

Новый отечественный процессор обработки сигналов 1879BM4 семейства NeuroMatrix

В.М. Черников, П.Е. Виксне, А.М. Шелухин, А.В. Черников, Д.Е. Косоруков

ЗАО НТЦ «Модуль», pvixne@module.ru

Аннотация — В статье представлены результаты разработки нового высокопроизводительного сигнального процессора. Подробно описана структура и система команд. Приведены оценки эффективности новых архитектурных решений. Приведены основные электрические и динамические параметры процессора. Рассмотрены перспективы развития семейства NeuroMatrix.

Ключевые слова — NeuroMatrix, DSP, SIMD, VLIW, матрично-векторный сопроцессор, обработка потоков данных.

I. ВВЕДЕНИЕ

Процессор 1879BM4 (NM6405) является дальнейшим развитием семейства отечественных процессоров семейства NeuroMatrix® ЗАО НТЦ "Модуль": Л1879BM1(NM6403) и 1879BM2(NM6404). Данные процессоры относятся к представителям нового класса векторно-конвейерных DSP. Их отличает высокая производительность на задачах обработки больших потоков данных при относительно небольших аппаратных затратах и малом потреблении питания.

Процессор 1879BM4 представляет собой высокопроизводительный матрично-векторный микропроцессор, имеющий оригинальную RISC-архитектуру с элементами VLIW (Very Long Instruction Word), SIMD (Single Instruction Multiple Data) и суперскаляра. Отличительными особенностями 1879BM4 являются:

- Аппаратная поддержка матричных и векторных операций, включающая в себя:

- § Выполнение двухвекторной АЛУ-операции за один процессорный такт.

- § Умножение предварительно загруженной матрицы данных (весовых коэффициентов) на вектор данных за один процессорный такт. Данная базовая операция позволяет существенно увеличить число операций умножения с накоплением (MAC), приходящихся на долю одной операции ввода/вывода.

- § Выполнение на проходе функции насыщения над элементами векторов с целью исключения переполнения при арифметических операциях.

- § Выполнение операции произвольной коммутации отдельных элементов в векторе и даже от-

дельных разрядов в любом элементе вектора за один процессорный такт.

Все перечисленные операции выполняются над векторами, представляющими собой 64-разрядные слова, в которых упакованы данные, представленные в дополнителном коде с фиксированной точкой.

- Программная настройка исполнительных узлов для работы с векторами данных, содержащих требуемое количество элементов требуемой разрядности. В общем случае количество элементов в векторе и их разрядность могут принимать любое значение в пределах от 1 до 64. Единственное ограничение заключается в том, что суммарная разрядность всех элементов каждого вектора должна быть равна 64 бит. Данное свойство позволяет даже в пределах одной задачи варьировать соотношение производительность/точность - повышать производительность процессора за счет снижения точности вычислений и, наоборот, повышать точность за счет снижения производительности.

- Многотактный характер векторных команд, которые содержат специальное поле, задающее количество повторений их выполнения (от 1 до 32). Такое решение позволяет аппаратно поддерживать короткие циклы и увеличить плотность программного кода. Например, одной командой можно задавать операцию перемножения двух матриц данных.

Сохраняя программную совместимость с предшественниками и их основные архитектурные особенности, процессор 1879BM4 имеет следующие отличия:

- § Повышена тактовая частота процессора по сравнению с предшественниками за счет увеличения глубины конвейера.

- § Повышена реальная производительность ядра в 3 – 5 раз (в зависимости от класса решаемых задач) по сравнению с предшественниками за счет следующих архитектурных и структурных усовершенствований:

- Обеспечение загрузки матрицы весовых коэффициентов в векторный операционный узел в темпе один вектор за один процессорный такт.

- Обеспечение одновременного выполнения до восьми операций ввода/вывода за один процессорный такт.

- Введение в структуру аппаратной вершины системного стека с целью ускорения процесса возврата из подпрограммы (процедуры обработки прерывания).
- Обеспечение модификации содержимого адресного регистра в первом такте при выполнении многотактных команд ввода/вывода, что позволяет ускорить начало выполнения очередной команды, использующей содержимое данного адресного регистра.
- Введение механизма очередей в конвейер процессора 1879BM4 с целью достижения максимальной производительности при работе с банками синхронной памяти, имеющими различную глубину конвейера.
- Уменьшение аппаратной сложности процессорного ядра за счет использования единого адресного генератора при выполнении различных команд ввода/вывода.

Перечисленные выше технические решения позволяют эффективно использовать процессор 1879BM4 для построения на его основе высокопроизводительных параллельных вычислительных систем, ориентированных на решение задач обработки больших потоков данных в реальном масштабе времени (цифровая обработка сигналов, эмуляция нейронных сетей и т.д.).

II. ОБЩАЯ СТРУКТУРА ПРОЦЕССОРА 1879BM4

Структурная схема процессора представлена на рис. 1. Процессор состоит из следующих функциональных узлов:

PROCESSOR CORE – процессорное ядро, состоящее из двух основных частей: RISC-ядра (**RISC**), выполняющего скалярные арифметические, логические операции и операции сдвига над 32-разрядными данными, а также осуществляющего управление выполнением программ, и матрично-векторного сопроцессора (**VECTOR COPROCESSOR**), осуществляющего арифметические и логические операции над 64-разрядными векторами данных – упакованными в 64-разрядные блоки данных произвольной разрядности. Обмен процессорного ядра с внешним миром осуществляется с помощью шести 64-разрядных шин: команд (**IB**), скалярных данных (**SDB**), векторных входных данных (**VDIB**), весов (**WB**), векторного регистра (**VRB**) и векторных выходных данных (**VDOB**).

AGU (Address Generator Unit) – блок адресных генераторов, который получает запрос на обмен данными с памятью от процессорного ядра (начальный адрес (**ADDR**), смещение (**BIAS**), число обращений) и затем сам формирует адреса, освобождая от этого процессор. Благодаря 6 шинам ядра NMC и наличию в блоке 6 адресных генераторов данных возможно

осуществить до шести операций ввода-вывода данных за один такт. Кроме того, в AGU имеется генератор адреса команд, осуществляющий последовательную предвыборку команд с помощью адресной шины **IAB (Instruction Address Bus)**. В случае, когда происходит переход, по шине **BIAS** процессорное ядро выдаёт адрес перехода, после чего осуществляется предвыборка команд по новому адресу.

ICACHE (Instruction Cache) – кэш команд объёмом $1K \times 64$ разряда. При обращении во внешнюю память выбранные команды по шине **MEM_IB** попадают в кэш и одновременно перетранслируются по шине **IB** в процессорное ядро (**PROCESSOR CORE**). При повторном обращении по тому же адресу команды будут выбираться из кэша по шине **IB**, а не из внешней памяти. Программы, расположенные во внутренней памяти, не кэшируются. Кэш команд в 1879BM4 позволяет эффективно работать с программами, расположенными во внешней памяти, и освободить программиста от дополнительной работы по трансляции программного кода из внешней во внутреннюю память.

LMI и **GMI** – два идентичных 64-разрядных интерфейса во внешнюю память – локальную и глобальную. Каждый интерфейс имеет возможность работать с четырьмя банками внешней памяти типа SRAM/SSRAM/SDRAM и поддерживает режим работы с общей памятью с другим процессором. Процессор использует 32-разрядный адрес, причём обмен данными с памятью осуществляется по 32 или 64 разряда. Таким образом, доступное адресное пространство равно 16 Гбайт. Если старший разряд адреса равен нулю, идёт обращение в локальную память, если равен единице – в глобальную. Младший разряд адреса используется только при обмене 32-разрядными данными.

SRAM (Static Random Access Memory) – четыре одинаковых банков памяти, каждый из которых представляет собой однопортовую статическую память, имеющую организацию $8K \times 64$ бит, и может использоваться для оперативного хранения 32- и 64-разрядных слов данных, коэффициентов и команд. Все банки памяти работают как псевдодвухпортовые с расслоением по младшим адресам памяти, что позволяет одновременно производить доступ к памяти как со стороны процессорного ядра, так и со стороны каналов ПДП. Два банка памяти – **LB0** и **LB1** – принадлежат локальной памяти, два других – **GB0** и **GB1** – глобальной.

CP0 и **CP1** – два байтовых синхронных коммуникационных порта ввода/вывода для межпроцессорного обмена типа точка-точка.

DMA (Direct Memory Access) – блок каналов ПДП, осуществляющих обмен 64-разрядными данными между локальной внешней и локальной внутренней па-

мятью, между глобальной внешней и глобальной внутренней памятью, между коммуникационными портами и любой локальной или глобальной внешней или внутренней памятью.

COMMUTATOR 6->5 – коммутатор шин данных, позволяющий динамически связать одну из шин данных банков внутренней или внешней памяти с одной из шести внутренних: **MEM_IB**, **SDB**, **VDIB**, **WB**, **VRB** и **VDOB**.

PCR (Peripheral Control Register) – регистр управления периферией.

GPIO (General Purpose Input/Output) – восемь портов ввода/вывода общего назначения.

TIMER0 и **TIMER1** – два 32-разрядных таймера.

INTERRUPT – контроллер внешних прерываний.

OCG (Output Clock Generator) – генератор внешнего синхросигнала, который может использоваться для тактирования внешних микросхем или внешних шин самого процессора.

JTAG – интерфейс с 5-выводным тестовым портом, реализованным согласно стандарту IEEE Std 1149.1-1990, который позволяет тестировать процессор 1879BM4 как самостоятельно, так и в составе законченного вычислительного модуля.

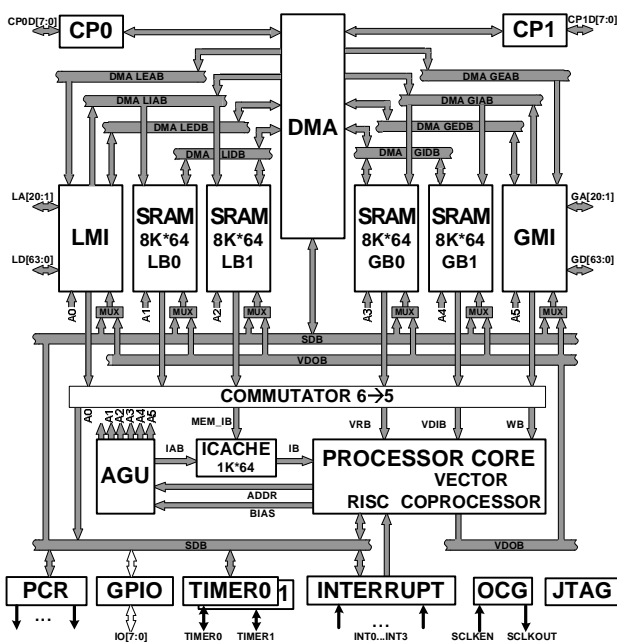


Рис. 1. Структурная схема процессора 1879BM4

III. СИСТЕМА КОМАНД ПРОЦЕССОРА

Команды процессора делятся на две основные группы: скалярные команды (см. рис. 2) и векторные (см. рис. 3).

Скалярные команды являются обычными RISC-командами, которые содержат следующие поля: P – содержит разрешение параллельной работы (можно

или нет выполнить данную команду, не дожидаясь окончания более ранних команд); MOVE – задаёт ввод/вывод данных с одновременной модификацией адресных регистров, условные переходы/переходы к подпрограмме и возвраты из подпрограммы/прерывания; OPER – определяет арифметическую, логическую операцию или операцию сдвига. Скалярные команды могут использовать 32-разрядные константы (поле CONST), которые могут грузиться в регистры или использоваться для задания адреса или смещения при обращении к памяти.

Векторные команды имеют похожие поля: P (разрешение параллельной работы); VMOVE – задаёт ввод/вывод векторных данных, VOPER – определяет арифметическую или логическую операцию над векторными данными. Также имеются следующие поля: W – управляет загрузкой в теневую матрицу весов; L – задаёт перезапись теневой матрицы весов в рабочую; COUNT – определяет число повторов выполнения данной команды (от 1 до 32), что позволяет аппаратно поддерживать организацию коротких циклов и значительно увеличить плотность кода.

Таким образом, процессор использует команды типа VLIW, задающие одновременно операции обмена с памятью, модификацию адресных регистров и арифметическую операцию, причём это относится как к скалярным, так и к векторным командам. Объединение в одной команде операций ввода-вывода и арифметической операции позволяет увеличить производительность скалярных команд на реальных задачах до 40%, но для векторных команд это решение по нашим оценкам особый выигрыш не даёт.

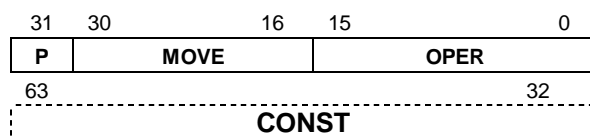


Рис. 2. Кодировка скалярных команд

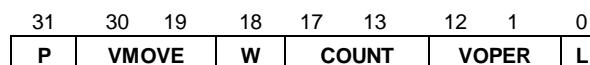


Рис. 3. Кодировка векторных команд

IV. МНОГОАКТОВЫЕ ВЕКТОРНЫЕ КОМАНДЫ И ВЕКТОРНО-КОНВЕЙЕРНАЯ ОРГАНИЗАЦИЯ ВЫЧИСЛЕНИЙ (ДИНАМИЧЕСКИЙ VLIW)

Принцип организации выполнения одновременно нескольких операций, заданных разными векторными командами, в разных функциональных узлах одного или нескольких сопроцессоров показан на рис. 4. Если команды не имеют зависимости по данным и используют разные функциональные узлы, они последовательно занимают соответствующие функциональные

узлы и начинают свою работу. Если одна из команд, например, на рис. 4 это команда 2, должна использовать результат, получаемый предыдущей командой 1, ей не надо ждать полного окончания выполнения последней. Достаточно дожидаться получения первого результата (1.1), чтоб выполнить первую операцию (2.1). Данный механизм называется «зацеплением по данным» (chaining).



Рис. 4. Векторно-конвейерная организация вычислений (динамический VLIW)

Несмотря на то, что команды поступают на выполнение по одной и в строгой последовательности, за счёт использования многотактовых команд достигается эффект суперскаляра (одновременно выполняется несколько операций в разных функциональных узлах над разными данными). Также поддерживается внеочередное выполнение команд (out-of-order execution), когда позже выбранная команда может закончить своё выполнение раньше, чем ранее выбранные команды.

V. ОСОБЕННОСТИ РАБОТЫ КОНВЕЙЕРА КОМАНД ПРИ ОБМЕНЕ ДАННЫХ С ПАМЯТЬЮ

Принципы организации конвейера команд процессора показан на рис. 5. Его основными особенностями являются:

- наличие общих первой и второй ступеней для всех команд – скалярных и векторных, причем на первой ступени осуществляется вычисление адреса первого данного для команды, последнего адреса и смещения (для векторных команд), модификации адресных регистров; на второй организуется единая очередь команд, ожидающих своих данных перед выполнением;
- несколько параллельных подконвейеров на третьей ступени (стадии выполнения операций), причём ввод и вывод данных осуществляется именно на данной ступени (реализуется Late Write). Конвейер для скалярных операций на данной ступени один и имеет глубину единица, конвейеров для векторных операций

несколько, и они имеют переменную глубину в зависимости от типа выполняемой операции.

Наличие очереди команд, ожидающих своих данных, позволяет обеспечить эффективную работу с банками внутренней и внешней памяти, имеющими различную глубину конвейера без потери производительности. Данная очередь имеет глубину восемь, что позволяет эффективно работать с внешней синхронной памятью в конвейерном режиме - имеется возможность выставить до 8-ми запросов на чтение, прежде чем придут первые данные.

Если процессор работает только с внутренней памятью, очередь команд заполнена не полностью. Если произвести хотя бы одно обращение во внешнюю память, то команда, ожидающая своих данных из внешней памяти, не может уйти из очереди, и очередь начинает заполняться. Как только данные придут, очередь начинает разгружаться. Тем самым реализуется конвейер переменной глубины в зависимости от требуемого числа тактов обращения в память, и это даёт возможность эффективно работать одновременно как с внешней, так и внутренней памятью.

Реализация выдачи и приёма данных на одной и той же ступени конвейера (Late Write) резко снижает количество конфликтов по данным. Так, если первая команда читает данные из памяти, а вторая должна записать эти данные в память, последняя может выполняться в следующем такте после выполнения первой. В случае, если это векторные команды, вторая команда может начать выполняться в следующем такте после получения первого данного первой командой.

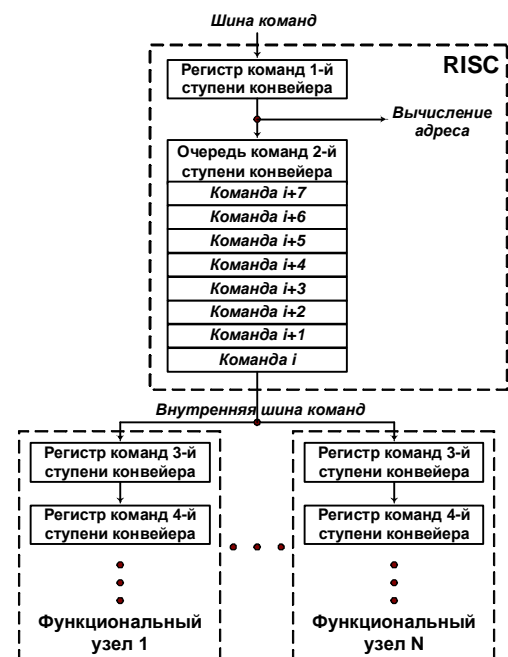


Рис. 5. Принципы организации конвейера команд процессора 1879BM4

VI. БАЗОВЫЕ ОПЕРАЦИИ ОБРАБОТКИ ДАННЫХ

Архитектура NeuroMatrix® даёт уникальную возможность варьировать между производительностью и точностью вычислений для базовой процедуры:

$$Y_m = U_m + \sum_{n=1}^N X_n \times W_{n,m}$$

В зависимости от приложения можно выбрать необходимую разрядность входных данных и результата (точность вычислений). Число умножений и сложений (MAC), выполняемых за один такт, зависит от разрядности операндов. Наибольшая производительность – 224 MAC – достигается при работе с 2-разрядными операндами. Имеется возможность поднять точность вычислений, если увеличить разрядность операндов до 32-х. В этом случае достигается производительность 2 MAC с получением 64-разрядного результата. Векторный сопроцессор содержит операционное устройство регулярной структуры, похожее на матричный умножитель (см. рис. 6).

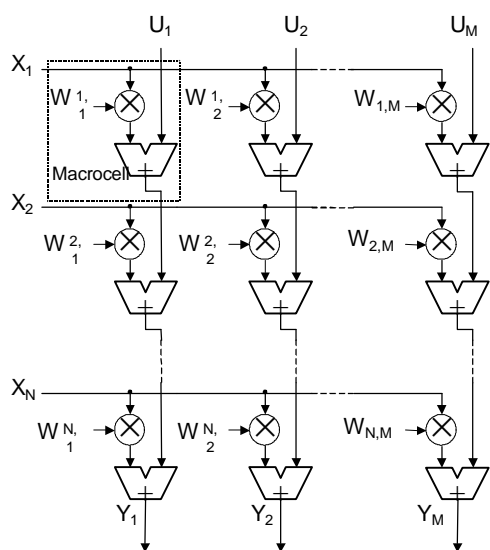


Рис. 6. Операционное устройство векторного сопроцессора

VII. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ПРОЦЕССОРА

- Формат обрабатываемых данных – 32-разрядные скалярные данные, а также векторы данных программируемой разрядности от 1 до 64 разрядов, упакованные в 64-разрядные слова.
- Команды размером 32-разряда и 64-разряда. Каждая команда может задавать несколько различных операций.

- Адресуемое пространство – 4×2^{30} 32-разрядных слов (16 Гбайт).
- Аппаратная поддержка операций умножения вектора на матрицу и матрицы на матрицу.
- Аппаратная поддержка функции насыщения.
- Единый поток команд для задания векторных и скалярных операций, операций ввода/вывода.
- Многотактные векторные команды (возможность одновременного выполнения до 5-ти векторных команд).
- Одновременное выполнение до 8-ми операций ввода/вывода за один процессорный такт.
- Производительность (количество операций «умножение с накоплением», выполняемых за один такт)
 - 2 MAC для 32-разрядных данных,
 - 4 MAC для 16-разрядных данных,
 - 24 MAC для 8-разрядных данных,
 - 80 MAC для 4-разрядных данных,
 - 224 MAC для 2-разрядных данных.
- Четыре двухпортовых банка внутренней памяти объёмом 8К*64 бит (16К*32 бит) каждый.
- Кэш-память команд объёмом 1К*64 бит.
- Две внешних 64-разрядных шины с рабочей частотой до 125 МГц и возможностью работы в многопроцессорном режиме.
- Два синхронных байтовых коммуникационных порта с пропускной способностью до 150 Мбайт/с каждый.
- JTAG-порт для отладки и тестирования.
- Блок ПДП, обеспечивающий обмен данными между внешней и внутренней памятью, а также между коммуникационными портами и внешней/внутренней памятью.
- Восемь битовых портов ввода/вывода общего назначения.
- Технология изготовления – 0.25 мкм КМОП.
- Тактовая частота – до 150 МГц.

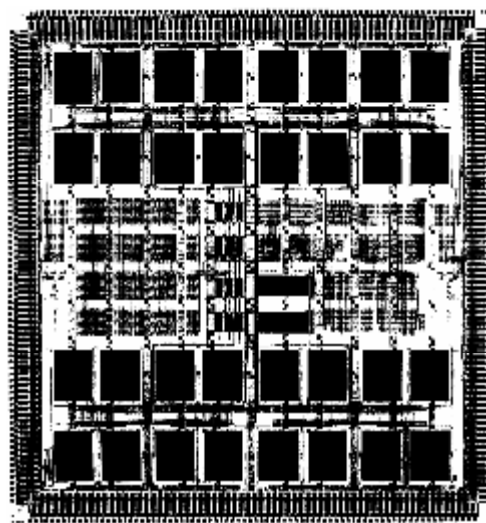


Рис. 7. Топология процессора 1879BM4

Наличие у процессора 1879BM4 двух 64-разрядных внешних шин с встроенной поддержкой многопроцессорного режима и двух высокопроизводительных коммуникационных портов позволяет на его базе строить многопроцессорные вычислительные комплексы без больших аппаратных затрат.

Топология процессора приведена на рис. 7.

VIII. ОБЛАСТИ ПРИМЕНЕНИЯ ПРОЦЕССОРА

Отечественные процессоры семейства NeuroMatrix являются представителями нового класса векторно-конвейерных DSP. Их отличает высокая производительность на задачах обработки больших потоков данных при относительно небольших аппаратных затратах и малом потреблении питания.

Благодаря описанным выше архитектурным и структурным усовершенствованиям реальная производительность нового процессора 1879BM4 по сравнению с L1879BM1 и 1879BM2 должна повыситься соответственно в 3 – 12 и 1,5 – 6 раз в зависимости от класса решаемых задач. В качестве примера в табл. 1 приведена реальная производительность процессоров семейства NeuroMatrix при решении задачи преобразования Уолша-Адамара 1024 точки (одновременно обрабатываются четыре вектора, каждый из которых содержит 1024 16-разрядных элемента).

Процессор 1879BM4 может применяться в качестве цифрового процессора сигналов в одно- и многопроцессорном режиме. Также он может быть применен в качестве сопроцессора для векторно-матричных вычислений в комплексных вычислительных системах под управлением ведущего процессора общего назначения или микроконтроллера.

Таблица 1

Сравнительная производительность процессоров семейства NeuroMatrix при решении задачи преобразования Уолша-Адамара 1024 точки (одновременно обрабатываются четыре вектора)

Процессор	L1879BM1 (NM6403)	1879BM2 (NM6404)	1879BM4 (NM6405)
Тактовая частота	40 МГц	80 МГц	150 МГц
Технология изготовления	КМОП 0.5 мкм	КМОП 0.25 мкм	КМОП 0.25 мкм
Число тактов	4.860	4.572	2.513
Время выполнения	121,5 мкс	57,15 мкс	16,74 мкс

Характерными областями применения процессора 1879BM4 являются:

- Обработка изображений, включая различные виды фильтрации и MPEG кодирование и декодирование.
- Обработка радиолокационных сигналов, в том числе, различные виды цифровой фильтрации, преобразование Фурье, Адамара и прочее.
- Высокопроизводительная коммутация сигналов.

IX. ПЕРСПЕКТИВЫ РАЗВИТИЯ ПРОЦЕССОРОВ СЕМЕЙСТВА NEUROMATRIX

Дальнейшее развитие процессоров семейства NeuroMatrix связано с завершением разработки и запуском на изготовление в 2011 году процессора четвертого поколения 1879BM5 (NM6406). Он будет реализован по КМОП технологии с топологическими нормами проектирования 90 нм, в нём используется то же процессорное ядро, что и в 1879BM4, но работающее уже на частотах до 300 МГц. Кроме того, объём внутренней памяти увеличивается в 2 раза (4 Мбит против 2 Мбит у 1879BM2 и 1879BM4).

ЛИТЕРАТУРА

- [1] Мушкаев С.В. Реализация ранжирующих и медианных фильтров на процессоре NM6403 (L1879BM1) // Цифровая обработка сигналов, - 2005. - № 1. - С. 52-56.
- [2] Мушкаев С.В., Ландышев С.В. Применение процессора L1879BM1 для сжатия изображений // Цифровая обработка сигналов. - 2002. - № 1. - С. 12-18.
- [3] Кашкаров В.А., Мушкаев С.В. Организация параллельных вычислений в алгоритмах БПФ на процессоре NM6403 // Цифровая обработка сигналов. - 2001. - № 1. - С. 53-58.
- [4] Борисов Ю.И., Комплекс "Трафик-Монитор" на базе процессора L1879BM1. Особенности разработки // Электроника: НТБ. - 2003. - №6. - С. 58-61.
- [5] Черников В.М., Вискне П.Е., Шелухин А.М., Шевченко П.А., Панфилов А.П., Косоруков Д.Е., Черников А.В. Семейство процессоров обработки сигналов с векторно-матричной архитектурой NeuroMatrix // Электронные компоненты. - 2006. - № 6. - С. 79-84.
- [6] Черников В.М., Вискне П.Е., Шелухин А.М., Панфилов А.П. Отечественные высокопроизводительные процессоры цифровой обработки сигналов векторно-матричной архитектуры, перспективы развития // Материалы конференции «Перспективы развития высокопроизводительных архитектур. История, современность и будущее отечественного компьютеростроения». Сборник научных трудов ИТМиВТ. - М.: ИТМиВТ им С.А.Лебедева РАН. - 2008. - вып. №1. - С. 52-59.