

РЕАЛИЗАЦИЯ СПЕЦИАЛИЗИРОВАННЫХ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВСТРАИВАЕМЫХ СИСТЕМ НА ОСНОВЕ ПЛИС

В.М. Ефименко

1. Введение. Микросхемы программируемой логики (ПЛИС) благодаря прогрессу микроэлектроники перешагнули рубеж в несколько миллионов свободно коммутируемых логических элементов и триггеров. Н.П.Бехтерева [1] писала о человеческом мозге: «Итак, множество клеток, множество связей... нейронные популяции реально или потенциально полифункциональны, связи функционируют, уровень активации различных зон мозга легко меняется в соответствии с необходимостью. Все они особенно важны для мозга с его почти астрономическим количеством клеточных элементов и связей между ними... один и тот же нейрон может участвовать в разных ансамблях... – то, что мы определяли как пространственно-временное кодирование.»

Конечно, реальным ПЛИС еще далеко до мозга человека, но аналогия напрашивается: нейроны – логические блоки с триггером, связи – внутренняя система коммутации, память – внутреннее ОЗУ, и т.д.

Кратко оценим вычислительные ресурсы микросхемы ПЛИС семейства «STRATIX V» фирмы ALTERA, не самой большой этого семейства, стоимостью порядка тысячи долларов.

На кристалле имеются следующие ресурсы (тактовая частота 500МГц):

- Около миллиона эквивалентных логических элементов и триггеров. Эти блоки позволяют организовать порядка 20 тысяч регистров общего назначения разрядностью 32 бит, и около каждого регистра поставить свой операционный блок (сумматор, схема сравнения, сдвига, и др.). Автоматическая организация ускоренных переносов при компиляции и трассировке проектов. Разрядность всех регистров произвольная и ограничена общими ресурсами. Вполне реальна разрядность сдвигового регистра 1Кбит, а для простейших операций (суммирование с фиксированной запятой) – 64 бита с организацией параллельно-последовательного двух – пяти ступенчатого переноса (задается опционально).
- Набор аппаратных целочисленных умножителей с переменной разрядностью (от 9 до 36 бит) с умножением за один такт – порядка 700 умножителей (18x18) и 350 умножителей (27x27). Все умножители работают независимо и параллельно.
- ОЗУ порядка 7Мбайт. Часть ОЗУ имеет двухпортовую организацию (запись и чтение одновременно по разным адресам), допускает конфигурирование (разделение) на отдельные блоки произвольного размера и разрядности, что позволяет организовать множество параллельно работающих локальных ОЗУ для каждого вычислительного процесса (блока).
- Важный ресурс – система коммутаций. Множество локальных и глобальных, горизонтальных и вертикальных шин, позволяющее организовать статические каналы связи между регистрами и блоками, т.е. структуру системы. Передача информации между блоками производится одновременно по многим шинам. Динамические изменяемые связи организуются посредством соответствующих мультиплексоров.
- Высокоскоростной внешний интерфейс. Организация внешних каналов связи со скоростями порядка десятки Гбайт/с.

Оценка производительности на операциях с плавающей запятой одинарной точности на примере перемножения матриц 8x8 дает 6,3Gflops на частоте 420МГц [2]. На упомянутом кристалле можно разместить порядка 20 таких блоков перемножения матриц. При этом суммарная производительность составит 126Gflops, однако при этом вступит в силу ограничение по рассеиваемой мощности. Так что реально будем иметь не более 30-40 Gflops на кристалл. Следует отметить, что по соотношению Gflops/Ватт ПЛИС дает наилучшие результаты [3].

2. Организация вычислений. В процессе разработки с применения ПЛИС применялись две основные архитектуры организации вычислений.

А. Структурная организации вычислений. Алгоритм вычислений реализован аппаратно в виде соответствующей конвейерной структуры [4]. Такая организация пригодна для относительно простых, но требовательных к объемам вычислений алгоритмов. Вычислительная структура, вместе с соответствующими регистрами распределена пространственно по площади кристалла. Реализация фильтров, систем управления, просто вычислений по формулам. В случае недостаточного быстродействия отдельные функциональные блоки разделяются на сегменты, с последующей конвейеризацией, либо блоки тиражируются (горизонтальное масштабирование).

Разработка структуры производится на языках HDL группы, и достаточно трудоемка. Заслуживает интереса применение пакета «Simulink» (системы «Matlab»), позволяющее непосредственно переносить функциональную блок – схему на структуру кристалла ПЛИС.

Б. Организация вычислений на основе управляющего двоичного автомата.

Исходными данными является программа на языке «С». Любой алгоритм, выраженный на языке процедурного или объектного программирования, реализуется с помощью двоичного автомата [5]. Автомат с n состояниями реализуется на основе $(n+1)$ – разрядного регистра сдвига с единицей в одном из разрядов регистра. При переходе из одного состояния в другое единица перемещается из одного разряда регистра в другой. Эти выходы регистра непосредственно используются для активизации операционных блоков и межрегистровых передач. Таким образом, двоичный автомат совмещает функции реализации алгоритма и управления процессом вычисления, а благодаря пространственному разделению выходов регистра, не требуется дешифрации команд. Просто в процессе разработки необходимо ответить на вопрос, какие операции выполняются в каждом состоянии автомата. Для этого составляются соответствующие таблицы занятости [4].

Все внутренние переменные объявляются регистрами. Это позволяет получить одновременный доступ ко всем переменным. Если есть внутренние массивы, то вводится внутренняя, локальная память для каждого массива за счет ресурсов ПЛИС. Массивы, по возможности располагаются не в разных областях одной памяти, а физически в разных ОЗУ. В этом случае каждое ОЗУ имеет меньше «пользователей» и допускается параллельное обращение к разным массивам. Для каждого регистра синтезируется отдельный операционный блок.

Ускорение вычислений достигается за счет

- Отсутствия стадий выборки и дешифрирования команд.
- Отсутствия перезагрузки переменных из ОЗУ в регистры процессора.
- Совмещения переменных, размещенных в регистрах с операционными блоками и шинами межрегистровых передач.
- Возможности активизации одновременно многих операционных элементов (до сотни).
- Совмещение стадий выполнения, условного перехода или вызова процедуры.
- Более эффективная реализация условного перехода. В одном состоянии автомата может быть проверено сколько угодно условий. Может быть сколько угодно вариантов переходов из одного узла автомата.

Актуальной задачей является разработка компилятора со стандартного языка высокого уровня (С) для реализации непосредственно в ПЛИС. Разрабатываемое в настоящее время в ООО «Электронные системы» (Ульяновск) программное обеспечение по синтезу структуры ПЛИС использует в качестве исходных данных программу на языке «С». При запуске компилятора инструкции промежуточного кода модифицируются с тем, чтобы снять ограничение по количеству регистров общего назначения для воображаемого процессора. Полученная совокупность инструкций промежуточного кода подвергается операции «сжатия» во времени путем совмещения операций. Ограничения совмещения связаны только с исходной программой, а не с недостатком аппаратных средств.

В компилятор закладывается способ реализации, перевода конструкций языка высокого уровня в архитектуре исполняющего вычислительного оборудования (синтезируемого специализированного процессора).

Трансляция программы можно рассматривать как отображение в структуру. Язык «С» переводится в структуру, описанную на языке VHDL. При трансляции с VHDL в ПЛИС идет автоматическое исключение лишних переменных.

В случае объектного программирования каждый объект имеет свою локальную память, внутренние переменные (в регистрах) и управляющий автомат – получается наиболее естественная реализация объекта. У объекта «все свое», при этом снимается ограничение по защите памяти – никакой другой объект не имеет физического доступа к памяти объекта.

Преимущество ПЛИС состоит в том, что она содержит все необходимые функциональные блоки, необходимые для организации сбора и обработки информации в реальном времени, реализация задач управления объектом (традиционные схемы управления легко переносятся в ПЛИС), параллельной реализации вычислительных алгоритмов, сопряжения со средствами отображения графики (непосредственная стыковка с ЖК панелью по LVDS каналу) и высокоскоростного интерфейса. Это позволяет организовать всю встраиваемую систему в одном кристалле.

Термин «reconfigurable computation» подразумевает способ реализации вычислений путем задания структуры (конфигурации). Однако это не означает, что в процессе вычислений структура может быть динамически изменена. Если необходимо изменить проект, то необходимо вновь оттранслировать проект, эта трансляция длительная, т.к. сопряжена с размещением и трассировкой. Полная трансляция в исполнимый модуль в сотни раз превышает по времени трансляцию для процессора, и реально составляет десятки минут. Поэтому эта реконфигурация занимает много времени. Получается выигрыш по быстродействию при исполнении, но длительная подготовка исполнимого файла, загружаемого в ПЛИС. Для встроенных систем это ограничение не существенно, т.к. работает одна и та же разработанная структура.

Если же сама реконфигурация вычислений заложена в проект, то она легко реализуется путем

коммутации каналов с помощью логики, и структура реализации вычислений может быть динамически изменена в процессе работы за один такт. Но для этого программист (разработчик) должен предусмотреть заранее все варианты реконфигурации в проекте.

ЛИТЕРАТУРА:

1. Н.П.Бехтерева Магия мозга и лабиринты жизни. АСТ Издательство, 2007, 351 с. ISBN: 978-5-17-042130-5
2. Floating-Point Megafunctions. User Guide. UG-01058-5.0 Altera 11.0 May 2011.
3. <http://www.altera.com/products/ip/dsp/arithmetic/m-alt-float-point.html>.
4. П.М.Коуги Архитектура конвейерных ЭВМ. М., Радио и связь, 1985.
5. В.М. Ефименко Аппаратная реализация программного обеспечения в ПЛИС для встраиваемых систем реального времени. -Труды Международной суперкомпьютерной конференции (20-25 сентября 2010 г., г. Новороссийск). - М.: Изд-во МГУ, 2010. с 18-22.