системы взаимосвязи блоков в СнК.

Практическая ценность исследования состоит в реализации микроархитектурной платформы взаимосвязи блоков микропроцессора, на базе которой возможно создание серии СнК, отличающихся составом компонентов и соответствующих различным сегментам рынка ВТ.

Указанные принципы внедрены в микроархитектуру СнК МЦСТ-R1000, серийно выпускаемой в настоящее время.

Апробация работы

Результаты работы докладывались на различных научно-технических конференциях:

- 52-ая и 54-ая научные конференции МФТИ (2009, 2011 гг.);
- Научно-техническая конференция «Перспективные направления развития средств вычислительной техники» (Москва, ОАО «Концерн «ВЕГА», ОАО «НИЦЭВТ», 2011 г.);
- XXXVII Международная молодежная научная конференция «Гагаринские чтения» (Москва, МАТИ, 2011 г.);
- Всероссийская научно-техническая конференция «Прикладные научно-технические проблемы современной теории управления системами и процессами» (Москва, ОАО «Концерн «Вега», 2012 г.).

Публикации

По теме диссертационной работы опубликованы 8 печатных работ, в том числе 3 в издании, рекомендованном ВАК РФ.

Структура и объем диссертации

Диссертация состоит из введения, пяти глав, заключения и списка литературы, насчитывающего 64 наименования. Работа изложена на 114 страницах, содержит 23 рисунка и 4 таблицы.

Основное содержание работы

Во введении дана общая характеристика работы, раскрыта ее актуальность, сформулированы основные задачи и цели исследования, определены научная новизна и практическая значимость полученных результатов, а также основные результаты, выносимые на защиту.

В первой главе определяется назначение системы взаимосвязи блоков в СнК МЦСТ-R1000 и требования к ней. Раскрываются принципы, используемые при ее проектировании. Анализируются методы, применяемые для решения подобных задач в других работах, выделяются те из них, которые целесообразно применить при разработке СнК МЦСТ-R1000.

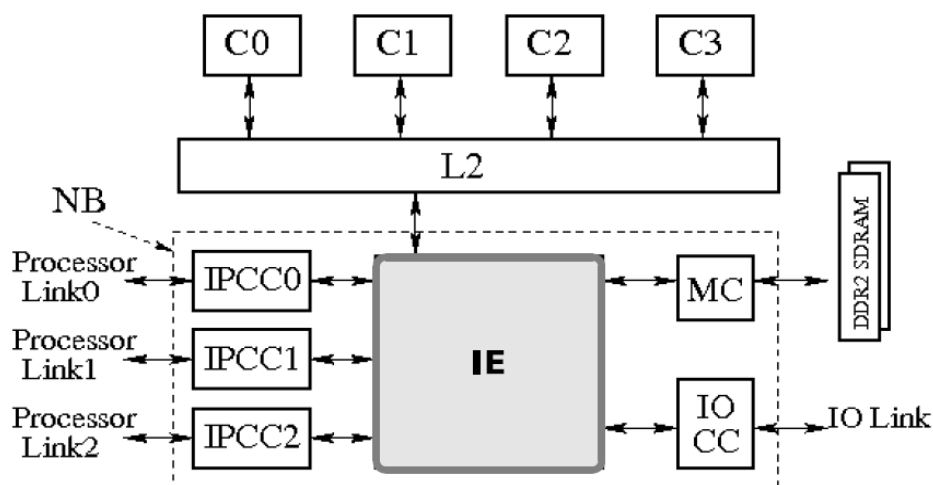


Рисунок 1: Структурная схема СнК МЦСТ-R1000

Структурная схема СнК МЦСТ-R1000, представленная на рисунке 1, включает в себя четыре процессорных ядра (C0...3); общую кэш-память второго уровня (L2); контроллер оперативной DDR2 SDRAM памяти (MC); три контроллера межпроцессорных каналов (IPCC0...2); контроллер канала ввода-вывода (IOCC) для связи с «южным мостом». Оборудование системы взаимосвязи (Interconnection Equipment, IE) обеспечивает взаимодействия L2, IOCC, MC между собой и их доступ к межпроцессорным каналам IPCC0..2 для построения многопроцессорного ссNUMA-комплекса (поддерживаются любые связные топологии с числом СнК до четырех). Таким образом, в многопроцессорном комплексе IE и IPCC0..2 четырех СнК обеспечивают взаимодействие 12 блоков. Структурно IE является составной частью подсистемы памяти СнК, которая также включает L2 и MC.

Производительность IE должна быть достаточной для загрузки пропускной способности внешних интерфейсов СнК. В первую очередь это касается канала оперативной

памяти. К такому требованию производительности добавляется требование минимизации физических ресурсов микросхемы (площадь, рассеиваемая мощность), потребляемых ИЕ. Необходимость использования микроархитектурной платформы ИЕ в модификациях СнК МЦСТ-R1000 требует универсальности аппаратных решений, принимаемых при проектировании ИЕ. Микроархитектура ИЕ должна быть применима для объединения блоков СнК вне зависимости от их функционального назначения. Прогнозируемое увеличение количества функциональных компонентов в составе микропроцессоров будущих поколений семейства МЦСТ-R требует возможности развития системы взаимосвязи, разработанной для СнК МЦСТ-R1000. Это развитие предполагает как увеличение числа объединяемых блоков, так и топологическую эволюцию ИЕ.

Для обеспечения универсальности аппаратных решений в ИЕ анализ функций взаимосвязи должен производиться на некоторой модели блоков СнК. В этой модели не должно конкретизироваться назначение объединяемых блоков. Общее количество интегрируемых на кристалле компонентов хотя и нуждается в ограничении, однако, не должно быть фиксированным. Кроме того, микроархитектурные элементы ИЕ необходимо разделить на общие ресурсы и те, которые непосредственно обслуживают только одного абонента. Такое разделение может быть выполнено путем разбиения всех функций взаимосвязи на уровни. Возможность масштабирования числа объединяемых блоков в СнК и дальнейшего развития топологии ИЕ требует определения границ между функциональными уровнями таким образом, чтобы при топологических изменениях затрагивалась аппаратура, реализующая функции только одного уровня.

Необходимый уровень производительности ИЕ должен обеспечиваться за счет параллельного выполнения транзакций обмена данными между блоками СнК. Степень параллелизма может быть определена из времени реализации одной транзакции и величины пропускной способности, которую должно гарантировать ИЕ.

В литературе встречаются примеры использования методов объединения блоков микропроцессоров, позволяющих улучшить его физические характеристики (Dally W., Towles B.: «Route packets, not wires: On-chip interconnection networks»; Micheli G., Benini L.: «Networks on chips»; Kumar S., Jantsch A.: «A network on chip architecture and design methodology» и др.). Путем внедрения многоузловых регулярных топологий *коммуникационных фабрик* (Communaction Fabric, CF) — оборудования, отвечающего за доставку сообщений между блоками, ограничиваются длины межблочных связей в СнК. Это приводит к улучшению временных характеристик микросхемы. Однако, многоузловые топологии в микропроцессорах целесообразно применять при объединении на одном кристалле большого числа блоков. В СнК МЦСТ-R1000 рациональнее реализовать более

простую топологию СФ. Применение пакетной передачи данных и использование общих разделяемых физических каналов позволяют повысить *степень утилизации* межблочных связей микропроцессора, сокращая ресурсы микросхемы, необходимые на их реализацию. Предположим, что для передачи объема данных S между определенной парой абонентов им необходимо обменяться P типами сообщений через СФ. В случае использования отдельной шины шириной w_p , где $p = \overline{0, P-1}$, для отправки сообщения p -го типа за l_p тактов утилизация межблочных соединений U определяется как

$$U_{gw} = \frac{\sum_{p=0}^{P-1} w_p l_p}{l \sum_{p=0}^{P-1} w_p} = \frac{\sum_{p=0}^{P-1} w_p l_p}{\frac{SF}{B} \sum_{p=0}^{P-1} w_p} = \frac{B}{SF} \frac{\sum_{p=0}^{P-1} w_p l_p}{\sum_{p=0}^{P-1} w_p} \quad (1),$$

где l — общее время (в тактах синхросигнала) транспортировки объема данных S , определяющееся пропускной способностью абонента-приемника данных B и тактовой частотой F . Переход от использования выделенных шин для каждого типа сообщения к применению общего канала передачи пакетов преобразует (1) в

$$U_{sh} = \frac{B}{SF} \sum_{p=0}^{P-1} l_p \quad (2).$$

Здесь l_p — время передачи сообщения p -го типа по общему каналу. В предельном случае можно получить оценку повышения степени утилизации межблочных связей при переходе к указанному методу

$$\frac{U_{sh}}{U_{gw}} \approx P \quad (3).$$

Выражение (2) верно, если суммарное время передачи всех сообщений по одному каналу, не превышает l , то есть

$$\sum_{p=0}^{P-1} l_p \leq l = \frac{SF}{B} \quad (4)$$

В результате анализа требований к проектируемой СнК МЦСТ-R1000 были сделаны выводы о том, что условие (4) может быть удовлетворено. Это позволяет повысить утилизацию межблочных связей разрабатываемой СнК путем перехода от использования выделенных шин к общим каналам передачи сообщений.

Указана стратегия проектирования, необходимая для удовлетворения предъявленным требованиям. Она заключается в анализе функций взаимосвязи некоторой модели СнК, разбиении необходимых для реализации функций на уровни и независимой разработке аппаратных решений, соответствующих каждому выделенному функциональному уровню.

Принято решение использовать пакетную передачу данных через общие коммуникационные каналы для повышения степени утилизации межблочных связей СнК.

Во второй главе на основе обобщенного представления блоков СнК анализируются функции, реализация которых необходима в системе взаимосвязи СнК МЦСТ-R1000. Результатом анализа является формирование функциональной иерархии ИЕ.

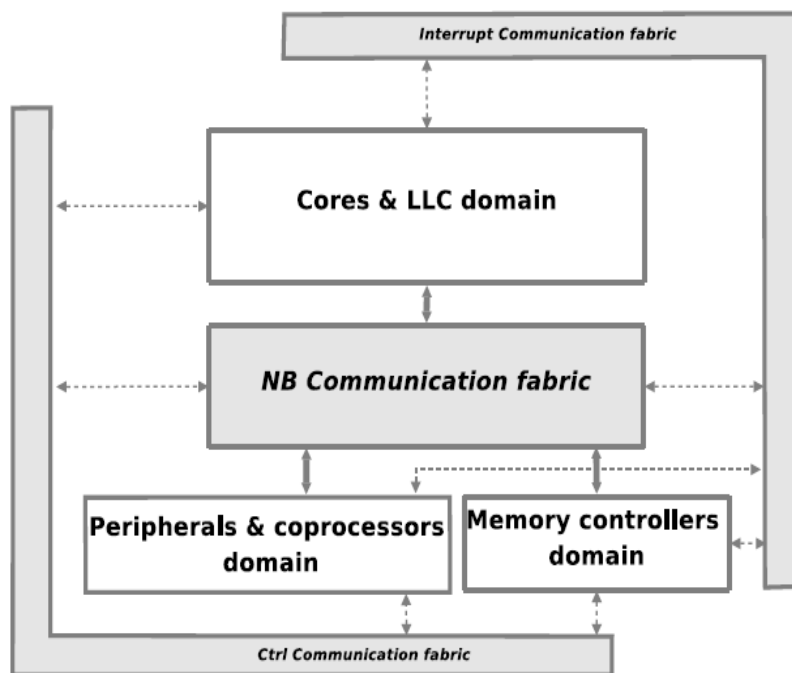


Рисунок 2: Обобщенное представление конфигурации блоков СнК и связей между ними

На рисунке 2 показано обобщенное представление традиционной конфигурации блоков СнК и связей между ними. В этой модели процессорный домен содержит основные вычислительные ядра (Cores) и общую кэш-память последнего уровня (Last Level Cache, LLC). Множество контроллеров оперативной памяти (Memory controllers) выделены в отдельный домен. Сопроцессоры (Coprocessors) и периферийные адаптеры (Peripherals) представлены в виде одной структуры, так как предполагается, что они взаимодействуют с другими компонентами СнК по одинаковому сценарию — через общую память, как управляемые генераторы обращений прямого доступа к памяти (Direct Memory Access, DMA). В таком представлении блоков СнК не конкретизируются их функциональные сущности и не фиксируется их количество. Коммуникационная инфраструктура представлена в виде трех коммутационных фабрик. Задача коммуникационной фабрики «северного моста» (NBCF) — связь процессорного домена, домена периферийных контроллеров и сопроцессоров с доменом контроллеров памяти. Управляющая коммуникационная фабрика (CtrlCF) предназначена для доступа к управляющим и конфигурационным регистрам

устройств СнК. Коммуникационная фабрика прерываний (InterruptCF) обеспечивает обмен сообщениями о прерываниях между блоками СнК. Все три коммуникационные фабрики могут иметь пересечения. CtrlCF и InterruptCF в данной работе не рассматриваются. Многопроцессорная конфигурация также может быть рассмотрена в рамках представленной модели с той лишь разницей, что NBCF в ней помимо собственного оборудования СнК будет включать и межпроцессорные каналы передачи сообщений.

Задача NBCF заключается в пересылке сообщений нескольких типов — *классов трафика* (КТ), которыми блоки обмениваются в процессе своего взаимодействия. Для этого NBCF необходимо выполнять набор сетевых функций, таких как маршрутизация и коммутация сообщений. При использовании одних и тех же коммуникационных ресурсов для обработки различных КТ возможны ситуации блокировки системы, когда обработка сообщений одного КТ может замедлить до критического уровня или полностью остановить пересылку сообщений других КТ, вследствие нехватки ресурсов. Для предотвращения подобных ситуаций необходимо выполнение функции обеспечения качества обслуживания (Quality of Service, QoS), гарантирующей определенную скорость обработки сообщений каждого КТ. Три перечисленные функции: маршрутизация, коммутация и обеспечение качества обслуживания — формируют отдельный *сетевой функциональный уровень* системы взаимосвязи. Этот уровень соответствует коммутационной фабрике северного моста (NBCF). Обособление перечисленных функций позволяет изолировать остальные элементы IE от топологических особенностей NBCF, которые влияют только на микроархитектурные решения, соответствующие сетевому функциональному уровню.

Процесс обмена регламентированной единицей данных между блоками СнК состоит из последовательности фаз, называемой *транзакцией*. Фаза заключается в пересылке определенного сообщения через NBCF. Транзакции выполняются на *транспортном уровне* системы взаимосвязи СнК. Каждому блоку СнК (Unit), взаимодействующему с другими через IE, соответствует свой экземпляр оборудования, реализующего функции транспортного уровня (Transport LE) (см. рисунок 3). Эти функции включают контроль хода транзакций и преобразование протокола обмена данными соответствующего блока СнК во внутренний протокол подсистемы памяти. Таким образом, каждый блок может взаимодействовать с окружением по своему протоколу, вне зависимости от протоколов обмена данными других блоков. Переход на транспортном уровне к внутреннему протоколу подсистемы памяти также позволяет ввести единый интерфейс доступа к оборудованию сетевого уровня (Network LE), которое, как уже упоминалось, представляет собой NBCF. Это еще больше изолирует топологические аспекты NBCF и вопросы числа объединяемых блоков от микроархитектуры оборудования, реализующего функции других уровней IE.

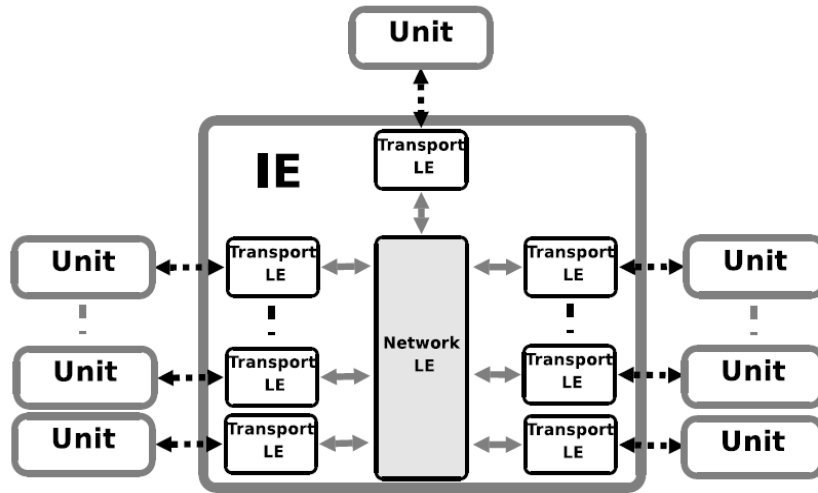


Рисунок 3: Взаимодействие блоков СнК с оборудованием сетевого и транспортного уровней системы взаимосвязи

Для повышения производительности подсистемы памяти и максимально эффективного использования ресурсов NBCF, функции транспортного уровня реализуются с использованием протокола с раздельным выполнением фаз. Каждый экземпляр оборудования транспортного уровня способен параллельно обслуживать некоторое число транзакций, которые в рамках выполнения своей текущей фазы конкурируют друг с другом за право доступа к NBCF. Грант на использование ресурсов NBCF выдается транзакции для осуществления только одной фазы, после чего происходит очередной раунд арбитража.

Для оценки необходимого числа параллельно обрабатываемых транзакций N на транспортном уровне рассмотрим взаимодействие блоков как обращения инициаторов транзакций к общему разделяемому ресурсу. Оценка времени реализации L транзакций одного инициатора при протоколе с раздельным выполнением фаз составляет

$$T_{Ltr} \approx \frac{\sum_{i=0}^{L-1} T_i}{N} \quad (5),$$

где $T_i, i = \overline{0, L-1}$ — количество тактов синхросигнала, требующееся для выполнения i -ой транзакции. Определяя производительность B_{ref} как количество транзакций, выполняемых за один такт,

$$B_{ref} = \frac{L}{T_{Ltr}} = \frac{LN}{\sum_{i=0}^{L-1} T_i} = \frac{N}{\langle T \rangle} \quad (6).$$

В классических терминах пропускной способности (объем данных в единицу времени) выражение (5) принимает вид

$$B = \frac{NF_{clock}S_{tr}}{\langle T \rangle} \quad (7),$$

где B — пропускная способность системы, F_{clock} — тактовая частота синхросигнала, T — количество тактов, необходимое для обработки одной транзакции S_{tr} — объем данных, передаваемый за одну транзакцию. Из (7) следует, что N компенсирует среднюю длину транзакции для достижения требуемой пропускной способности и определяется, как

$$N = \frac{B \langle T \rangle}{F_{clock} S_{tr}} \quad (8).$$

Далее выражение (8) используется для оценок необходимого числа одновременно выполняемых транзакций в ИЕ СнК МЦСТ-R1000.

Для корректного обмена данными между блоками СнК необходимо соблюдение определенного порядка выполнения транзакций в рамках этого обмена. В функции транспортного уровня не входит контроль порядковых отношений между транзакциями. Гарантия соблюдения подобных отношений в системе взаимосвязи — функция синхронизации, выделяемая в *сеансовый функциональный уровень* ИЕ.

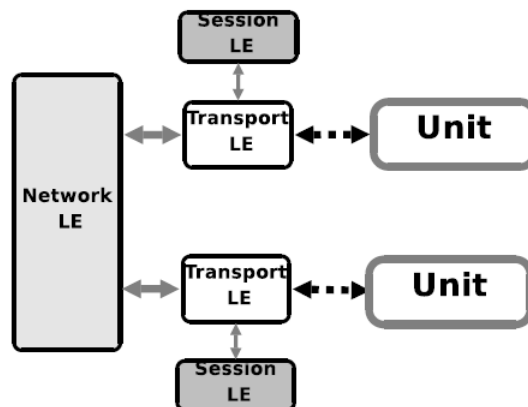


Рисунок 4: Взаимодействие оборудования сеансового уровня с другими компонентами ИЕ

Блоки, объединяемые ИЕ, делятся на инициаторов транзакций и их приемников. Первые входят в состав процессорного домена, вторые — домена контроллеров памяти. Домен периферийных контроллеров и сопроцессоров содержит блоки, являющиеся как инициаторами транзакций (DMA-обращения), так и их приемниками (транзакции, генерируемые процессорным доменом для управления ходом DMA). Внутренний протокол подсистемы памяти требует синхронизации обращений к блокам-приемникам. Протокол обмена данными определенного блока-инициатора может требовать соблюдения порядка выполнения транзакций, генерируемых им. Сопроцессор или периферийный контроллер требует синхронизации DMA-обращений, генерируемых им с запросами от процессора, управляющими ходом его DMA. Для выполнения отношений порядка в трех описанных

случаях вводятся три *подуровня синхронизации*: *протокольный, поточный, и межпоточный*. С соответствующим какому-либо блоку оборудованием транспортного уровня взаимодействует свой собственный экземпляр оборудования сеансового уровня (Session LE). Взаимодействие между экземплярами оборудования сеансового уровня отсутствует (см. рисунок 4). Разделение функций организации транзакций и соблюдения порядка их выполнения упрощают разработку системы взаимосвязи блоков в СнК.

Для представленной модели СнК необходимые функции системы взаимосвязи блоков были разделены на три уровня: сетевой, транспортный и сеансовый.

В третьей главе раскрываются детали реализации микроархитектурных элементов системы взаимосвязи СнК МЦСТ-R1000, соответствующих функциям каждого выделенного уровня.

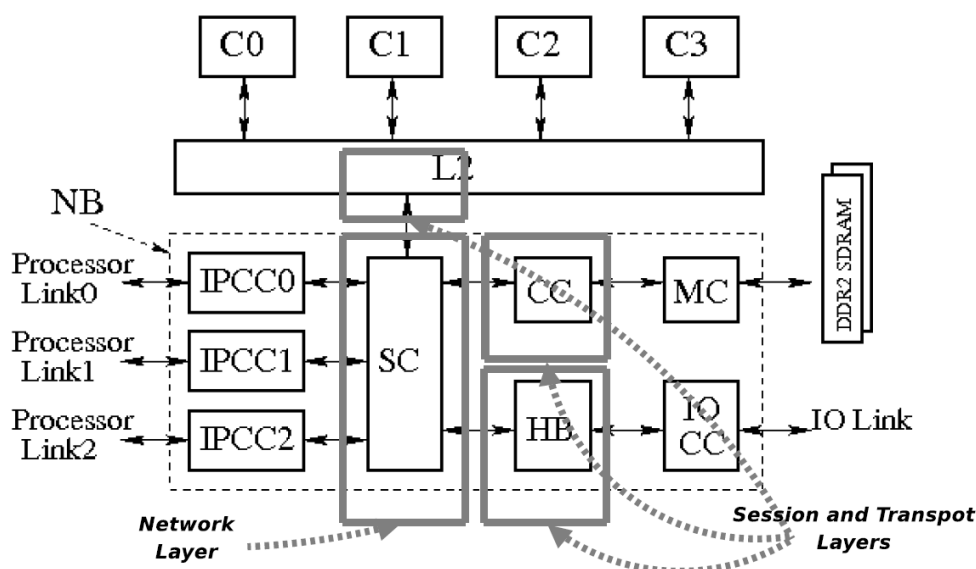


Рисунок 5: Структурная схема СнК МЦСТ-R1000 с системой взаимосвязи

На рисунке 5 изображена структурная схема СнК МЦСТ-R1000 с системой взаимосвязи. В нее входит системный коммутатор (SC), реализующий функции сетевого уровня (Network Layer); контроллер CC, выполняющий функции транспортного и сеансового уровней (Session and Transport Layer) для MC; контроллер HB, осуществляющий функции транспортного и сеансового уровней для IOCC. Оборудование, реализующие функции двух верхних уровней для L2, интегрировано в состав его контроллера. IPCC0...2 напрямую соединяются с SC, так как они не являются ни инициаторами, ни приемниками транзакций (IPCC0..2 — расширение коммуникационной фабрики СнК, позволяющее строить многопроцессорные системы, как показано на рисунке 6).

Системный коммутатор обслуживает четыре класса трафика: *первичные*

запросы, когерентные запросы, ответы, данные. Для передачи сообщений первых трех КТ используется единый формат пакетов (командный пакет). Данные передаются пакетами другого формата (пакет данных), существенное отличие которого — разбиение на несколько посылок (flit). Основным квантом передачи данных в подсистеме памяти является кэш-строка — 64 байта, которая передается за 8 посылок. В соответствии с идеями 2-ой главы интерфейс системного коммутатора с блоками, выполняющими функции транспортного и сеансового уровня (арбитр L2, СС и НВ) и IPCC0...2 унифицирован. В каждом направлении, к коммутатору и от коммутатора, он представляет собой два порта – один для командных пакетов, другой для пакетов данных. Контроль ресурсов принимающей стороны осуществляется при помощи кредитного механизма. Введение унифицированного интерфейса с оборудованием сетевого уровня позволяет изолировать детали реализации последнего, включая топологические, от остальных компонентов СнК как на уровне одной СнК, так и на уровне многопроцессорной системы.

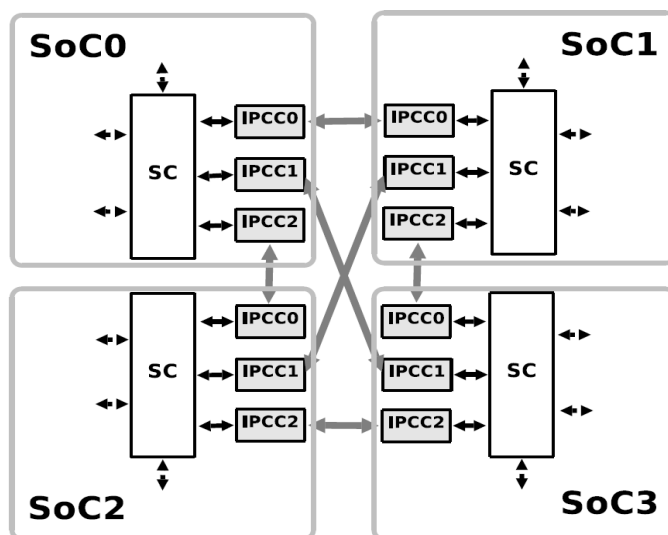


Рисунок 6: Коммуникационная фабрика в случае многопроцессорной системы

В SC реализована статическая конфигурируемая маршрутизация. Идентификатором блока в системе является совокупность идентификатора блока в пределах одной СнК (unit_id) и идентификатора СнК, в которой расположено устройство (node_id). unit_id жестко привязан к каждому блоку: L2, МС, ЮСС. Во время выполнения ПНЗ определяется node_id каждого СнК многопроцессорной системы и формируются карта памяти и соответствующие таблицы маршрутизации. unit_id и node_id блока-приемника транзакции определяются исходя из того, в какую область карты памяти попадает адрес запроса, инициирующего транзакцию. Необходимость передачи пакета блоку локального СнК или в один из трех межпроцессорных каналов определяется исходя из node_id и таблицы маршрутизации.

Описанные решения позволяют создавать многопроцессорные конфигурации с любым количеством СнК до четырех с произвольной связной топологией.

В SC реализована сквозная коммутация пакетов (cut-through packet switching) традиционная для коммуникационных фабрик в составе СнК, где время передачи пакетов между абонентами является критичным параметром. SC разделен на два коммутатора. Первый — коммутатор команд (SCC) обрабатывает командные пакеты. Второй — коммутатор данных (SCD) передает пакеты данных. Это решение позволяет устранить конфликты за коммуникационные ресурсы между классом трафика *данных* и другими КТ и упростить микроархитектуру обоих коммутаторов, исключив из нее необходимость поддержки пакетов различной длины. SCC реализует архитектуру коммутатора с несколькими буферами на входном порте, каждый из которых соответствует выходному порту (multiple input cross-point buffer switch architecture). Каждый буфер статически разделен между пакетами трех КТ. Разделение конфигурируется во время выполнения ПНЗ, что позволяет оптимальным образом распределить буферное пространство между пакетами трех КТ. Равноправное распределение пропускной способности каналов передачи данных к выходным портам регулируется на основе отдельного алгоритма назначения (*separable allocator*). Принятые меры гарантируют равномерное распределение коммуникационных ресурсов SCC между тремя КТ и исключают возможность их захвата пакетами только одного класса трафика. Это необходимо для обеспечения качества обслуживания. SCD реализует архитектуру коммутатора с одним буфером на каждом входе, формирующим единую очередь пакетов (single buffer single queue switch architecture). Это архитектура проще реализованной в SCC, так как SCD обрабатывает пакеты одного КТ, безусловно принимаемые блоками СнК (*данные* освобождают ресурсы блоков, в то время, как сообщения других КТ их захватывают).

На рисунке 7 представлена структурная схема оборудования, выполняющего функции транспортного уровня IE СнК МЦСТ-R1000. Оборудование спроектировано единообразно для всех блоков, объединяемых IE. Различия определяются разницей интерфейсов и протоколов обмена данными этих блоков (если такие имеются). TRN_REG — оборудование, обрабатывающее одну транзакцию, конечный автомат, контролирующий ход ее выполнения. TRN_REG преобразует протокол обмена данными блока во внутренний протокол подсистемы памяти; контролирует последовательность фаз обрабатываемой транзакции; формирует сообщения для отправки смежным модулям (SC или обслуживаемый блок) в рамках той или иной фазы. Конфликты между TRN_REG за право использования внешних интерфейсов разрешаются при помощи арбитров (ARB_LL и ARB_HL). Входящие сообщения декодируются в блоках DEC_LL и DEC_HL. Буфер данных (DATA_BUF) служит для

хранения данных, передаваемых в рамках как транзакций чтения, так и транзакций записи. Число параллельно обрабатываемых транзакций N определяется количеством TRN_REG в экземпляре оборудования. Подобная структура модуля, реализующего функции транспортного уровня, позволяет параметризовать степень параллелизма при обработке транзакций в ИЕ.

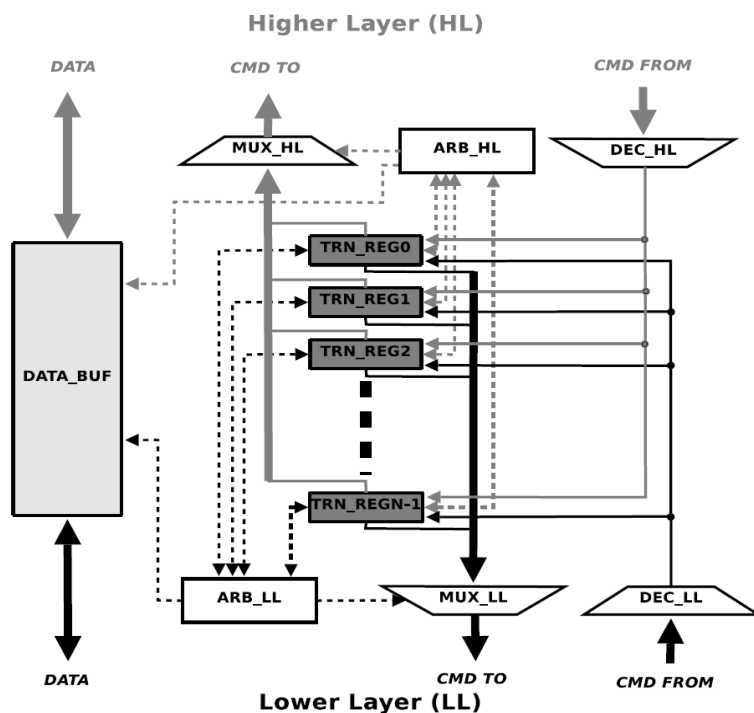


Рисунок 7: Структурная схема оборудования, выполняющего функции транспортного уровня

Определение числа параллельно обрабатываемых транзакций N основывается на необходимости утилизации пропускной способности канала памяти $B = 6,4$ Гбайт/с. Тактовая частота оборудования системы взаимосвязи $F_{clock} = 1$ ГГц, объем данных, передаваемый за одну транзакцию обращения к памяти $S_{tr} = 64$ байта. Принимая во внимание высокую частоту синхронизации, конкуренцию сообщений различных транзакций за общие коммуникационные ресурсы в SC, задержки обработки запросов в MC, оценка среднего времени выполнения одной транзакции обращения к памяти была сделана как (130 ± 20) тактов. В таких условиях N определяется из (8) как

$$N = \frac{B \langle T \rangle}{F_{clock} S_{tr}} = (13 \pm 2) \quad (9).$$

В СнК МЦСТ-R1000 на транспортном уровне одновременно обрабатываются 16 транзакций.

Аппаратура, реализующая функции сеансового уровня для определенного блока СнК, занимает малую площадь по сравнению с соответствующим оборудованием транспортного уровня и взаимодействует только с ним. Чтобы не усложнять структуру СнК МЦСТ-R1000

оборудование, выполняющее функции двух верхних уровней для определенного блока (МС, IOCC, L2), интегрировано в составе одного модуля (СС, НВ, контроллер L2). Функции сеансового уровня реализуются в механизмах трех подуровней синхронизации.

Протокольный (Protocol Synch Level, PSL) — механизм синхронизации приемника, гарантирующий, что в любой момент времени приемником обрабатывается только одна транзакция с определенным идентификатором (требуется в силу специфики внутреннего протокола подсистемы памяти). Приемниками транзакций в СнК МЦСТ-R1000 являются МС и IOCC. IOCC в качестве приемника обрабатывает процессорные запросы к устройствам «южного моста» (Processor Input-Output, PIO). В случае МС идентификатором является адрес запроса, инициировавшего транзакцию. В случае IOCC — совокупность node_id, unit_id блока, инициировавшего транзакцию, и номера TRN_REG, обрабатывающего данную транзакцию в этом блоке.

Поточный (Stream Synch Level, SSL) — механизм синхронизации, гарантирующий порядок выполнения транзакций блока-источника, диктуемый собственным протоколом обмена данными этого блока. Механизм реализуется в оборудовании сеансового уровня блоков-источников. В случае СнК МЦСТ-R1000 только IOCC требует соблюдения некоторого порядка выполнения транзакций (DMA-обращений), генерируемых им. Механизм, контролирующий данный порядок, основывается на гарантии выполнения транзакций, инициированных до определенного события, раньше транзакций, инициированных после этого события. Рассматриваемое событие называется *точкой синхронизации*. В НВ введено несколько точек синхронизации. Наиболее существенная из них — генерация транзакции чтения после транзакции записи. Ее введение позволяет разрешить конфликты записи-чтения обращений к оперативной памяти.

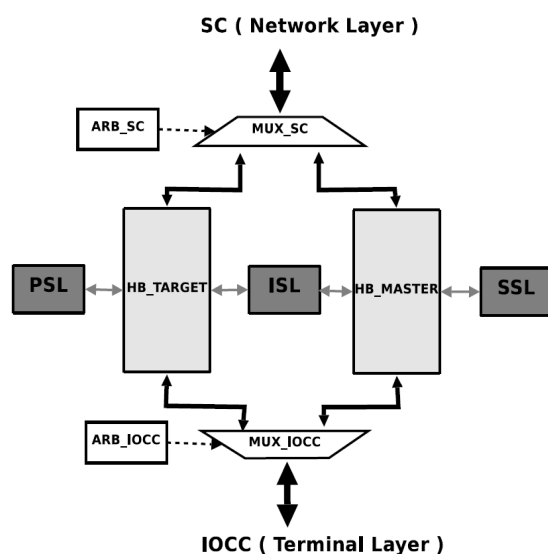


Рисунок 8: Структурная схема НВ

Межпоточный (Interstream Synch Level, ISL) — механизм синхронизации, необходимый для корректной работы блоков домена периферийных контроллеров и сопроцессоров. В СнК МЦСТ-R1000 этот домен представлен ЮСС. Следовательно, механизм ISL реализован только в НВ, где они синхронизируют PIO-запросы и DMA-обращения. Данный механизм частично основывается на SSL, вводя точку синхронизации, соответствующую приходу из ЮСС ответа на PIO-чтение. Этот ответ не выдается процессору до завершения всех DMA-обращений, инициированных до введенной точки синхронизации. Такая синхронизация необходима для корректной работы с устройствами «южного моста».

Структурная схема НВ изображена на рисунке 8. Как отмечалось, НВ является одновременно инициатором и приемником транзакций. Он включает два экземпляра оборудования транспортировки, обрабатывающих PIO-обращения — НВ_TARGET и DMA-запросы — НВ_MASTER. Механизмы PSL, согласно назначению, взаимодействует только с НВ_TARGET, SSL — с НВ_MASTER. Механизмы ISL, обеспечивая синхронизацию двух потоков, взаимодействует и с тем и с другим.

Таким образом в системе взаимосвязи СнК МЦСТ-R1000 реализованы микроархитектурные элементы, соответствующие каждому выделенному функциональному уровню.

В четвертой главе дан количественный и качественный анализ эффективности решений, принятых при проектировании системы взаимосвязи блоков СнК МЦСТ-R1000, представлены сравнительные характеристики разработанной системы и ее зарубежных аналогов.

Эксперименты проведены на базе двух стендов: исполнения тестовых программ на синтезируемой *glt*-модели всего микропроцессора; автономного моделирования поведения системного коммутатора при синтетических нагрузках на него. Предметом измерений являлись время выполнения транзакций обращения к памяти, задержки прохождения сообщений через SC, относительная пропускная способность системного коммутатора (throughput). Последняя определялась как

$$Thr_{in \rightarrow out} = \frac{\Delta N_{in \rightarrow out}}{\Delta t} \cdot \frac{1}{R_{in \rightarrow out}} \quad (10),$$

где $\Delta N_{in \rightarrow out}$ — приращение количества пакетов прибывших от входа *in* к выходу *out* за время измерения Δt . $R_{in \rightarrow out}$ — темп генерации сообщений на входе *in*, местом назначения которых является *out*. Темп генерации является управляемой — известной величиной.

С помощью первого стенда получены значения задержек прохождения пакетов через SC. Измерения производились при выполнении процедуры массового копирования. Данная

задача является наиболее критичной для производительности подсистемы памяти и ИЕ, как ее составной части. Полученные результаты представлены на рисунке 9. При выполнении указанной процедуры средняя задержка — 13 тактов. Около 75% пакетов обрабатывается SC за три такта. Таким образом, подавляющее большинство пакетов проходит коммутатор за минимальное время.

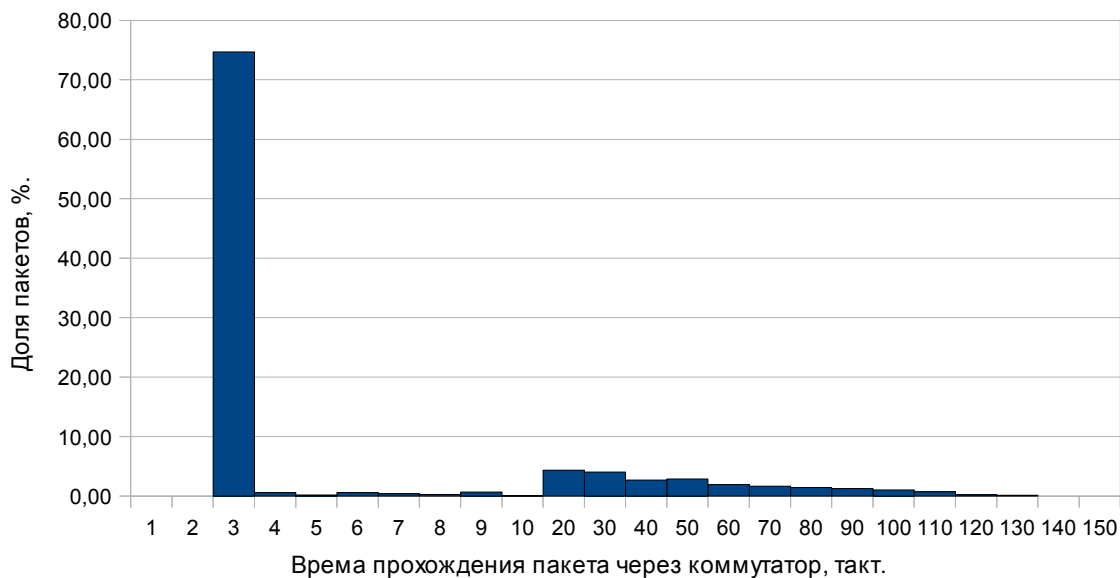


Рисунок 9: Распределение времени прохождения пакетов через SC

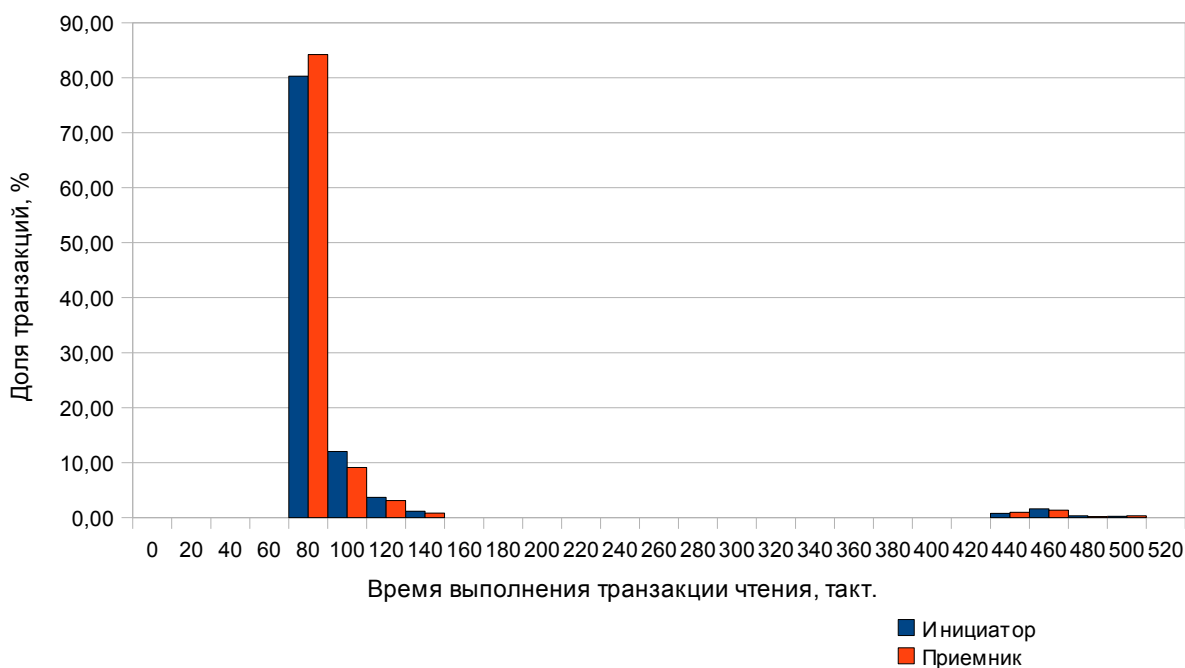


Рисунок 10: Распределение времени выполнения транзакций чтения памяти

Одновременно были произведены замеры времени реализации транзакций обращения

к памяти. Измерения производились как в инициаторе (L2), так и в приемнике (МС) транзакций. Распределение времени обработки транзакций чтения представлено на рисунке 10. Из него следует, что среднее время чтения — 85 тактов. Полученные результаты подтверждают корректность оценки длительности транзакций обращения к памяти, сделанные в главе 3.

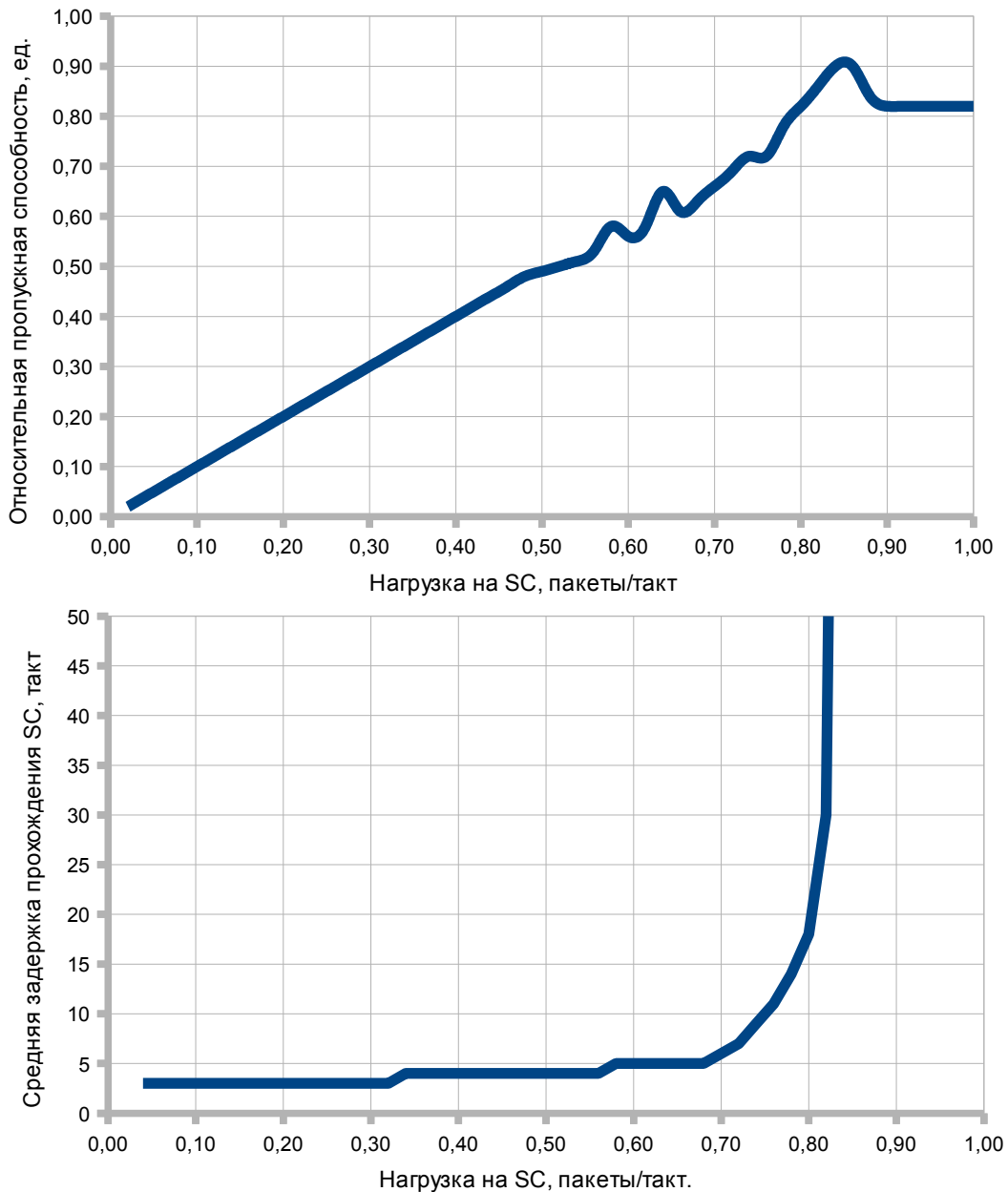


Рисунок 11: Параметры производительности SC

При помощи второго измерительного стенда были получены параметры производительности одного канала SC при фоновой нагрузке на остальные каналы (см. рисунок 11). Зависимость пропускной способности одного канала SC от нагрузки на него — линейная до значения аргумента, равного $(0,83 \pm 0,02)$, что определяют предельную пропускную способность коммутатора. Факт состояния насыщения SC также отображается в

графике зависимости задержек прохождения пакетов от нагрузки, который имеет вертикальную асимптоту $X=(0,83\pm 0,02)$. Снижение предельной пропускной способности до 0,83 объясняется фоновой нагрузкой на порты SC, не превосходящей 0,1, и неоптимальным в некоторых случаях решением задачи назначения в коммутаторе, что характерно для аппаратно реализуемых алгоритмов. Так как нагрузка на SC ограничивается пропускной способностью интерфейса с памятью, можно показать что она не превосходит 0,4. Таким образом, SC всегда находится в состоянии ниже точки насыщения.

В таблице 1 сравниваются характеристики подсистемы памяти СнК МЦСТ-R1000 и аналогичных систем: ARM CCI-400 и Freescale CoreNet. Первая является IP, предоставляемым ARM вместе с лицензируемыми ядрами Cotex-A15, Cortex-A7 и т.п.. CoreNet — внутренняя платформа Freescale, на базе которой созданы такие многоядерные СнК, как QorIQ P4040, P4080, T4160, T4240. Кроме МЦСТ-R1000 в таблице приведены характеристики расширенного варианта этой СнК — МЦСТ-R1000E, проработанного на данный момент в ЗАО «МЦСТ». В МЦСТ-R1000E два четырехъядерных кластера и два канала памяти.

	ARM CCI-400	Freescale CoreNet	МЦСТ R1000	МЦСТ R1000E
Количество объединяемых ядер, число 4-х ядерных кластеров.	2	2†	1	2
Пропускная способность на один 4-х ядерный кластер, Гбайт/с	14,4	25	6,8	27,2
Характеристики каналов памяти	2x32 бит DDR3-1600*	2x64 бит DDR3-2133†	64бит DDR2-800	2x64бит DDR3-1600
Ширина шин данных, байт	16	32	8	16
Тактовая частота, ГГц	0,9*	0,9†	1	2
Технология производства	28нм*	28нм†	90нм	28нм
Поддержка многопроцессорных конфигураций	Нет	Нет	Да	Да

Таблица 1: Сравнительные характеристики подсистем памяти ARM CCI-400, Freescale CoreNet, СнК МЦСТ-R1000 и СнК МЦСТ-R1000E.

* — на примере СнК Allwinner UltraOcta A80.

† — на примере СнК Freescale QorIQ T4169.

Из таблицы 1 видно, что подсистема памяти МЦСТ-R1000E соответствует представленным зарубежным аналогам. За счет более высокой тактовой частоты требуемая пропускная способность в ней достигается при вдвое меньшей ширине шин данных. Необходимо отметить, что подсистема памяти МЦСТ-R1000 (R1000E) поддерживает многопроцессорные конфигурации в то время, как представленные зарубежные аналоги лишены этого свойства.

Анализ характеристик разработанной системы взаимосвязи показал, что она

удовлетворяет изначально поставленным требованиям производительности. Время прохождения пакетов через коммутатор в большинстве случаев близко к минимальному. Это подтверждает выполнение сетевым уровнем функций обеспечения качества обслуживания трафика. Сравнение характеристик разработанной системы с ее зарубежными аналогами показывает, что она соответствует уровню разработок передовых западных компаний.

В пятой главе рассматриваются варианты развития СнК МЦСТ-R1000 и перспективы внедрения коммуникационных фабрик с многоузловыми топологиями в будущие поколения многоядерных микропроцессоров.

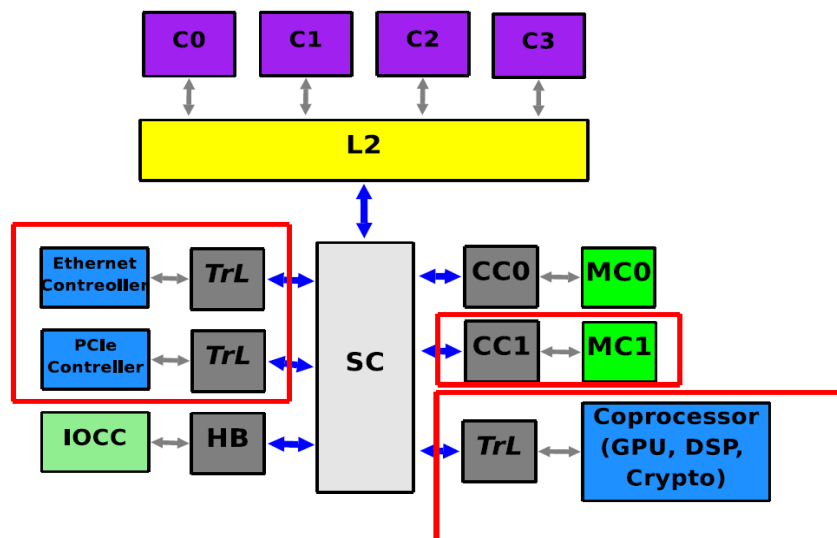


Рисунок 12: Вариант модификации СнК МЦСТ-R1000 на базе разработанной системы взаимосвязи

Разработанное оборудование может быть использовано в модификациях СнК МЦСТ-R1000, не выходящих за рамки модели СнК из главы 2. На рисунке 12 представлен вариант микропроцессора на основе СнК МЦСТ-R1000 с интегрированным сопроцессором и набором контроллеров современных периферийных интерфейсов. Сопроцессором может быть графическое ядро (GPU), сигнальный процессор (DSP), крипто-ускоритель (Crypto) и т.п., взаимодействующие с процессорными ядрами микросхемы через общую память. Сопроцессор является DMA-генератором, что позволяет его логически объединить с остальными модулями СнК, так же как HB.

В состав СнК включены два периферийных контроллера (PCIe и Ethernet), интегрирование которых в состав микропроцессора является общепринятой практикой в СнК для встроенных применений. Введение второго канала памяти (контроллера MC1) в состав СнК подразумевает создание дубликата соответствующего оборудования транспортного и сеансового уровней (CC1), которое идентично оригинальному контроллеру когерентности. Создание описанной СнК возможно на базе уже разработанной системы взаимосвязи блоков.

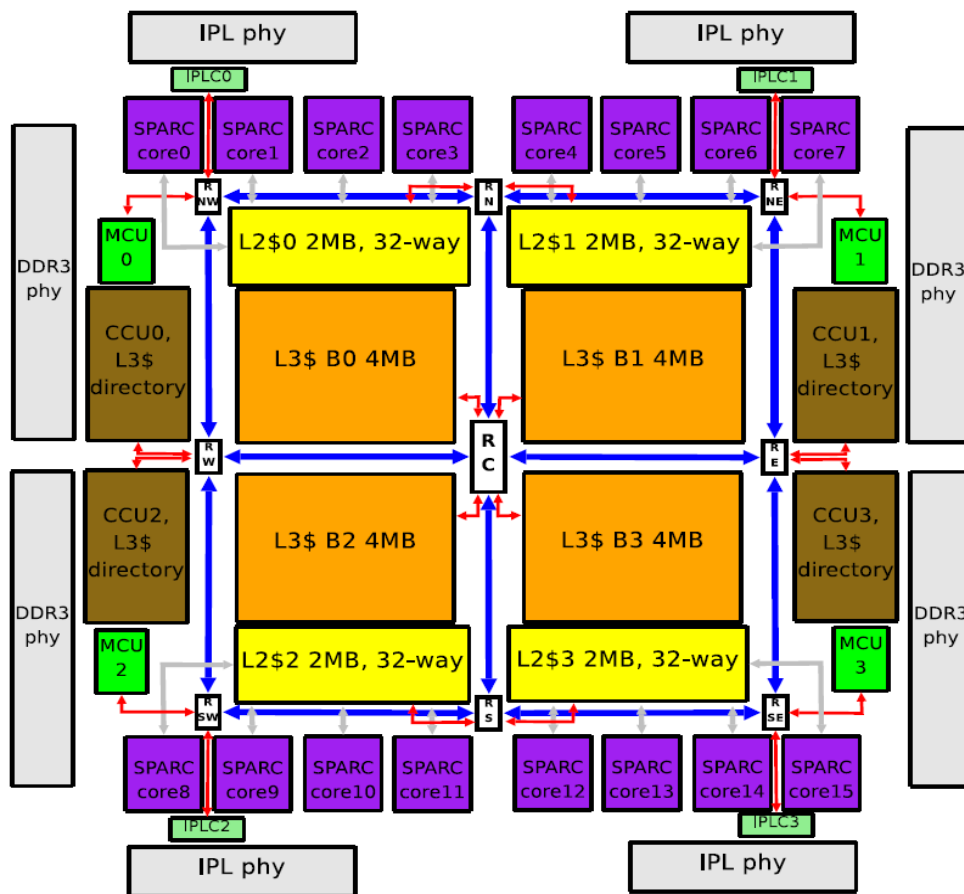


Рисунок 13: Вариант модификации СнК МЦСТ-R1000, использующий многоузловую топологию коммуникационной фабрики

Развитие СнК МЦСТ-R1000, предполагающее наращивание количества процессорных ядер и увеличение масштабов многопроцессорных конфигураций столкнется с необходимостью обеспечения взаимосвязи до 20 блоков в СнК и перехода к многоузловым топологиям коммутационной фабрики СнК. В отличие от CMP (Chip Multi Processor), которые являются популярным примером рассмотрения многоузловых топологий *сетей на кристаллах* (Network on Chips, NoC), микропроцессоры общего назначения обладают нерегулярной компоновкой блоков и характеризуются несимметричным шаблоном нагрузок со стороны модулей на систему взаимосвязи. Таким образом, топологии коммуникационных фабрик СнК с нерегулярной структурой блоков становятся ключевой темой исследования, открывающего пути применения NoC в серверных микропроцессорах.

На рисунке 13 представлен вариант СнК, включающей 16 процессорных ядер (core0...15), разделенных на четыре домена, с общей кэш-памятью второго уровня (L2\$0...3) в каждом. Разделяемая всеми ядрами кэш-память третьего уровня организована в виде четырех независимых банков (L3\$ B0...3). Устройство поддержки когерентного состояния кэш-памяти (CCU0...3) логически распределено в структуре микропроцессора. В состав СнК также входят четыре контроллера DDR3 SDRAM памяти (MCU0...3) и четыре контроллера

межпроцессорных каналов (IPLC0...3), при помощи которых возможно построение многопроцессорного комплекса с шестнадцатью СнК. В представленной СнК топология коммуникационной фабрики — 3x3 решетка (mesh). В ее узлах располагаются 9 коммутаторов трех различных видов. Четыре коммутатора 2x2: RNE, RSE, RSW и RNW. Четыре — 2x3: RN, RE, RS и RW. Один — 4x4, RC. Разбиение L3 на четыре банка дает возможность эффективно использовать выбранную топологию, размещая ее ребра между блоками СнК. Распределение оборудования ССУ и симметричное расположение остальных абонентов сети обеспечивают равномерную нагрузку на коммуникационную фабрику.

Система взаимосвязи представленной СнК строится на основе IE СнК МЦСТ-R1000. В последней сетевые функции изолированы на одноименном уровне. Это решение позволяет оградить влияние топологии коммуникационной фабрики СнК на оборудование транспортного и сеансового уровней IE. Благодаря этому в приведенном примере возможно использование микроархитектурных элементов двух верхних уровней IE СнК МЦСТ-R1000, в то время как оборудование сетевого уровня здесь претерпевает кардинальные изменения.

Приведены варианты модификации СнК МЦСТ-R1000, построенные на базе разработанной системы взаимосвязи блоков. Показан пример использования разработанной системы для реализации NoC в будущих поколениях семейства микропроцессоров МЦСТ-R.

Основные результаты работы

Представленная работа направлена на исследование принципов построения системы аппаратных средств, обеспечивающих взаимодействие блоков СнК. Целью работы является разработка системы взаимосвязи, позволяющей объединять блоки СнК вне зависимости от их конкретного назначения. Результаты работы дают возможность на базе микроархитектурной платформы, разработанной в рамках одного поколения микропроцессоров, проектировать ряд СнК, соответствующих различным сегментам рынка ВТ.

Апробация результатов исследования произведена в процессе проектирования СнК МЦСТ-R1000. Для этого микропроцессора разработана микроархитектура системы взаимосвязи блоков, элементы которой могут быть использованы как для модификации СнК МЦСТ-R1000, так и в будущих поколениях микропроцессоров серии МЦСТ-R.

В процессе выполнения работы были получены следующие основные результаты, выносимые на защиту.

- Спецификация иерархической структуры функций системы взаимосвязи блоков в СнК.

Выделены сетевой, транспортный и сеансовый уровни. Указаны функции, соответствующие этим уровням.

- Разработка микроархитектурных элементов применительно к каждому из выделенных функциональных уровней.

Получены аппаратные решения, соответствующие каждому функциональному уровню системы взаимосвязи. Спроектированные механизмы реализованы в составе СнК МЦСТ-R1000.

- Определение принципов построения микроархитектуры системы взаимосвязи СнК, на основе которой возможно объединение блоков вне зависимости от их конкретного функционального назначения.

На базе рассмотренной модели СнК определены принципы формирования системы взаимосвязи ее блоков, применимые для СнК, соответствующих указанной модели. Данные принципы реализованы в СнК МЦСТ-R1000. Приведены примеры модификации этой микросхемы, заключающиеся в интеграции дополнительных вычислительных модулей на базе разработанной микроархитектуры системы взаимосвязи.

- Исследование вариантов применения разработанной системы взаимосвязи блоков в будущих поколениях микропроцессоров, использующих многоузловые топологии коммутационных фабрик.

Произведен анализ проблем, возникающих при использовании многоузловых топологий коммуникационных фабрик в перспективных микропроцессорах семейств МЦСТ-R. Рассмотрен вариант применения подобных топологий в рамках развития СнК МЦСТ-R1000, предусматривающего масштабирование количества процессорных ядер.

Список работ, опубликованных по теме диссертации

- Щербина Н.А. Разработка системного коммутатора для микропроцессора МЦСТ-4R // Труды 52-й научно конференции МФТИ, Часть 1, Том 1, М.:МФТИ, 2009, С. 77-79.
- Щербина Н.А. Системный коммутатор для микропроцессора «МЦСТ-4R» // Вопросы радиоэлектроники, серия ЭВТ, Выпуск 3, 2010, С. 93-102.
- Щербина Н.А. Факторы обеспечения качества обслуживания в сетях на кристалле // Научно-техническая конференция «Перспективные направления развития средств вычислительной техники». Сборник тезисов и докладов. М.: ОАО «Концерн «Вега», 2011, С. 45-46.
- Щербина Н.А. Проблемы реализации сетевых подходов при создании СнК и NUMA-

- систем на их основе (на примере микропроцессора «МЦСТ-4R») // XXXVII Гагаринские чтения. Научные труды Международной молодежной научной конференции в 8 т. Т. 4, М.: «МАТИ»-РГГУ, 2011, С. 125-126.
- Волин В.С., Щербина Н.А. Анализ потребляемой мощности СнК «МЦСТ-4R» // Труды 54-й научной конференции МФТИ, Радиотехника и кибернетика, Том 1, М.:МФТИ, 2011, С. 31-32.
 - Щербина Н.А., Черепанов С.А. Применение концепций сетей межсоединений при проектировании СБИС // Всероссийская молодежная научно-техническая конференция «Прикладные научно-технические проблемы современной теории управления системами и процессами». Сборник тезисов и докладов. М.: ФНПЦ ОАО «Концерн «Вега», 2012, С. 88-89.
 - Волин В.С., Черепанов С.А., Щербина Н.А. Организация поддержки когерентности в системе на кристалле «МЦСТ-R1000» // Вопросы радиоэлектроники, серия ЭВТ, Выпуск 3, 2012, С. 14-27.
 - Волин В.С., Щербина Н.А. Анализ потребляемой мощности для опытных образцов СнК «МЦСТ-R1000» // Вопросы радиоэлектроники, серия ЭВТ, Выпуск 3, 2013, С. 135-143.