

ЛЫСАКОВ КОНСТАНТИН ФЕДОРОВИЧ

**ИССЛЕДОВАНИЕ МЕТОДОВ РЕАЛИЗАЦИИ
АЛГОРИТМОВ ОБРАБОТКИ БОЛЬШИХ ПОТОКОВ ДАННЫХ
ЗА СЧЕТ КОНВЕЙЕРНОГО РАСПАРАЛЛЕЛИВАНИЯ**

05.13.18 «Математическое моделирование,
численные методы и комплексы программ»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Новосибирск – 2009

Работа выполнена в Учреждении Российской академии наук
Институте автоматики и электрометрии
Сибирского отделения Российской академии наук

Научный руководитель доктор физико-математических наук
Лаврентьев Михаил Михайлович

Официальные оппоненты: доктор технических наук
Золотухин Юрий Николаевич

кандидат технических наук
Коваленко Юрий Васильевич

Ведущая организация Московский государственный
технический университет
имени Н.Э. Баумана

Защита диссертации состоится «___» _____ 2010 г. в _____ час. на заседании диссертационного совета Д 003.005.01 при Институте автоматики и электрометрии СО РАН по адресу: Россия, 630090, Новосибирск, проспект Академика Коптюга, 1.

С диссертацией можно ознакомиться в библиотеке ИАиЭ СО РАН.

Автореферат разослан «__» _____ 2010 г.

Ученый секретарь диссертационного совета
д.ф.-м.н.

Насыров К.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Развитие современных научных и производственных технологий приводит к стремительному росту объемов информации, которую необходимо оперативно обрабатывать для получения результатов с минимальными временными задержками. К числу таких задач можно отнести оперативную обработку изображений в режиме поступления данных, обработку мультимедийных видеопотоков в телевидении высокой четкости (HD) и многие другие. Все возрастающими потоками данных характеризуются такие области, как геофизика (например, обработка данных скважинных измерений или сейсмического зондирования) и биоинформатика (анализ геномных последовательностей). В скважинной геофизике важно провести обработку в сжатые сроки в полевых условиях, когда высокопроизводительные многопроцессорные системы не доступны. Существующие методы реализации алгоритмов обработки данных с использованием универсальных и сигнальных процессоров зачастую не способны обеспечить обработку со скоростью поступления данных.

Во многих алгоритмах обработки (например, алгоритмы полного перебора) используются массивные однотипные вычисления. Такие алгоритмы могут исполняться в нескольких независимых потоках, что позволит уменьшить общее время обработки. Процесс распараллеливания на несколько вычислительных ядер принято называть построением широкого вычислительного конвейера, состоящего из нескольких однотипных ветвей исполнения. В качестве вычислительных систем для параллельных вычислений сегодня используются либо многоядерные и многопроцессорные персональные компьютеры, либо кластерные системы на основе универсальных процессоров.

При реализации различных алгоритмов обработки данных часто для достижения результата возникает необходимость выполнения нескольких этапов действий. Например, реализация функции вида (AX^5+B) распадается на 3 этапа: возведение в степень, умножение на коэффициент и добавление константы. Каждый этап может требовать от одного до десятков тактов процессора. Для увеличения производительности в процессорных устройствах применяются методы построения «глубоких конвейеров», в которых на каждом такте на вход обработчика подаются новые данные, а все промежуточные данные хранятся внутри конвейера. Такой способ позволяет существенно ускорить обработку данных, обеспечивая появление результатов с частотой поступления данных. Задержка, необходимая для получения результата, называется глубиной конвейера. В современных универсальных процессорах глубина конвейера составляет десятки стадий (для Pentium 4 на ядрах Prescott – 31). Однако жестко заданная глубина конвейера не позволяет инструкциям, требующим для исполнения лишь несколько тактов, исполняться быстрее, чем глубина конвейера, что приводит к уменьшению производительности при решении ряда задач.

Современные микросхемы программируемой логики FPGA (Field-Programmable Gate Array) обеспечивают параллельное исполнение до сотен тысяч одновременных потоков, при этом объем внутренней памяти достигает десятков Мбит. FPGA являются программно-конфигурируемыми вычислителями, то есть связи между вычислительными примитивами и внутренней памятью задаются программистом. Такая система делает возможным построение вычислительной архитектуры, максимально соответствующей реализуемым алгоритмам. При этом микросхемы FPGA можно неограниченное количество раз перепрограммировать, что позволяет использовать одно аппаратное устройство для решения различных задач.

В настоящее время существует два основных направления по разработке и созданию вычислителей на базе FPGA:

- высокопроизводительные вычислительные системы, состоящие из каскадируемых устройств, каждое из которых включает в себя от 5 до 20 FPGA;
- системы контроля и принятия решений, обладающие сравнительно невысокой производительностью. Для увеличения производительности решений используются сигнальные процессоры.

Вышеизложенное позволяет сделать вывод об актуальности работы по реализации алгоритмов обработки больших потоков данных на вычислительных системах, допускающих конвейерное распараллеливание исполнения, и созданию программно-аппаратных комплексов на базе FPGA для высокопроизводительной обработки потоковых данных.

Целью работы являются исследование особенностей применения FPGA в задачах потоковой обработки данных для повышения производительности за счет конвейерного распараллеливания и создание программно-аппаратного комплекса на базе FPGA, обеспечивающего реализацию алгоритмов обработки потоков данных до 10 Гбит/с.

Для достижения данной цели необходимо решить следующие *задачи*:

1. Исследовать особенности реализации задач потоковой обработки данных на примере фильтрации изображений и поиска объектов с использованием метода наименьших квадратов.
2. Разработать метод реализации задачи обработки последовательностей изображений, включающей фильтрацию, компенсацию целочисленного сдвига, компенсацию фона, выявление экстремумов и пороговая обработка; путем моделирования на ПК оценить эффективность предлагаемого метода реализации.
3. Разработать программно-аппаратную архитектуру вычислительных комплексов на базе FPGA для решения задач высокопроизводительной обработки больших потоков данных.
4. Создать макет программно-аппаратного комплекса для решения практических задач с потоками данных порядка 10 Гбит/с.

5. На базе созданного программно-аппаратного комплекса исследовать эффективность реализации задач обработки потока видеоданных 6 Гбит/с и поиска мотивов в нуклеотидных последовательностях генома.

Научная новизна

1. Разработана программно-аппаратная архитектура вычислительных комплексов на базе FPGA для обработки потоков данных порядка 10 Гбит/с, позволяющая оперировать данными со скоростью их поступления за счет оптимизации операций с памятью, организации программных модулей и создания специального программного обеспечения.
2. Предложен метод реализации задачи обработки последовательностей изображений, включающей такие разнородные операции как фильтрация, компенсация целочисленного сдвига, компенсация фона, выявление экстремумов и пороговая обработка.
3. Предложен метод реализации задачи поиска транскрипционных факторов в регуляторных выборках генома, обеспечивающий производительность $1,67 \cdot 10^{13}$ операций целочисленного сравнения в секунду, позволяющий уменьшить время решения задачи в 20 000 раз по сравнению с использованием стандартного ПК.

Практическая ценность

1. Разработан программно-аппаратный комплекс, позволяющий моделировать работу алгоритмов, реализованных на языках описания аппаратуры (HDL – Hardware Description Language), для их тестирования и выявления факторов, ограничивающих производительность.
2. Создан макет бортового спецвычислителя для обработки последовательностей изображений и поиска малоразмерных объектов, способный в режиме поступления обрабатывать поток данных 1,5 Гбит/с, что в 50 раз превышает возможности существующего решения на сигнальном процессоре ADSP21060.
3. Создан программно-аппаратный комплекс для одновременной обработки семи видеопотоков формата HD (1,5 Гбит/с), что в сумме составляет около 10 Гбит/с.

Основные положения, выносимые на защиту

1. Программно-аппаратная архитектура устройств на базе FPGA, основанная на разделении функциональных программных модулей, позволяет обрабатывать в режиме поступления потоки данных порядка 10 Гбит/с и реализовывать алгоритмы перебора с производительностью до $3 \cdot 10^{13}$ целочисленных операций в секунду, обеспечивая решение задач в различных областях: от обработки потоковых видеоданных до задач биоинформатики.
2. Реализация на базе FPGA задачи обработки последовательностей изображений, включающей фильтрацию, компенсацию целочисленного сдвига, компенсацию фона, выявление экстремумов и пороговую обработку, сокращает время выполнения до 50 раз по сравнению с реализацией на базе сигнального

процессора за счет организации вычислительных конвейеров и распараллеливания их исполнения.

3. Алгоритм поиска транскрипционных факторов в регуляторных выборках, основанный на применении таблиц истинности в программируемой логике и использовании встроенной памяти, позволяет достичь производительность $1,67 \cdot 10^{13}$ операций сравнения в секунду, что недостижимо для вычислительных систем на основе универсальных процессоров или графических ускорителей.
4. Программно-аппаратный комплекс на базе FPGA за счет создания специального программного обеспечения дает возможность обрабатывать потоки данных порядка 10 Гбит/с на стандартном ПК.

Личный вклад автора. Выносимые на защиту результаты получены соискателем лично. В опубликованных работах участие автора заключалось в проведении исследовательских работ, реализации алгоритмов и тестировании. Постановка задач и выбор конкретного метода решения осуществлялись коллективом исполнителей при непосредственном участии соискателя.

Внедрение полученных результатов. Результаты работы использованы при разработке и создании макета бортового спецвычислителя на базе FPGA, предназначенного для обработки последовательностей изображений в режиме реального времени, применяющегося в ФГУП «ЦНИИ «Комета», г. Москва.

Результаты работы использованы при разработке и создании программно-аппаратного комплекса на базе FPGA для обработки четырех видеопотоков HD-SDI, используемого в составе виртуальной студии «Фокус» производства ЗАО «СофтЛаб-НСК».

Апробация работы. Результаты диссертации докладывались международных, всероссийских и региональных научных конференциях, в том числе:

- International Conference on Pattern Recognition and Image Analysis: New Information technologies (PRIA). St. Petersburg, 2004; Yoshkar-Ola, 2007.
- IASTED International Multi-Conference AUTOMATION, CONTROL, AND APPLICATIONS (ACIT-ACA). Novosibirsk, 2005.
- Всероссийская конференция по математическому моделированию и информационным технологиям. Кемерово 2005.
- Конференция «Информационно-вычислительные системы анализа и синтеза изображений». Новосибирск, 2006.
- IEEE International Siberian Conference On CONTROL AND COMMUNICATIONS SIBCON-2007. Tomsk, 2007.
- Международная научно-техническая конференция и выставка ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ЕЕ ПРИМЕНЕНИЕ – DSPA. Москва 2008, 2009.

Публикации. По теме диссертации опубликовано 20 работ, из которых 12 публикаций в трудах и материалах международных конференций и 2 статьи в рецензируемых журналах.

Структура и объем работы. Диссертационная работа состоит из введения, пяти глав и заключения, изложенных на 101 странице, включает 39 рисунков и 6 таблиц. Список литературы содержит 56 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении приводится обоснование актуальности выбранной темы, формулируются цель и задачи работы, описываются научная новизна и практическая ценность работы, приведены основные положения, выносимые на защиту, и информация о внедрении полученных результатов.

Глава 1 «Особенности реализации задач потоковой обработки данных на различных типах вычислительных систем». В главе приводятся результаты анализа и классификация аппаратных устройств для различной обработки данных. При этом изучаются особенности реализации алгоритмов потоковой обработки со скоростью поступления данных. Рассмотрены следующие аспекты:

- процессорные устройства,
- архитектура вычислительных комплексов на базе различных процессорных устройств,
- особенности применения FPGA в качестве процессорных устройств.

В качестве современных вычислительных систем в главе рассмотрены ПК, кластеры, графические ускорители в составе ПК, комплексы на базе сигнальных процессоров и устройства на базе программируемой логики. Все рассматриваемые системы включали в себя универсальный процессор для реализации постобработки данных. В качестве особенностей применения FPGA в качестве процессорного устройства освещены аспекты создания программных моделей на языках описания аппаратуры (HDL), а также их моделирование и отладка.

Производительность вычислительных систем зависит от типа используемого процессорного устройства: универсальный процессор, сигнальный процессор или программируемая логика. Все эти типы процессоров используют различные АЛУ (арифметико-логические устройства), являющиеся вычислительными ядрами процессорного устройства. АЛУ могут исполнять различные наборы команд: от реализации булевой функции над входными данными до операций быстрого фурье-преобразования. Помимо этого, производительность процессорного устройства зависит от:

- тактовой частоты работы АЛУ;
- количества одновременно исполняемых потоков;
- наличия внутренней памяти, работающей на частоте процессора.

В главе рассмотрены наиболее распространенные на сегодняшний день процессорные устройства и вычислительные системы на их основе, что дало возможность сделать оценку возможностей современных систем при решении задач потоковой обработки информации. Выбор FPGA в качестве процессорного средства для таких задач обоснован возможностью построения широких и

глубоких вычислительных конвейеров для эффективной реализации алгоритмов обработки, что позволит существенно повысить производительность по сравнению с другими типами процессорных устройств.

Глава 2 «Программная реализация на языке VHDL задачи обработки последовательности изображений». В Институте автоматики и электрометрии СО РАН под руководством В.С. Киричука был разработан набор алгоритмов для обработки последовательности изображений, получаемых при съемке поверхности Земли из космоса, с целью обнаружения малоразмерных объектов. В главе описываются методы реализации этих алгоритмов на языке VHDL (Very high speed integrated circuits Hardware Description Language), позволяющие за счет существенного распараллеливания вычислительных конвейеров повысить производительность в десятки раз по сравнению с их реализацией на базе сигнальных процессоров. Оценка производительности предложенной программной модели производилась путем ее моделирования на ПК.

Задача обработки последовательности изображений условно разделяется на три этапа: две внутрикадровых обработки (ВКО) и одна межкадровая обработка (МКО) (рис. 1).

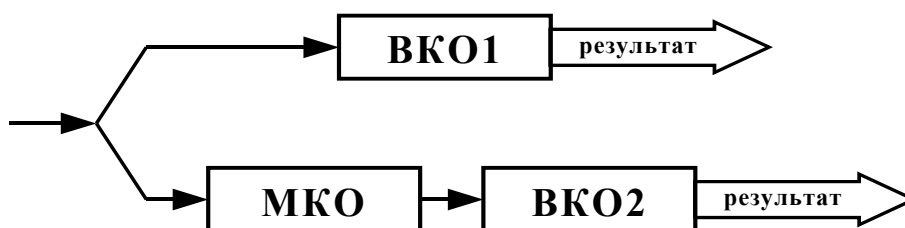


Рис. 1. Общая блок-схема алгоритма обработки последовательности изображений
Алгоритмы ВКО1.

1. Построение разностного изображения

$$C_{diff}(i, j) = C(i, j) - C(i - 1, j), \quad (1)$$

где j – номер строки изображения; i – номер пикселя в строке.

2. Одномерная свертка с импульсным откликом (коэффициенты K_l) и коэффициентом нормировки N .

$$C_{filtr}(i, j) = \frac{1}{N} \sum_{l \in S} K_l C_{diff}(i + l, j), \quad (2)$$

где j – номер строки изображения; i – номер пикселя в строке; l – индекс коэффициента импульсного отклика.

3. Поиск локальных экстремумов. Точка с координатами (i, j) является экстремумом, если выполняются следующие условия:

$$C(i, j) \geq C(i + 1, j) \cup C(i, j) \geq C(i - 1, j). \quad (3)$$

Для реализации всех трех алгоритмов ВКО1 был разработан единый вычислительный конвейер, обеспечивающий обработку данных со скоростью их

поступления (рис. 2). При использовании импульсного отклика размером 7 точек производится одновременно 7 умножений, 1 деление и 7 сложений. Объем хранения промежуточных данных составляет 16 регистров разрядностью от 7 до 22 бит.

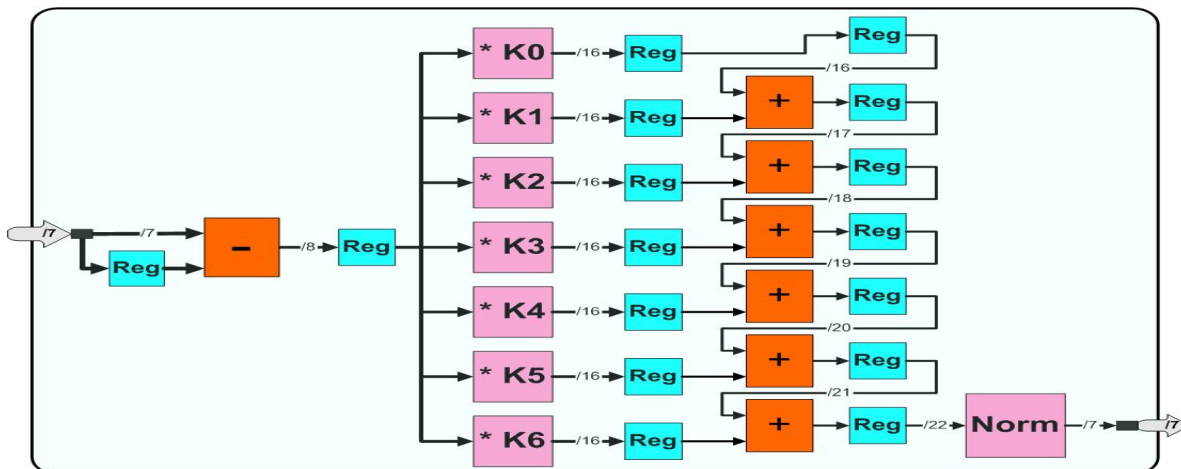


Рис. 2. Блок-схема конвейера внутрикадровой обработки (ВКО)

При межкадровой обработке (МКО) необходимо выделить те объекты, координаты которых изменились по сравнению с предыдущим кадром, учитывая возможность относительного смещения кадров. Для этого совершаются следующие действия:

1. Целочисленная привязка кадров.
2. Целочисленная привязка фрагментов изображения.
3. Алгоритм компенсации фона, включающий в себя компенсацию возможного субпиксельного смещения кадров.

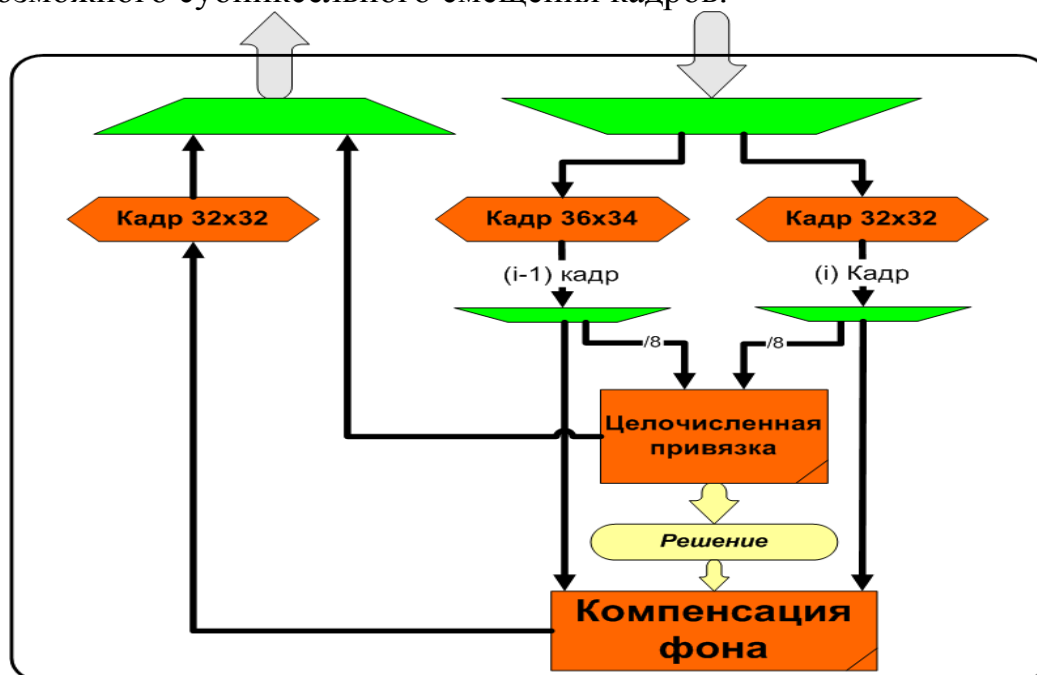


Рис. 3 Блок-схема конвейера межкадровой обработки (МКО)

Особенностью алгоритма МКО является возможное ветвление исполнения в зависимости от результатов каждого этапа. Применение программируемой логики позволяет одновременно исполнить все ветви, а в качестве результата использовать

данные лишь одной ветви. Такой подход не уменьшает производительность каждой ветви в отдельности, но существенно ускоряет обработку в целом. Блок-схема конвейера МКО приведена на рис. 3.

Алгоритмы ВКО2 используются для нахождения объектов на изображении, полученном после МКО, и идентичны ВКО1.

Все эти алгоритмы были реализованы автором на языке описания аппаратуры VHDL в виде единой программной системы, моделирование которой производилось на ПК с использованием программного пакета Active-HDL (Aldec).

По результатам моделирования была проведена оценка производительности предложенной реализации алгоритмов на архитектуре FPGA. Эта реализация оказалась в 50 раз эффективней используемого в настоящее время решения на основе сигнального процессора ADSP 21060 и макета на базе двух векторных процессоров NM6403 (табл. 1).

Таблица 1. Сравнение производительности аппаратных решений

	Время обработки кадра 1536*6200 (сек.)	Производительность (млн. операций в секунду)
ADSP 21060	50,2	120
2x NM6403	28,8	210
Xilinx VirtexE	1	6 024

Далее в главе приводится метод реализации алгоритма двумерной свертки, используемого для решения задач поиска объектов на изображениях. Архитектура FPGA дает возможность производить двумерную свертку со скоростью поступления данных за счет широкого распараллеливания вычислительного конвейера и использования встроенной памяти. Предложенный метод реализации позволяет производить двумерную свертку изображения 2000*2000 пикселей с импульсным откликом 20*10 пикселей за 40 мс при тактовой частоте 100 МГц. На основе описанного метода реализации можно построить произвольный конвейер двумерной свертки, ограниченный лишь логической емкостью используемого кристалла FPGA.

В задачах обработки изображений часто необходимо проверить наличие объекта на изображении, используя критерий среднеквадратичного отклонения. При этом необходимо вычислить коэффициенты для всех возможных положений объекта и найти минимальный, который и будет соответствовать возможному положению объекта на изображении.

$$\text{Min}\{Kor(x, y)\} = \sum_{i,j} (Pic(i, j) - Mask(i, j))^2, \quad (4)$$

где Pic(i, j) – значение пикселя с координатами i, j в исходном изображении; Mask(i, j) – значение пикселя с координатами i, j в маске.

Автором предложен метод реализации на FPGA алгоритма поиска объектов, позволяющий осуществлять обработку маски $X*Y$ в строке $A*Y$ со скоростью поступления данных за счет распараллеливания операций сравнения (сравнение каждого столбца происходит за 1 такт). Таким образом, расчет коэффициентов для $(A-X)$ положений осуществляется за (A) тактов. Расчет всех возможных положений занимает $(A)*(B-Y)$ тактов.

Решение задачи поиска объекта размером $32*32$ на изображении $2000*2000$ займет около $3,8*10^6$ тактов, что при тактовой частоте 200 МГц составляет около 20 мс. Это дает возможность при наличии внутренней памяти программно-аппаратного комплекса осуществлять поиск объектов, непрерывно обрабатывая поток данных 1,5 Гбит/с, получаемый при видеосъемке с разрешением $2000*2000*8b$ и частотой 50 кадров/с.

Необходимо заметить, что все предложенные в данной главе методы реализации алгоритмов на FPGA являются масштабируемыми, т.е. легко модифицируются под параметры задачи и возможности используемой FPGA. Предложенные автором схемы реализации вычислительных конвейеров позволяют достичь высокой производительности и обрабатывать данные в режиме поступления, обеспечивая при этом минимальные задержки при получении результатов обработки.

Глава 3 «Программно-аппаратная архитектура вычислительных комплексов на базе FPGA для потоковой обработки данных». При обработке больших потоков данных на универсальных аппаратных архитектурах часто возникают многочисленные сложности, связанные с пропускными способностями шин, объемом памяти и скоростью передачи данных.

В данной главе сформулированы рекомендации по построению программной модели и аппаратной архитектуры, соблюдение которых существенно повышает производительность аппаратно распараллеливаемых алгоритмов обработки данных, сохраняя универсальность. Все они основаны на проведенных исследованиях особенностей и принципов построения вычислительных комплексов на базе FPGA. Анализ существующих программно-аппаратных решений и общение с разработчиками позволил не только объединить существующие наработки и рекомендации, но и сформулировать выводы, основанные на практическом решении ряда математических задач обработки изображений на базе FPGA.

Анализ производительности различных аппаратных устройств на ряде приложений, связанных с потоковой обработкой данных, показал, что заявленная пропускная способность интерфейсных шин не соответствует экспериментальным результатам. Для выбора того или иного стандарта интерфейсной шины в задачах, где требуется гарантированная скорость передачи данных, необходимо знать реальную пропускную способность. Поэтому было проведено сравнительное исследование реальных пропускных способностей шин PCI-X, PCI-E x1 и PCI-E x4.

По результатам проведенных исследований производительности интерфейсных шин и динамической памяти были сформулированы ряд принципов схемотехнического проектирования аппаратных устройств для задач

высокопроизводительной обработки потоковых данных. Соблюдение следующих рекомендаций позволит избежать возникновения факторов, ограничивающих производительность всей вычислительной системы:

- при необходимости передачи потоков данных более 2,5 Гбит/с между устройством и ПК необходимо использовать шины PCI-E x4, x8 либо x16, так как шины PCI-X и PCI-E x1 в силу особенностей реализации контроллеров на материнских платах не обеспечивают передачу таких потоков;
- для достижения максимальной скорости обмена данными с ПК необходимо применять контроллеры на базе FPGA, поскольку существующие внешние микросхемы контроллеров шин обладают рядом недостатков при потоковой передаче данных;
- наличие динамической памяти, связанной непосредственно с кристаллом FPGA, позволит с максимальной скоростью осуществлять обмен данными, минуя промежуточные контроллеры и ограничения внутренних шин;
- наличие нескольких кварцевых генераторов позволит каждому программному модулю функционировать на максимальной частоте, тем самым обеспечивая максимальную производительность системы.

Представленное в главе структурное решение программных модулей было разработано автором с целью повышения эффективности решения различных задач при сохранении общей программной архитектуры системы. Это достигается за счет выделения следующих программных модулей (рис. 4).

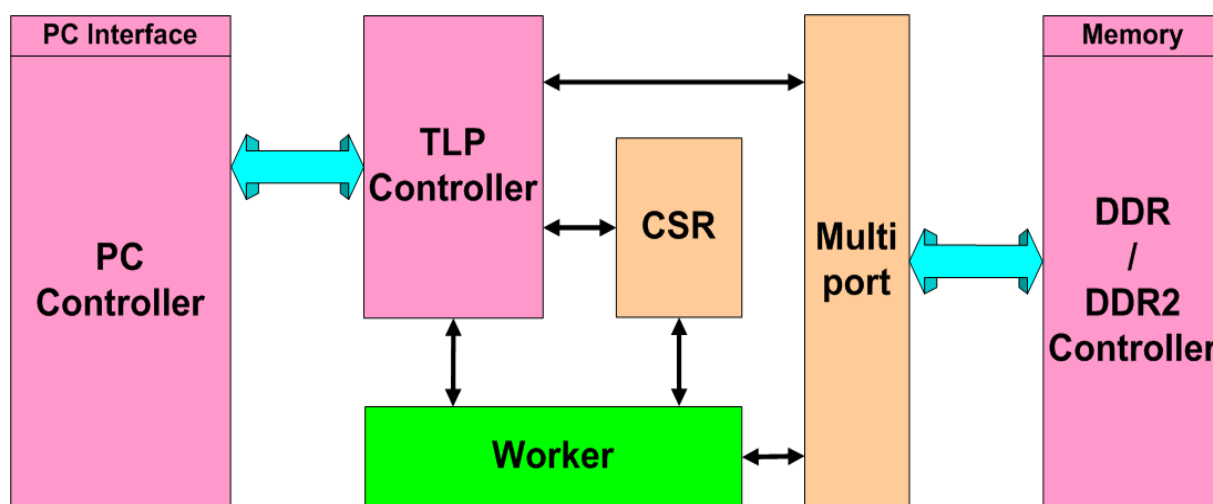


Рис. 4. Программная архитектура комплексов на базе FPGA

Worker – компонент, реализующий обработку данных и имеющий унифицированный интерфейс для управления его работой. При этом можно сказать, что данный компонент представляет собой уровень реализации алгоритмов обработки, а все описанные ниже компоненты обеспечивают уровень интерфейсного взаимодействия.

- PC Controller – компонент для реализации физического уровня интерфейса общения с ПК. В случае PCI-X включает в себя буферы данных и управление базисами тактовых сигналов, а в случае PCI-E реализует физический уровень протокола.
- TLP Controller – модуль обработки запросов шины ПК, также реализующий метод Scatter/Gather Lists, который обеспечивает непрерывную работу с блоками данных до 4 МБ в оперативной памяти ПК на аппаратном уровне.
- CSR – Control Status Registers. Модуль выполняет функцию управления работой всего устройства, что позволяет даже в случае замены интерфейсной шины или памяти оставить без изменения драйвер устройства и пользовательское приложение.
- Multiport – модуль, предназначенный для выполнения функции арбитра общего ресурса – память в составе программно-аппаратного комплекса.
- DDR/DDR2 Controller – компонент контроллера динамической памяти.

Особенности шин семейства PCI таковы, что добиться максимальной пропускной способности возможно лишь при пакетном чтении и записи. Стандартные обращения операционной системы к устройству на шине разбиваются на одиночные или пакетные операции записи и чтения на уровне контроллера шины. При этом на инициализацию каждого обращения к устройству существуют накладные расходы инициализации обращения (вызов функции драйвера, запрос шины, выставление команды и адреса операции), часто занимающие гораздо большее время, чем само обращение.

При работе с памятью в режиме прямого доступа DMA у применяемых шин PCI-X и PCI-E существует ограничение на максимальный размер блока данных, которое составляет 4 КБ (размер одной страницы памяти). Это означает, что для прочтения или записи блока данных размером более 4 КБ необходимо несколько раз инициализировать транзакцию, указав при этом физический адрес. Для обеспечения возможности оперировать с данными до 4 ГБ автором был предложен и реализован метод адресации, заключающийся в организации дополнительного уровня логики в стандартной схеме «Scatter/Gather Lists» и реализации «списка списков» страниц памяти. Практическое тестирование, проведенное в рамках экспериментов по определению пропускной способности шин данных, подтвердило эффективность использования такого метода адресации данных, позволившего повысить производительность передачи данных примерно на 15%, по сравнению со стандартным методом Scatter/Gather Lists.

Описанные принципы построения спецвычислителей на базе FPGA позволяют использовать преимущества архитектуры FPGA и получать увеличенную производительность по сравнению со стандартными решениями. При таком разделении на функциональные программные модули сохраняется гибкость применения всего аппаратного решения, что позволяет использовать его для решения различных задач и реализации различных алгоритмов обработки данных.

Глава 4 «Семейство программно-аппаратных комплексов SLSP». На основе сформулированных в предыдущей главе рекомендаций по проектированию программной и аппаратной архитектуры комплексов на базе FPGA было разработано семейство программно-аппаратных комплексов SLSP для решения задач высокопроизводительной потоковой обработки данных. Все спецвычислители используют в качестве процессора микросхемы программируемой логики архитектуры FPGA Xilinx семейства Virtex. В каждом спецвычислителе используется внутренняя динамическая память стандарта DDR (SLSP-1 и SLSP-2) и DDR2 (HDG). На сегодняшний день семейство представлено тремя комплексами: **SLSP-1** (FPGA Xilinx XC2V1000, 32-bit PCI-X, 128 МБ DDR), **SLSP-2** (FPGA Xilinx XC2V3000, 64-bit PCI-X, 256 МБ DDR) и **HDG** (FPGA Xilinx XC5LX50T, PCI-E x4, 4 SDI/HD-SDI, 2 x 128 МБ DDR2).

В таблице 2 представлена пропускная способность каждого программно-аппаратного решения при работе с внутренней памятью, а также при записи данных в оперативную память ПК и чтении из нее.

Таблица 2. Максимальная пропускная способность комплексов SLSP

	Чтение / запись собственной памяти, МБайт/с	Передача в ПК, МБайт/с	Чтение из ПК, МБайт/с
SLSP-1	850	88	71
SLSP-2	850	354	284
HDG	2 x 1,9	683	536

Созданное семейство программно-аппаратных комплексов SLSP может применяться для решения следующих задач:

- обработка и передача телевизионных изображений;
- математическое моделирование трудоемких алгоритмов и задач;
- автоматизация физических экспериментов;
- исследование генома в биоинформатике;
- бортовые спутниковые спецвычислители для обработки и анализа изображений.

Описанные в главе программно-аппаратные комплексы были разработаны автором в соответствии со сформулированными рекомендациями по схемотехническому проектированию и архитектуре программных модулей, связанными с выделением слоев интерфейсной обвязки и обработки данных, и позволили достичь высокой производительности и надежности, что подтверждается двумя актами внедрения.

Глава 5 «Применение программно-аппаратного комплекса HDG». В главе исследуется эффективность применения разработанного комплекса HDG при решении следующих задач:

- обработка четырех потоков формата HD-SDI (каждый поток 1,5 Гбит/с) и передача результирующего потока 3,9 Гбит/с в память ПК;
- исследование генома для нахождения возможных транскрипционных факторов.

Первое решение позволит виртуальным студиям, предназначенным для размещения актера в трехмерном виртуальном пространстве, перейти на стандарт видео высокой четкости (HD) в рамках стандартного ПК. Для обработки видеоданных был использован программно-аппаратный комплекс HDG, программная архитектура которого представлена на схеме (рис. 5).

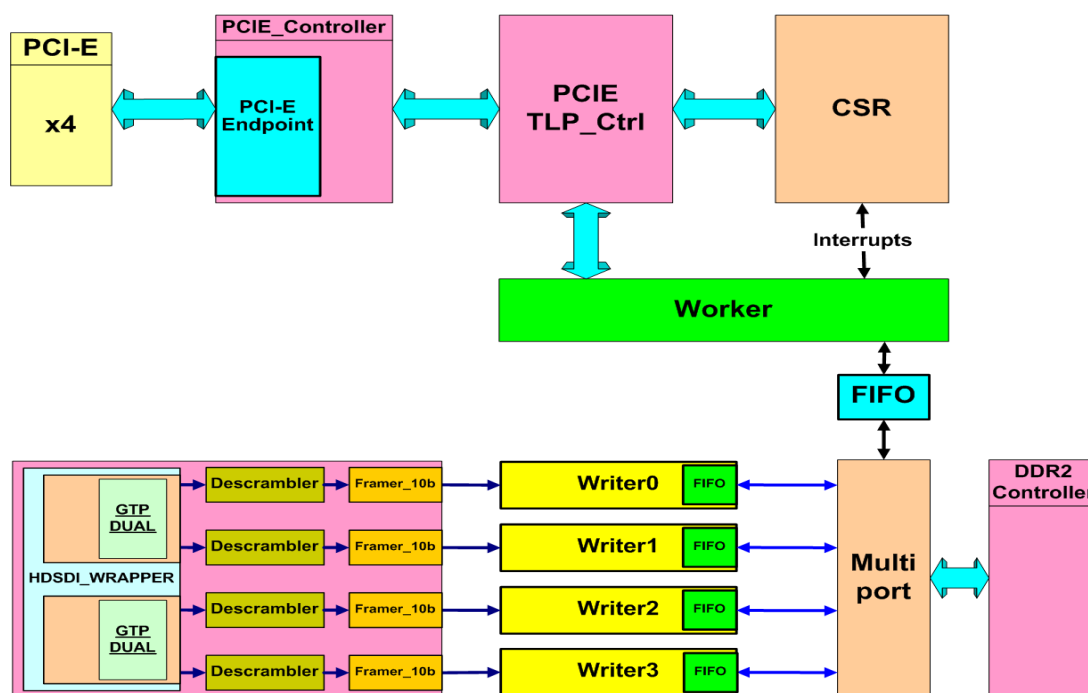


Рис. 5. Блок-схема программной архитектуры виртуальной студии на базе HDG

Было разработано и реализовано пользовательское приложение, позволившее не только передавать данные в оперативную память ПК со скоростью 3,9 Гбит/с, но и производить необходимую обработку видеоданных для дальнейшего воспроизведения или передачи в эфир.

Программная архитектура отличается от описанной в третьей главе (рис. 4) лишь тем, что добавилось 4 клиента, которые записывают видеоданные во внутреннюю память устройства. Таким образом, эффективность предлагаемых рекомендаций построения программной архитектуры была подтверждена на примере решения практической задачи.

Второе решение используется в биоинформатике, где существует потребность анализа регуляторных выборок, являющихся набором нуклеотидных последовательностей. При анализе необходимо осуществить поиск возможных

транскрипционных факторов, что реализуется методом полного перебора всех возможных факторов (рис. 6).

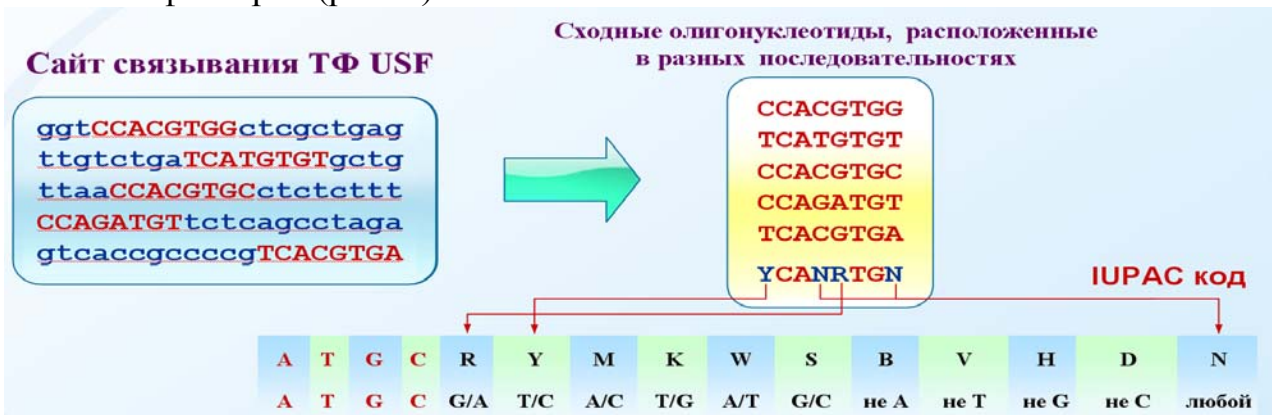


Рис. 6. Анализ регуляторных выборок

В настоящее время задача решается для следующих параметров:

- выборка 1000x64 символов в алфавите А-Т-Г-С;
- длина транскрипционного фактора (далее мотива) – 8 символов;
- алфавит мотива – 15-буквенный (для компенсации вырождения).

Решение такой задачи требует провести 1000*57 сравнений 8-буквенных слов, записанных в разных алфавитах, для каждого возможного мотива. Количество возможных мотивов – 15⁸. Таким образом, трудоемкость задачи составляет порядка 10¹⁵ сравнений. Существующая реализация на ПК решает задачу за 14 суток.

Предложенная реализация алгоритма на базе FPGA семейства Virtex5 позволяет осуществлять сравнение одной 64-символьной строки с 8-символьным мотивом за 1 такт (рис. 7). Это достигается за счет использования 6-входовых таблиц истинности: символ в 4-буквенном алфавите записывается 2-битным словом, а в 15-буквенном – 4-битным.

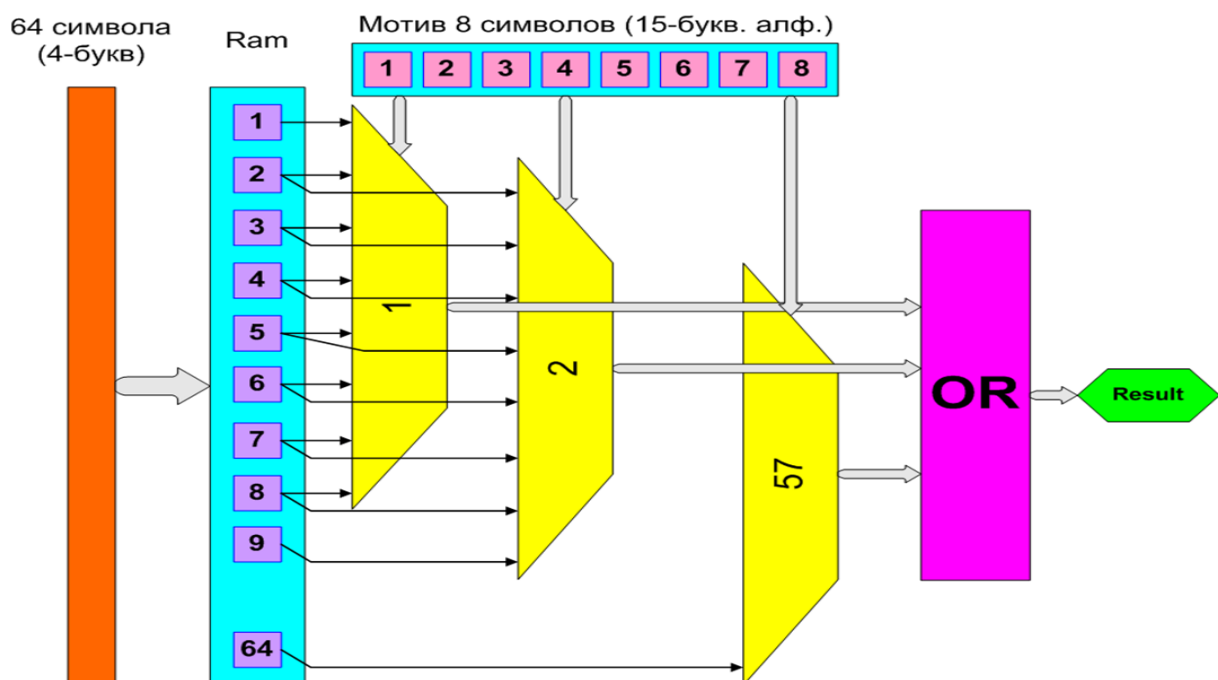


Рис. 7. Сравнение 64-символьной строки с 8-символьным мотивом

Путем моделирования установлено, что применение программно-аппаратного комплекса HDG позволит решить поставленную задачу за 9 минут (при частоте работы 250 МГц) за счет одновременного сравнения 20 строк с одним мотивом на каждом такте. При этом вся выборка исходных данных может храниться во внутренней памяти кристалла, а генератор мотивов (так как необходим полный перебор) может быть реализован внутри FPGA. Таким образом, поток входных данных будет равен нулю. Применение разработанного программно-аппаратного комплекса HDG позволит в 2 000 раз ускорить решение задачи по сравнению с универсальным ПК на базе Core2Duo, а при использовании кристалла XC5VLX330T – в 20 000 раз.

В заключении сформулированы основные результаты диссертации:

1. Разработана программно-аппаратная архитектура вычислительных комплексов на базе FPGA для обработки потоков данных порядка 10 Гбит/с, позволяющая оперировать данными со скоростью их поступления за счет оптимизации операций с памятью, организации программных модулей и создания специального программного обеспечения.
2. Предложен метод реализации задачи обработки последовательностей изображений, включающей такие разнородные операции как фильтрация, компенсация целочисленного сдвига, компенсация фона, выявление экстремумов и пороговая обработка.
3. Предложен метод реализации задачи поиска транскрипционных факторов в регуляторных выборках генома, который обеспечивает производительность $1,67 \cdot 10^{13}$ операций целочисленного сравнения в секунду и позволяет уменьшить время решения задачи в 20 000 раз по сравнению с использованием стандартного ПК.
4. Разработан программно-аппаратный комплекс, позволяющий моделировать работу алгоритмов, реализованных на языках описания аппаратуры (HDL), для их тестирования и выявления факторов, ограничивающих производительность.
5. Создан макет бортового спецвычислителя для обработки последовательностей изображений и поиска малоразмерных объектов, способный в режиме поступления обрабатывать поток данных 1,5 Гбит/с, что в 50 раз превышает возможности существующего решения на сигнальном процессоре ADSP21060.
6. Создан программно-аппаратный комплекс для одновременной обработки семи видеопотоков формата HD (1,5 Гбит/с), что в сумме составляет около 10 Гбит/с.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Шадрин М.Ю., Лысаков К.Ф., Девятайкин А.М. Реализация алгоритмов обработки данных на базе программируемой логики для решения задач обнаружения и наблюдения за природными явлениями // Материалы международной конференции «Вычислительные и информационные технологии в науке, технике и образовании». Алматы: КазНУ им. Аль-Фараби, 2004. С. 270–279.

2. Gromilin G.I., Devjataikin A.M., Lysakov K.F., Shadrin M.J. Hi-performance co-processor based on FPGA // Proceedings of the Second IASTED International Multi-Conference AUTOMATION, CONTROL, AND APPLICATIONS. Novosibirsk: ACTA Press, 2005. P. 89–92.
3. Devjataikin A.M., Lysakov K.F., Shadrin M.J. FPGA implementation of sub-pixel images matching algorithm // Proceedings of the Second IASTED International Multi-Conference AUTOMATION, CONTROL, AND APPLICATIONS. Novosibirsk: ACTA Press, 2005. P. 93–97.
4. Лысаков К.Ф. Применение ВМПП для создания высокопроизводительных вычислительных устройств и систем // Сборник трудов по итогам XI Международной научной конференции «Современные проблемы информатизации в моделировании и программировании». Воронеж: Издательство «Научная книга», 2006. С. 238–239.
5. Лысаков К.Ф., Шадрин М.Ю. Универсальный модуль высокопроизводительной обработки данных в составе ПК — SLSP-2 // Труды международной конференции «Вычислительные и информационные технологии в науке, технике и образовании». Павлодар: ТОО НПФ «ЭКО», 2006. С. 11–17.
6. Lysakov K.F., Shadrin M.Y. SLSP – Special Processor for Image and Video Processing // Proceedings of the IEEE International Siberian Conference On CONTROL AND COMMUNICATIONS SIBCON-2007. Tomsk: TUSUR, 2007. P. 140–144.
7. Lysakov K.F., Shadrin M.Y. SLSP-2 – FPGA-based standalone system supporting work as PCIX-X module // Proceedings of the 8th International Conference “Pattern Recognition and Image Analysis: New Information Technologies” (PRIA-8-2007):. Vol.2. Russia, Yoshkar-Ola: Mari State Technical University, 2007. P. 192–195.
8. Лысаков К.Ф., Рудаков А.В., Шадрин М.Ю. Программно-аппаратный комплекс HDG с интерфейсами SDI / HD-SDI и PCI-E с возможностью работы вне ПК // Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. Выпуск: XI-2. Москва. 2009. С. 563–565.
9. Лысаков К.Ф., Рудаков А.В., Шадрин М.Ю. Применение FPGA для решения задач биоинформатики и исследования генома // Тезисы докладов Всероссийской конференции, приуроченной к 80-летию академика С. К. Годунова «Математика в приложениях» Новосибирск, 2009. С. 176–177.
10. Lysakov K.F., Shadrin M.Yu. Principles for Implementing and Optimizing Mathematical Image Processing Algorithms Based on Programmable Logic // Pattern Recognition and Image Analysis, Vol. 19, No. 1. 2009. P. 52–58.
11. Лысаков К.Ф., Шадрин М.Ю. Особенности применения аппаратных устройств на базе FPGA для задач потоковой обработки изображений // Вестник Новосибирского Государственного Университета, Серия: Информатика. Том 7, выпуск 3. 2009. С. 15–22.

Научное издание

Лысаков Константин Федорович

**ИССЛЕДОВАНИЕ МЕТОДОВ РЕАЛИЗАЦИИ
АЛГОРИТМОВ ОБРАБОТКИ БОЛЬШИХ ПОТОКОВ ДАННЫХ
ЗА СЧЕТ КОНВЕЙЕРНОГО РАСПАРАЛЛЕЛИВАНИЯ**

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Подписано в печать 07.05.2010 г.
Формат 60x84 1/16. Офсетная печать.
Усл. печ. л. **1,1**. Уч.-изд. л. **1**.
Тираж 100 экз.
Заказ № ____

Редакционно-издательский центр НГУ
630090, Новосибирск-90, ул. Пирогова, 2.