Improving Performance Lab

Introduction

This lab introduces various techniques and directives which can be used in Vivado HLS to improve design performance. The design under consideration accepts an image in a (custom) RGB format, converts it to the Y’UV color space, applies a filter to the Y’UV image and converts it back to RGB.

Objectives

After completing this lab, you will be able to:

* Add directives in your design
* Understand the effect of INLINE directive
* Improve performance using PIPELINE directive
* Distinguish between DATAFLOW directive and Configuration Command functionality

# Procedure

This lab is separated into steps that consist of general overview statements that provide information on the detailed instructions that follow. Follow these detailed instructions to progress through the lab.

This lab comprises 6 primary steps: You will create a new project using Vivado HLS command prompt, analyze the created project and generated solution, turn off inlining and apply the TRIPCOUNT, PIPELINE, and DATAFLOW directives and command configuration, and finally export and implement the design.

# General Flow for this Lab

Step 1:

Create a   
Project using CLI

Step 2:

Analyze Project and Results

Step 3:

Apply TRIPCOUNT Directive

Step 4:

Apply  
PIPELINE Directive

Step 5:

Apply DATAFLOW Directive

Step 6:

Export & Implement the Design

1. Create a Vivado HLS Project from Command Line Step 1
   1. Validate your design using Vivado HLS command line window. Create a new Vivado HLS project from the command line.
      1. Launch Vivado HLS: Select **Start > All Programs > Xilinx Design Tools > Vivado 2015.4 > Vivado HLS > Vivado HLS 2015.4 Command Prompt**.
      2. In the Vivado HLS Command Prompt, change directory to **c:\xup\hls\labs\lab2.**
      3. A self-checking program (yuv\_filter\_test.c) is provided**.** Using that we can validate the design. A Makefile is also provided. Using the Makefile, the necessary source files can be compiled and the compiled program can be executed. You can examine the contents of these files and the project directory. In the Vivado HLS Command Prompt, type **make** to compile and execute the program.

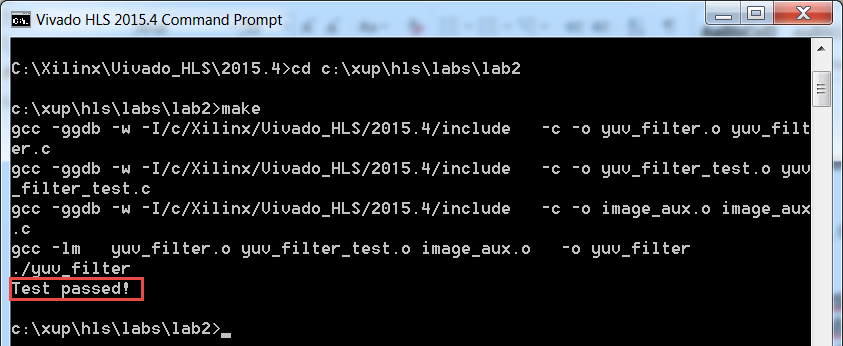


Figure 1. Validating the design

Note that the source files (yuv\_filter.c, yuv\_filter\_test.c, and image\_aux.c) were compiled, then yuv\_filter executable program was created, and then it was executed. The program tests the design and outputs Test Passed message.

* + 1. A Vivado HLS tcl script file (yuv\_filter.tcl) is provided and can be used to create a Vivado HLS project. Type **vivado\_hls –f zed\_yuv\_filter.tcl** in the Vivado HLS Command Prompt window to create the project targeting the ZedBoard or type **vivado\_hls –f zybo\_yuv\_filter.tcl** in the Vivado HLS Command Prompt window to create the project targeting the Zybo.

The project will be created and Vivado HLS.log file will be generated.

* + 1. Open the **vivado\_hls.log** file from **c:\xup\hls\labs\lab2** using any text editor and observe the following sections:
* Creating directory and project called yuv\_filter.prj within it, adding design files to the project, setting solution name as solution1, setting target device (Zynq-z020 for ZedBoard or Zynq-z010 for Zybo), setting desired clock period of 10 ns (for ZedBoard) or 8 ns (for Zybo), and importing the design and testbench files (Figure 2).
* Synthesizing (Generating) the design which involves scheduling and binding of each functions and sub-function (Figure 3).
* Generating RTL of each function and sub-function in SystemC, Verilog, and VHDL languages (Figure 4).

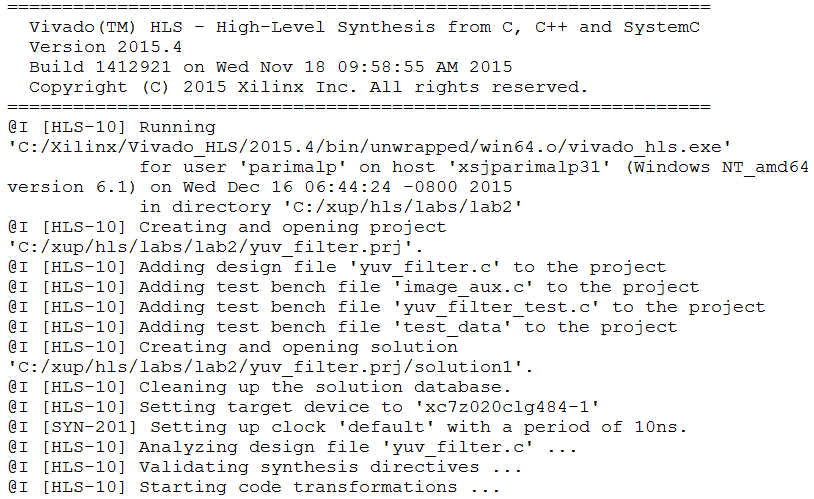


Figure 2. Creating project and setting up parameters

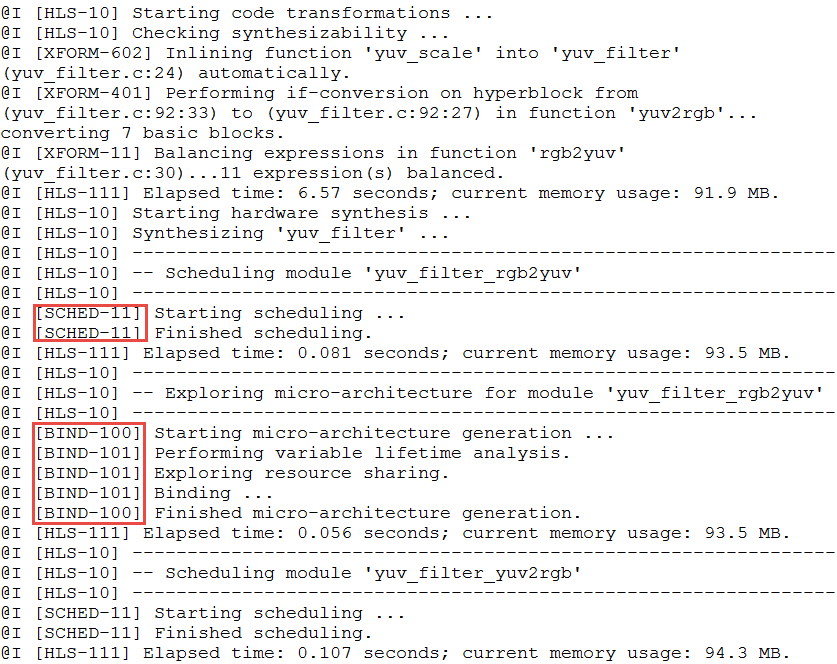


Figure 3. Synthesizing (Generating) the design

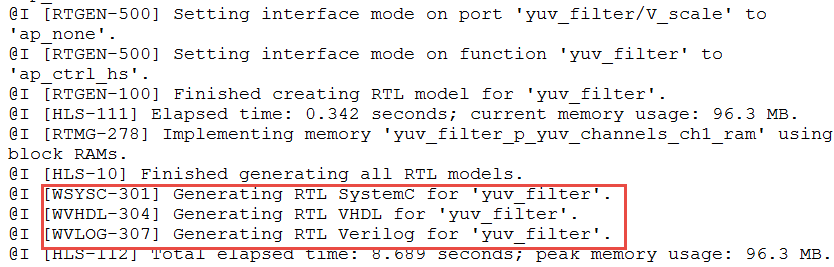


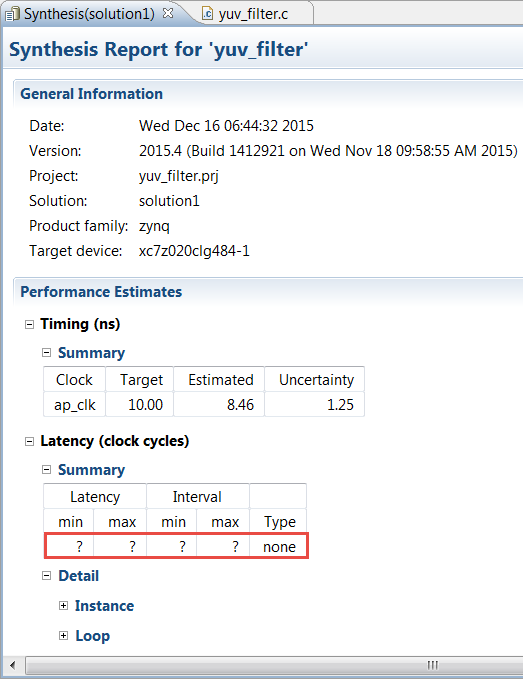
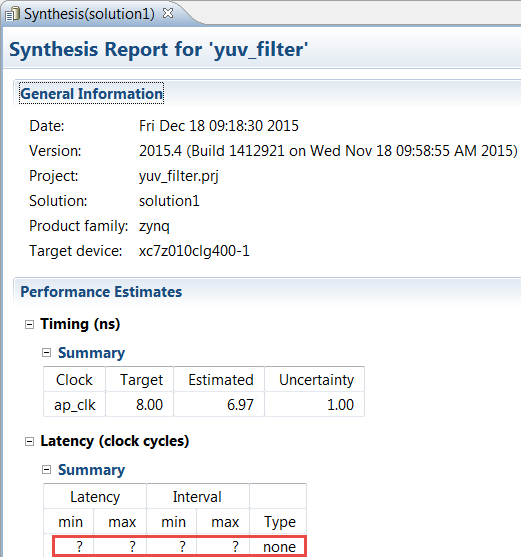
Figure 4. Generating RTL

* + 1. Open the created project (in GUI mode) from the Vivado HLS Command Prompt window, by typing **vivado\_hls –p yuv\_filter.prj**.

The Vivado HLS will open in GUI mode and the project will be opened.

1. Analyze the Created Project and Results Step 2
   1. Open the source file and note that three functions are used. Look at the results and observe that the latencies are undefined (represented by ?).
      1. In Vivado HLS GUI, expand the **source** folder in the Explorer view and double-click **yuv\_filter.c** to view the content.

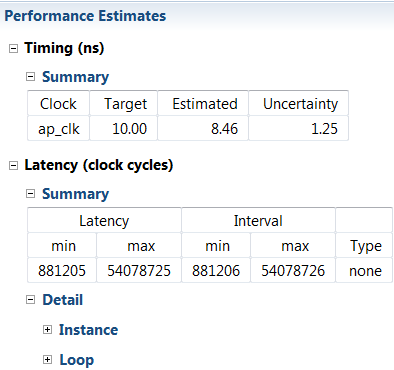
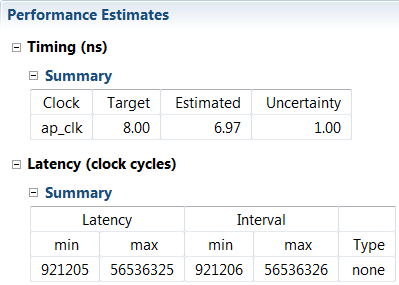
* The design is implemented in 3 functions: rgb2yuv, yuv\_scale and yuv2rgb.
* Each of these filter functions iterates over the entire source image (which has maximum dimensions specified in image\_aux.h), requiring a single source pixel to produce a pixel in the result image.
* The scale function simply applies individual scale factors, supplied as top-level arguments to the Y’UV components.
* Notice that most of the variables are of user-defined (typedef) and aggregate (e.g. structure, array) types.
* Also notice that the original source used malloc() to dynamically allocate storage for the internal image buffers. While appropriate for such large data structures in software, malloc() is not synthesizable and is not supported by Vivado HLS.
* A viable workaround is conditionally compiled into the code, leveraging the \_\_SYNTHESIS\_\_ macro. Vivado HLS automatically defines the \_\_SYNTHESIS\_\_ macro when reading any code. This ensure the original malloc() code is used outside of synthesis but Vivado HLS will use the workaround when synthesizing.
  + 1. Expand the **syn > report** folder in the Explorer view and double-click **yuv\_filter\_csynh.rpt** entry to open the synthesis report.
    2. Each of the loops in this design has variable bounds – the width and height are defined by members of input type image\_t. When variables bounds are present on loops the total latency of the loops cannot be determined: this impacts the ability to perform analysis using reports. Hence, “?” is reported for various latencies.

(a) ZedBoard (b) Zybo

Figure 5. Latency computation

1. Apply TRIPCOUNT Pragma Step 3
   1. Open the source file and uncomment pragma lines, re-synthesize, and observe the resources used as well as estimated latencies. Answer the questions listed in the detailed section of this step.
      1. To assist in providing loop-latency estimates, Vivado HLS provides a TRIPCOUNT directive which allows limits on the variables bounds to be specified by the user. In this design, such directives have been embedded in the source code, in the form of #pragma statements.
      2. Uncomment the #pragma lines (50, 53, 90, 93, 130, 133) to define the loop bounds and save the file.
      3. Synthesize the design by selecting **Solution > Run C Synthesis > Active Solution.** View the synthesis report when the process is completed.

(a) ZedBoard (b) Zybo

Figure 6. Latency computation after applying TRIPCOUNT pragma

* + 1. Looking at the report, and answer the following question.

Question 1

Estimated clock period:

Worst case latency:

Number of DSP48E used:

Number of BRAMs used:

Number of FFs used:

Number of LUTs used:

* + 1. Scroll the Console window and note that yuv\_scale function is automatically inline into the yuv\_filter function.

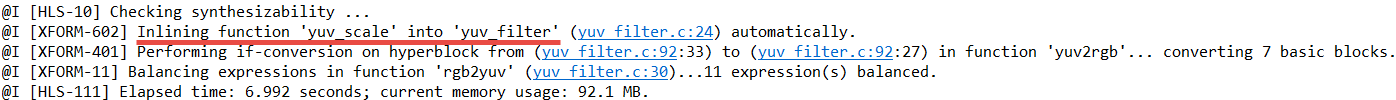
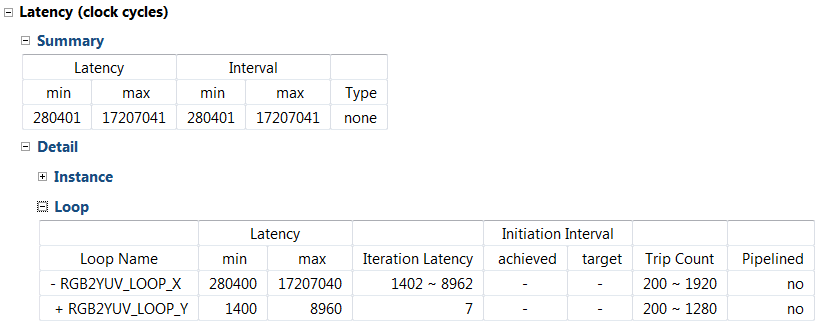


Figure 7. Vivado HLS automatically inlining function

* + 1. Observe that there are three entries – rgb2yuv.rpt, yuv\_filter.rpt, and yuv2rgb.rpt under the **syn** report folder in the Explorer view. There is no entry for yuv\_scale.rpt since the function was inlined into the yuv\_filter function.

You can access lower level module’s report by either traversing down in the top-level report under components (under Utilization Estimates > Details > Component) or from the reports container in the project explorer.

* + 1. Expand the Summary of loop latency and note the latency and trip count numbers for the yuv\_scale function. Note that the YUV\_SCALE\_LOOP\_Y loop latency is 7X the specified TRIPCOUNT, implying that 7 cycles are used for each of the iteration of the loop.



(a) ZedBoard

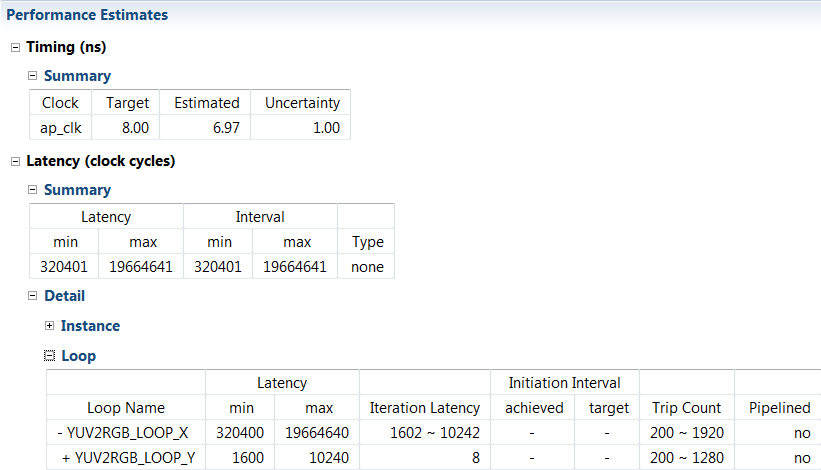
 (b) Zybo

Figure 8. Loop latency

* + 1. You can verify this by opening an analysis perspective view, expanding the **YUV\_SCALE\_LOOP\_X** entry, and then expanding the **YUV\_SCALE\_LOOP\_Y** entry.

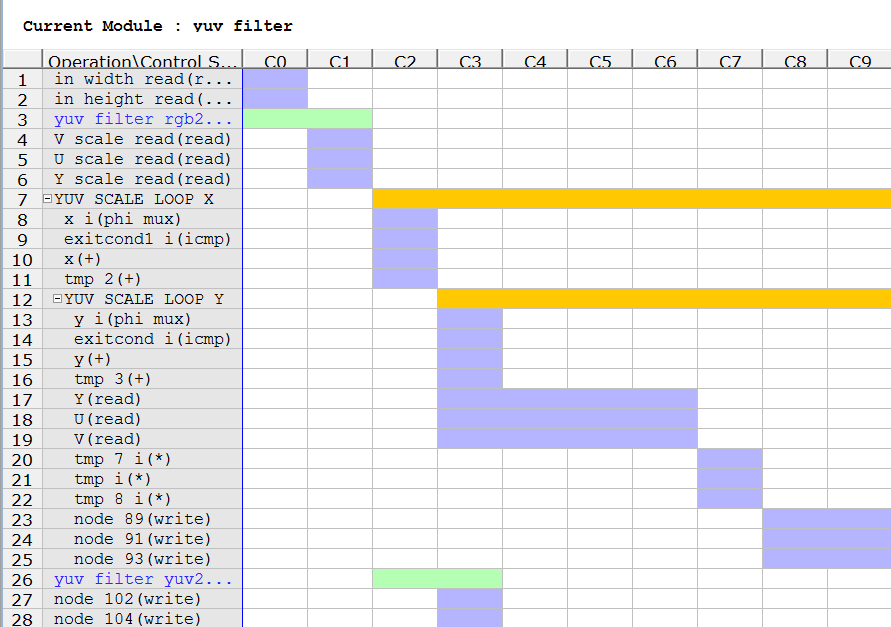


Figure 9. Design analysis view of the YUV\_SCALE\_LOOP\_Y loop

* + 1. In the report tab, expand **Detail > Instance** section of the *Utilization Estimates* and click on the **grp\_rgb2yuv\_fu\_244** (rgb2yuv) entry to open the report.
    2. Answer the following question pertaining to rgb2yuv function.

Question 2

Estimated clock period:

Worst case latency:

Number of DSP48E used:

Number of FFs used:

Number of LUTs used:

* + 1. Similarly, open the yuv2rgb report.
    2. Answer the following question pertaining to yuv2rgb function.

Question 3

Estimated clock period:

Worst case latency:

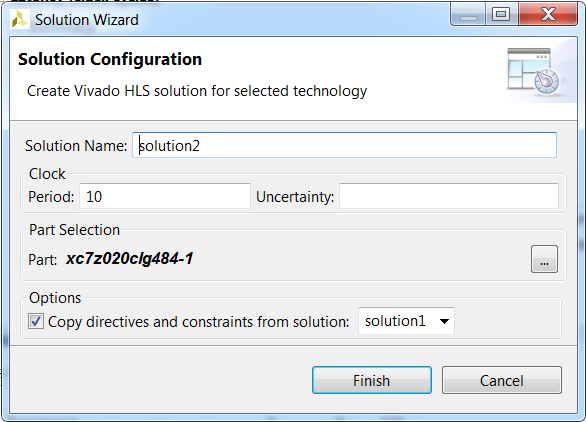
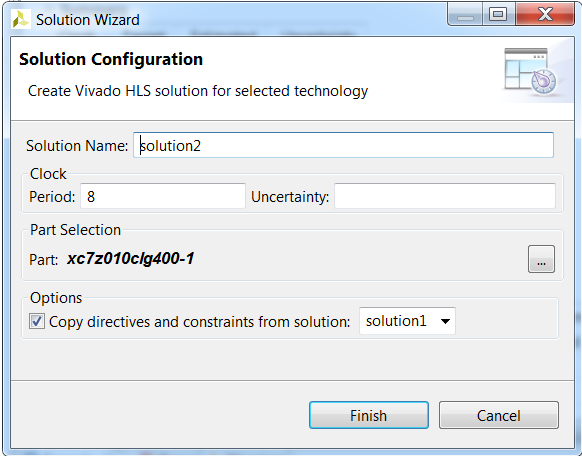
Number of DSP48E used:

Number of FFs used:

Number of LUTs used:

* + 1. For the rgb2yuv function, in case of ZedBoard, the worst case latency is reported as 17207040 clock cycles. The reported latency can be estimated as follows.
* RGB2YUV\_LOOP\_Y total loop latency = 7 x 1280 = 8960 cycles
* 1 entry and 1 exit clock for loop RGB2YUV\_LOOP\_Y = 8962 cycles
* RGB2YUV\_LOOP\_X loop body latency = 8962 cycles
* RGB2YUV\_LOOP\_X total loop latency = 8962 x 1920 =17207040 cycles
  + 1. For the rgb2yuv function, in case of ZYBO, the worst case latency is reported as 19664640 clock cycles. The reported latency can be estimated as follows.
* RGB2YUV\_LOOP\_Y total loop latency = 8 x 1280 = 10240 cycles
* 1 entry and 1 exit clock for loop RGB2YUV\_LOOP\_Y = 10242 cycles
* RGB2YUV\_LOOP\_X loop body latency = 10242 cycles
* RGB2YUV\_LOOP\_X total loop latency = 10242 x 1920 =19664640 cycles

1. Turn OFF INLINE and Apply PIPELINE Directive Step 4
   1. Create a new solution by copying the previous solution settings. Prevent the automatic INLINE and apply PIPELINE directive. Generate the solution and understand the output.
      1. Select **Project > New Solution** or click on () from the tools bar buttons.
      2. A *Solution Configuration* dialog box will appear. Note that the check boxes of *Copy existing directives from solution* and *Copy* *custom constraints directives from solution* are checked with Solution1 selected. Click the **Finish** button to create a new solution with the default settings.

(a) ZedBoard (b) Zybo

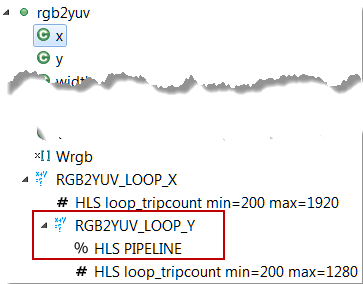
Figure 10. Creating a new Solution after copying the existing solution

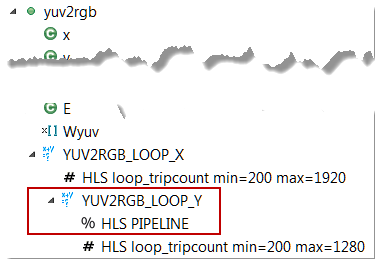
* + 1. Make sure that the **yuv\_filter.c** source is opened and visible in the information pane, and click on the **Directive** tab.
    2. Select function **yuv\_scale** in the directives pane, right-click on it and select *Insert Directive...*
    3. Click on the drop-down button of the *Directive* field. A pop-up menu shows up listing various directives. Select **INLINE** directive.
    4. In the *Vivado HLS Directive Editor* dialog box, click on the **off** option to turn OFF the automatic inlining. Make sure that the Directive File is selected as destination. Click **OK**.



Figure 11. Turning OFF the inlining function

* When an object (function or loop) is pipelined, all the loops below it, down through the hierarchy, will be automatically unrolled.
* In order for a loop to be unrolled it must have fixed bounds: all the loops in this design have variable bounds, defined by an input argument variable to the top-level function.
* Note that the TRIPCOUNT directive on the loops only influences reporting, it does not set bounds for synthesis.
* Neither the top-level function nor any of the sub-functions are pipelined in this example.
* The pipeline directive must be applied to the inner-most loop in each function – the inner-most loops have no variable-bounded loops inside of them which are required to be unrolled and the outer loop will simply keep the inner loop fed with data
  + 1. Expand the *yuv\_scale* in the Directives tab, right-click on *YUV\_SCALE\_LOOP\_Y* object and select insert directives …, and select **PIPELINE** as the directive.
    2. Leave **II** (Initiation Interval) blank as Vivado HLS will try for an II=1, one new input every clock cycle.
    3. Click **OK**.
    4. Similarly, apply the PIPELINE directive to *YUV2RGB\_LOOP\_Y* and *RGB2YUV\_LOOP\_Y* objects. At this point, the Directive tab should look like as follows.





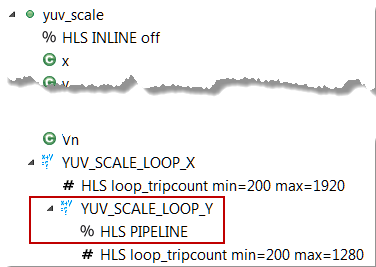
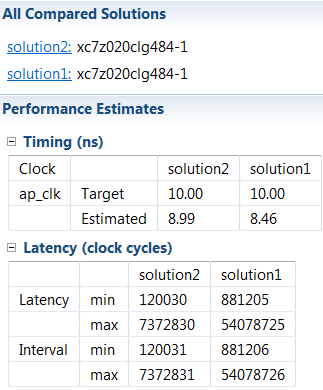
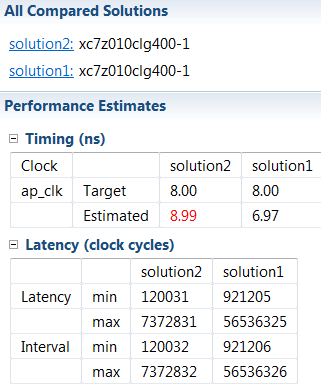


Figure 12. PIPELINE directive applied

* + 1. Click on the **Synthesis** button.
    2. When the synthesis is completed, select **Project > Compare Reports…** or click on  to compare the two solutions.
    3. Select *Solution1* and *Solution2* from the **Available Reports,** and click on the **Add>>** button.
    4. Observe that the latency reduced from 56536325 to 7372831 clock cycles.

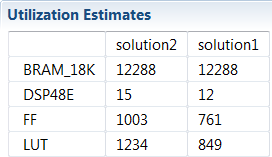
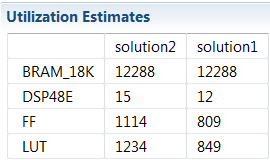
 

(a) ZedBoard (b) Zybo

Figure 13. Performance comparison after pipelining

In Solution1, the total loop latency of the inner-most loop was loop\_body\_latency x loop iteration count, whereas in Solution2 the new total loop latency of the inner-most loop is loop\_body\_latency + loop iteration count.

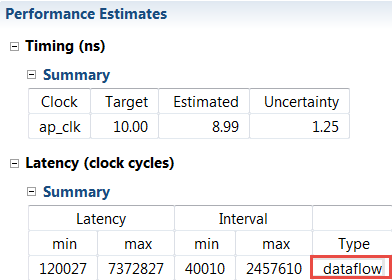
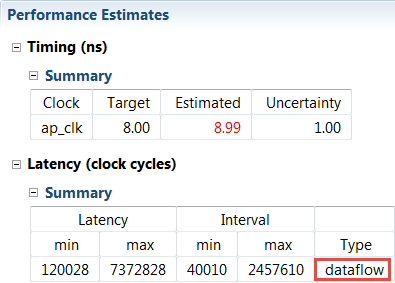
* + 1. Scroll down in the comparison report to view the resources utilization. Observe that the FFs, LUTs, and DSP48E utilization increased whereas BRAM remained same.

(a) ZedBoard (b) Zybo

Figure 14. Resources utilization after pipelining

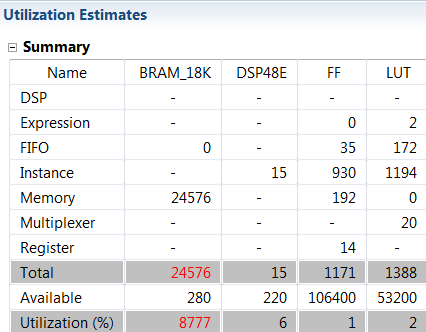
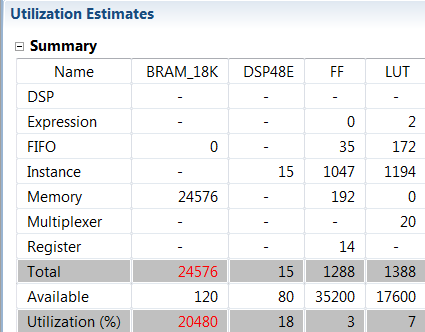
1. Apply DATAFLOW Directive and Configuration Command Step 5
   1. Create a new solution by copying the previous solution (Solution2) settings. Apply DATAFLOW directive. Generate the solution and understand the output.
      1. Select **Project > New Solution** or click on () from the tools bar buttons.
      2. A *Solution Configuration* dialog box will appear. Click the **Finish** button (with copy from Solution2 selected).
      3. Close all inactive solution windows by selecting **Project > Close Inactive Solution** **Tabs.**
      4. Make sure that the *yuv\_filter.c* source is opened in the information pane and select the Directive tab.
      5. Select function **yuv\_filter** in the directives pane, right-click on it and select *Insert Directive...*
      6. A pop-up menu shows up listing various directives. Select **DATAFLOW** directive and click **OK**.
      7. Click on the **Synthesis** button.
      8. When the synthesis is completed, the synthesis report is automatically opened.
      9. Observe additional information, Dataflow Type, in the Performance Estimates section is mentioned.

(a) ZedBoard (b) Zybo

Figure 15. Performance estimate after DATAFLOW directive applied

* The Dataflow pipeline throughput indicates the number of clocks cycles between each set of inputs reads. If this throughput value is less than the design latency it indicates the design can start processing new inputs before the currents input data are output.
* While the overall latencies haven’t changed significantly, the dataflow throughput is showing that the design can achieve close to the theoretical limit (1920x1280 = 2457600) of processing one pixel every clock cycle.
  + 1. Scrolling down into the Utilization Estimates, observe that the number of BRAMs required has doubled. This is due to the default dataflow ping-pong buffering.

(a) ZedBoard (b) Zybo

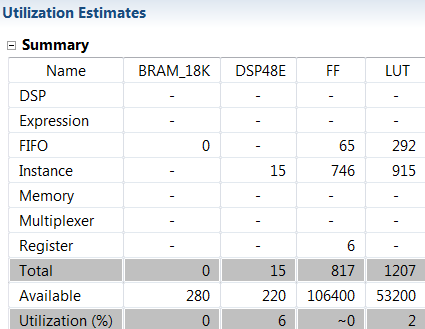
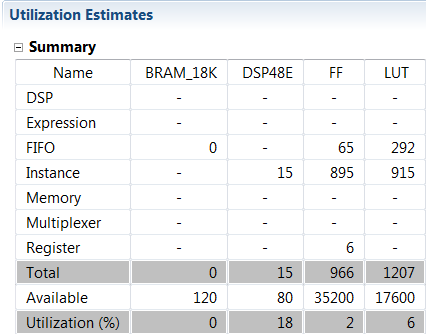
Figure 16. Resource estimate with DATAFLOW directive applied

* When DATAFLOW optimization is performed, memory buffers are automatically inserted between the functions to ensure the next function can begin operation before the previous function has finished. The default memory buffers are ping-pong buffers sized to fully accommodate the largest producer or consumer array.
* Vivado HLS allows the memory buffers to be the default ping-pong buffers or FIFOs. Since this design has data accesses which are fully sequential, FIFOs can be used. Another advantage to using FIFOs is that the size of the FIFOs can be directly controlled (not possible in ping-pong buffers where random accesses are allowed).
  + 1. The memory buffers type can be selected using Vivado HLS Configuration command.
  1. Apply Dataflow configuration command, generate the solution, and observe the improved resources utilization.
     1. Select **Solution > Solution Settings…** or click on  to access the configuration command settings.
     2. In the *Configuration Settings* dialog box, select **General** and click the **Add…** button.
     3. Select *config\_dataflow* as the command using the drop-down button and **fifo** as the default\_channel. Enter **2** as the fifo\_depth. Click **OK**.



Figure 17. Selecting Dataflow configuration command and FIFO as buffer

* + 1. Click **OK** again.
    2. Click on the **Synthesis** button.
    3. When the synthesis is completed, the synthesis report is automatically opened.
    4. Note that the performance parameter has not changed; however, resource estimates show that the design is not using any BRAM and other resources (FF, LUT) usage has also reduced.

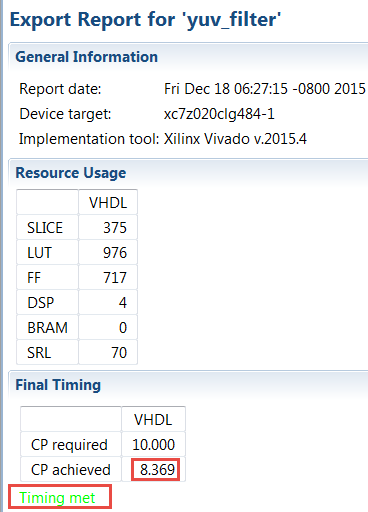
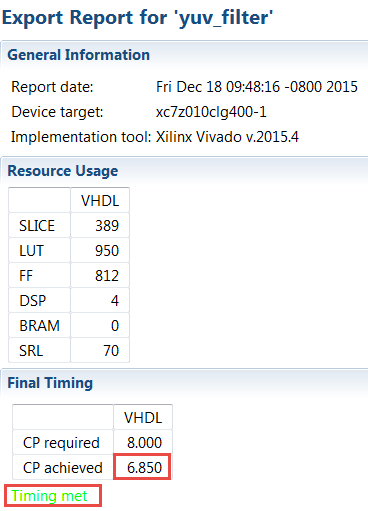
(a) ZedBoard (b) Zybo

Figure 18. Resource estimation after Dataflow configuration command

1. Export and Implement the Design in Vivado HLS Step 6
   1. In Vivado HLS, export the design, selecting VHDL as a language, and run the implementation by selecting Evaluate option.
      1. In Vivado HLS, select **Solution > Export RTL** or click on the  button to open the dialog box so the desired implementation can be run.

An Export RTL Dialog box will open.

* + 1. Click on the drop-down button of the **Option** field and select **VHDL** as the language and tick **Evaluate**.
    2. Click **OK** and the implementation run will begin. You can observe the progress in the Vivado HLS Console window. When the run is completed the implementation report will be displayed in the information pane.

(a) ZedBoard (b) Zybo

Figure 19. Implementation results in Vivado HLS

Note that the implementation was successful in case of ZedBoard but failed in case of Zybo.

* + 1. Close Vivado HLS by selecting **File > Exit.**

Conclusion

In this lab, you learned that even though this design could not be pipelined at the top-level, a strategy of pipelining the individual loops and then using dataflow optimization to make the functions operate in parallel was able to achieve the same high throughput, processing one pixel per clock. When DATAFLOW directive is applied, the default memory buffers (of ping-pong type) are automatically inserted between the functions. Using the fact that the design used only sequential (streaming) data accesses allowed the costly memory buffers associated with dataflow optimization to be replaced with simple 2 element FIFOs using the Dataflow command configuration.

Answers

1. Answer the following questions for yuv\_filter:

Estimated clock period: 8.46 ns (ZedBoard) 6.97 ns (Zybo)

Worst case latency: 54078725 (ZedBoard) 56536325 (Zybo) clock cycles

Number of DSP48E used: 12

Number of BRAMs used: 12288

Number of FFs used: 761 (ZedBoard) 809 (Zybo)

Number of LUTs used: 849 (ZedBoard) 849 (Zybo)

1. Answer the following questions rgb2yuv:

Estimated clock period: 8.34 ns (ZedBoard) 6.38 ns (Zybo)

Worst case latency: 17207041 (ZedBoard) 19664641 (Zybo) clock cycles

Number of DSP48E used: 5

Number of FFs used: 246 (ZedBoard) 257 (Zybo)

Number of LUTs used: 282 (ZedBoard) 282 (Zybo)

1. Answer the following questions for yuv2rgb:

Estimated clock period: 8.46 ns (ZedBoard) 6.97 ns (Zybo)

Worst case latency: 19664641 (ZedBoard) 19664641 (Zybo) clock cycles

Number of DSP48E used: 4

Number of FFs used: 200 (ZedBoard) 237 (Zybo)

Number of LUTs used: 289 (ZedBoard) 289 (Zybo)