

Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

«Московский государственный технический университет имени Н.Э. Баумана

(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ	«Информатика и системы управления»
КАФЕДРА	«Компьютерные системы и сети (ИУ6)»
НАПРАВЛЕНИЕ ПОДГОТОВКИ _	«09.03.04 Программная инженерия»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №5 по курсу «Архитектура ЭВМ»

«Разработка ускорителей вычислений средствами САПР высокоуровневого синтеза Xilinx Vitis HLS»

Студент:	<u>ИУ7-53Б</u> (группа)	(подпись, дата)	<u>М. Д. Маслова</u> (И. О. Фамилия)
Преподаватель:		(подпись, дата)	Е. Н. Дубровин (И. О. Фамилия)

Содержание

Bı	Введение					
1	Осн	овные	теоретические сведения	4		
	1.1	Метод	дология ускорения вычислений на основе ПЛИС	4		
	1.2	Оптим	мизация времени обработки и пропускной способности .	5		
		1.2.1	Конвейерная обработка циклов	5		
		1.2.2	Разворачивание циклов	6		
		1.2.3	Потоковая обработка	6		
2 Практическая часть		7				
3	Кон	трольн	ные вопросы	16		
3a	ключ	нение		17		

Введение

Целью данной работы является изучение методики и технологии синтеза аппаратных устройств ускорения вычислений по описаниям на языках высокого уровня.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- рассмотреть маршрут проектирования устройств, представленных в виде синтаксических конструкций ЯВУ С/С++;
 - изучить принципы работы IDE Xilinx Vitis HLS;
 - изучить методику анализа и отладки устройств;
 - разработать ускоритель вычислений по индивидуальному заданию;
 - разработать код для тестирования ускорителя;
- реализовать ускоритель с помощью средств высоко-уровненного синтеза;
 - выполнить его отладку.

1 Основные теоретические сведения

В данном разделе описаны методология ускорения вычислений на основе ПЛИС и способы оптимизации времени обработки и пропускной способности на ускорителе вычислений **Xilinx Alveo**.

1.1 Методология ускорения вычислений на основе ПЛИС

Ускорение вычислительных алгоритмов с использованием программируемых логических интегральных схем (ПЛИС) имеет ряд преимуществ по сравнению с их реализацией на универсальных микропроцессорах, или графических процессорах. В то время, как традиционная разработка программного обеспечения связана с программированием на заранее определенном наборе машинных команд, разработка программируемых устройств — это создание специализированной вычислительной структуры для реализации желаемой функциональности.

Микропроцессоры и графические процессоры имеют предопределенную архитектуру с фиксированным количеством ядер, набором инструкций, и жесткой архитектурой памяти, и обладают высокими тактовыми частотами и хорошо сбалансированной конвейерной структурой. Графические процессоры масштабируют производительность за счет большого количества ядер и использования параллелизма SIMD/SIMT (рисунок 1.1). В отличие от них, программируемые устройства представляют собой полностью настраиваемую архитектуру, которую разработчик может использовать для размещения вычислительных блоков с требуемой функциональностью. В таком случает, высокий уровень производительности достигается за счет создания длинных конвейеров обработки данных, а не за счет увеличения количества вычислительных единиц. Понимание этих преимуществ является необходимым условием для разработки вычислительных устройств и достижения наилучшего уровня ускорения.

Методологию создания ускорителей на ПЛИС с применением средств синтеза высокого уровня (High Level Synthesis, HLS) можно представить в виде трех этапов:

Микропроцессор СРИ — func1 — func2(a) — func2(b) — func2(c) — func3 Графический процессор ПЛИС СРИ — func1 — func2 — func3 Func2(a) — func2 — func3 FPGA — func2 — func3 — func

Рисунок 1.1 – Принципы организации вычислений на различных платформах

- создание архитектуры приложения;
- разработка ядра аппаратного ускорителя на языках C/C++;
- анализ производительности и выявление способов ее повышения.

1.2 Оптимизация времени обработки и пропускной способности

Существует несколько подходов к оптимизации.

1.2.1 Конвейерная обработка циклов

Полагая, что одна итерация цикла занимает более одного такта (фаза выборки данных, фаза вычисления, фаза записи результата) в таком варианте может быть организован конвейер выполнения. Для этого необходимо поместить в тело цикла прагму <HLS PIPELINE>.

Без конвейерной обработки каждая последующая итерация цикла начинается в каждом третьем такте. При конвейерной обработке цикл может начинать последующие итерации цикла менее чем за три такта, например, в каждом втором такте или в каждом такте. Для указания конкретного значения интервала запуска заданий используется свойство Initiation inteval (II). Например: #pragma HLS PIPELINE II = 1

Стоит отметить, что конвейерная обработка цикла приводит к разворачиванию любых циклов, вложенных внутрь конвейерного цикла. Если внутри

цикла существуют зависимости по данным, может оказаться невозможным достичь запуска новой итерации в каждом такте, и результатом может быть больший интервал инициации.

1.2.2 Разворачивание циклов

Разворачивание циклов является общепризнанным механизмом снижения времени выполнения циклов. Этот механизм может быть описан самим разработчиком вручную, если он просто повторит вычисления и сократит количество итераций цикла. С помощью прагмы <HLS UNROLL> компилятор v++ позволяет запустить механизм автоматического разворачивания цикла, частично или полностью.

1.2.3 Потоковая обработка

Реализация механизма потоковой обработки (#pragma HLS DATAFLOW) также опирается на представление вычислительных действий в виде многостадийного конвейера. Однако, в то время как директива конвейеризации (#pragma HLS PIPELINE) используется для реализации конвейера выполнения операций внутри функций или циклов, потоковая обработка данных позволяет сформировать конвейер из более крупных вычислительных блоков: нескольких функций или нескольких последовательных циклических конструкций. Таким образом, директива HLS DATAFLOW позволяет сформировать вычислительный конвейер на уровне задач. Для этих целей компилятор HLS Vitis выполнит анализ зависимостей по данным между задачами и реализует структуру обрабатывающих блоков и FIFO очередей между ними.

2 Практическая часть

Листинг 2.1 – Программа варианта без оптимизаций

```
1 extern "C"
2 {
3
      void var012_no_pragmas(int* c, const int* a, const int* b, const int len)
4
5
           int ptr = 0;
6
           for (int i = 0; i < len; i++)
7
8
               ptr = b[i] % len;
9
               c[i] = a[ptr] + i;
10
           }
11
       }
12 }
```

Листинг 2.2 – Программа варианта с конвейерной организацией

```
1 extern "C"
2 {
3
      void var012_pipelined(int* c, const int* a, const int* b, const int len)
 4
       {
5
           int ptr = 0;
6
           for (int i = 0; i < len; i++)
7
8
               #pragma HLS PIPELINE
9
               ptr = b[i] % len;
10
               c[i] = a[ptr] + i;
11
           }
12
       }
13 }
```

Листинг 2.3 – Программа варианта с разворачиванием цикла

```
1 extern "C"
2 {
3
      void var012_unrolled(int* c, const int* a, const int* b, const int len)
4
      {
5
           int ptr = 0;
           for (int i = 0; i < len; i++)
6
7
8
               #pragma HLS UNROLL
9
               ptr = b[i] % len;
10
               c[i] = a[ptr] + i;
```

Листинг 2.3 (продолжение)

Листинг 2.4 – Программа варианта с конвейерной организацией и

разворачиванием

```
1 extern "C"
2 {
3
      void var012_pipe_unroll(int* c, const int* a, const int* b, const int len)
4
5
           int ptr = 0;
6
           for (int i = 0; i < len; i++)
7
8
               #pragma HLS DATAFLOW
9
               ptr = b[i] % len;
10
               c[i] = a[ptr] + i;
11
           }
12
       }
13 }
```

Листинг 2.5 – Результаты Emulation-SW

```
1 [Console output redirected to file:/iu_home/iu7072/workspace/Alveo_lab2/
    Emulation-SW/SystemDebugger_Alveo_lab2_system_Alveo_lab2.launch.log]
2 Found Platform
3 Platform Name: Xilinx
4 INFO: Reading /iu_home/iu7072/workspace/Alveo_lab2_system/Emulation-SW/
    binary_container_1.xclbin
5 Loading: '/iu_home/iu7072/workspace/Alveo_lab2_system/Emulation-SW/
    binary_container_1.xclbin'
6 Trying to program device[0]: xilinx_u200_xdma_201830_2
7 Device[0]: program successful!
                         Wall-Clock Time (ns) |
9 | Kernel
                      10
11 | var012_no_pragmas
                      12
13 | var012 urolled
15 | var012 pipelined
                      16
17 | var012_pipe_unroll
18 | | ------
19 Note: Wall Clock Time is meaningful for real hardware execution only, not for
```

Листинг 2.5 (продолжение)

emulation.
20 Please refer to profile summary for kernel execution time for hardware emulation
.
21 TEST PASSED.

Name	Compute Units	Memory	SLR	Protocol Checker	Data Transfer	Execute Profiling	Stall Profiling
▼ 👼 binary_container_1		Auto	Auto		Default (Counters + Trace)	~	
▼ Ø var012_no_pragmas	1	Auto	Auto		Default (Counters + Trace)	~	
▼ 🔡 var012_no_pragmas_1		Auto	Auto		Default (Counters + Trace)	~	I
□ c		Auto			Default (Counters + Trace)	~	
■ a		Auto			Default (Counters + Trace)	~	
■ b		Auto			Default (Counters + Trace)	→	
■ len							
▼ 🚱 var012_pipelined	1	Auto	Auto		Default (Counters + Trace)	~	İ
▼ var012_pipelined_1		Auto	Auto		Default (Counters + Trace)	~	
≪ c		Auto			Default (Counters + Trace)	~	
■ a		Auto			Default (Counters + Trace)	~	
■ b		Auto			Default (Counters + Trace)	~	
len							

Рисунок 2.1 – Копия экрана Assistant View для Emulation-HW (часть 1)

▼ Ø var012_unrolled	1	Auto	Auto	Default (Counters + Trace)	•	
The second secon	1			CORPORATION CONTRACTOR SECURIOR SECURIO		
▼ 🔠 var012_unrolled_1		Auto	Auto	Default (Counters + Trace)	~	
⋖ c		Auto		Default (Counters + Trace)	•	
<□ a		Auto		Default (Counters + Trace)	•	
⋖ b		Auto		Default (Counters + Trace)	•	
<□ len						
▼ Ø var012_pipe_unroll	1	Auto	Auto	Default (Counters + Trace)	•	
▼ 🔡 var012_pipe_unroll_1		Auto	Auto	Default (Counters + Trace)	•	
≪ c		Auto		Default (Counters + Trace)	✓	
<□ a		Auto		Default (Counters + Trace)	•	
⋖ b		Auto		Default (Counters + Trace)	•	
■ len						

Рисунок 2.2 – Копия экрана Assistant View для Emulation-HW (часть 2)

Листинг 2.6 – Результаты Emulation-HW

Листинг 2.6 (продолжение)

```
9 | ------
10 | Kernel
                   Wall-Clock Time (ns) |
11 | | ------
                 1
                              15008592508 |
12 | var012_no_pragmas
13
14 INFO::[ Vitis-EM 22 ] [Time elapsed: 3 minute(s) 38 seconds, Emulation time:
    0.200175 msl
15 Data transfer between kernel(s) and global memory(s)
WR =
    4.000 KB
17 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                  RD = 0.000 \text{ KB}
                                                       WR =
    0.000 KB
18 var012_pipelined_1:m_axi_gmem-DDR[1] RD = 0.000 KB
                                                      WR =
    0.000 KB
19 var012_unrolled_1:m_axi_gmem-DDR[1] RD = 0.301 KB
                                                     WR =
    0.125 KB
20
21 INFO::[ Vitis-EM 22 ] [Time elapsed: 8 minute(s) 39 seconds, Emulation time:
    0.466535 ms]
22 Data transfer between kernel(s) and global memory(s)
WR =
    4.000 KB
24 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                  RD = 0.000 \text{ KB}
                                                       WR =
    0.000 KB
25 var012 pipelined 1:m axi gmem-DDR[1] RD = 0.000 KB
                                                      WR =
    0.000 KB
26 var012_unrolled_1:m_axi_gmem-DDR[1] RD = 3.750 KB
                                                     WR =
    1.812 KB
27
28 INFO::[ Vitis-EM 22 ] [Time elapsed: 13 minute(s) 39 seconds, Emulation time:
    0.771576 msl
29 Data transfer between kernel(s) and global memory(s)
WR =
    4.000 KB
                                  RD = 0.000 \text{ KB}
31 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                                       WR =
    0.000 KB
WR =
    0.000 KB
33 var012_unrolled_1:m_axi_gmem-DDR[1] RD = 7.699 KB
                                                     WR =
    3.812 KB
34
               1
35 | var012 urolled
                       653277479951
36
                   - 1
37 | var012 pipelined
                             24012090705 |
38
39 INFO::[ Vitis-EM 22 ] [Time elapsed: 18 minute(s) 40 seconds, Emulation time:
    1.05055 ms]
```

Листинг 2.6 (продолжение)

```
40 Data transfer between kernel(s) and global memory(s)
41 var012_no_pragmas_1:m_axi_gmem-DDR[1]
                                         RD = 8.000 \text{ KB}
                                                                         WR =
     4.000 KB
42 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                             RD = 2.270 \text{ KB}
                                                                          WR =
      1.133 KB
43 var012_pipelined_1:m_axi_gmem-DDR[1] RD = 8.000 KB
                                                                        WR =
     4.000 KB
44 var012_unrolled_1:m_axi_gmem-DDR[1]
                                           RD = 8.000 \text{ KB}
                                                                       WR =
     4.000 KB
45
46 INFO::[ Vitis-EM 22 ] [Time elapsed: 23 minute(s) 40 seconds, Emulation time:
     1.25502 msl
47 Data transfer between kernel(s) and global memory(s)
48 var012 no pragmas 1:m axi gmem-DDR[1]
                                                                         WR =
     4.000 KB
49|var012_pipe_unroll_1:m_axi_gmem-DDR[1] RD = 4.195 KB
                                                                          WR =
      2.094 KB
50 var012_pipelined_1:m_axi_gmem-DDR[1] RD = 8.000 KB
                                                                        WR =
     4.000 KB
51 var012 unrolled 1:m axi gmem-DDR[1]
                                           RD = 8.000 \text{ KB}
                                                                       WR =
     4.000 KB
52
53 INFO::[ Vitis-EM 22 ] [Time elapsed: 28 minute(s) 46 seconds, Emulation time:
     1.51806 ms]
54 Data transfer between kernel(s) and global memory(s)
55 var012 no pragmas 1:m axi gmem-DDR[1]
                                             RD = 8.000 \text{ KB}
                                                                         WR =
     4.000 KB
56 | var012\_pipe\_unroll\_1:m\_axi\_gmem-DDR[1] RD = 6.672 KB
                                                                          WR =
      3.332 KB
57 var012_pipelined_1:m_axi_gmem-DDR[1] RD = 8.000 KB
                                                                        WR =
     4.000 KB
58 var012_unrolled_1:m_axi_gmem-DDR[1]
                                           RD = 8.000 \text{ KB}
                                                                       WR =
     4.000 KB
59
60 | var012_pipe_unroll
                         1028355914876 |
61
62 Note: Wall Clock Time is meaningful for real hardware execution only, not for
     emulation.
63| Please refer to profile summary for kernel execution time for hardware emulation
64 TEST PASSED.
65 INFO::[ Vitis-EM 22 ] [Time elapsed: 33 minute(s) 46 seconds, Emulation time:
     1.73059 msl
66 Data transfer between kernel(s) and global memory(s)
WR =
     4.000 KB
                                              RD = 8.000 \text{ KB}
68 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                                                          WR =
```

Листинг 2.6 (продолжение)

```
4.000 KB
69 var012_pipelined_1:m_axi_gmem-DDR[1]
                                            RD = 8.000 \text{ KB}
                                                                        WR =
     4.000 KB
70|var012\_unrolled\_1:m\_axi\_gmem-DDR[1] RD = 8.000 KB
                                                                       WR =
     4.000 KB
71
72 INFO::[ Vitis-EM 22 ] [Time elapsed: 35 minute(s) 39 seconds, Emulation time:
     1.79479 ms]
73 Data transfer between kernel(s) and global memory(s)
74 var012_no_pragmas_1:m_axi_gmem-DDR[1]
                                              RD = 8.000 \text{ KB}
                                                                         WR =
     4.000 KB
75 var012_pipe_unroll_1:m_axi_gmem-DDR[1]
                                             RD = 8.000 \text{ KB}
                                                                         WR =
      4.000 KB
76 var012_pipelined_1:m_axi_gmem-DDR[1]
                                            RD = 8.000 \text{ KB}
                                                                        WR =
     4.000 KB
WR =
     4.000 KB
78
79 INFO: [HW-EMU 06-0] Waiting for the simulator process to exit
```

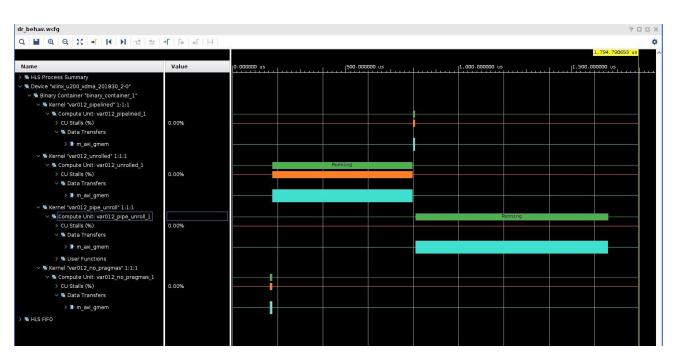


Рисунок 2.3 – Окно внутрисхемного отладчика Vivado для сборки в режиме Emulation-HW

Листинг 2.7 – Результаты Hardware

```
1 [Console output redirected to file:/iu_home/iu7072/workspace/Alveo_lab2/Hardware
    /SystemDebugger_Alveo_lab2_system_Alveo_lab2.launch.log]
2 Found Platform
3 Platform Name: Xilinx
4 INFO: Reading /iu_home/iu7072/workspace/Alveo_lab2_system/Hardware/
    binary_container_1.xclbin
5 Loading: '/iu home/iu7072/workspace/Alveo lab2 system/Hardware/
    binary_container_1.xclbin'
6 Trying to program device[0]: xilinx u200 xdma 201830 2
7 Device[0]: program successful!
9 | Kernel
                        Wall-Clock Time (ns) |
                     10
11 | var012_no_pragmas
                                  45233097 I
                     12
13 | var012_urolled
                     662424184
15 | var012_pipelined
                     16
17 | var012_pipe_unroll
                     -
                                 913769872 I
18
19 Note: Wall Clock Time is meaningful for real hardware execution only, not for
    emulation.
20| Please refer to profile summary for kernel execution time for hardware emulation
21 TEST PASSED.
```

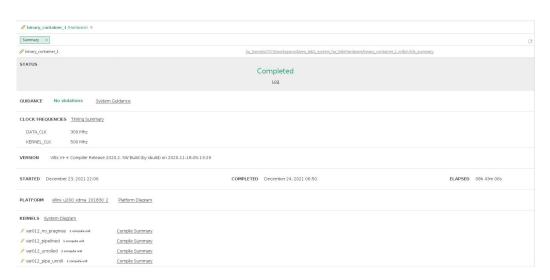


Рисунок 2.4 – Копия экрана вкладки Summary

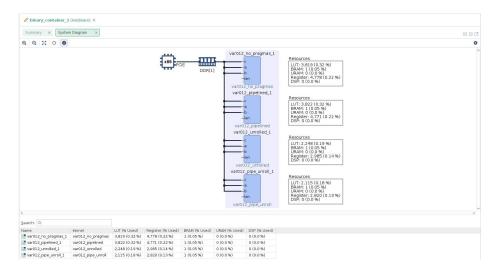


Рисунок 2.5 – Копия экрана вкладки System Diagram

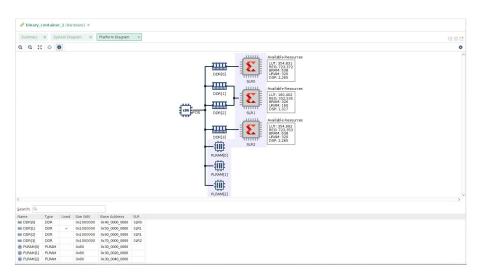


Рисунок 2.6 – Копия экрана вкладки Platform Diagram



Рисунок 2.7 – Копия экрана вкладки HLSSynthesis (без прагм)



Рисунок 2.8 – Копия экрана вкладки HLSSynthesis (конвейер)



Рисунок 2.9 – Копия экрана вкладки HLSSynthesis (разворачивание)



Рисунок 2.10 – Копия экрана вкладки HLSSynthesis (конвейер + разворачивание)

3 Контрольные вопросы

Заключение