

УСКОРИТЕЛЬ ЧИСЛЕННЫХ РАСЧЕТОВ НА БАЗЕ МОДУЛЯРНО-СИСТОЛИЧЕСКОГО МИКРОПРОЦЕССОРА

И.П. Осинин

ФГУП «РФЯЦ-ВНИИЭФ», г. Саров Нижегородской обл.

Введение

Прогресс в области проектирования и производства вычислительной техники связан с повышением ее быстродействия. Физический предел повышения тактовой частоты стимулирует поиск принципиально новых подходов для решения этой задачи. Параллельные системы обработки информации являются эволюционным шагом в повышении быстродействия устройств вычислительной техники.

Однако отсутствие «параллельной» математики, сложность программной и аппаратной реализации таких систем существенно ограничивают их применение [1]. Одним из подходов к преодолению этих трудностей является использование модулярной арифметики (МА) и непозиционной системы счисления – системы остаточных классов (СОК). Её основной особенностью является отсутствие переносов при выполнении операций над остатками операндов, что обеспечивает существенное повышение быстродействия без перехода на новые технологические нормы.

При этом остатки являются малоразрядными (7-8 бит) позиционными числами, поэтому для предельного распараллеливания вычислений предлагается использование концепции однородной вычислительной среды (ОВС). ОВС представляет собой набор однотипных ячеек, объединенных регулярными связями. Её основными достоинствами [2] являются:

- пространственно-временной параллелизм (систолическое выполнение операций);
- простота масштабирования при изменении разрядности операндов [3].

Далее представлено описание основных технических решений, применённых в разработанном модулярно-систолическом процессоре (МСП).

1 Технические решения модулярно-систолического процессора

Прототип МСП представляет собой микропроцессор с четырьмя ядрами, выполняющими операции, как в позиционной системе счисления (ПСС), так и в системе остаточных классов. Он имеет МIMD-архитектуру с распределенной кэш-памятью и ориентирован на массовую арифметико-логическую обработку чисел с фиксированной точкой. Под массовой обработкой понимается высокая скорость и большой объем входного потока данных.

МСП обладает уникальной микроархитектурой, которая позволяет динамически изменять разрядность обрабатываемых операндов, что бывает необходимо для повышения точности в процессе вычислений. Например, каждое из четырех процессорных ядер может вести независимую обработку 32-разрядных операндов, либо они могут группироваться парами для обработки 64-разрядных чисел, либо все ядра могут объединиться для обработки 128-разрядных операндов.

Даная группировка возможна как для ПСС, так и для СОК, причем, она не требует временных затрат, так как реконфигурация процессора происходит динамически и может выполняться с приходом каждой следующей пары операндов. Подобное техническое решение отсутствует как в современных универсальных микропроцессорах, так и в сопроцессорах.

Другой отличительной особенностью микропроцессора является распараллеливание обработки каждой пары операндов на двух уровнях.

Первый уровень обеспечивается независимостью обработки по каждому основанию числа в СОК. Так, например, время вычисления 128-разрядного числа сокращается в 16 раз за счет одновременного вычисления по 20 восьмиразрядным остаткам исходных чисел (модулярность структуры).

Второй уровень обеспечивается пространственно-временным параллелизмом вычислений по каждому основанию на базе ОВС (систоличность структуры).

Таким образом, параллельно-конвейерное выполнение операций над независимыми остатками чисел позволяет сократить время вычислений до времени срабатывания одного логического элемента. Например, вычисления над числами разрядности 128, 64 и 32 бита выполняются одинаково быстро и на предельной частоте работы, что недостижимо в известных аналогах, функционирующих в ПСС. В них высокая скорость работы обеспечивается наращиванием количества ядер, что в итоге сильно сужает класс задач, так как необходимо избегать зависимостей по данным. Для обеспечения своевременной подачи исходных операндов на входы арифметико-логического устройства предусмотрено повышение скорости работы кэш-памяти за счет 96-кратного расслоения доступа за счет частично-ассоциативного доступа и одновременного доступа в распределенные банки памяти, которые соответствуют независимым основаниям СОК.

Основные технические характеристики МСП представлены в таблице 1.

Таблица 1 - Основные технические характеристики МСП

Теоретическая пиковая производительность, Гоп/с	2
Количество вычислительных ядер, шт.	4
Рабочая тактовая частота, ГГц	0,5
Поддерживаемый объём оперативной памяти, Гбайт	32
Количество видов адресации, шт.	4
Размер кэш-памяти одного ядра, кбайт	80
Уровней вложенности прерываний, шт.	16
Разрядность операндов в ПСС, бит	32/64/128
Разрядность операндов в СОК, бит	40/80/160
Разрядность команды, бит	64
Ширина выборки из оперативной памяти, бит	64
Потребляемая мощность, Вт	2

Представление разрядной сетки для чисел в ПСС стандартно: старший бит определяет знак числа, остальные биты содержат само число. Для кодирования чисел в СОК необходимо большее число двоичных разрядов, чем в ПСС, поэтому 32-,64-,128-разрядным сеткам позиционных чисел соответствуют 40-,80-,160-разрядные сетки чисел в СОК. Пример 40-разрядной сетки приведен на рисунке 1. Она содержит остатки числа по пяти различным основаниям, разрядность которых не превышает восьми бит. При этом два младших бита разрядной сетки определяют группу оснований. Сочетания различных групп оснований позволяют кодировать числа большей разрядности. Таким образом, для представления 128-разрядного числа в СОК потребуется задействовать 160-разрядную сетку, содержащую двадцать остатков, входящих в четыре различные группы оснований.

39	32	31	24	23	16	15	9	8	2	1	0
Остаток по модулю №5	Остаток по модулю №4	Остаток по модулю №3	Остаток по модулю №2	Остаток по модулю №1	Группа оснований						

Рисунок 1 – Остаточное представление 32-х битного числа в процессоре

В системе команд процессора используется универсальный формат всех микроопераций. В 64-битном формате команды (рисунок 2) присутствуют следующие поля: смещения

операндов (C1-C3), вид адресации (ВА), код операции (КОП), указатель векторной команды (В).

63	48	47	32	31	16	15	10	9	3	2	1	0
C3			C2		C1		ВА		КОП		В	*

Рисунок 2 – Универсальный формат команд

В процессоре используются основные типы адресации (таблица 2), при этом задействовано 80 команд, среди которых арифметико-логические команды, передачи управления и другие (таблица 3).

Таблица 2 – Способы адресации

№	Способ адресации	Операнд
1	Непосредственная	Операнд в команде
2	Регистровая прямая	Операнд в регистре константы
3	Индексная автоинкрементная	Смещение в адресном счетчике
4	Относительная прямая	Смещение в команде

Таблица 3 – Общие сведения о командах

Тип команды	Адресность	Количество команд
Команды пересылки данных и ввода/ вывода	2	3
Команды арифметико-логические	3	13
	2	17
Команды управления	1	2
	0	11
Команды передачи управления	1	18
	0	3
Привилегированные команды	1	8
Команды работы с аппаратурой прерываний	0	5
Всего команд	-	80

При этом используется в процессоре иерархическая структура памяти. Существует как предвыборка команд в очередь команд из оперативной памяти, так и предвыборка операндов в очереди данных (ОД) из КЭШ-памяти для повышения скорости их обработки. После выполнения операции в арифметико-логическом устройстве (АЛУ), её результат поступает в очередь результатов, откуда записываются в КЭШ память по адресу, извлеченному из очереди адресов результатов (ОАР). Подобные очереди представляют собой высокоскоростные буферы FIFO.

Обмен информацией между ядрами, оперативной памятью и внешними устройствами производится по системной шине. Структурная схема процессора представлена на рисунке 3.

Так как процессор работает не только в ПСС, но и в СОК, то он содержит аппаратный быстродействующий преобразователь кодов, который в параллельно-конвейерном режиме преобразуют числа из ПСС в СОК и параллельно с этим из СОК в ПСС.

Обработка информации в процессоре конвейеризирована на двух уровнях для оптимального распределения частот синхронизации, соответствующих узлам процессора:

- микроконвейер арифметико-логических преобразований операндов (предельное распараллеливание позволяет установить частоту синхронизации равную времени срабатывания элемента ИСКЛЮЧАЮЩЕЕ ИЛИ);
- макроконвейер стадий подготовки операндов и записи результата (частота синхронизации определяется временем срабатывания самой медленной ступени макроконвейера).



Рисунок 3 - Структурная схема процессора

В связи с этим устройство управления процессора децентрализовано. Каждое ядро содержит три устройства управления (УУ), все они отвечает за соответствующие ступени работы макроконвейера. Причем, микроконвейер АЛУ не нуждается в собственном УУ.

При этом АЛУ каждого ядра может выполнять операции, как в СОК, так и в ПСС, благодаря универсальному вычислительному полю ОВС, задача которого – суммирование пар операндов в параллельно-конвейерном режиме. В связи с этим все арифметические операции сводятся к суммированию в дополнительном коде. При этом операции умножение и деление выполняются итерационно.

2 Вычислительное ядро

Условимся называть часть АЛУ, выполняющую только арифметические операции, вычислительным ядром.

Свойства модулярной арифметики позволяют свести выполнение таких операций как суммирование, вычитание, умножение нацело и деление к операции суммирования остатков. Для этого используются таблицы подстановок (ТП), представляющие собой ассоциативную память, которая выполняет однократную табличную выборку. Общая структурная схема вычислительного ядра [4] в общем виде представлена на рисунке 4.

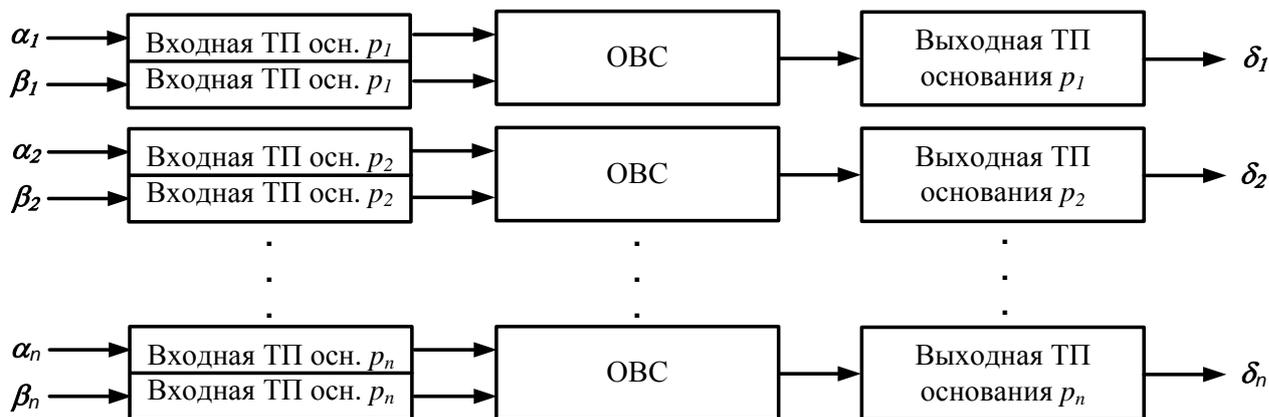


Рисунок 4 – Структурная схема вычислительного ядра

Вычислительное ядро состоит из вычислительных модулей, число которых равно n , где n – число оснований СОК. Каждый вычислительный модуль ведет вычисления по заранее определенному основанию и содержит две входные таблицы подстановок, выходную таблицу подстановок и ОВС.

ОВС имеет размерность $m+1$ строк и $m+1$ столбцов базовых элементов (БЭ) и предназначена для суммирования операндов с распараллеливанием конвейерного типа вплоть до разряда операнда, где m – разрядность остатка p_i . Структурная схема ОВС представлена на рисунке 5.

Базовый элемент, осуществляющий суммирование бита операнда с битом переноса, построен на базе полусумматора, содержащего логический элемент И и логический элемент ИСКЛЮЧАЮЩЕЕ ИЛИ, буферные триггерах Т1 и Т2 для фиксации текущего состояния бита операнда и бита переноса в старший разряд (рисунок 5).

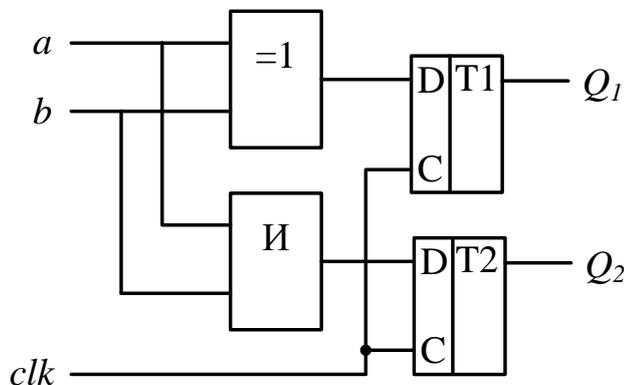


Рисунок 5 – Структурная схема базового элемента

Таким образом, время одного такта синхронизации вычислительного ядра $t_{овс}$ сводится к времени срабатывания БЭ. При этом время заполнения конвейера вычислительного ядра T составит:

$$T = t_{вх.мн} + (m+1) \cdot t_{я} + t_{вых.мн},$$

где $t_{вх.мн}$ – время задержки входной таблицы подстановок, $t_{вых.мн}$ – время задержки выходной таблицы подстановок, $t_{я}$ – время задержки ячейки ОВС, m – разрядность остатка.

В случае конвейерного режима работы результат операции будет выдаваться каждый такт работы устройства, независимо от типа арифметической операции.

Временная сложность Q_{n1} вычислительного ядра, выполняющего операции сложение и вычитание в ПСС, определяется произведением времени такта работы ОВС $T_{овс}$ на количество разрядов операндов n . Причем, в конвейерном режиме возможно совмещение во времени обработки $m+1$ пары операндов, то есть

$$Q_{n1} = \frac{T_{овс} \cdot n}{m+1},$$

где m – разрядность оснований СОК.

Так как операции умножение и деление в ПСС сводятся к суммированию в дополнительном коде, для выполнения каждой требуется n итераций, в этом случае временная сложность Q_{n2} составит

$$Q_{n2} = \frac{T_{овс} \cdot n \cdot n}{m+1} = \frac{T_{овс} \cdot n^2}{m+1}.$$

Временная сложность Q_c вычислительного ядра, выполняющего операции в СОК, одинакова для всех арифметических операций и в конвейерном режиме на каждую операцию необходим один такт работы ОВС, то есть

$$Q_c = T_{овс}.$$

На рисунке 6 представлен график зависимости временной сложности работы ОВС от разрядности операндов в предлагаемом процессоре, где разрядность оснований $m=8$.

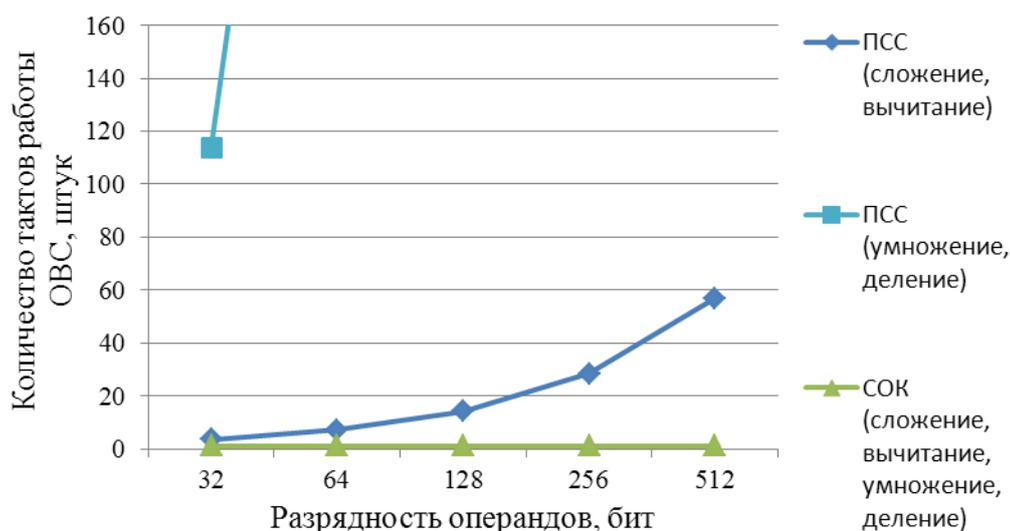


Рисунок 6 – График зависимости временной сложности от разрядности операндов

Таким образом, полученное выражение временной сложности устройства показывает пропорциональную зависимость от разрядности операндов для операций сложение и вычитание, выполняемых в ПСС. Для операций умножение и деление эта зависимость является квадратичной.

При выполнении операций в СОК временная сложность не зависит ни от типа операции, ни от разрядности операндов. Так, например, скорость суммирования двух 64-разрядных чисел в предлагаемом процессоре 7,11 раз выше при выполнении операции в СОК по сравнению с ПСС. Аналогично, скорость по сравнению с традиционным способом умножения двух 64-разрядных чисел в 455,11 раз выше при выполнении операции в СОК [5].

Так как вычислительное поле ОВС универсально для выполнения операций, как в СОК, так и в ПСС, то аппаратная сложность R ОВС определяется произведением размерности ОВС (n строк и $m+1$ столбцов, где n – разрядность операнда, m – разрядность основания p_i), то есть

количеством БЭ, на количество логических элементов в каждом БЭ (14 элементов при представлении в базе И-НЕ), то есть

$$R = 14 \cdot n \cdot (m + 1).$$

На рисунке 7 представлен график зависимости аппаратной сложности работы ОВС от разрядности операндов в предлагаемом процессоре.



Рисунок 7 – График зависимости аппаратной сложности от разрядности операндов

Аппаратная сложность ОВС одинакова для ПСС и СОК и прямо пропорциональна разрядности операндов.

В результате оценки эффективности предлагаемого арифметического устройства установлено, что при одинаковых аппаратных затратах скорость вычислений в СОК после заполнения $m+1$ ступеней конвейера будет в $n/(m+1)$ раз выше для операций сложение и вычитание и в $n^2/(m+1)$ раз выше для операций умножение и деление нацело по сравнению с ПСС, где n – разрядность операндов, m – разрядность оснований СОК.

3 Прототип модулярно-систолического процессора

Прототип МСП реализован в качестве прошивки ПЛИС 4CGX15N семейства Cyclone IV фирмы Altera. Данная ПЛИС входит в состав отладочной платы DK-START-4CGX15N, внешний вид которой приведен на рисунке 8. Ввиду относительно небольшого количества логических элементов (14400 штук) была запрограммирована лишь часть модулярно-систолического процессора, отвечающая за арифметические команды, а также за преобразование операндов из ПСС в СОК и обратно.



Рисунок 8 – Внешний вид отладочной платы DK-START-4CGX15N

Исходные данные и команды загружаются в флеш-память отладочной платы с помощью программного средства (рисунок 9).

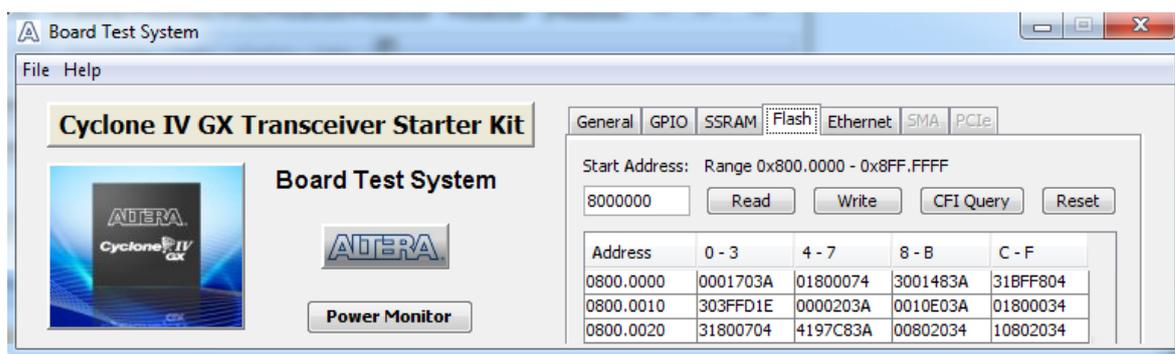


Рисунок 9 – Программное средство для работы с отладочной платой

Прототип МСП преобразует исходные данные в СОК, выполняет требуемые операции над ними и преобразует обратно в ПСС, после чего результат записывается в память отладочной платы. Таким образом, прототип МСП подтверждает работоспособность предлагаемых технических решений.

Заключение

Разработанный модулярно-систолический четырехядерный микропроцессор имеет MIMD-архитектуру с распределенной кэш-памятью и ориентирован на массовую арифметико-логическую обработку чисел с фиксированной точкой.

Ему присущ ряд уникальных особенностей, например, динамическая группировка процессорных ядер для наращивания разрядности обрабатываемых чисел. При этом время выполнения операций неизменно, а сама обработка предельно распараллелена, что достигается благодаря выполнению операций в параллельно-конвейерном режиме над непоозиционными числами.

Прототип МСП реализован на базе ПЛИС. Пиковая производительность составляет 2 Гоп/с при энергопотреблении 2 Вт. При этом стоит учитывать, что скорость выполнения вычислительных операций ограничивается максимальной тактовой частотой кристалла, а количество ядер – его площадью, поэтому при реализации МСП на СБИС ожидается повышение производительности, как за счет увеличения тактовой частоты (вплоть до 4 ГГц), так и за счет количества ядер (до 16 штук). В этом случае расчетная пиковая производительность составляет 64 Гоп/с.

МСП ориентирован на высокопроизводительные вычисления в виде сопроцессора, либо в качестве универсального процессора. При этом следует подчеркнуть, что создание отечественного микропроцессора является важнейшей стратегической задачей государства.

Список литературы

1. Барский А. Б. Параллельные процессы в вычислительных системах. – М.: Радио и связь, 1990. – 255 с.
2. Акушский, И.Я. Машинная арифметика в остаточных классах / Акушский И.Я., Юдицкий Д.М. — М.: Советское радио, 1968. 440с.
3. Модулярные параллельные вычислительные структуры нейропроцессорных систем / Под ред. Н. И. Червякова ; Ред. Н. И. Червяков. - М. : Физматлит, 2003. - 288 с.

4. Осинин И. П., Князьков В. С. Однородная модулярно-систолическая структура для массовых арифметических вычислений // IV международная научная конференция «Суперкомпьютерные системы и их применение (SSA-2012)»: сборник материалов. – Минск: ОИПИ НАН Беларуси, 2012. – С. 122-126
5. Осинин И. П., Князьков В. С. Однородный арифметический систолическо-модулярный ПЛИС-процессор с реконфигурируемой микроархитектурой // Всероссийская конференция «Проведение научных исследований в области обработки, хранения, передачи и защиты информации»: сборник тезисов докладов. – М.: МСП ИТТ, 2011. – С. 121-122.