УДК 004.382

Применение справочника для поддержки когерентности памяти в системе "Эльбрус-2S"

А.Е. Шерстнёв
ОАО «ИНЭУМ», ЗАО «МЦСТ»
andrewsh84@gmail.com

"Эльбрус-3S" - это созданный ЗАО "МЦСТ" многопроцессорный вычислительный комплекс, каждый процессорный модуль которого представляет собой систему - на - кристалле (System-on-Chip, SOC) [1, 4]. На одном кристалле размещаются процессор и оборудование распределённой интерфейсной логики (chipset) - "северного моста". При помощи высокоскоростных последовательных каналов межпроцессорного обмена от двух до четырёх процессоров могут быть объединены в когерентную систему. Каналы связывают каждый процессор со всеми остальными. Система не имеет общего контроллера памяти, на одном кристалле с каждым процессором интегрирован собственный контроллер, работающий с подключенной к процессору памятью [3]. Таким образом, "Эльбрус-3S" является системой с неоднородным доступом в память (NUMA) [5]. Когерентность при доступе в память в NUMA-системе "Эльбрус-3S" поддерживается на аппаратном уровне при помощи техники снупирования. Это означает, что при каждом обращении процессора в память формируются запросы проверки когерентности (снуп-запросы), направляемые в контроллер кэша каждого процессора [2].

Разработанный протокол межпроцессорного обмена ориентирован на эффективное взаимодействие процессоров внутри кластера, включающего не более четырёх процессорных модулей. К плюсам рассмотренного протокола можно отнести его относительную простоту реализации и верификации, а также независимость от числа процессоров. Основным минусом является то, что в системах с числом процессоров >4, организуемых при помощи дополнительно чипа - межкластерного коммутатора, время исполнения каждого запроса резко увеличивается из-за необходимости опроса кэшей всех процессоров большой системы. При этом, за счёт необходимости передачи снуп-запросов снижается эффективная пропускная способность каналов межпроцессорного обмена.

Проблему перегрузки межпроцессорных каналов снуп-запросами можно решить, если в момент отработки очередного запроса точно знать кэшированы ли запрашиваемые данные, и, если да, то в каком процессоре. Для этого в систему вводится дополнительный модуль - справочник, хранящий информацию о местоположении кэш-строк. При отработке запроса команда чтения отсылается в контроллер памяти, если данные не кэшированы ни в одном процессоре, или формируется один снуп-запрос владельцу модифицированных данных. Грубый подсчёт показывает, что суммарное число отосланных сообщений сокращается в 4 раза.

Справочник можно реализовать одним из двух способов: полный справочник, хранящий информацию о каждой строке оперативной памяти системы, и усечённый справочник, представляющий собой кэш-память, ячейки которой хранят информацию о некотором множестве данных, кэшированных процессорами системы [1]. Полный справочник подразумевает зависимость от общего объёма оперативной памяти (ОП) процессоров системы и поэтому рациональнее всего его размещение в самой ОП. Усечённый справочник может содержаться в заранее определённом на стадии проектирования массиве памяти, аналогичной основной кэш-памяти процессора. Главные преимущества и недостатки справочников обоих типов собраны в таблице.

	Преимущества	Недостатки
Полный справочник	Хранит информацию о всей памяти системы. При запросе по любому адресу формируется минимально	Большие затраты ресурсов на реализацию. Объём справочника определяется максимальным

	необходимое число запросов поддержания когерентности.	объёмом оперативной памяти во всей системе.
Усечённый справочник	По сравнению с полным справочником на несколько порядков сокращается объем памяти, необходимый для организации справочника.	Влияние эффекта «старения» строк в кэш-памяти справочника на состояние кэш-памяти процессоров в кластере.

Технология справочника применяется в процессоре следующего поколения "Эльбрус-2S". Основные характеристики процессора таковы:

Архитектура - Эльбрус (VLIW)

Частота - 800МГц

Количество ядер - 4

Тип памяти - DDR3-1600

Количество каналов памяти - 3

В результате анализа преимуществ и недостатков справочников двух типов был сделан выбор в пользу полного справочника, хранящегося в ОП. Распределение информации справочника между процессорами соответствует распределению ОП. То есть информация о кэш-строках (64-байтовых блоках) памяти, принадлежащих определённому процессору, хранится в ОП этого процессора. Такого рода локализация позволяет относительно быстрый доступ к информации справочника. На основании того, что каждый запрос в ОП за данными порождает чтение справочника из этой же области ОП, применяется более глубокая оптимизация, ориентированная на наиболее распространённую на данный момент оперативную динамическую память страничной организации, - данные и относящаяся к ним часть справочника хранятся в одной странице памяти. Это позволяет сократить накладные расходы на интерфейсе DRAM на переоткрытие страниц.

Считывание данных из памяти, обработку информации справочника и рассылку необходимых снуп-запросов выполняет "северный мост", интегрированный на одном кристалле с процессором. Таким образом, в процессе работы каждый исходный процессорный запрос порождает как минимум одно обращение в ОП - за данными справочника. Время считывания из ОП составляет примерно 25ns (DDR2,DDR3)[6, 7]. Для избегания такого рода затрат в "северный мост" вводится кэш для данных справочника, аналогичный кэшу памяти процессоров. Использование кэша преследует две основные цели: доступ к данным справочника с минимальной задержкой (эта задержка определяет также полное время выполнения исходного запроса) и уменьшение нагрузки на интерфейс с ОП. Уменьшение задержки получения информации справочника играет роль в случае, если модифицированная копия запрашиваемых данных содержится в кэше какого-либо процессора. Время исполнения запроса при этом сокращается на время считывания справочника, т.е. на 25ns, что составляет порядка 30% от общего времени выполнения запроса в рассматриваемой системе с неоднородным доступом к памяти. Уменьшение нагрузки на интерфейс с ОП имеет постоянный характер: из ОП будут считываться только "полезные" данные, необходимые инициатору исходного запроса, данные справочника берутся из кэша.

В кэше справочника используется протокол MOSI [8]. Структура кэша справочника определялась на компьютерной модели системы при прогоне тестов, описывающих разнообразные задачи. В итоге, объём кэша справочника равен 512КБайт, ассоциативность - 16, размер кэш-строки - 64Байта, размер элемента справочника (информация, относящаяся к одной процессорной кэш-строке) – 1Байт для 4-процессорной системы, 1.5Байта для 16-процессорной системы. Хранение справочника в ОП требует <1.6% от общего объёма ОП. Кэш справочника встраивается в "северный мост" рядом с 4-стадийным конвейером запросов. С первой стадии конвейера в кэш справочника выдаётся адрес исходного запроса, принимаемого к исполнению, на 4-й стадии кэш справочника выдаёт информацию о местонахождении строки или сообщает, что требуемая информация содержится в основной памяти справочника.

"Эльбрус-2S" представляет собой высокопроизводительный процессор, ориентированный главным образом на применения в больших серверных системах. Использование справочника

для оптимизации процесса снупирования уже применяется в последних разработках основных производителей процессоров. К примеру, Intel в новом поколении процессоров Itanium [9] реализует усечённый справочник объёмом 2МБайта. Подход, применяемый в "Эльбрус-2S", фактически обладает как преимуществами полного справочника – хранится информация о всех строках ОП, так и усечённого – быстрый доступ к данным справочника. В режиме расширенного справочника в дополнение к информации о своём кластере сохраняется информация о соседних кластерах. Возможность отключения справочника делает доступным весь объём ОП системы и позволяет использование процессора в небольших системах.

Список литературы

1. Зайцев А.И., Шерстнёв А.Е. Организация межпроцессорного обмена в многокластерных системах на базе микропроцессоров "Эльбрус-S" и "МЦСТ-4R" // Вопросы радиоэлектроники, серия ЭВТ, 2009 г., вып. 3.
2. Недбайло Ю.А., Шерстнёв А.Е. Оптимизация доступа к памяти в вычислительном комплексе "Эльбрус-3S", //«ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ» № 11/2008
3. Шерстнёв А.Е. Контроллер памяти DDR2 SDRAM ВК «Эльбрус-3S» //МЦСТ, 2009
4. Шерстнёв А.Е. Системный коммутатор ВК «Эльбрус-3S» //МЦСТ, 2009
5. A. Ahmed et al., “AMD Opteron Shared-Memory MP Systems”, http://www.hotchips.org/archive/hc14/program/28_AMD_Hammer_MP_HC_v8.pdf
6. DDR2 SDRAM Specification. Revision JESD79-2B, JEDEC Solid State Technology Association, January 2005, http://www.jedec.org
7. DDR3 SDRAM Specification. Revision JESD79-3D, JEDEC Solid State Technology Association, September 2009, http://www.jedec.org
8. Jim Handy, “The Cache Memory Book”, //Morgan Kaufmann 2nd edition 1998
9. L. Schaelicke and E. DeLano Intel Itanium Quad-Core Architecture for Enterprise, http://www.cgo.org/cgo2010/epic8/slides/schaelicke.pdf