赛灵思工业物联网研讨会
XILINX IIoT SEMINAR

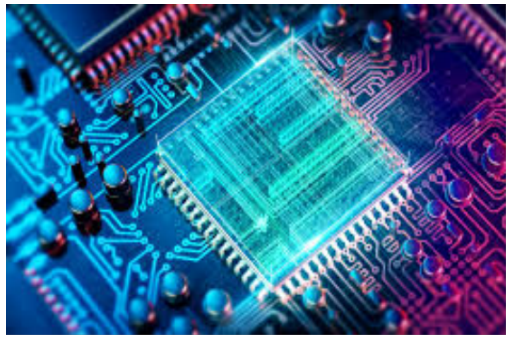
融入Python生态的开源Zynq软硬件设计框架

PYNQ™

陆佳华
joshual@Xilinx.com
学术与创新生态

XILINX

来源: 崔爱国@华为 “边缘计算：对基础软件提出的新挑战” 第三届边缘计算技术研讨会， 杭州，2019年5月



Compute Acceleration

Computational Storage

SmartNICs &
Network Acceleration

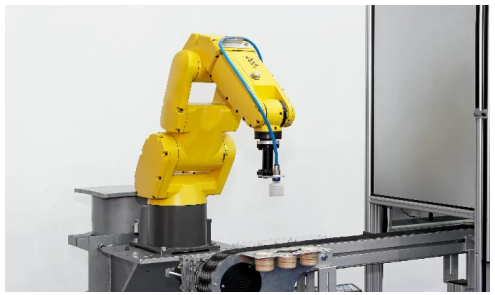
Automotive

Wireless Infrastructure

Wired Communications

Audio, Video, &
Broadcast

Aerospace & Defense



Industrial, Scientific & Medical

Test & Measurement,
and Emulation

Consumer

Three Big Trends

1

Explosion of Data

- >90% unstructured
- >Video & image content
- >Needs higher throughput & real-time computing

2

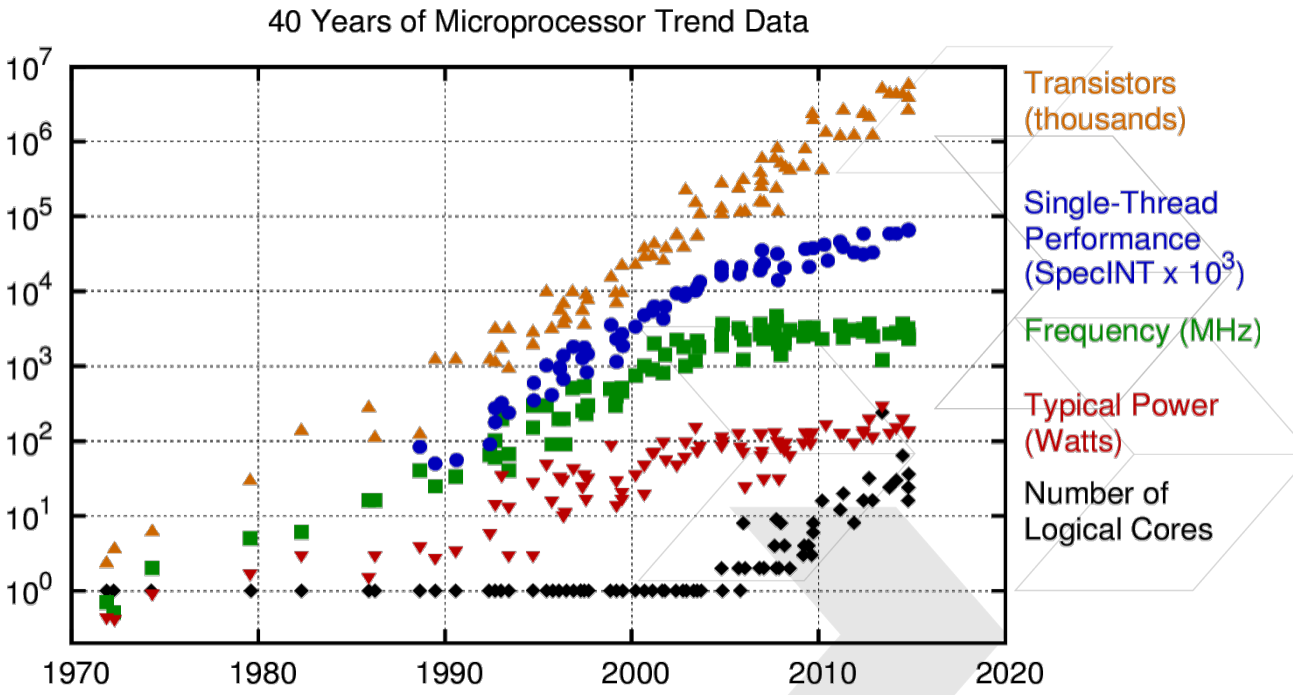
Dawn of AI

- >Adoption across all industries
- >Injecting new intelligence into apps
- >From endpoints to edge to cloud

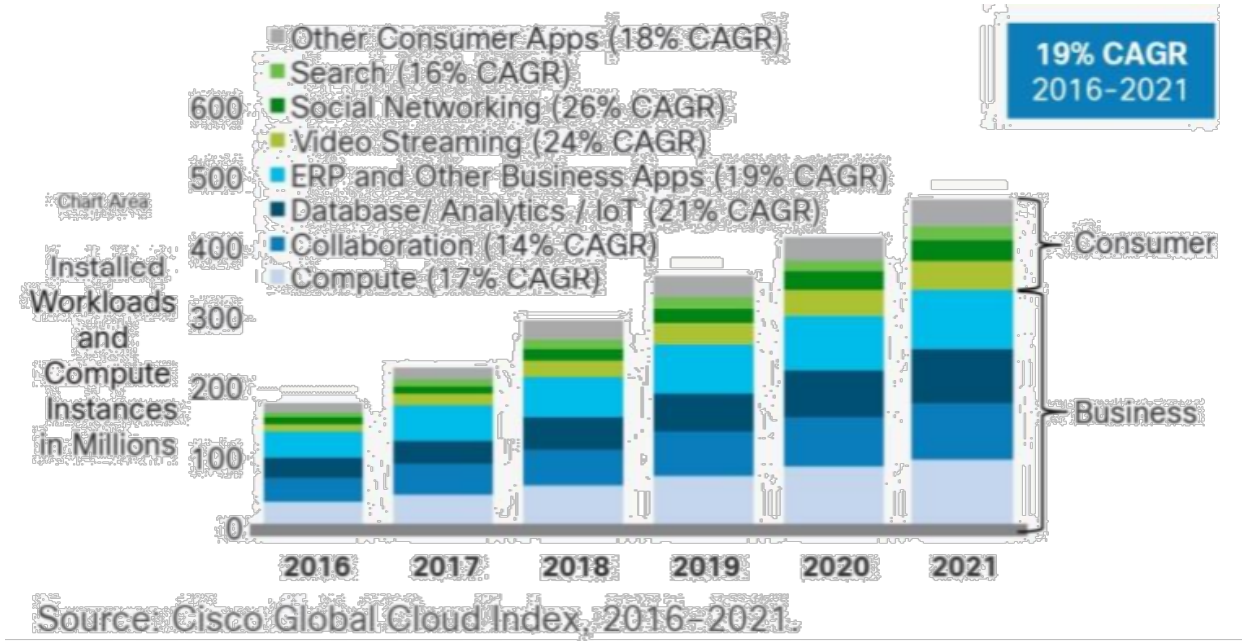
3

Computing After Moore's law

- >Heterogeneous computing with accelerators
- >Breadth of apps require different architectures
- >Speed of innovation outpacing silicon cycles



Diverse and Evolving Applications Needs



- Top 50 hottest applications only ~60% of data centre compute cycles
  - No killer applications
- Large number of diverse and evolving workloads → **Adaptable Acceleration**
  - Compute (Database, Analytics, AI, Video …), Storage, Networking, Security, etc.
- AI infusion into traditional workloads driving additional need for adaptable accel

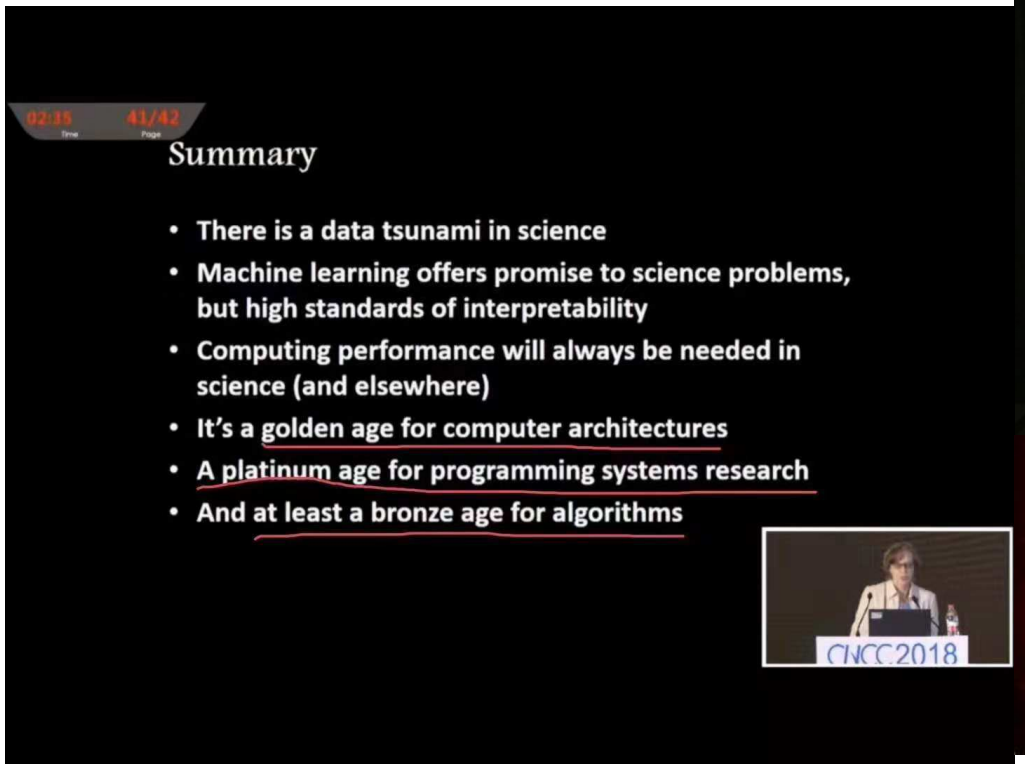
Mountains of Unstructured Data

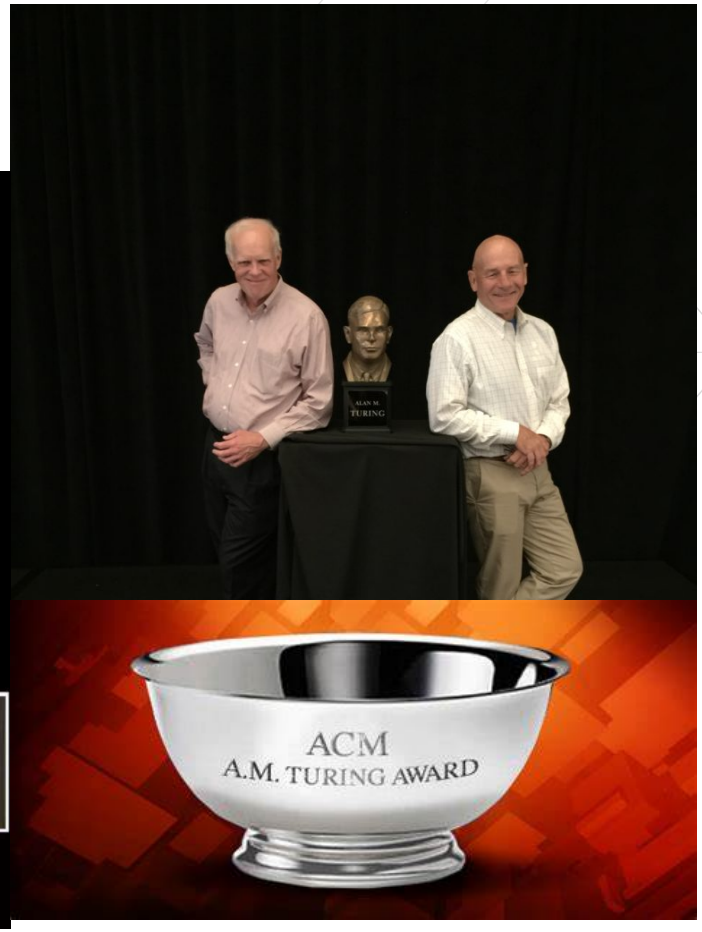
One Architecture Can't Do It Alone

This is the Era of Heterogeneous Compute

Hennessy & Patterson:
A New Golden Age for Computer Architecture

- Hardware/Software Co-Design for High-Level and Domain-Specific Languages
- Enhancing Security
- Free and Open Architectures and Open-Source Implementations
- Agile Chip Development





2017 Electronics Resurgence Initiative

- ***Software Defined Hardware (SDH)***
- ***Domain-specific System on Chip (DSSoC)***
- ***Intelligent Design of Electronic Assets (IDEA)***
- ***Posh Open Source Hardware (POSH)***
- ***Three Dimensional Monolithic System-on-a-Chip (3DSoC)***
- ***Foundations Required for Novel Compute (FRANC)***

注：互联网、半导体、个人计算机操作系统UNIX、激光器、全球定位系统（GPS）等都源于DARPA

HotChip 30 Years

Hot Chips: A Symposium on High Performance Chips

Conference Sponsor IEEE Technical Committee on Microprocessors and Microcompu
SIGARCH.

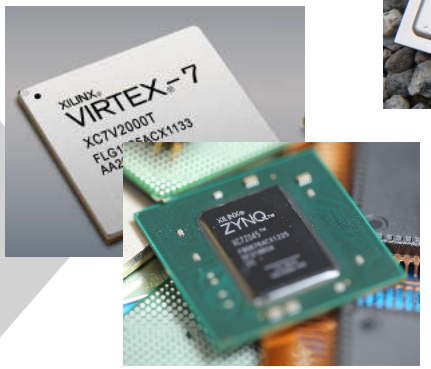
Transformation Through Innovation

World's First FPGA

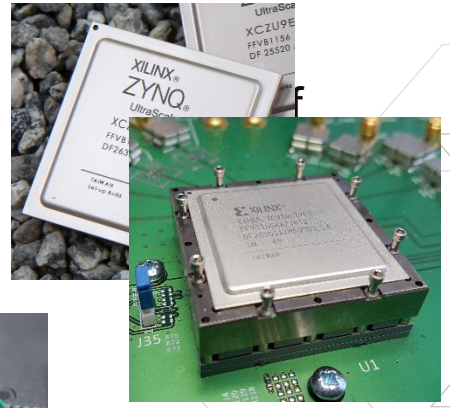
First Virtex FPGA

Virtex-2 Pro

First 3D FPGA & HW/SW Programmable SoC



First MPSoC & RFSoC



ACAP

1980 1990 2000 2010 2020

Driving Adaptive Computing with Versal

The World's First Adaptive Compute Acceleration Platform

- Scalar Processing Engines
- Adaptable Hardware Engines
- Intelligent Engines
  SW Programmable, HW Adaptable
- Breakout Integration of Advanced
  Protocol Engines

Harnessing the Flexibility of Programmable Hardware with Design Tools

SDx™ Environments

Enables Full Programmability for Software Developers

> CPU and GPU-like development environments

> Automated acceleration in programmable logic

HLx™ Editions

15X Productivity for Hardware Developers

> Graphical plug-and-play IP Integration

> High Level Synthesis in C, C++, and System C

VIVADO®

State-of-the-Art Tools for Traditional Hardware Design

> Optimizes for performance, low power, and density

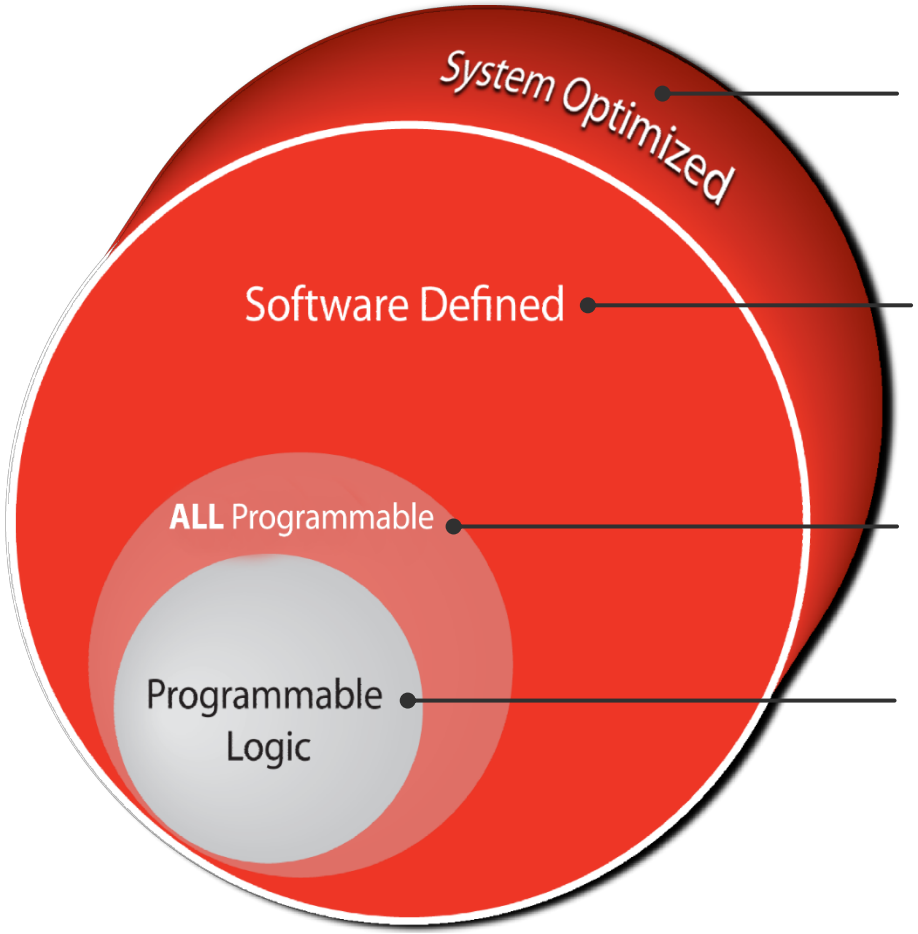
> Advanced flows in-system updates and remote debug

Enabling Innovation
Open Source Pynq Framework

XILINX

Setting the Course for the Next 5 Years

Goal: 5X Potential Users in 5 Years



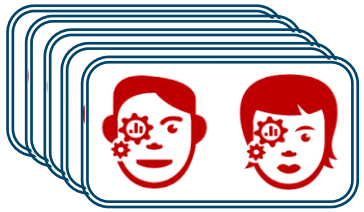
For Highest Growth Megatrends

>250,000 SW Engineers

1000s of SoC Teams with SW *and* HW Engineers

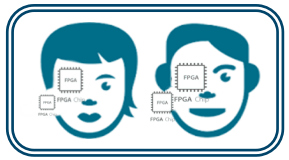
>50,000 HW Engineers

The Goal is Proliferation



Domain Experts,
Software engineers

Emb. software
Engineers



Hardware
Engineers

Modern projects have many, more software engineers

How do we get them to use Xilinx devices?

How can we let them experience Xilinx advantages?

We must speak to them in their ‘language'

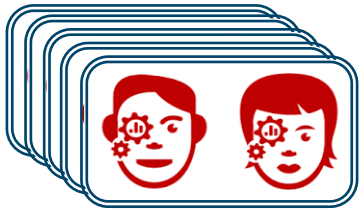
They do not use CAD tools

PYNQ™

Is a framework for creating platforms that software engineers can use to experience the benefits of Xilinx silicon

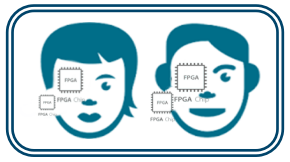
PYNQ enables rapid development for FPGA designers and embedded s/w engineers also

PYNQ Python Productivity for Zynq



Domain Experts

Embedded software Engineers



Hardware Engineers

Targeting the data center artificial intelligence, machine learning, data science

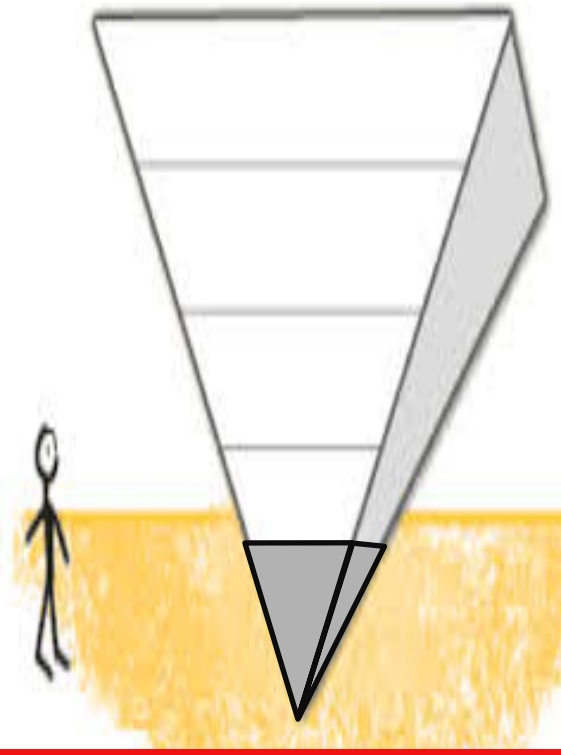
New users are not hardware designers, or embedded systems designers

PYNQ

Enable more people to program Xilinx processing platforms, more productively

Productivity Languages & Hardware Overlays

Zynq / Zynq UltraScale+



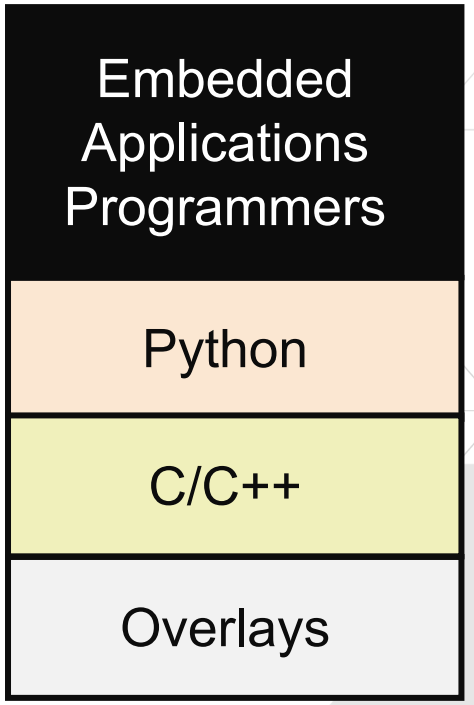
Users:

Apps Programmers:

Programmers(embedded):

Device driver writers:

Kernel developers:



Small group of experts create APSoC overlays and C API/drivers
Many more users build applications in C/Python

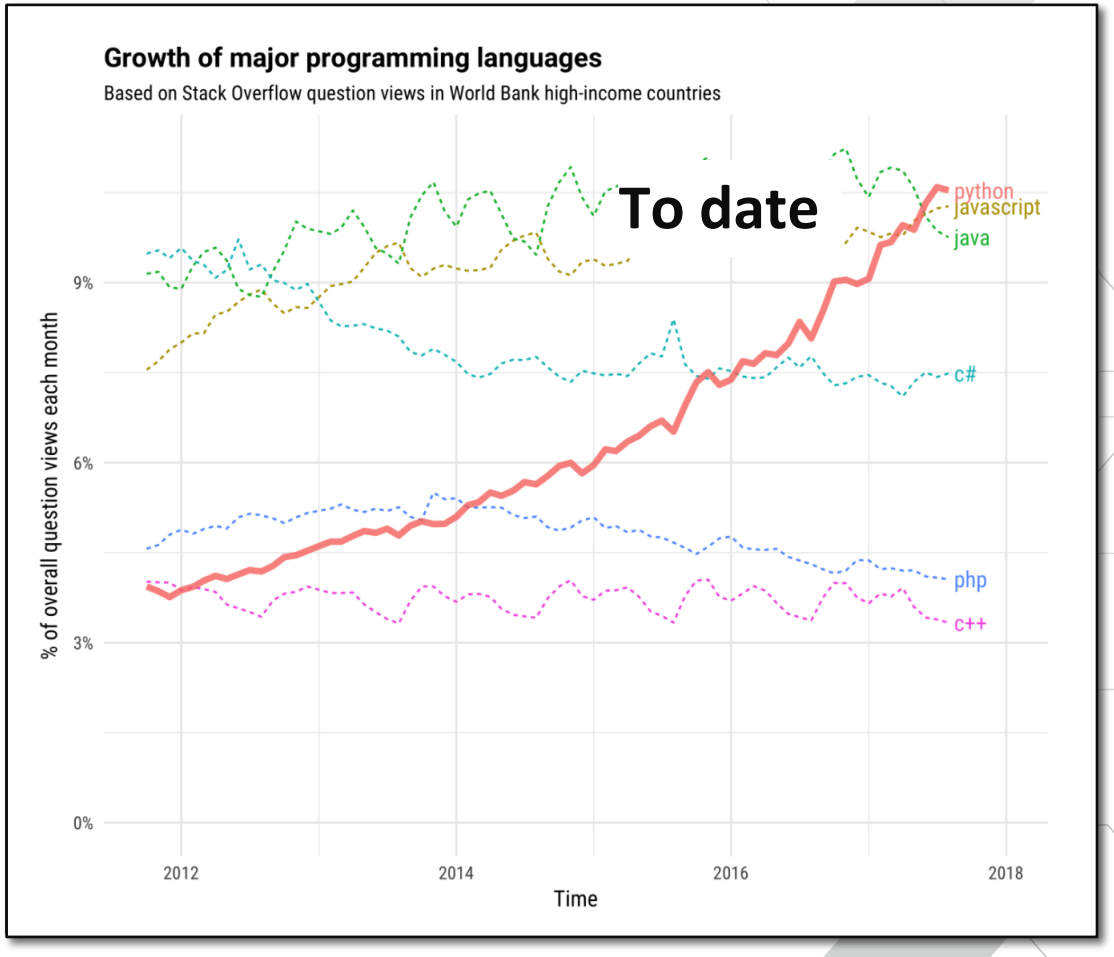
Python is increasingly the Language of Choice

Top Programming Languages,
IEEE Spectrum, July'18

Language Rank	Types	Spectrum Ranking
1. Python		100.0
2. C++		98.4
3. C		98.2
4. Java		97.5
5. C#		89.8
6. PHP		85.4
7. R		83.3
8. JavaScript		82.8
9. Go		76.7
10. Assembly		74.5

Python is listed as an embedded language for the first time

https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages



https://stackoverflow.blog/2017/09/06/incredible-growth-python/

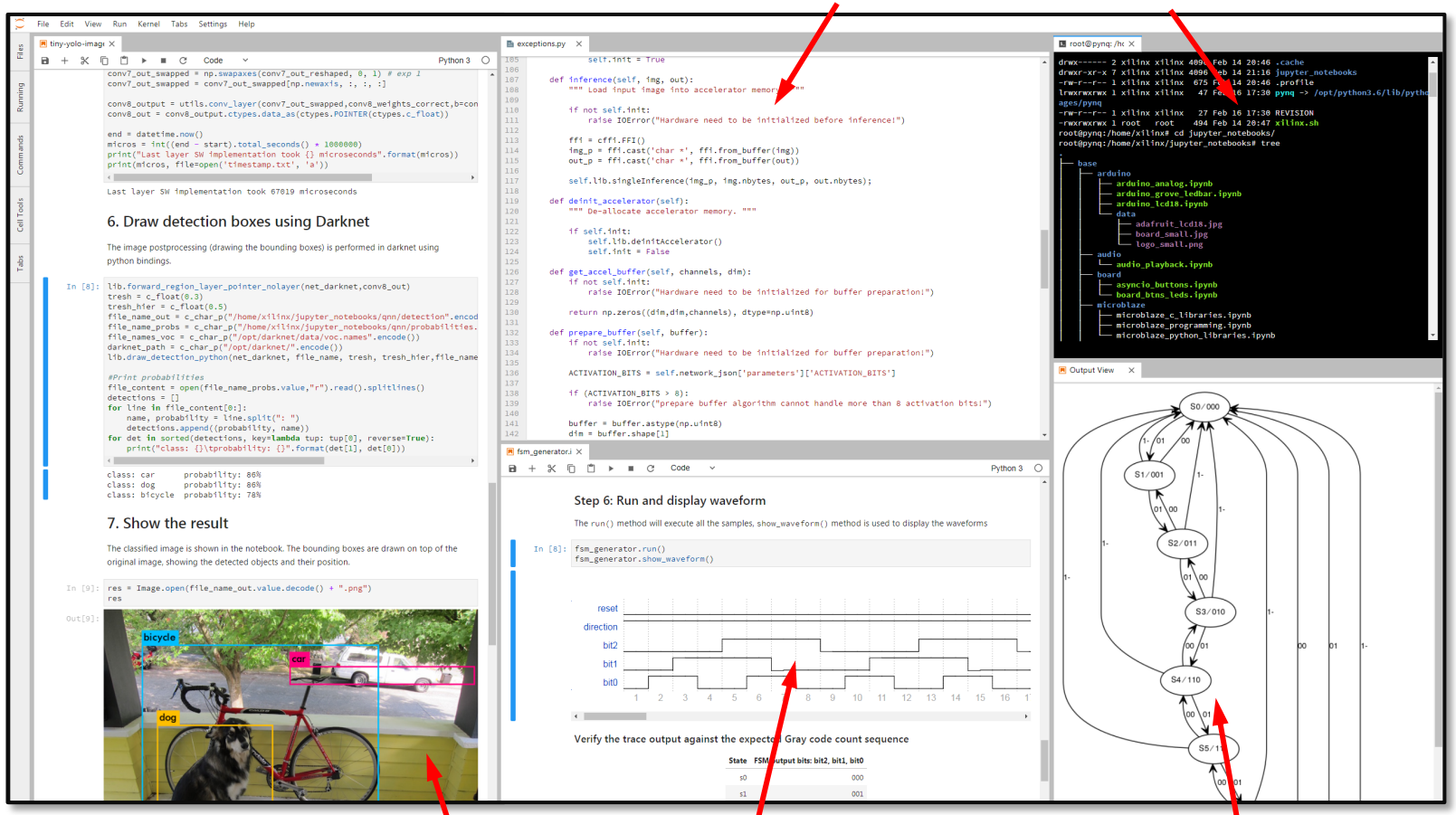
Python is the fastest growing language: driven by data science, AI, ML and academia

Jupyter Notebooks to JupyterLab IDE

2017 ACM
Software System Award

Code editor

Terminal



Jupyter notebooks

Visualization

Jupyter … Julia, Python, R

Default engine of data science

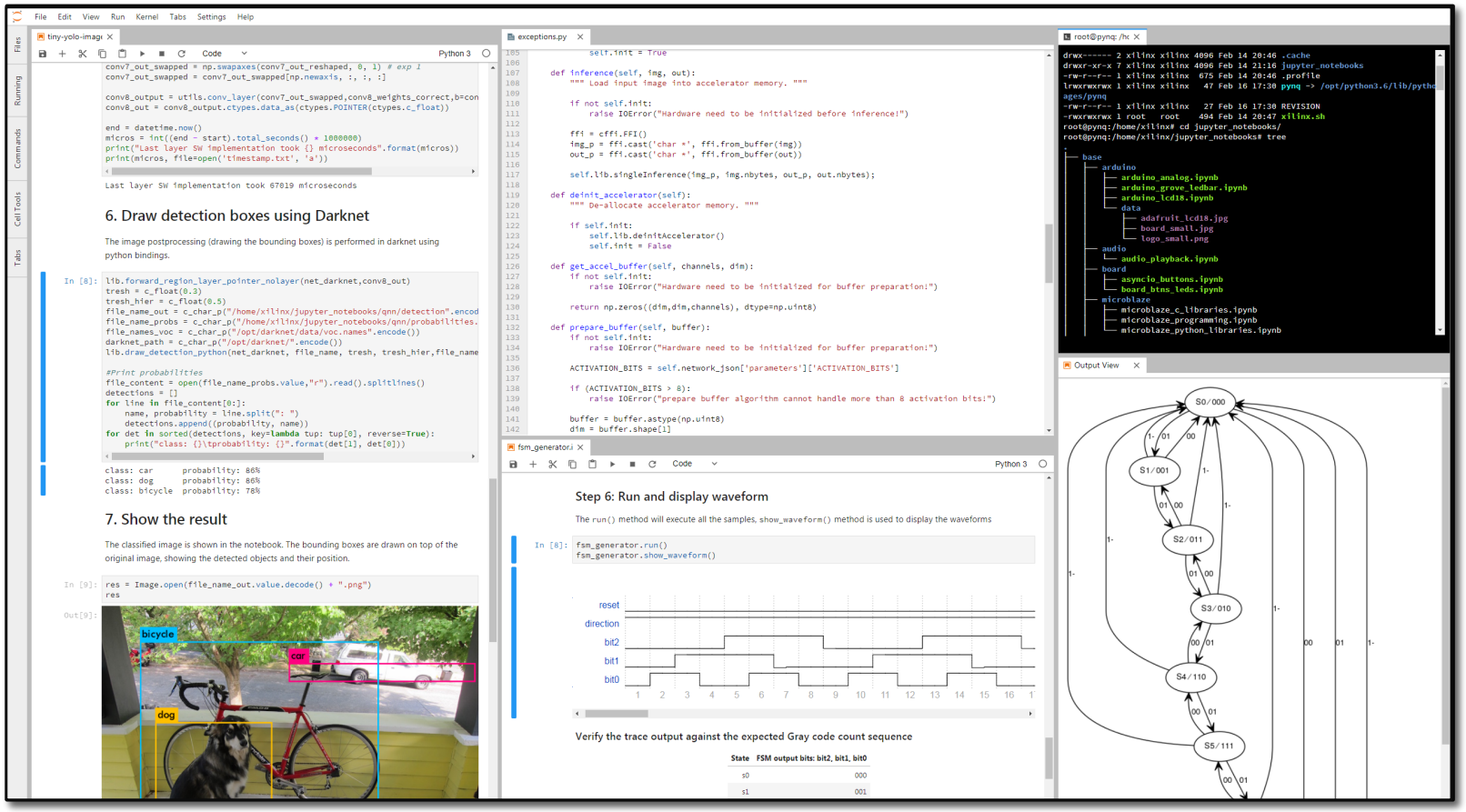
Taught to 1,000+ Berkeley students every semester

2+ million notebooks on GitHub

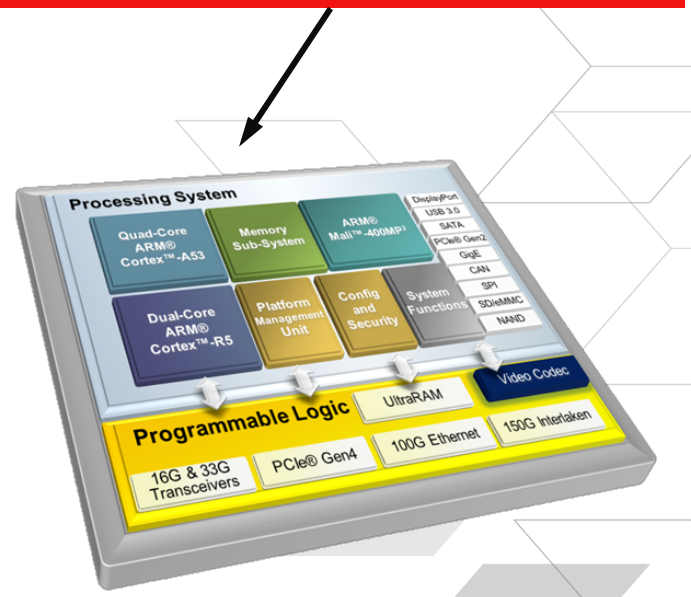
Next-gen browser IDE

Includes Jupyter Notebooks

JupyterLab runs on Zynq and Zynq UltraScale+



Runs natively on ARM A9 and ARM A53 processors

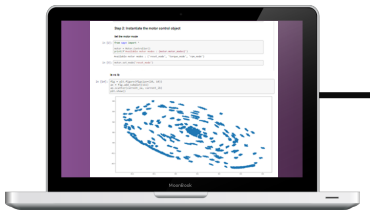


Performant on Zynq 32-bit ARM A9 at 650 MHz, 512MB DRAM, no GPU

Python productivity for Zynq

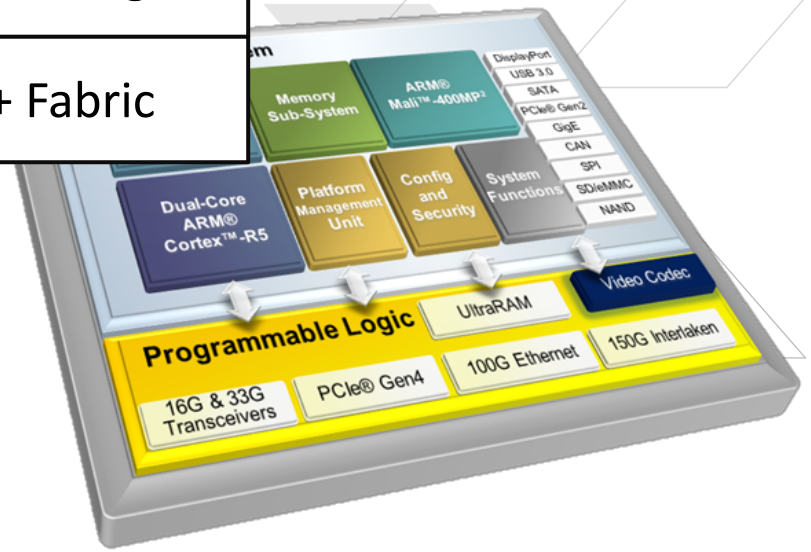
Jupyter notebooks/lab, browser-based interface

PYNQ enables Jupyter on Zynq and ZU+



Jupyter web server	
IPython kernel	
Ubuntu-based Linux	Overlays/designs
PS: ARM A53s	**PL**: ZU+ Fabric

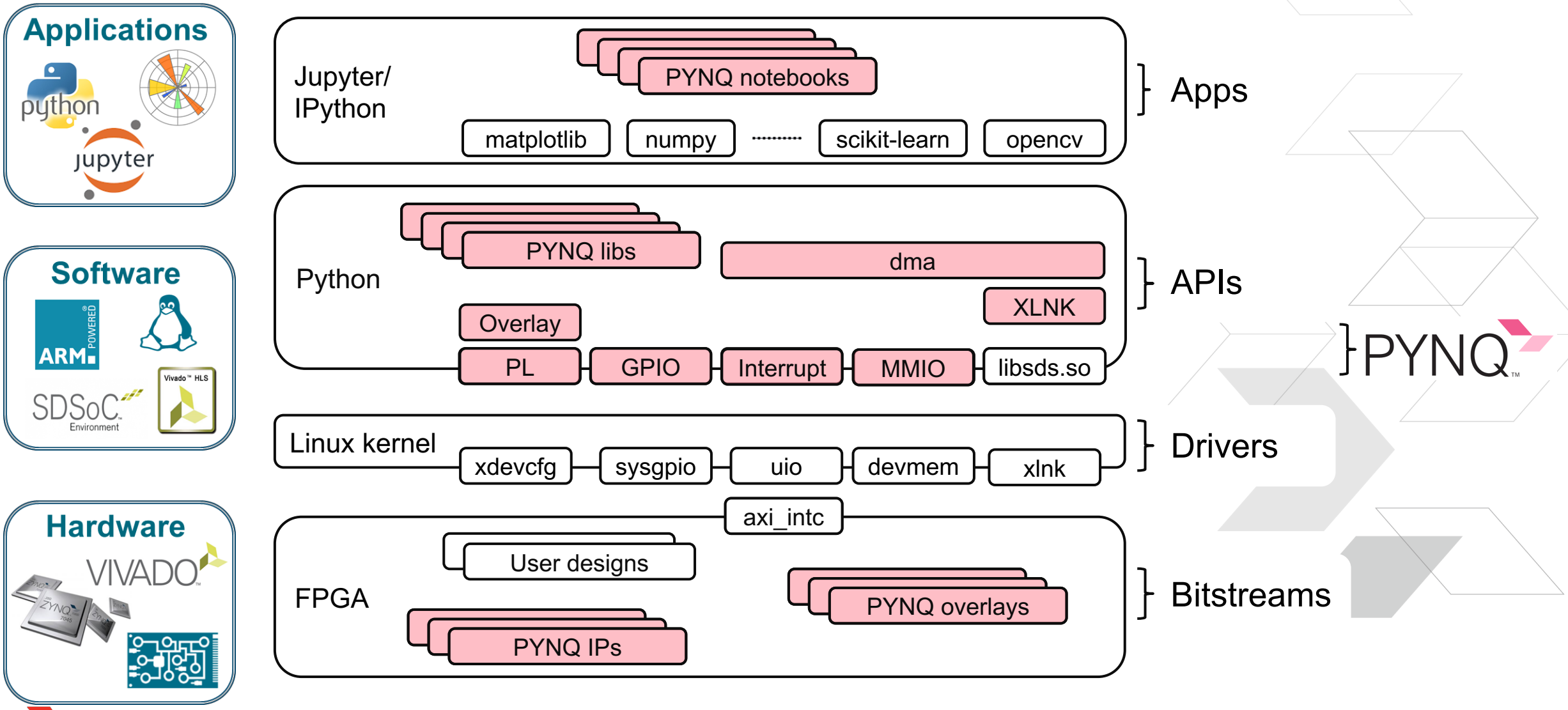
Hardware drivers wrapped in Python libraries



Optimized for developer productivity

- All the Linux libraries and drivers you expect
- Pre-built SD image
- Ubuntu/Debian ecosystem & community
  - >> 110,000,000 Google hits

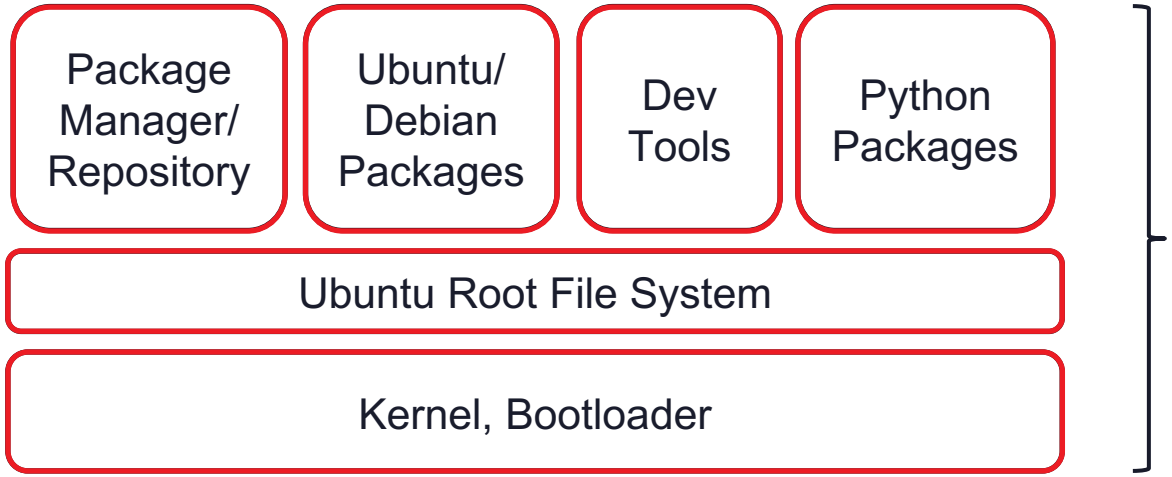
PYNQ is a Framework



PYNQ's Ubuntu-based Linux

PYNQ uses Ubuntu's:
- Root file system (RFS)
- Package manager (*apt-get*)
- Repositories

PYNQ bundles :
- Development tools
  - Cross-compilers
- Latest Python packages



PYNQ's Ubuntu-based Linux

PYNQ uses the PetaLinux build flow and board support package:
- Access to all Xilinx kernel patches
- Works with any Xilinx supported board
- Configured with additional drivers for PS-PL interfaces

Ubuntu-based Linux versus embedded Linux

Ubuntu-based Linux

Optimized for developer productivity

- All the Linux libraries and drivers you expect
- Pre-built SD card image
- Ubuntu/Debian ecosystem & community
  - >>145,000,000 Google hits

3 orders of magnitude difference

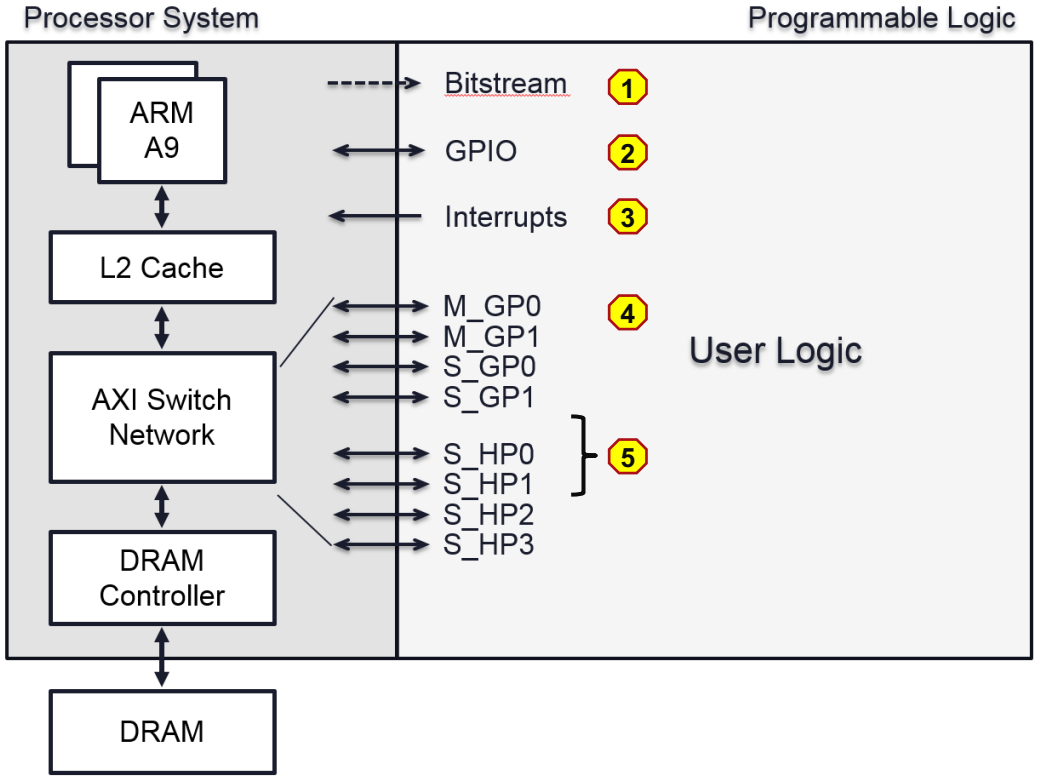
Embedded Linux

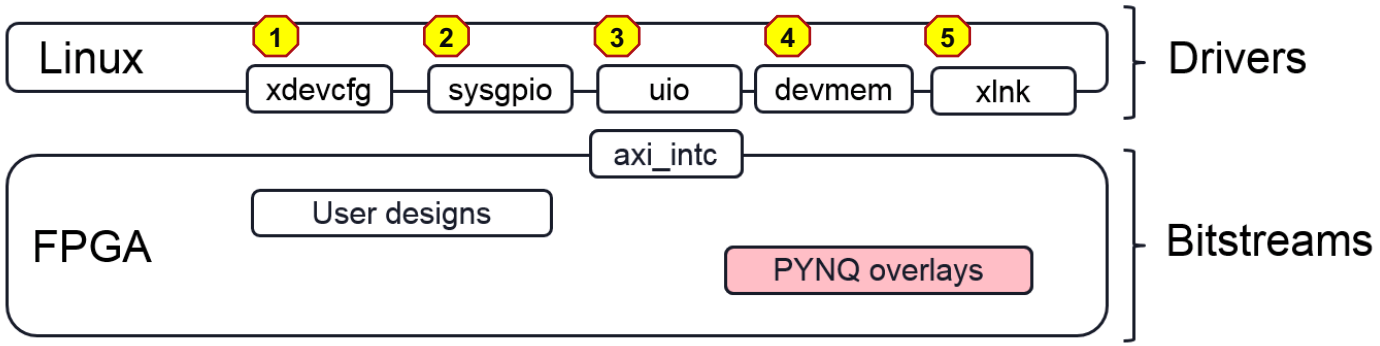
Optimized for deployment efficiency

- Selective Linux libraries and drivers
- Commonly delivered in flash memory on board
- PetaLinux ecosystem:
  - >> 143,000 Google hits

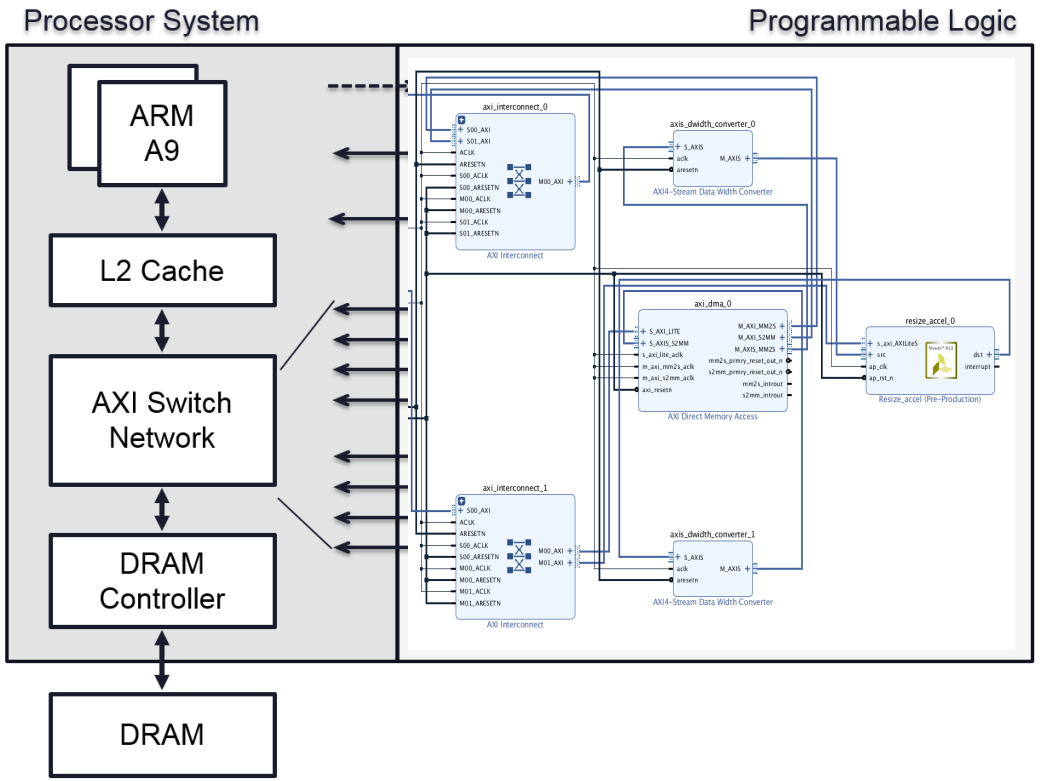
PYNQ provides Linux drivers for PS-PL i/fs …

wrapped in Python libraries

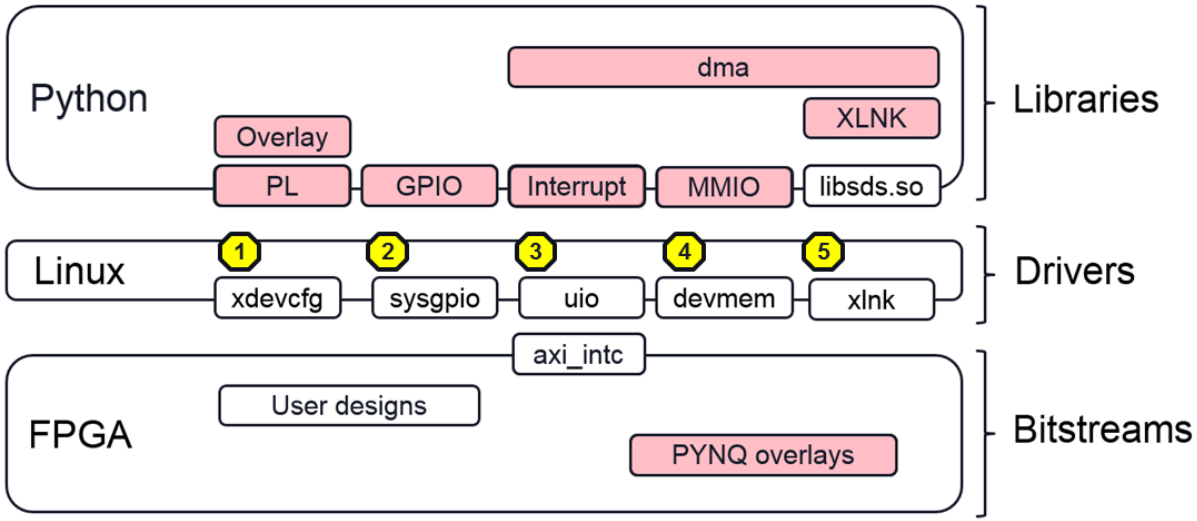




Loading a design into Zynq using PYNQ

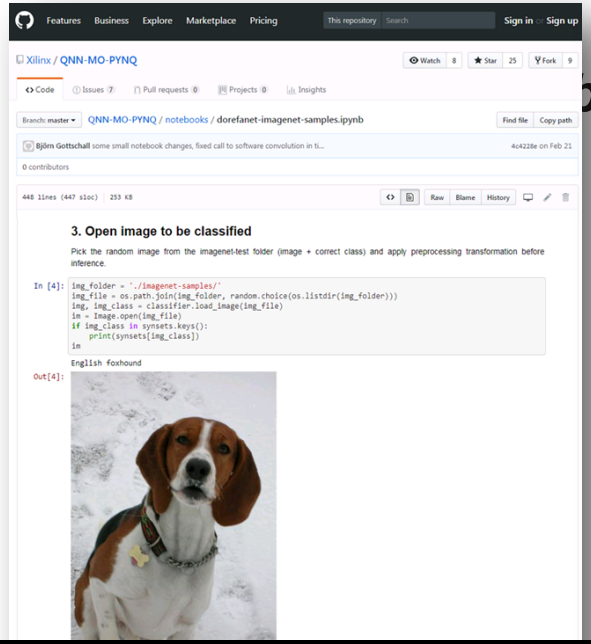


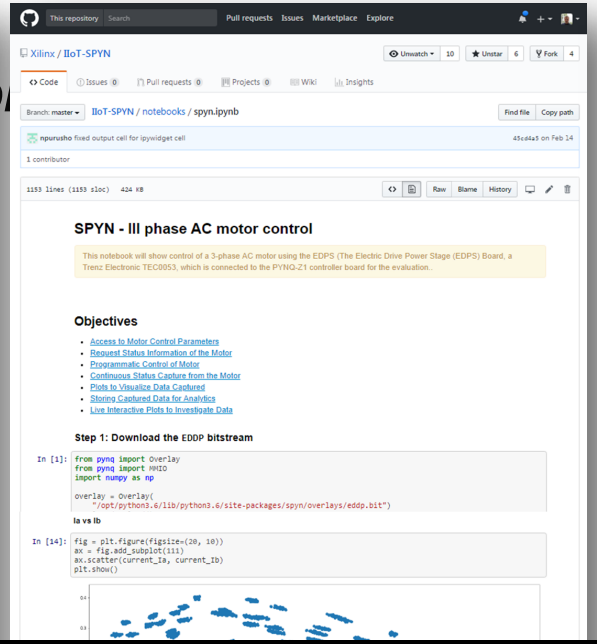

```
from pynq import Overlay
resizer = Overlay('./resizer.bit')
```

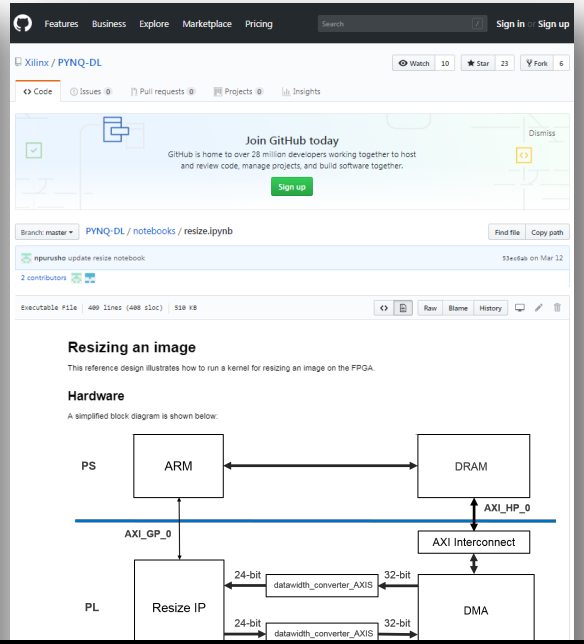


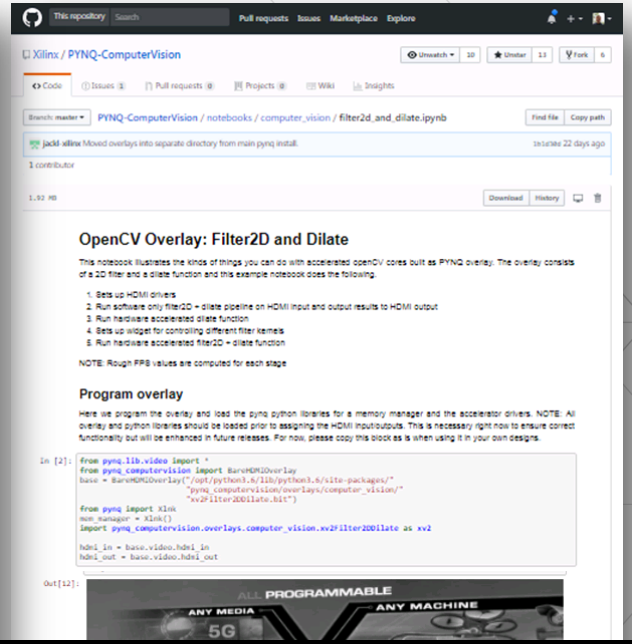
PYNQ automatically configures many design parameters based on data parsed from hybrid library

Software-style packaging & distribution of designs










```
Download a design from GitHub with a single Python command:

pip install git+https://github.com/Xilinx/pynqDL.git
```

Next steps: scaling across platforms and domains

ISM ML

python

jupyter

matplotlib

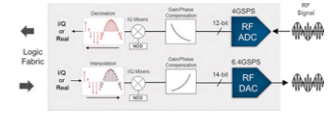
IIoT vision

OpenCV

jupyter

python Package Index

SDR



SciPy

jupyter

Software

ARM POWERED

Vivado™ HLS

Software

ARM POWERED

SDSoC Environment

Software

SDAccel Environment

ARM POWERED

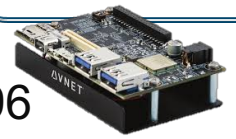
Zynq

VIVADO

ZynqU+

VIVADO

Avnet Ultra96



RFSoC

VIVADO

Zynq UltraScale+ RFSoC

How Numpy interacts with Programmable Logic?

NumPy

Array, Matrix

```
In [7]: import numpy as np
        import pynq

        def get_pynq_buffer(shape, dtype):
            """ Simple function to call PYNQ's memory allocator with numpy attributes

            """
            return pynq.Xlnk().cma_array(shape, dtype)
```

Virtual memory

PYNQ

Xlnk

Physical memory

Contiguous Array in DDR Memory
Shared Object

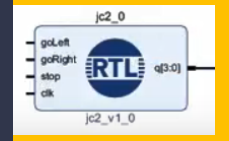
Physical memory

Hard Real Time
Offload CPU and OS
Dedicated “Accelerator”

A X I H P

VIVADO HLx Editions

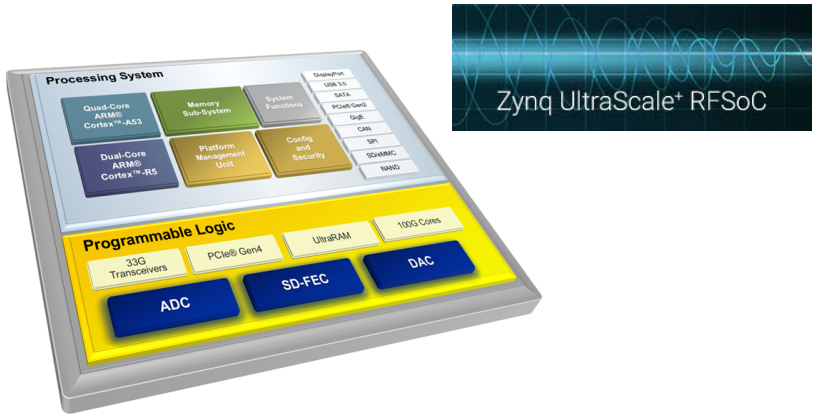
SDSoC Environment



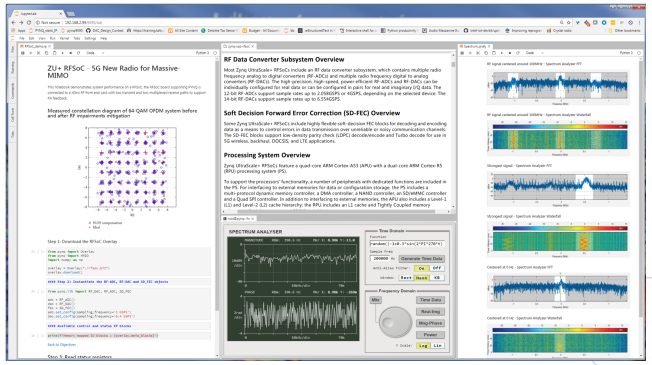
PL

BIG DATA tools for BIG DATA devices

BIG DATA device

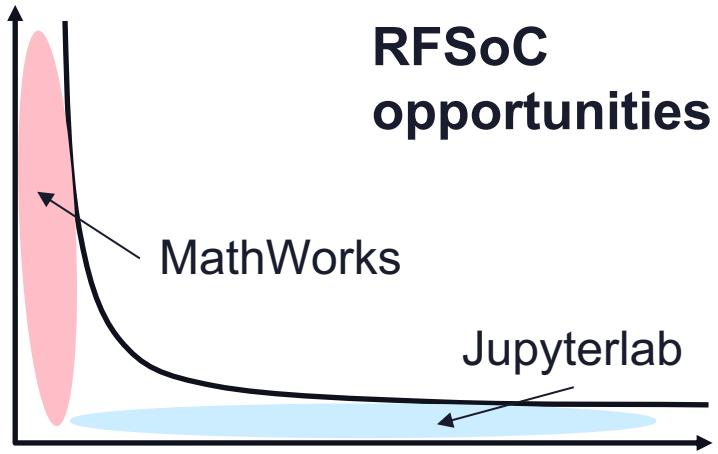


BIG DATA IDE



jupyter

JupyterLab

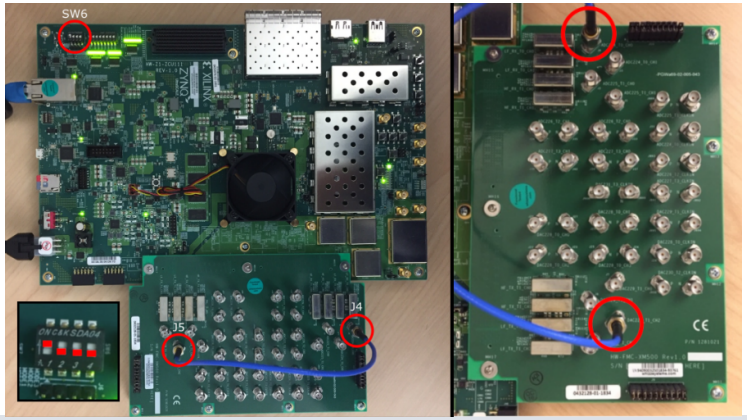


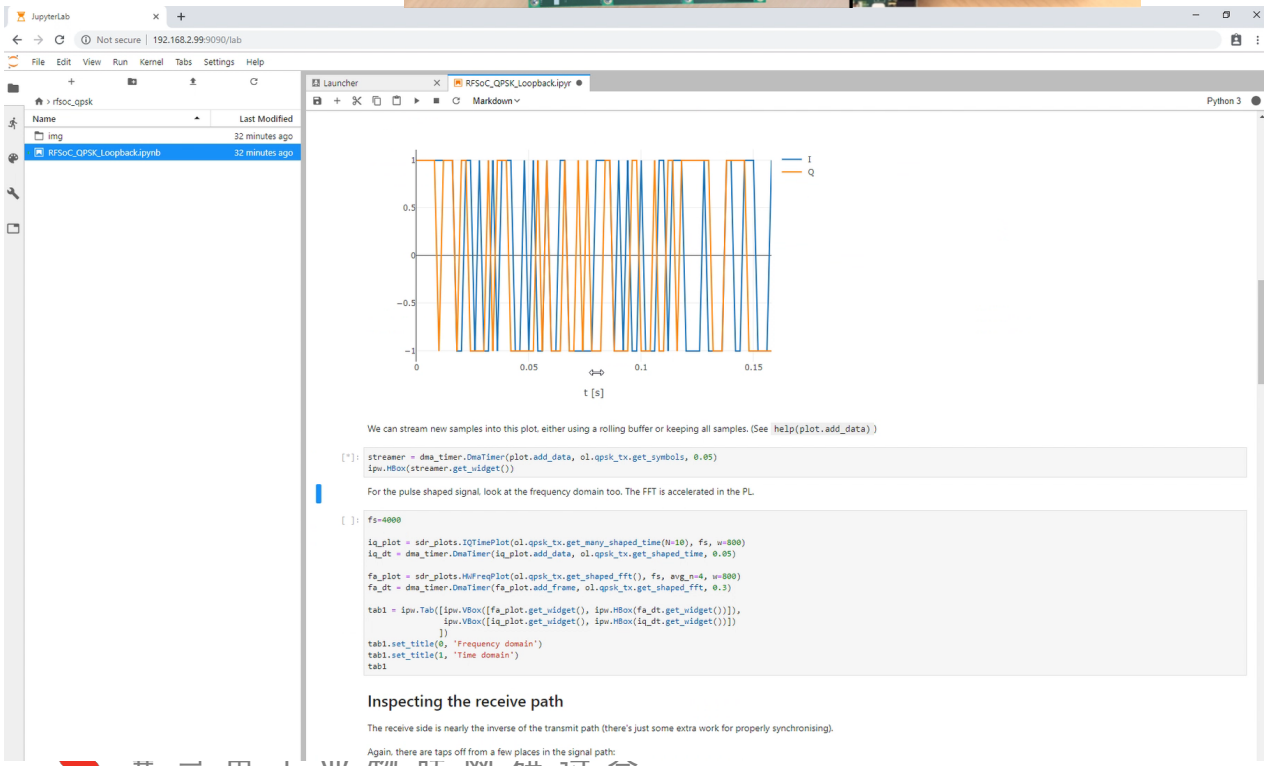
RFSoC opportunities

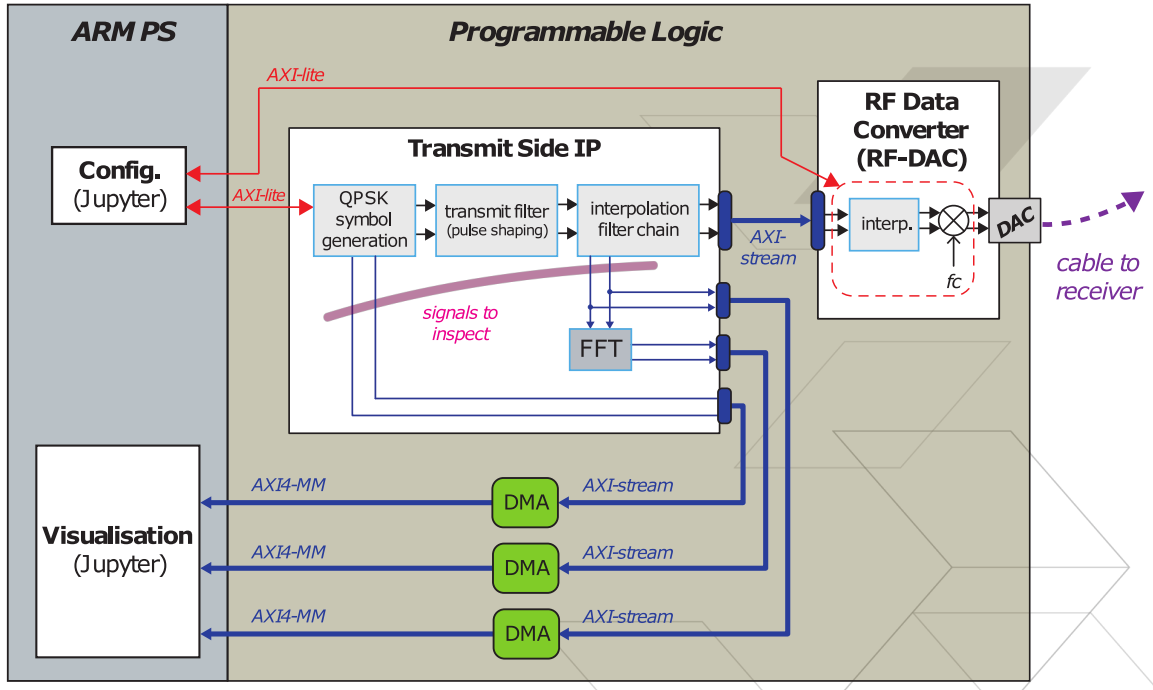
- State-of-the-art BIG DATA analysis
- State-of-the-art BIG DATA interactive visualization
- Opportunities: ML and SDR
  - Cognitive radio, etc

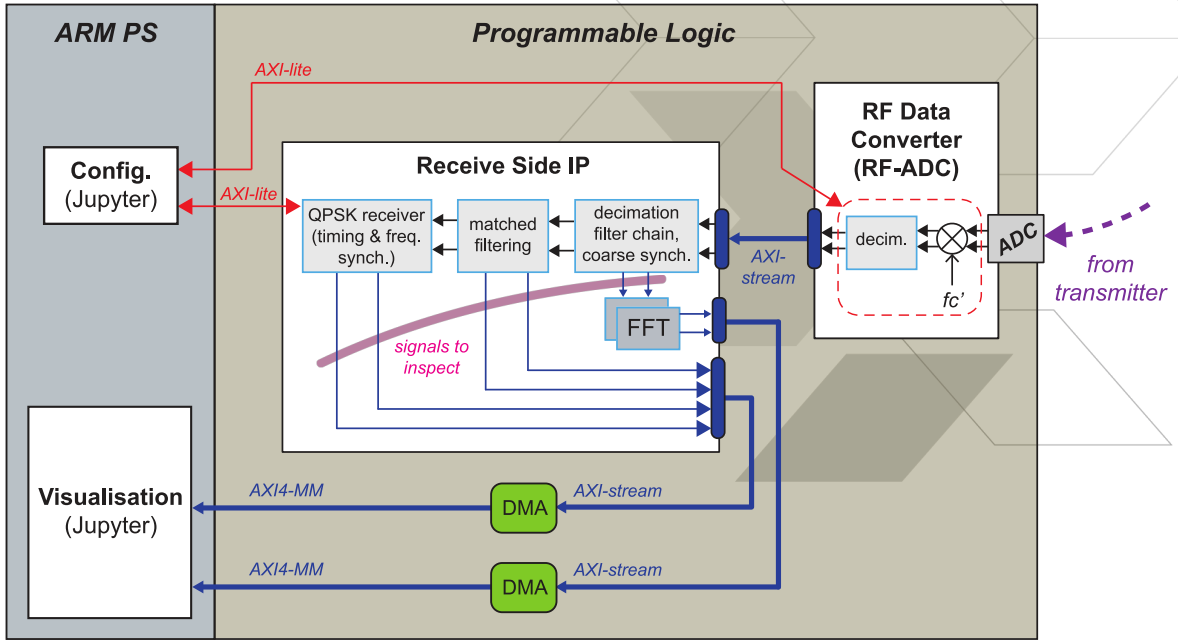
University of Strathclyde

RF_QPSK Demo





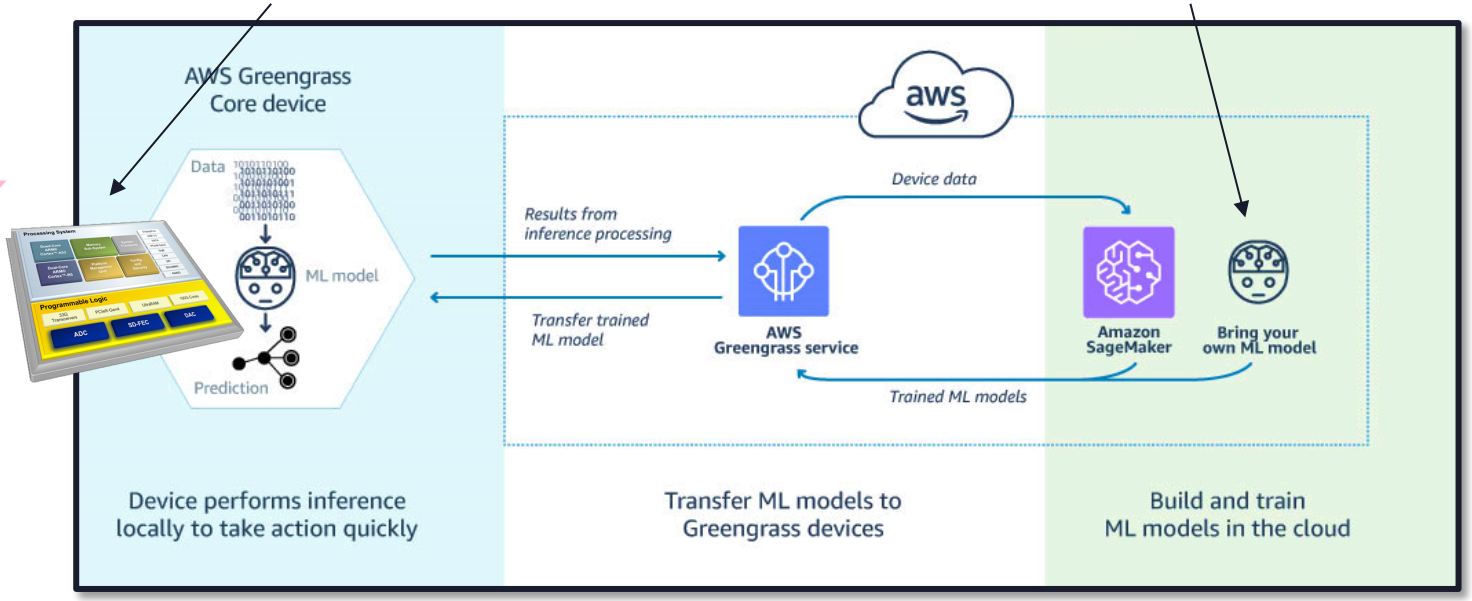




Edge-to-cloud co-design

Common JupyterLab tooling at edge and cloud

PYNQ

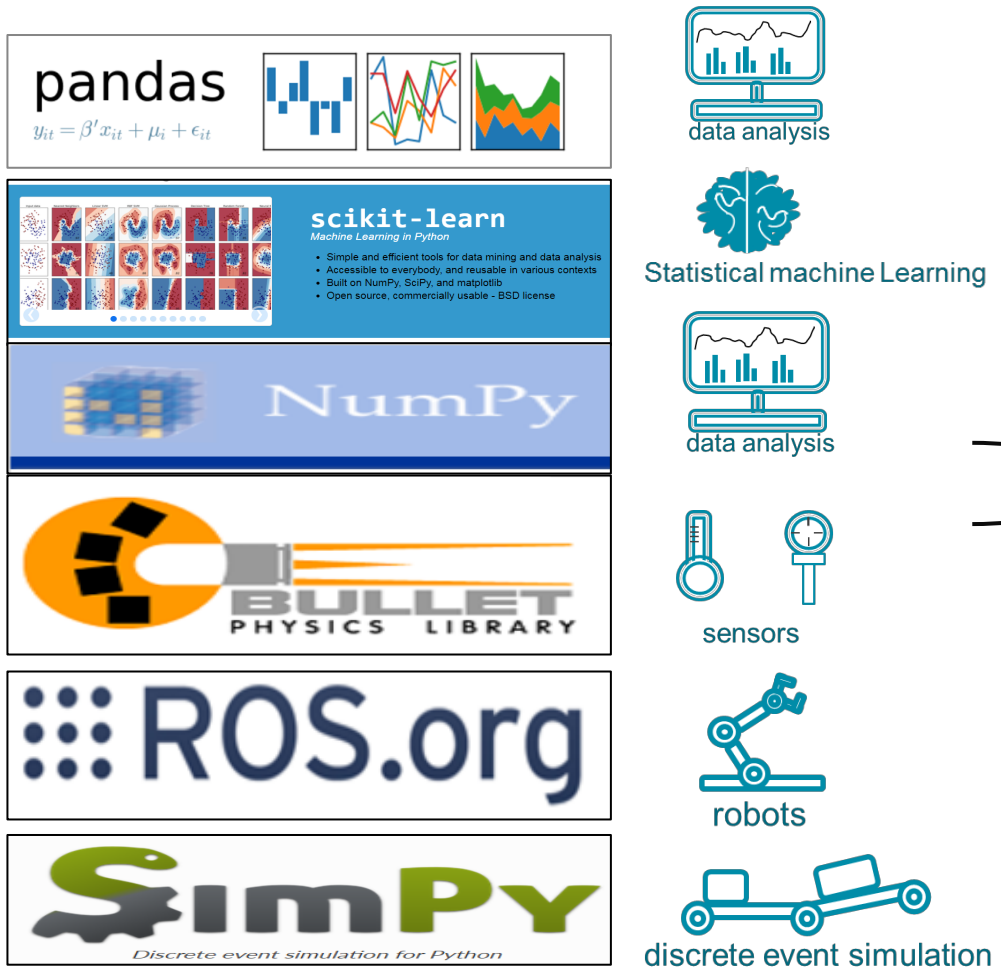


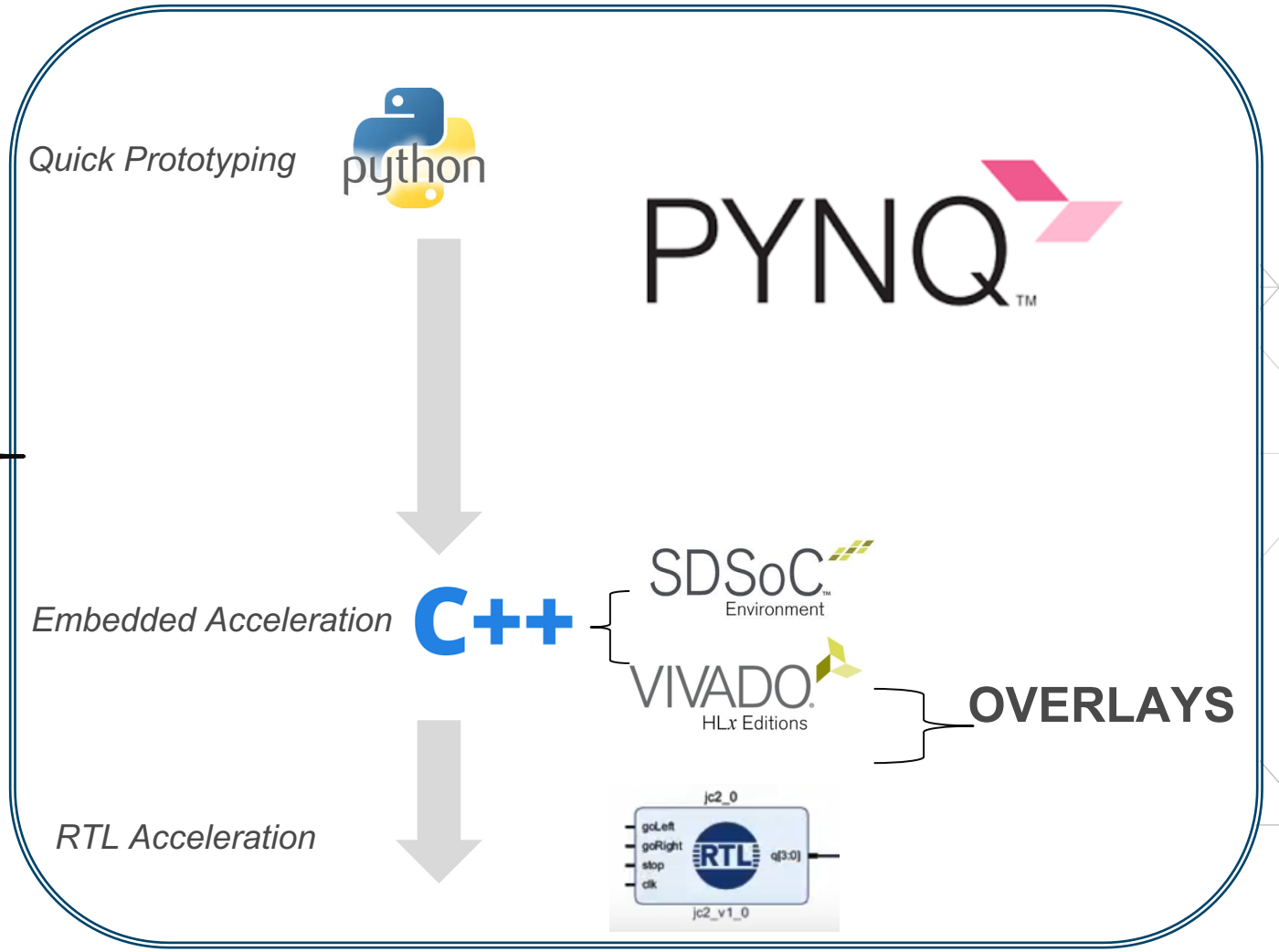
PYNQ enables ML experts and radio engineers to focus on their ‘value-add’

Edge-to-cloud co-design trade-offs:

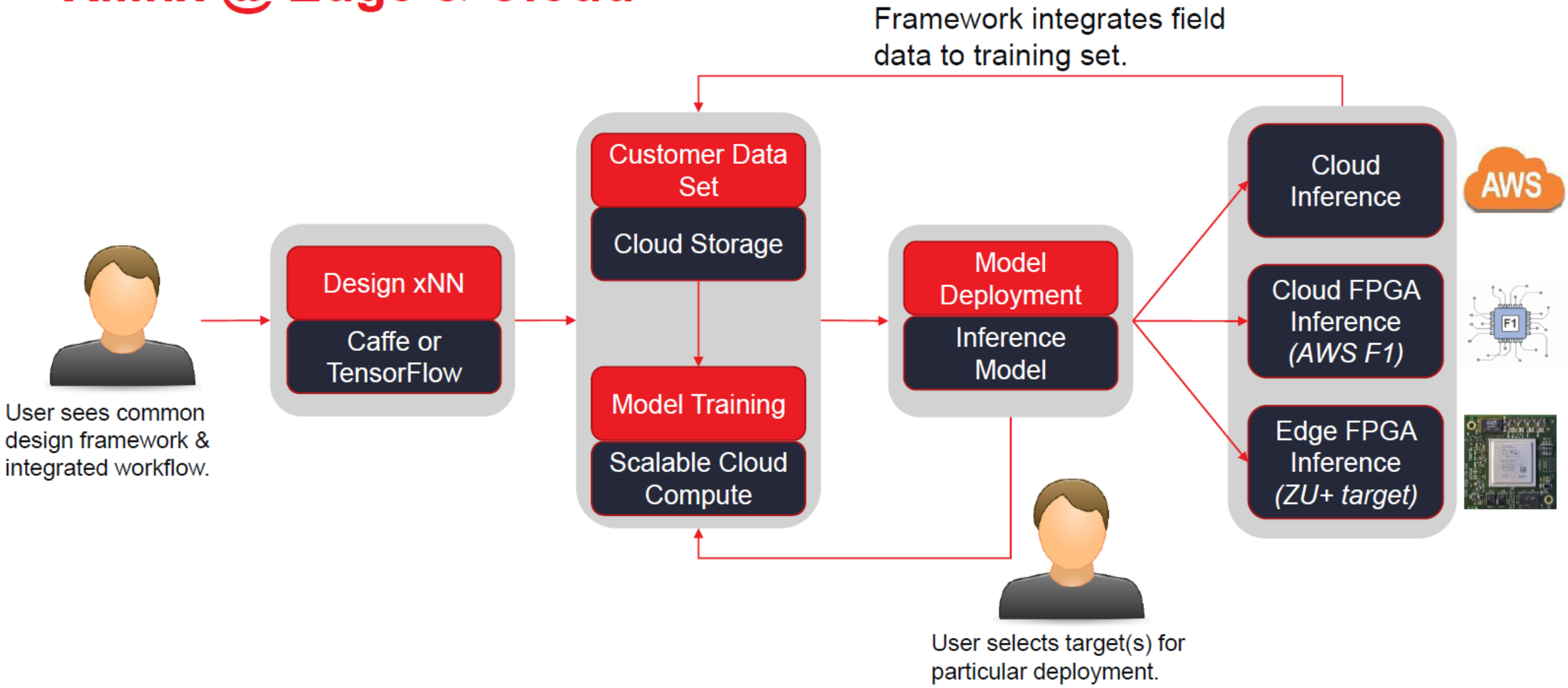
- Maximize on-chip processing
- Minimize edge-to-cloud data exchange
- Exploit scalability of cloud processing
- Aggregate intelligence between and across multiple edge nodes
- Co-optimize the above for best system performance

Connect other Python Libraries -





AWS + Xilinx - SageMaker ML Framework target
Xilinx @ Edge & Cloud



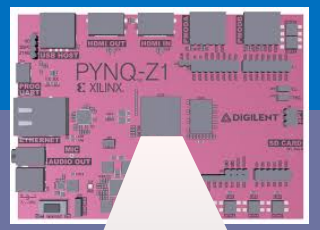
XILINX.

How PYNQ is bridging hardware and software

Software

Open Source

Hardware



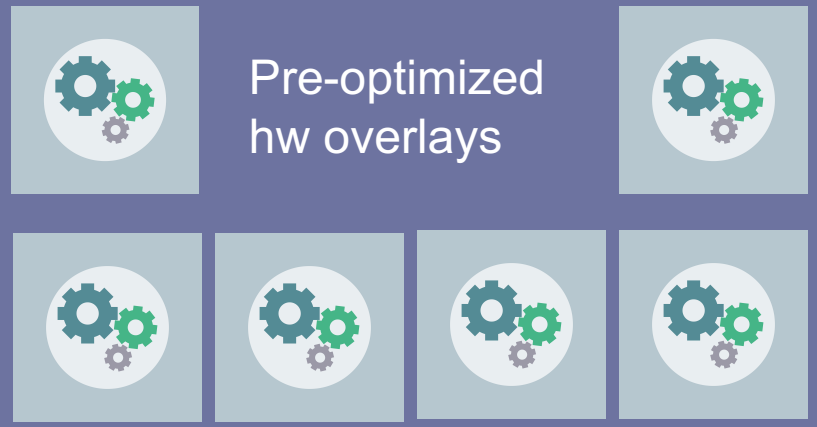
Python stack

1001011100111111…

All programmable hardware

Python software libraries

+100,000 Python libraries



Pre-optimized hw overlays

Efficient porting PYNQ to any Zynq-based platform

3 **Target-specific PYNQ components**
- Board-specific porting
- Common interfaces available for re-use

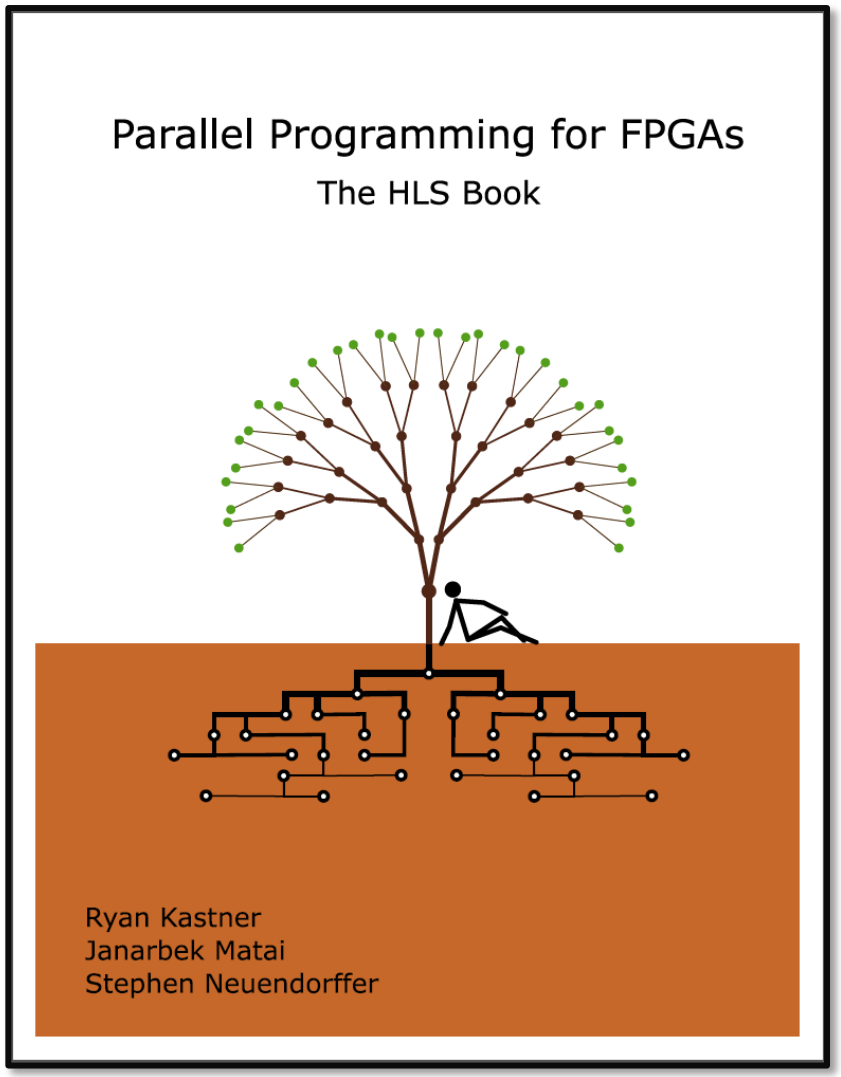
2 **PYNQ Software Core (board independent)**
- Pre-built images available
- Expected to take less than 1 day to port

1 **PetaLinux build tools/BSP**
- Available for any Xilinx-supported board;
- Instructions for building for custom boards available

ALINX® MYiR 威视锐® V3 Technology, Ltd.

Actions

Open source HLS book



Parallel Programming for FPGAs is an open-source book aimed at teaching hardware and software developers how to efficiently program FPGAs using high-level synthesis

公众号回复 pp4fpgas

CECA@PKU FPGA Accelerator

PYNQ & HLS | 跟着北大CECA学FPGA加速器设计

Pooterko Xilinx学术合作 1 week ago

本文的目标是帮助对于深度学习硬件加速器设计感兴趣的朋友快速上手
习加速器设计。

准备

以下是阅读本文的基础，请做好下列基础准备后再上手加速器设计：

1.C语言设计：熟练掌握C语言语法。

2.计算机体系结构知识：参考书《计算机组成与设计》，不需要熟读全
器设计相关的基础概念有比较清晰的理解和认识，如流水线、数据并行

高层次综合

利用高层次综合工具，开发者只需要编写高级语言的代码完成程序功
编写的代码综合成相同功能的RTL级实现(基于Verilog或VHDL)。开发
些pragma的方式来指示和调整高层次综合工具生成的硬件模块的架构。
层次综合工具进行FPGA硬件开发的过程，应该是利用软件语言的表达
程。目前，高层次综合的代码都是基于C/C++/OpenCL的，所以对于
朋友来说，利用高层次综合工具可以大幅度地降低学习难度，缩短开
代速度。

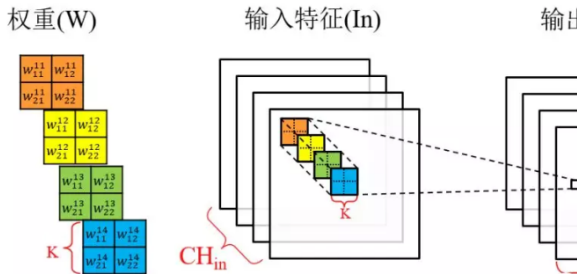
3天入门实例

我们需要使用一个简单的实例来进行入门学习。既然目标是让大家
器的硬件设计，那么我们的实战示例就选择使用目前最火爆、最具
——卷积神经网路(CNN)。

我们选取卷积神经网络前向计算中耗时最长的卷积操作作为我们
来，需完成以下步骤：

1. 准备工作(1天)下载Xilinx Vivado HLS或Xilinx SDx工具链，利
软件工具的使用，包括：新建工程、配置工程参数、综合流程等。

2. 软件实现(1天)实现卷积层的软件版本(C语言版本)，并封装成一
现，结合高层次综合工具的report和analyze工具分析理解所生成的
卷积运算的流程如下图所示：



在HLS工具中重新综合，发现延迟降到了仅仅**1782**个时钟周期(具体数字可能略有差异)，相对于原始无优化版本的加速达到了**141.06**倍！HLS的综合报告里显示加速器调用也已经完全流水化了：

Detail

Instance

Loop

	Latency			Initiation Interval			
Loop Name	min	max	Iteration Latency	achieved	target	Trip Count	Pipelined
- Kernel_Row_Row_Column	1782	1782	20	1	1	1764	yes

综上，我们就完成了一个高效的卷积运算加速器，而基本上利用HLS工具设计FPGA硬件加速器的入门也就完成了。总结来说，使用HLS设计FPGA加速器的一般化设计流程如下：

1.熟悉并理解目标算法，通过软件运行目标算法，分析性能瓶颈所在；

2.实现加速目标的软件版本，分析其中可以并行、流水化的部分，并构想可行的硬件架构；

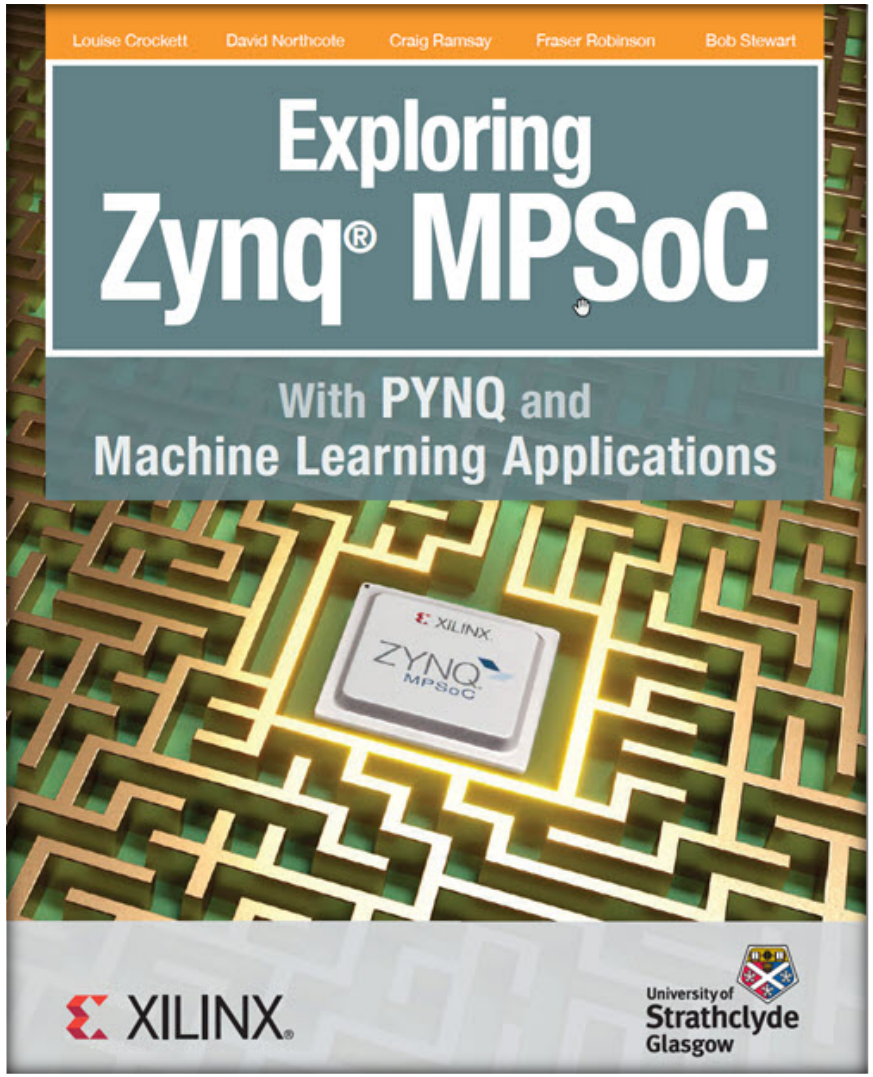
3.通过代码重构，加pragma等方法在HLS工具中描述目标架构，此过程需注意保证改写的代码功能性上与原代码严格保持一致；

4.调整硬件参数配置，最大化利用硬件资源(计算资源如DSP、存储资源如BRAM)来最大化加速器的性能。

30天精通学习

在完成了上面的3天入门实例后，大家可以进一步学习和实践FPGA加速器的设计，这一部分我们推荐大家利用3到4周的时间对相关知识进行详细、系统的学习。高层次综合的相关

Open source MPSOC book



The book covers the architecture of the device, the design tools and methods, conventional hardware/software co-design approach, and the newer software-defined methodology, as well as hardware and software development, multiprocessing, safety, security and platform management, system booting, PYNQ and machine learning applications

XILINX
University Program

公众号回复 mpsocbook

Curriculums All in One (Pynq- Z2)

Digital Logic

Computer Architecture

On Chip Networks

Machine Learning

Image and Video Processing

Speech Recognition

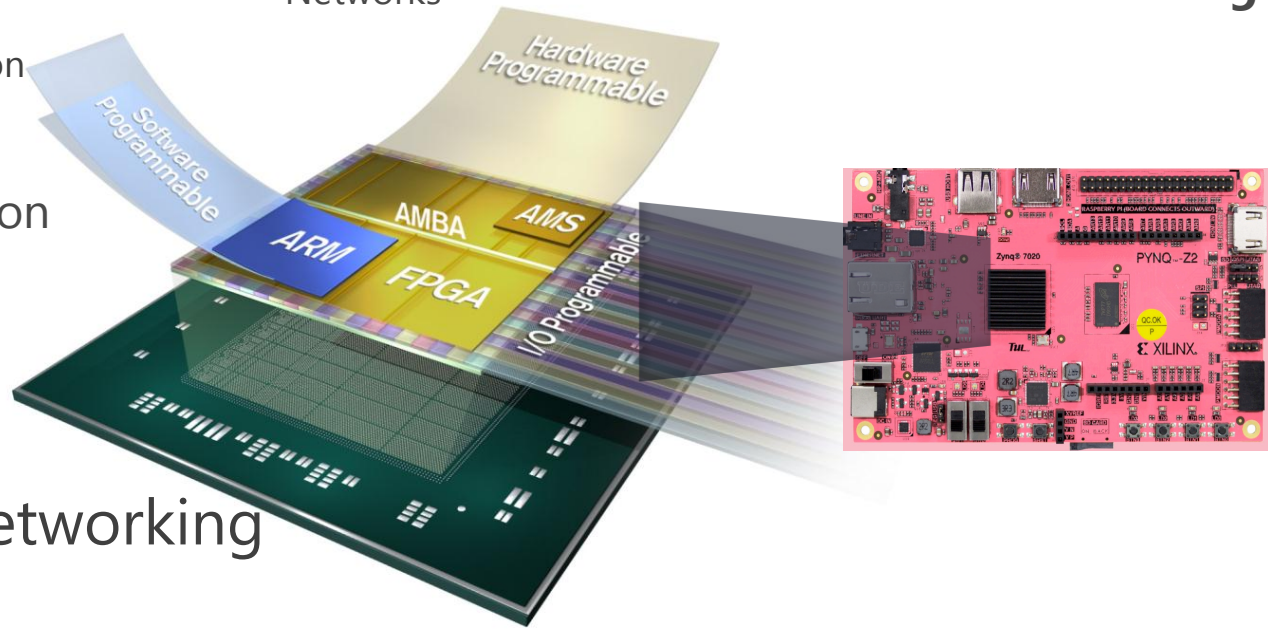
FEC Coding

Hardware Software Co-Design

Encryption

CAD Tools

Dynamically Reconfigurable Systems



High Performance Computing

Networking

Configurable Computing

Digital Signal Processing

Embedded Systems

Robotics

Ideal for Teaching and Research

Adaptable.
Intelligent.

赛灵思工业物联网研讨会
XILINX IIoT SEMINAR

XILINX