Justin Pacella

Lab 5 Report

Trial 1: No optimization

For trial 1, I modified the reference matrix multiplication function so as to fully unroll the dot-product (innermost) loop. My reason for not unrolling the other two loops is that the memory utilization scales exponentially with each nested unrolling. For instance, fully unrolling one loop requires the HLS synthesizer to create DIM (DIM=1024) VHDL blocks to be run concurrently. A second unrolling would require 1024 blocks for each of the existing iterations, for a total of 1024^2 concurrently running blocks. This drastically increases memory utilization, so I decided having only one unrolled loop was a good trade-off. Below is the synthesis result for trial 1.

Synthesis Report for 'HLS_accel'

General Information

Date:	Sun Apr 2 13:20:01 2023
Version:	2018.3 (Build 2405991 on Thu Dec 06 23:56:15 MST 2018)
Project:	lab_5
Solution:	sln1
Product family:	zynq
Target device:	xc7z020clg400-1

Performance Estimates

Timing (ns)

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	10.00	8.372	1.25

Latency (clock cycles)

Summary

Latency		Interval		
min	max	min	max	Type
174664	174664	174664	174664	none

Detail

Instance

Loop

Utilization Estimates

Summary

Name	BRAM_18K	DSP48E	FF	LUT
DSP	-	-	-	-
Expression	-	-	0	865
FIFO	-	-	-	-
Instance	0	8	527	1072
Memory	6	-	0	0
Multiplexer	-	-	-	1784

Trial 2: Full Pipelining

For trial 2, I modified the function from trial 1 to include a pipeline directive in the second (column) loop. This allows the hardware accelerator to concurrently execute certain parts of each loop iteration concurrently. Using an initiation interval of 1 means that a new loop iteration is initiated each clock cycle. This, combined with the unrolling from trial 1 makes it so that the number of clock cycles required to execute the inner two (column & dot-product) loops is little more than the number of iterations specified by DIM. Below is the synthesis result for trial 2.

Synthesis Report for 'HLS_accel'

General Information

Date:	Sun Apr 2 13:32:38 2023
Version:	2018.3 (Build 2405991 on Thu Dec 06 23:56:15 MST 2018)
Project:	lab_5
Solution:	sln1
Product family:	zynq
Target device:	xc7z020clg400-1

Performance Estimates

Timing (ns)

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	10.00	8.317	1.25

Latency (clock cycles)

Summary

Latency		Interval		
min	max	min	max	Type
19615	19615	19615	19615	none

Detail

Instance

Loop

Utilization Estimates

Summary

Name	BRAM_18K	DSP48E	FF	LUT
DSP	-	-	-	-
Expression	-	-	0	3295
FIFO	-	-	-	-
Instance	0	10	732	1462
Memory	6	-	0	0
Multiplexer	-	-	-	1221

Trial 3: Full Pipelining & Array Partitioning

For trial 3, I uncommented the two array_partition directives as instructed. The first line operates on dimension 2 of the 'a' operand and the second on dimension 1 of the 'b' operand. Both apply a partioning factor of DIM/2 (DIM/2=512), so the 'a' operand is divided into 512 subarrays, each with row-by-column shape 1024 by 2. Similarly, the 'b' operand is divided into 512 subarrays with shape 2 by 1024. The synthesized hardware will now process all of the subarrays concurrently, further reducing the expected latency. Below is the synthesis result from trial 3.

Synthesis Report for 'HLS_accel'

General Information

Date: Sun Apr 2 13:41:29 2023
Version: 2018.3 (Build 2405991 on Thu Dec 06 23:56:15 MST 2018)
Project: lab_5
Solution: sln1
Product family: zynq
Target device: xc7z020clg400-1

Performance Estimates

Timing (ns)

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	10.00	8.092	1.25

Latency (clock cycles)

Summary

Latency		Interval		
min	max	min	max	Type
4270	4270	4270	4270	none

Detail

Instance

Loop

Utilization Estimates

Summary

Name	BRAM_18K	DSP48E	FF	LUT
DSP	-	-	-	-
Expression	-	-	0	538
FIFO	-	-	-	-
Instance	0	160	11172	22792
Memory	66	-	0	0
Multiplexer	-	-	-	884