

Заказные блоки памяти на технологии 28нм для кэш-памяти первого уровня микропроцессора серии «Эльбрус».

28nm custom memory units for first level cache of “Elbrus” line microprocessor.

Гунтис Викторович Кристовский, Юрий Лаврентьевич Погребной, Юрий Михайлович Потовин, Сергей Анатольевич Соин

Guntis V. Kristovskiy, Yury L. Pogrebnoy, Yury M. Potovin, Sergey A. Soin.

Yury.M.Potovin@mcst.ru

Акционерное общество «МЦСТ»

117105, г. Москва, ул. Нагатинская, д.1, стр.23. <http://mcst.ru>

В статье изложен подход к разработке заказных блоков памяти для кэша первого уровня 8-ядерного микропроцессора серии «Эльбрус». Для обеспечения высокого быстродействия и повышенной помехоустойчивости использована схемотехника, при которой все транзисторы работают с полным логическим перепадом. Описаны основные схемотехнические и топологические решения.

The paper deals with generation of custom memory units for 8 core “Elbrus” line microprocessor first level cache. A full swing approach is used to achieve high performance and enhanced noise margin. Main features of circuitry and layout of custom blocks are described.

Ключевые слова: кэш, заказное проектирование, ячейка памяти, многопортовая память, топология.

Key words: cache, custom design, memory cell, multi-port memory, layout.

1. Введение.

Основной тренд современной КМОП технологии - это постоянное уменьшение размеров элементов интегральных схем. Получаемые выгоды очевидны, однако уменьшение размеров порождает ряд проблем, вызванных увеличенным технологическим разбросом характеристик транзисторов и линий связи. Особенно ярко эти проблемы проявляются при проектировании встроенных устройств памяти. В данной работе излагается подход к проектированию блоков памяти на технологии 28нм для кэша первого уровня 8-ядерного микропроцессора серии "Эльбрус". При разработке методики проектирования блоков памяти были учтены следующие особенности технологии 28нм: уменьшенное напряжение питания – 0.9В, увеличенное сопротивление линий связи, изменившееся соотношение между усилительными свойствами n-канальных и p-канальных транзисторов, жесткие правила топологического проектирования. Особое внимание было уделено обеспечению надежной работы блоков памяти в условиях повышенных технологических разбросов.

В работе проведен анализ помехоустойчивости запоминающих ячеек. Исходя из результатов исследования, выбрана оптимальная конструкция ячеек памяти. Определены требования к кэшу первого уровня, предложены схемотехнические решения и разработана топология заказных блоков памяти.

2. Анализ помехоустойчивости запоминающих ячеек.

Классическая 6-ти транзисторная ячейка статической памяти (6Т) представлена на рис. 1. Два инвертора с перекрестными обратными связями образуют элемент, хранящий информацию. Линия WL (Word Line) управляет двумя транзисторами доступа M5, M6 (Pass transistors). Линии BL, BLB (Bit Line) — битовые линии, используются как для записи данных, так и для их чтения. Поскольку для входных и выходных данных используются одни и те же шины, операции чтения и записи должны быть разнесены во времени.

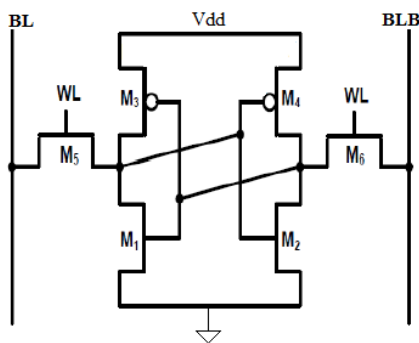


Рис.1. Электрическая схема 6-ти транзисторной ячейки памяти.

Область устойчивой работы запоминающих ячеек зависит от вариаций параметров транзисторов и линий связи, обусловленных технологией изготовления. Технологические разбросы приводят к трудно предсказуемым изменениям таких важнейших характеристик транзисторов, как порог, усилительные свойства и емкости. В какой-то степени эти изменения учитываются при создании моделей транзисторов для разных технологических углов. Однако если учесть, что на разброс технологии накладываются отклонения напряжения питания, а также работа в широком температурном диапазоне, то становится ясно, что для обеспечения устойчивой работы блоков памяти необходимо принимать специальные меры.

Общепринятой оценкой устойчивости ячейки памяти при чтении является статический запас помехоустойчивости (Static Noise Margin-SNM) [1,2]. На рис. 2 представлен способ определения SNM. Суть его сводится к наложению графиков статических передаточных характеристик двух инверторов, образующих триггер. Величина SNM равна длине стороны квадрата вписанного полученную картинку.

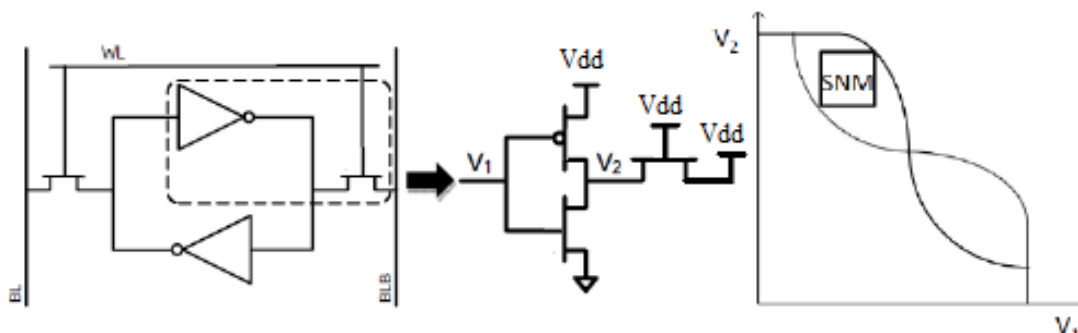


Рис. 2. Эквивалентная схема определения SNM

Запас помехоустойчивости по записи находится путем построения так называемой токовой характеристики. Для этого на битовых шинах фиксируются напряжения соответствующие записываемой информации (например,

$BL=V_{ss}$, $BLB=V_{dd}$, $WL=V_{dd}$, рис.1). Далее строится зависимость разности токов транзисторов M5 (Pass transistor) и M3 (Pull up transistor) от напряжения в узле между ними. Запас помехоустойчивости по записи равен величине тока, соответствующей точке переключения 2-го инвертора. При токе меньше или равном 0 запись не состоится.

Для обеспечения неразрушающего чтения, транзисторы M1 и M2 (Pull down transistors) должны быть «больше» транзисторов M5 и M6 (Pass transistors). Для устойчивой записи транзистор M5 (Pass transistor) должен быть «больше» M3 (Pull up transistor). Из приведенных рассуждений видно, что для устойчивой работы 6-и транзисторной ячейки в режимах записи и чтения необходимо выдержать жесткие соотношения между размерами всех транзисторов. В блоках памяти ячейки объединяются в массивы, в результате сигналы на шинах отличаются от идеальных. Кроме того, отличается от номинального и напряжение питания, поступающее на конкретную запоминающую ячейку. Следует отметить, что КМОП логические элементы продолжают сохранять работоспособность при снижении напряжения питания до 2/3 от номинала. На рисунке приведены SNM характеристики 6-и транзисторной ячейки при напряжениях питания 0.9В и 0.6В. Из рисунка видно, что понижение питания до 0.6V, существенно уменьшает SNM. Графики рисунка получены путем моделирования на программе HSPICE с использованием идеальных моделей транзисторов технологии 28nm фирмы TSMC.

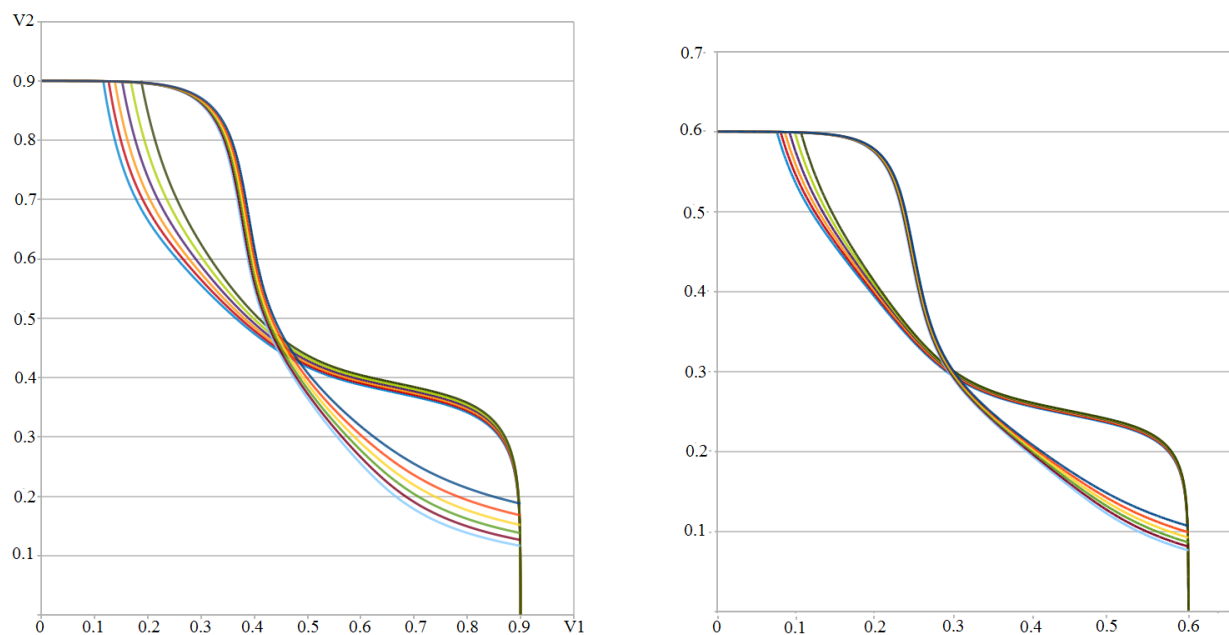


Рис.3. Графики определения SNM ячейки 6Т при уменьшении напряжения питания до 0.6В.

Таким образом, в условиях отсутствия достоверной информации о технологических разбросах практически невозможно гарантировать достаточные запасы помехоустойчивости 6-и транзисторной ячейки. Решением этой проблемы является использование 8-ми транзисторной ячейки (рис.4), в которой операции записи и чтения разделены [3]. Временная диаграмма порта чтения состоит из двух фаз. На фазе предзаряда осуществляется заряд битовых линий (RBL). На фазе считывания предзаряд отключается и подается сигнал на линию RWL (Read Word Line). В зависимости от того, что хранилось в ячейке, на RBL сохраняется потенциал, либо появляется ток, разряжающий до потенциала земли. Порт записи реализован также как в 6-и транзисторной ячейке.

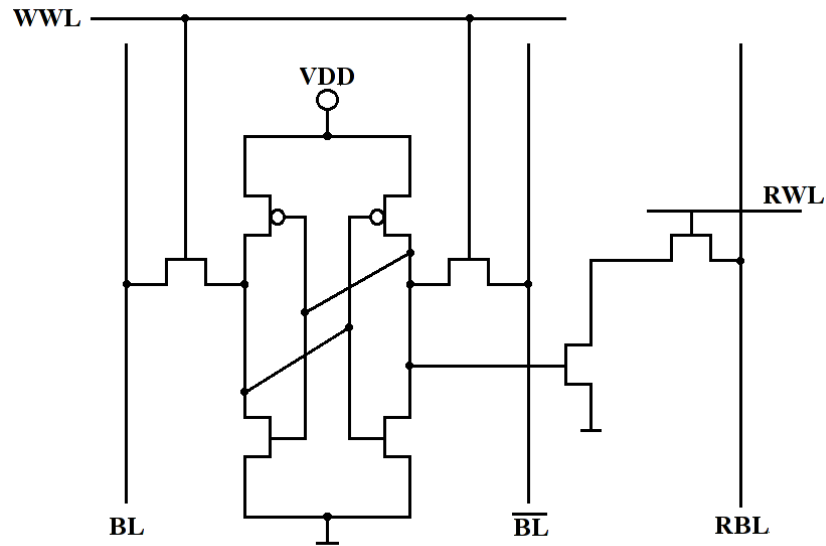


Рис.4.Электрическая схема 8-и транзисторной ячейки (8Т).

Аналогичное моделирование этой ячейки с использованием моделей транзисторов технологии 28нм показывает, что даже при уменьшенном напряжении питания и увеличении соотношения транзисторов M5 и M6 (Pass transistors) к M1 и M2 (Pull down transistors), SNM остается высоким (рис.5). Необходимо отметить, что разделение портов записи и считывания позволяет в этой ячейке добавлять порты чтения с минимальным увеличением площади.

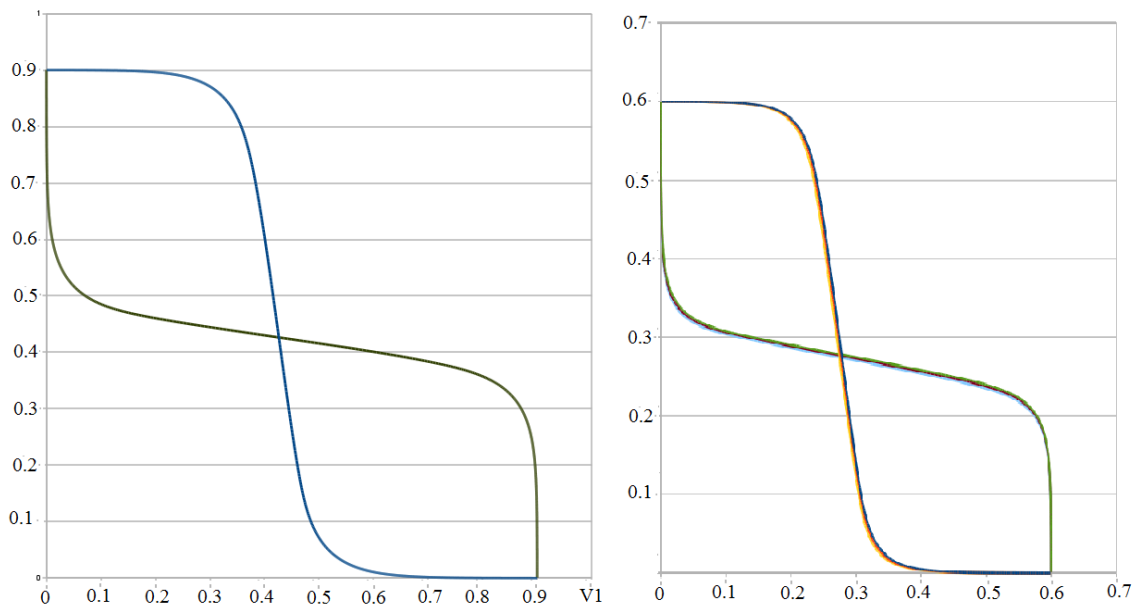


Рис. 5. SNM ячейки 8Т при напряжениях питания 0.9V и 0.6V.

3. Требования к блокам памяти в L1.

К кэшу первого уровня (L1) микропроцессоров серии «Эльбрус» предъявляется целый ряд достаточно жестких требований, так как в нем хранятся наиболее часто используемые данные. Функциональная схема кэша первого уровня (L1) показана на рисунке 6. L1 имеет 4-столбцовую, частично ассоциативную организацию. Память данных организована в виде 4 колонок по 2к 78 разрядных слов. Каждой колонке приписан свой блок памяти тегов емкостью 512 слов разрядностью 53. Строка адресуется индексом - разряды 13-5 адреса, номер слова определяют разряды 4-3, а разряды 2-0 используются для обращения к конкретному байту.

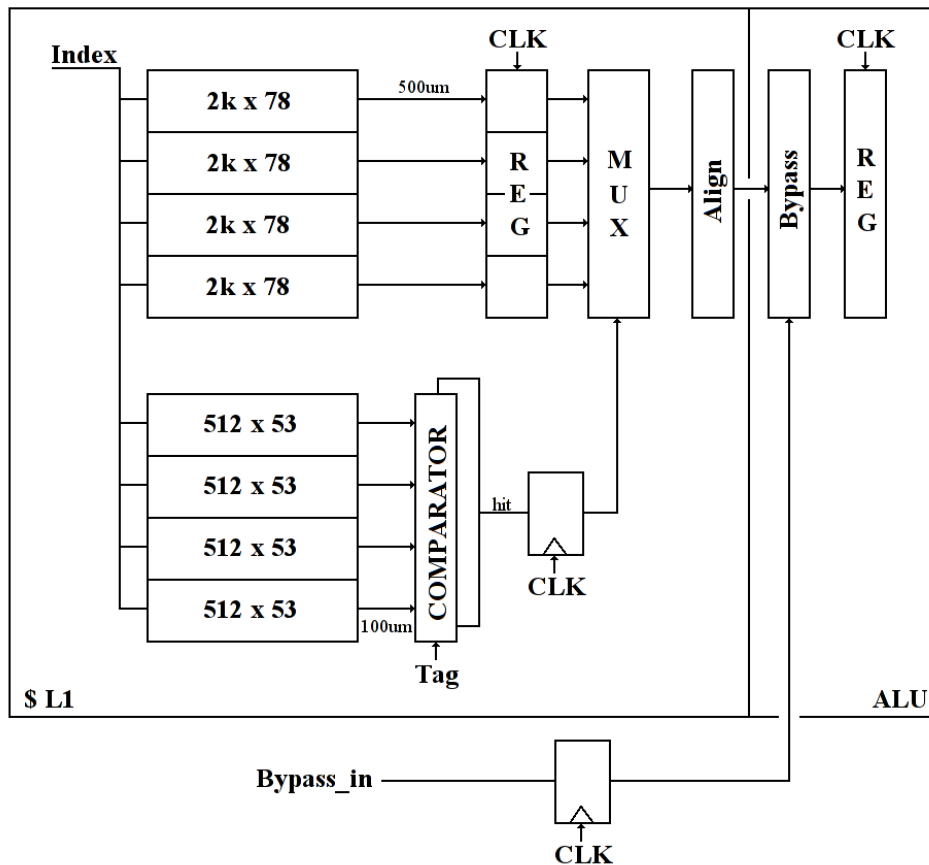


Рис.6. Функциональная схема КЭШа L1 процессора.

Время чтения из L1 должно составлять 2 такта. При чтении происходит одновременное обращение ко всем столбцам памяти данных и к памяти тегов. 4 слова, считанные с памяти данных, запоминаются на регистрах, максимально приближенных к блокам памяти. 4 тега поступают на 4 компаратора. Если какой-либо из прочитанных тегов совпадает с теговской частью адреса, генерируется сигнал hit, который управляет мультиплексором.

При записи данные модифицируются в столбце, для которого имел место "hit". Существенно, что блоки, используемые в памяти данных, должны позволять модифицировать данные с точностью до байта.

В рассматриваемом проекте L1 должен иметь 4 порта по чтению и 2 порта записи. По предварительным оценкам исполнительные устройства могут работать на частоте 1.5 ГГц, следовательно, такой же должна быть рабочая частота L1, что соответствует периоду синхронизации длительностью 670пс. Для получения чистого времени, приходящегося на работу кэша, из длительности периода синхронизации необходимо вычесть разброс синхросигнала, время установки и время задержки триггера. Предположим, что эти накладные расходы не превышают 100пс, тогда часть такта, отведенная для работы L1, составляет 570пс. Однако, необходимо учесть время распространения данных с выходов блоков памяти до приемных регистров. Согласно предварительным оценкам, длина связи примерно равна 500 мкм, что приводит к дополнительной задержке 70-80 пс, следовательно, время чтения из блока памяти данных должно быть меньше 500пс. Еще более жесткие временные требования предъявляются к блокам памяти тегов, так как в такте чтения дополнительно необходимо выполнить сравнение тегов. Моделирование схемы компаратора при помощи программы Spice дало задержку 110пс. Для памяти тегов в силу меньшей площади потери на связях могут быть меньше, чем для памяти данных, – предварительная оценка 50пс. Таким образом, задержка чтения из блока памяти тегов должна быть меньше 410пс.

Это требование накладывает жесткие ограничения на размеры блоков памяти. Предварительные оценки показали, что для памяти данных наилучшим образом подходят блоки, хранящие 512 слов с разрядностью 78, имеющие 4 порта чтения и 2 порта записи. Блоки памяти тегов также должны иметь 4 порта чтения, но только 1 порт записи и хранить 512 слов с разрядностью 53.

4.Схемотехнические и топологические решения, использованные при создании блоков памяти для кэша первого уровня.

Применение идеологии 8-транзисторной ячейки позволяет упростить схемотехнику блоков памяти и, тем самым, повысить быстродействие. Идея заключается в том, чтобы построить двухуровневую систему чтения: на первом уровне небольшие группы ячеек памяти подключаются к локальным битовым шинам. На втором уровне локальные битовые шины объединяются при помощи глобальной битовой шины. При таком подходе как

локальные, так и глобальные битовые шины имеют сравнительно небольшую емкость, что позволяет развить на этих шинах полный логический перепад. Отказ от малосигнального считывания позволяет исключить из тракта чтения усилители чтения.

На рисунке 7 приведена электрическая схема ячейки памяти, в которой реализованы 4 порта чтения и 2 порта записи (4R/2W).

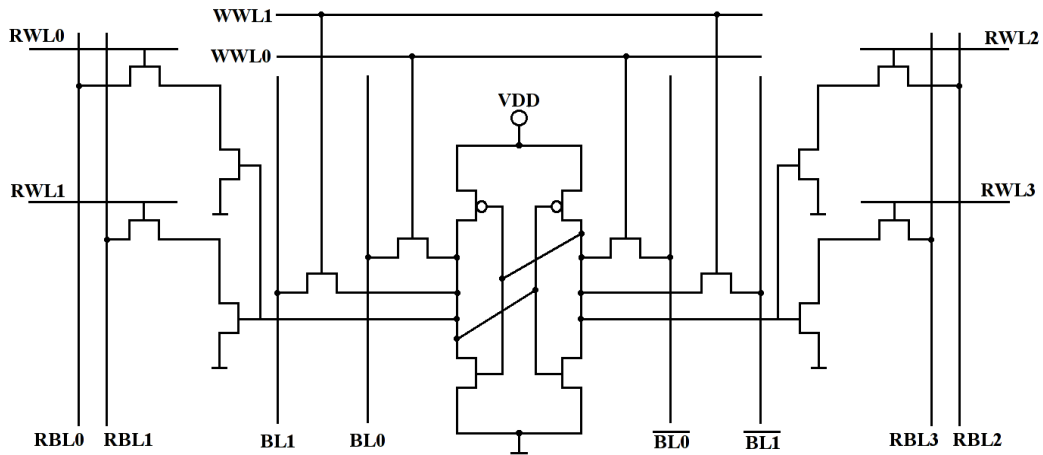


Рис. 7 Запоминающая ячейка с 4R/2W.

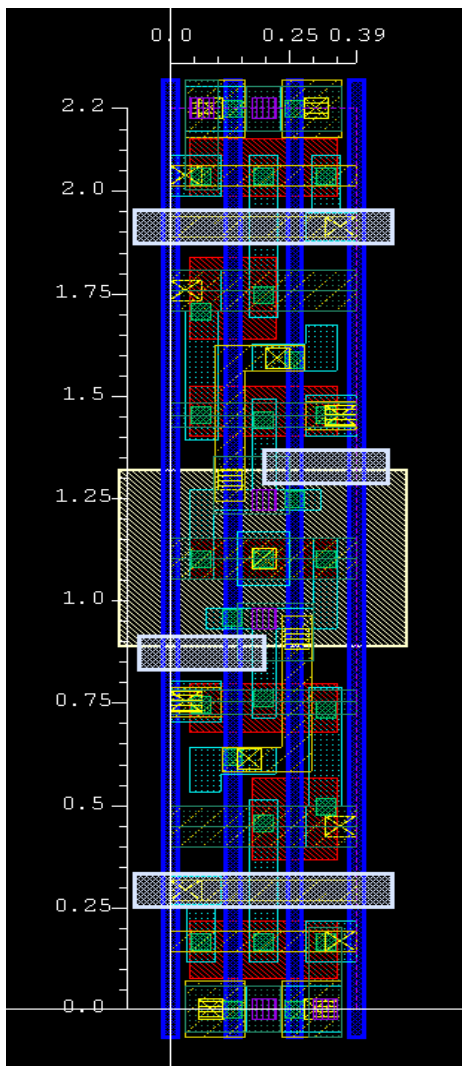


Рис.8 Топологический рисунок запоминающей ячейки 4R/2W.

Основным достоинством этой ячейки является ее простота, что позволяет минимизировать площадь. Словарные шины 4-х портов чтения и порта записи выполнены в вертикальных слоях металла M4, M6. Битовые шины портов чтения и записи - горизонтально в слоях M2, M3, M5. Топология выполнена в рекомендуемых проектных нормах технологии TSMC 28nm. Ширина запоминающей ячейки всего 0.39 мкм, что уменьшает длину как локальных, так и глобальных битовых шин.

Для дальнейшего уменьшения длины битовых шин в рамках единой конструкции блок разбивается на 2 подблока емкостью 256x78. Это решение позволило еще вдвое уменьшить длину локальных и глобальных битовых шин. В подблоке группы из 16 запоминающих ячеек подключаются к локальным битовым шинам, которые объединяются в пары. Восемь таких пар через усилительные транзисторы подключаются к глобальной битовой шине, которая управляет триггером типа SDL (Set Dominant Latch). Как на уровне локальной битовой шины, так и для глобальной битовой шины, используются схемы с предзарядом (рис. 9).

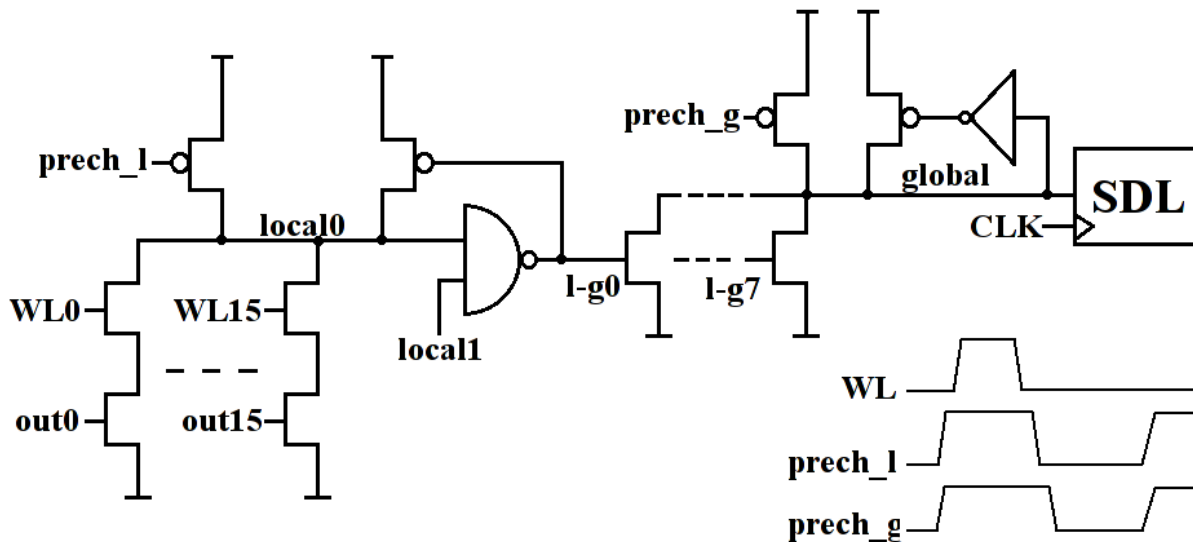


Рис.9. Схема тракта чтения

При разработке блока для памяти тегов была использована та же схемотехника, что и для блока памяти данных. Отличие этих блоков заключается в том, что для работы с тегами требуется только 1 порт записи. Кроме того, блок для хранения тегов имеет меньшую разрядность. Эти два упрощения позволили уменьшить задержку блока до 350пс.

5. Результаты.

Изложенный подход позволил создать блоки для памяти данных и памяти тегов с требуемыми электрическими и топологическими характеристиками. Основные параметры этих блоков приведены в таблице 1.

Таблица 1.

Область применения	Емкость	Площадь, кв.мкм	Время задержки, пс Сн=50фФ	Время установки адреса, пс
Память данных 4R/2W (заменяет 2 блока DP 1024x39)	512 x 78	51668	395	30
Станд. блок DP (Dolphin)	1024 x 39	26650	600	50
Память тегов (заменяет 2 блока DP 512x53)	512 x 53	37216	350	30
Станд. блок DP (Dolphin)	512 x 53	21261	540	50

Из таблицы видно, что разработанные блоки обладают необходимыми временными характеристиками. В отличие от стандартных блоков памяти, вновь разработанные блоки допускают одновременное обращение по портам записи и чтения, за исключением случая, когда адрес чтения совпадает с адресом записи. Независимость портов записи и чтения позволяет существенно упростить методику проверки работоспособности этих блоков.

На моделях, экстрагированных из топологии, было проведено исследование работоспособности разработанных блоков для допустимых отклонений параметров транзисторов, напряжения питания и температуры. Моделирование при помощи программы Finesim (SPICE подобная программа, разработанная фирмой Synopsys) подтвердило работоспособность для всех допустимых сочетаний названных параметров.

6. Выводы.

Указанные характеристики достигнуты за счет разделения портов записи и чтения и отказа от малосигнальной схемотехники, используемой в коммерческих скомпилированных блоках памяти. Мало-сигнальная схемотехника позволяет минимизировать площадь блоков памяти, так как ячейки памяти проектируются изготовителем по нормам, отличными от тех, которые используются при проектировании логических схем. Однако в настоящее время для технологии 28нм фирмы TSMC доступны только однопортовая и двухпортовая ячейки памяти. Наличие

четырёхпортовых блоков для памяти данных и тэгов дало возможность существенно оптимизировать топологию кэша L1. Прежде всего, это позволило отказаться от дублирования блоков памяти и тэгов. Кроме того, упростилась коммутация данных, выходящих из блоков памяти, и в 2 раза уменьшилось количество «получателей» адреса входящего обращения, а это один из самых напряженных критических путей кэша L1. Результатом упрощений стало уменьшение площади L1 на 0.32 мм² (15%). Сыграла свою роль и меньшая (по сравнению с блоками, получаемыми при помощи компилятора) задержка, она позволила достичь необходимой по ТЗ частоты без изменения конвейера L1. Стоит отметить, что разработанные блоки обладают повышенной помехозащищенностью. Таким образом, применение заказных блоков памяти в разрабатываемом 8-и ядерном микропроцессоре серии «Эльбрус» позволило получить необходимые временные характеристики кэша первого уровня и гарантировать повышенную помехоустойчивость этого устройства.

Авторы считают, что в данной работе новыми являются следующие положения и результаты:

- Применение в КЭШе микропроцессора блоков памяти, построенных на основе схем с полным логическим перепадом вместо традиционно используемых компилируемых блоков памяти на основе схем с малым сигналом, что позволило достичь заданной по ТЗ частоты без изменения конвейера L1, а так же сократить площадь на 15% и повысить помехозащищенность устройства.

Список использованной литературы

1. Grossar E. et al. “ Read Stability and Write-ability Analysis of SRAM Cells for Nanometer Technologies” // IEEE Journal of Solid -State Circuits, Vol. 41, pp 2577-2588, Nov 2006.
2. Guo Z. et al. “Large Scale SRAM Variability Characterization in 45 nm CMOS” // IEEE Journal of Solid -State Circuits, Vol. 44, pp 3174-3192, Nov 2009.
3. Chang L. et al. “An 8T SRAM for Variability Tolerance and Low-Voltage Operation in High Performance Caches” // IEEE Journal of Solid -State Circuits, Vol. 43, pp 956-963, March,2008.