Αρχιτεκτονική Προηγμένων Υπολογιστών και Επιταχυντών

Σχεδίαση και προσομοίωση επιταχυντή υλικού με χρήση FPGA

Α. ΑΘΑΝΑΣΙΑΔΗΣ

Τμημα Ηλεκτρολογων Μηχανικων
& Μηχανικων Υπολογιστων
Αριστοτελειο Πανεπιστημιο Θεσσαλονικης

HDL → HLS

➢ Παρελθόν → Επιταχυντές Υλικού (accelerators) σχεδιάζονταν κυρίως με τις γλώσσες περιγραφής υλικού (Hardware Description Languages - HDLs)

  ✓ Verilog, VHDL

➢ Σήμερα → Σχεδιάζονται και με μεθόδους σύνθεσης υψηλού επιπέδου (High-Level Synthesis - HLS)

  ✓ C, C++, SystemC

Πλεονεκτήματα

☝Μείωση του χρόνου ανάπτυξης

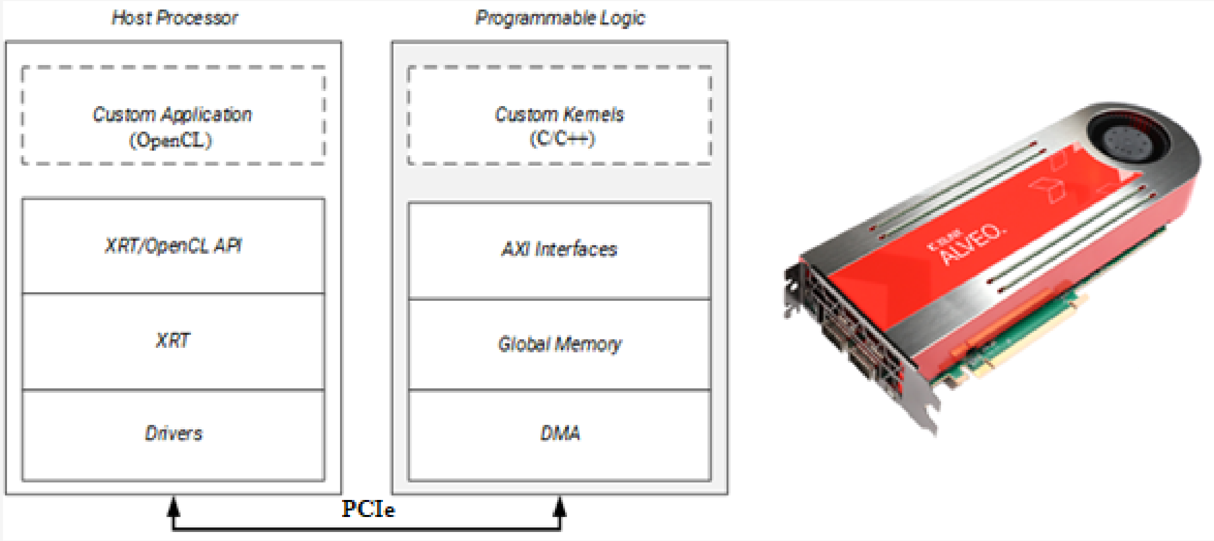
☝Πολύ μεγάλη ευκολία διαχείρισης και μεταβολής της αρχικής σχεδίασης

☝Καλή ποιότητα της παραγόμενης σχεδίασης

Βήματα Σχεδίασης

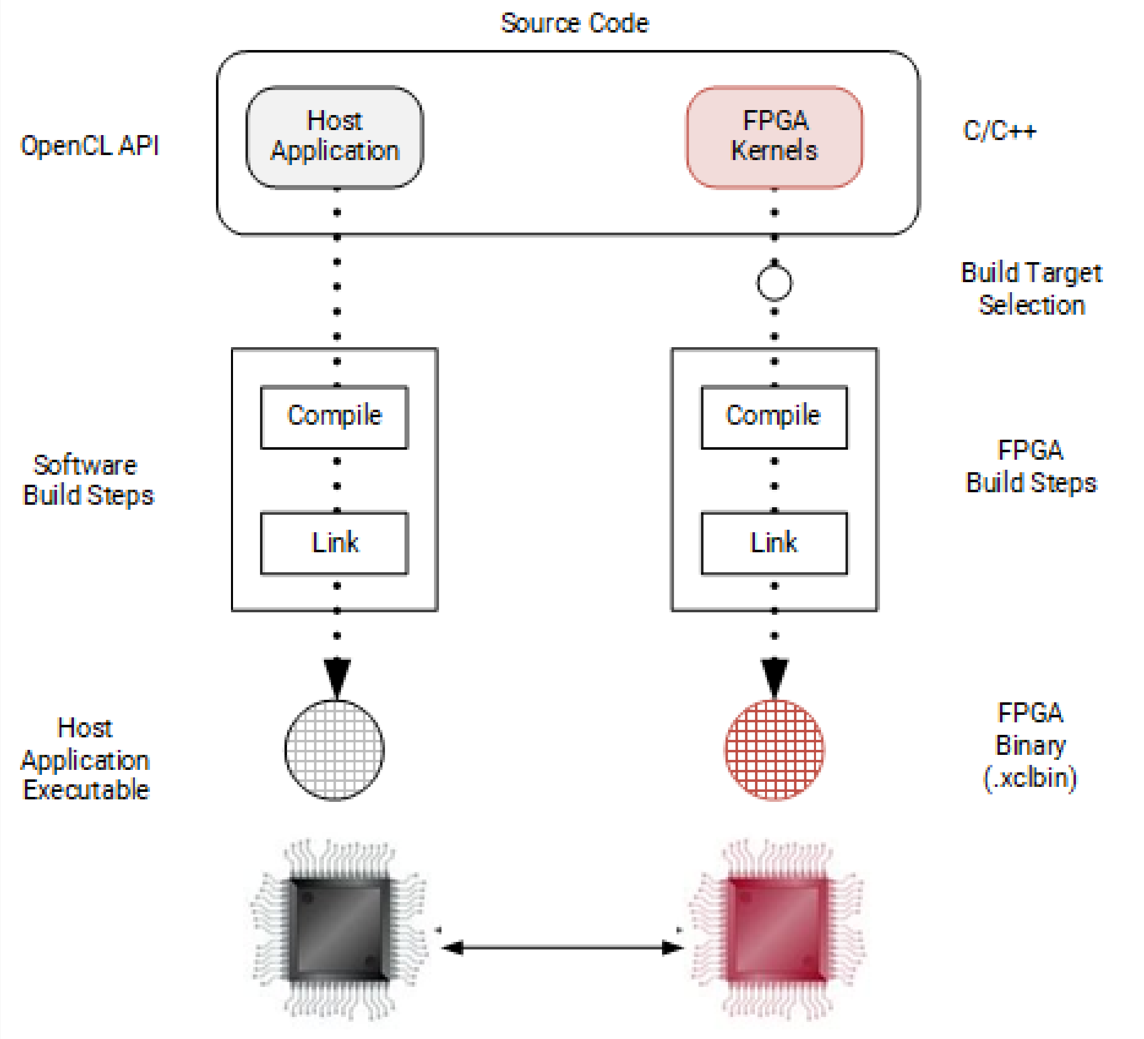
1. Σχεδίαση (Design)
2. Προσομοίωση (Simulation)
3. Υλοποίηση σε πραγματικό Υλικό (π.χ. πλακέτα που περιλαμβάνει ένα FPGA SoC)

Υλοποίηση **εφαρμογής** (π.χ. σε C/C++/ OpenCL), η οποία τρέχει στο επεξεργαστικό σύστημα (Processing System - **PS**, δηλ. τη CPU) του FPGA SoC και **καλεί τον επιταχυντή** που έχει υλοποιηθεί στο τμήμα προγραμματιζόμενης λογικής (Programmable Logic - **PL**, δηλ. το FPGA) του FPGA SoC.

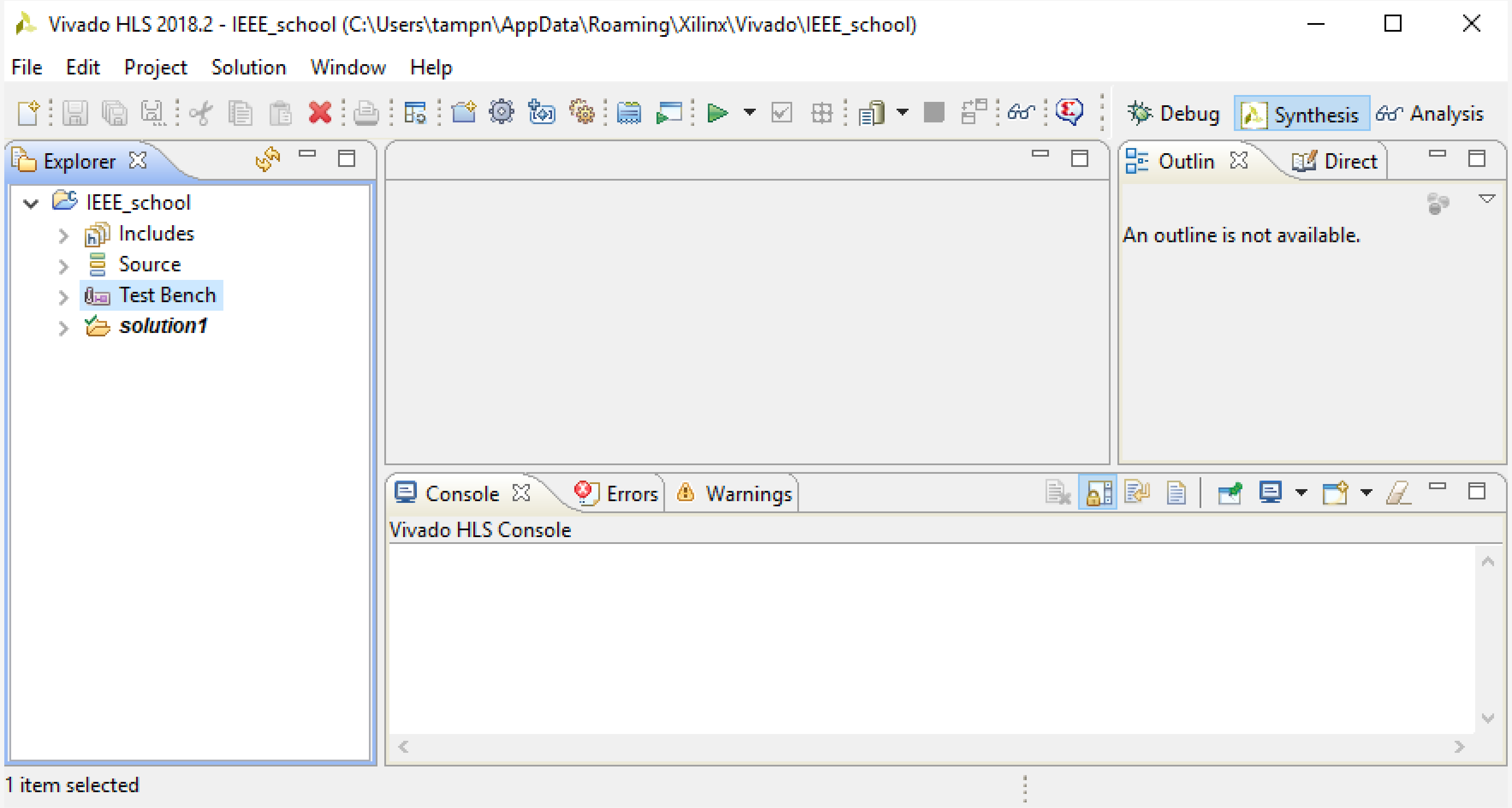


Αρχιτεκτονική Xilinx Alveo U200

Διαδικασία εκτέλεσης της εφαρμογής



Step 1. Project Creation



Vivado HLS 2018.2 - IEEE_school (C:\Users\tampn\AppData\Roaming\Xilinx\Vivado\IEEE_school)

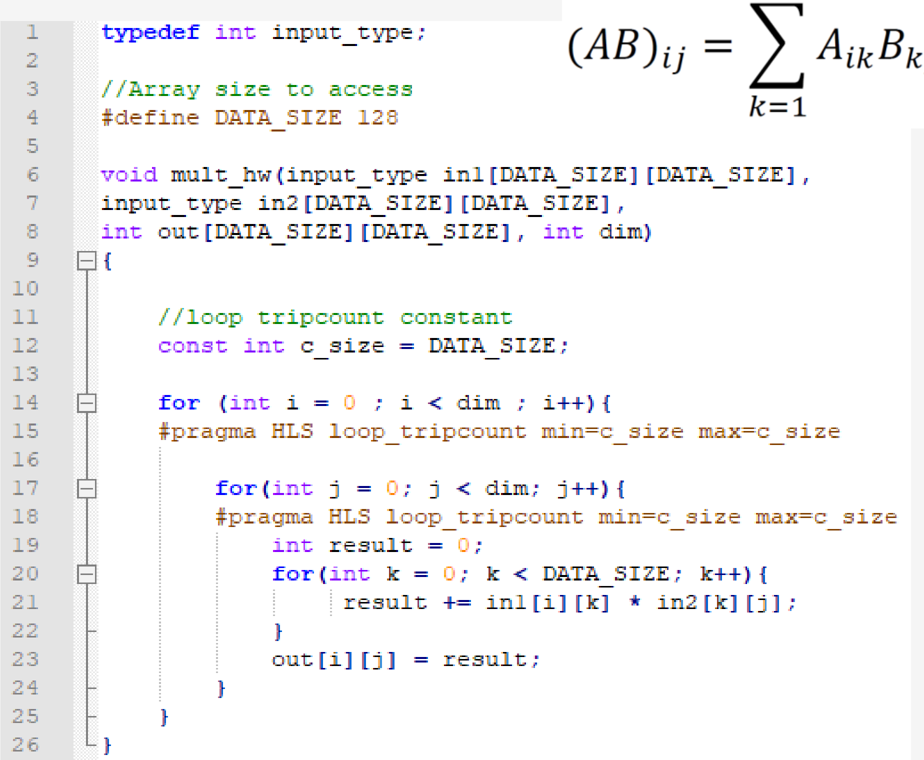
An outline is not available.

Vivado HLS Console

1 item selected

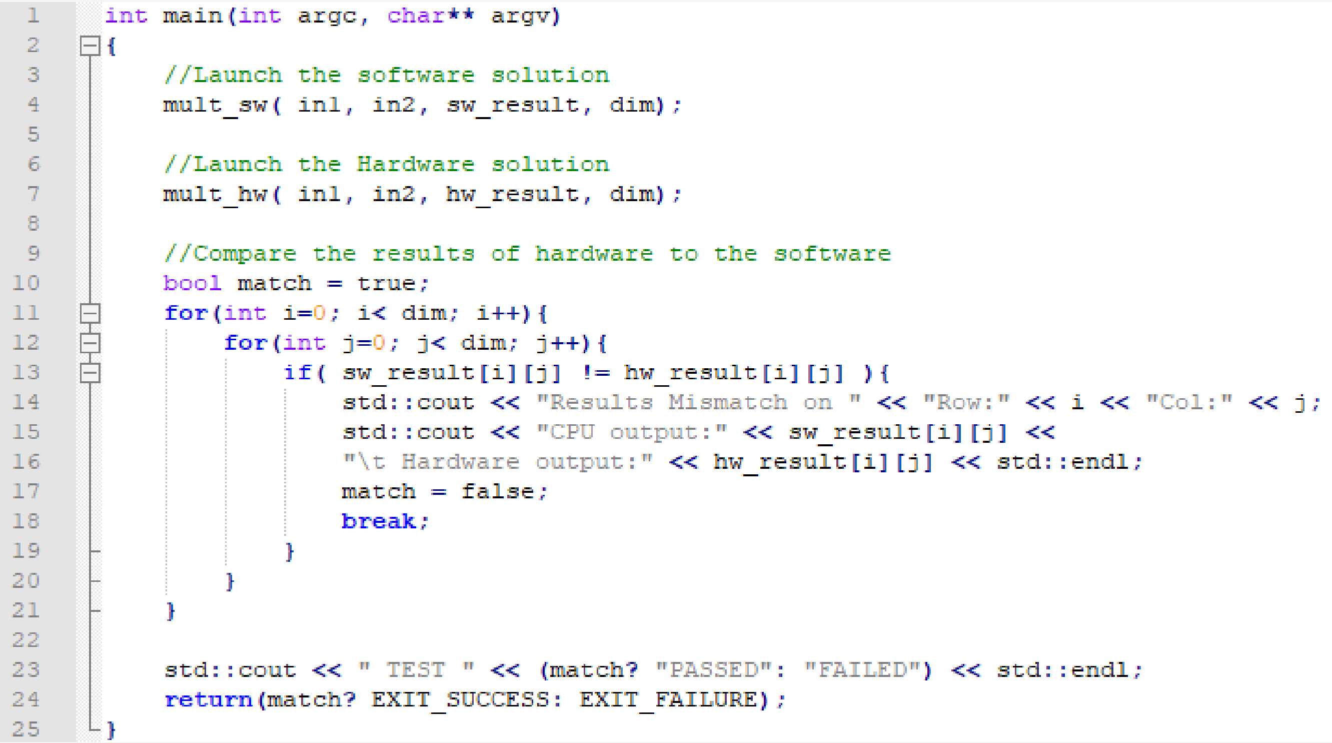
Design under consideration

$$(AB)_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}$$



```
typedef int input_type;

//Array size to access
#define DATA_SIZE 128

void mult_hw(input_type in1[DATA_SIZE][DATA_SIZE],
input_type in2[DATA_SIZE][DATA_SIZE],
int out[DATA_SIZE][DATA_SIZE], int dim)
{

    //loop tripcount constant
    const int c_size = DATA_SIZE;

    for (int i = 0 ; i < dim ; i++){
    #pragma HLS loop_tripcount min=c_size max=c_size

        for(int j = 0; j < dim; j++){
        #pragma HLS loop_tripcount min=c_size max=c_size
            int result = 0;
            for(int k = 0; k < DATA_SIZE; k++){
                result += in1[i][k] * in2[k][j];
            }
            out[i][j] = result;
        }
    }
}
```

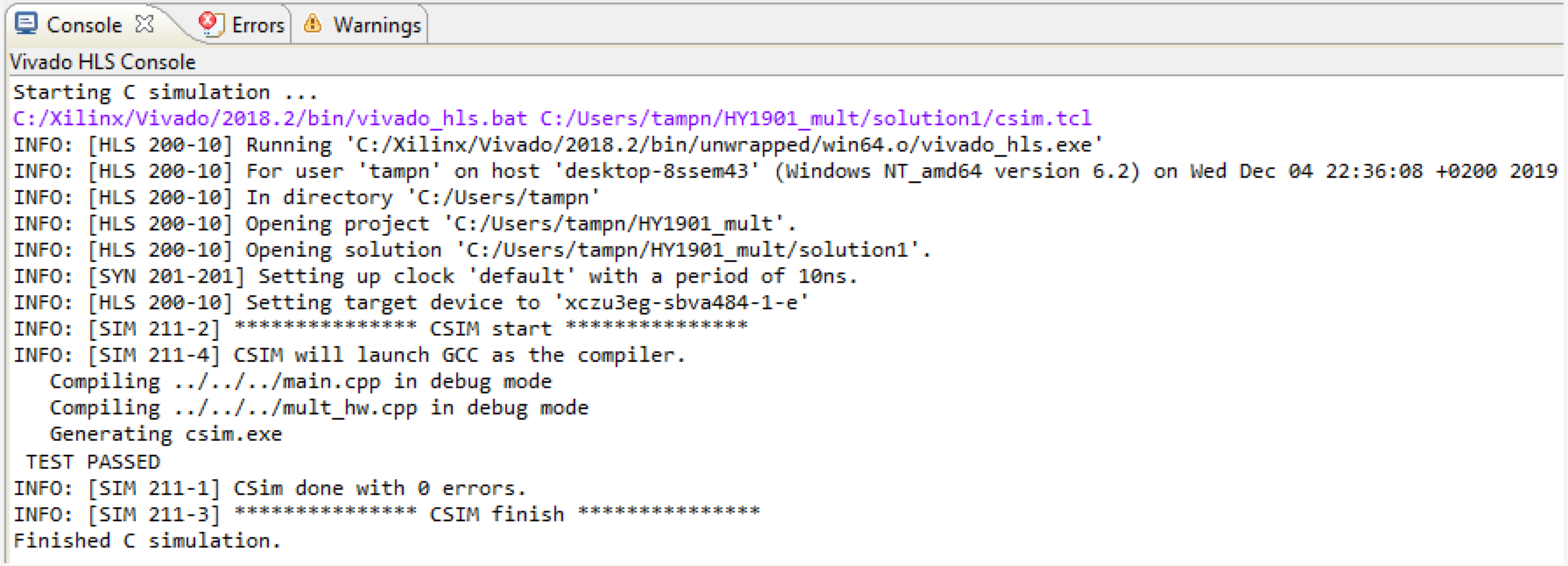
Design (source code)



```
int main(int argc, char** argv)
{
    //Launch the software solution
    mult_sw( in1, in2, sw_result, dim);

    //Launch the Hardware solution
    mult_hw( in1, in2, hw_result, dim);

    //Compare the results of hardware to the software
    bool match = true;
    for(int i=0; i< dim; i++){
        for(int j=0; j< dim; j++){
            if( sw_result[i][j] != hw_result[i][j] ){
                std::cout << "Results Mismatch on " << "Row:" << i << "Col:" << j;
                std::cout << "CPU output:" << sw_result[i][j] <<
                "\t Hardware output:" << hw_result[i][j] << std::endl;
                match = false;
                break;
            }
        }
    }

    std::cout << " TEST " << (match? "PASSED": "FAILED") << std::endl;
    return(match? EXIT_SUCCESS: EXIT_FAILURE);
}
```

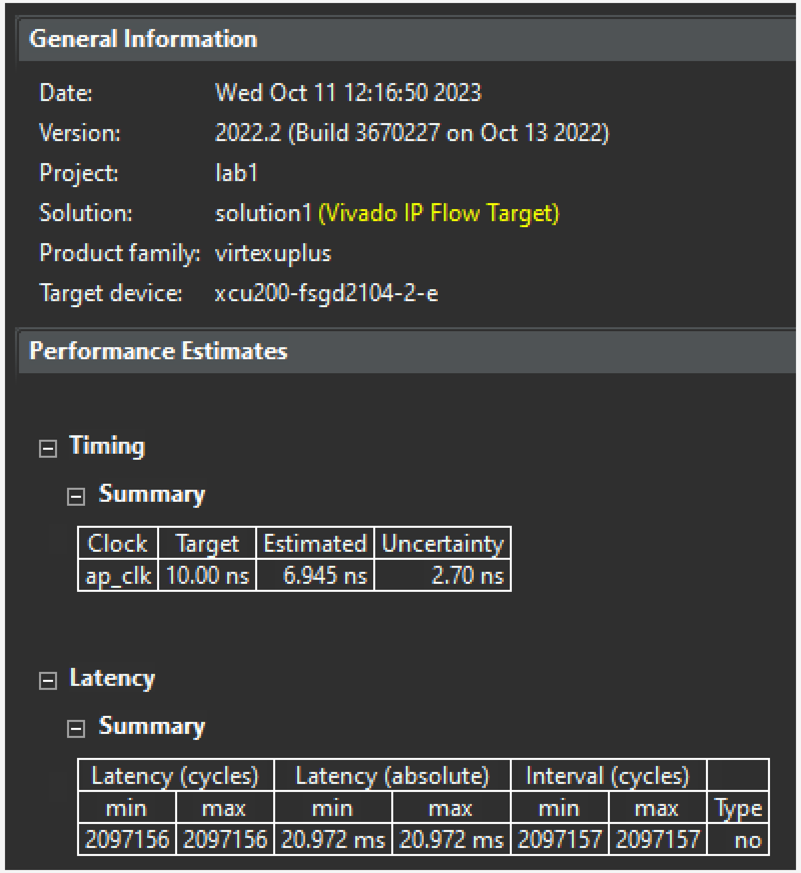
Testbench

Step 2. Run C-Simulation



```
Starting C simulation ...
C:/Xilinx/Vivado/2018.2/bin/vivado_hls.bat C:/Users/tampn/HY1901_mult/solution1/csim.tcl
INFO: [HLS 200-10] Running 'C:/Xilinx/Vivado/2018.2/bin/unwrapped/win64.o/vivado_hls.exe'
INFO: [HLS 200-10] For user 'tampn' on host 'desktop-8ssem43' (Windows NT_amd64 version 6.2) on Wed Dec 04 22:36:08 +0200 2019
INFO: [HLS 200-10] In directory 'C:/Users/tampn'
INFO: [HLS 200-10] Opening project 'C:/Users/tampn/HY1901_mult'.
INFO: [HLS 200-10] Opening solution 'C:/Users/tampn/HY1901_mult/solution1'.
INFO: [SYN 201-201] Setting up clock 'default' with a period of 10ns.
INFO: [HLS 200-10] Setting target device to 'xczu3eg-sbva484-1-e'
INFO: [SIM 211-2] *************** CSIM start ***************
INFO: [SIM 211-4] CSIM will launch GCC as the compiler.
   Compiling ../../../main.cpp in debug mode
   Compiling ../../../mult_hw.cpp in debug mode
   Generating csim.exe
 TEST PASSED
INFO: [SIM 211-1] CSim done with 0 errors.
INFO: [SIM 211-3] *************** CSIM finish ***************
Finished C simulation.
```

Step 3. Design Synthesis



General Information

Date:	Wed Oct 11 12:16:50 2023
Version:	2022.2 (Build 3670227 on Oct 13 2022)
Project:	lab1
Solution:	solution1 (Vivado IP Flow Target)
Product family:	virtexuplus
Target device:	xcu200-fsgd2104-2-e

Performance Estimates

⊟ **Timing**

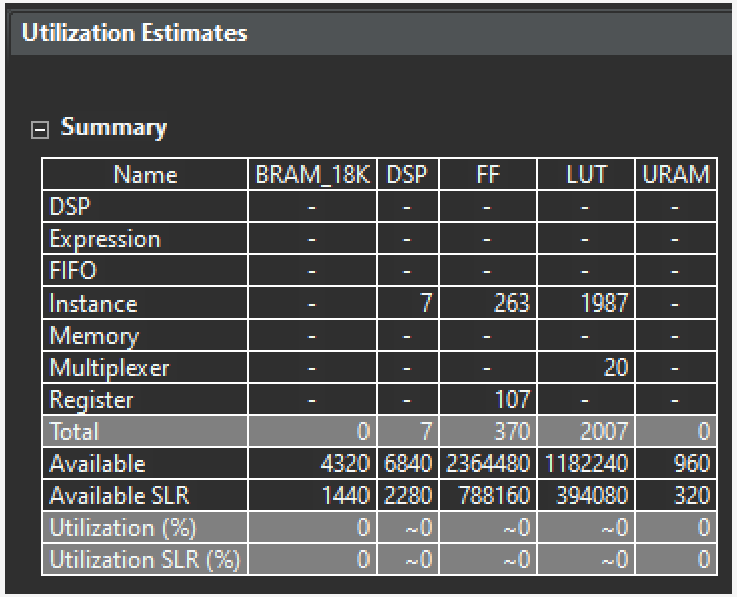
⊟ **Summary**

Clock	Target	Estimated	Uncertainty
ap_clk	10.00 ns	6.945 ns	2.70 ns

⊟ **Latency**

⊟ **Summary**

Latency (cycles)		Latency (absolute)		Interval (cycles)		
min	max	min	max	min	max	Type
2097156	2097156	20.972 ms	20.972 ms	2097157	2097157	no

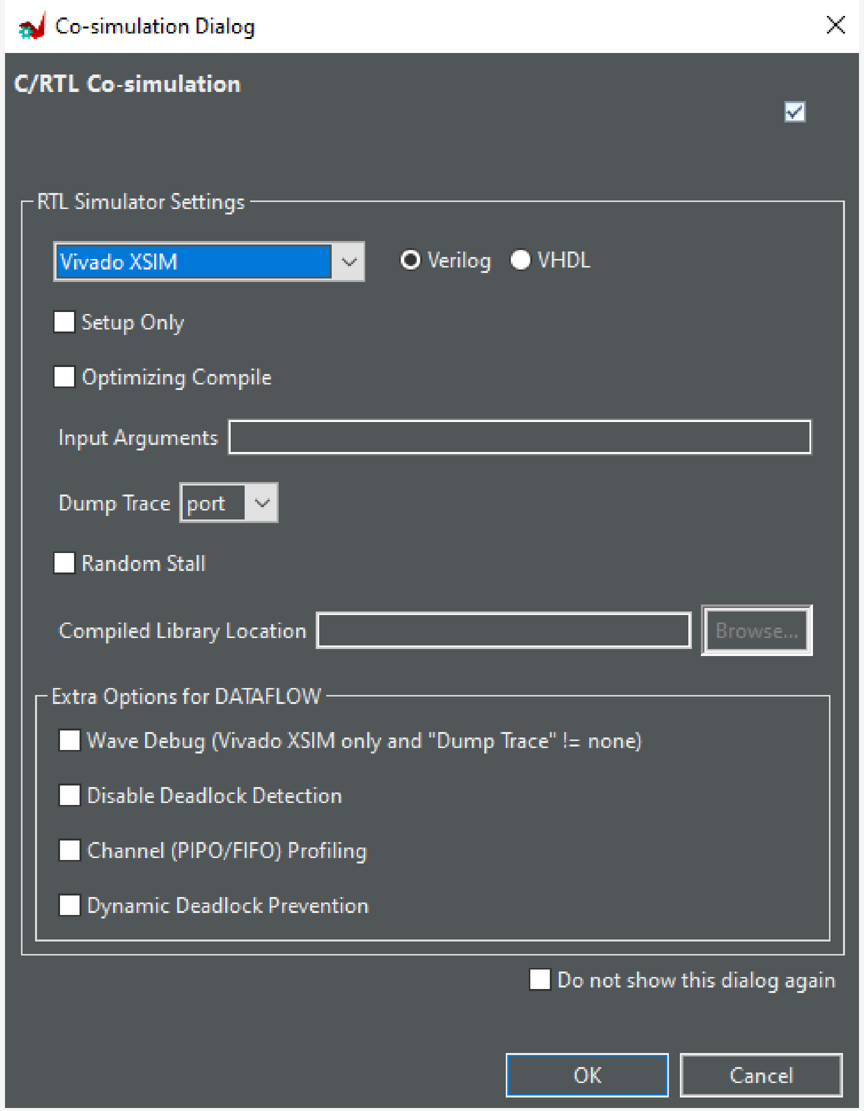


Utilization Estimates

⊟ **Summary**

Name	BRAM_18K	DSP	FF	LUT	URAM
DSP	-	-	-	-	-
Expression	-	-	-	-	-
FIFO	-	-	-	-	-
Instance	-	7	263	1987	-
Memory	-	-	-	-	-
Multiplexer	-	-	-	20	-
Register	-	-	107	-	-
Total	0	7	370	2007	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	0	~0	~0	~0	0
Utilization SLR (%)	0	~0	~0	~0	0

Step 4. Run C/RTL Co-Simulation



Co-simulation Dialog

C/RTL Co-simulation ☑

RTL Simulator Settings

Vivado XSIM ○ Verilog ● VHDL

□ Setup Only

□ Optimizing Compile

Input Arguments

Dump Trace: port

□ Random Stall

Compiled Library Location

Extra Options for DATAFLOW

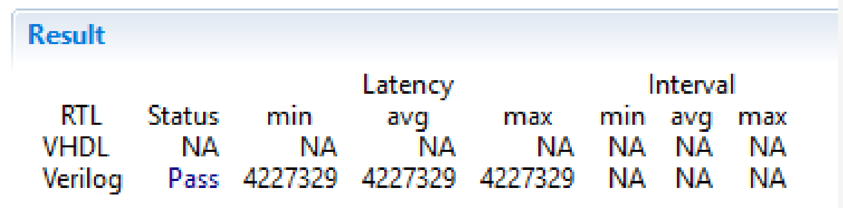
□ Wave Debug (Vivado XSIM only and "Dump Trace" != none)

□ Disable Deadlock Detection

□ Channel (PIPO/FIFO) Profiling

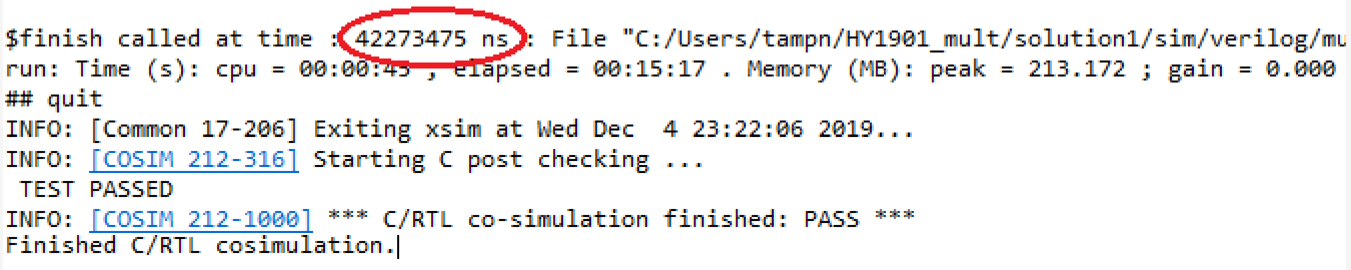
□ Dynamic Deadlock Prevention

□ Do not show this dialog again



Result

			Latency			Interval	
RTL	Status	min	avg	max	min	avg	max
VHDL	NA	NA	NA	NA	NA	NA	NA
Verilog	Pass	4227329	4227329	4227329	NA	NA	NA



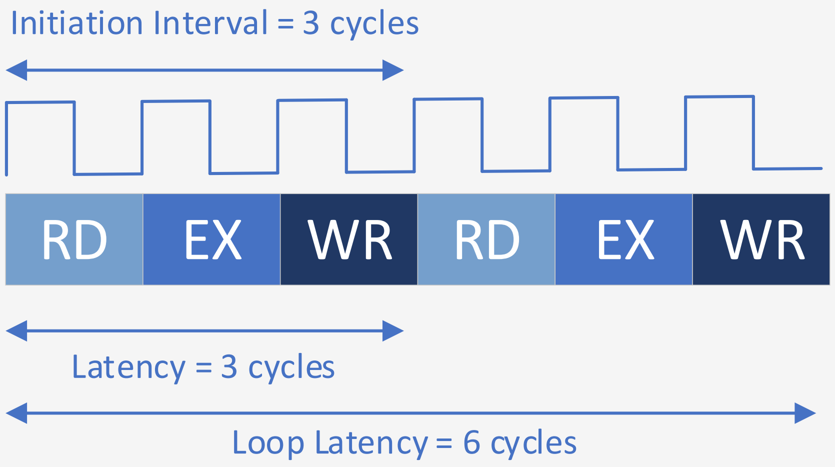
```
$finish called at time : 42273475 ns : File "C:/Users/tampn/HY1901_mult/solution1/sim/verilog/mu
run: Time (s): cpu = 00:00:45 ; elapsed = 00:15:17 . Memory (MB): peak = 213.172 ; gain = 0.000
## quit
INFO: [Common 17-206] Exiting xsim at Wed Dec  4 23:22:06 2019...
INFO: [COSIM 212-316] Starting C post checking ...
 TEST PASSED
INFO: [COSIM 212-1000] *** C/RTL co-simulation finished: PASS ***
Finished C/RTL cosimulation.
```

Step 5. Optimizing the Design (Loop Unrolling)

Without Unrolling

```
for (i=0; i<2; i++){
  Op_Read;                RD
  Op_Execute;             EX
  Op_Write;               WR
}
```

With Unrolling

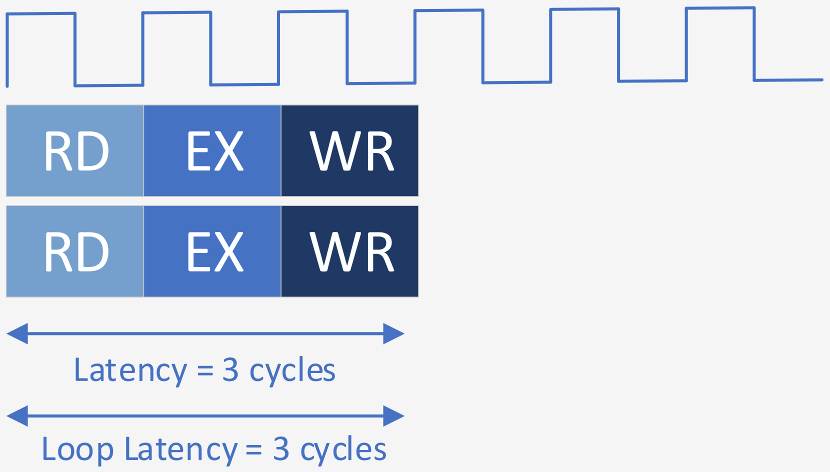


Initiation Interval = 3 cycles

RD EX WR RD EX WR

Latency = 3 cycles

Loop Latency = 6 cycles



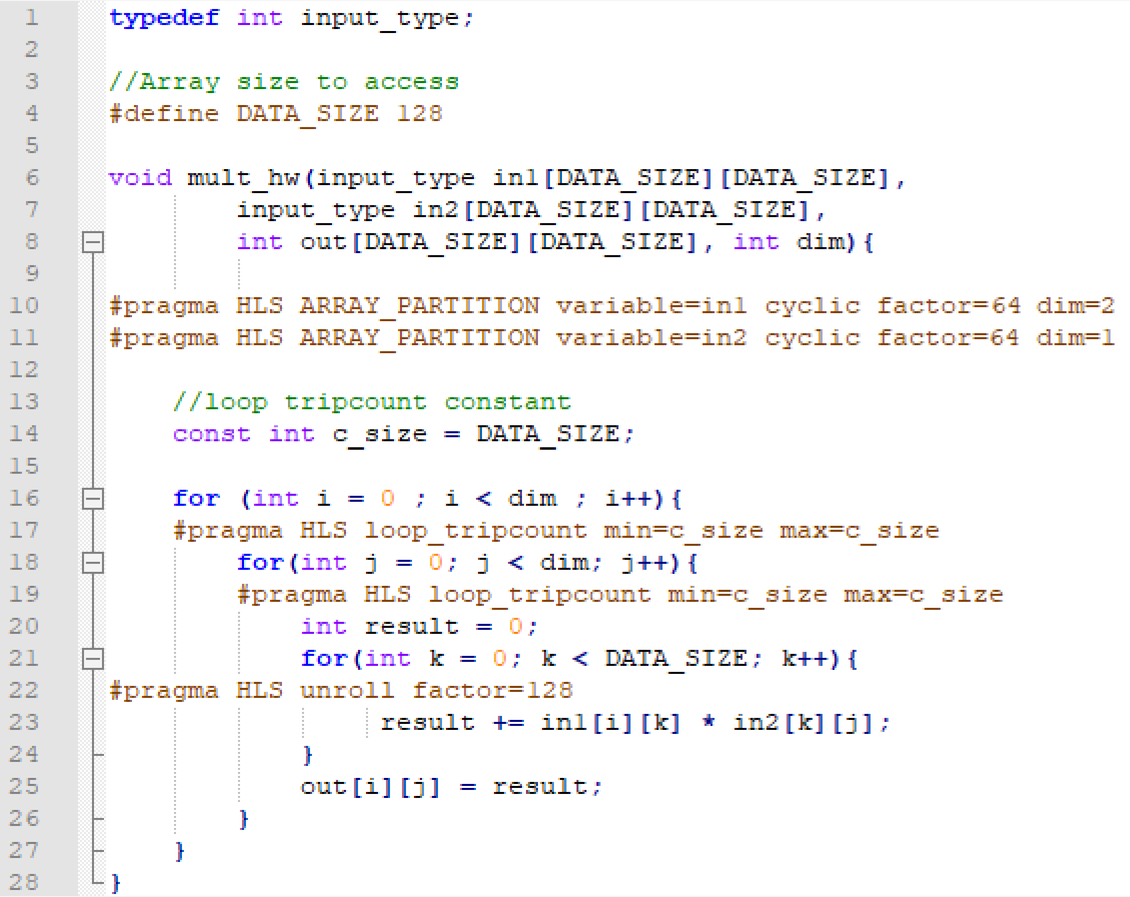
RD EX WR

RD EX WR

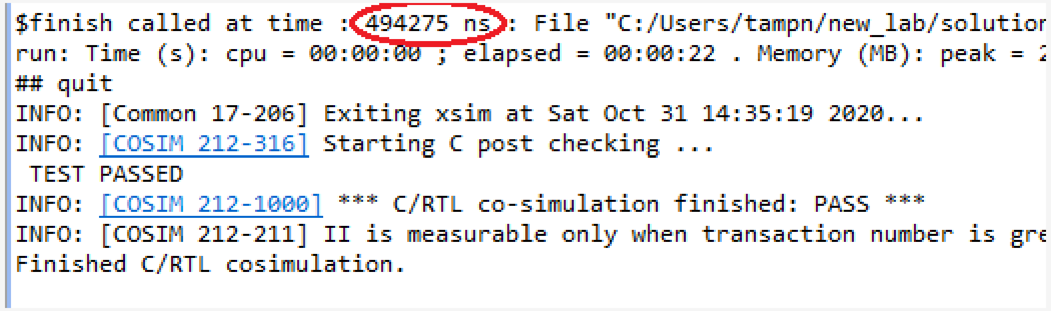
Latency = 3 cycles

Loop Latency = 3 cycles

Step 5. Optimizing the Design (Loop Unrolling)

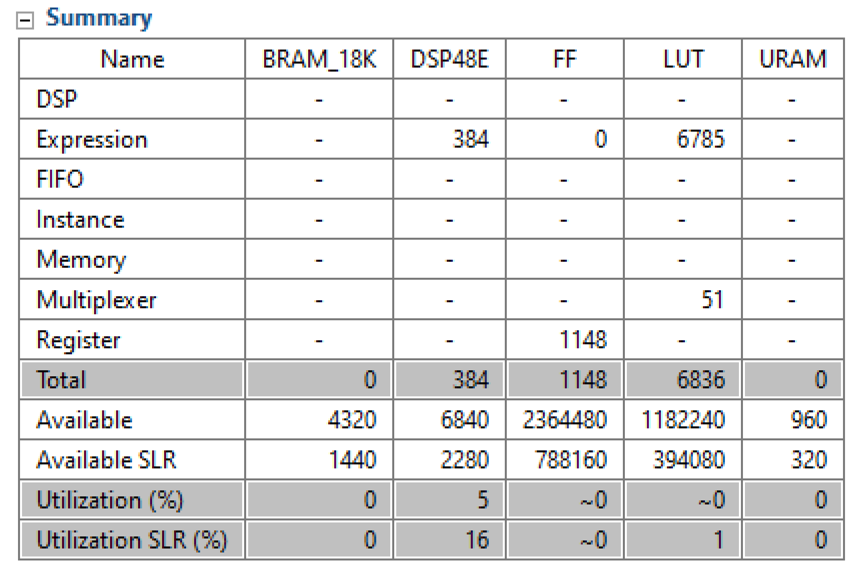


```
typedef int input_type;

//Array size to access
#define DATA_SIZE 128

void mult_hw(input_type in1[DATA_SIZE][DATA_SIZE],
        input_type in2[DATA_SIZE][DATA_SIZE],
        int out[DATA_SIZE][DATA_SIZE], int dim){

#pragma HLS ARRAY_PARTITION variable=in1 cyclic factor=64 dim=2
#pragma HLS ARRAY_PARTITION variable=in2 cyclic factor=64 dim=1

    //loop tripcount constant
    const int c_size = DATA_SIZE;

    for (int i = 0 ; i < dim ; i++){
    #pragma HLS loop_tripcount min=c_size max=c_size
        for(int j = 0; j < dim; j++){
        #pragma HLS loop_tripcount min=c_size max=c_size
            int result = 0;
            for(int k = 0; k < DATA_SIZE; k++){
#pragma HLS unroll factor=128
                result += in1[i][k] * in2[k][j];
            }
            out[i][j] = result;
        }
    }
}
```



```
$finish called at time : 494275 ns : File "C:/Users/tampn/new_lab/solutior
run: Time (s): cpu = 00:00:00 ; elapsed = 00:00:22 . Memory (MB): peak = 2
## quit
INFO: [Common 17-206] Exiting xsim at Sat Oct 31 14:35:19 2020...
INFO: [COSIM 212-316] Starting C post checking ...
 TEST PASSED
INFO: [COSIM 212-1000] *** C/RTL co-simulation finished: PASS ***
INFO: [COSIM 212-211] II is measurable only when transaction number is gre
Finished C/RTL cosimulation.
```

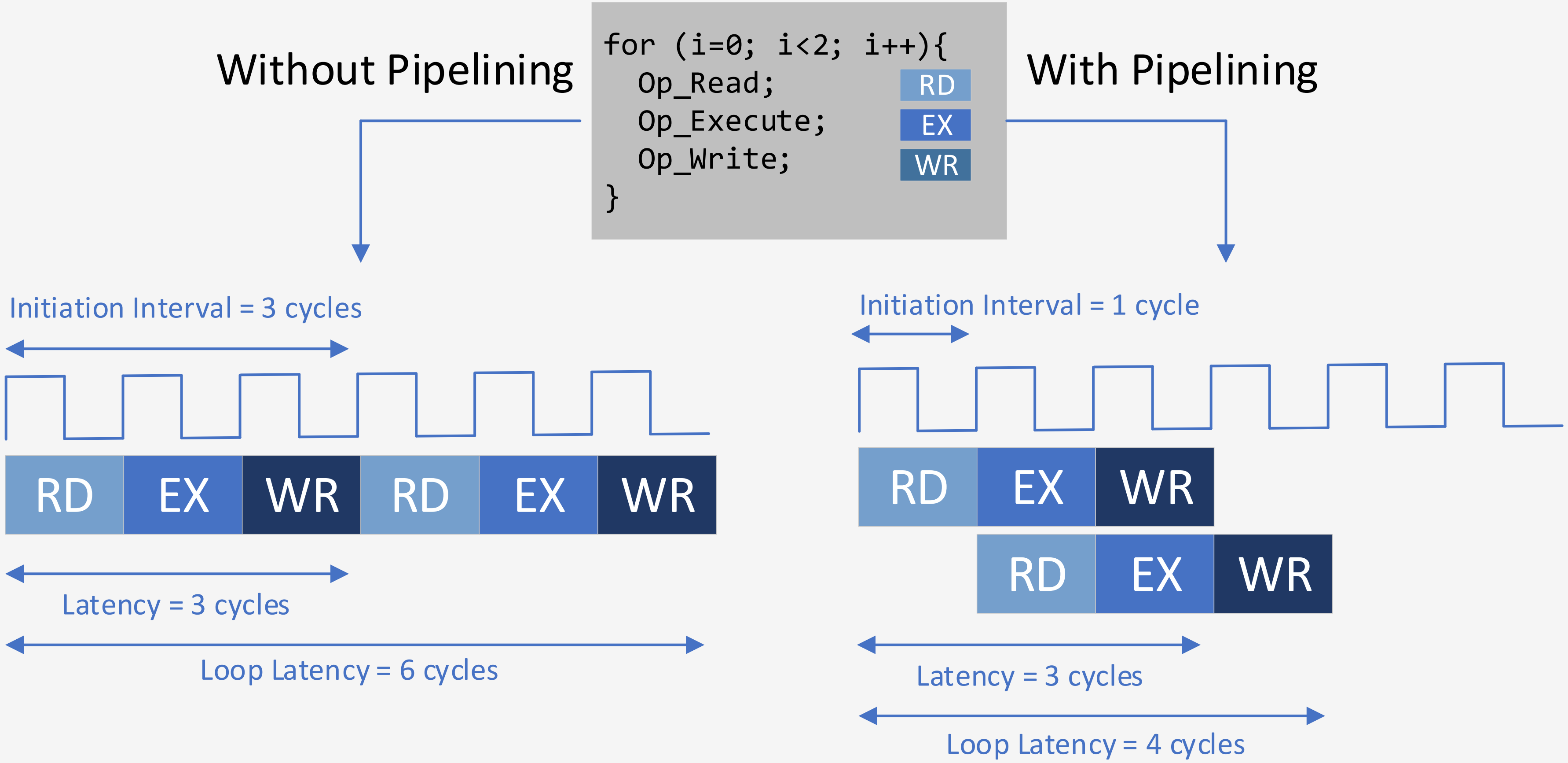
(128*128*3)



⊟ **Summary**

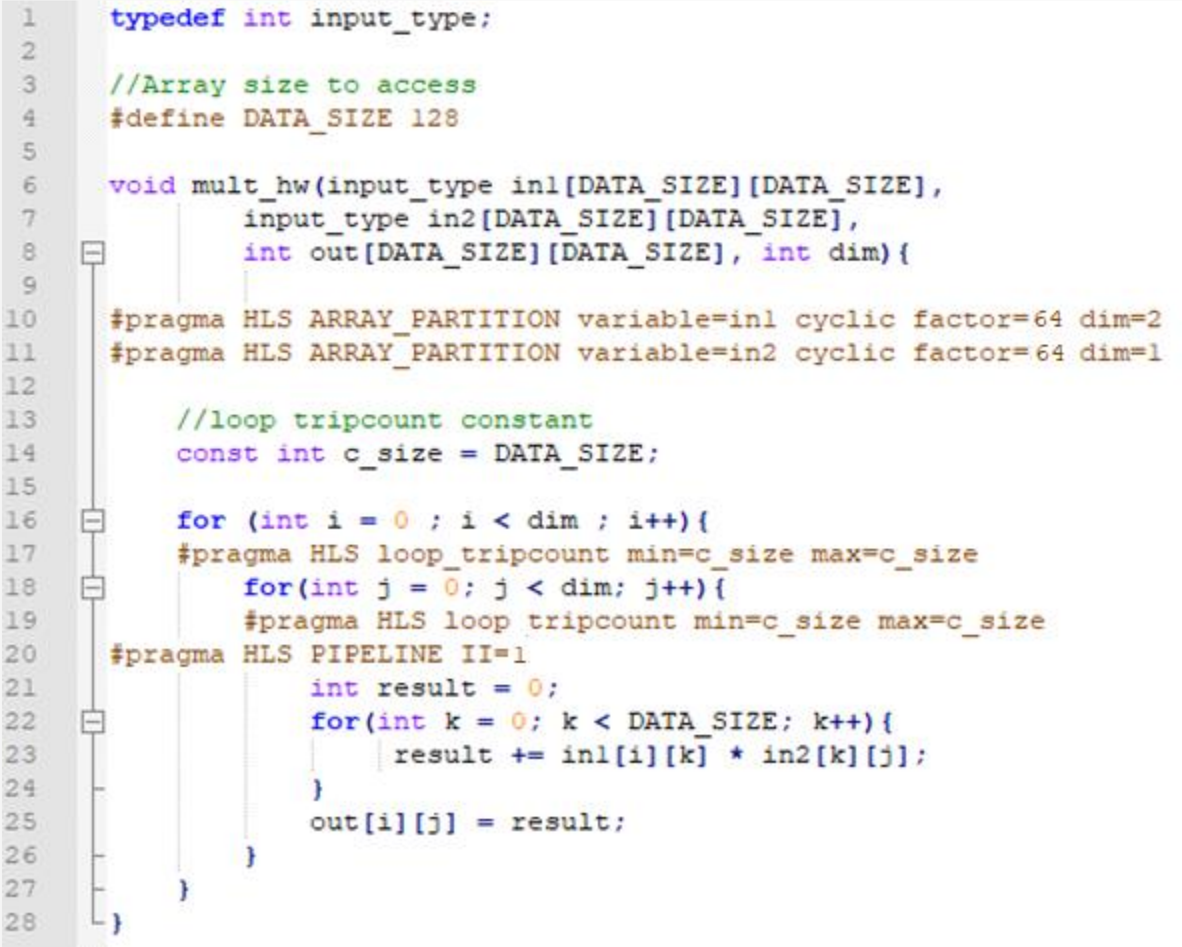
Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	-	-	-	-
Expression	-	384	0	6785	-
FIFO	-	-	-	-	-
Instance	-	-	-	-	-
Memory	-	-	-	-	-
Multiplexer	-	-	-	51	-
Register	-	-	1148	-	-
Total	0	384	1148	6836	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	0	5	~0	~0	0
Utilization SLR (%)	0	16	~0	1	0

Step 5. Optimizing the Design (Loop Pipelining)

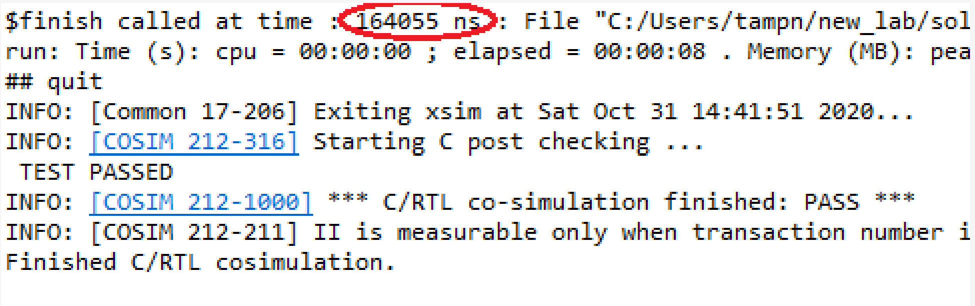


```
for (i=0; i<2; i++){
  Op_Read;
  Op_Execute;
  Op_Write;
}
```

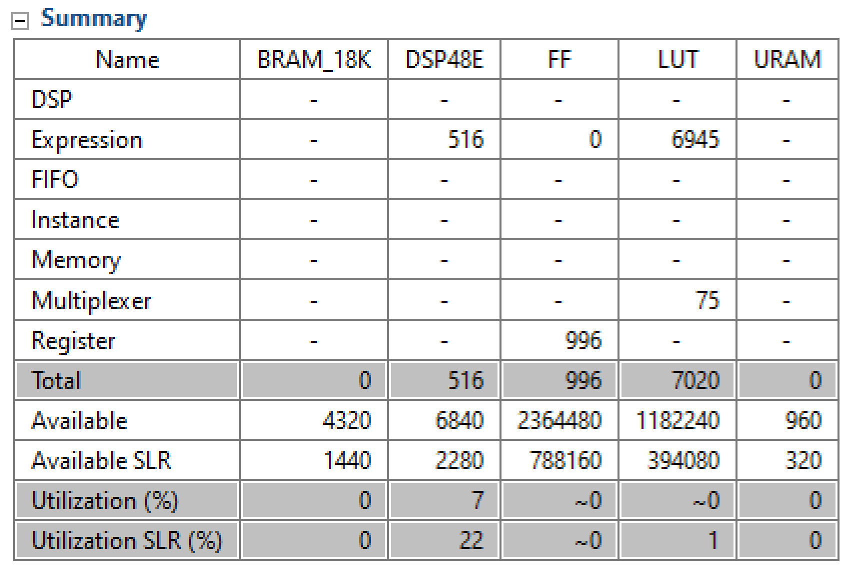
Step 5. Optimizing the Design (Loop Pipelining)



```
typedef int input_type;

//Array size to access
#define DATA_SIZE 128

void mult_hw(input_type in1[DATA_SIZE][DATA_SIZE],
    input_type in2[DATA_SIZE][DATA_SIZE],
    int out[DATA_SIZE][DATA_SIZE], int dim){

#pragma HLS ARRAY_PARTITION variable=in1 cyclic factor=64 dim=2
#pragma HLS ARRAY_PARTITION variable=in2 cyclic factor=64 dim=1

  //loop tripcount constant
  const int c_size = DATA_SIZE;

  for (int i = 0 ; i < dim ; i++){
  #pragma HLS loop_tripcount min=c_size max=c_size
    for(int j = 0; j < dim; j++){
    #pragma HLS loop_tripcount min=c_size max=c_size
#pragma HLS PIPELINE II=1
      int result = 0;
      for(int k = 0; k < DATA_SIZE; k++){
        result += in1[i][k] * in2[k][j];
      }
      out[i][j] = result;
    }
  }
}
```



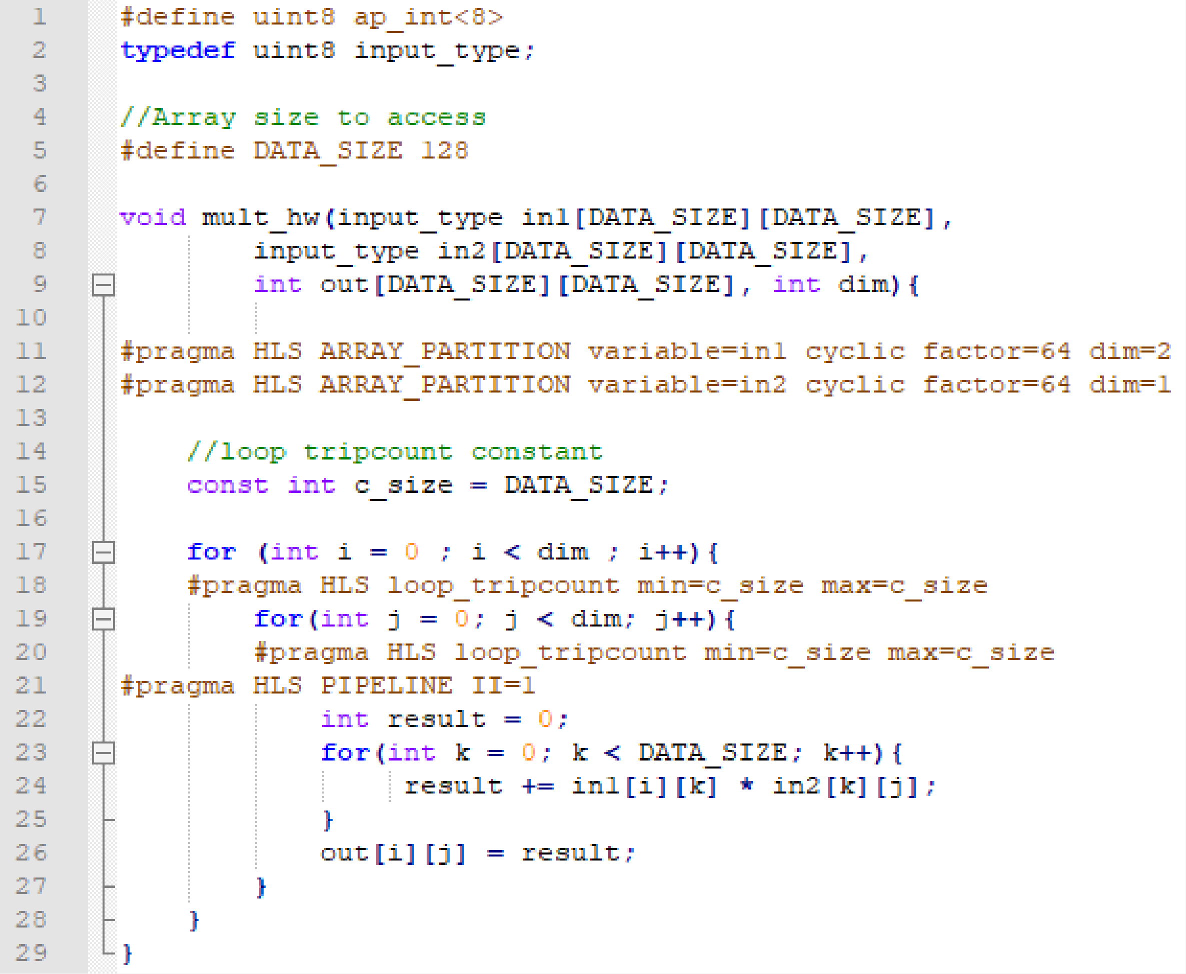
```
$finish called at time : 164055 ns : File "C:/Users/tampn/new_lab/sol
run: Time (s): cpu = 00:00:00 ; elapsed = 00:00:08 . Memory (MB): pea
## quit
INFO: [Common 17-206] Exiting xsim at Sat Oct 31 14:41:51 2020...
INFO: [COSIM 212-316] Starting C post checking ...
 TEST PASSED
INFO: [COSIM 212-1000] *** C/RTL co-simulation finished: PASS ***
INFO: [COSIM 212-211] II is measurable only when transaction number i
Finished C/RTL cosimulation.
```



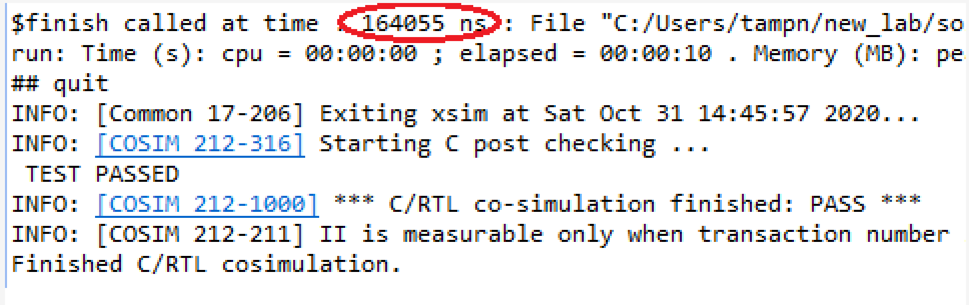
⊟ **Summary**

Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	-	-	-	-
Expression	-	516	0	6945	-
FIFO	-	-	-	-	-
Instance	-	-	-	-	-
Memory	-	-	-	-	-
Multiplexer	-	-	-	75	-
Register	-	-	996	-	-
Total	0	516	996	7020	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	0	7	~0	~0	0
Utilization SLR (%)	0	22	~0	1	0

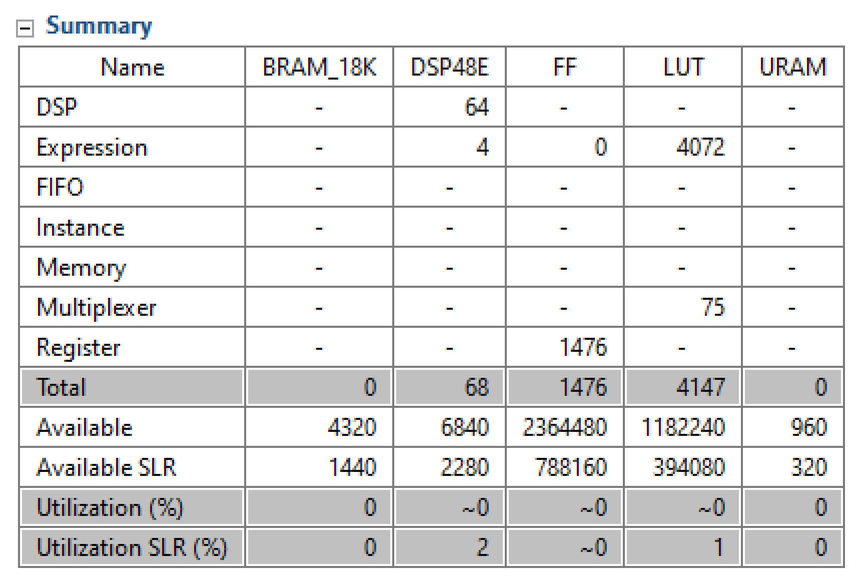
Step 5. Optimizing the Design (Arbitrary bit-widths)



```
#define uint8 ap_int<8>
typedef uint8 input_type;

//Array size to access
#define DATA_SIZE 128

void mult_hw(input_type in1[DATA_SIZE][DATA_SIZE],
        input_type in2[DATA_SIZE][DATA_SIZE],
        int out[DATA_SIZE][DATA_SIZE], int dim){

#pragma HLS ARRAY_PARTITION variable=in1 cyclic factor=64 dim=2
#pragma HLS ARRAY_PARTITION variable=in2 cyclic factor=64 dim=1

    //loop tripcount constant
    const int c_size = DATA_SIZE;

    for (int i = 0 ; i < dim ; i++){
    #pragma HLS loop_tripcount min=c_size max=c_size
        for(int j = 0; j < dim; j++){
        #pragma HLS loop_tripcount min=c_size max=c_size
#pragma HLS PIPELINE II=1
            int result = 0;
            for(int k = 0; k < DATA_SIZE; k++){
                result += in1[i][k] * in2[k][j];
            }
            out[i][j] = result;
        }
    }
}
```



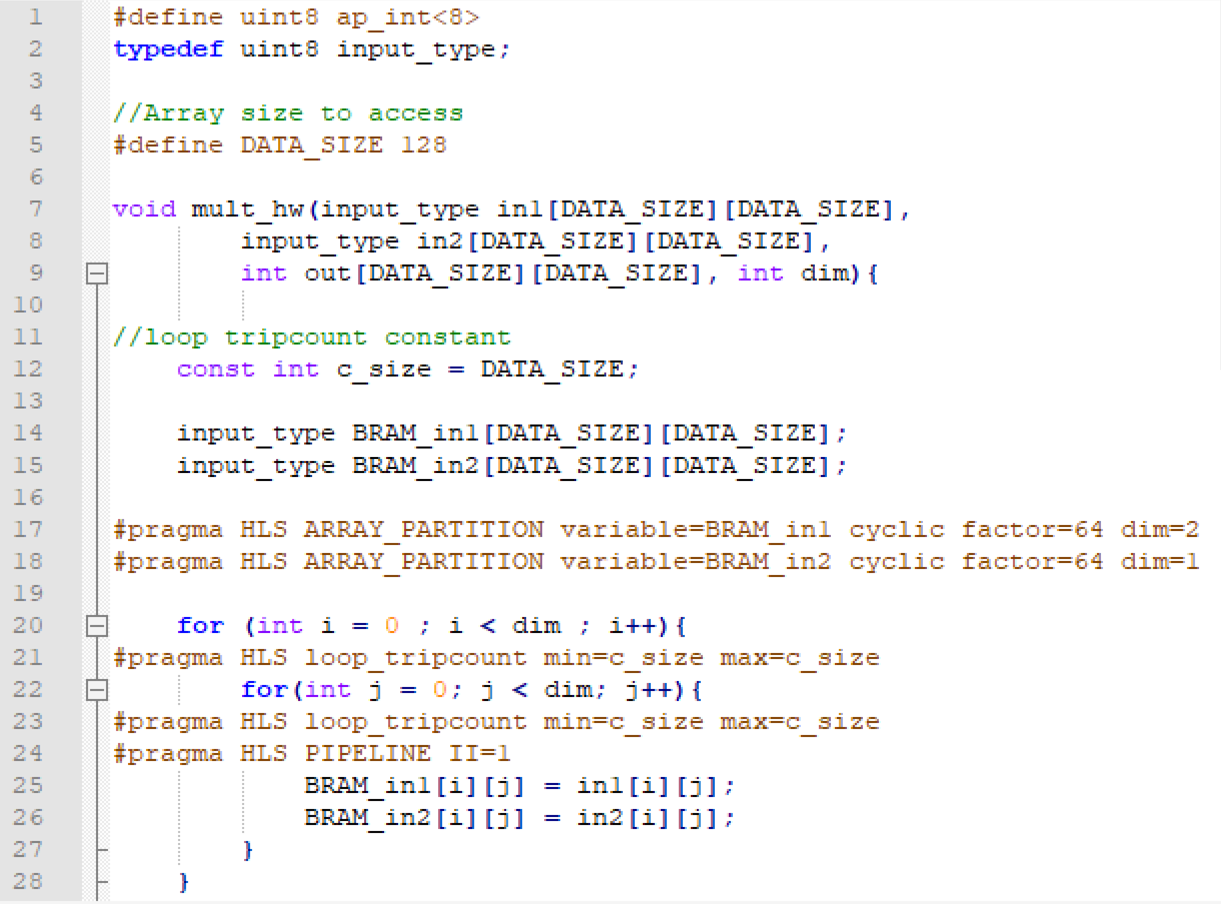
```
$finish called at time  164055 ns: File "C:/Users/tampn/new_lab/so
run: Time (s): cpu = 00:00:00 ; elapsed = 00:00:10 . Memory (MB): pe
## quit
INFO: [Common 17-206] Exiting xsim at Sat Oct 31 14:45:57 2020...
INFO: [COSIM 212-316] Starting C post checking ...
 TEST PASSED
INFO: [COSIM 212-1000] *** C/RTL co-simulation finished: PASS ***
INFO: [COSIM 212-211] II is measurable only when transaction number
Finished C/RTL cosimulation.
```



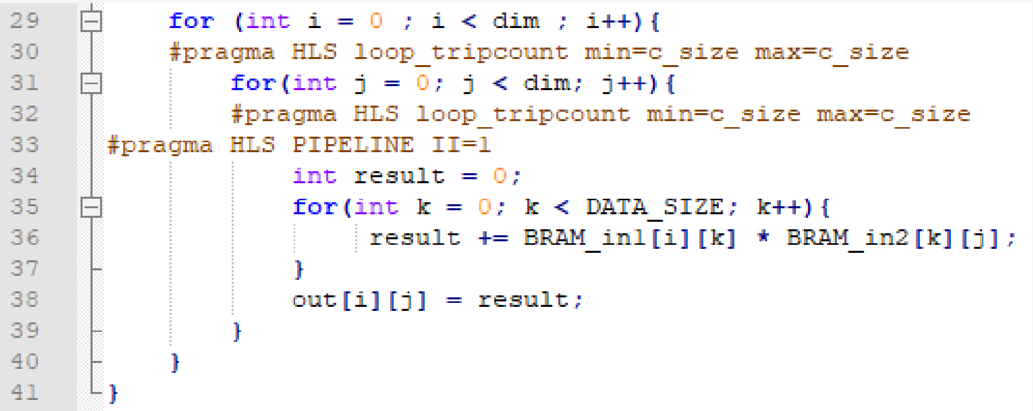
⊟ Summary

Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	64	-	-	-
Expression	-	4	0	4072	-
FIFO	-	-	-	-	-
Instance	-	-	-	-	-
Memory	-	-	-	-	-
Multiplexer	-	-	-	75	-
Register	-	-	1476	-	-
Total	0	68	1476	4147	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	0	~0	~0	~0	0
Utilization SLR (%)	0	2	~0	1	0

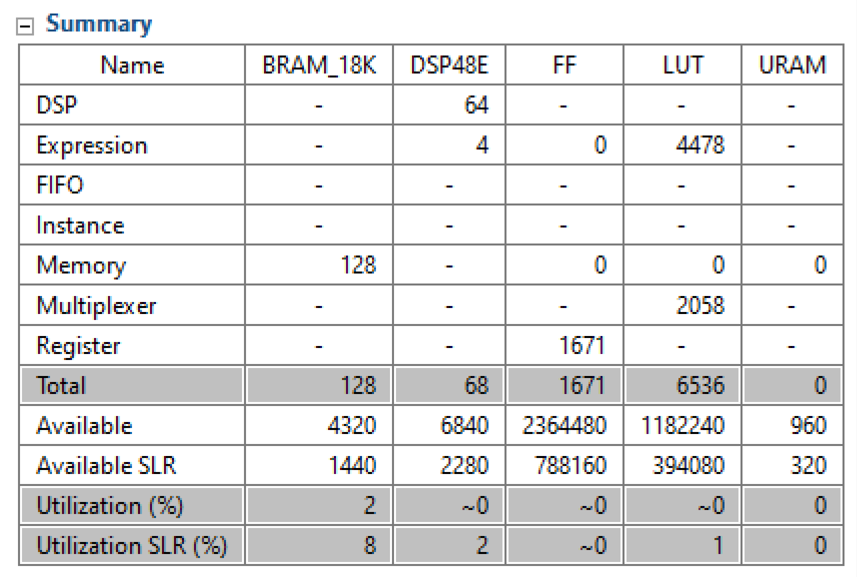
Step 5. Optimizing the Design (BRAMs)



```
#define uint8 ap_int<8>
typedef uint8 input_type;

//Array size to access
#define DATA_SIZE 128

void mult_hw(input_type in1[DATA_SIZE][DATA_SIZE],
        input_type in2[DATA_SIZE][DATA_SIZE],
        int out[DATA_SIZE][DATA_SIZE], int dim){

//loop tripcount constant
    const int c_size = DATA_SIZE;

    input_type BRAM_in1[DATA_SIZE][DATA_SIZE];
    input_type BRAM_in2[DATA_SIZE][DATA_SIZE];

#pragma HLS ARRAY_PARTITION variable=BRAM_in1 cyclic factor=64 dim=2
#pragma HLS ARRAY_PARTITION variable=BRAM_in2 cyclic factor=64 dim=1

    for (int i = 0 ; i < dim ; i++){
#pragma HLS loop_tripcount min=c_size max=c_size
        for(int j = 0; j < dim; j++){
#pragma HLS loop_tripcount min=c_size max=c_size
#pragma HLS PIPELINE II=1
            BRAM_in1[i][j] = in1[i][j];
            BRAM_in2[i][j] = in2[i][j];
        }
    }
```

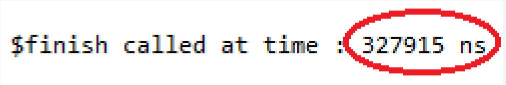


```
    for (int i = 0 ; i < dim ; i++){
    #pragma HLS loop_tripcount min=c_size max=c_size
        for(int j = 0; j < dim; j++){
        #pragma HLS loop_tripcount min=c_size max=c_size
#pragma HLS PIPELINE II=1
            int result = 0;
            for(int k = 0; k < DATA_SIZE; k++){
                result += BRAM_in1[i][k] * BRAM_in2[k][j];
            }
            out[i][j] = result;
        }
    }
}
```



Summary

Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	64	-	-	-
Expression	-	4	0	4478	-
FIFO	-	-	-	-	-
Instance	-	-	-	-	-
Memory	128	-	0	0	0
Multiplexer	-	-	-	2058	-
Register	-	-	1671	-	-
Total	128	68	1671	6536	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	2	~0	~0	~0	0
Utilization SLR (%)	8	2	~0	1	0



```
$finish called at time : 327915 ns
```