Санкт-Петербургский Политехнический Университет Петра Великого

Институт компьютерных наук и кибербезопасности

Высшая школа компьютерных технологий и информационных систем

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ Lab7\_Z4

Дисциплина: Проектирование реконфигурируемых гибридных вычислительных систем

Тема: Optimizing for Area and Logic

Выполнил студент гр. Шеметов С.А.

гр.5140901/21502

Руководитель, доцент Антонов А.П.

«\_\_» \_\_\_\_\_\_ 2023

Санкт-Петербург

2023

Оглавление

[\_Toc152424039](#_Toc152424039)[1. Задание 3](#_Toc152424040)

[2. Исходный код функции 3](#_Toc152424041)

[3. Исходный код теста 4](#_Toc152424042)

[3.1. Моделирование 4](#_Toc152424043)

[4. Создание скрипта 6](#_Toc152424044)

[5. Исследование решений 7](#_Toc152424045)

[5.1. Решение 1 7](#_Toc152424046)

[5.1.1. Синтез 7](#_Toc152424047)

[5.2. Решение 1\_1 8](#_Toc152424048)

[5.2.1. Синтез 8](#_Toc152424049)

[5.3. Решение 2 9](#_Toc152424050)

[5.3.1. Синтез 9](#_Toc152424051)

[5.3.2. C/RTL моделирование 10](#_Toc152424052)

[6. Сравнение результатов решений 11](#_Toc152424053)

[6.1. Сравнение временных параметров, производительности и аппаратных затрат 11](#_Toc152424054)

[6.2. Анализ результатов 12](#_Toc152424055)

[7. Измерение времени выполнения 13](#_Toc152424056)

[7.1. Измерение времени выполнения на ПК 13](#_Toc152424057)

[7.2. Измерение времени выполнения на аппаратной реализации 16](#_Toc152424058)

[7.3. Сравнение результатов 16](#_Toc152424059)

[8. Вывод 17](#_Toc152424060)

1. Задание

* Создать на языке С++ функцию (lab7\_z4.cpp), выполняющую умножение двух массивов поэлементно. Тип данных ap\_int<256>.
* Создать на языке С++ тест (lab7\_z4\_test.cpp) для проверки работы функции.
* Создать скрипт автоматизирующий процесс создания проекта и работы с ним.
* Создать, провести исследование и сравнительный анализ трёх решений.
  + Микросхема: xc7a100tcsg324-2;
  + clock period 10; clock\_uncertainty 1;
  + Solution1:
    - Выключить конвейеризацию цикла;
  + Solution1\_1:
    - Выключить конвейеризацию цикла;
    - Для переменной temp\_mult задать директиву BIND\_OP: op=mul, impl=fabric.
  + Solution2:
    - Задать максимально возможный Unroll Factor цикла MULT (чтобы использовались только DSP);
    - Использовать Array Partition (или Array Reshape) нужного типа и фактора, обеспечивающего балансировку производительности умножителей и чтения/записи данных (использование одно портовой или двух портовой памяти на выбор);
    - Включить конвейеризацию цикла;
    - Для переменной temp\_mult задать директиву BIND\_OP: op=mul, impl=dsp, latency=1.
  + Синтез.
  + С/RTL cosimulation.
* Измерить время выполнения на ПК (однопоточная и многопоточная реализация) и на аппаратной реализации для Solution2.

2. Исходный код функции

Исходный код синтезируемой функции приведен на рисунке 1. Код заголовочного файла указан на рисунке 2.

*Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание*

*Рис.1. Исходный код функции* *lab7\_z4*

Изображение выглядит как текст, снимок экрана, Мультимедийное программное обеспечение, программное обеспечение

Автоматически созданное описание

*Рис.2. Исходный код заголовочного файла lab7\_z4*

Функция принимает три вектора типа ap\_int<256> (длинные целые числа) и размера N=128: два входных и один выходной. Затем в цикле перемножаются элементы входных массивов и результат записывается в выходной массив.

3. Исходный код теста

Исходный код теста для проверки функции lab7\_z4, исправленный в соответствии с заданием, приведен на рисунке 3. Тест обеспечивает: запуск функции 3 раза; формирование входных массивов; очистку выходного массива между запусками функции; проверку правильности вычисленного результата и формирование признака успешного/неуспешного выполнения для каждого запуска функции.

3.1. Моделирование

Результаты моделирования исходного кода синтезируемой функции приведены на рисунке 4. Результаты моделирования показывают, что тест успешно пройден — так как условия теста выполняются (функция *main* возвращает 0) и результат расчета совпадает с ожидаемыми значениями.

Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

*Рис.3. Исходный код теста*

*Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание*

*Рис.4. Лог результата выполнения моделирования*

Для проверки правильности работы теста в него было внесено изменение в ожидаемый результат - expted\_A[j] = in\_B[j] \* in\_C[j] + 1. Результат моделирования приведён на рисунке 5. Он показывает, что тест отрабатывает ошибку корректно.

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

*Рис.5. Лог результата при ошибке в тесте*

4. Создание скрипта

Созданный скрипт для автоматизации работы с проектом представлен на рисунке 6.

Данный скрипт создаёт проект, добавляет в него файл с исходным кодом и тестовый файл, а также устанавливает функцию верхнего уровня. Затем по очереди создаются 3 решения. Для каждого из них задаётся одинаковая микросхема, период 10 нс и неопределённость 1 нс.

Первое решение не имеет директив оптимизации и в нём отключена конвейеризация. В sol1\_1 конвейреизация также отключена и используется директива BIND\_OP с параметрами mul (умножение целых чисел), fabric (связывание операции с логическими элементами).

В решении sol2 используются директивы UNROLL и ARRAY\_PARTITION для обеспечения балансировки производительности умножителей и чтения/записи данных, включена конвейеризация и используется директива BIND\_OP с параметрами mul (умножение целых чисел), dsp (связывание операции с умножителями DSP) и latency (задержка).

Запуск скрипта выполняется в командной строке Vitis HLS 2021.2 Command Prompt с помощью команды *Vitis\_hls -f lab7\_z4.tcl*.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

*Рис.6. Скрипт для работы с проектом*

5. Исследование решений

5.1. Решение 1

5.1.1. Синтез

Решение sol1 не имеет директив для оптимизации и в нём отключена конвейеризация для цикла Mult. Временные параметры, производительность и аппаратные затраты решения представлены на рисунке 7.

Изображение выглядит как снимок экрана, текст, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

*Рис.7. Результаты синтеза для sol1*

Estimated timing для решения составило 8.609 нс. Для реализации потребовалось 55 модулей умножения DSP.

Результаты планирования для sol1 представлены на рисунке 8.

Изображение выглядит как снимок экрана, текст, Мультимедийное программное обеспечение, Графическое программное обеспечение

Автоматически созданное описание

*Рис.8. Результаты планирования для sol1*

Как можно видеть на рисунке 8, в решении sol1 итерация цикла длится 5 тактов, операция умножения занимает 1 такт.

5.2. Решение 1\_1

5.2.1. Синтез

В решении sol1\_1 используется директива bind\_op с параметрами op=mul, impl=fabric для переменной temp\_mult. Эта директива позволяет связать запланированную операцию умножения целых с логическими элементами. Временные параметры, производительность и аппаратные затраты решения представлены на рисунке 9.

Изображение выглядит как программное обеспечение, Мультимедийное программное обеспечение, текст, Графическое программное обеспечение

Автоматически созданное описание

*Рис.9. Результаты синтеза для sol1\_1*

Как можно видеть на рисунке 9, в результате синтеза возникла ошибка Timing Violation. Такая ошибка возникает из-за того, что Estimated timing (24.261 нс) превысило заданное время Target (10 нс).

Для решения данной проблемы можно увеличить период, чтобы время выполнения в него укладывалось, либо связывать операцию умножения с DSP (impl=dsp) как в решении sol2.

5.3. Решение 2

5.3.1. Синтез

В решении sol2 используется директива unroll с factor=4 для того, чтобы использовались только DSP. Также вместе с unroll используется директива array\_partition factor=4 и cyclic. Для такого решения по умолчанию будет применяться однопортовая память. Директива bind\_op используется с параметрами op=mul, impl=dsp, latency=1 для переменной temp\_mult. Включена конвейеризация. Временные параметры, производительность и аппаратные затраты решения представлены на рисунке 10.

Изображение выглядит как текст, Мультимедийное программное обеспечение, программное обеспечение, Графическое программное обеспечение

Автоматически созданное описание

*Рис.10. Результаты синтеза для sol2*

Estimated timing для решения составило 8.232 нс. Для реализации потребовалось 220 модулей умножения DSP.

Результаты планирования для sol2 представлены на рисунке11.

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

*Рис.11. Результаты планирования для sol2*

Как можно видеть на рисунке 11, итерация цикла занимает 3 такта. Операция умножения занимает 1 такт. Параллельно считываются 4 значения b и 4 значения c. Также 4 операции умножения выполняются параллельно и их результат записывается в 4 элементов массива a. В данном цикле была выполнена конвейеризация, что видно также видно на рисунке. II составляет 35 тактов.

5.3.2. C/RTL моделирование

Временная диаграмма для sol2 представлена на рисунке 12.

Изображение выглядит как снимок экрана, текст, электроника, программное обеспечение

Автоматически созданное описание

*Рис.12. Временная диаграмма C/RTL для sol2*

На временной диаграмме видно, что каждый из массивов b и c был разбит на 4 блока по 32 элемента. На каждом такте из блока считывается по одному элементу и выполняется их перемножение.

6. Сравнение результатов решений

6.1. Сравнение временных параметров, производительности и аппаратных затрат

В данной работе были реализованы 3 решения с разными параметрами bind\_op. Результаты сравнения приведены на рисунке 13.

Изображение выглядит как текст, снимок экрана, диаграмма, Шрифт

Автоматически созданное описание

*Рис.13. Сравнение всех решений*

6.2. Анализ результатов

На основе результатов сравнения решений была составлена электронная таблица и построен график, в котором для всех решений отображены: Latency, II и использованные ресурсы. Результаты представлены на рисунках 14–15.

Изображение выглядит как текст, число, снимок экрана, Шрифт

Автоматически созданное описание

*Рис.14. Электронная таблица для всех решений*

Изображение выглядит как линия, График, диаграмма, снимок экрана

Автоматически созданное описание

*Рис.15. График для всех решений*

С точки зрения производительности оптимальным решением является sol2, так как оно имеет наименьшие значения Latency и затрачивает меньше времени. Это обусловлено использованием директив bind\_op, unroll, array\_partition и конвейеризации.

7. Измерение времени выполнения

7.1. Измерение времени выполнения на ПК

Измерение времени выполнения синтезируемой функции было выполнено на ПК со следующими характеристиками:

|  |  |  |  |
| --- | --- | --- | --- |
| **Тип процессора** | **Базовая частота работы** | **Максимальная частота работы** | **ОЗУ** |
| 12th Gen Intel(R) Core(TM) i5-12450H | 2 ГГц | 4,4 ГГц | 16 Гб |

На базе теста lab7\_z4\_test.cpp был создан отдельный, модернизированный, тест lab7\_z4\_testSW.cpp, который содержит операторы измерения времени выполнения и имеет большее число запусков равное 32. Код теста представлен на рисунке 16.

Изображение выглядит как текст, снимок экрана

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

*Рис.16. Код теста* *lab7\_z4\_testSW.cpp*

Также в исходный код функции для реализации многопоточности была добавлена pragma omp со значением 8, что соответствует числу ядер на ПК (рисунок 17).

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

*Рис.17. Исходный код многопоточной функции lab7\_z4*

Также для реализации на ПК больших чисел был использован класс BigInt (рисунок 18).

Изображение выглядит как текст, Шрифт, линия, снимок экрана

Автоматически созданное описание

*Рис.18. Код заголовочного файла lab7\_z4*

Для компиляции теста был установлен компилятор GCC версии 13.2.0 (рисунок 19). После этого была выполнена компиляции была выполнена на одном ядре и на 8. Запуск теста проводился для четырёх значений N – 8192, 16384, 32768, 65536.

Изображение выглядит как текст, снимок экрана, Шрифт, черный

Автоматически созданное описание

*Рис.19. Выполнение компиляции теста*

Результаты запуска для одного потока.

|  |  |  |  |
| --- | --- | --- | --- |
| **N** | **Минимальное время, нс** | **Максимальное время. нс** | **Медиана, нс** |
| 8192 | 4.146e+06 | 4.4545e+06 | 4.3025e+06 |
| 16384 | 8.123e+06 | 8.99583e+06 | 8.69845e+06 |
| 32768 | 1.5595e+07 | 1.83685e+07 | 1.6904e+07 |
| 65536 | 3.17447e+07 | 3.4292e+07 | 3.25034e+07 |

Результаты запуска для 8 потоков.

|  |  |  |  |
| --- | --- | --- | --- |
| **N** | **Минимальное время, нс** | **Максимальное время, нс** | **Медиана, нс** |
| 8192 | 4.14067e+06 | 4.70225e+06 | 4.55589e+06 |
| 16384 | 8.241e+06 | 9.0775e+06 | 8.73707e+06 |
| 32768 | 1.5293e+07 | 1.7138e+07 | 1.69887e+07 |
| 65536 | 3.0129e+07 | 3.36242e+07 | 3.20115e+07 |

7.2. Измерение времени выполнения на аппаратной реализации

Для измерения времени необходимо использовать решение sol2.

Синтез решения sol2 был осуществлён для четырёх значений N – 8192, 16384, 32768, 65536. Время выполнения = II \* Estimated period для каждого случая:

|  |  |  |  |
| --- | --- | --- | --- |
| **N** | **II, такт** | **Estimated period, нс** | **Время выполнения, нс** |
| 8192 | 8.232 | 2052 | 16892.064 |
| 16384 | 8.232 | 4100 | 33751.2 |
| 32768 | 8.232 | 8196 | 67469.472 |
| 65536 | 8.232 | 16388 | 134906.016 |

7.3. Сравнение результатов

На основе результатов, полученных при измерениях времени выполнения на ПК и аппаратной реализации, была составлена электронная таблица и построен график, в котором отражено соотношение размера массивов данных задачи (N) ко времени выполнения. Результаты представлены на рисунках 20–21.

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

*Рис.20. Электронная таблица для ПК (1 и 8 ядер) и аппаратной реализации*

Изображение выглядит как снимок экрана, График, линия, текст

Автоматически созданное описание

*Рис.21. График для ПК (1 и 8 ядер) и аппаратной реализации*

На графике можно видеть, что результаты аппаратной реализации растут медленнее и значительно выигрывают по производительности у обеих реализаций на ПК. Это связанно с тем, что в аппаратной реализации выполнение цикла было распараллелено, а также использована конвейеризация и директива bind\_op. При этом многопоточная реализация не даёт преимуществ перед однопоточной, что связано с реализацией больших чисел на ПК.

8. Вывод

В результате выполнения данной лабораторной работы были получены навыки работы с директивой bind\_op, которая позволяет связать запланированную операцию с имеющимися аппаратными ресурсами. Для этого была создана программа на языке C++, выполняющая поэлементное умножение двух массивов размером N и типом ap\_int<256>.

Для создания проекта и работы с ним был создан TCL скрипт, содержащий 2 решений. Первое решение является базовым, в нём отсутствуют директивы оптимизации и отключена конвейеризация. В решении sol1\_1 также отключена конвейеризация и использована директива bind\_op, которая связывает операцию умножение с логическими элементами. Данное решение не укладывается в заданный временной период и вызывает проблему Timing Violation, для решения которой следует увеличить период или изменить аппаратные ресурсы в bind\_op. В решении sol2 помимо bind\_op были использованы директивы unroll и array\_partition для распараллеливания цикла и конвейеризация. Оптимальным с точки зрения производительности оказалось sol2.

Также было выполнено измерение времени выполнения программы на для однопоточной и многопоточной реализации на ПК и аппаратной реализации для sol2. Результаты показали, что аппаратная реализация значительно выигрывает по производительности у ПК за счёт оптимизации. При этом использование многопоточной реализации на ПК не даёт преимуществ перед однопоточной.