

Блок аппаратных ускорителей “ACC_Cores” в составе системы-на-кристалле. Многофункциональный аппаратный ускоритель быстрого преобразования Фурье – FFT_RT_Core

Д.С. Фролов, П.П. Пирогов, Ю.Н. Александров, Ю.И. Грибов, А.А. Беляев

ОАО НПЦ «ЭЛВИС», dfrolov@elvees.com

Аннотация — Представлена архитектура IP-блока аппаратных ускорителей, обеспечивающая параллельную бесконфликтную работу нескольких блоков аппаратных ускорителей. Рассмотрена реализация блока аппаратного ускорителя, выполняющего функцию быстрого преобразования Фурье. Проведен сравнительный анализ с другими реализациями аппаратных ускорителей данной функции.

Ключевые слова — блок аппаратных ускорителей, FFT, быстрое преобразование Фурье, Radix-4, DSP, цифровая обработка сигналов.

I. ВВЕДЕНИЕ

В области цифровой обработки сигналов и изображений за все время ее существования всегда предъявлялись повышенные требования к производительности и многофункциональности аппаратно-программных платформ. Данные характеристики в основном и определяют все остальные параметры системы, такие как качество передаваемой информации по линии связи, её объем, качество сжатия изображений, количество обрабатываемых кадров в единицу времени и т.д.

Разработанные на данный момент стандарты связи и обработки изображений представляют собой алгоритмы, требующие исполнения операций в режиме реального времени. Для решения этой задачи необходимо определить и классифицировать узкие места исполнения той или иной операции. В основном, можно выделить несколько типов узких мест: большое количество вычислительных операций; побитовые операции; необходимость неоднородных обращений к внешней памяти; необходимость во внешней памяти с высокой пропускной способностью.

Например, алгоритмы помехоустойчивого кодирования и декодирования требуют как определенных вычислительных мощностей, так и особого обращения с памятью. В существующих алгоритмах нередко используется побитовое перемежение, деперемежение, кодирование, декодирование Хаффмана и другие функции, требующие специфических побитовых операций и нелинейного обращения к памяти. В стандартах телерадиовещания (DVB-T, DVB-T2) одной из основных операций является быстрое преобразование

Фурье. Она используется для реализации функции OFDM, которая довольно требовательна к вычислительным ресурсам.

Решение проблем, упомянутых ранее, возможно с помощью встраивания в аппаратно-программную платформу аппаратных ускорителей вычислений и обработки информации, предназначенных для минимизации затрат по площади и потреблению систем-на-кристалле, освобождения от регулярных операций программируемых устройств сигнальной обработки, таких как DSP, и достижения требуемой производительности для решения задачи в реальном времени.

II. СТРУКТУРА IP-БЛОКА АППАРАТНЫХ УСКОРИТЕЛЕЙ

Структура блока аппаратных ускорителей представлена на Рис. 1.

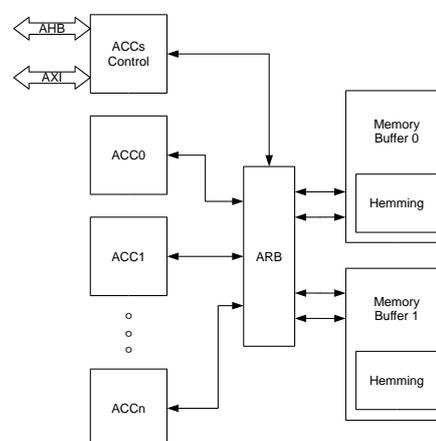


Рис. 1. Структурная схема IP-блока аппаратных ускорителей

IP-блок состоит из следующих элементов:

- ACCs Control – модуль управления блоком ACC. Осуществляется формирование управляющих сигналов для блоков ускорителей и модулей защиты памяти, прерываний, преобразование входных и выходных данных (при необходимости).

- ACC0...n – модули аппаратных ускорителей. Являются самостоятельными устройствами, включающими в себя блок формирования запросов во внутреннюю память и вычислительные ядра, отвечающие за обработку данных.

- ARB – арбитр обращений во внутреннюю память. Позволяет обеспечить параллельную работу нескольких аппаратных ускорителей, блокируя запросы по одинаковым адресам в соответствии с выбранными приоритетами блоков.

- Memory Buffer 0,1 – внутренние блоки памяти, предназначенные для хранения входных, промежуточных и выходных данных.

- Memming – модуль защиты памяти. Обеспечивает исправление ошибки 1 бита информации на одно 32-разрядное слово и сигнализацию о наличии ошибки более чем в одном бите.

Данная архитектура IP-блока аппаратных ускорителей позволяет стандартизовать интерфейсы блоков ускорителей, обеспечить легкую интеграцию новых блоков, производить параллельную работу нескольких аппаратных ускорителей, сократить количество памяти, размещаемой в системе-на-кристалле. Схема интеграции IP-блока в систему-на-кристалле представлена на Рис. 2.

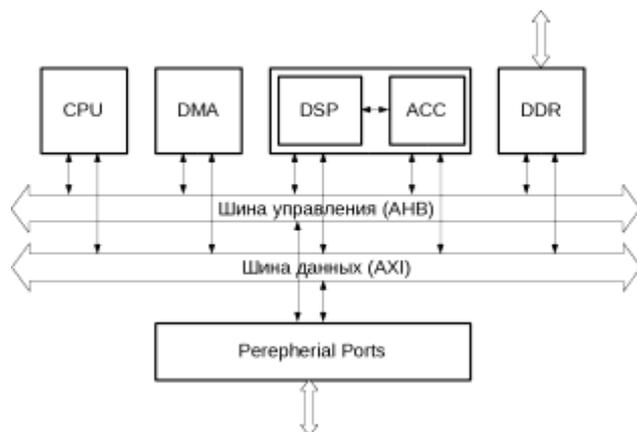


Рис. 2. Схема интеграции блока ACC_Cores в составе SoC

IP-блок аппаратных ускорителей работает под управлением центрального процессора CPU или DSP. Загрузка входных данных и выгрузка результата во внешнюю память осуществляется с помощью внешнего DMA-контроллера.

III. АППАРАТНЫЙ УСКОРИТЕЛЬ БПФ FFT_RT_CORE

A. Дискретное преобразование Фурье

Дискретное преобразование Фурье (в англоязычной литературе DFT, Discrete Fourier Transform) — это одно из преобразований Фурье, широко применяемых в алгоритмах цифровой обработки сигналов (его моди-

фикации применяются в сжатии звука в MP3, сжатии изображений в JPEG и др.), а также в других областях, связанных с анализом частот в дискретном (к примеру, оцифрованном аналоговом) сигнале.

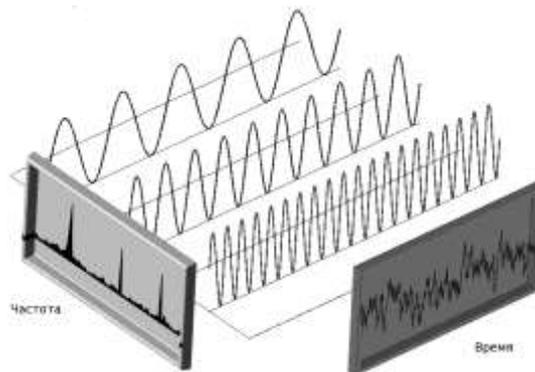


Рис. 3. Физический смысл дискретного преобразования Фурье

Физический смысл дискретного преобразования Фурье - представление некоторого дискретного сигнала в виде суммы гармоник. Поясняющее изображение представлено на рис. 3.

B. Структурная схема FFT_RT_Core

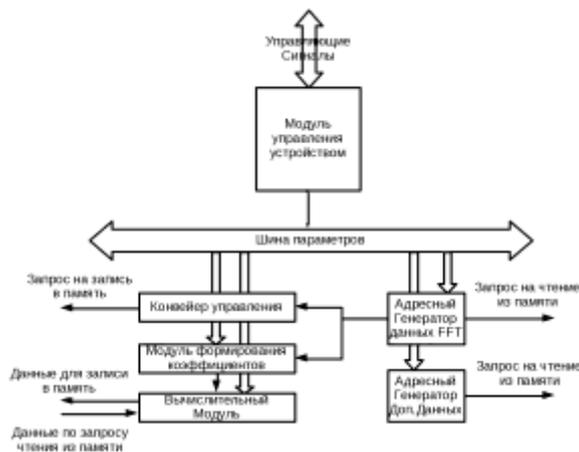


Рис. 4. Структурная схема аппаратного ускорителя FFT_RT_Core

Блок аппаратного ускорителя быстрого преобразования Фурье FFT_RT_Core состоит из следующих модулей:

- 1) Модуль управления устройством. В данном модуле содержится набор программируемых регистров для управления режимами аппаратного ускорителя. Также им формируются прерывания об окончании работы блока аппаратного ускорителя.
- 2) Адресный генератор данных FFT. Модуль предназначен для формирования запросов к внутренней памяти для чтения/записи данных. Модуль

автоматически генерирует адреса в зависимости от выбранного режима работы блока ускорителя.

3) Адресный генератор дополнительных данных. Производится формирование запросов к внутренней памяти для чтения данных второго массива в таких режимах работы блока ускорителя как: поэлементное перемножение, быстрая свертка.

4) Модуль формирования коэффициентов. В данном модуле производится генерация поворачивающих коэффициентов, значение и порядок генерации которых зависит от выбранного режима работы устройства ускорителя.

5) Вычислительный модуль. Представляет собой набор умножителей, сумматоров и дополнительной логики для исполнения необходимой операции в зависимости от режима работы устройства ускорителя. Данный модуль конвейеризован с целью повысить быстродействие операции.

6) Конвейер управления. Для повышения производительности процесс исполнения всех операций был конвейеризован. По конвейеру управления передается информация о режиме работы, адреса записи результата исполнения операции и управляющие сигналы для вычислительного модуля устройства ускорителя.

C. Области применения

Радиолокация, гидроакустика, связь, телевизионное вещание, спутниковая ретрансляция, сжатие информации, обработка сигналов и изображений.

D. Особенности блока FFT_RT_Core

1) Форматы действительных/мнимых компонент входных и выходных данных задаются программно:

- 32-разрядная плавающая точка (стандарт IEEE-754);
- 32-разрядное целое число (дополнительный код);
- 16-разрядное целое число (дополнительный код).

2) Формат вычислений: 32-разрядная плавающая точка. Преобразование типов происходит автоматически с помощью специального конвертера. Преобразование формата производится только при загрузке данных во внутреннюю память и при выгрузке из нее.

3) Способ округления при вычислениях и при преобразовании форматов – к ближайшему числу, а при равноудаленности – к четному.

4) Максимальный размер непосредственно выполняемого преобразования – 4096, минимальный – 16.

5) Размер памяти БПФ-сопроцессора – 8192x64 разрядных слов, она расположена в общем пространстве памяти DSP-процессора. Память используется в 2-буферном режиме, граница буферов соответствует середине памяти. 2-буферный режим предназначен для непрерывного выполнения преобразований в реальном времени. Он позволяет

внешним устройствам (например, DSP-процессору, DMA-каналу, RISC-ядру) параллельно с текущим преобразованием данных одного буфера работать с другим буфером памяти, например, выгрузить результаты предыдущего преобразования или загрузить данные для последующего преобразования.

Максимальный размер преобразования равен 4096. Номер буфера с преобразуемым массивом задается старшим битом начального адреса массива, который программируется в регистрах устройства. При неработающем БПФ-сопроцессоре вся его память доступна внешним устройствам.

Данные, подлежащие обработке, загружаются в БПФ-сопроцессор в прямом порядке. Результаты обработки выгружаются также в прямом порядке. Действительным компонентам соответствуют нечетные адреса памяти, а квадратурным (мнимым) – четные адреса памяти.

6) Тригонометрические коэффициенты формируются во встроенных блоках. Они обеспечивают как фазовые повороты выполняемого Фурье-преобразования, так и матричные фазовые повороты для матричного наращивания размера преобразования. Предельный размер наращиваемого преобразования – 256К.

7) За один такт выполняются максимально 40 арифметических операций с плавающей точкой (24 сложения / вычитания и 16 умножений). Например, при тактовой частоте 160 МГц производительность БПФ-сопроцессора составит 6400 Мфлоп / с.

Блок FFT_RT_Core спроектирован для выполнения нескольких операций. Основными операциями являются прямое и обратное преобразование Фурье. Размер комплексного массива $N = 2^n$, $n = 4:12$. Порядок следования отсчетов на входе и на выходе – прямой. Логарифмический размер преобразования задается в регистре устройства.

Также блок FFT_RT_Core способен выполнять другие операции на том же оборудовании. Ниже приведены сопутствующие операции и их описание:

1) Нормировка результатов преобразования. Мультипликативный коэффициент нормировки – константа в формате плавающей точки задается в регистре устройства. Выполняется как при обратном, так и при прямом преобразовании.

2) Фазовые матричные повороты результатов преобразования. Предназначены для матричного наращивания размера преобразования до величины, превышающей 4096. Предельный размер наращиваемого преобразования – 256К.

Выполняется как при прямом, так и при обратном преобразовании. Преобразуемый массив может быть только один. Логарифмический размер строки матрицы наращивания и текущий номер преобразуемого столбца задаются в регистрах устройства.

3) Расчет мощностей результатов преобразования. На месте действительных компонент формируется сумма квадратов обеих компонент комплексных выходов, на

месте мнимых размещаются нули. Выполняется как при прямом, так и при обратном преобразовании.

4) Одновременное преобразование $M = 2m$ комплексных массивов. Размер каждого из преобразуемых массивов равен $N = 2n$, $N \cdot M \leq 4096$. Логарифмические размеры массивов и их логарифмическое число задаются параметрами в регистрах устройства. Массивы размещаются последовательно, без разрывов.

Выполняются все сопутствующие операции, кроме матричного наращивания размера преобразования.

5) Поэлементное перемножение (с сопряжением) двух комплексных массивов. Перемножаемые массивы должны располагаться в разных буферах внутренней памяти. Результирующий массив, расположен на месте одного из преобразуемых массивов.

6) Быстрая БПФ–свертка (ковариация). Последовательно при одном запуске БПФ-сопроцессора выполняются три процедуры:

- прямое БПФ;
- перемножение результата преобразования на комплексный массив частотной характеристики (с сопряжением последнего);
- обратное БПФ.

Результат перемножения и выходной массивы располагаются на месте входного (преобразуемого) массива. Входной преобразуемый массив и массив частотной характеристики фильтра должны располагаться в разных буферах внутренней памяти БПФ-сопроцессора.

Оба входных массива загружаются в прямом порядке, выходной выгружается также в прямом порядке.

7) Вычисление скалярного произведения комплексных массивов. Выполняется поэлементное перемножение (с сопряжением) двух комплексных массивов, частичные произведения накапливаются в комплексном аккумуляторе устройства. Перемножаемые массивы должны располагаться в разных буферах внутренней памяти. Результирующий массив расположен на месте одного из преобразуемых массивов.

Е. Обеспечение высокой производительности

Побочным эффектом выполнения быстрого преобразования Фурье является перемешивание выходных данных относительно исходного прямого порядка входных данных.

Во многих сторонних аппаратных ускорителях для решения проблемы доступа к выходным данным по прямым адресам предлагается модуль, который автоматически восстанавливает прямой порядок выходных данных.

Как можно увидеть из [3], на данную операцию затрачивается некоторое количество тактов (времени),

что естественно сказывается на общей производительности ускорителя.

Другим решением является установка дополнительной памяти для размещения в ней выходных результатов в прямом порядке, что увеличивает площадь устройства.

В FFT_RT_Core предусмотрена система преобразования адресов входных, промежуточных и выходных данных, которая автоматически размещает входные данные в памяти таким образом, что дальнейшая обработка может производиться, не выходя за пределы данной памяти.

Для обеспечения высокой производительности требуется широкий доступ к памяти. Так как FFT_RT_Core выполняет преобразование Фурье по схеме Radix-4, из памяти необходимо загружать 4 комплексных числа за один такт. В связи с этим память была разбита на 4 банка памяти по 64 разряда каждый. Чтобы обеспечить исполнение преобразования «на месте» был предусмотрен механизм расположения данных в памяти таким образом, что данные по запросу на чтение всегда располагаются в разных банках.

При чтении результатов из внутренней памяти ускорителя модуль преобразования адресов автоматически конвертирует адреса запроса и выдает результат в прямом порядке.

Данная процедура позволяет сократить накладные расходы на перекладку данных из инверсного порядка в прямой, не увеличивая объем памяти и не затрачивая на это дополнительное время.

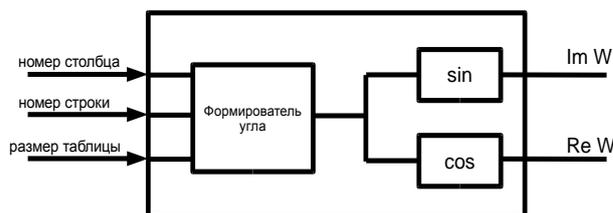


Рис. 5. Модуль формирования поворачивающих коэффициентов

В IP-блок аппаратного ускорителя быстрого преобразования Фурье был внедрен модуль формирования поворачивающих коэффициентов.

Структурная схема модуля формирования коэффициентов приведена на Рис. 5.

Данный модуль позволил сократить некоторое количество памяти, предназначенной для хранения коэффициентов, и организовать матричное наращивание преобразования Фурье (в текущей реализации до 256К).

Вычислительные ядра модуля формирования коэффициентов содержат следующие функциональные блоки:

- Формирователь угла.
- Вычислитель синуса угла.
- Вычислитель косинуса угла.

В формирователь поворачивающих коэффициентов от блока генератора адреса, который является организатором всех транзакций с памятью, подаются параметры для формирования угла, такие как номер столбца, номер строки, размер таблицы. Значение этих параметров напрямую зависит от текущей стадии исполнения операции БПФ/оБПФ.

Угол формируется в диапазоне от 0 до 2π разрядностью 18 бит, 16 младших которого идут непосредственно на вычисление синуса и косинуса в целях экономии аппаратуры вычислителей, а 2 старших бита идут на уточнение результата (определение квадранта). После вычисления косинуса и синуса угла результат переводится в формат плавающей точки (IEEE 754) и поступает на вход основного вычислительного модуля аппаратного ускорителя БПФ.

Таблица 1

Сводная таблица характеристик блока аппаратного ускорителя быстрого преобразования Фурье

Параметр	Значение
Размер преобразования Фурье	Локально: 16...4096; В режиме матричного наращивания: до 256К;
Размер поэлементного перемножения	до 4096
Размер быстрой свертки	16, 64, 256, 1024, 4096
Производительность	40 флоп/такт, на частоте 160 МГц - 6,4 Гфлоп/с
Быстродействие	БПФ размера N: $(\log_4 N) * N/4$ тактов; Быстрая свертка размера N: $2 * (\log_4 N) * N/4$ тактов; Поэлементное перемножение размера N: $N/4$ тактов;
Формат входных и выходных данных	32-разрядная плавающая точка (IEEE-754); 32-разрядная фиксированная точка; 16-разрядная фиксированная точка;
Формат вычислений	32-разрядная плавающая точка (IEEE-754);
Технология, нм	180 нм
Тактовая частота, МГц.	160 МГц
Площадь, мм ²	8 мм ²

Таблица 2

Таблица сравнения характеристик

	FFT_RT_Core	Xilinx LogicCore	Texas Instruments HWAFFT	Altera MegaCore
N	1К	1К	1К	1К
Количество тактов	1360*	3722*	5244+2071*	1024+1024*
Время исполнения	8.5 мкс	25.5 мкс	73 мкс	17 мкс
Max N	256К	64К	1К	4096
Дополнительные вычисления	Перемножение массивов, нормировка результата, мощность, Быстрая свертка	Нормировка результата	Нормировка результата	Нормировка результата

* - в данную цифру входят количество тактов, необходимых для восстановления прямого порядка результатов.

IV. СРАВНИТЕЛЬНЫЙ АНАЛИЗ

Сводная таблица характеристик FFT_RT_Core приведена в Таблица 3. Время исполнения отдельных операций приведено в Таблица 3.

Таблица 3

Время исполнения операции БПФ/обратное БПФ

Размер преобразования (число отсчетов)	Время преобразования (такты)	Время преобразования (мкс)
4096	6240	38.4 мкс
2048	3168	19.2 мкс
1024	1360	8.5 мкс
512	720	4.5 мкс
256	320	2 мкс *
128	192	1.2 мкс *
64	96	0.6 мкс *
32	72	0.3 мкс *
16	40	0.15 мкс *

* - время указано для неоптимизированного варианта.

Оптимизацией является загрузка нескольких массивов во внутреннюю память сопроцессора и запуск его в режиме исполнения БПФ для нескольких массивов. За счет этого удается выиграть время на загрузку конвейера, соответственно увеличить производительность устройства.

В Таблица 4. приведено время исполнения дополнительных операций.

Таблица 4

Время исполнения дополнительных операций

Размер преобразования (число отсчетов)	Время преобразования (такты)	Время преобразования (мкс)
Время выполнения преобразования большого размера		
65536	133120	19.2 мкс
Время поэлементного перемножения двух комплексных массивов		
4096	1032	6.5 мкс
Время вычисления БПФ – свертки комплексного массива (пример)		
4096	12480	77 мкс

Сводная таблица сравнительных характеристик приведена в Таблица 1.

В качестве IP-блоков для проведения сравнительного анализа были выбраны следующие реализации быстрого преобразования Фурье:

- Texas Instruments HWFFT [3],
- Xilinx LogicCore FFT [1],
- Altera MegaCore FFT [2].

Таблица сравнения временных характеристик приведена в Таблица 2. Тактовая частота – 160 МГц.

V. ЗАКЛЮЧЕНИЕ

Оценивая эффективность выбранной системы объединения аппаратных ускорителей в кластер с общей памятью необходимо учитывать особенности аппаратных ускорителей, требуемую пропускную способность внутренней и внешней памяти, способ загрузки и выгрузки данных. В дальнейшем в блок ускорителей будет встроена многоканальная DMA-контроллер и организована система автоматической загрузки и выгрузки данных. Также в будущих реализациях система аппаратных ускорителей пополнится новыми разработками, такими как кодер/декодер Viterbi, RS, BCH, блоками шифрования AES, DES и т.д.

С помощью предложенных методов ускорения вычислений быстрого преобразования Фурье, таких как: схема вычисления Radix-4, преобразование адреса при чтении/записи данных во внутреннюю память устройства, удалось достичь следующих результатов:

- Повышение быстродействия.
- Автономность вычислений.
- Возможность исполнения сопутствующих операций.
- Прямой порядок входных и выходных данных без затрат дополнительной логики, памяти и времени.

ЛИТЕРАТУРА

- [1] URL: http://www.xilinx.com/support/documentation/ip_documentation/xfft/v9_0/pg109-xfft.pdf.
- [2] URL: www.altera.com/literature/ug/ug_fft.pdf.
- [3] URL: www.ti.com/lit/an/sprabb6b/sprabb6b.pdf