

# LLM ON FPGA

## DEPLOYMENT ON THE FPGA

# A BRIEF OVERVIEW

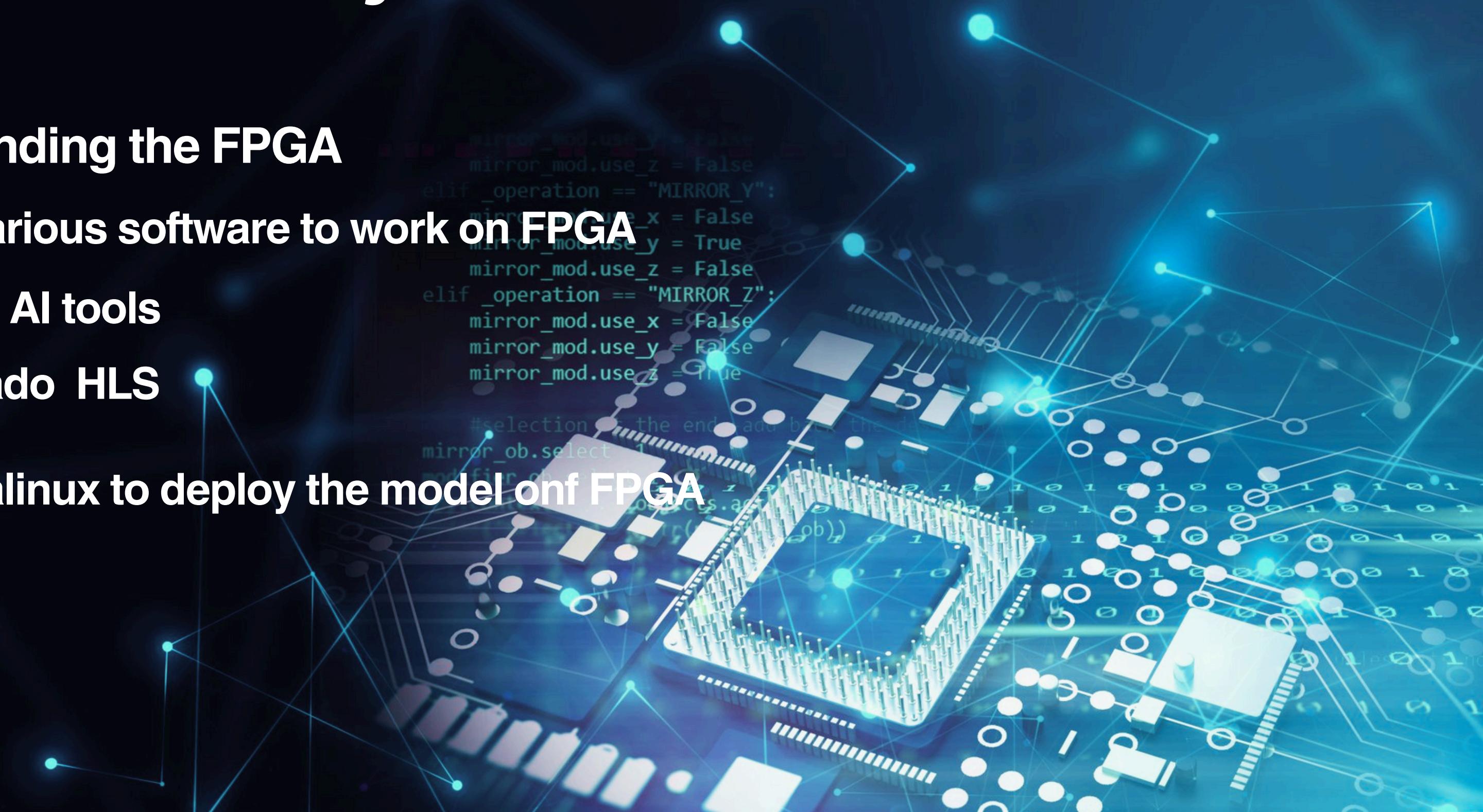
# *In the previous semester we looked upon:*

- performance of different AI models on CPU
  - performance of different AI models on GPU
  - estimated performance of different AI models on FPGA
  - According to found results ,we estimated that the use of FPGA for AI models would have a better impact on its latency,throughput and overall performance.

# CURRENT FLOW

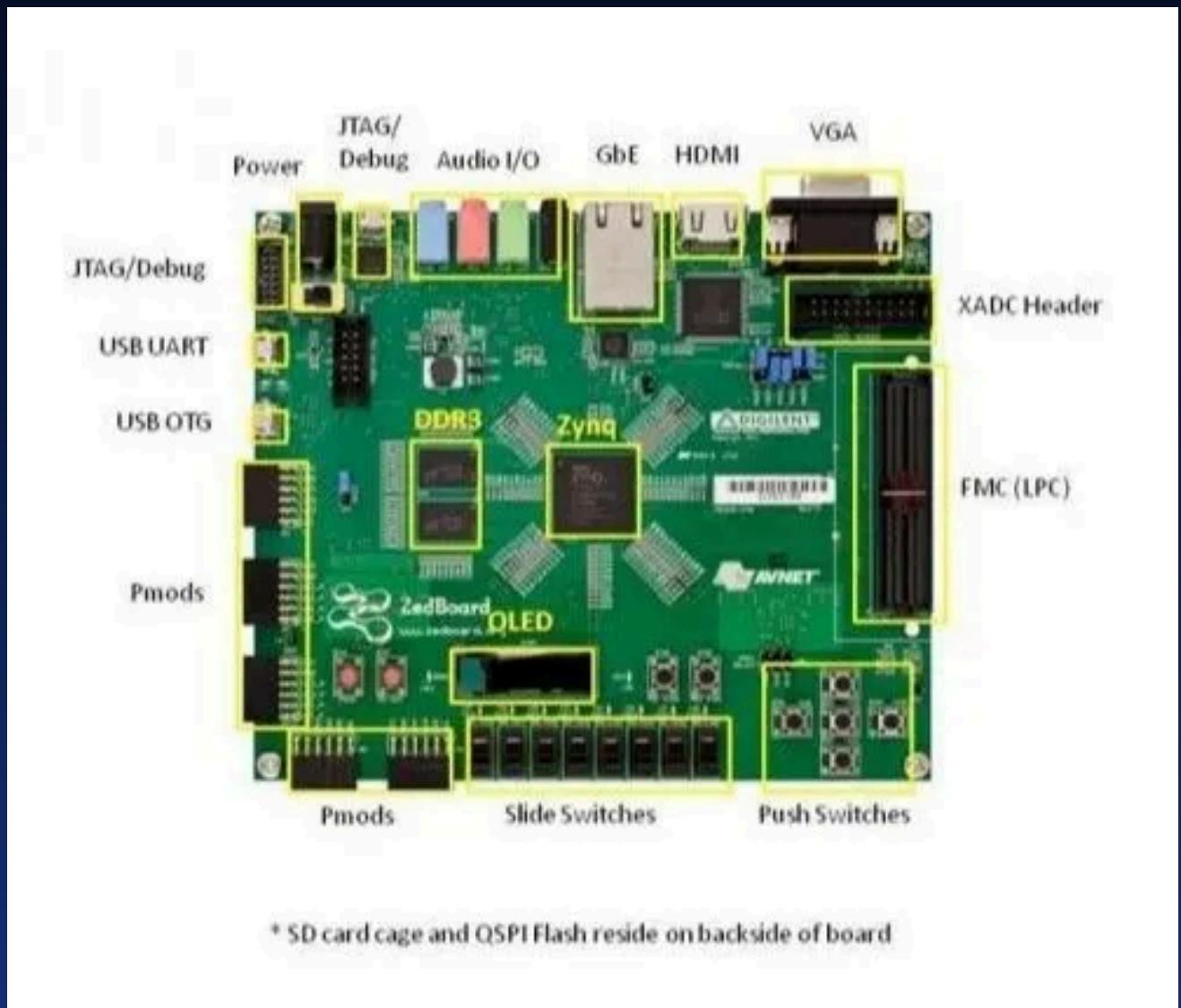
## Deployment and analysis of LLM on FPGA

- Understanding the FPGA
- Using the various software to work on FPGA
  - vitis AI tools
  - Vivado HLS
- Using Petalinux to deploy the model onf FPGA



# ZNYQ 7000 ZEDBOARD

- Dual-Core ARM Cortex-A9 Processor
- Optimized for Real-Time Applications
- 512 MB DDR3 RAM
- 256 MB QSPI Flash (ROM)
- Supports SD cards up to 32 GB or more
- Lower Latency Compared to CPU/GPU Solutions



# ZYNQ ULTRASCALE+ ZU19 BOARD

- Quad-core ARM Cortex-A53: 64-bit processors
- Dual-core ARM Cortex-R5
- Mali-400 MP2 GPU
- Up to 8 GB DDR4 RAM
- UltraRAM and Block RAM
- 32 MB QSPI Flash
- eMMC/SD Card Slot
- High-speed transceivers up to 32.75 Gbps



## VIVADO DESIGN SUITE

A comprehensive toolset designed for FPGA and SoC development, offering an integrated development environment (IDE) that supports a wide range of design activities. It features High-Level Synthesis (HLS) to convert high-level code (C, C++, SystemC) into RTL, and the IP Integrator for easy integration of pre-built IP cores into designs.



## XILINX VITIS AI

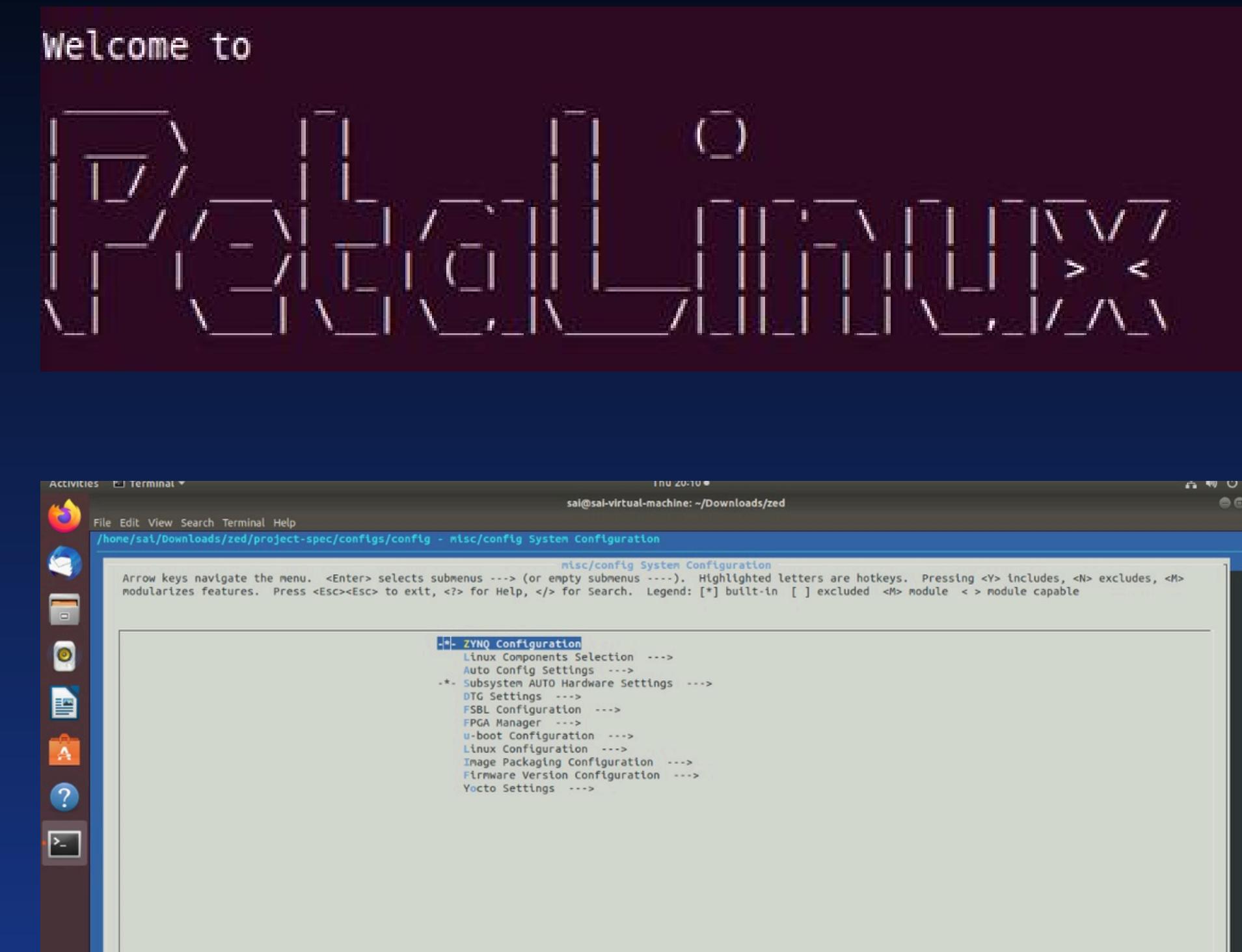
A Unified Software Platform is an integrated development environment designed for creating and deploying applications on Xilinx FPGAs and SoCs. Vitis also integrates with the Xilinx Vitis tool for converting high-level code into hardware description language (HDL), allowing for seamless hardware/software co-design. The platform supports a range of Xilinx devices, including the UltraScale+ series, and facilitates rapid development with its extensive ecosystem of pre-built IP cores and development kits.



# PETALINUX

Slide 07

- Embedded Linux development for Xilinx FPGAs and SoCs
- Tools for configuring, building, and deploying custom Linux distributions
- Support for custom kernel configurations and device tree generation
- Root filesystem creation and customization
- Cross-compilation for software development on target devices
- Integration with Vivado and Vitis for hardware/software co-design
- Compatible with Xilinx FPGA and SoC families, including UltraScale+ series
- Access to development kits and pre-built components



# PLAN OF ACTION

learning the FPGA:

- led toggling using a zed board -  
.xsa file using vivado and petalinux

Deployment of ML model on the ultrascale board(due to memory constraints of Zed board)-

- Generation of .xsa file using Vivado for PL's
- deployment of ONNX model on the ARM processor using petalinux.
- Generating an executable file(.elf file) of ONNX model using Vitis AI.
- Integration of application which is generated .elf file with petalinux project.
- Testing: booting image kernal using SD card (boot.bin ,image.ub ,boot.scr)

THANK YOU  
FOR YOUR ATTENTION

Presented by:

Y.Tharun (1602-22-735-099)

U.Sathwik(1602-22-735-103)

P.Supriya Sree(1602-22-735-118)