

ghp\_vgeV2JopdBps60XxHHDLVKfwAcYU1i4eyB3d

# AI-accelerator-course-preparation tutorial

---

## AI-accelerator-course-preparation tutorial

### Lab 1 Software platform Description

- 1.1 Goal
- 1.2 Vivado® Design Suite
  - Introduction
  - Component
    - Vivado High-Level Synthesis compiler
    - Vivado Simulator
    - Vivado IP Integrator
    - Vivado Tcl Store
- 1.3 Resource downloading
- 1.4 Final installation and licensing
  - Windows version
  - Linux version
  - Licensing

### Lab 2 Verilog & VHDL:

- 2.1 Goal
- 2.1 Design Flow
  - Project creating
  - Source Adding
  - RTL Analysis
- 2.3 Finite State Machine
- 2.4 Testbench
  - Timescale definition and variable declaration
  - DUT Instantiating
  - Generating clock
  - Simulation
- 2.5 Assignment
  - Assignment 1
  - Assignment 2
  - Assignment 3

### Lab 3 Convolutional Layer

- 3.1 Goal
- 3.2 Introduction
  - What is **convolution**?
    - Linear formula
    - Discrete formula
  - Convolution in Image Processing
- 3.3 Assignment
  - Assignment 1
  - Assignment 2
  - Assignment 3

### Lab 4 HLS

- 4.1 Goal
- 4.2 Introduction
  - HLS introduction
  - Hardware description-AVNET Ultra96-V2
  - Software development kit

# Lab 1 Software platform Description

---

## 1.1 Goal

---

This part of the documentation provides an introduction for fresh users about how to install the Vivado® Design Suite on Windows or Linux OS(operating system).

## 1.2 Vivado® Design Suite

---

### Introduction

Vivado® Design Suite is a software suite produced by [Xilinx](#) for synthesis and analysis of [HDL](#)(Hardware description language) designs, superseding [Xilinx ISE](#) with additional features for SoC([system on a chip](#) ) development and HLS([high-level synthesis](#)).

It delivers a SoC-strength, IP-centric and system-centric, next generation development environment that has been built from the ground up to address the productivity bottlenecks in system-level integration and implementation. It comes in three editions:

- Vivado HL WebPack Edition
- Vivado HL Design Edition
- Vivado HL System Edition

### Component

#### Vivado High-Level Synthesis compiler

The **Vivado High-Level Synthesis** compiler enables [C](#), [C++](#) and [SystemC](#) programs to be directly targeted into Xilinx devices without the need to manually create RTL. Vivado HLS is widely reviewed to increase developer productivity, and is confirmed to support C++ classes, templates, functions and operator overloading.

#### Vivado Simulator

The **Vivado Simulator** is a component of the Vivado Design Suite. It is a compiled-language simulator that supports mixed-language, [Tcl](#) scripts, encrypted IP and enhanced verification.

#### Vivado IP Integrator

The **Vivado IP Integrator** allows engineers to quickly integrate and configure IP from the large Xilinx IP library. The Integrator is also tuned for [MathWorks Simulink](#) designs built with Xilinx's System Generator and Vivado High-Level Synthesis.

#### Vivado Tcl Store

The **Vivado Tcl Store** is a scripting system for developing add-ons to Vivado, and can be used to add and modify Vivado's capabilities. Tcl is the scripting language on which Vivado itself is based. All of Vivado's underlying functions can be invoked and controlled via Tcl scripts.

## 1.3 Resource downloading

---

Please check the [link](#) to download the required version of Vivado® Design Suite.

**Note:**

1. Download from the official website may require you register for an Xilinx account.
2. Suggested version of Vivado® Design Suite is **2020.1**, which i used to design and verify the lab and code. It can not be guaranteed that the lab code can work on all the versions of suite.
3. Strongly suggest reading the **release note** file under the documentation catalog, it will save your time for the platform preparation including the downloading, installation and the later licensing.

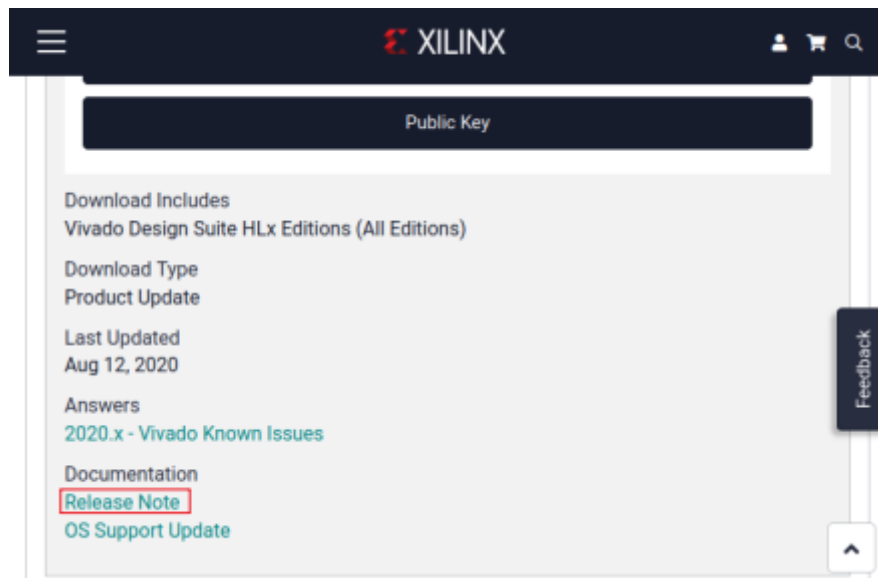


Figure 1.1

## 1.4 Final installation and licensing

After successful downloading of Vivado® Design Suite installer, it can be ready for installation. Please check the **release note** again to make sure your OS version can support the downloaded Vivado® Design Suite installer.

Here the documentation for Windows and Linux Ubuntu based installation instruction, you can choose from it, and skip another OS's description.

### Windows version

If you downloaded the lightweight installer, launch the downloaded file. You are prompted to log in and use your regular Xilinx login credentials to continue with the installation process.

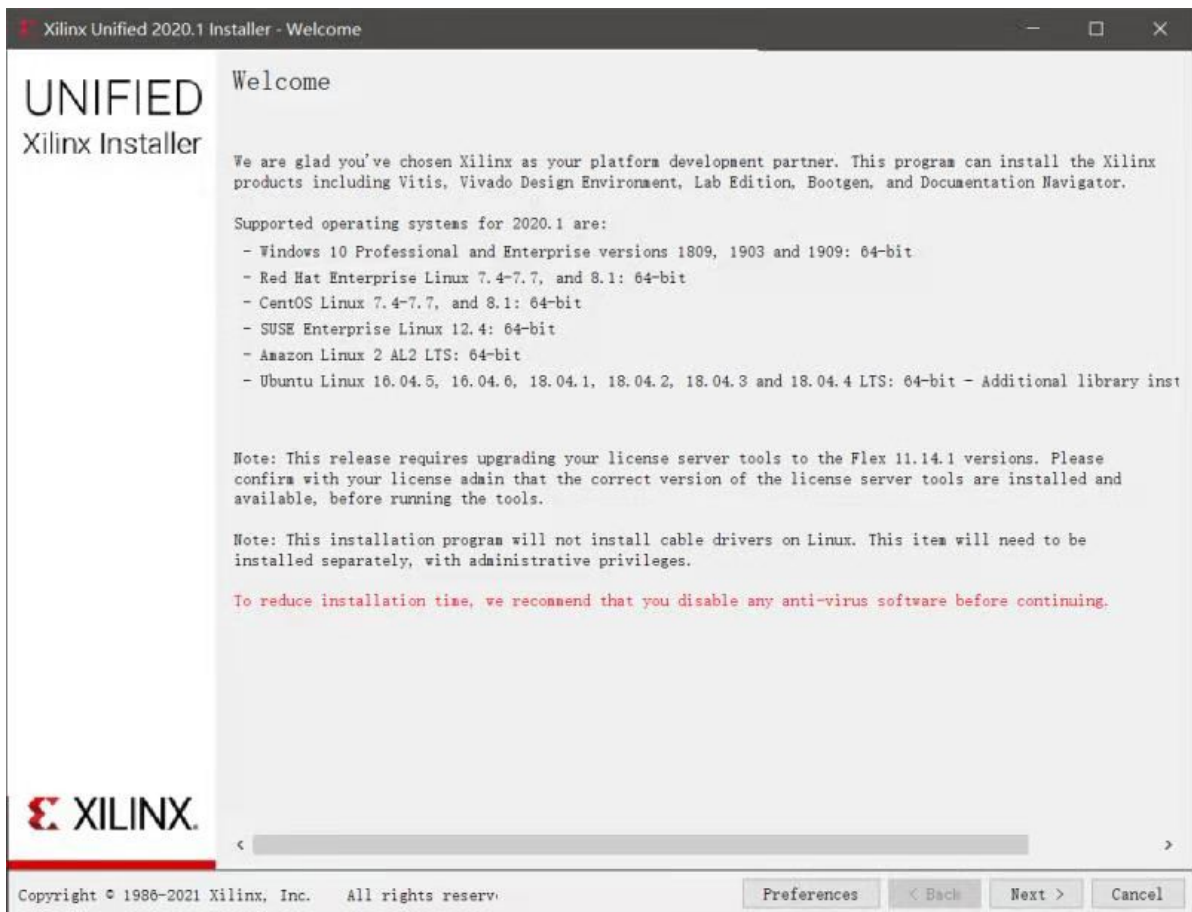


Figure 1.2

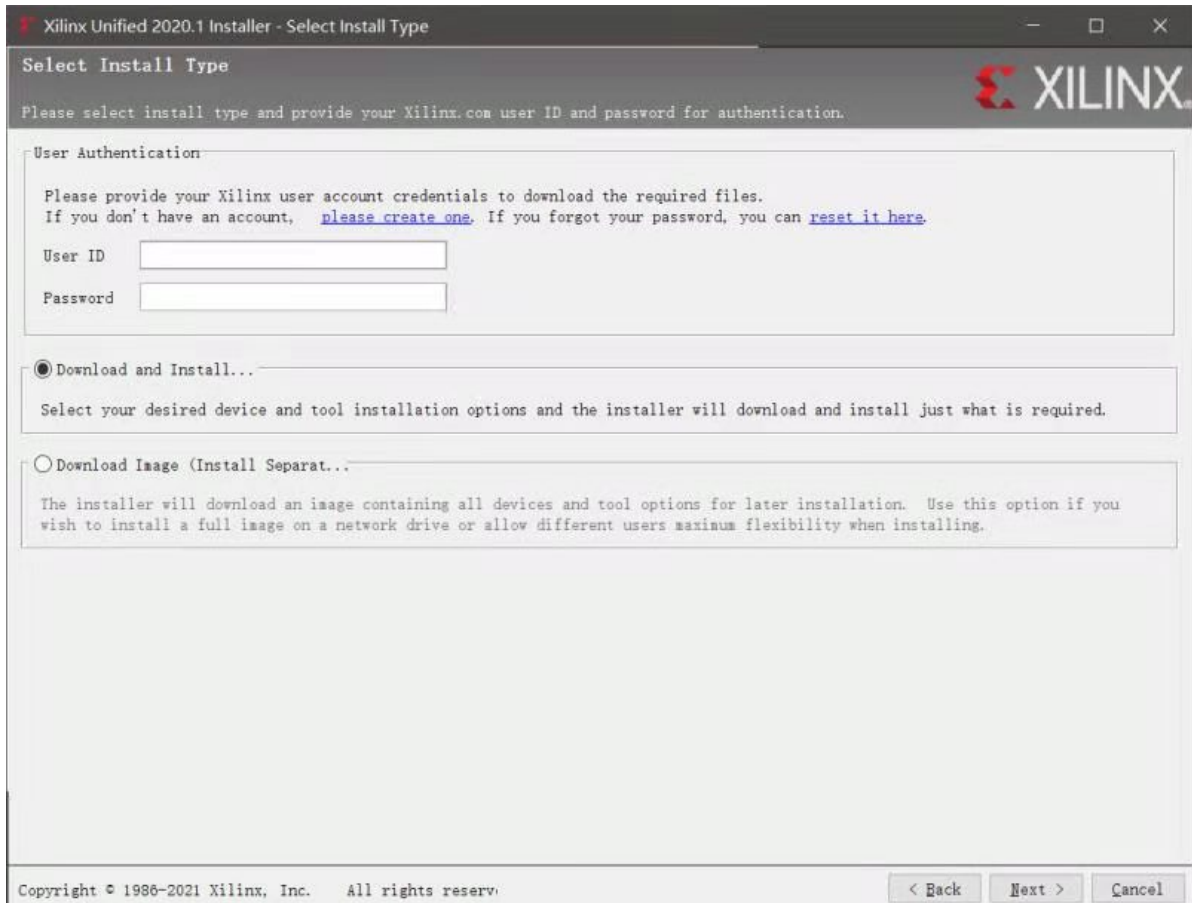


Figure 1.3

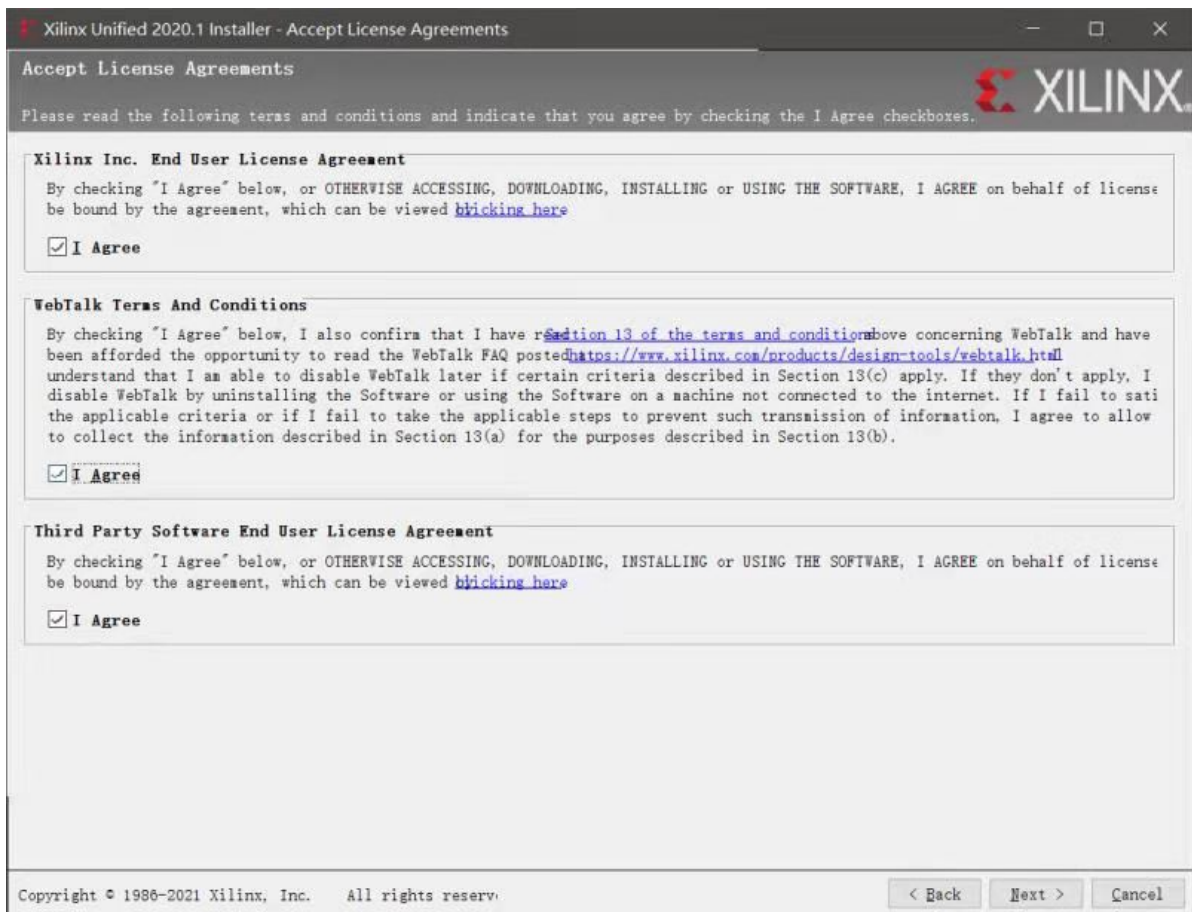


Figure 1.4

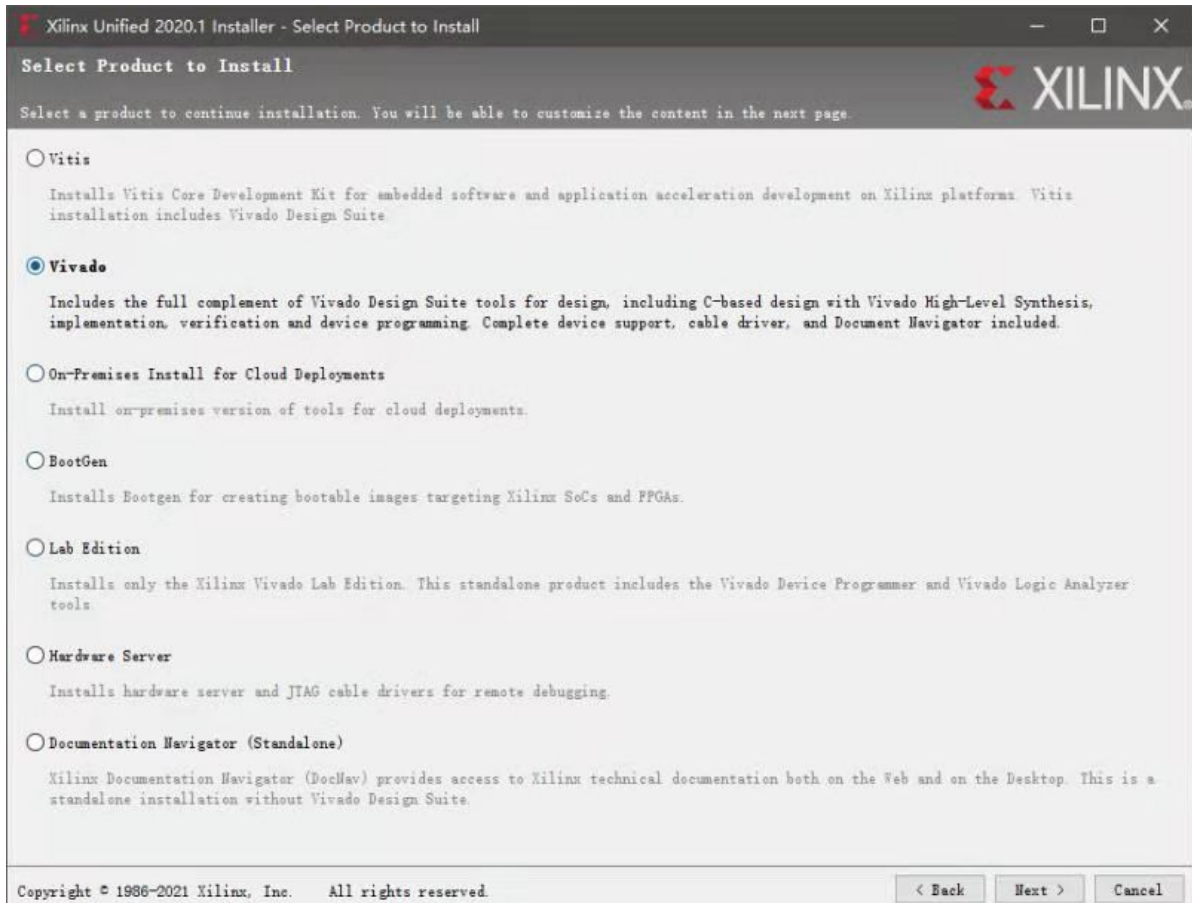


Figure 1.5

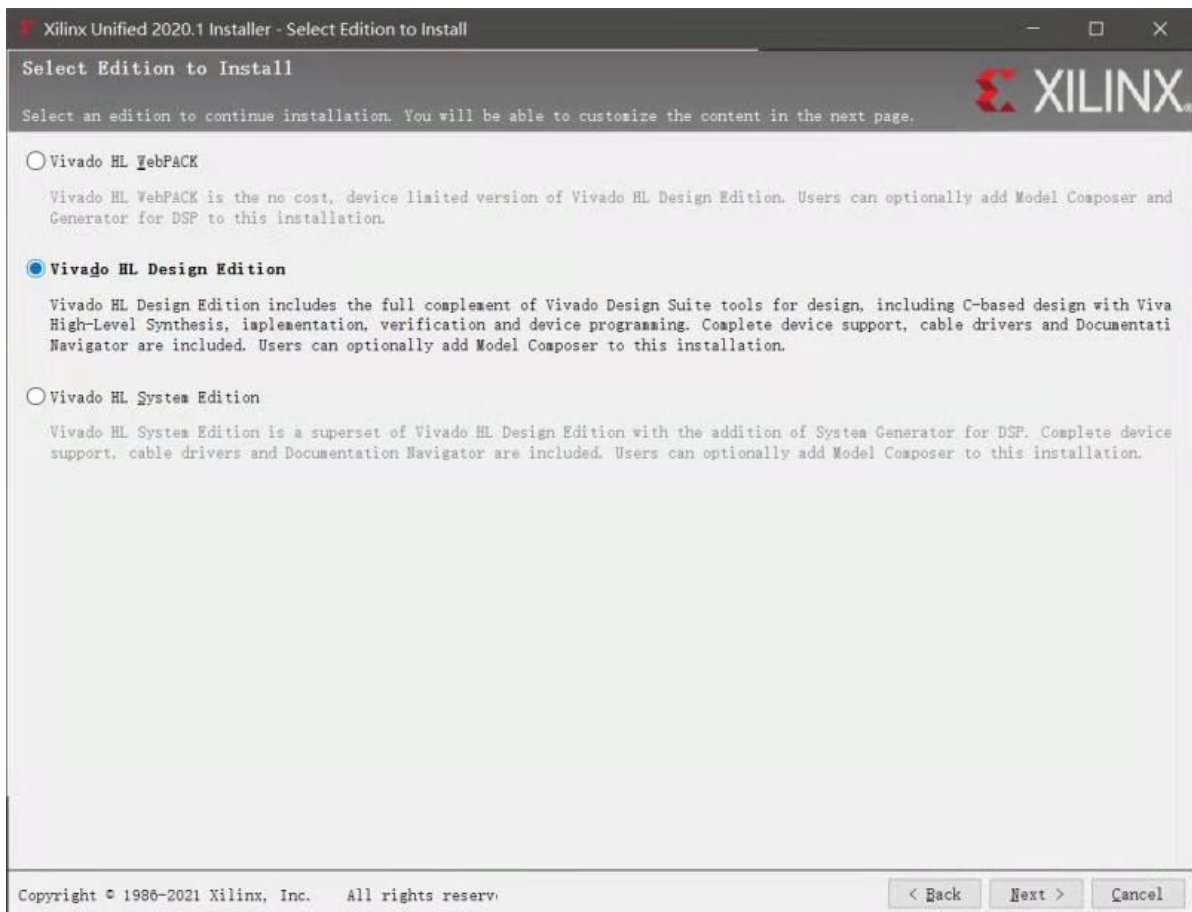


Figure 1.6

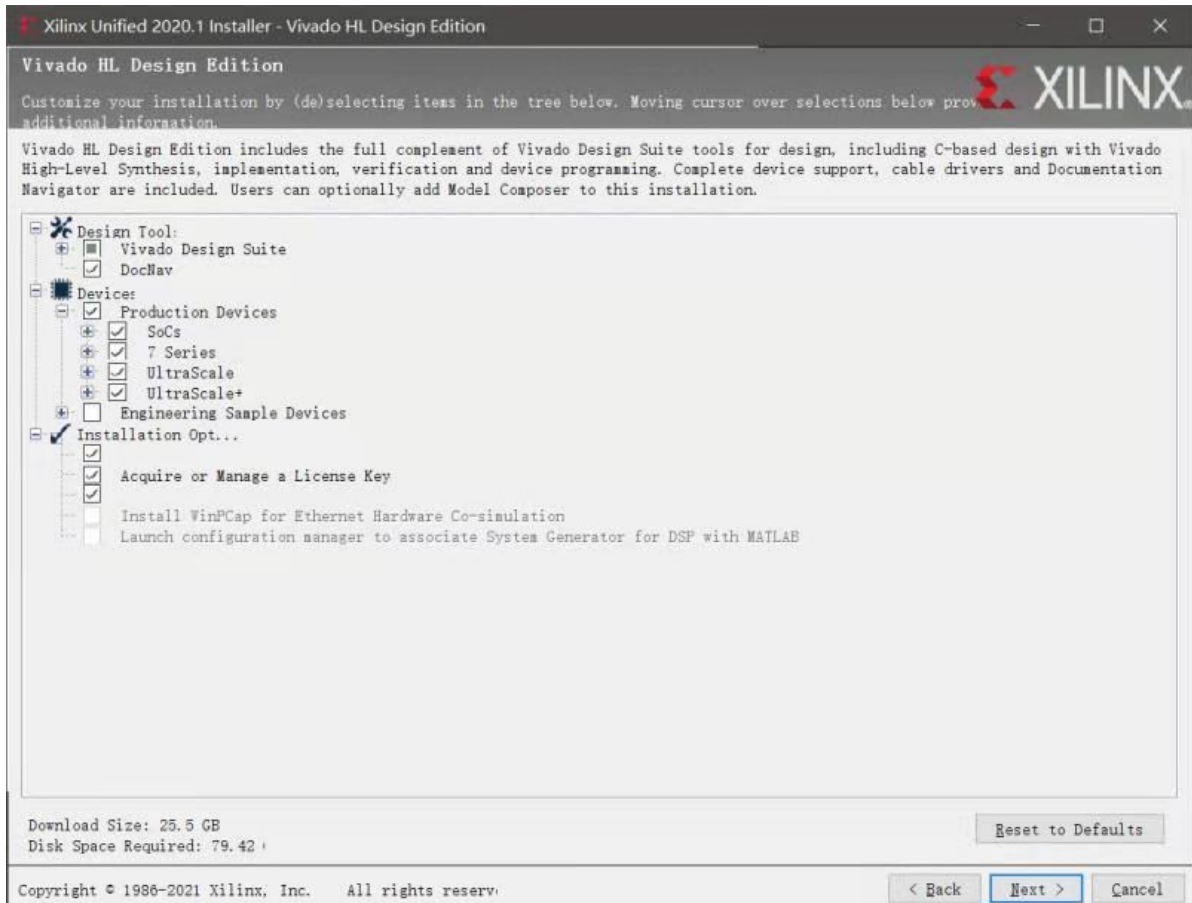


Figure 1.7



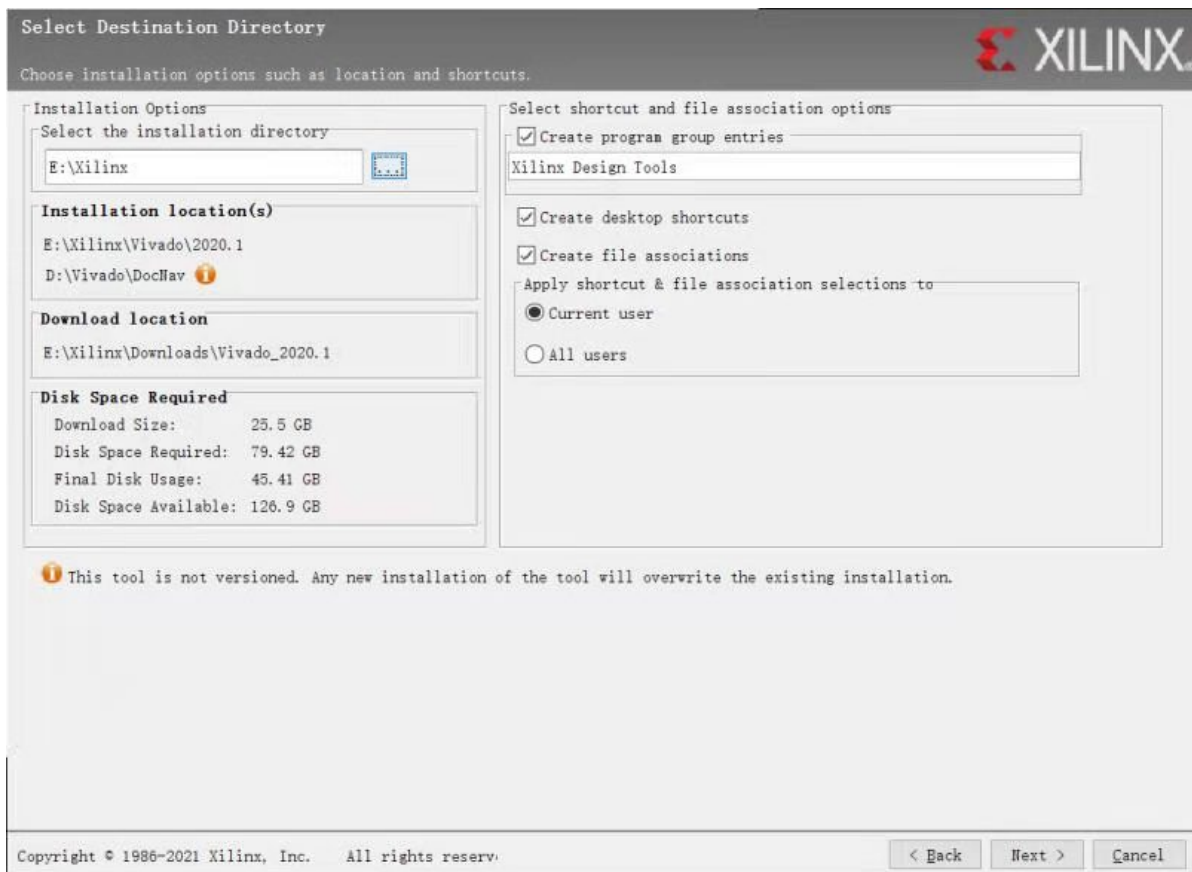


Figure 1.8

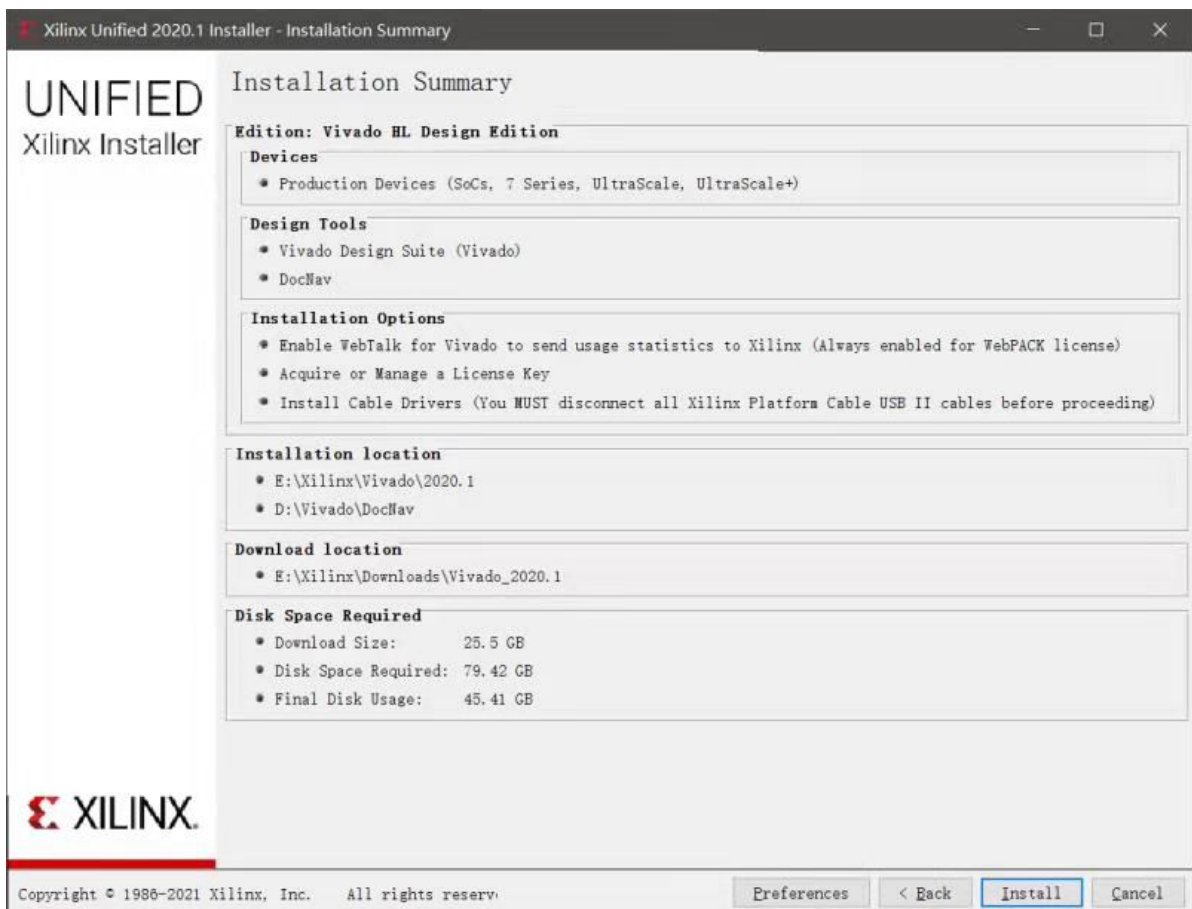


Figure 1.9



## Linux version

Find the mirror image in the [resoure link](#): Ubuntu Linux's installation is not difficult. Two options are available:

1. Direct installation on your PC or Laptop, but it will erase all the files from your previous OS.
2. Virtual Machine(Oracle or VM workstation).

For option 1, please refer to this [guide](#) for USB booting or DVD formatting.

If you downloaded the lightweight installer, launch the downloaded file. You are prompted to log in and use your regular Xilinx login credentials to continue with the installation process.

**Note:** On Linux the file is a .bin file and can be launched by running `./bin` in the shell terminal, as can be seen in [Figure 1.10][#110] .

Please ensure that you have changed the file permissions to execute.

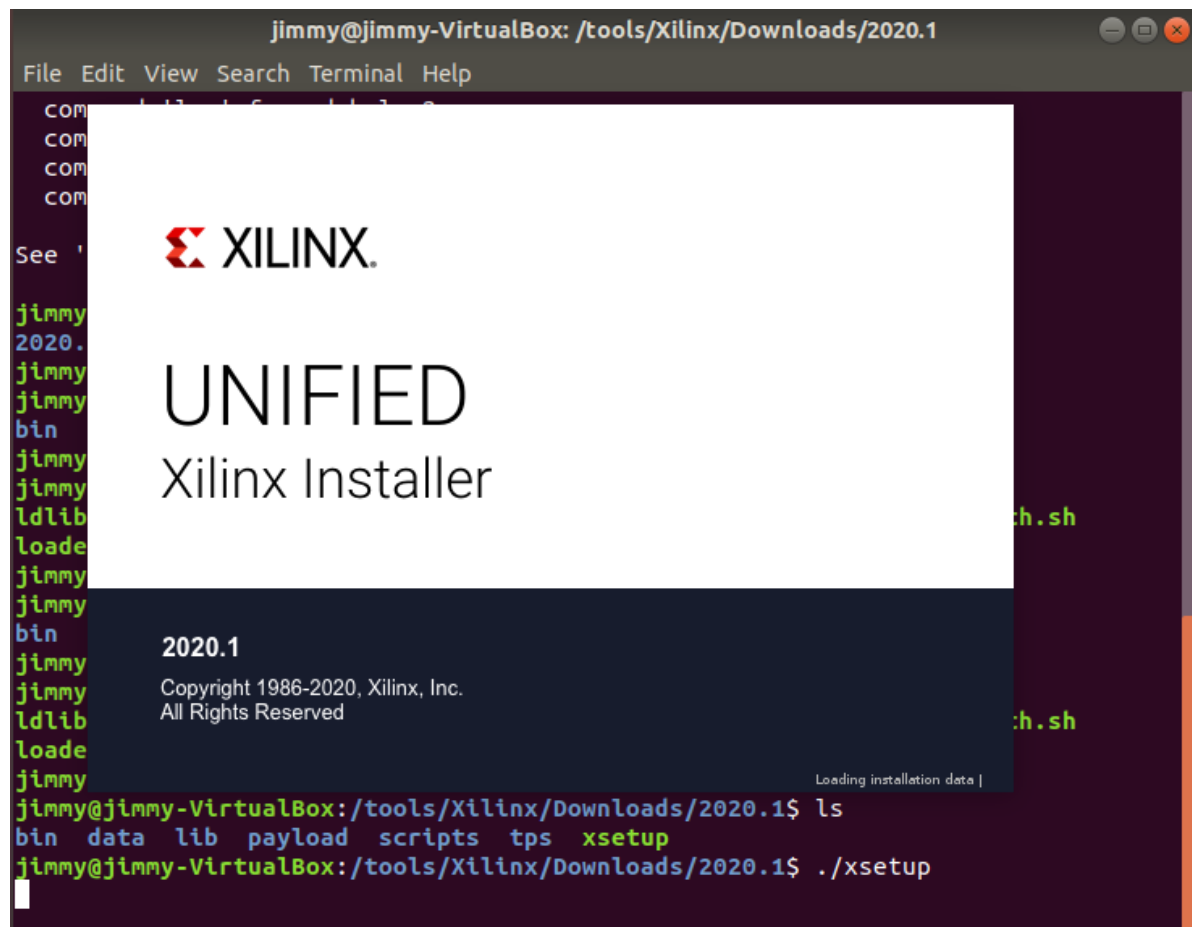


Figure 1.10

Following the initializing of the installer, a welcome window will appear with the necessary information about the supported operating systems, shown in [Figure 1.11][111].

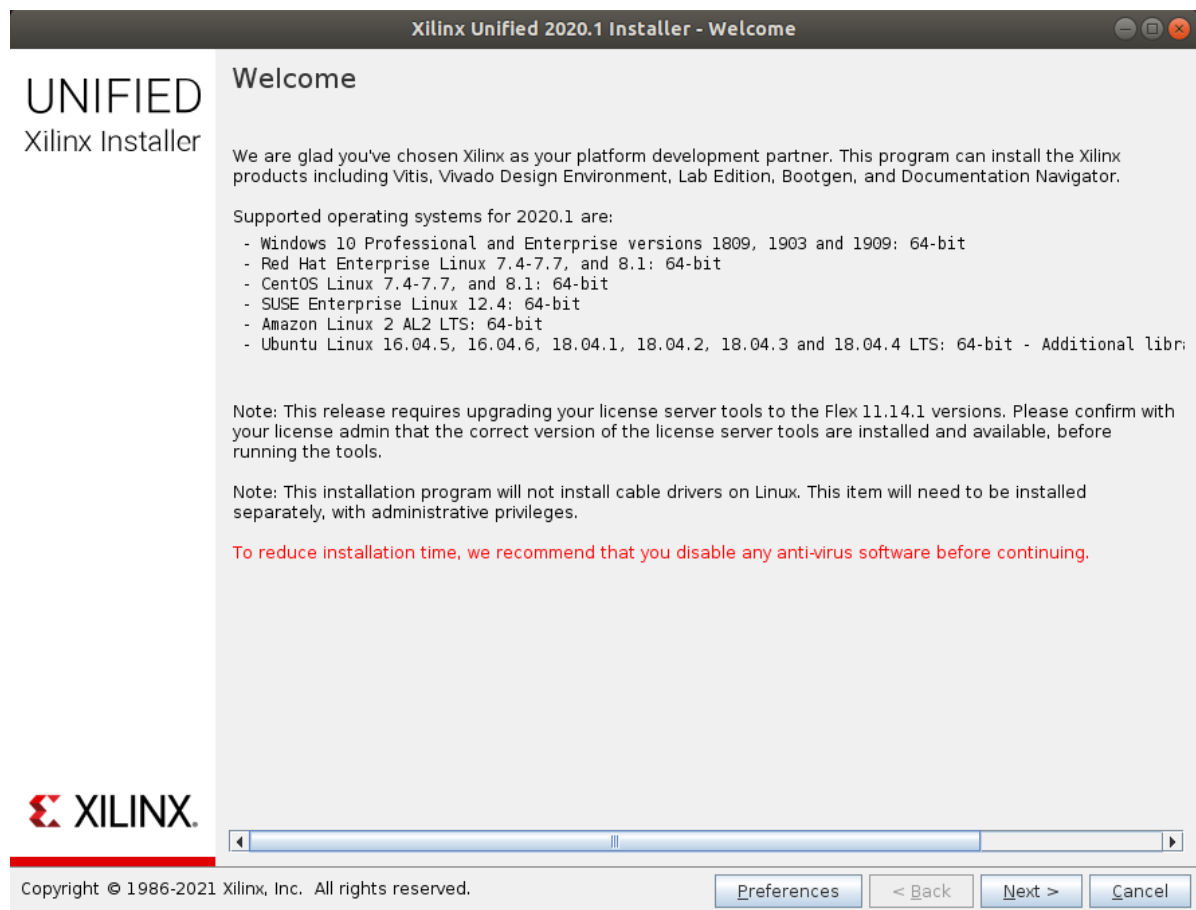


Figure 1.11

Accept all the license agreement, otherwise the installation will not proceed, displayed in Figure 1.12.

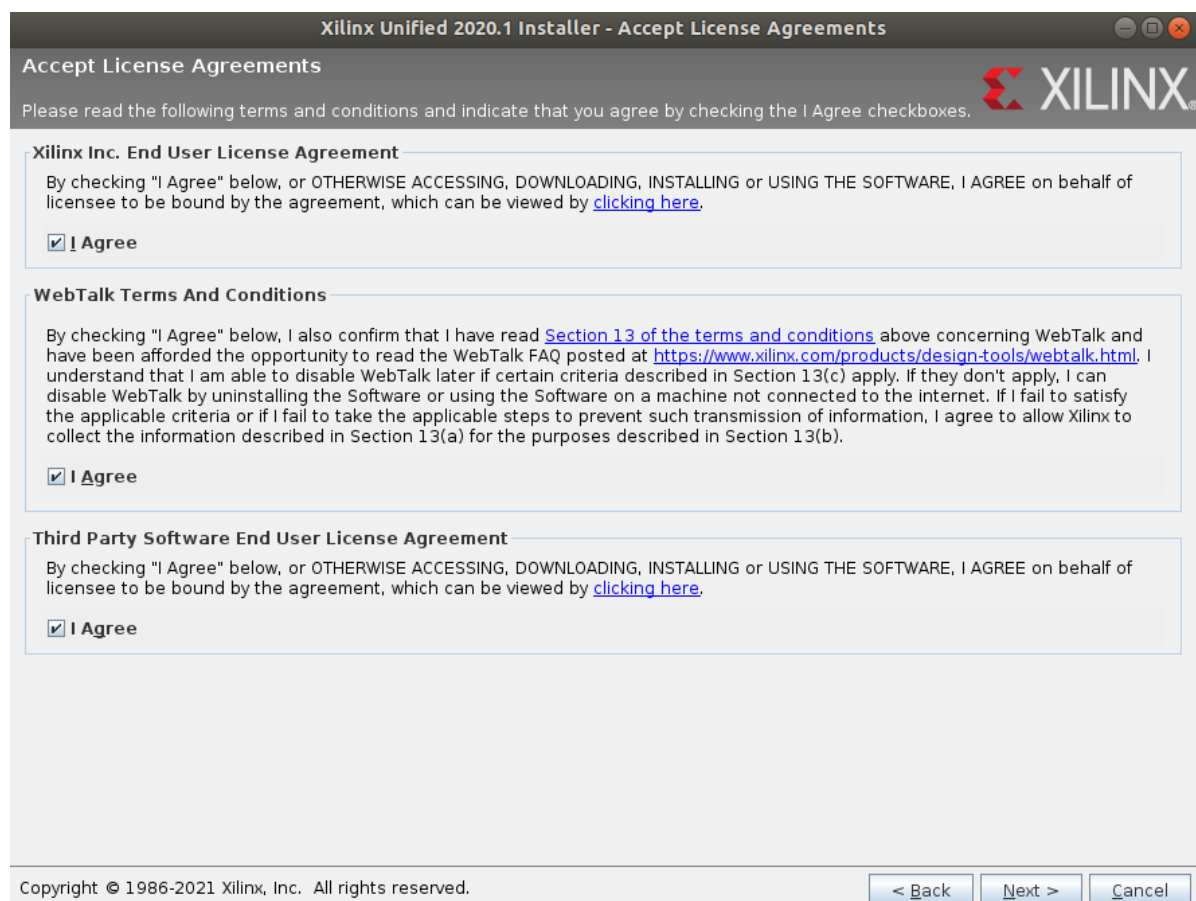


Figure 1.12

Shown in Figure 4, take the Vivado product, including all the required toolkits and library. To experience the Machine learning and AIoT function, you can take the Vitis option, which also includes the Vivado Design Suite.

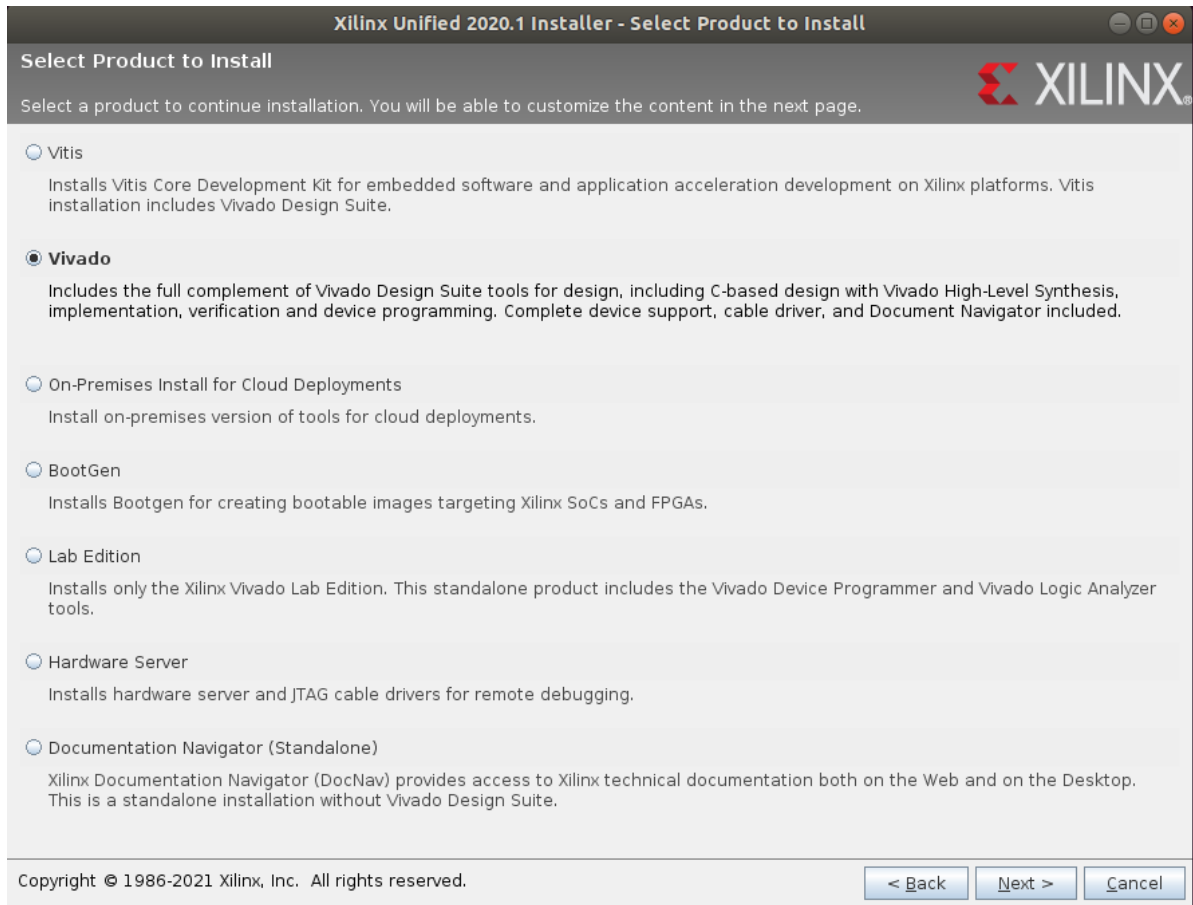


Figure 1.13

Choose the Vivado HL System Edition.

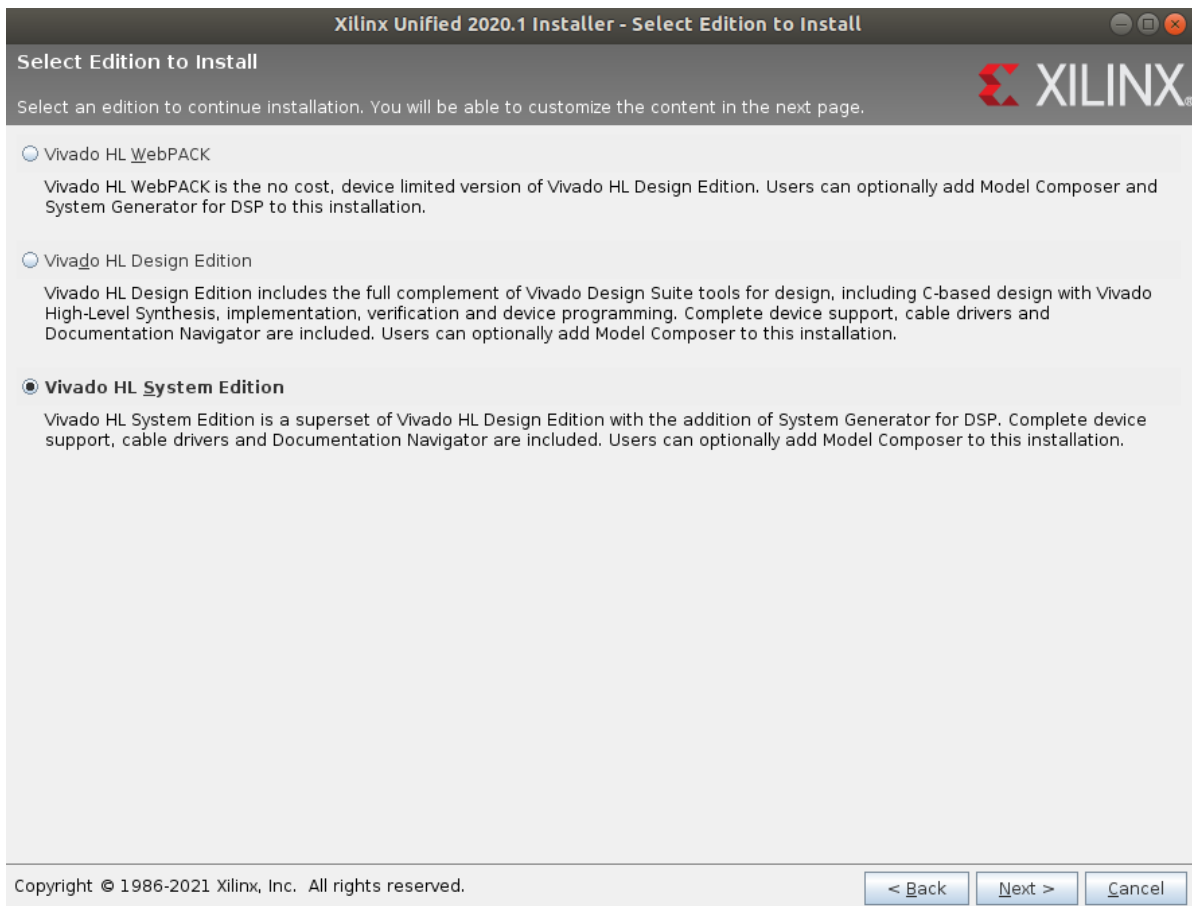


Figure 1.14

After the Edition version is set, the installer will display the available design tools for your reference. Here i suggest selecting all the tools and devices, shown in Figure 6.

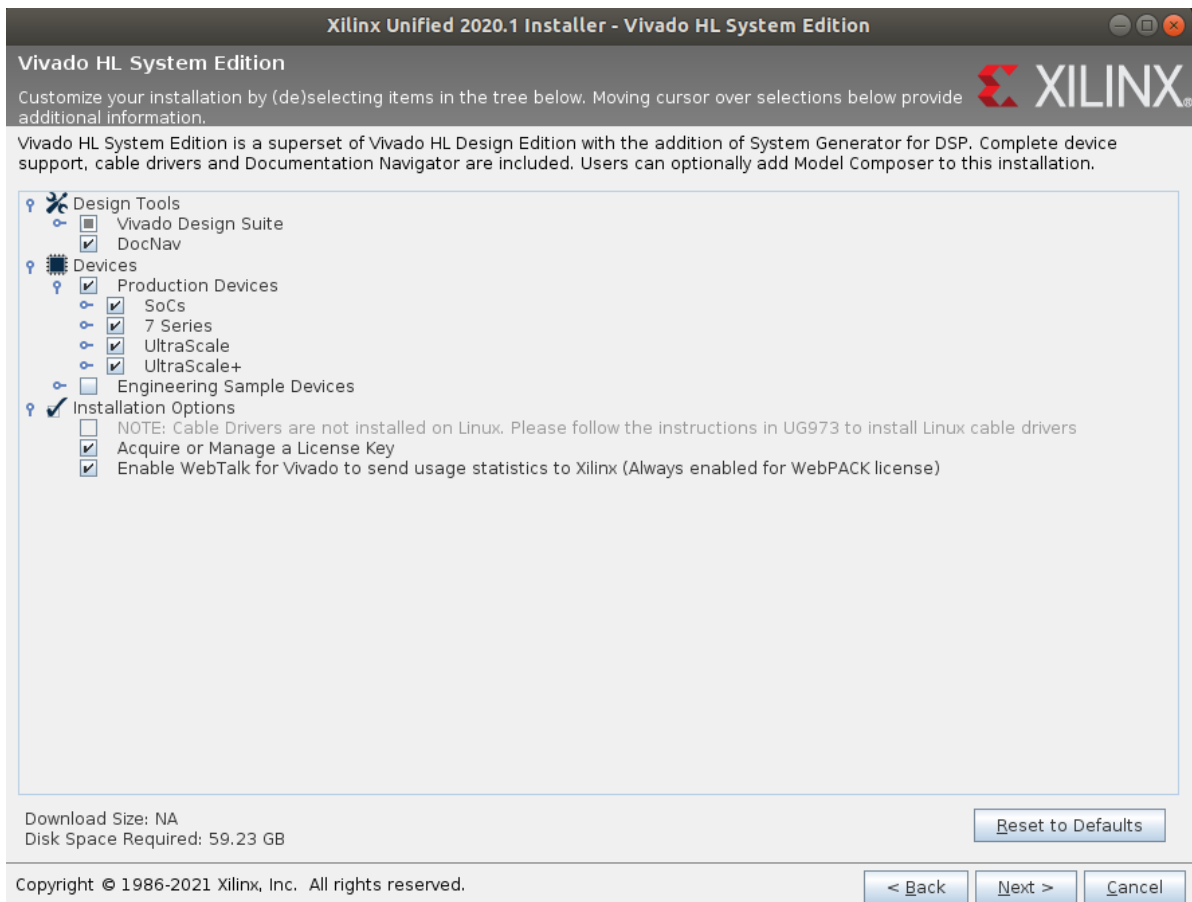


Figure 1.15

Choose the installation location.

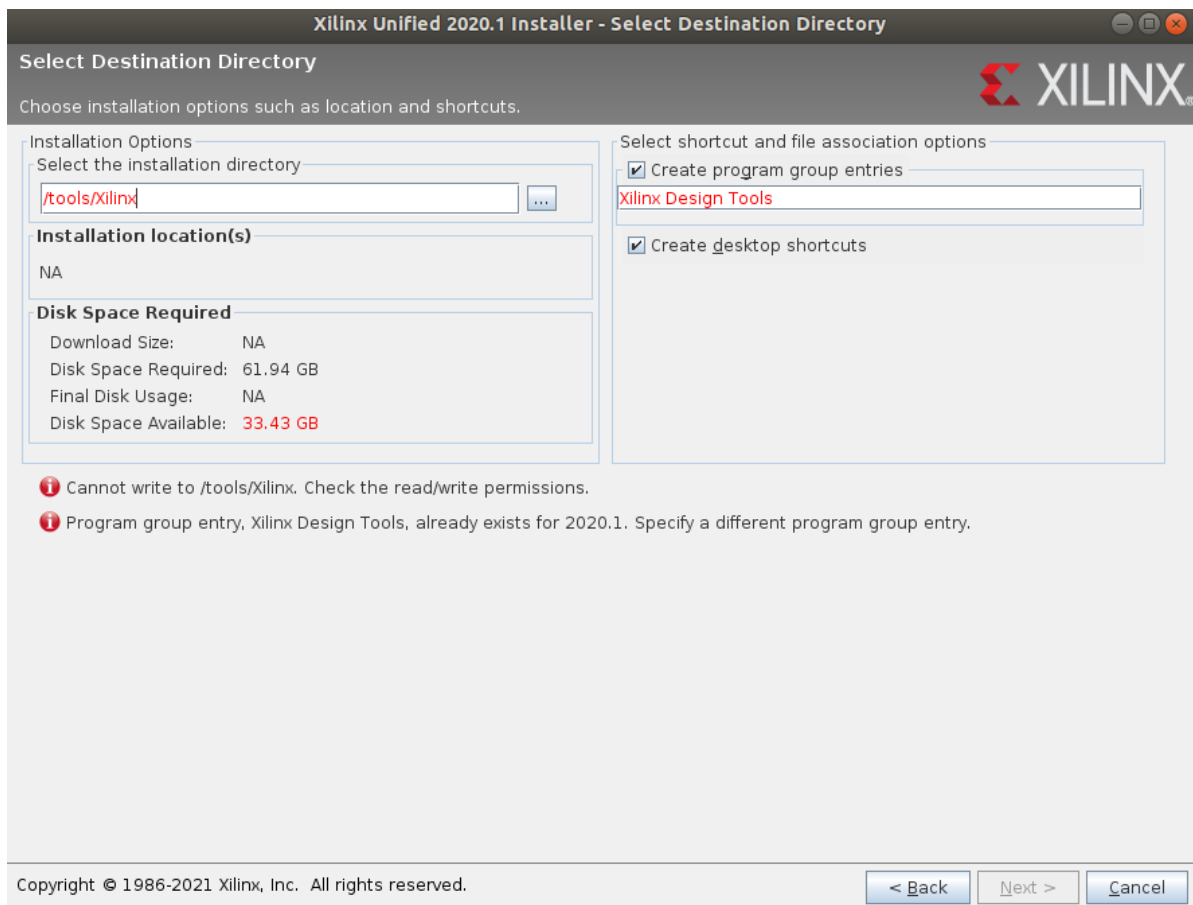


Figure 1.16

Finally the installation will automatically progress and logically it will succeed. If not, the problem might link to the disk room for storage and processing or the system's compatibility.

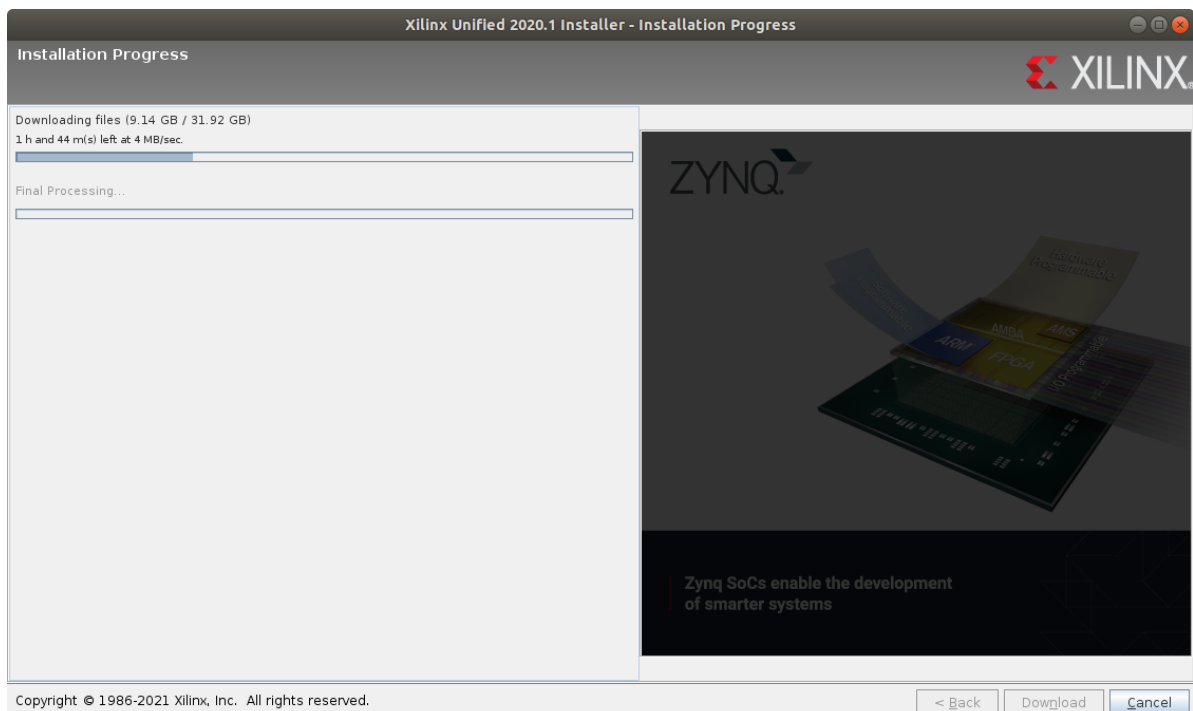


Figure 1.17

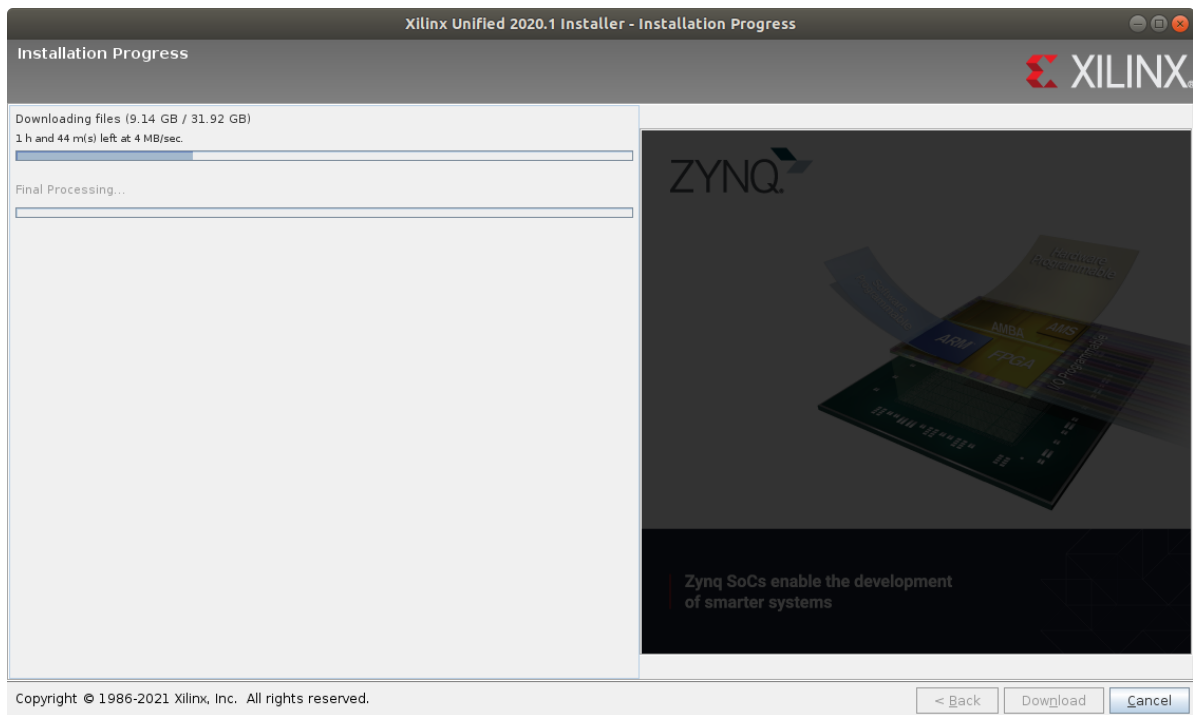


Figure 1.18

## Licensing

Xilinx charges for IP design and software and profits from the business. Currently, you can have one month's trial free of charge with your registered information.

After completing the installation of Vivado, SDx or ISE Design Suite, the Xilinx License Configuration Manager (XCLM) will start automatically and guide you through the licensing process, displayed in Figure 9. You may also go directly to the [Xilinx Product Licensing Site](#) to obtain licenses for free or evaluation products if you decided to skip this step during product installation.

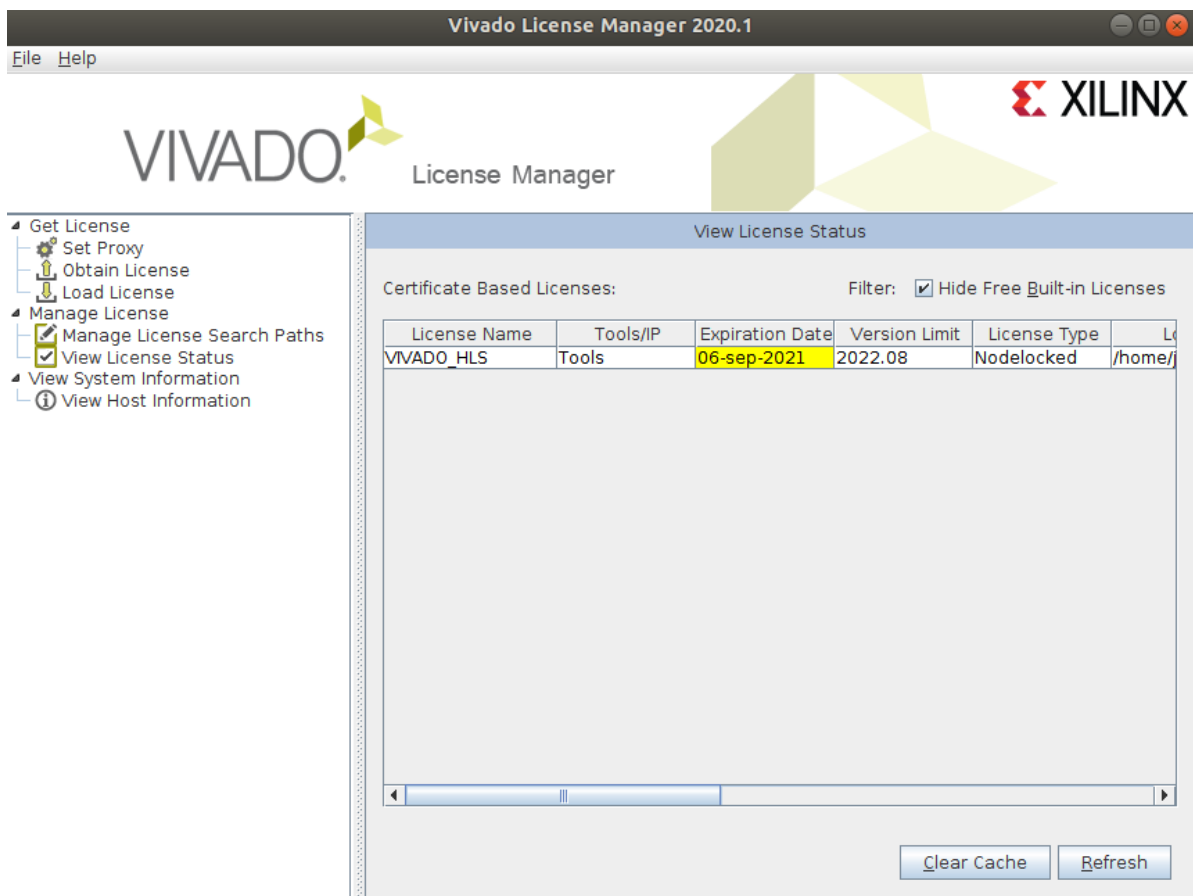


Figure 1.19

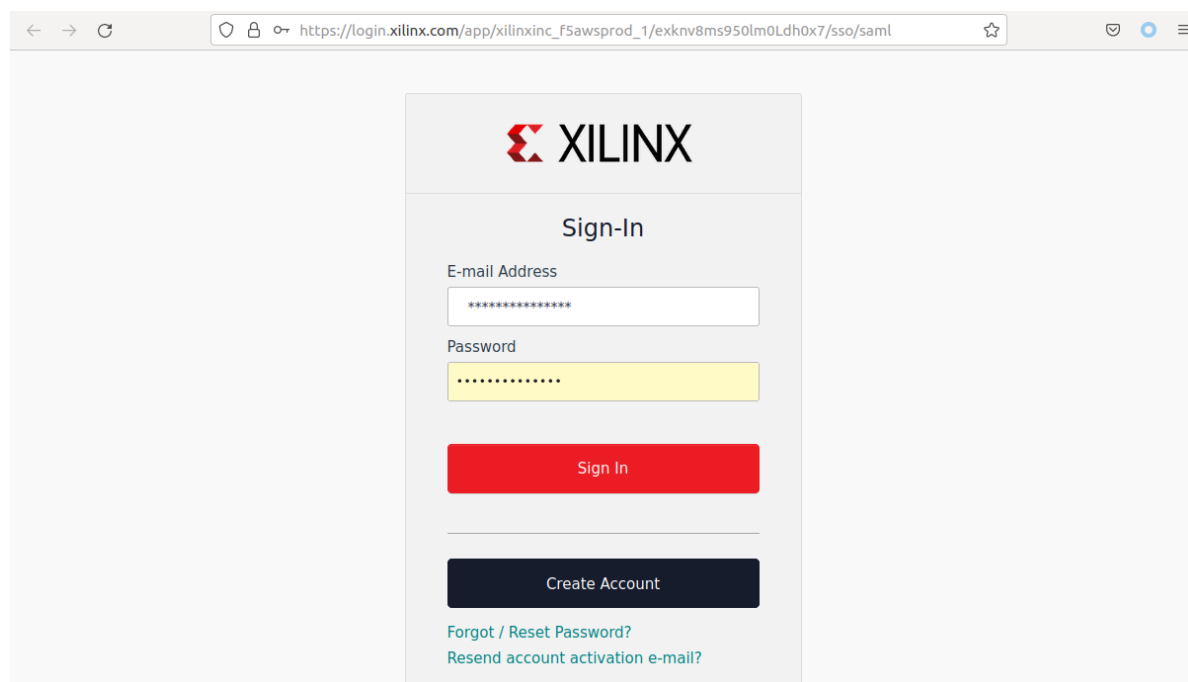
Click the **obtain license** part on the left side, choose the second option **Get Vivado or IP Evaluation Licenses**, you will have 30 days' free use of Vivado Design Suite. And confirm with clicking the **Connect Now**.



Figure 1.20



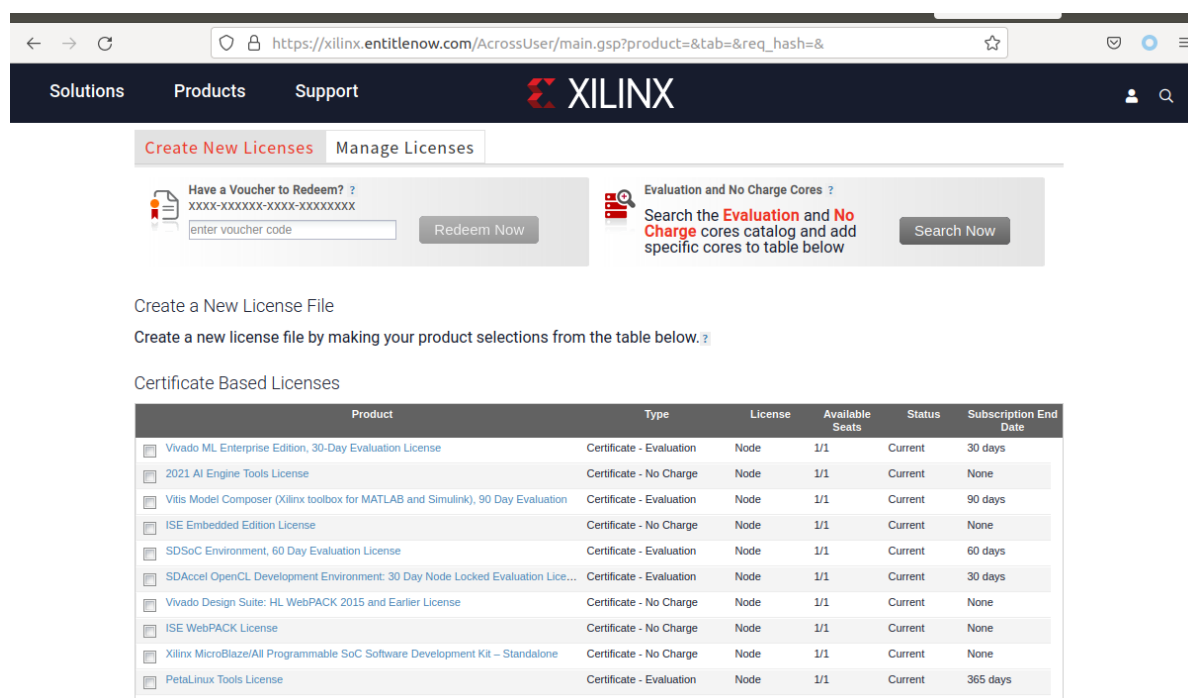
The program will ask you to sign in with your pre-registered user information, including e-mail address and password.



The image shows the Xilinx Sign-In page in a web browser. The URL is [https://login.xilinx.com/app/xilinxinc\\_f5awsprod\\_1/exknv8ms950lm0Ldh0x7/sso/saml](https://login.xilinx.com/app/xilinxinc_f5awsprod_1/exknv8ms950lm0Ldh0x7/sso/saml). The page features the Xilinx logo at the top, followed by the heading "Sign-In". Below this, there are two input fields: "E-mail Address" and "Password". The "E-mail Address" field contains a series of asterisks. The "Password" field is highlighted in yellow and also contains a series of asterisks. Below the input fields is a red "Sign In" button. Underneath the button is a dark blue "Create Account" button. At the bottom of the form, there are two links: "Forgot / Reset Password?" and "Resend account activation e-mail?".

Figure 1.21

After successful login or signup, you will see a list of certificate based licenses, shown in Figure . Find the Vivado HLS Evaluation Evaluation, select it and confirm the generation. The license with file name **Xilinx.lic** will be downloaded to you local computer.



The image shows the Xilinx Manage Licenses page. The URL is [https://xilinx.entitlenow.com/AcrossUser/main.gsp?product=&tab=&req\\_hash=&](https://xilinx.entitlenow.com/AcrossUser/main.gsp?product=&tab=&req_hash=&). The page has a dark blue header with the Xilinx logo and navigation links for Solutions, Products, and Support. Below the header, there are two tabs: "Create New Licenses" and "Manage Licenses". The "Manage Licenses" tab is active. Below the tabs, there are two sections. The first section is titled "Have a Voucher to Redeem? ?" and contains a text input field for "enter voucher code" and a "Redeem Now" button. The second section is titled "Evaluation and No Charge Cores ?" and contains a "Search the Evaluation and No Charge cores catalog and add specific cores to table below" text and a "Search Now" button. Below these sections, there is a table titled "Certificate Based Licenses".











Product	Type	License	Available Seats	Status	Subscription End Date
 Vivado ML Enterprise Edition, 30-Day Evaluation License	Certificate - Evaluation	Node	1/1	Current	30 days
 2021 AI Engine Tools License	Certificate - No Charge	Node	1/1	Current	None
 Vitis Model Composer (Xilinx toolbox for MATLAB and Simulink), 90 Day Evaluation	Certificate - Evaluation	Node	1/1	Current	90 days
 ISE Embedded Edition License	Certificate - No Charge	Node	1/1	Current	None
 SDSoc Environment, 60 Day Evaluation License	Certificate - Evaluation	Node	1/1	Current	60 days
 SDAccel OpenCL Development Environment: 30 Day Node Locked Evaluation License	Certificate - Evaluation	Node	1/1	Current	30 days
 Vivado Design Suite: HL WebPACK 2015 and Earlier License	Certificate - No Charge	Node	1/1	Current	None
 ISE WebPACK License	Certificate - No Charge	Node	1/1	Current	None
 Xilinx MicroBlaze/All Programmable SoC Software Development Kit – Standalone	Certificate - No Charge	Node	1/1	Current	None
 PetaLinux Tools License	Certificate - Evaluation	Node	1/1	Current	365 days

Figure 1.22

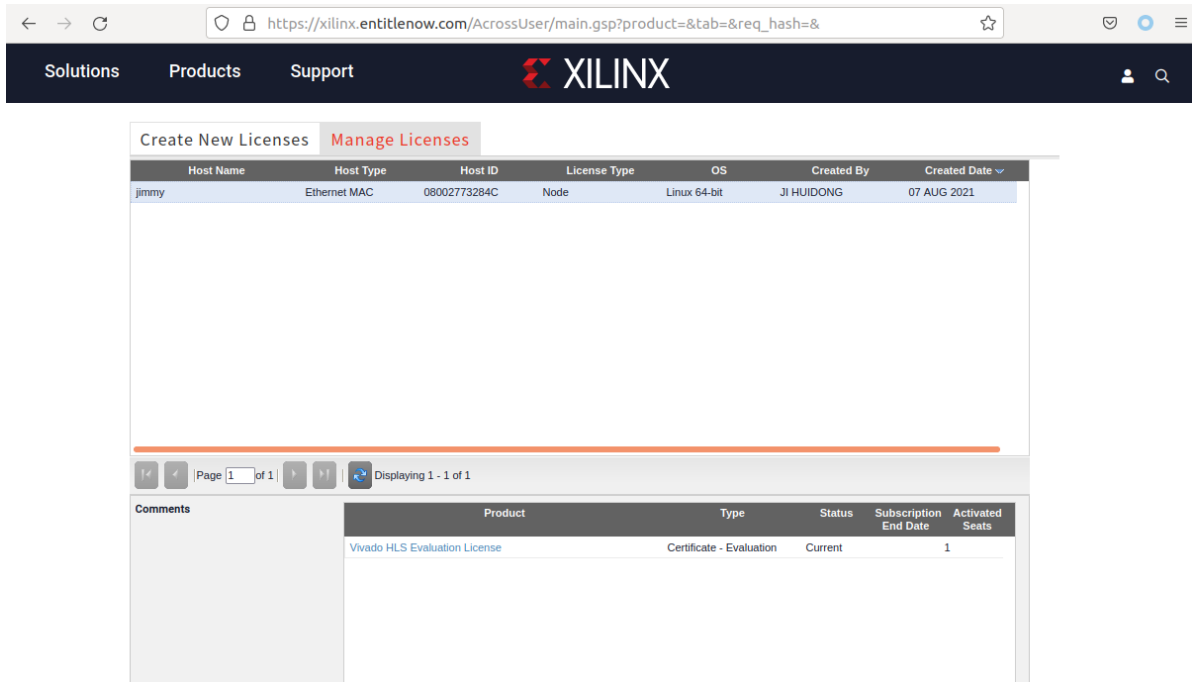


Figure 1.23

Shown in Figure , **Load license** with the selected license file in your local machine. The License manager will verify the file to check the credential. If the verification passed, the license status window will show the license's detail.

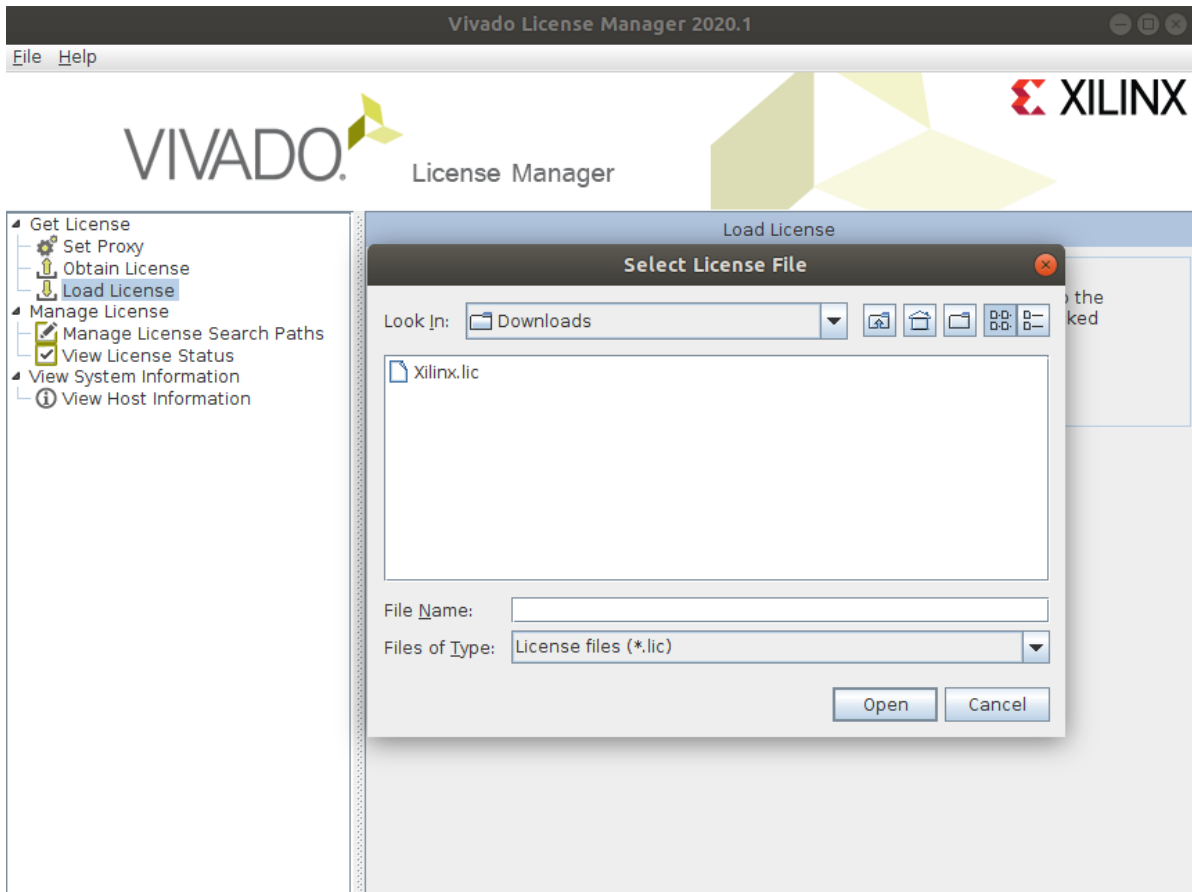


Figure 1.24

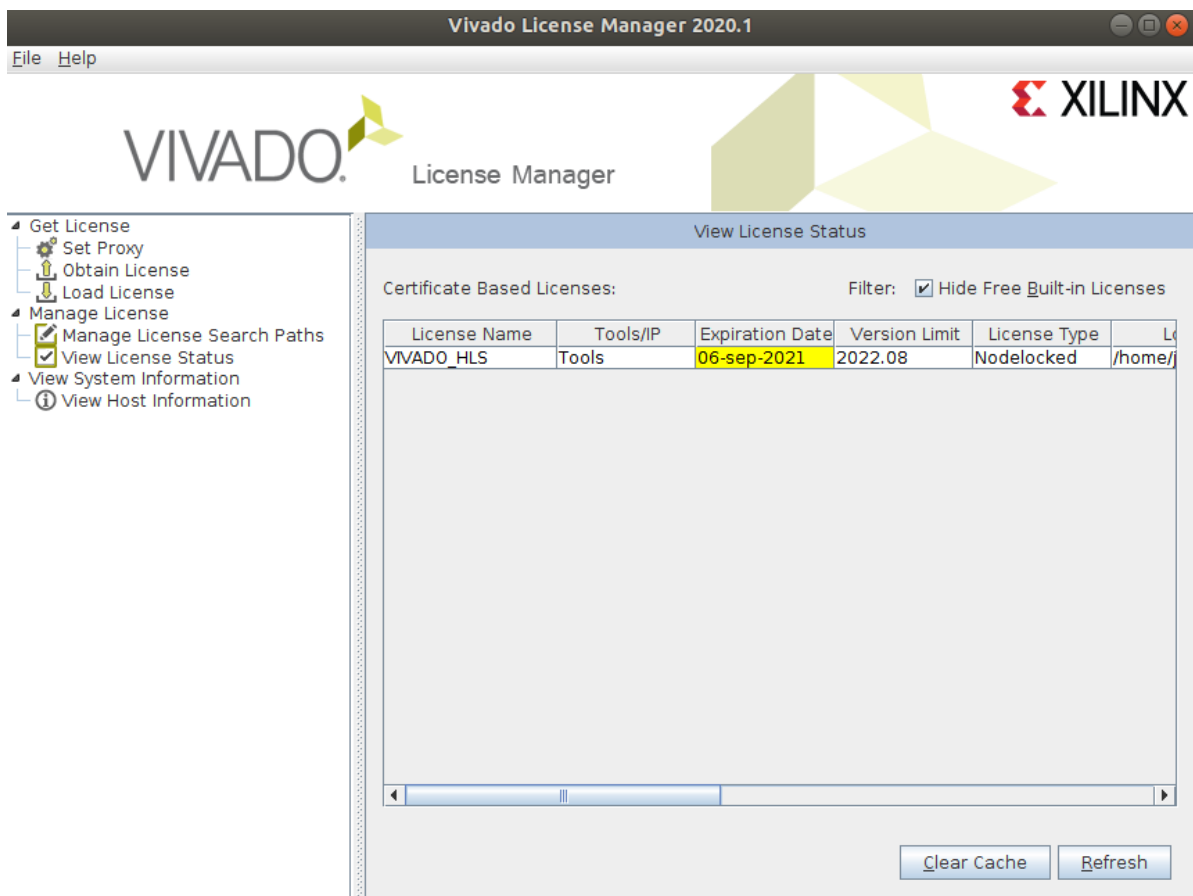


Figure 1.25

Any problems concerning the installation, please feel free to contact me.

In the next lab session, we will write the first project on Vivado Design Suite, therefore, before the start of the lab, please learn yourself the basics of VHDL and Verilog.

Maybe there are students who are not familiar with Verilog and VHDL, two popular hardware description languages. Don't worry, Here are some websites helping you get start with these.

# Lab 2 Verilog & VHDL:

## 2.1 Goal

This part of the documentation gives a step by step instruction about how to create a project and verify the result. It is expected the users can learn how to create a simple digital circuit using Verilog HDL and finish the synthesis and generate the bitstream for the designated hardware.

## 2.1 Design Flow

Shown in Figure 2.1, a typical design flow consists of model(s) creating, user constraint creating file(s), Vivado project creating, the created models importing, assigning created constraint file(s), optionally running behavioral simulation, synthesizing the design, implementing the design, generating the bitstream, and finally verifying the functionality in the hardware by downloading the generated bitstream file.

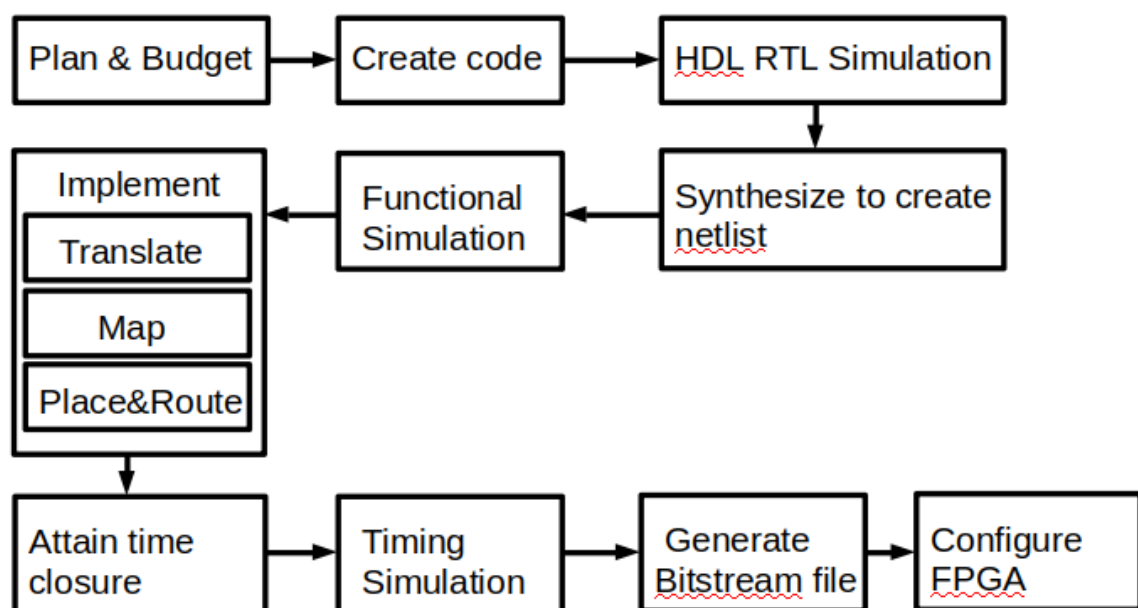


Figure 2.1

## Project creating

Figure to Figure is the new project, lab2's creating process. You can locate the project on the **vivado\_lab** folder, which is provided and can be downloaded from the github.

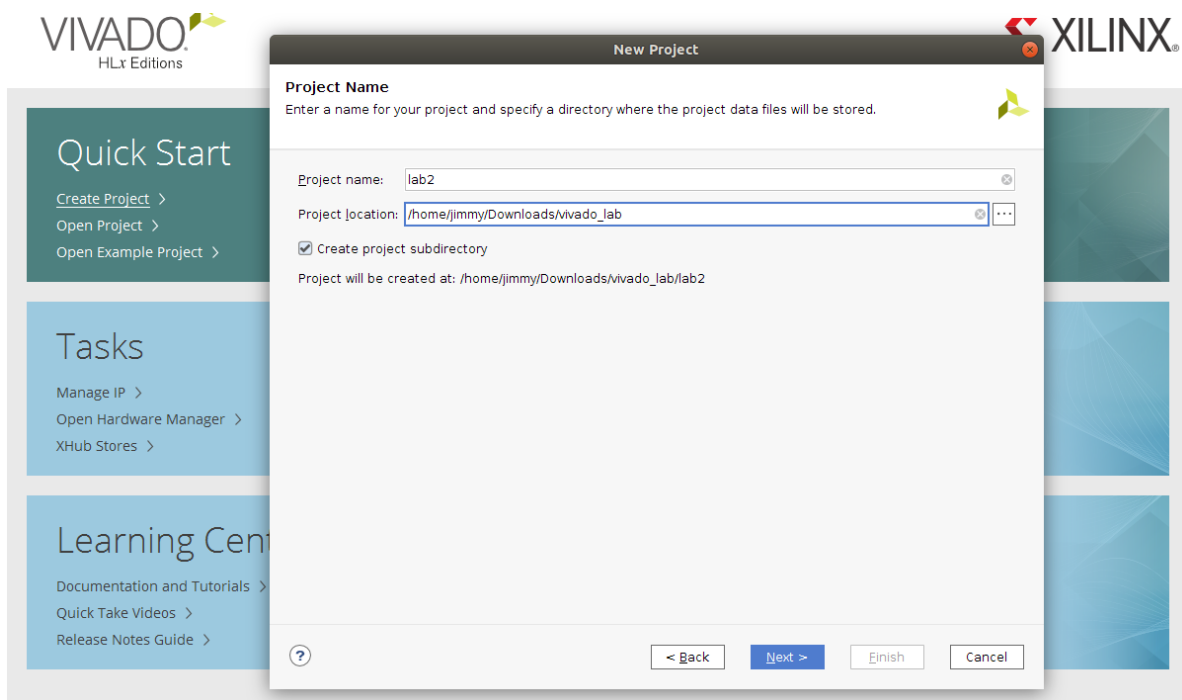


Figure 2.2

Choose RTL project as the project type.

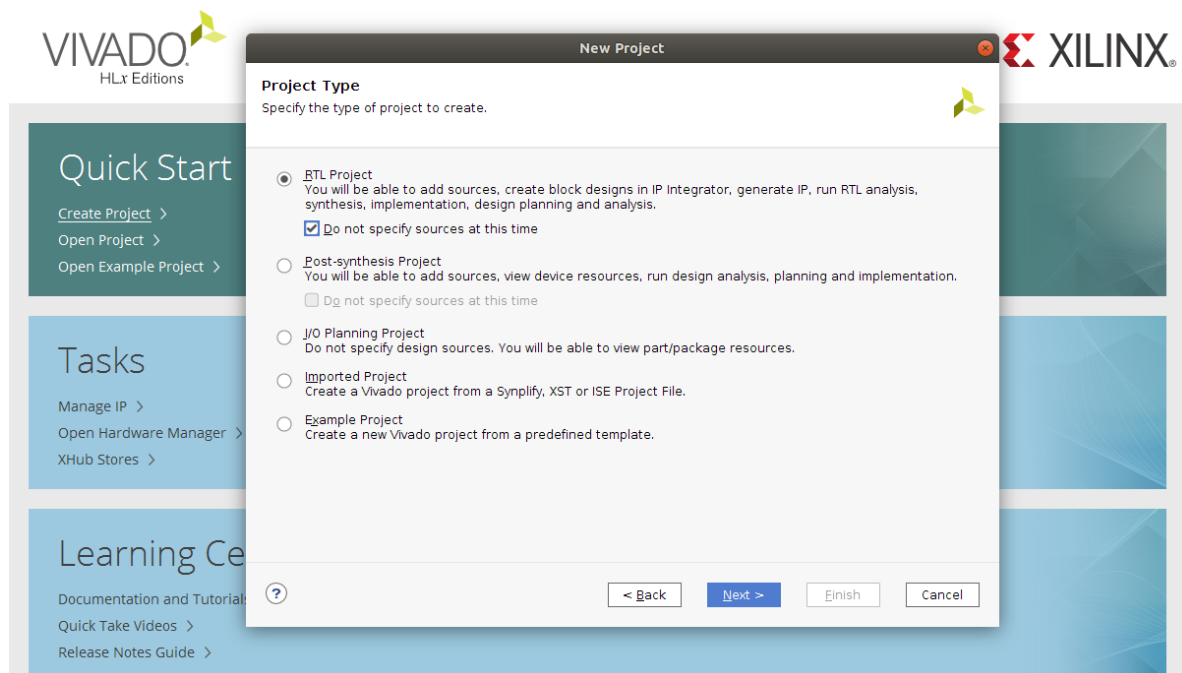


Figure 2.3

After the type selection, decide the board part.

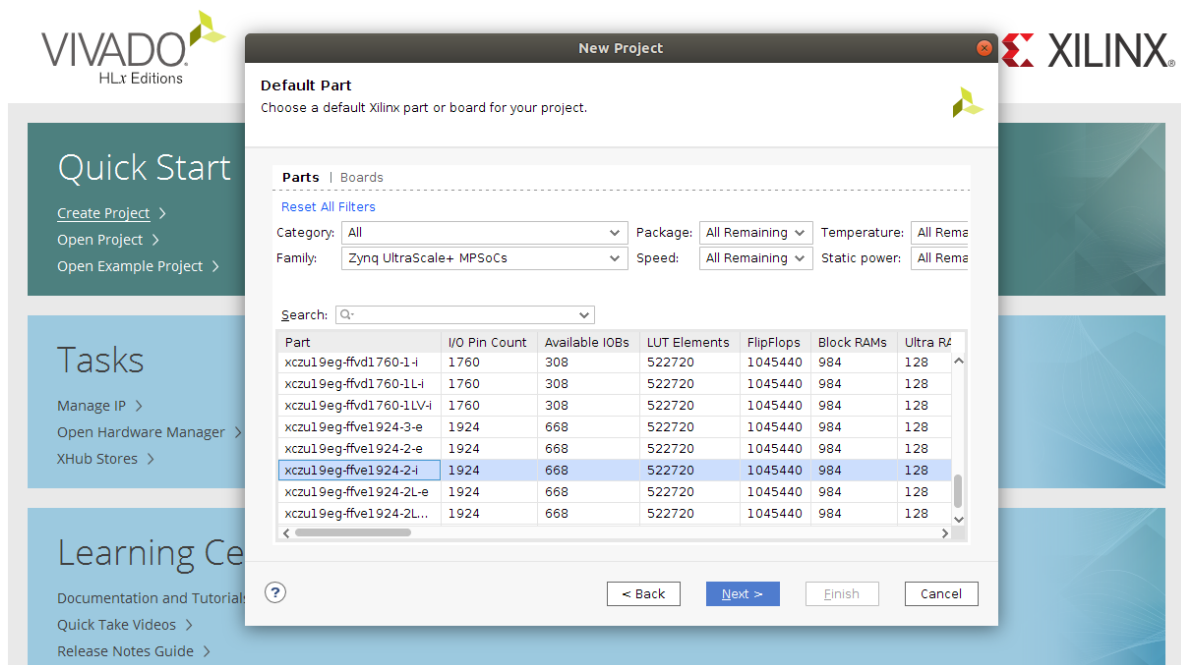


Figure 2.4

We can see the layout of the project platform and the sources hierarchy.

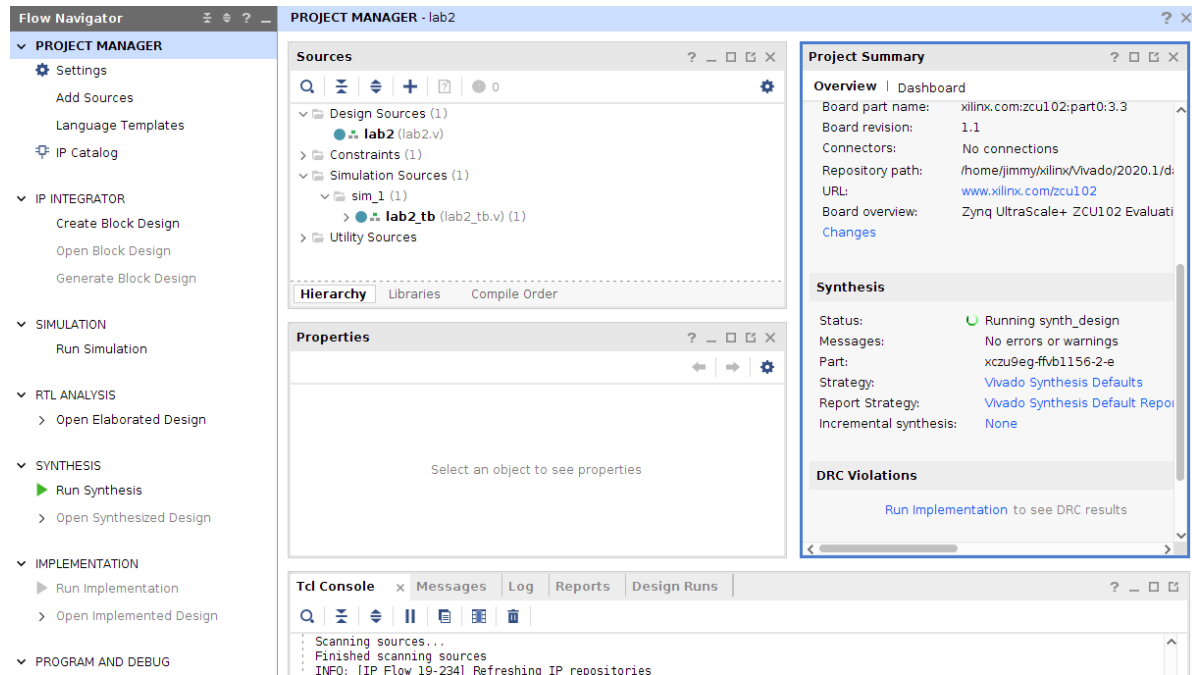


Figure 2.5



Figure 2.6

## Source Adding

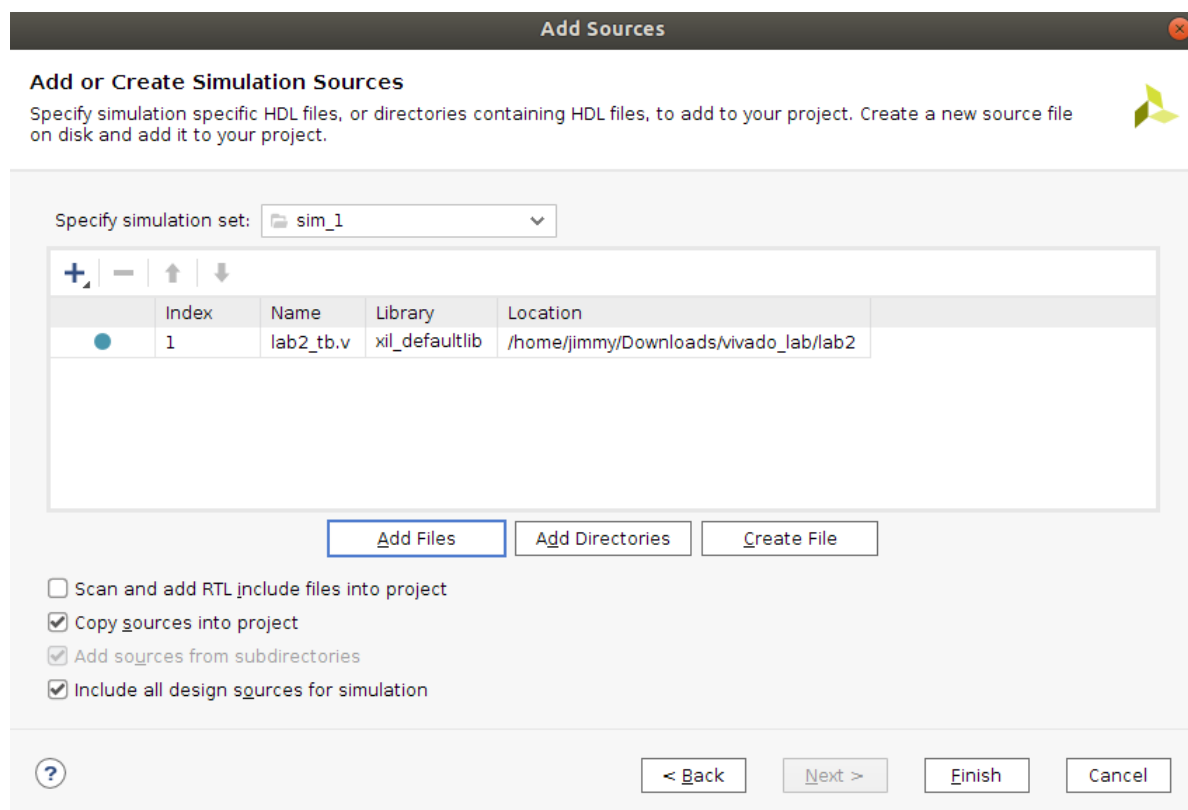


Figure 2.7



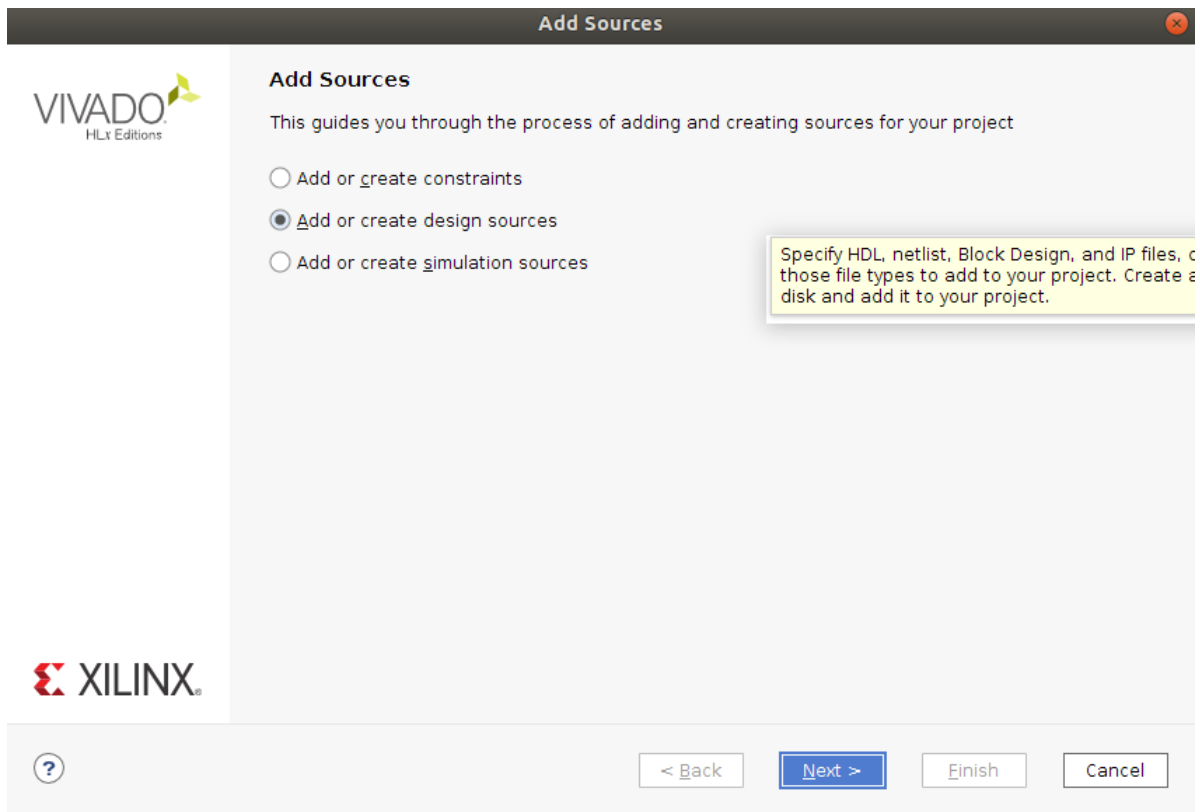


Figure 2.8

In the **source** window, we can find the verilog code for lab2 project. double click it and check the code, which is also pasted below.

```
`timescale 1ns / 1ps
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
/
// Module Name: lab2
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
/

module lab2(
    input [7:0] swt,
    output [7:0] led
);

    assign led[0] = ~swt[0];
    assign led[1] = swt[1] & ~swt[2];
    assign led[3] = swt[2] & swt[3];
    assign led[2] = led[1] | led[3];

    assign led[7:4] = swt[7:4];

endmodule
```

1. The first line ``timescale 1ns / 1ps` defines the timescale directive specifies the time unit and precision for the modules for the later functional and timing simulation. Please note: it has no effect on the **synthesis** and also **hardware implementation**.

```

module lab2()
    /*main content*/
endmodule

```

2. The `module` and `endmodule` defines the beginning and end of a module.

```

input [7:0] swt,
output [7:0] led

```

3. The input and output define the I/O port of the module, `[7:0]` is the length of data array passing through the port, for example `0x00101011`.
4. The follow part defines the logical relationship between the input **swt** and the output **led**. The first bit of the output, **led[0]**, is linked with the value of the first bit of the input, **swt[0]** with **assign**. `~` in the verilog code is the bitwise operator for negation.
5. `&` and `|` are also the bitwise operator working as AND and OR logics.

## RTL Analysis

After the code detail explanation, we can get a functional circuit with the output array of **led** control by the input **swt**. To get out of the abstract description, we can use RTL analysis to get the schematic drawing for the logic circuit, shown in Figure 2.9.

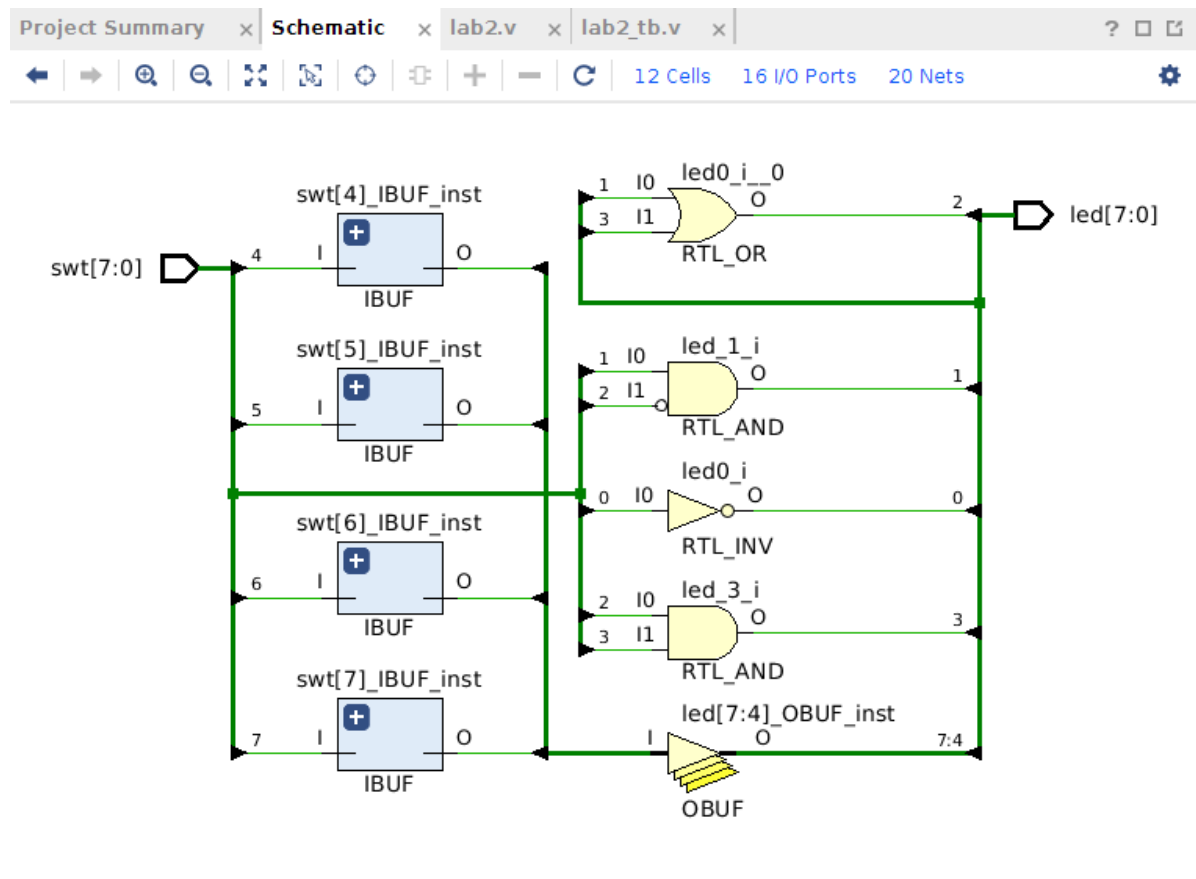


Figure 2.9

## 2.3 Finite State Machine

Generally circuits can be divided into two modes, combinational and sequential mode. In combinational circuits, the output depends on the current values of inputs only. Whereas in sequential circuits, the output depends on the current values of the inputs along with the previously stored states.

Shown in Figure 2.10, the FSM consists of:

1. Initial state (**STATE\_Initial**).
2. States with conditional outward transitions(**STATE\_1/STATE\_2/STATE\_3**).
3. States with non-conditional outward transitions(**STATE\_Initial**).
4. States looping back into themselves(**STATE\_1/STATE\_2/STATE\_3**).
5. States with no outward transitions(**STATE\_4**).

All the individual state in some cases share more than one characteristics.

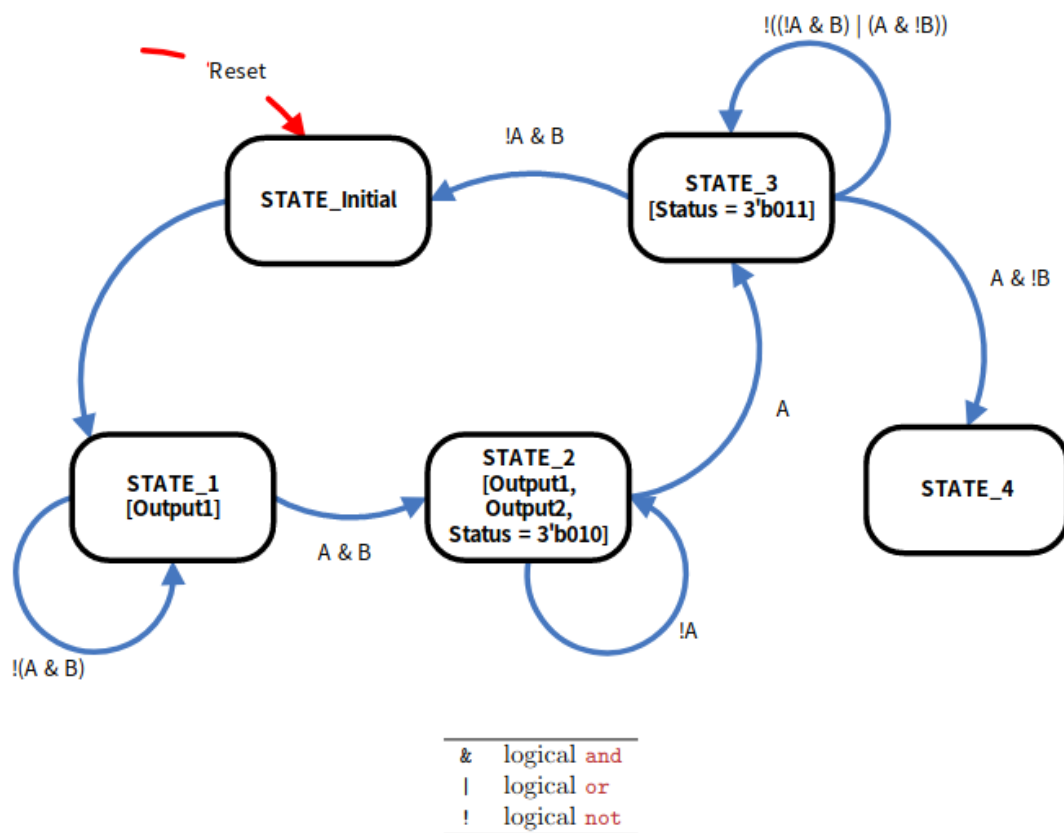


Figure 2.10

To construct the logics in verilog, we need to:

1. A state representation for each state.
2. A mechanism for keeping track of the current state.
3. Transitions or jumps from state to state.
4. Output values based on the current state.

Below is the template code for your reference.

```

module BasicFsm(
//-----Inputs & Outputs declaration-----
input wire Clock ,
input wire Reset ,
input wire A,

```

```

input wire B,
output wire Output1 ,
output wire Output2 ,
output reg [2:0] Status
// -----
);
//----- State Encoding-----
localparam STATE_Initial = 3'd0 ,
           STATE_1 = 3'd1 ,
           STATE_2 = 3'd2 ,
           STATE_3 = 3'd3 ,
           STATE_4 = 3'd4 ,
           STATE_5_PlaceHolder = 3'd5 ,
           STATE_6_PlaceHolder = 3'd6 ,
           STATE_7_PlaceHolder = 3'd7;

// -----
// -----State reg Declarations-----
reg [2:0] CurrentState;
reg [2:0] NextState;
// -----
// -----Outputs-----
// -----1-bit outputs-----
assign Output1 = (CurrentState == STATE_1) | (CurrentState == STATE_2);
assign Output2 = (CurrentState == STATE_2);
// -----multi -bit outputs-----
always@( * ) begin
    Status = 3'b000;
    case (CurrentState)
        STATE_2: begin
            Status = 3'b010;
        end
        STATE_3: begin
            Status = 3'b011;
        end
    endcase
end

// -----
// Synchronous State -Transition always@(posedge Clock) block
always@(posedge Clock) begin
    if (Reset) CurrentState <= STATE_Initial;
    else CurrentState <= NextState;
end

// -----
// Conditional State -Transition always@( * ) block
always@( * ) begin
    NextState = CurrentState;
    case (CurrentState)
        STATE_Initial: begin
            NextState = STATE_1;
        end
        STATE_1: begin
            if (A & B) NextState = STATE_2;
        end
        STATE_2: begin
            if (A) NextState = STATE_3;
        end
        STATE_3: begin
            if (!A & B) NextState = STATE_Initial;

```

```

        else if (A & !B) NextState = STATE_4;
    end
    STATE_4: begin
    end
    STATE_5_PlaceHolder: begin
        NextState = STATE_Initial;
    end
    STATE_6_PlaceHolder: begin
        NextState = STATE_Initial;
    end
    STATE_7_PlaceHolder: begin
        NextState = STATE_Initial;
    end
endcase
end
// -----
endmodule

```

In the code, the declaration part describes the output variables and input variables, which corresponds to the flow diagram of the FSM.

To output the value of register **Status**, the program implements **case** statement to accomplish the conditional selection. Also the same method of **case** statement applies to the state register value of **NextState**.

In Synchronous State part, the value of **CurrentState** is determined only by the input **reset** value.

## 2.4 Testbench

To validate the performance of the module, a testbench script file is needed.

Testbench is a program or model written in any language for the purposes of exercising and verifying the functional correctness of a hardware model during the simulation.

Add the testbench verilog file into the source as the create simulation sources.

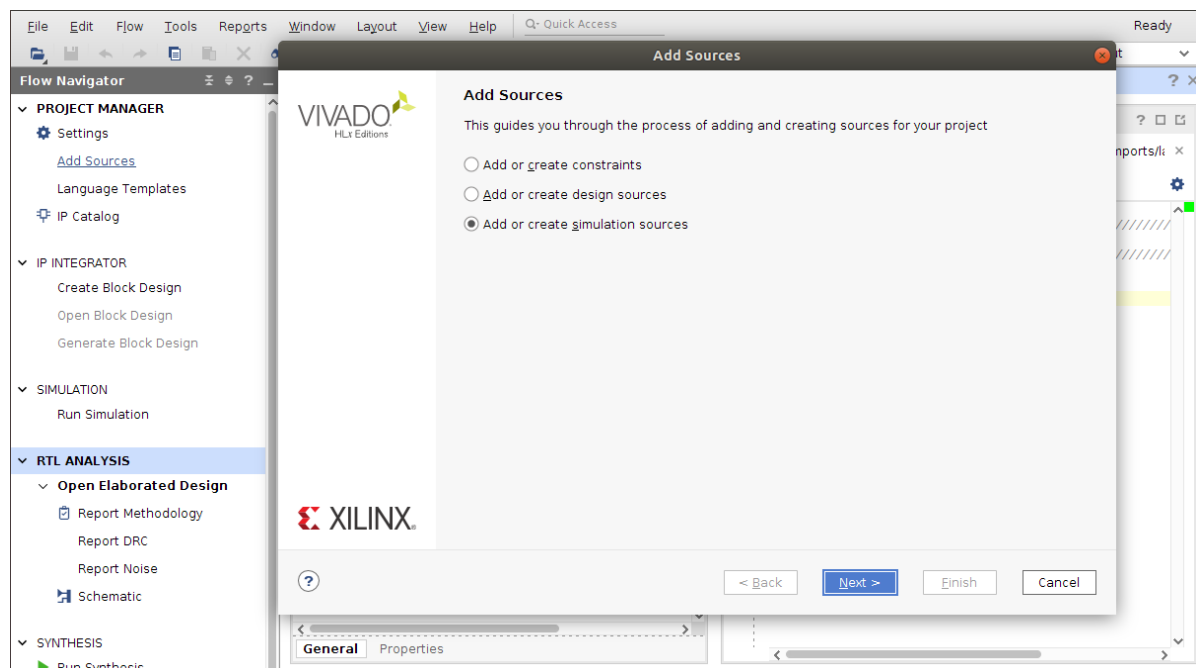


Figure 2.11

**Add Sources**

**Add or Create Simulation Sources**

Specify simulation specific HDL files, or directories containing HDL files, to add to your project. Create a new source file on disk and add it to your project.

Specify simulation set: sim\_1

	Index	Name	Library	Location
●	1	lab2_tb.v	xil_defaultlib	/home/jimmy/Downloads/vivado_lab/lab2

☐ Scan and add RTL include files into project  
☒ Copy sources into project  
☒ Add sources from subdirectories  
☒ Include all design sources for simulation

Figure 2.12

Below is the code block for your reference.

```

`timescale 1ns / 1ps
//////////////////////////////////////////////////////////////////
// Module Name: lab1_tb
//////////////////////////////////////////////////////////////////
module lab2_tb(

);

    reg [7:0] switches;
    wire [7:0] leds;
    reg [7:0] e_led;

    integer i;

    lab2 dut(.led(leds),.swt(switches));

    function [7:0] expected_led;
        input [7:0] swt;
    begin
        expected_led[0] = ~swt[0];
        expected_led[1] = swt[1] & ~swt[2];
        expected_led[3] = swt[2] & swt[3];
        expected_led[2] = expected_led[1] | expected_led[3];
        expected_led[7:4] = swt[7:4];
    end
    endfunction

    initial

```

```

begin
    for (i=0; i < 255; i=i+2)
    begin
        #50 switches=i;
        #10 e_led = expected_led(switches);
        if(leds == e_led)
            $display("LED output matched at", $time);
        else
            $display("LED output mis-matched at ", $time, ": expected: %b,
actual: %b", e_led, leds);
        end
    end
end

endmodule

```

## Timescale definition and variable declaration

```

`timescale 1ns / 1ps

reg [7:0] switches;
wire [7:0] leds;
reg [7:0] e_led;
integer i;

```

Comparing with the functional model module, the Testbench module usually has no port. And in the first part, the Testbench defines the variables for future references and assignments.

## DUT Instantiating

```

lab2 dut(.led(leds), .swt(switches));

```

Following the variables' definition, the Testbench instantiating a DUT(Design Under Design), which is the duplicates of the functional module waiting to be tested(in our project, the function module is **lab2**). and DUT regulates that the input variable must be **reg** type, and the output the **wire** type, because the **reg** type can be assigned with value, while **wire** type can not.

## Generating clock

```

initial
begin
    for (i=0; i < 255; i=i+2)
    begin
        #50 switches=i;
        #10 e_led = expected_led(switches);
        if(leds == e_led)
            $display("LED output matched at", $time);
        else
            $display("LED output mis-matched at ", $time, ": expected: %b,
actual: %b", e_led, leds);
        end
    end
end

```

This Testbench program uses a for loop to counter the frequency or clocks of running. `#50` and `#10` are the delay controls the for-loop used to generate the signal.



In between the clock generating and DUT instantiating, the program starts another function block to repeat the logic of the lab2 circuit and the link the **swt** with output **expected\_led**.

```
function [7:0] expected_led;
    input [7:0] swt;
    begin
        expected_led[0] = ~swt[0];
        expected_led[1] = swt[1] & ~swt[2];
        expected_led[3] = swt[2] & swt[3];
        expected_led[2] = expected_led[1] | expected_led[3];
        expected_led[7:4] = swt[7:4];
    end
endfunction
```

The value of **expected\_led** will be used to compare with the DUT output to check the correctness of the lab2 function. The result will be published with `$display()` function on the **Tcl console**.

## Simulation

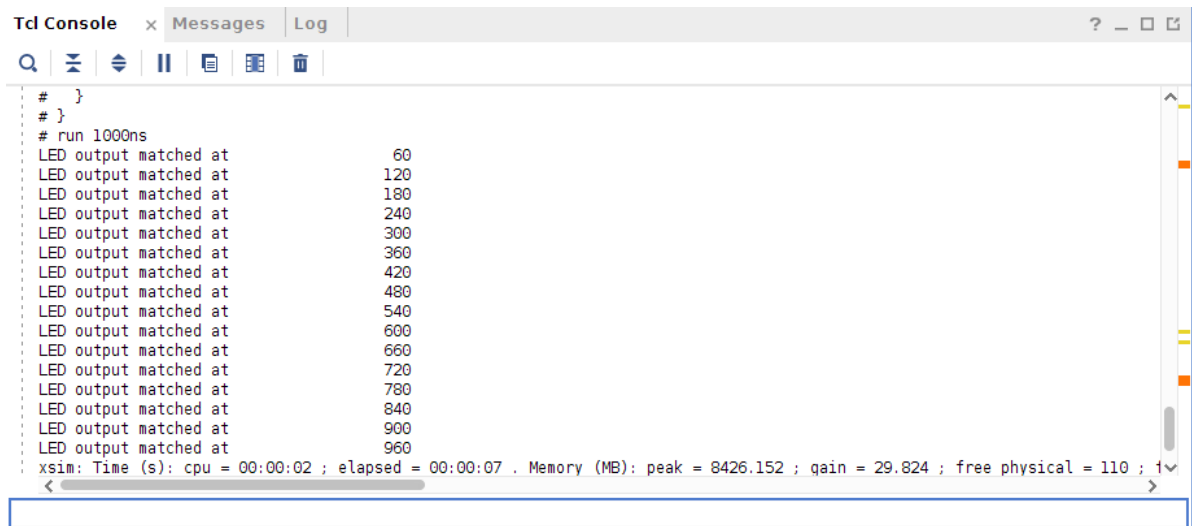


Figure 2.13

To verify the logic function's correctness, the waveform generator from simulation also can provide the result, shown in Figure [2.14](#).



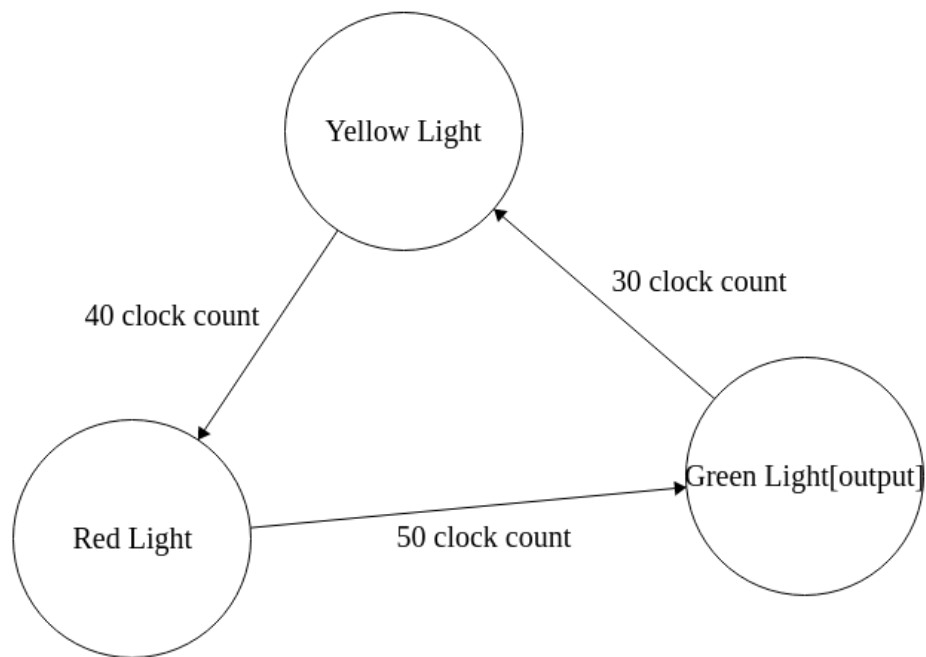


Figure 2.15

The flow diagram shows the sequential logic of the traffic lights, red,yello,green. The transisition from one light to another depends on the time delay, such as the 40 clock count. When the light turns green, the output will turn 1.

In this assignment, you are required to:

1. write a verilog module using FSM to realize the traffic light system, where the Yellow Light is the initial state.
2. write a verilog testbench to check the correctness of the module, generate the waveform with Vivado Design Suite, and store it with the form **wcfg**.

# Lab 3 Convolutional Layer

## 3.1 Goal

Design a module functioning as a layer, which operates convolution on the input data array.

## 3.2 Introduction

### What is convolution?

Convolution is a function that outputs an integral that expresses the amount of **overlap** of one function  $g$  as it is shifted over another function  $f$ . It therefore "blends" one function with another.

#### Linear formula

In mathematics, a convolution is defined as a product of functions  $f$  and  $g$  that are objects in the algebra of [Schwartz functions](#) in  $\mathbb{R}^n$ . Convolution of two functions  $f(t)$  and  $g(t)$  over a finite range  $[0, t]$  is given by

$$\begin{aligned} f * g &= \int_{-\infty}^{+\infty} f(\tau)g(t - \tau)d\tau \\ &= \int_{-\infty}^{+\infty} g(\tau)f(t - \tau)d\tau \end{aligned}$$

#### Formula 1

And the convolution  $f * g$  can also be expressed as  $f \otimes g$ .

#### Discrete formula

In many engineering fields, such as the computer science, the input signals come with the discrete form. Convolution of two discrete functions  $x(n)$  and  $h(n)$  over a finite count  $i$  is given by

$$\begin{aligned} x * h &= \sum_{i=-\infty}^{\infty} x(i)h(n - i) \\ &= \sum_{i=-\infty}^{\infty} h(i)f(n - i) \end{aligned}$$

#### Formula 2

#### 2-Dimensional formula

The 2-D Convolution computes the two-dimensional convolution of two input 2-D arrays, in another word, matrix. Assume that *matrix A* has dimensions  $(M_a, N_a)$  and matrix B has dimensions  $(M_b, N_b)$ . When the convolution calculates the full output size, the equation for the 2-D discrete convolution is:

$$C(i, j) = \sum_{m=0}^{M_a-1} \sum_{n=0}^{N_a-1} A(m, n) * B(i - m, j - n)$$

#### Formula 3

## Convolution in Image Processing

In image processing, convolutional filtering can be used to implement algorithms such as edge detection, image sharpening, and image blurring.

With the convolution process, many important features can be extracted from the edges of an image (e.g., corners, lines, curves). In image process, Features which distinguish them from the surrounding display the color intensity change. The feature can later be used for further deep learning training or traditional computer vision object recognition and tracking.



Figure 3.1

## 3.3 Assignment

### Assignment 1

In this lab, we plan to build with **Verilog** based on Vivado Design Suite to realize a 3 by 3 matrix's convolutional operation.

Imagine **Matrix A** is the pixel value for the upper left corner

$$A = \begin{bmatrix} 3 & 3 & 1 \\ 0 & 1 & 0 \\ 1 & 4 & 2 \end{bmatrix}$$

Matrix B is the image kernel for convolution

$$B = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 0 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

Following the *Formula 3* above,  $A * B$  can be separated into two parts:

1. Element-wise Multiplication

$$M_{11} = A_{11} * B_{11}, M_{12} = A_{12} * B_{12}, M_{13} = A_{13} * B_{13};$$

$$M_{21} = A_{21} * B_{21}, M_{22} = A_{22} * B_{22}, M_{23} = A_{23} * B_{23};$$

$$M_{31} = A_{31} * B_{31}, M_{32} = A_{32} * B_{32}, M_{33} = A_{33} * B_{33};$$

2. Summation of the multiplication

$$S = M_{11} + M_{12} + M_{13} + M_{21} + M_{22} + M_{23} + M_{31} + M_{32} + M_{33}$$

1. Please calculate the result
2. Please simulate the whole convolution process using the mathematical operators in hardware mode, such as the **multipliers** and **adders**. Write the verilog-based module and testbench, store the waveform.

Hint

1. Vivado Design Suite has already provided functional IPs for use, such as the **RAM** or **ROM** for memory reading and writing, **Adder** for mathematical adding and **Multiplier** for multiplication. Of course you can also build your own IP module with verilog and Vivado, this is encouraged.
2. As is known in the lectures, the convolution layer in deep learning or machine learning in effect is the result of continuous convolution between matrixes. To accelerate the computing, please take parallel computing in consideration, and take use of the advantage of Xilinx FPGA's reconfigurable hardware to realize it.
3. FSM(Finite State Machine) or Behavioral Tree can be used to create case-based function flow, which helps the project management.

## Assignment 2

As shown in [Ass](#), the matrix's convolution will reduce the size and information, which meets the requirement of subsampling.

For the input matrix's size outnumber the kernel's, the output will be a matrix also:

0	0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0	
0	55	121	75	78	95	88	0	
0	35	24	204	113	109	221	0	
0	3	154	104	235	25	130	0	
0	15	253	225	159	78	233	0	
0	68	85	180	214	245	0	0	
0	0	0	0	0	0	0	0	0

WEIGHT			
1	0	1	139
0	1	0	
1	0	1	

Step 1

0	0	0	0	0	0	0	0	
0	18	54	51	239	244	188	0	
0	55	121	75	78	95	88	0	
0	35	24	204	113	109	221	0	
0	3	154	104	235	25	130	0	
0	15	253	225	159	78	233	0	
0	68	85	180	214	245	0	0	
0	0	0	0	0	0	0	0	0

WEIGHT			
1	0	1	139 184
0	1	0	
1	0	1	

## Step 2

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250
-----	-----	-----

## Step 3

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409
-----	-----	-----	-----

## Step 4

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409	410
-----	-----	-----	-----	-----

## Step 5

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409	410	283
-----	-----	-----	-----	-----	-----

## Step 6



0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409	410	283
133					

Step 7

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409	410	283
133	429				

Step 8

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

WEIGHT

1	0	1
0	1	0
1	0	1

139	184	250	409	410	283
133	429	505	686	856	441
310	261	792	412	640	341
280	633	653	851	751	317
254	608	913	713	657	503
321	325	592	517	637	78

Step 9

In [Step 1](#), the kernel mapping the first 9 element in the top left corner and calculates the output integer. Following it, from [Step 2](#) to [Step 5](#) , the kernel moves rightward to finish the mapping until the last effective index. After that, the kernel jump to the starting index of the **second** row of the input, and operates the same as in the first row.

In the end, the final convolution will be like in [Step 9](#).

Here is the definition for the images and kernel:

O: The size of output image,

I: The size of input image,

K: The size of the kernel,

S: The stride value of the kernel's single step moving. In the example above, the stride is 1, therefore, the kernel moves one index every step.

P: The padding number,

The output size is calculated with formula:

$$O = \frac{I - K + 2P}{S} + 1$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 0 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

In this assignment, the input Matrix A will be with size 5 by 5, and the kernel matrix B will remain the same with the size 3 by 3. Please calculate the convoluted output with **0 padding** and **stride = 1**.

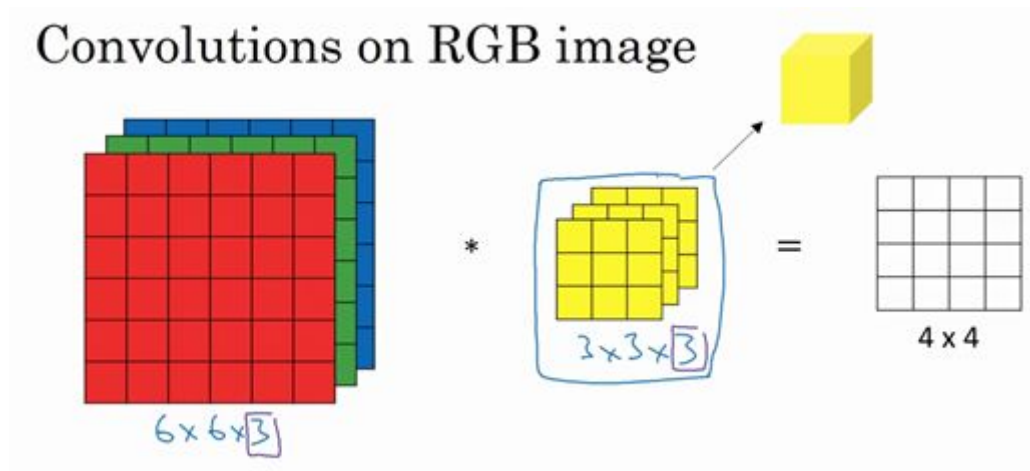
Again with verilog and Vivado Design Suite, write the code for the convolution function, and store the waveform of the final matrix result.

### Assignment 3

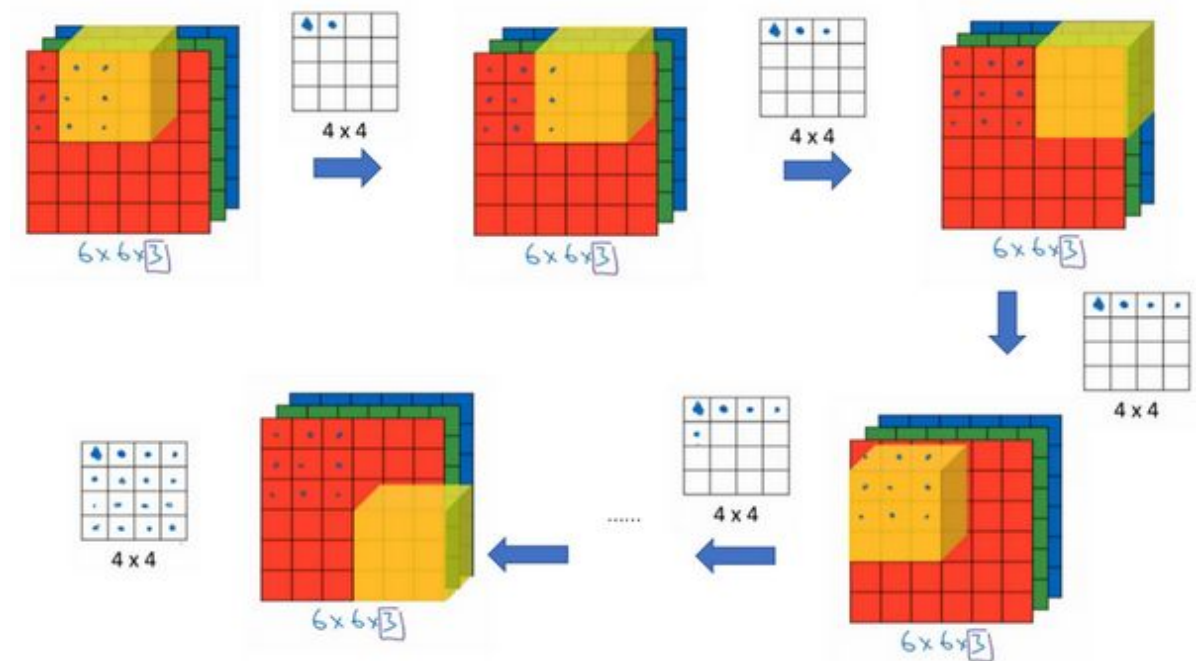
In deep learning, the input images will be more than 2 dimensions, containing the width, height and the depth or the channels of the images. Another term to introduce the multi-dimensional input image is call **tensor**.

Also for the kernel matrix, to get the features for one channel of the input images, kernel with multiple channels will be designed. To get wanted features, such as horizontal edges or vertical edges, multiple kernels will be implemented on the same input image for convolution.

For example, there is a 3D matrix with the size 3 x 3 x 3. Still the same with the 2D convolution in [Assignment 2](#), the total **27** weight parameters of the kernel is fetched and implemented on the **top left 3 x 3 corner**, and the convolution will output an integer, which is the first element for the final 4 x 4 matrix.

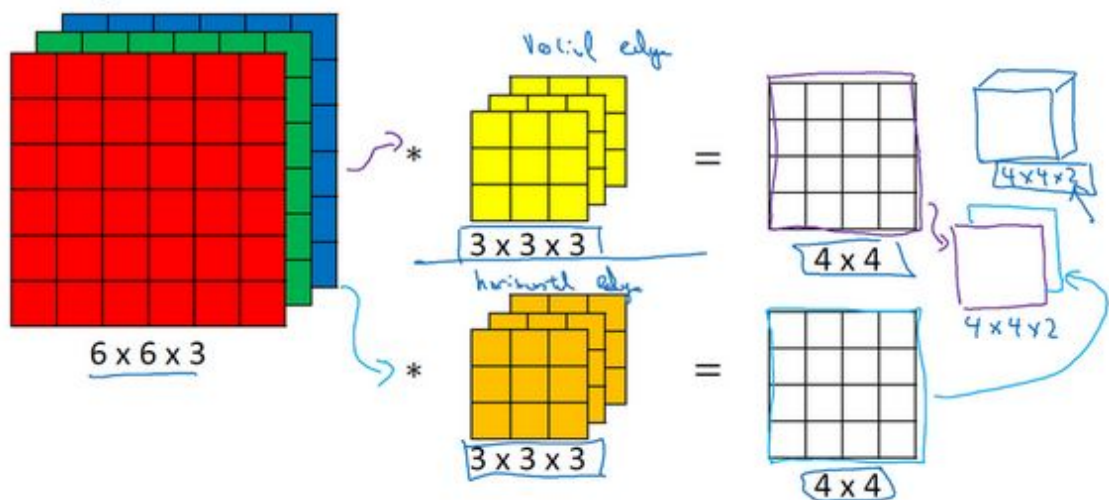


Repeating the above operation, the 3D convolution will finally output the 4 x 4 matrix, whose size can be determined by the **formular**



As mentioned above, multiple filters can be implemented to get different features, such as vertical edges and horizontal edges. If the input image is with the size  $6 \times 6 \times 3$ , and the kernel size is  $3 \times 3 \times 2$ , where the last number mean the number of filters, then the output convolution size is  $4 \times 4 \times 4$ . Therefore, the number of the output matrix equals the number of the kernel.

## Multiple filters



In the task, you are required to perform a convolution on Matrix A with the size:  $6 \times 6 \times 3$  with the kernel with the size:  $3 \times 3 \times 3 \times 2$ . Calculate the output matrix.

Again with verilog and Vivado Design Suite, write the code for the convolution function, and store the waveform of the final matrix result.

$$A_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A_1 = \begin{bmatrix} 1 & 3 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 1 & 4 & 2 & 0 \\ 0 & 3 & 3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{array}{lll}
\textit{Kernel}B_0: & B_{00} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} & B_{01} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} & B_{02} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \\
\textit{Kernel}B_1: & B_{10} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} & B_{11} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} & B_{12} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}
\end{array}$$

# Lab 4 HLS

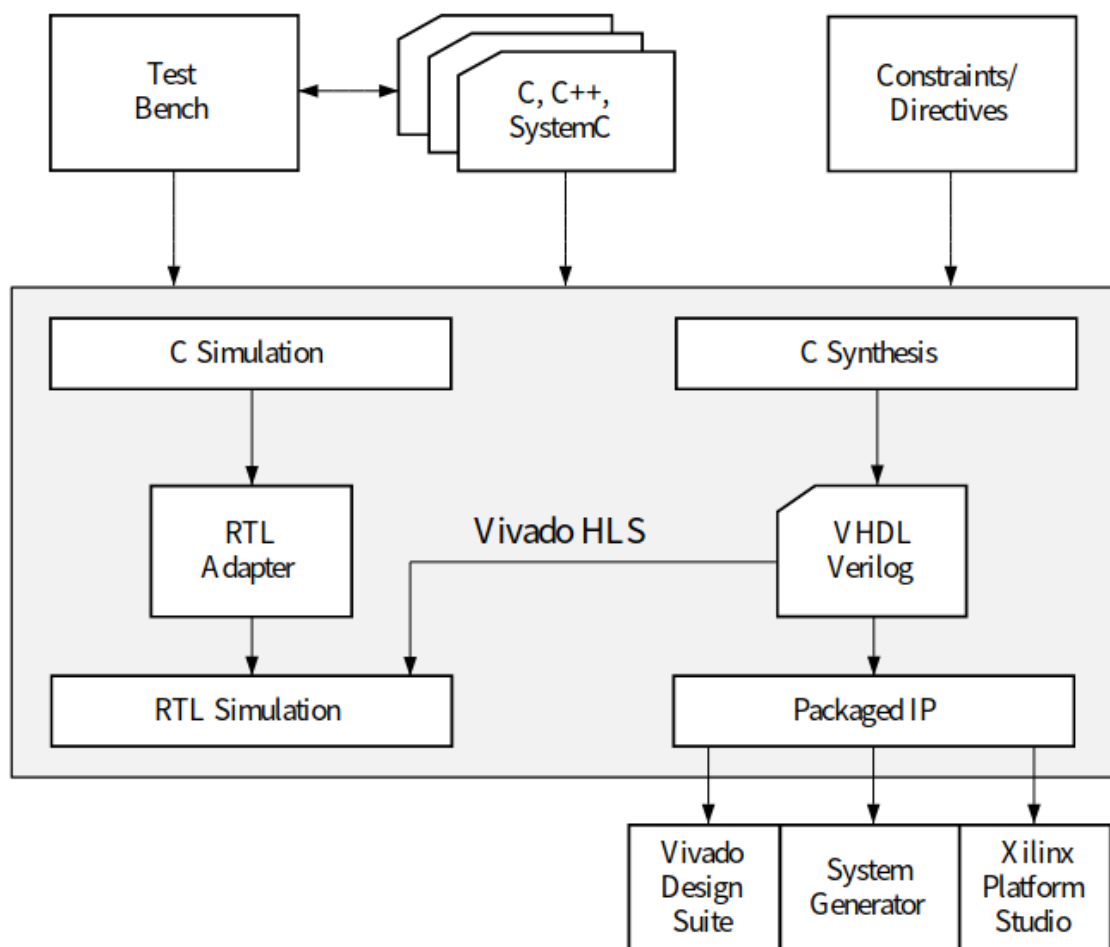
## 4.1 Goal

## 4.2 Introduction

### HLS introduction

HLS, abbreviated for High-Level-Synthesis, is a high level abstraction tool to automatically create RTL(Register Transfer Level) design.

The following figure shows an overview of the Vivado HLS input and output files. The Xilinx Vivado HLS tool synthesizes a C function into an IP block that you can integrate into a hardware system. It is tightly integrated with the rest of the Xilinx design tools and provides comprehensive language support and features for creating the optimal implementation for your C algorithm.



In two parts, HLS shows its advantage:

1. HLS synthesizes the function's content into RTL statement with its algorithm synthesis.
2. HLS helps transform the function's parameters, inputs and outputs into ports(**output**, **input**, **reg**, **wire**, **RAM**).

In all, HLS saves the developers from the details design of the circuit with verilog/VHDL.

## Hardware description-AVNET Ultra96-V2

Ultra96-V2 is an Arm-based, Xilinx Zynq UltraScale+™ MPSoC development board based on the Linaro 96Boards Consumer Edition (CE) specification. the Ultra96-V2 boots from the provided Delkin 16 GB microSD card. Engineers have options of connecting to Ultra96-V2 through a Webserver using integrated wireless access point capability or to use the provided PetaLinux desktop environment which can be viewed on the integrated Mini DisplayPort video output.

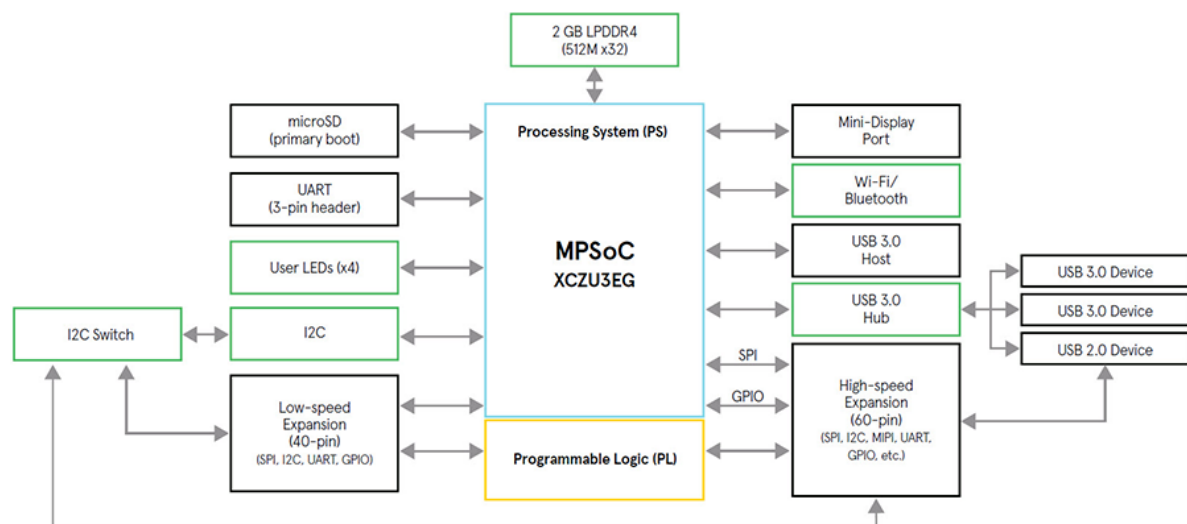
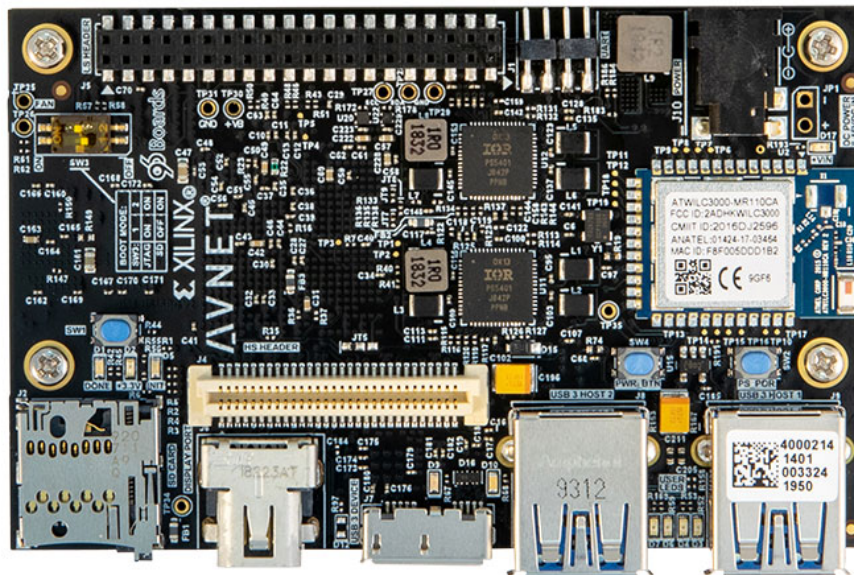
Here the feature table and the application fields are given below. For more details, you can refer to the product's [homepage](#).

Feature Table
Xilinx Zynq UltraScale+ MPSoC
ZU3EG A484 Micron 2 GB (512M x32) LPDDR4
Memory Delkin 16 GB microSD card + adapter
PetaLinux environment available for download
Microchip Wi-Fi / Bluetooth
IDT VersaClock 6E clock generator
Infineon high efficiency power management
Aavid/Boyd Heatsink
Mini DisplayPort (MiniDP or mDP)
1x USB 3.0 Type Micro-B upstream port

Table 4.1

Target Applications
Artificial Intelligence
Machine Learning
IoT/Cloud connectivity for add-on sensors
Embedded Computing
Robotic
Entry level Zynq UltraScale+ MPSoC development environment
Training, prototyping and proof-of-concept demo platform
Wireless design and demonstrations using Wi-Fi and Bluetooth

Table 4.2



## Software development kit

The Deep Neural Network Development Kit (**DNNDK**) package can be freely downloaded after registration from the Xilinx website:

<https://www.xilinx.com/products/design-tools/ai-inference/edge-ai-platform.html>.

Deep Neural Network Development Kit (DNNDK) is a full-stack deep learning SDK for the Deep-learning Processor Unit (DPU). It provides a unified solution for deep neural network inference applications by providing pruning, quantization, compilation, optimization, and run time support.

Key

highlights and features are listed below:

Innovative full-stack solution for deep learning inference application development

- A complete set of optimized tool chains, including compression, compilation and runtime.
- Lightweight C/C++ and Python programming APIs.
- Easy-to-use with gradual learning curve.

As shown in the figure below, DNNDK framework is composed of Deep Compression Tool (DECENT), Deep Neural Network Compiler (DNNC), Deep Neural Network Assembler (DNNAS), Neural Network Runtime (N2Cube), DPU Simulator, Profiler, DExplorer and DDump. Since v3.0, DNNDK supports both Caffe and TensorFlow with a unified framework. The toolchain flow for deploying Caffe and TensorFlow network models are almost the same, which delivers a familiar user experience.

#### 4.3