

Разработка контроллера оперативной памяти для применения в современных микропроцессорах семейства «Эльбрус»

И. А. Петров¹

¹ АО «МЦСТ», Москва, Россия

Реализация нового поколения микропроцессоров семейства «Эльбрус», предусматривающего внедрение сети-на-кристалле, требует разработки эффективных средств доступа узлов сети в каналы оперативной памяти (ОП) DDR. В статье описывается решение этой проблемы применительно к взаимодействию с ОП DDR4 16-ядерного микропроцессора «Эльбрус-16СВ», который предъявляет повышенные требования к доступному объему и пиковой пропускной способности каналов памяти. Одной из основных целей при проектировании микропроцессора «Эльбрус-16СВ» также является улучшение показателей энергоэффективности и надежности. Важной составляющей при выполнении поставленных задач стала доработка контроллера памяти, успешно используемого в выпускаемых микропроцессорах АО «МЦСТ», до поддержки стандарта DDR4 3DS, взятого за основу для использования в ряде новых разработок и обеспечивающего увеличение доступного объема ОП в четыре раза без прямо пропорционального роста энергопотребления. Приведены структура контроллера памяти и принятые решения, позволяющие увеличить целевую частоту работы устройства на 30% до 800 МГц, а также повысить надежность работы канала памяти.

Ключевые слова: сеть-на-кристалле, контроллер памяти, DDR4 3DS, физический дизайн

Для цитирования:

Петров И. А. Разработка контроллера оперативной памяти для применения в современных микропроцессорах семейства «Эльбрус» // Радиопромышленность. 2019. Т. 29, № 3. С. 41–47. DOI: 10.21778/2413-9599-2019-29-3-41-47

© Петров И. А., 2019



Development of memory controller for today's Elbrus microprocessors

I. A. Petrov¹

¹MCST JSC, Moscow, Russia

The introduction of a new generation of microprocessors that belong to the Elbrus family and involve the introduction of a network-on-chip, requires the development of efficient means of access to DDR random access memory channels for network nodes. The paper includes a solution to this issue related to the interaction between DDR4 RAM and Elbrus-16CB, 16-core microprocessor, which demands higher standards of an available capacity and peak bandwidth of memory channels. When designing Elbrus-16CB microprocessor, higher energy efficiency and reliability are also between main objectives. When performing the tasks set, an important component was adaptation of the memory controller, successfully applied in the microprocessors produced by MCST JSC, to DDR4 3DS standard compliance, taken as a basis for the use in a number of recent developments. It provides a four-time higher available RAM capacity without a directly proportional growth of energy consumption. The paper includes a structure of the memory controller and made decisions. These make it possible to increase the target frequency in operations of the device by 30% up to 800 MHz and increase operation reliability of the memory channel.

Keywords: network-on-chip, memory controller, DDR4 3DS, design

For citation:

Petrov I. A. Development of memory controller for today's Elbrus microprocessors. Radio industry (Russia), 2019, vol. 29, no. 3, pp. 41–47. (In Russian). DOI: 10.21778/2413-9599-2019-29-3-41-47

Введение

Постоянное увеличение числа блоков в составе кристалла микропроцессора стало одной из определяющих проблем современной микроэлектроники. Это связано и с принципиальным развитием современных компьютерных архитектур (наращиванием числа процессорных ядер в кристалле или распределением по кристаллу банков кэш-памяти последнего уровня), так и с расширением сферы практического применения вычислительных систем, что приводит к увеличению взаимодействующего с микропроцессором внешнего поля. В связи с этим важнейшим фактором проектирования становится выбор связывающей блоки коммуникационной среды кристалла, которая обладала бы должными масштабируемостью, производительностью и гибкостью. Современным решением этой задачи является сеть-на-кристалле (network-on-chip, NoC), в одной из своих реализаций наиболее соответствующая архитектуре разрабатываемого микропроцессора [1].

Подобное решение стало парадигмой проектной работы АО «МЦСТ», направленной на создание отечественных микропроцессоров высокой производительности. В частности, оно было принято и при завершающейся в настоящее время разработке микропроцессора «Эльбрус-16СВ», где в кристалле объединяются 16 процессорных ядер, распределенный L3-кэш и набор контроллеров, обеспечивающих

доступ к оперативной памяти (ОП) и к устройствам ввода-вывода I/O (рис.1).

Одна из важнейших задач при проектировании первого микропроцессора семейства «Эльбрус» с технологией NoC – разработка узлов, обеспечивающих взаимодействие в сети. С принципиальной точки зрения выполняемые каждым из них функции подразделяются на две группы. Функции первой группы обеспечивают прием и передачу сетевых пакетов с предусмотренным качеством услуг (функции роутера), функции второй группы – эффективную связь с подключенным к узлу архитектурным блоком микропроцессора. В этом смысле особое значение приобретает взаимодействие узлов микропроцессора с модулями подсистемы ОП, которое оказывает решающее влияние на итоговую производительность микропроцессора. Цель статьи – рассмотрение данной проблемы.

Согласно свойственной промышленным разработкам тенденции использовать по возможности существенно развитые решения предыдущих проектов во вновь создаваемой продукции, описанные в статье результаты получены на основании того, что при организации доступа к ОП использовался контроллер собственной разработки [2], эволюционировавший до поддержки стандарта DDR4 (JEDEC Standard DDR4 SDRAM, JESD79-4B) в микропроцессоре «Эльбрус-8СВ». В статье приведен обзор улучшений и доработок этого устройства, которые позволяют

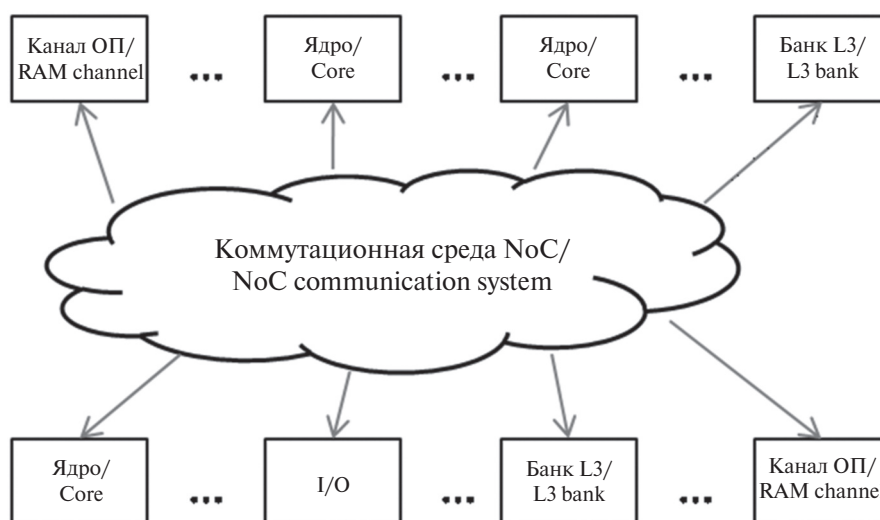


Рисунок 1. Схематическое представление микропроцессора «Эльбрус-16СВ»: ОП – оперативная память; NoC (Network-on-Chip) – сеть-на-кристалле

Figure 1. Schematic diagram of Elbrus microprocessor 16CB: RAM – a random access memory, NoC – network-on-chip

получить показатели, установленные в проекте «Эльбрус-16СВ» [3], и будут использованы в серии запланированных проектов. Применительно к этому рассматриваются повышенные требования к доступному объему и пиковой производительности ОП, а также к функциям категории RAS (надежность, доступность, ремонтпригодность).

Оперативная память микропроцессора «Эльбрус-16СВ»

Поддержка стандарта DDR4 3DS

Микросхема, выполненная по стандарту DDR4 3DS [4], предусматривающему существенно увеличенный по сравнению с предыдущими разработками АО «МЦСТ» объем ОП, – это собранная в одном корпусе вертикальная сборка (стек) из обычных чипов SDRAM DDR4, построенная на технологии «связи сквозь кремний» (through silicon via, TSV). Протокол 3DS вводит понятие «логический ранк» (один чип в стеке) в дополнение к физическому ранку. Максимальное количество логических ранков в стеке равно восьми, поэтому по сравнению с возможностями предыдущего микропроцессора «Эльбрус-8СВ» максимальный объем, поддерживаемый контроллером памяти в одном канале, увеличивается в восемь раз до 1 Тбайт. В данный момент основные производители микросхем памяти выпускают модули DDR4 3DS 4H Stack объемом 128 Гбайт, что позволяет набрать объем памяти 2 Тбайт для одного процессора «Эльбрус-16СВ» при заполнении двух слотов в каждом канале.

Механизм интерливинга

В каждом логическом банке (размером от 256 Мбайт до 2 Гбайт) можно одновременно задей-

ствовать лишь одну страницу памяти (8 Кбайт). Переход к другой странице в рамках одного логического банка возможен через 45–50 нс. Для увеличения количества одновременно использующихся страниц в канале памяти используется механизм интерливинга – процесс перемешивания изначально идущих подряд страниц памяти между логическими банками, находящимися в разных физических и логических ранках и группах логических ранков. На этапе инициализации системы в контроллере памяти задаются положения адресных разрядов запроса, определяющих номер логического банка в пространстве логических ранков, куда должен попасть запрос. Использование в большей степени межстекового интерливинга по сравнению с интерливингом между физическими ранками может повысить реальную пропускную способность канала из-за меньших задержек на переключениях между стеками.

Наращивание емкости модуля памяти вследствие использования DDR4 3DS не приводит к пропорциональному росту энергопотребления и снижению рабочей частоты, так как нагрузка на шину данных остается сопоставимой с обычными модулями DDR и не зависит от количества логических ранков в стеке [5].

Модули с поддержкой ECC-коррекции

При использовании модулей с поддержкой ECC-коррекции перед началом работы системы с ОП потребуется заполнение ECC-кода в ней корректными данными, так как после включения питания содержимое ячеек DRAM с ECC-кодом не соответствует содержимому ячеек DRAM с привязанным блоком данных [6]. Полная начальная инициализация памяти может осуществляться программно, но для сокра-

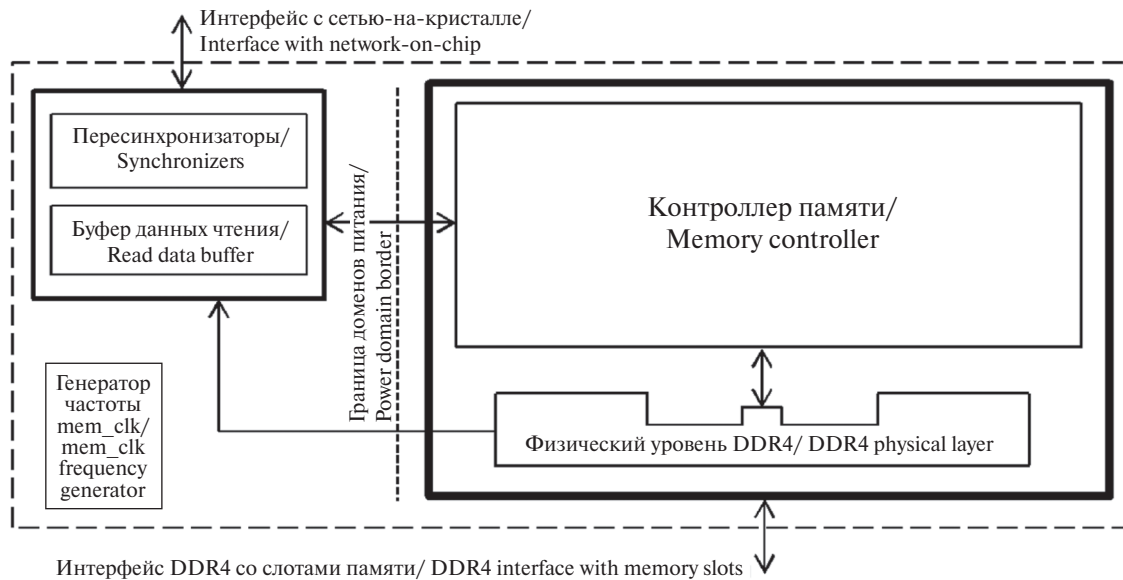


Рисунок 2. Упрощенная структура узла доступа в канал оперативной памяти
Figure 2. Simplified structure of RAM channel access point

щения времени исполнения в контроллер памяти заложена возможность сделать очистку ОП аппаратными средствами. Этот процесс, например для 32 Гбайт модуля DDR4-2400, занимает ~2 с. В связи с использованием более объемных 3DS-чипов алгоритм аппаратной чистки памяти был доработан. Контроллер памяти больше не блокирует исполнение системных запросов до окончания чистки. Она выполняется в фоновом режиме, а программа начальной загрузки системы (boot) получает возможность работать с участком памяти сразу после того, как он был проинициализирован.

Системные решения

Каналы ОП, управляемые независимыми устройствами, расположены в узлах сети-на-кристалле. Упрощенная структурная схема одного такого узла представлена на рис. 2.

В этой реализации контроллер памяти (memory controller, MC) и физический уровень DDR впервые в серии процессоров «Эльбрус» объединены структурно и пространственно в один блок. Эта интеграция имеет следующие преимущества:

- сокращение времени проведения физического дизайна и возможность выполнить его без дополнительных затрат в параллельных или последующих разработках процессоров «Эльбрус» с внедрением сетей-на-кристалле;
- упрощенное построение сетки синхросигнала mem_clk из-за локализации устройств, его использующих. При этом отпадает необходимость в использовании техники передачи синхросигнала вместе с данными – Source-Synchronous Clocking. Она применялась в «Эльбрус-8СВ» из-за пе-

редач на большие расстояния между составными частями устройства доступа в ОП и пересечения по пути передач независимых управляемых доменов питания VDD. В этом варианте синхросигнал становится асинхронным по отношению к исходному mem_clk, из-за чего требуется вставка дополнительного уровня пересинхронизации, увеличивающая задержку передачи данных.

Система питания физического уровня DDR4 чувствительна к асинхронным шумам, поэтому для устройства доступа в ОП выделяется независимый домен питания VDD. Переход через границу доменов питания осуществляется полутаковыми регистровыми передачами с запретом вставки промежуточной комбинационной логики.

Функционал контроллера Структура контроллера

Основные составляющие контроллера представлены на рис. 3.

Буфер данных записи расположен в контроллере, тогда как буфер данных чтения, в который с физического уровня DDR4 пересылаются данные, полученные из канала памяти (см. рис. 2), непосредственно в контроллер не входит. Оба буфера заполняются в соответствии с логикой, определяемой планировщиком, исходя из текущего состояния реестра запросов.

Особенностью интерфейса DDR4 является необходимость соблюдать задержки – протокольные блокировки между операциями на шине памяти, величина которых зависит от времени, занятого смежными операциями. Контроль блокировок производится независимо для каждого логического банка

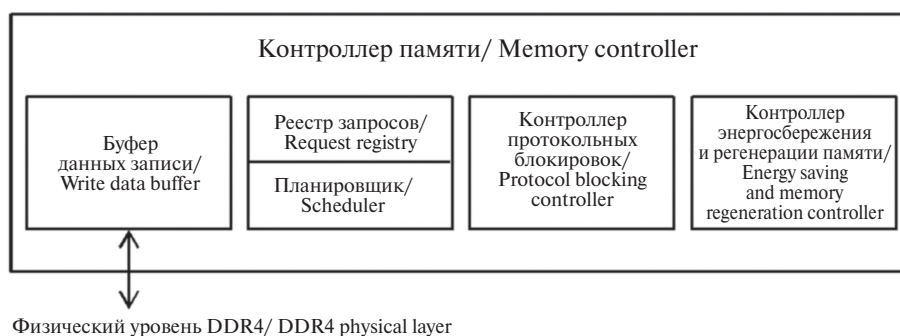


Рисунок 3. Функциональная структура контроллера
Figure 3. Controller's functional structure

при помощи однотипных наборов счетчиков-таймеров. Появление понятия логического ранка эквивалентно добавлению четвертой размерности в существовавшую трехмерную структуру массива логических банков: физический ранк–группа логических банков–логический банк внутри группы. В связи с этим в контроллере протокольных блокировок было выполнено экстенсивное расширение числа счетчиков-таймеров. Аналогичное расширение было произведено в контроллере энергосбережения и регенерации памяти, так как нельзя одновременно выдавать команду регенерации памяти в несколько логических ранков.

Планировщик и реестр заявок

В контроллере памяти планировщик объединен с реестром, который выполняет арбитраж заявок (выбор очередной заявки для исполнения). Выбор проводится на основе параметров, хранящихся в каждой ячейке реестра. К ним относятся как системные параметры (адрес, тип операции), так и параметры, модифицируемые контроллером памяти (признак готовности данных для операции записи, возраст запроса, стадия обработки запроса). Задача планировщика – изменение порядка обработки заявок таким образом, чтобы минимизировать среднее время, затрачиваемое на одну заявку, за счет сокращения числа вспомогательных операций открытия и предварительного заряда страниц, а также исключения чередования операций чтения и записи. Принцип работы планировщика состоит в обеспечении максимальной загрузки интерфейса DDR4 с модулями памяти, т. е. исключении пустых циклов на шине данных. Внутренняя организация планировщика представляет собой систему последовательно соединенных фильтров, на выходе которой размещен блок формирования элементарных операций для адресно-командной шины интерфейса DDR4. Используются следующие фильтры:

- фильтр ресурсов, задерживающий запросы, выполнение которых невозможно вследствие недо-

статочности ресурсов (если в варианте чтения отсутствуют места для сохранения считанных данных в буфере данных чтения, а в варианте записи – данные в буфере данных записи);

- фильтр адресной зависимости, обеспечивающий корректность последовательности обращений по одному адресу и определяющий конфликты типа WAR (запись после чтения) и RAW (чтение после записи), запрещая выполнение запросов, пришедших позднее;
- фильтр приоритета выполнения запроса в открытую страницу логического банка, исключающий дополнительные операции открытия страниц. Запрос отсеивается, если он попадает в закрытую страницу, хотя в то же время есть запросы, адресованные в открытую страницу;
- фильтр приоритета операций чтения, ориентированный на задержку запросов записи при наличии запросов чтения ввиду того, что операции чтения более критичны с точки зрения общей производительности системы;
- фильтр протокольных блокировок, на основании состояния таймеров контроллера протокольных блокировок задерживающий запросы, выполнение которых в данный момент нарушает протокол DDR4;
- фильтр возраста, пропускающий старший по возрасту запрос из всех имеющихся на входе.

Прирост количества ядер и поддерживаемого объема памяти требует обеспечить достаточное число позиций в реестре заявок. В совокупности с предусмотренным повышением пиковой пропускной способности канала памяти с 19,2 Гбайт/с ($f_{mem_clk} = 600$ МГц) до 25,6 Гбайт/с ($f_{mem_clk} = 800$ МГц) это привело бы, во-первых, к неприемлемо долгому времени выполнения физического синтеза контроллера памяти, во-вторых – к увеличению времени прохождения запросов через систему фильтров с одного целевого такта синхросигнала mem_clk до нескольких, что усложнило бы систему фильтров вследствие необходимости ее конвейеризации.

В связи с этим было определено, что проблемы с синтезом устройства обусловлены резким возрастанием комбинационной логики в фильтре адресной зависимости и фильтре приоритета выполнения запроса в открытую страницу логического банка при использовании компараторов адреса в формате «каждая ячейка с каждой». В результате массив адресов запросов фактически представлял собой ассоциативную память (content addressable memory, CAM) с N (число ячеек) портами для поиска. Во избежание этого плохо масштабируемого решения для каждого запроса к параметрам, модифицируемым контроллером памяти, добавлены признаки:

- 1) «в реестре есть запрос по пересекающемуся адресу с меньшим возрастом»;
- 2) «логический банк для данного запроса занят».

Оба признака обновляются на основе адресного сравнения при поступлении новых запросов с двух входных портов и после выдачи планировщиком запросов на интерфейс DDR4 (два порта). Таким образом, число портов CAM-памяти массива адресов уменьшилось до четырех. Это позволило преобразовать зависимость количества необходимых элементарных логических элементов от увеличения числа ячеек в реестре из квадратичной в линейную. Соответственно, физический синтез устройства стал выполняться в приемлемые сроки, а прохождение запросов через систему фильтров осталось однократным. В результате количество ячеек в реестре заявок было удвоено относительно этого показателя контроллера памяти процессора «Эльбрус-8СВ» с сохранением целевой частоты работы контроллера памяти 800 МГц (DDR4-3200).

Фактор RAS

В новой версии контроллера памяти большое внимание было уделено улучшению функциональности RAS (надежность, доступность ремонтпригодность) [7] вследствие того, что с увеличением показателей целевой частоты и доступного объема возрастает вероятность возникновения аппаратных сбоев в микросхемах памяти и при передачах по интерфейсу DDR4.

Добавлена механика исправления ошибок двух типов на DDR4-интерфейсе: ошибок четности на адресно-командной шине и ошибок CRC (циклический

избыточный код) для данных по записи. При детектировании ошибки контроллер вместо выдачи сигнала об аварийном завершении работы системы заново выдает на интерфейс DDR4 последовательность команд, исполнение которых не гарантируется в модуле памяти во временной окрестности сбоя, при этом обеспечивая когерентность памяти.

Для контроля целостности ОП реализован механизм патрулирования – периодического (настраиваемого программно) вычитывания с помощью операций чтение–модификация–запись всего объема памяти в целях исправления накопившейся одиночной ECC-ошибки. По практическому опыту предыдущих проектов было решено при возникновении множественной ошибки ECC расширить регистрацию места ее возникновения с точностью до номера чипа SDRAM и бита шины данных, тем самым, например, сократив время, необходимое для обнаружения устойчиво сбойной линии передачи данных на участке от модуля памяти до процессора.

В контроллер памяти добавлена возможность настройки на лету уровня опорного напряжения шины данных в чипах памяти, необходимая для получения глазковых диаграмм, характеризующих работоспособность канала памяти [8].

Выводы

Выбранная система коммутации в разрабатываемом микропроцессоре «Эльбрус-16СВ» требовала реорганизации подсистемы доступа в ОП. В результате был сформирован новый структурный блок – узел доступа сети-на-кристалле к ОП. Универсальный характер принятого решения позволяет быстро внедрить его в ряд дальнейших разработок компании. В ходе работы также была решена задача по адаптации контроллера памяти собственной разработки под увеличившиеся требования к объему, пропускной способности и надежности каналов памяти. Повышение объема ОП обеспечивается внедрением поддержки стандарта DDR4 3DS. Максимальная частота работы контроллера памяти доведена до 800 МГц, что стало возможно благодаря существенному сокращению оборудования для системы фильтров планировщика контроллера. При этом размер реестра заявок удалось удвоить, что способствует оптимальному использованию пропускной способности канала памяти.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Недбайло Ю. А. Проблемы масштабирования производительности подсистемы памяти многоядерного микропроцессора и методы их решения // Вопросы радиоэлектроники. 2018. № 2. С. 23–31.
2. Кожин А. С. Контроллер памяти DDR2 SDRAM и его система синхронизации в составе системы на кристалле «Эльбрус-S2» // Труды 52-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». Часть I. Радиотехника и кибернетика. Т. 1. М.: МФТИ, 2009. С. 55–58.

3. Ким А. К., Перекаатов В. И., Фельдман В. М. На пути к российской экзасистеме: планы разработчиков аппаратно-программной платформы «Эльбрус» по созданию суперкомпьютера эксафлопсной производительности // Вопросы радиоэлектроники. 2018. № 2. С. 6–13.
4. *Addendum No. 1 to JESD79-4, 3D Stacked DRAM*. JEDEC, 2017 [Электронный ресурс]. URL: <https://www.jedec.org/standards-documents/docs/jesd79-4-1> (дата обращения: 03.07.2019).
5. Choi J. S. Next Big Thing: DDR4 3DS. *Server Forum*, 2014. [Электронный ресурс]. URL: [https://www.jedec.org/sites/default/files/files/JS_Choi__Server_Forum_2014\(1\).pdf](https://www.jedec.org/sites/default/files/files/JS_Choi__Server_Forum_2014(1).pdf) (дата обращения: 03.07.2019).
6. Dice P. *Quick Boot: A Guide for Embedded Firmware Developers*, 2nd ed., Berlin, De Gruyter Press, 2017. 284 p.
7. Fruehe J. AMD EPYC Brings New RAS Capability. *Moor Insights & Strategy*, 2017. [Электронный ресурс]. URL: <https://www.amd.com/system/files/2017-06/AMD-EPYC-Brings-New-RAS-Capability.pdf> (дата обращения: 03.07.2019).
8. Билялетдинов И. Е., Ометов А. Е., Тимин Л. С. Оптимизация параметров высокоскоростных каналов процессора с целью повышения отказоустойчивости вычислительного комплекса // Вопросы радиоэлектроники. 2018. № 2. С. 87–92.

REFERENCES

1. Nedbailo Yu. A. Memory subsystem performance scaling problems in chip multiprocessors and their solution. *Voprosy radioelektroniki*, 2018, no. 2, pp. 23–31. (In Russian).
2. Kozhin A. S. Kontroller pamyati DDR2 SDRAM i ego sistema sinkhronizatsii v sostave sistemy na kristalle Elbrus-S2. *Trudy 52 nauchnoi konferentsii MFTI "Sovremennye problemy fundamentalnykh i prikladnykh nauk". Chast I. Radiotekhnika i kibernetika* [Modern Challenges of Fundamental and Applied Sciences. Part I. Radioengineering and Cybernetics. Proceedings of the 52nd Scientific Conference of MIPT]. Vol. 1. Moscow, MFTI, 2009, pp. 55–58. (In Russian).
3. Kim A. K., Perekatov V. I., Feldman V. M. On the way to russian exasistemes: plans of the Elbrus hardware-software platform developers on creation of an exaflops performance supercomputer. *Voprosy radioelektroniki*, 2018, no. 2. pp. 6–13. (In Russian).
4. *Addendum No. 1 to JESD79-4, 3D Stacked DRAM*. JEDEC, 2017. Available at: <https://www.jedec.org/standards-documents/docs/jesd79-4-1> (accessed 03.07.2019).
5. Choi J. S. Next Big Thing: DDR4 3DS. *Server Forum*, 2014. Available at: [https://www.jedec.org/sites/default/files/files/JS_Choi__Server_Forum_2014\(1\).pdf](https://www.jedec.org/sites/default/files/files/JS_Choi__Server_Forum_2014(1).pdf) (accessed 03.07.2019).
6. Dice P. *Quick Boot: A Guide for Embedded Firmware Developers*, 2nd ed., Berlin, De Gruyter Press, 2017. 284 p.
7. Fruehe J. AMD EPYC Brings New RAS Capability. *Moor Insights & Strategy*, 2017. Available at: <https://www.amd.com/system/files/2017-06/AMD-EPYC-Brings-New-RAS-Capability.pdf> (accessed 03.07.2019).
8. Bilyaletdinov I. E., Ometov A. E., Timin L. S. Optimization of parameters of high-speed channels of the processor in order to increase the fault tolerance of the computer complex. *Voprosy radioelektroniki*, 2018, no. 2, pp. 87–92. (In Russian).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Петров Игорь Александрович, старший инженер, АО «МЦСТ», 119334, Москва, ул. Вавилова, д. 24, тел.: +7 (499) 135-33-61, e-mail: petrov_i@mcst.ru.

AUTHOR

Igor A. Petrov, senior engineer, MCST JSC, 24, ulitsa Vavilova, Moscow, 119334, Russia, tel.: +7 (499) 135-33-61, e-mail: petrov_i@mcst.ru.

Поступила 29.04.2019; принята к публикации 22.06.2019; опубликована онлайн 21.08.2019.
Submitted 29.04.2019; revised 22.06.2019; published online 21.08.2019.