Inference Architectures @Xilinx

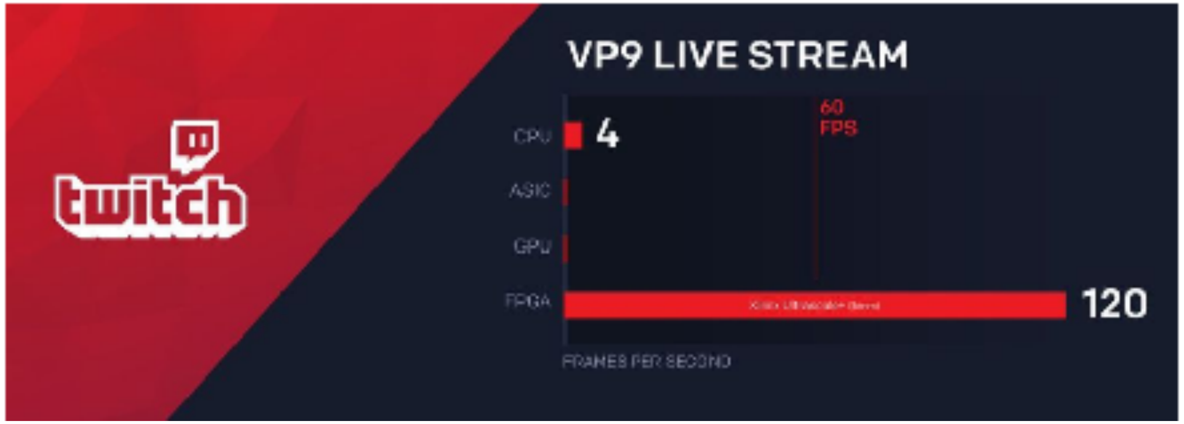
Graham Schelle, PhD
Principal Engineer
Xilinx Research Labs

Xilinx Headlines

When Databases Meet FPGA – Achieving 1 Million TPS with X-DB Heterogeneous Computing

When Databases Meet FPGA – Achieving 1 Million TPS with X-DB Heterogeneous Computing
alibabacloud.com

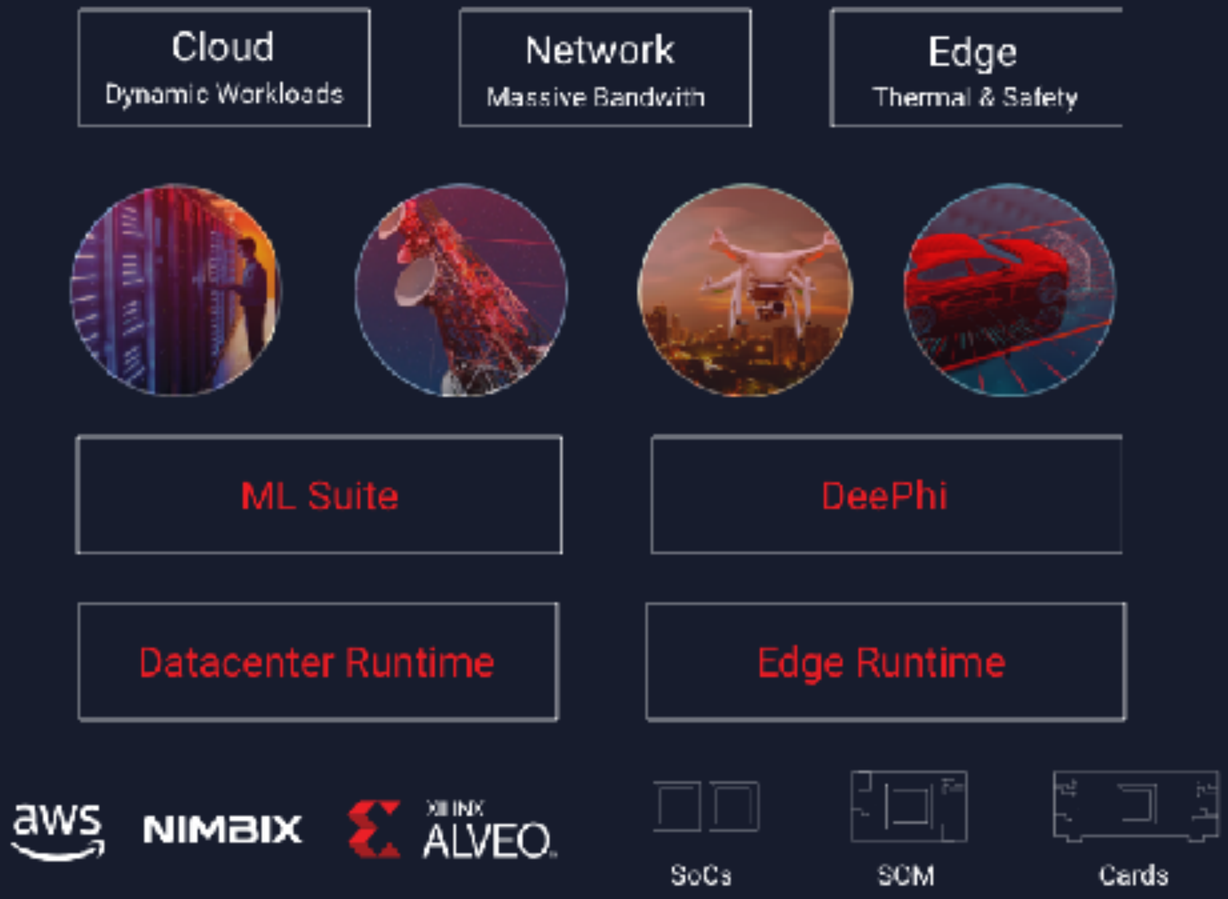
Children's Hospital of Philadelphia And Edico Genome Achieve Fastest-Ever Analysis Of 1,000 Genomes



Twitch Chooses Xilinx to Enable its Broadcast-quality Livestream of eSports

Agenda

- Xilinx Adaptive Architectures
- Inference Architectures
- Open Source



Cloud
Dynamic Workloads

Network
Massive Bandwith

Edge
Thermal & Safety

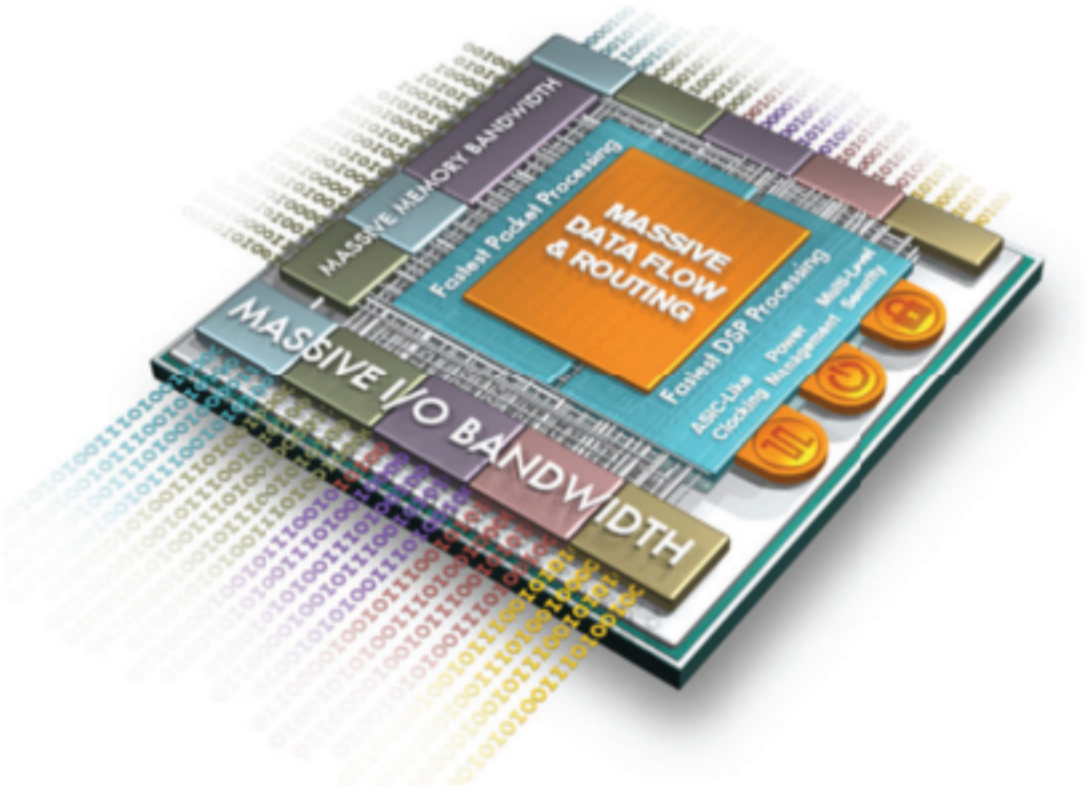
ML Suite

DeePhi

Datacenter Runtime

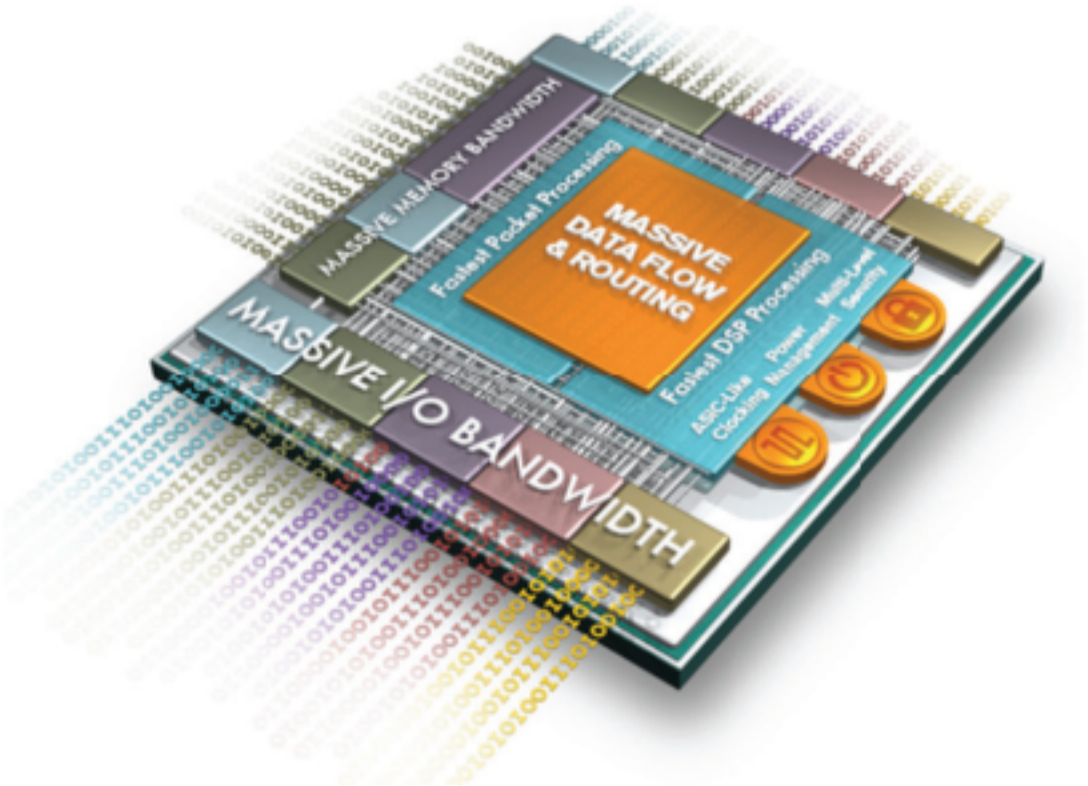
Edge Runtime

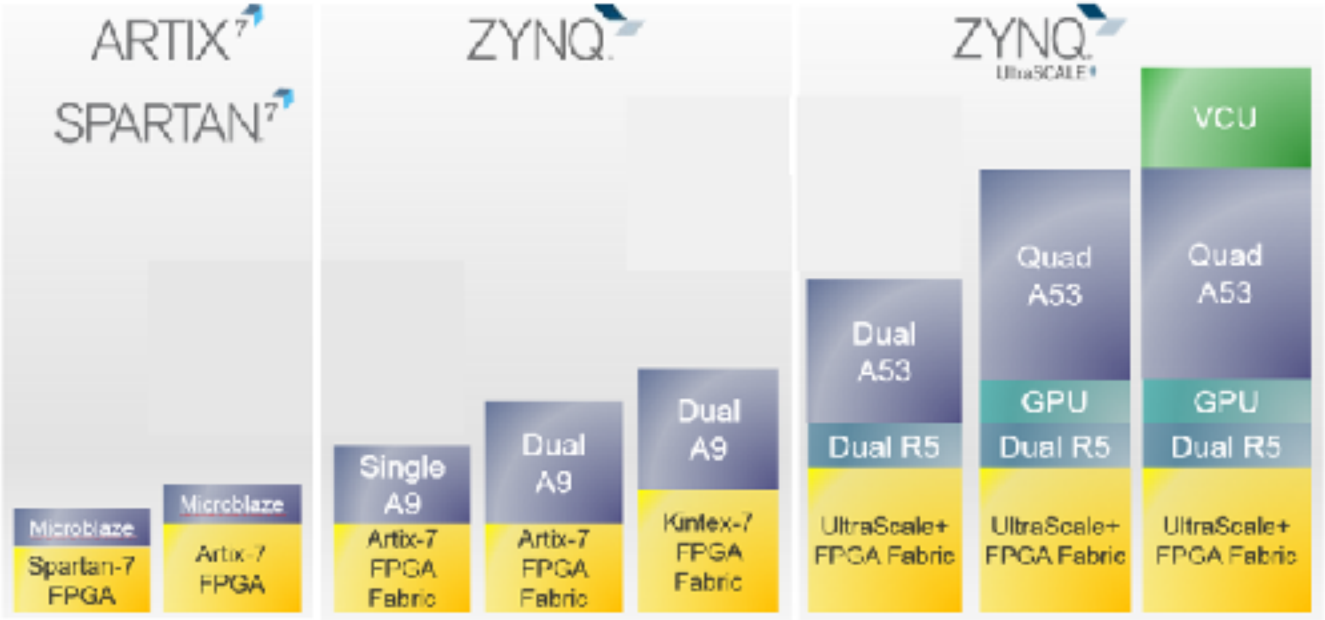
Xilinx Adaptive Architectures



Traditionally, FPGAs for massively data-parallel applications

Xilinx Adaptive Architectures



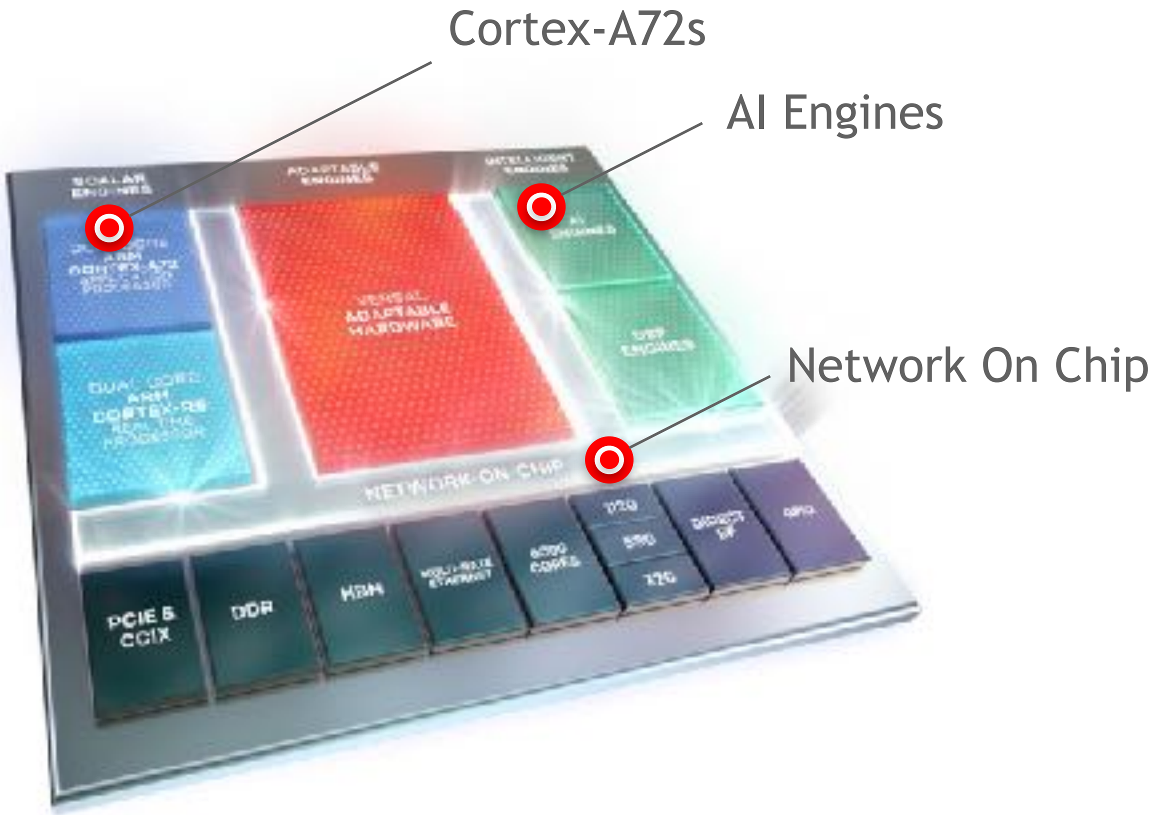


Traditionally, FPGAs for massively data-parallel applications

In 2011, Zynq introduced (ZU+ in 2015)
ARM CPUs added for embedded applications

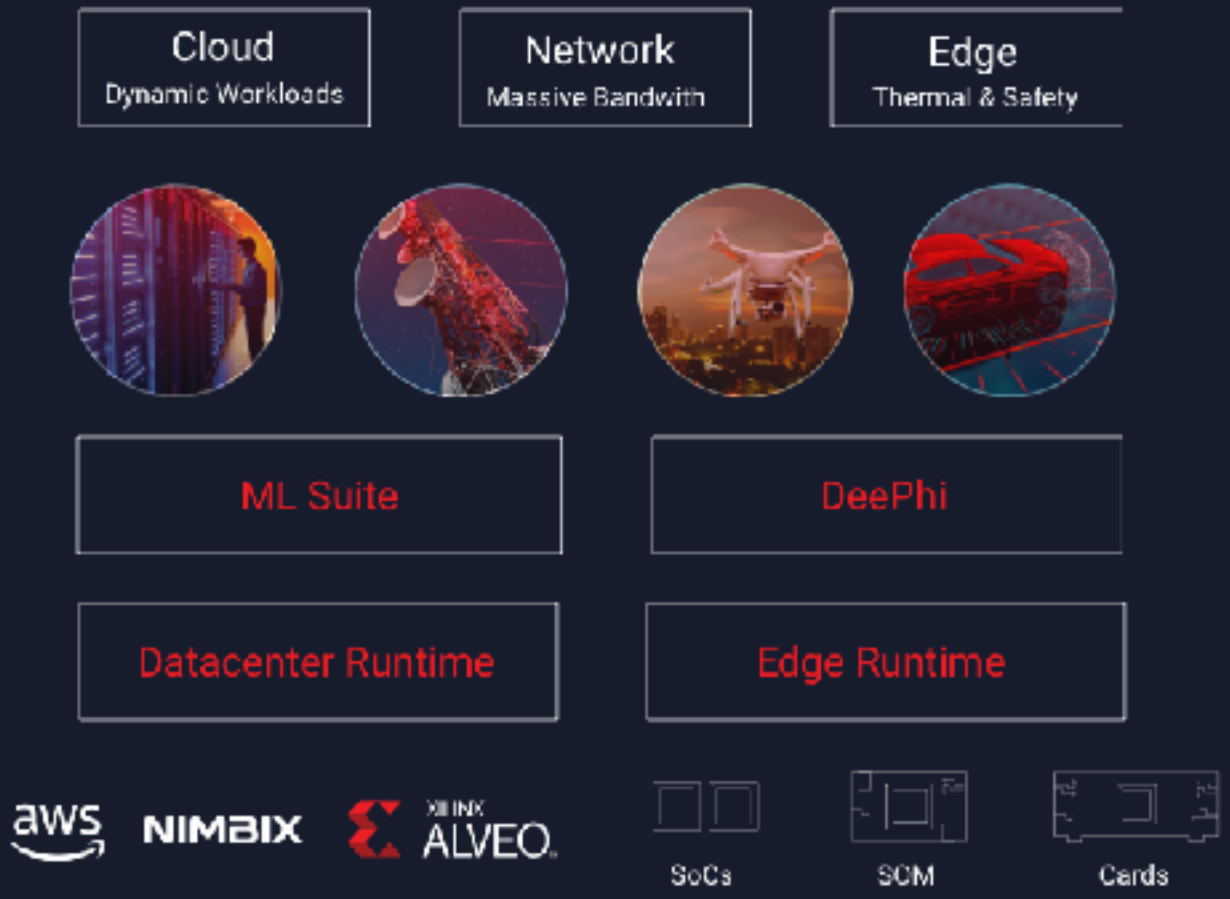
Xilinx Adaptive Architectures – Alveo & Versal

In 2018, Alveo introduced
Accelerator cards for data center workloads



Coming in 2019, Versal Platform
Adaptive compute acceleration platform (ACAP)

Inference Architectures



Cloud
Dynamic Workloads

Network
Massive Bandwith

Edge
Thermal & Safety

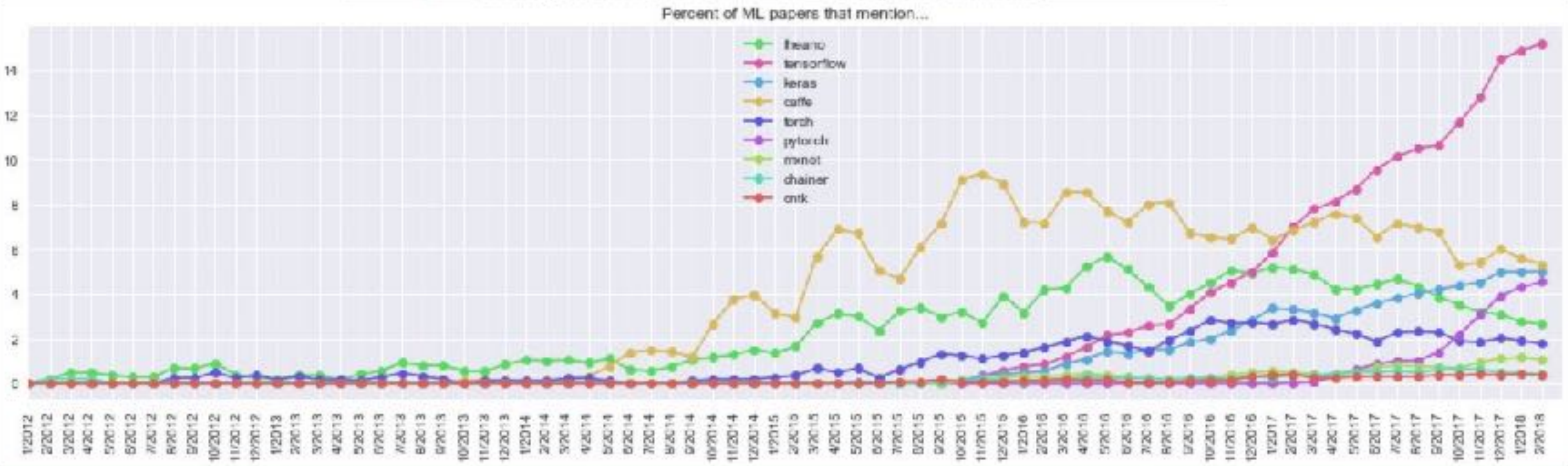
ML Suite

DeePhi

Datacenter Runtime

Edge Runtime

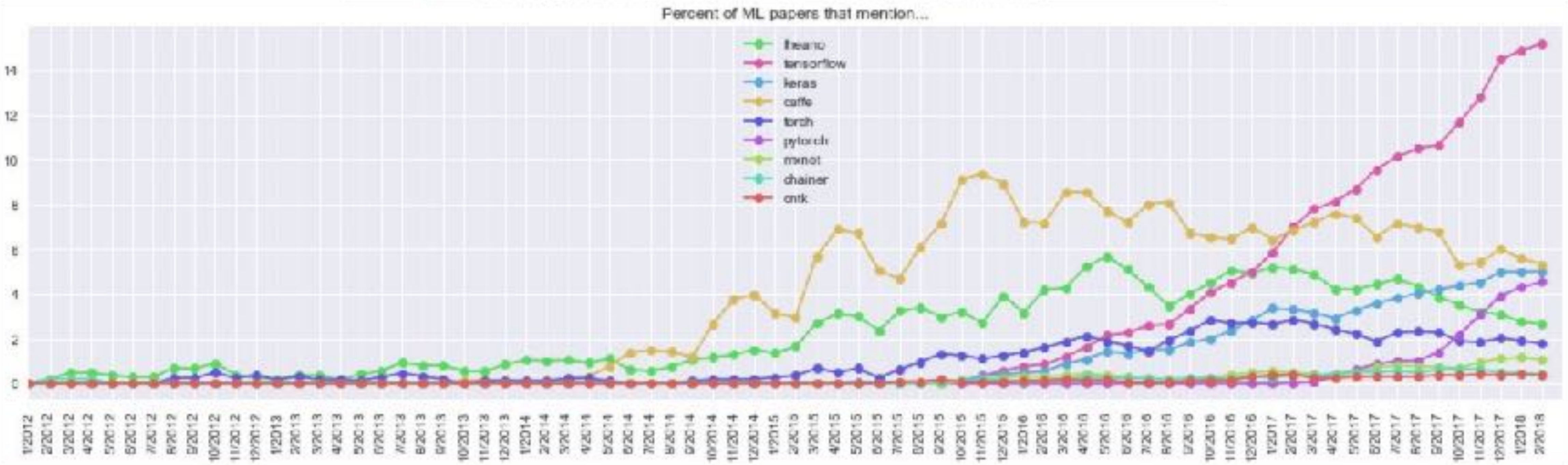
Inference Architectures – Evolving Frameworks



Andrej Karpathy on Twitter

- **Increasing, Evolving Workloads**
  - New acceleration needs & algorithms
  - ML “infused” in many applications
  - Adaptable HW a key benefit

Inference Architectures – Evolving Frameworks



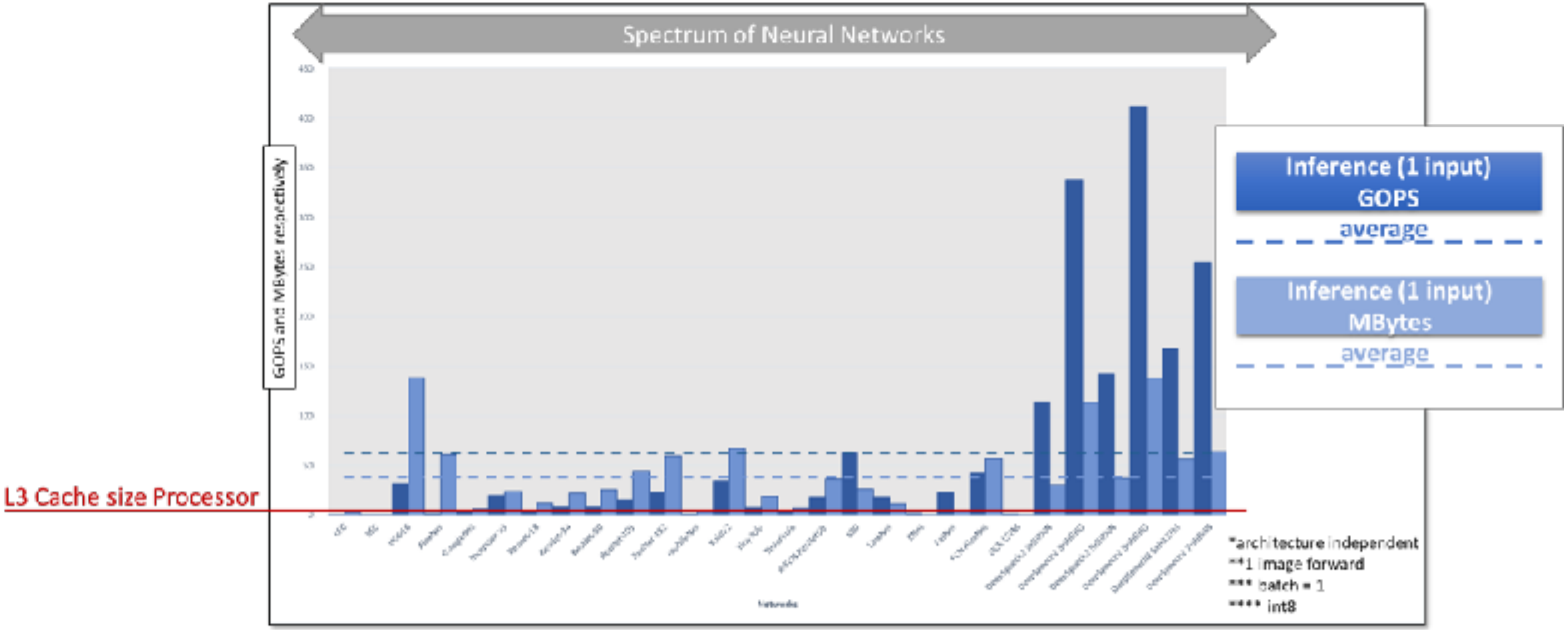
Andrej Karpathy on Twitter

- **Increasing, Evolving Workloads**
  - New acceleration needs & algorithms
  - ML “infused” in many applications
  - Adaptable HW a key benefit

- **Move to Lower Precision**
  - ML inference moving to INT8 & lower
  - Better Perf/W with similar accuracy
  - Xilinx devices natively support variable precision

- **Compressed Networks**
  - Higher performance with reduced compute / memory needs
  - Pruning & load balancing to match network requirements

Inference Architectures – Evolving Workloads



*architecture independent
**1 image forward
*** batch = 1
**** int8

> ## Increasing, Evolving Workloads

- New acceleration needs & algorithms
- ML "infused" in many applications
- Adaptable HW a key benefit

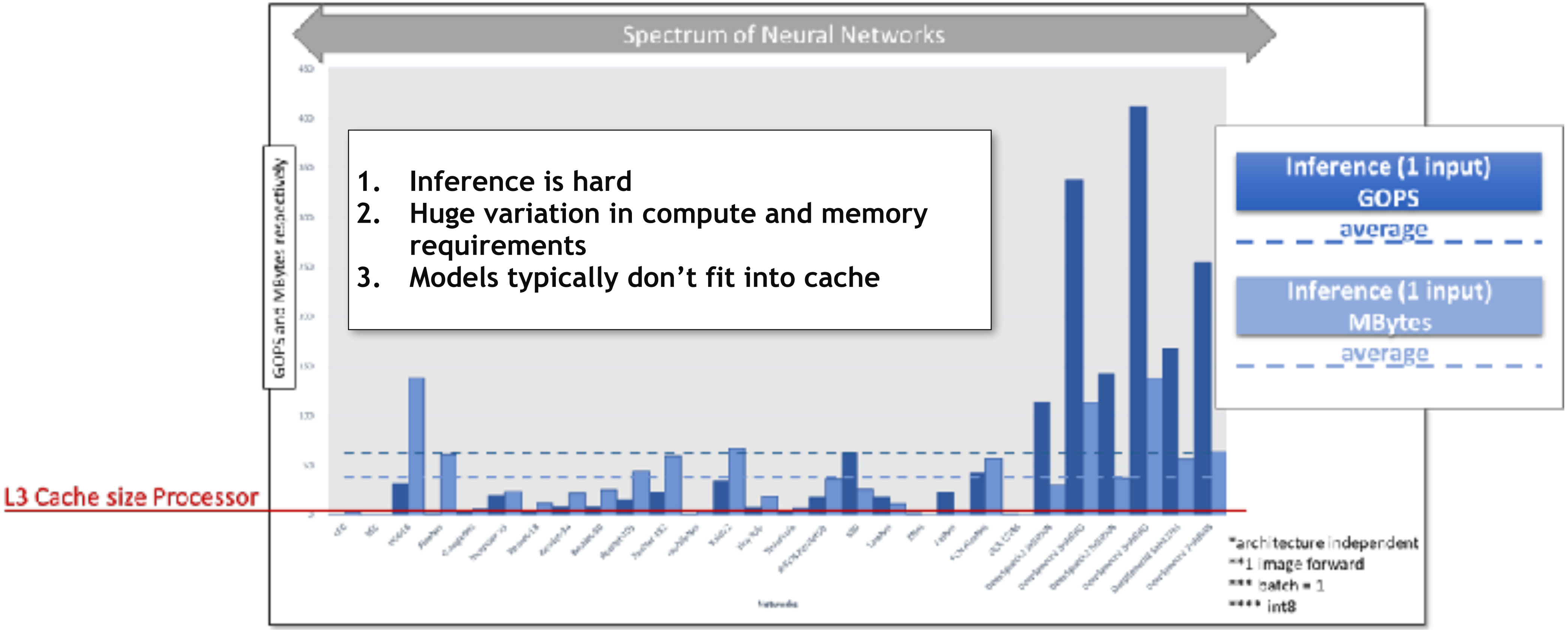
> ## Move to Lower Precision

- ML inference moving to INT8 & lower
- Better Perf/W with similar accuracy
- Xilinx devices natively support variable precision

> ## Compressed Networks

- Higher performance with reduced compute / memory needs
- Pruning & load balancing to match network requirements

Inference Architectures – Evolving Workloads



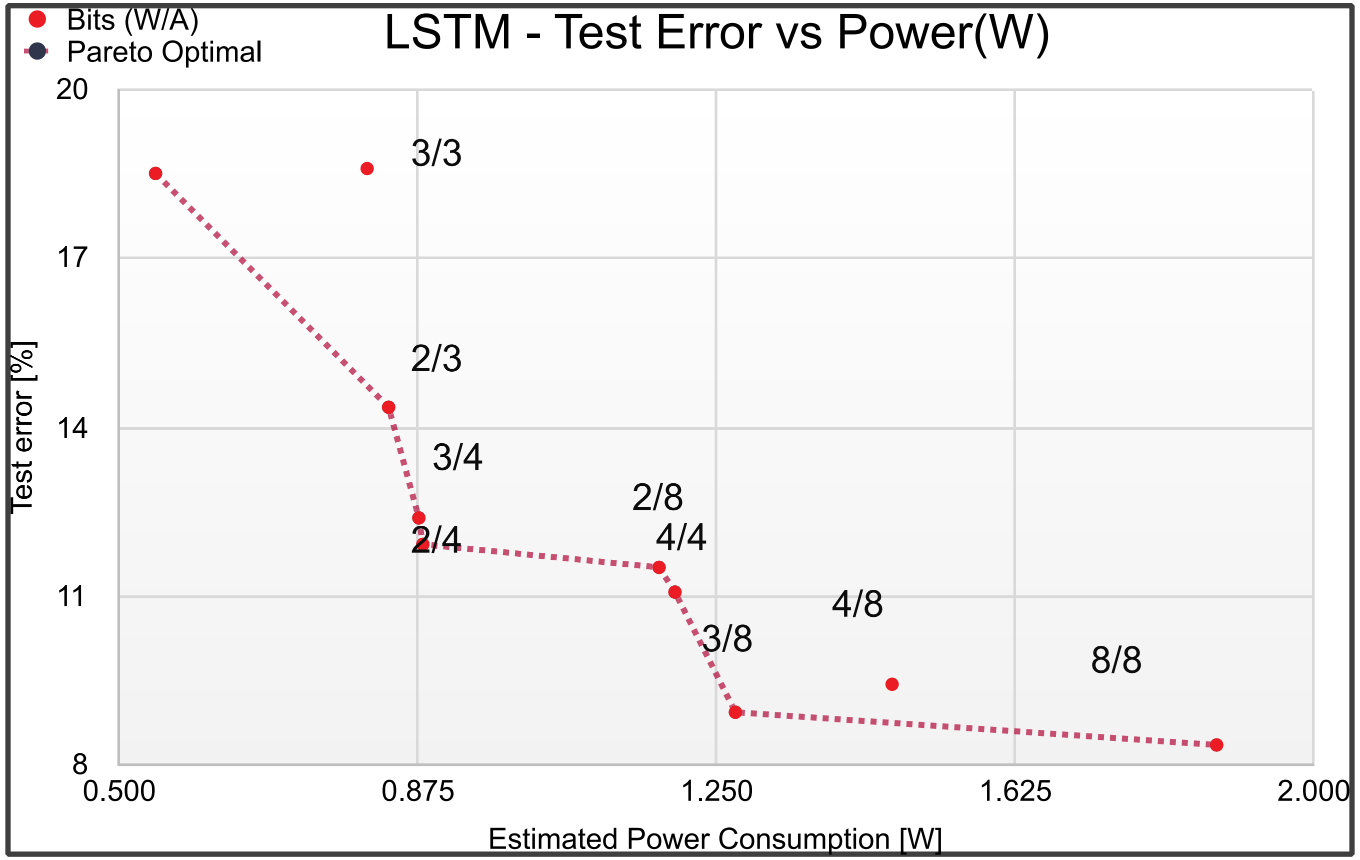
1. Inference is hard
2. Huge variation in compute and memory requirements
3. Models typically don’t fit into cache

*architecture independent
**1 image forward
*** batch = 1
**** int8

- **Increasing, Evolving Workloads**
  - New acceleration needs & algorithms
  - ML “infused” in many applications
  - Adaptable HW a key benefit
- **Move to Lower Precision**
  - ML inference moving to INT8 & lower
  - Better Perf/W with similar accuracy
  - Xilinx devices natively support variable precision
- **Compressed Networks**
  - Higher performance with reduced compute / memory needs
  - Pruning & load balancing to match network requirements

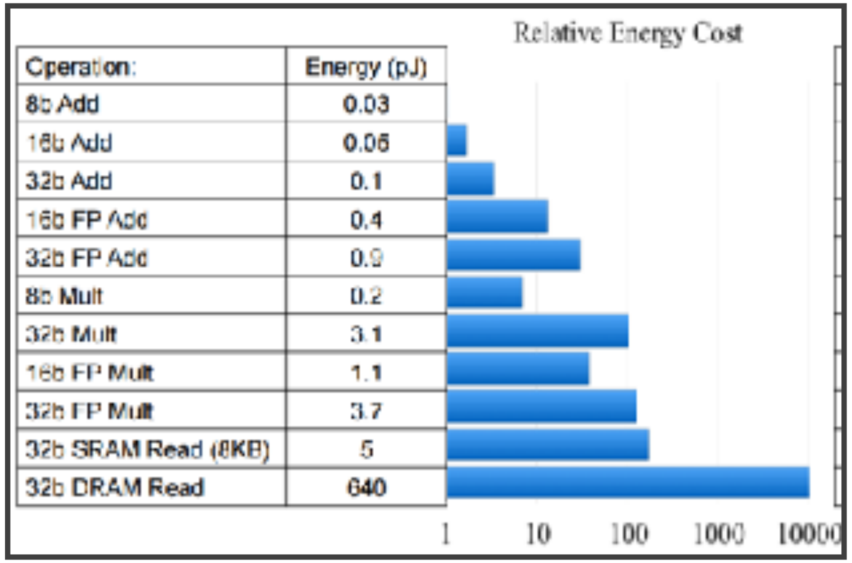
Inference Architectures – Precision vs Power

FPGA:



Target Device ZU7EV ● Ambient temperature: 25 °C ● 12.5% of toggle rate ● 0.5 of Static Probability ● Power reported for PL accelerated block only

ASIC:



Relative Energy Cost
1
10
100
1000
10000

Operation:	Energy (pJ)
8b Add	0.03
16b Add	0.05
32b Add	0.1
16b FP Add	0.4
32b FP Add	0.9
8b Mult	0.2
32b Mult	3.1
16b FP Mult	1.1
32b FP Mult	3.7
32b SRAM Read (8KB)	5
32b DRAM Read	640

Source: Bill Dally (Stanford), Cadence Embedded Neural Network Summit, February 1, 2017

Michaela Blott, Hot Chips 2018 Tutorial, "Overview of Deep Learning and Computer Architectures for Accelerating DNNs"

Rybalkin, V., Pappalardo, A., Ghaffar, M.M., Gambardella, G., Wehn, N. and Blott, M. "FINN-L: Library Extensions and Design Trade-off Analysis for Variable Precision LSTM Networks on FPGAs."

Xilinx Cloud Inference - ML Suite Overlays with xDNN

Adaptable

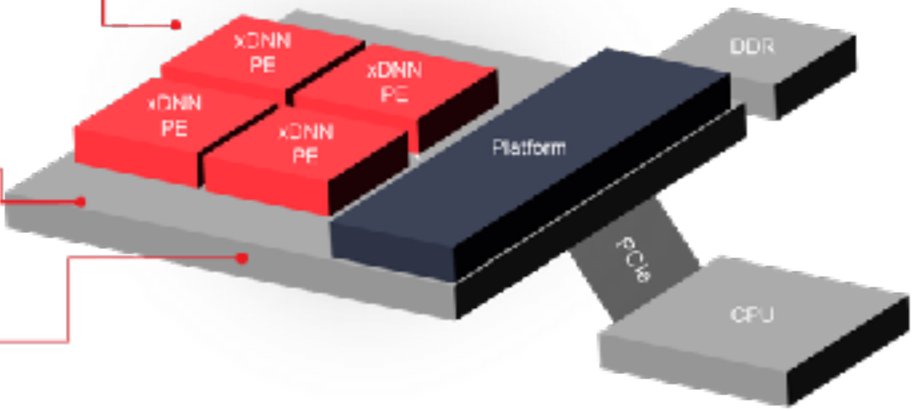
- > AI algorithms are changing rapidly
- > Adjacent acceleration opportunities

Realtime

- > Lower latency than CPU and GPU
- > Data flow processing

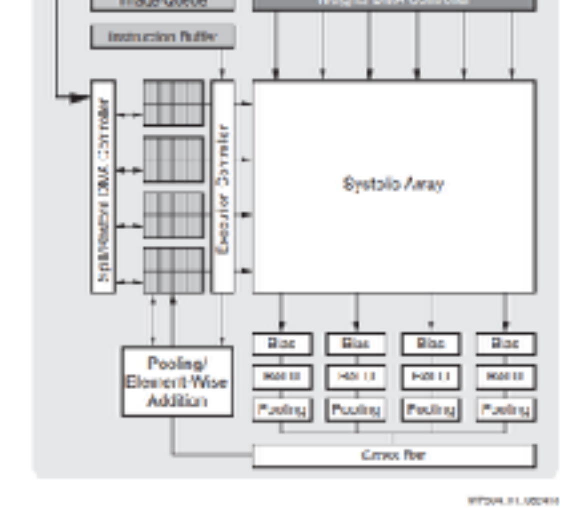
Efficient

- > Performance/watt
- > Low Power



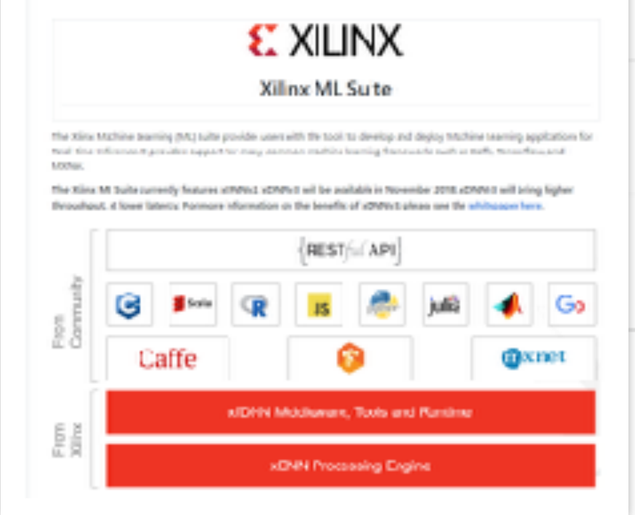
- Built in Programmable Logic
- High Utilization, Thput or Latency Variants
- CPU offload for new layer exploration

xDNN w/ xfDNN Compiler



On-prem and cloud boards

https://github.com/Xilinx/ml-suite



XILINX

Xilinx ML Suite

The Xilinx Machine Learning (ML) Suite provides users with the tools to develop and deploy Machine Learning applications for Real-time Inference. It provides support for many common machine learning frameworks such as Caffe, Tensorflow, and MXNet.

The Xilinx ML Suite currently features xDNNv2. xDNNv3 will be available in November 2018. xDNNv3 will bring higher throughput & lower latency. For more information on the benefits of xDNNv3 please see the whitepaper here.

Xilinx Edge Inference - DeePhi

(2013)

(2016)

“Learning both Weights and Connections for Efficient Neural Networks”, NeurIPS 2015

“EIE: Efficient Inference Engine on Compressed Deep Neural Network”, ISCA 2016

“ESE: Efficient Speech Recognition Engine with Compressed LSTM on FPGA”, FPGA 2017

Xilinx Edge Inference - DeePhi

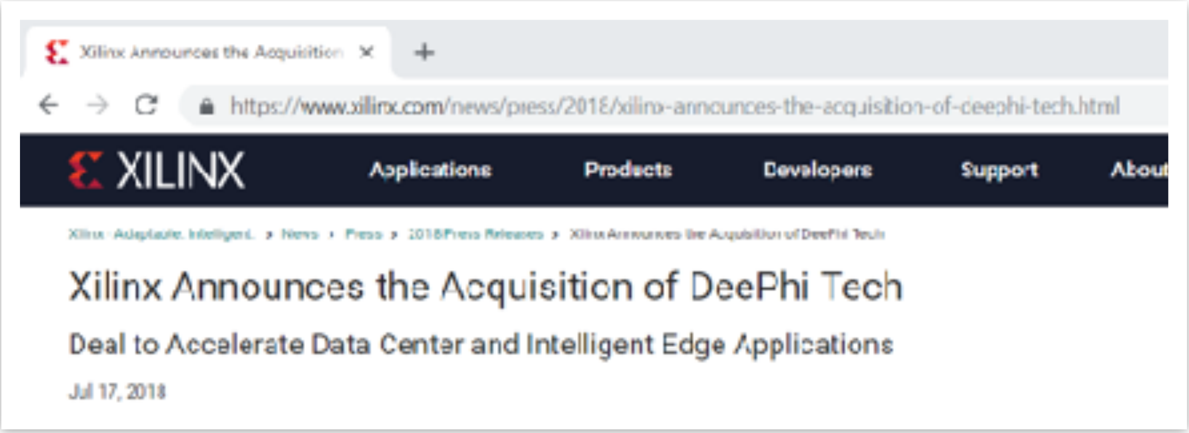
(2013)

(2016)

“Learning both Weights and Connections for Efficient Neural Networks”, NeurIPS 2015

“EIE: Efficient Inference Engine on Compressed Deep Neural Network”, ISCA 2016

“ESE: Efficient Speech Recognition Engine with Compressed LSTM on FPGA”, FPGA 2017



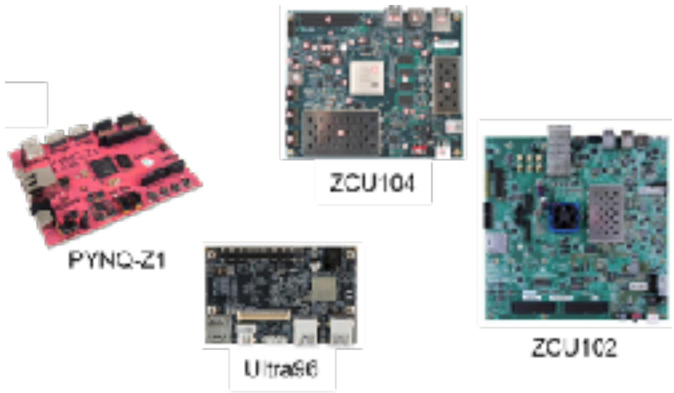
Xilinx Announces the Acquisition
https://www.xilinx.com/news/press/2018/xilinx-announces-the-acquisition-of-deephi-tech.html
XILINX
Applications
Products
Developers
Support
About

Xilinx: Adaptable. Intelligent. › News › Press › 2018 Press Releases › Xilinx Announces the Acquisition of DeePhi Tech

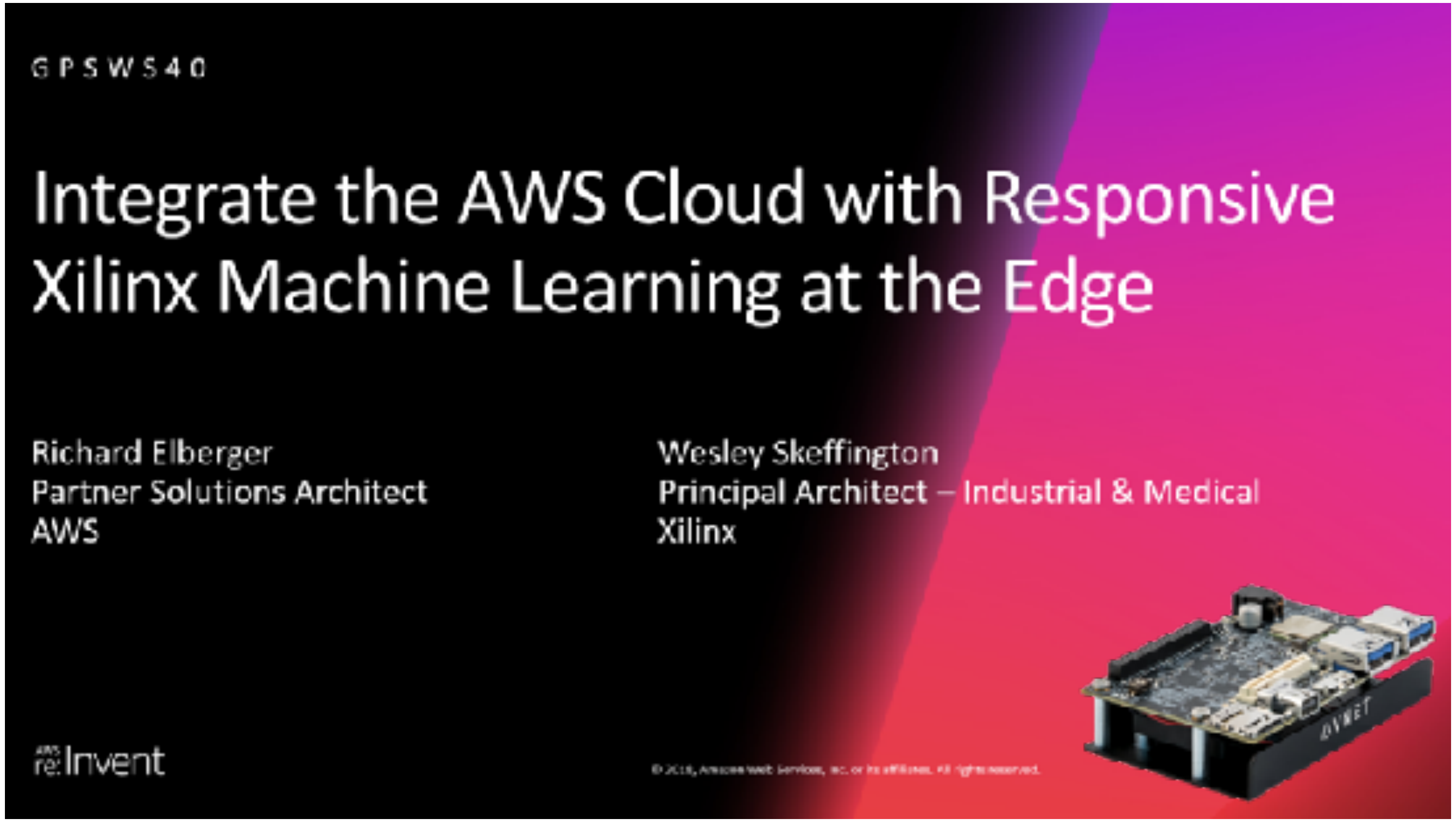
Xilinx Announces the Acquisition of DeePhi Tech

Deal to Accelerate Data Center and Intelligent Edge Applications

Jul 17, 2018



Cloud & Edge Integration



GPSWS40

Integrate the AWS Cloud with Responsive Xilinx Machine Learning at the Edge

Richard Elberger
Partner Solutions Architect
AWS

Wesley Skeffington
Principal Architect – Industrial & Medical
Xilinx

re:Invent
AVNET

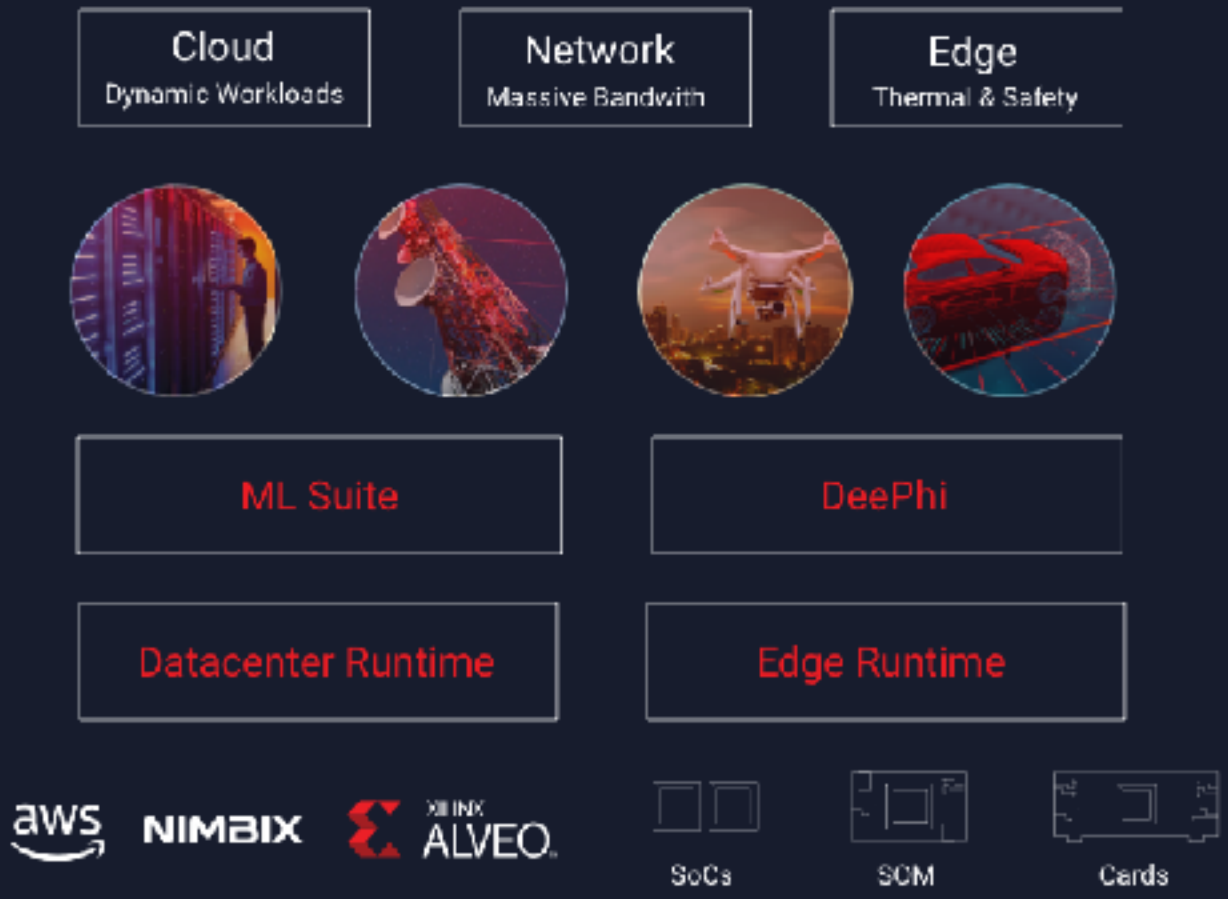
© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

David Hallgrove • 2nd
Manager Cloud Consulting at Deloitte
4d • Edited

Great workshop on 'SageMaker Machine Learning at the Edge with Xilinx and AWS', on a MicroZed and Ultra96

...see more

Xilinx and Open Source



Cloud
Dynamic Workloads

Network
Massive Bandwith

Edge
Thermal & Safety

ML Suite

DeePhi

Datacenter Runtime

Edge Runtime

Xilinx and Open Source

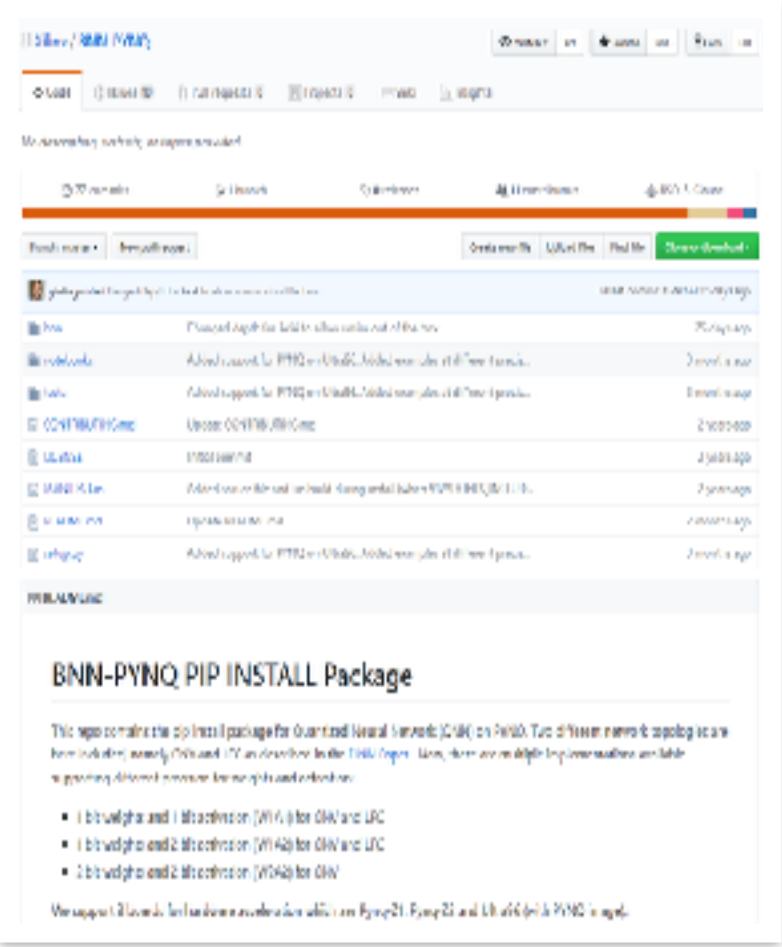
PYNQ

Xilinx / PYNQ

PYNQ

PYNQ is an open-source project from Xilinx that makes it easy to design embedded systems with Zynq All Programmable Systems on Chips (APSoCs). Using the Python language and libraries, designers can exploit the benefits of programmable logic and microprocessors in Zynq to build more capable and exciting embedded systems. PYNQ users can now create high performance embedded applications with

Quantized Neural Networks



Xilinx / BNN-PYNQ

README.md

BNN-PYNQ PIP INSTALL Package

This repo contains the pip install package for Quantized Neural Network (QNN) on PYNQ. Two different network topologies are here included, namely CNV and LFC as described in the FINN Paper. Then, there are multiple implementations available supporting different precision for weights and activation:

- 1 bit weights and 1 bit activation (W1A1) for CNV and LFC
- 1 bit weights and 2 bit activation (W1A2) for CNV and LFC
- 2 bit weights and 2 bit activation (W2A2) for CNV

We support 3 boards for hardware acceleration which are Pynq-Z1, Pynq-Z2 and Ultra96 (with PYNQ image).

Xilinx Runtime for PCIe Attached FPGAs

Xilinx / XRT

README.md

Xilinx Runtime

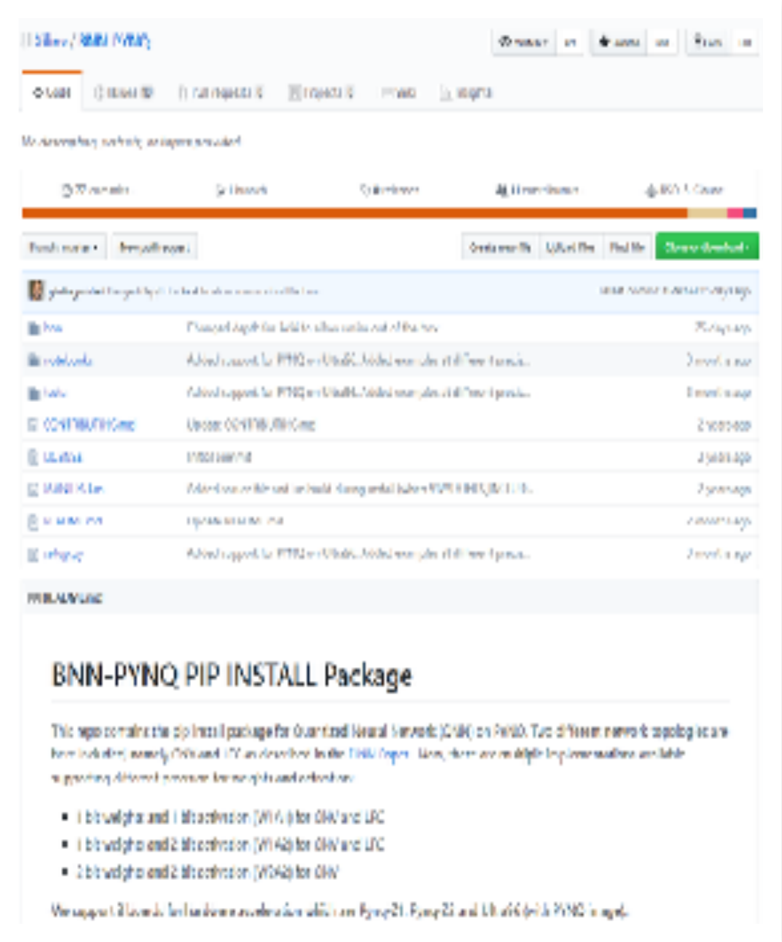
…More on www.github.com/Xilinx

Xilinx and Open Source

PYNQ

PYNQ

Quantized Neural Networks



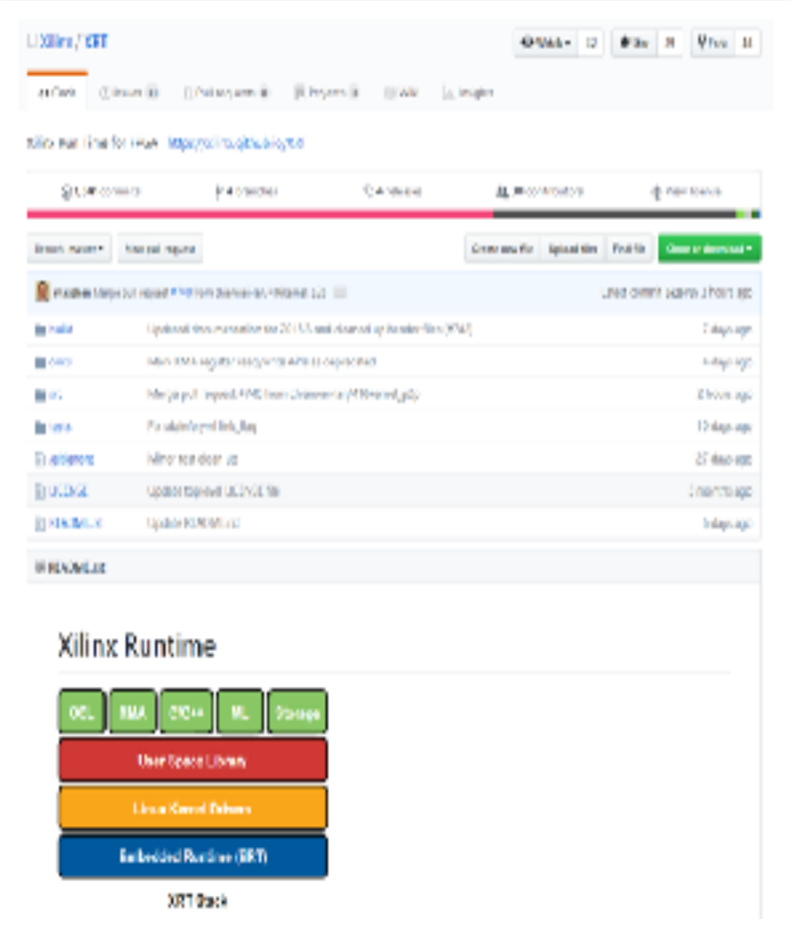
BNN-PYNQ PIP INSTALL Package

This repo contains the pip install package for Quantized Neural Network (QNN) on PYNQ. Two different network topologies are here included, namely CNV and LFC as described in the FINN Paper. Then, there are multiple implementations available supporting different precision for weights and activation:

- 1 bit weights and 1 bit activation (W1A1) for CNV and LFC
- 1 bit weights and 2 bit activation (W1A2) for CNV and LFC
- 2 bit weights and 2 bit activation (W2A2) for CNV

We support 3 boards for hardware acceleration which are Pynq-Z1, Pynq-Z2 and Ultra96 (with PYNQ image).

Xilinx Runtime for PCIe Attached FPGAs



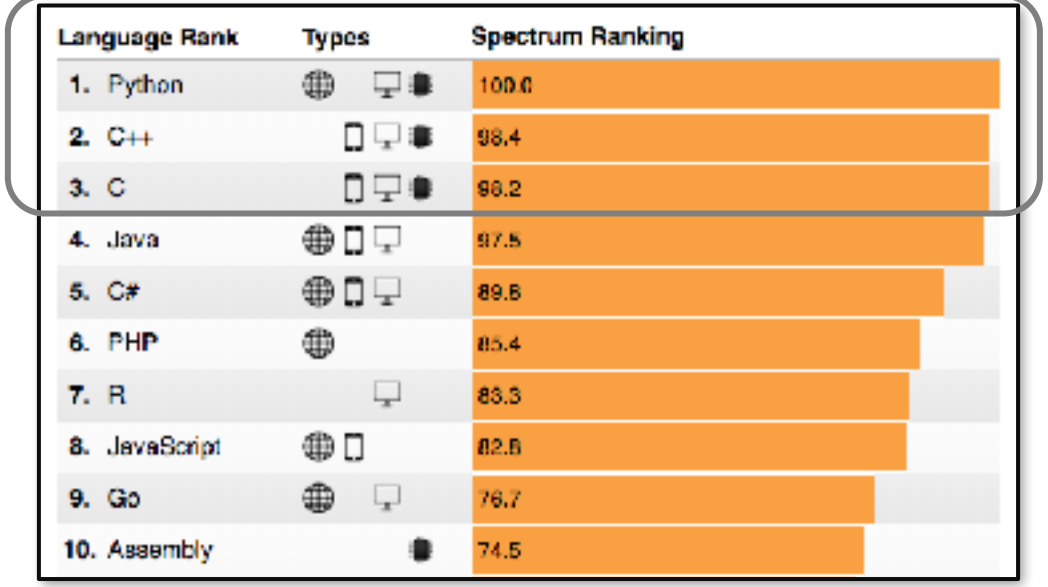
Xilinx Runtime

XRT Stack

…More on www.github.com/Xilinx

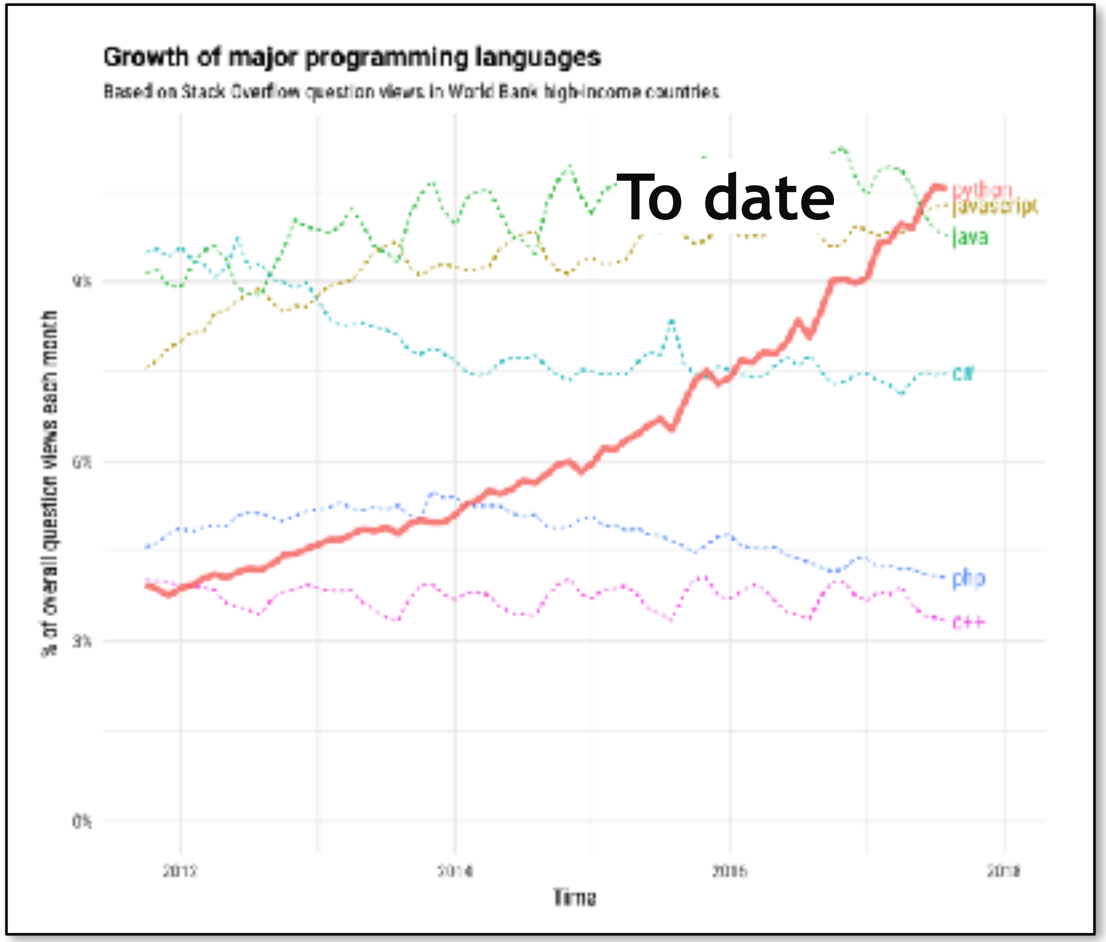
Python is increasingly the Language of Choice

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

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

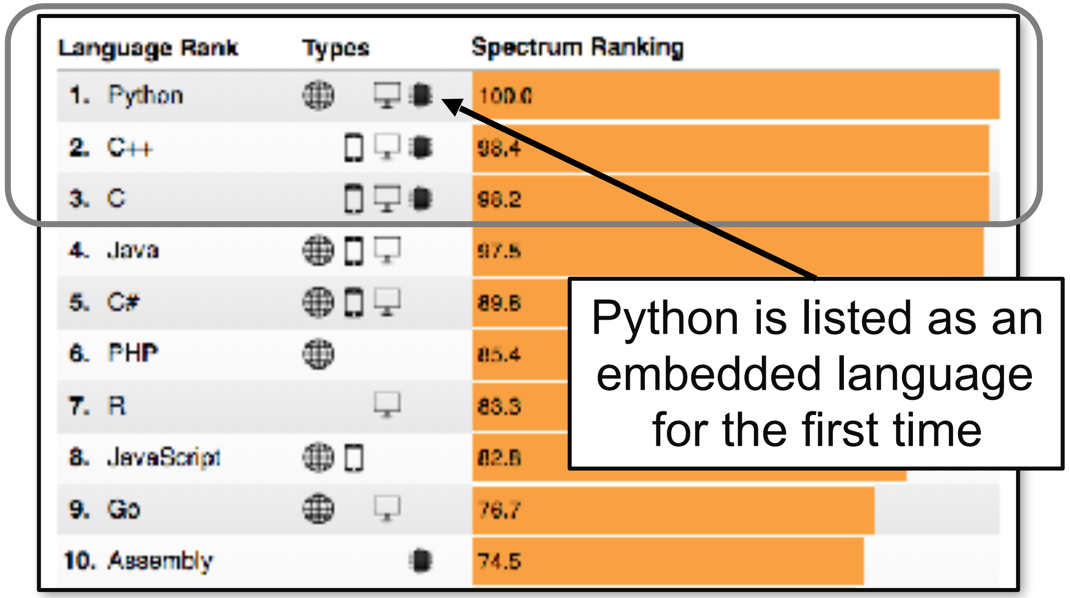


To date

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

Python is increasingly the Language of Choice

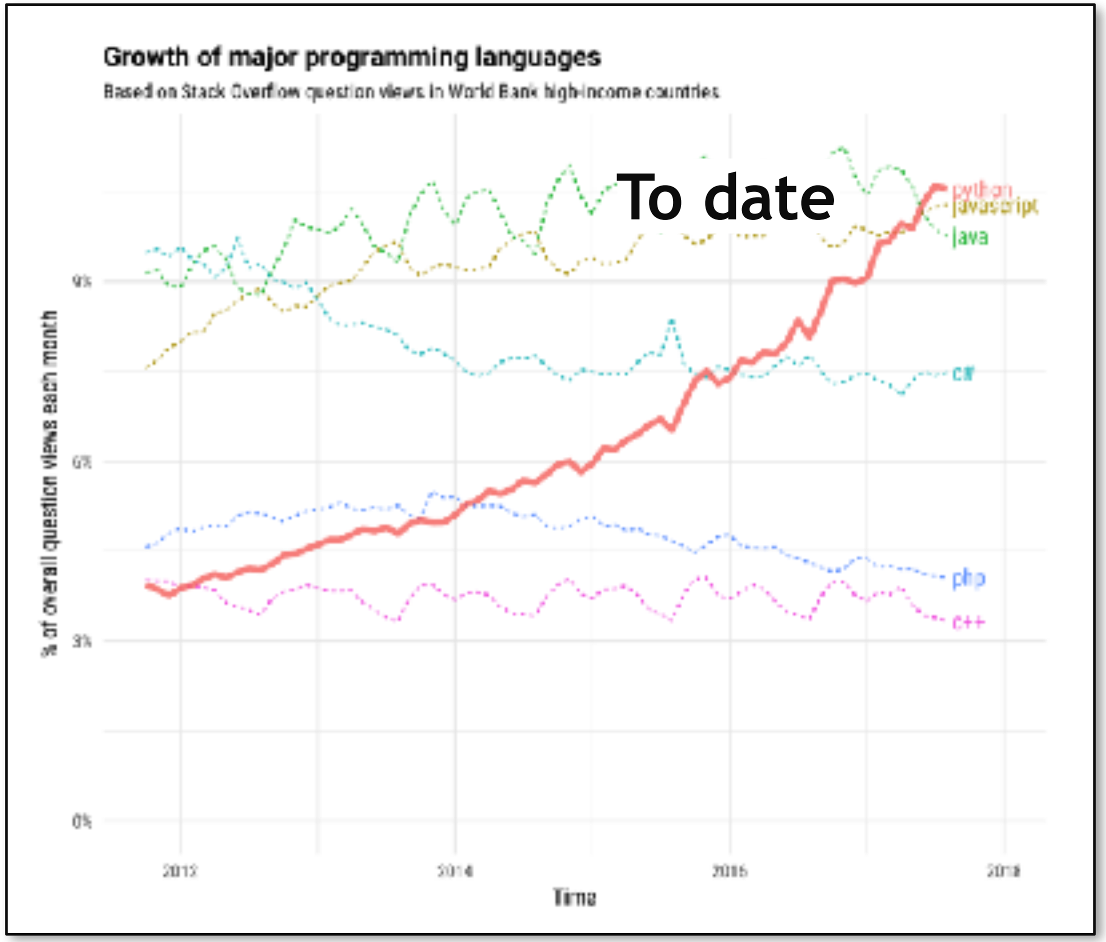
Top Programming Languages, IEEE Spectrum, July’18



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

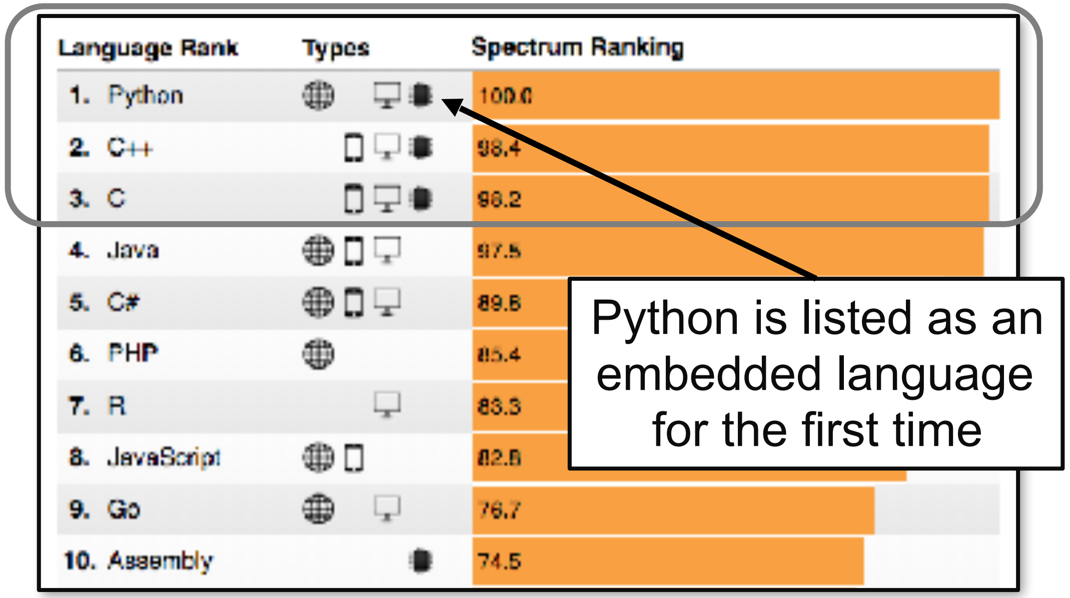


To date

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

Python is increasingly the Language of Choice

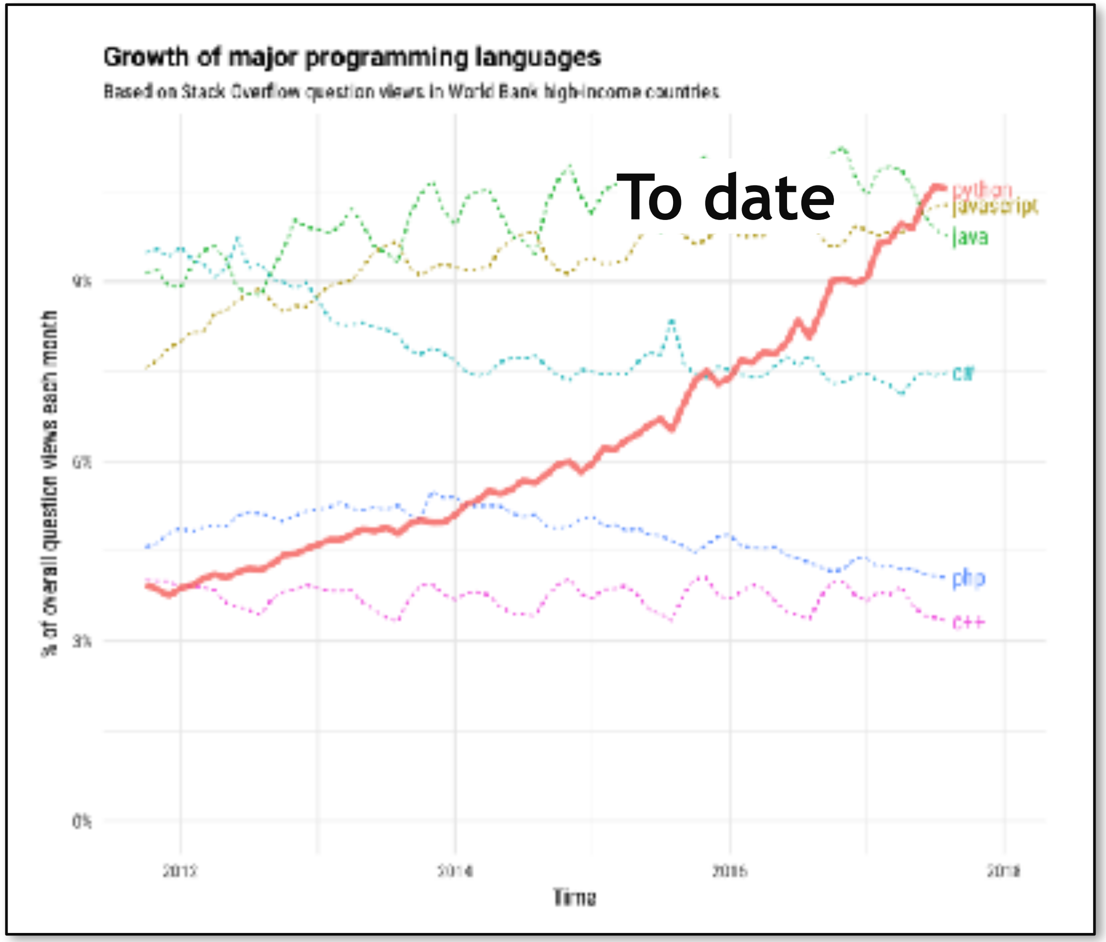
Top Programming Languages, IEEE Spectrum, July'18



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

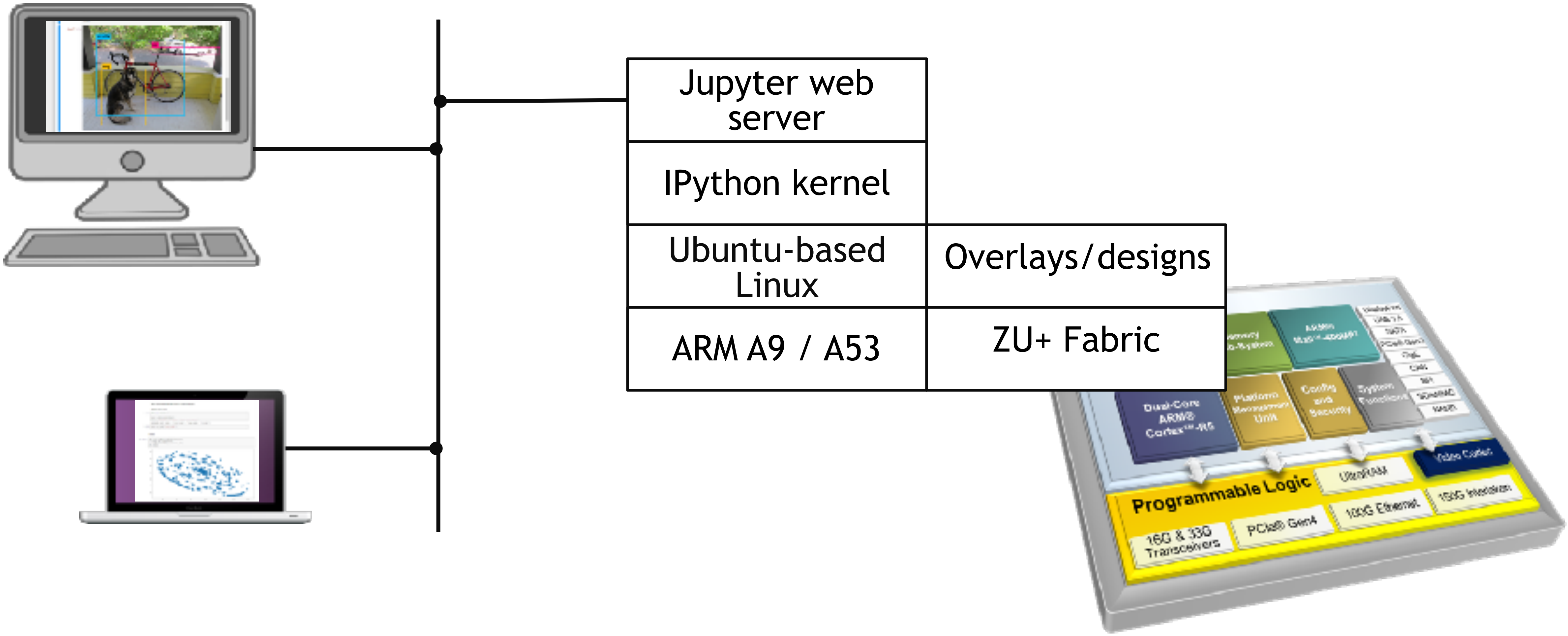


To date

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

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

PYNQ Python Productivity for Zynq



Jupyter web server

IPython kernel

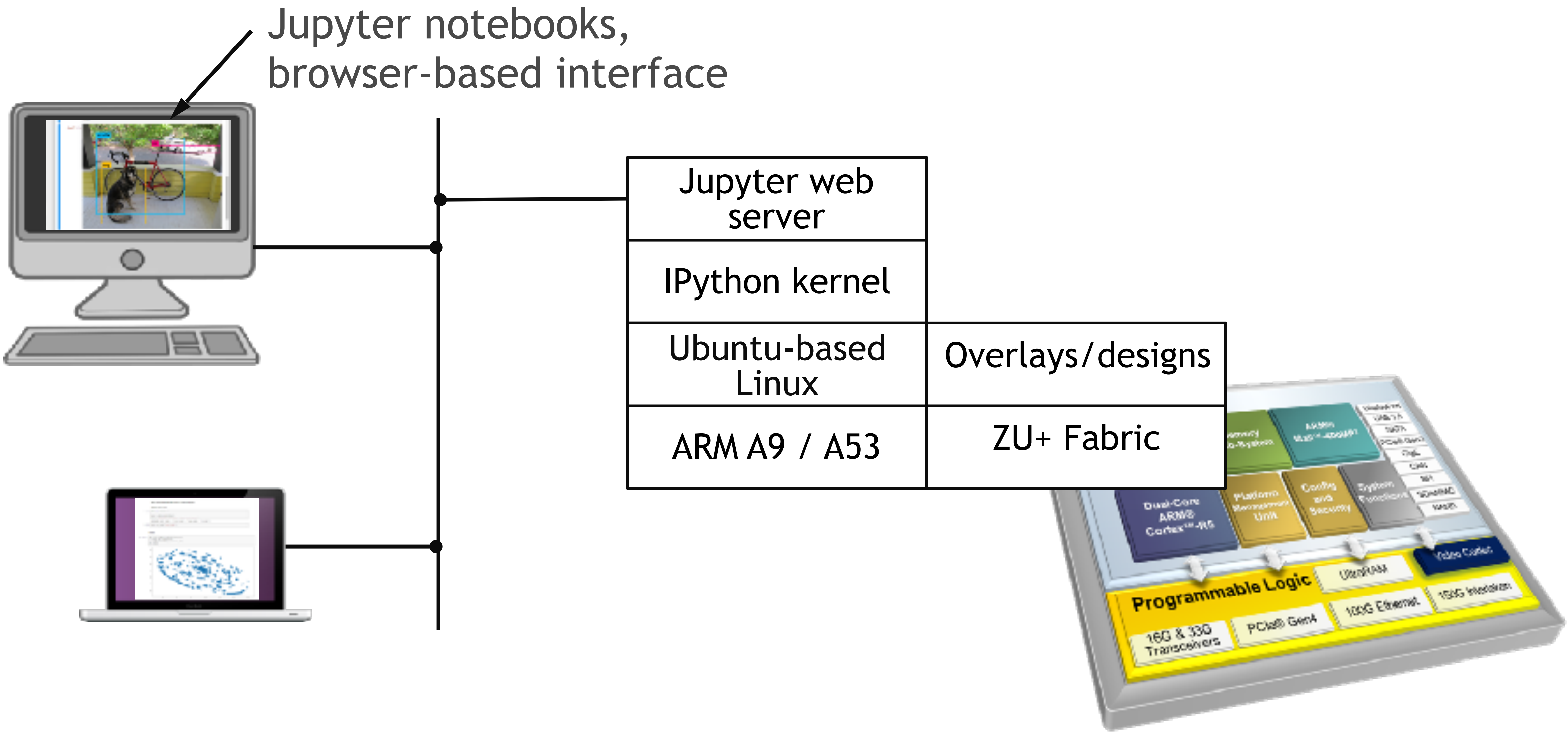
Ubuntu-based Linux

Overlays/designs

ARM A9 / A53

ZU+ Fabric

PYNQ Python Productivity for Zynq



Jupyter notebooks, browser-based interface

Jupyter web server

IPython kernel

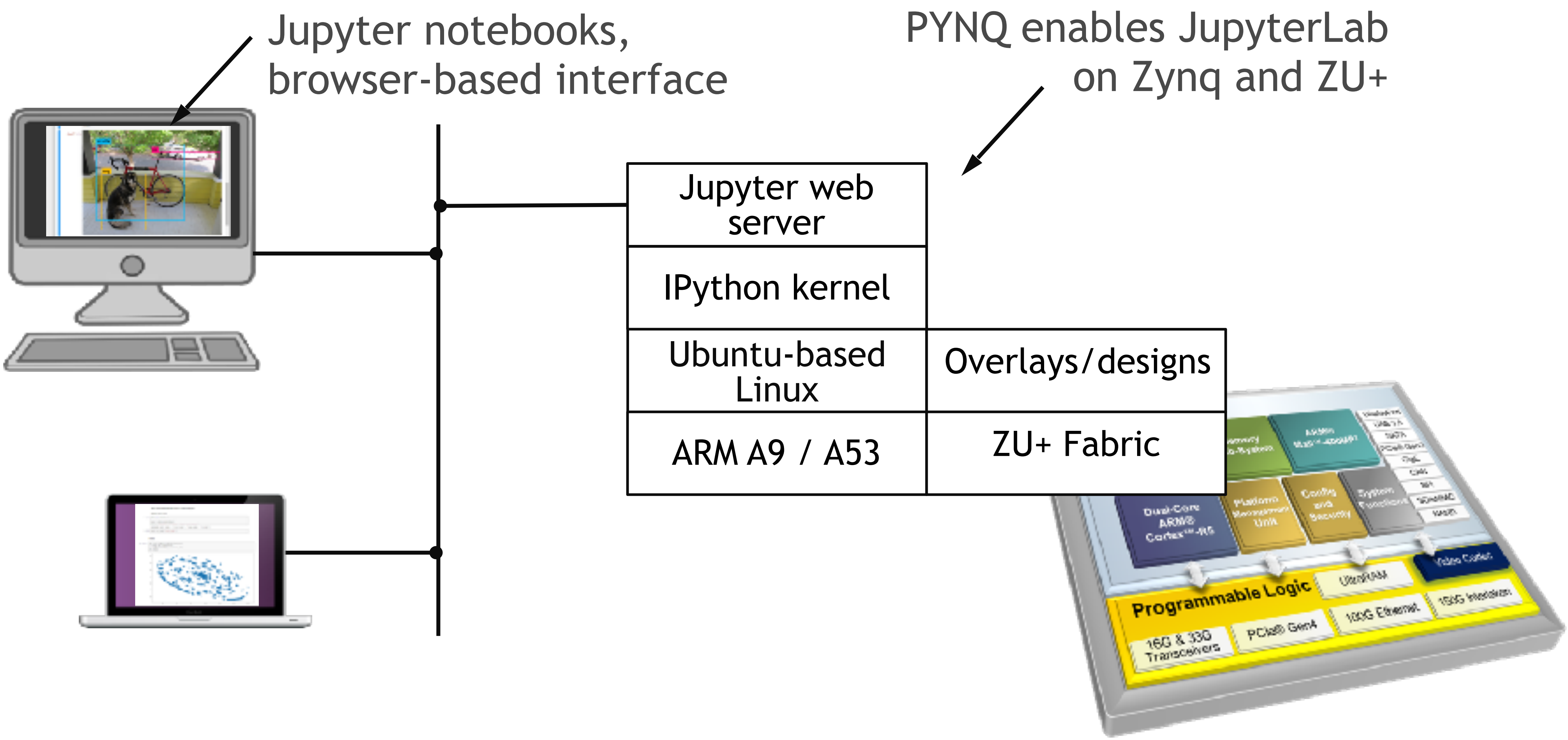
Ubuntu-based Linux

ARM A9 / A53

Overlays/designs

ZU+ Fabric

PYNQ Python Productivity for Zynq



Jupyter notebooks, browser-based interface

PYNQ enables JupyterLab on Zynq and ZU+

Jupyter web server

IPython kernel

Ubuntu-based Linux

ARM A9 / A53

Overlays/designs

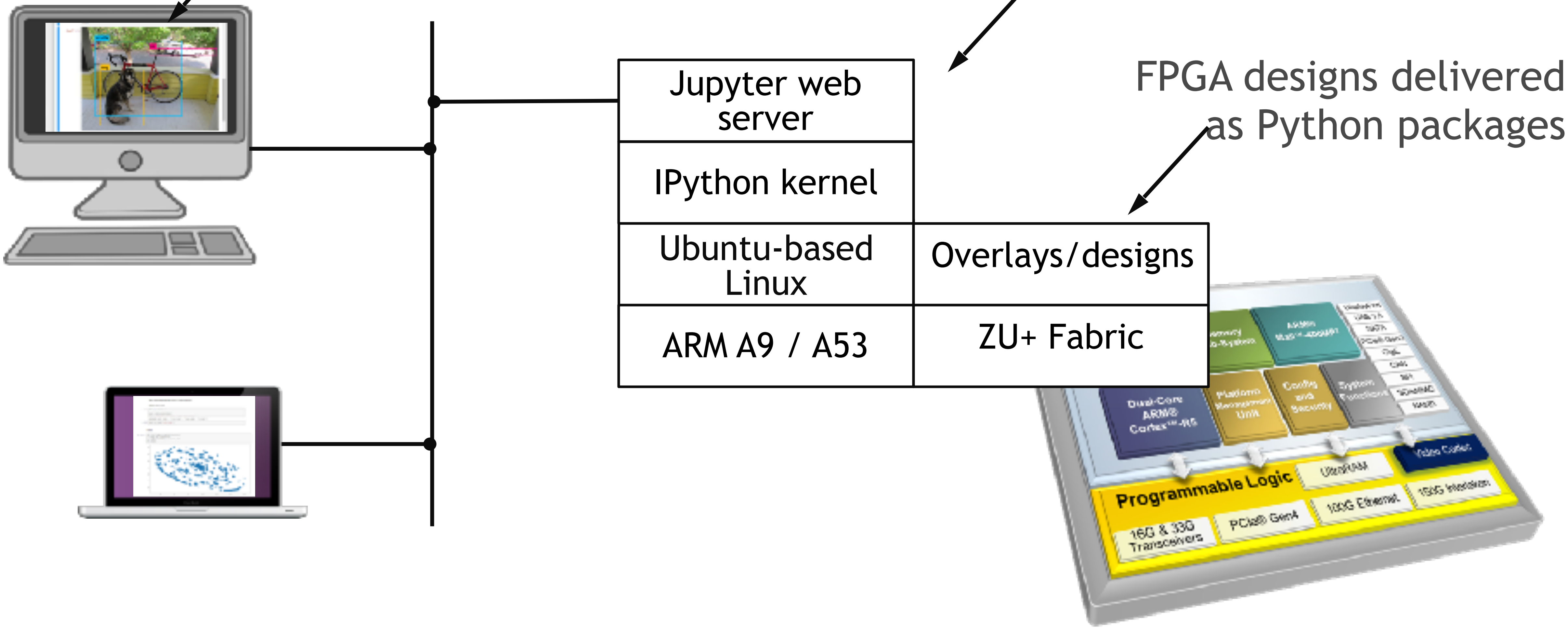
ZU+ Fabric

Python Productivity for Zynq

Jupyter notebooks, browser-based interface

PYNQ enables JupyterLab on Zynq and ZU+

FPGA designs delivered as Python packages



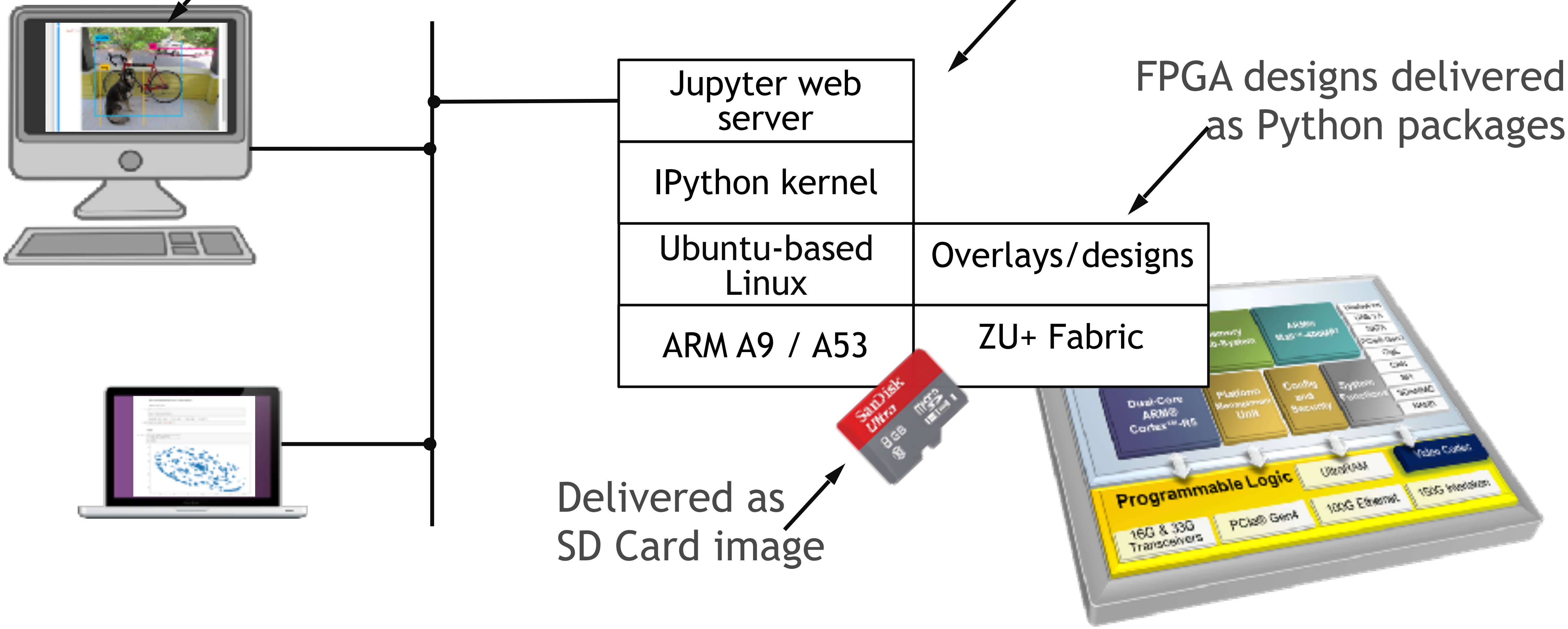
Jupyter web server	
IPython kernel	
Ubuntu-based Linux	Overlays/designs
ARM A9 / A53	ZU+ Fabric

PYNQ Python Productivity for Zynq

Jupyter notebooks, browser-based interface

PYNQ enables JupyterLab on Zynq and ZU+

FPGA designs delivered as Python packages



Jupyter web server	
IPython kernel	
Ubuntu-based Linux	Overlays/designs
ARM A9 / A53	ZU+ Fabric

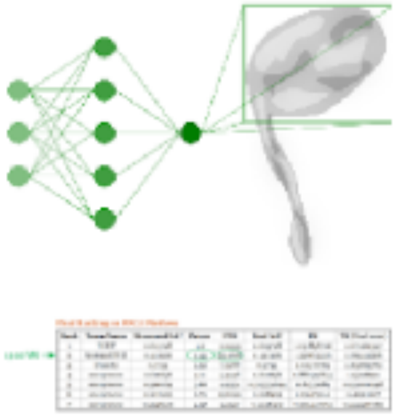
Delivered as SD Card image

PYNQ Community – ML, Non-ML & Academic Partners

Extended Kalman filter
University Sydney

K
Est. MEAN
Est. COVAR.

spoonNN
ETH Zurich
FPGA based neural network inference project

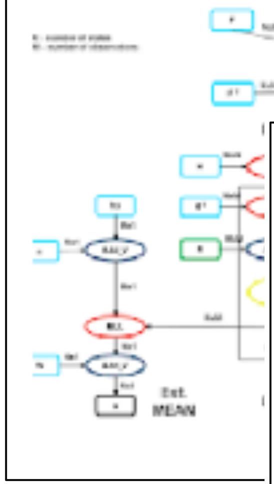


iSmart DNN
FPGA-based neural network inference for DAC 2018 contest

TGIIF
1st place in the DAC 2018 design contest for neural network object detection

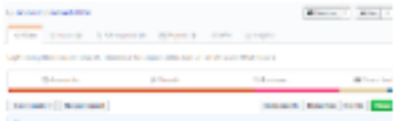
PYNQ Community – ML, Non-ML & Academic Partners

Extended Kalman filter
University Sydney



spoonNN
ETH Zurich
FPGA based neural network inference project

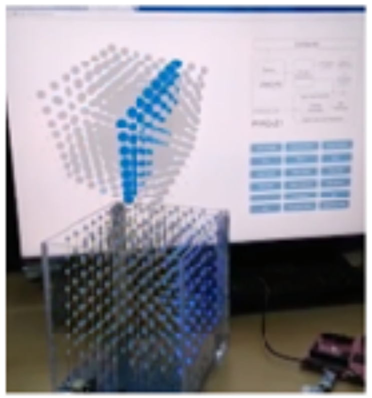
iSmart DNN
FPGA-based neural network inference for DAC 2018 contest



TGIIF
1st place in the DAC 2018 design contest for neural network object detection

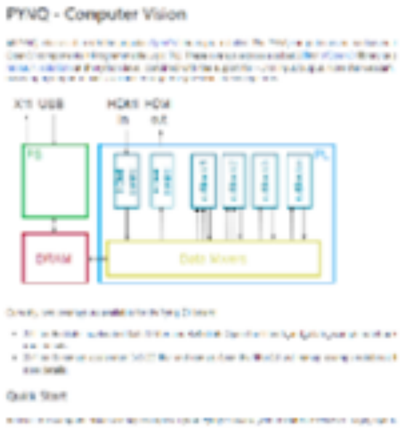
PYNQ bot
IT Tallaght
Control of robotic car from PYNQ

PYNQ LED cube
Fudan University, Xilinx China
Controlling an LED cube from PYNQ

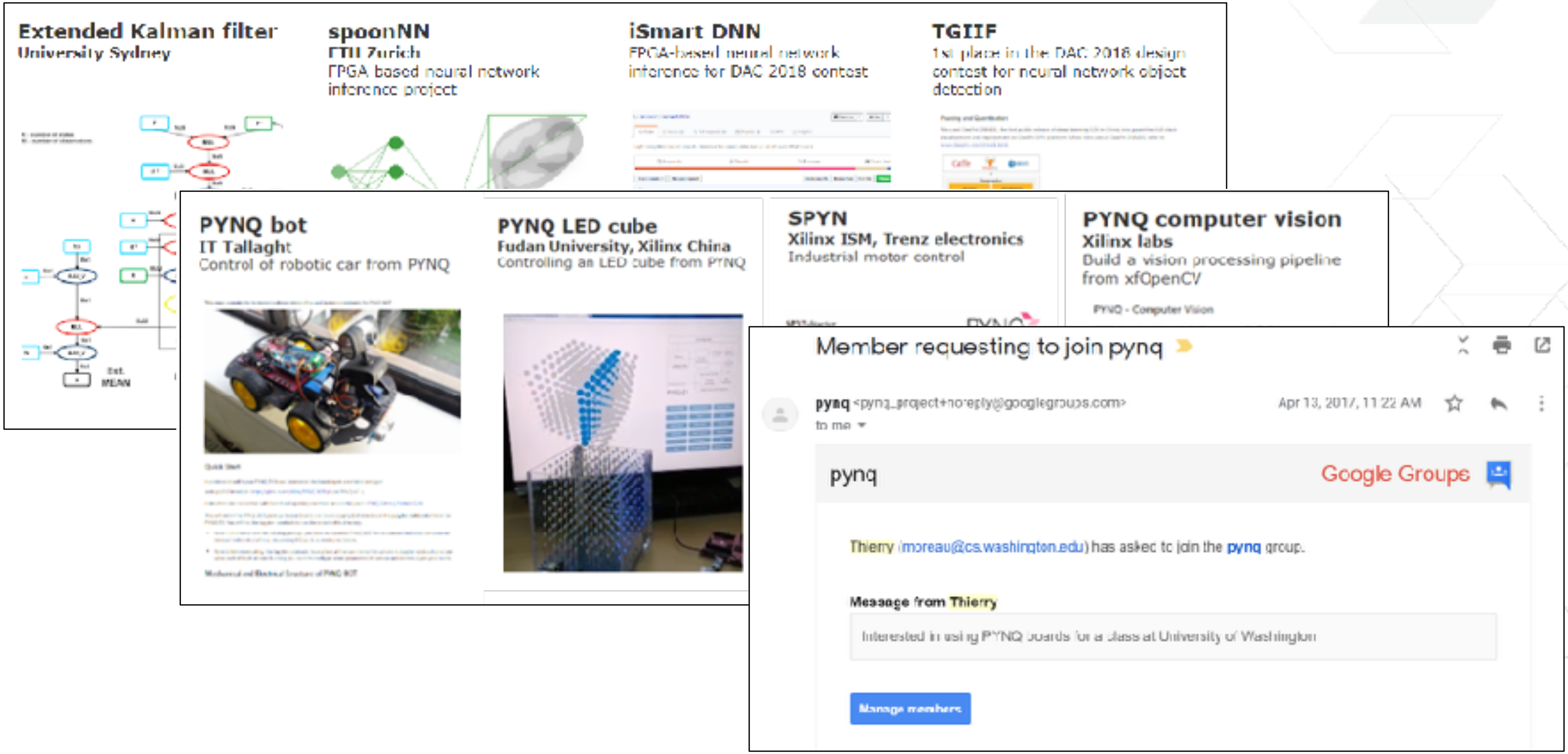


SPYN
Xilinx ISM, Trenz electronics
Industrial motor control

PYNQ computer vision
Xilinx labs
Build a vision processing pipeline from xfOpenCV



PYNQ Community – ML, Non-ML & Academic Partners



Extended Kalman filter
University Sydney

spoonNN
ETH Zurich
FPGA based neural network inference project

iSmart DNN
FPGA-based neural network inference for DAC 2018 contest

TGIIF
1st place in the DAC 2018 design contest for neural network object detection

PYNQ bot
IT Tallaght
Control of robotic car from PYNQ

PYNQ LED cube
Fudan University, Xilinx China
Controlling an LED cube from PYNQ

SPYN
Xilinx ISM, Trenz electronics
Industrial motor control

PYNQ computer vision
Xilinx labs
Build a vision processing pipeline from xfOpenCV

Member requesting to join pynq

pynq <pynq_project+noreply@googlegroups.com> Apr 13, 2017, 11:22 AM
to me

pynq Google Groups

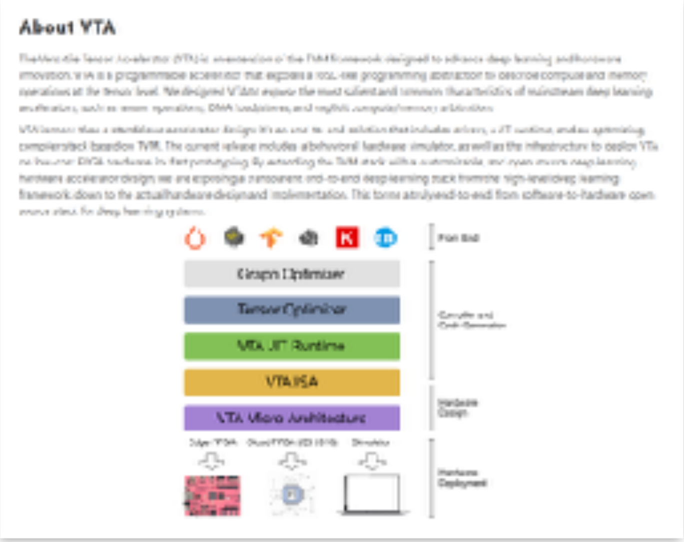
Thierry (moreau@cs.washington.edu) has asked to join the pynq group.

Message from Thierry

Interested in using PYNQ boards for a class at University of Washington

Manage members

Xilinx open source engagements related to today's TVM meeting



About VTA

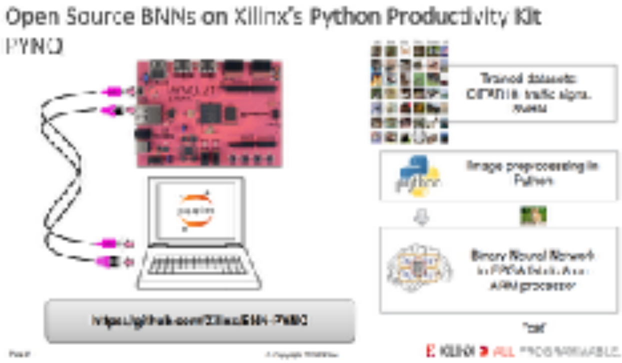
[illegible]

SOFTWARE DEFINED RADIO WITH RFSoC & PYNQ

Robert W Stewart, Louise Crockett, Craig Ramsay, Josh Goldsmith, David Northcote, Kenny Barlee

Department of Electronic and Electrical Engineering, University of Strathclyde, r.stewart@strath.ac.uk

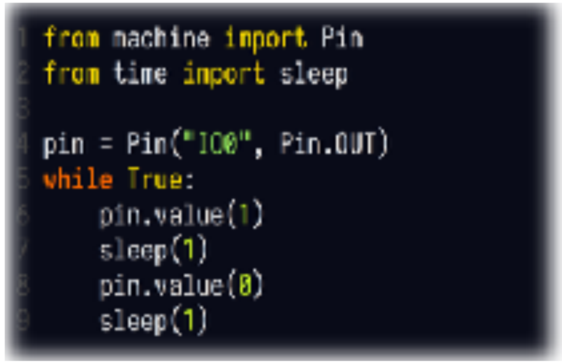
University of Strathclyde
XDF



Open Source BNNs on Xilinx's Python Productivity Kit

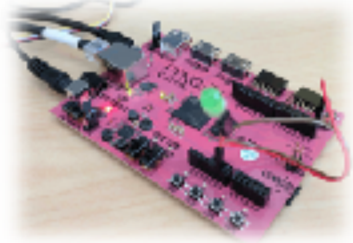
PYNQ

Trained datasets: CIFAR10, traffic signs, MNIST
Image pre-processing in Python
Binary Neural Network in FPGA fabric + ARM processor
https://github.com/Xilinx/BNN-PYNQ
XILINX ALL PROGRAMMABLE.

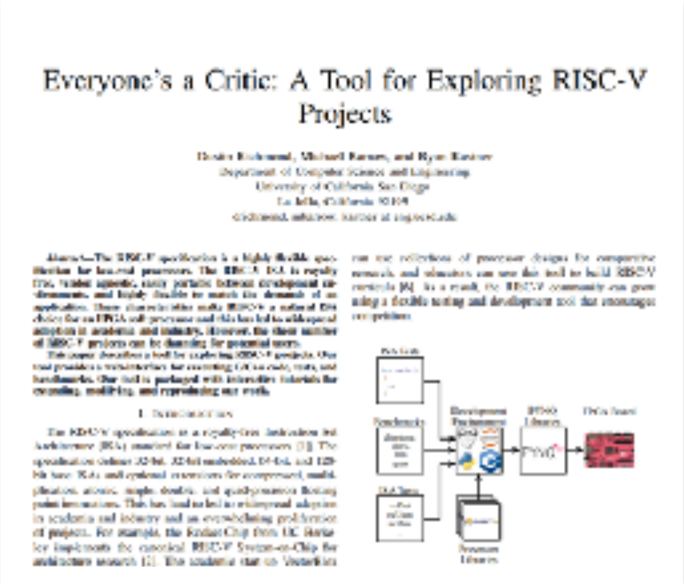


```
from machine import Pin
from time import sleep

pin = Pin("IO0", Pin.OUT)
while True:
    pin.value(1)
    sleep(1)
    pin.value(0)
    sleep(1)
```

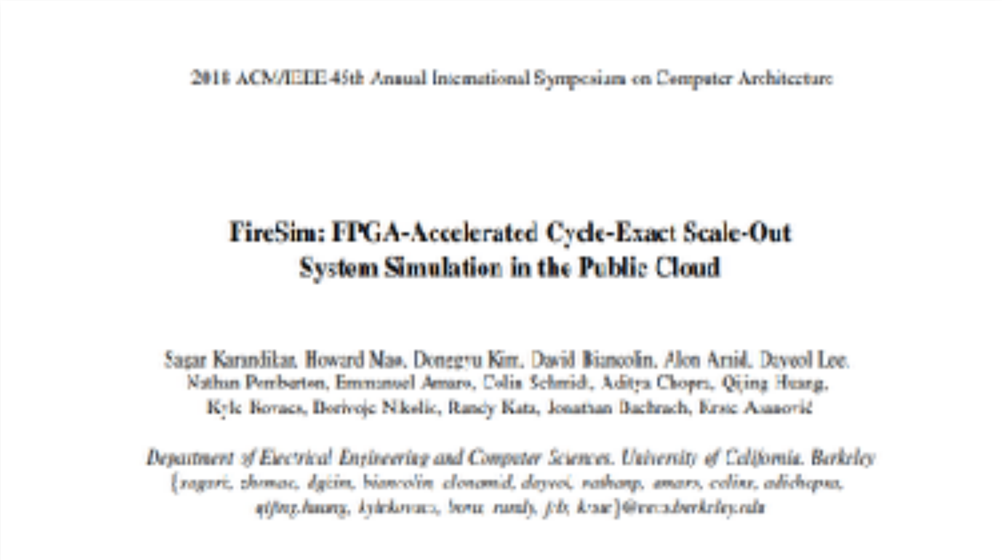


MicroPython



Everyone's a Critic: A Tool for Exploring RISC-V Projects

[illegible]



2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture

FireSim: FPGA-Accelerated Cycle-Exact Scale-Out System Simulation in the Public Cloud

Sagar Karandikar, Howard Mao, Donggyu Kim, David Biancolin, Alon Amid, Dayeol Lee, Nathan Pemberton, Emmanuel Amaro, Colin Schmidt, Aditya Chopra, Qijing Huang, Kyle Kovacs, Borivoje Nikolic, Randy Katz, Jonathan Bachrach, Krste Asanović

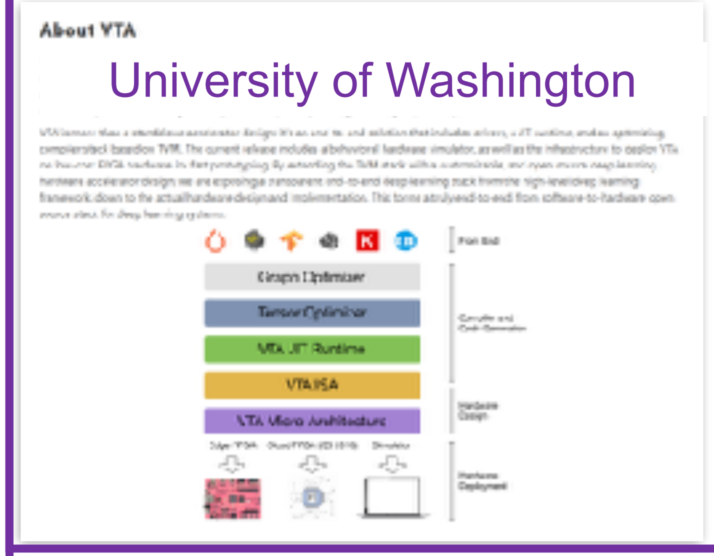
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
{sagark, zhemao, dgkim, biancolin, alonamid, dayeol, nathanp, amaro, colins, adichopra, qijing.huang, kylekovacs, bora, randy, jrb, krste}@eecs.berkeley.edu

Caffein-AI-tor

Double deep learning CNNs with face and emotion recognition, feeding predictive machine learning, to bring you the optimal caffeine kick.

Advanced · Full instructions provided · 10 hours · 716

Xilinx open source engagements related to today’s TVM meeting



About VTA

University of Washington

VTA (versatile tensor accelerator) is an open-source deep learning accelerator design. It's an end-to-end solution that includes drivers, a JIT runtime, and an optimizing compiler stack based on TVM. The current release includes a behavioral hardware simulator, as well as the infrastructure to deploy VTA on low-cost FPGA hardware for fast prototyping. By extending the TVM stack with a customizable, and open source deep learning hardware accelerator design, we are exposing a transparent end-to-end deep learning stack from the high-level deep learning framework, down to the actual hardware design and implementation. This forms a truly end-to-end, from software-to-hardware open source stack for deep learning systems.

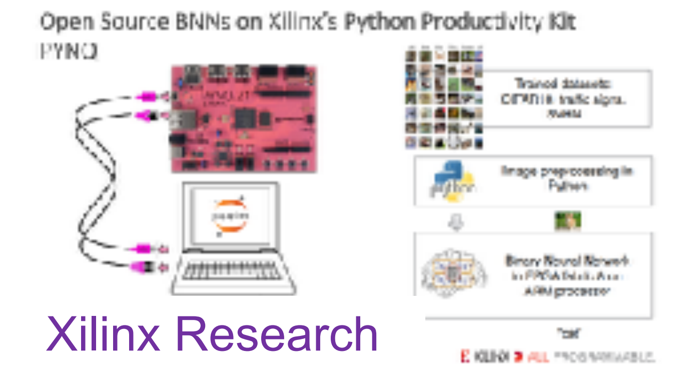
Front End
Graph Optimizer
Tensor Optimizer
VTA JIT Runtime
VTA ISA
VTA Micro Architecture
Compiler and Code Generation
Hardware Design
Edge FPGA
Cloud FPGA (2018)
Simulator
Hardware Deployment

SOFTWARE DEFINED RADIO WITH RFSoC & PYNQ

Robert W Stewart, Louise Crockett, Craig Ramsay, Josh Goldsmith, David Northcote, Kenny Barlee

Department of Electronic and Electrical Engineering, University of Strathclyde, r.stewart@strath.ac.uk

University of Strathclyde
XDF

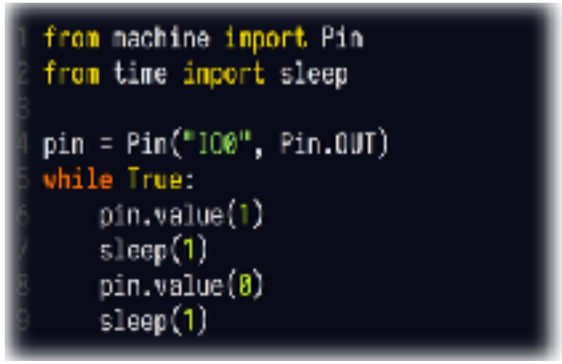


Open Source BNNs on Xilinx’s Python Productivity Kit

PYNQ

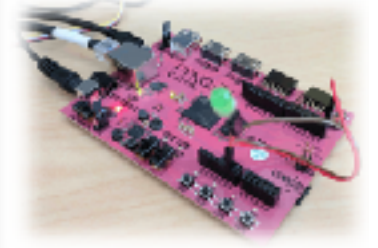
Trained datasets: CIFAR10, traffic signs, ...
Image pre-processing in Python
Binary Neural Network in FPGA fabric + ARM processor
"cat"
XILINX ALL PROGRAMMABLE.

Xilinx Research

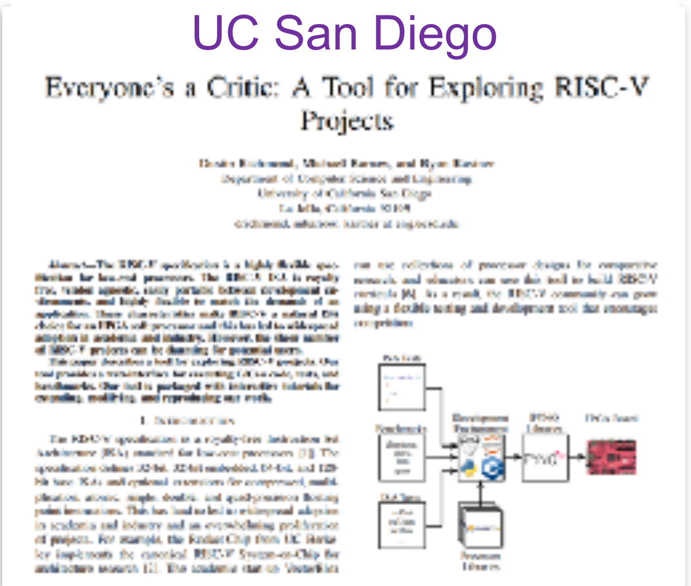


```
from machine import Pin
from time import sleep

pin = Pin("IO0", Pin.OUT)
while True:
    pin.value(1)
    sleep(1)
    pin.value(0)
    sleep(1)
```



MicroPython



UC San Diego

Everyone’s a Critic: A Tool for Exploring RISC-V Projects

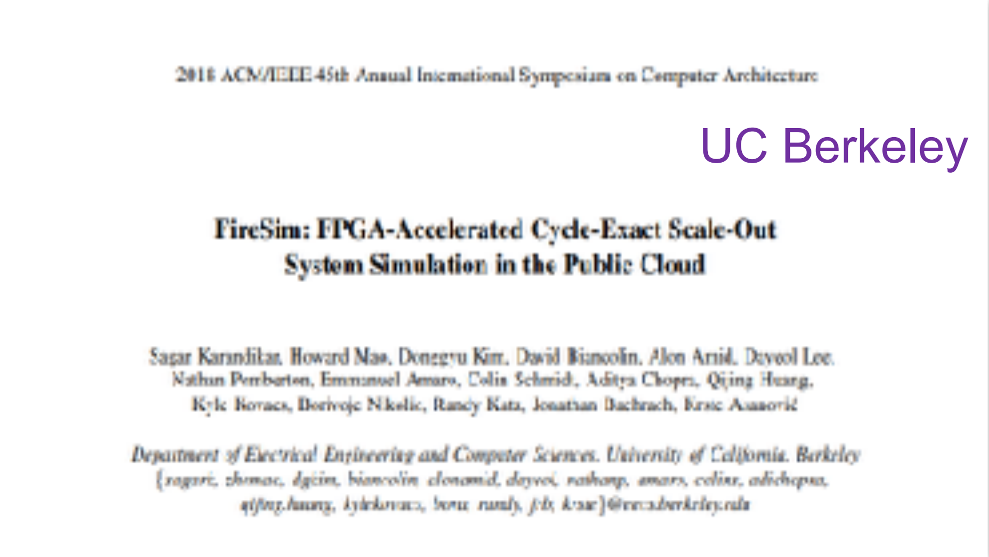
Dustin Richmond, Michael Barrow, and Ryan Kastner
Department of Computer Science and Engineering
University of California San Diego
La Jolla, California 92093
{drichmond, mbarrow, kastner}@eng.ucsd.edu

Abstract—The RISC-V specification is a highly flexible specification for low-cost processors. The RISC-V ISA is royalty free, vendor agnostic, easily portable between development environments, and highly flexible to match the demands of an application. These characteristics make RISC-V a natural ISA choice for an FPGA soft processor and this has led to widespread adoption in academia and industry. However, the sheer number of RISC-V projects can be daunting for potential users.

This paper describes a tool for exploring RISC-V projects. Our tool provides a testbench for executing C/C++ code, tests, and benchmarks. Our tool is packaged with interactive tutorials for extending, modifying, and reproducing our work.

I. Introduction

The RISC-V specification is a royalty-free Instruction Set Architecture (ISA) standard for low-cost processors [1]. The specification defines 32-bit, 32-bit embedded, 64-bit, and 128-bit base ISAs and optional extensions for compressed, multiplication, atomic, single, double, and quad-precision floating point instructions. This has lead to led to widespread adoption in academia and industry and an overwhelming proliferation of projects. For example, the RocketChip from UC Berkeley implements the canonical RISC-V System-on-Chip for architecture research [2]. The academic start-up VectorBlox [illegible] run our collections of processor designs for comparative research, and educators can use this tool to build RISC-V curricula [4]. As a result, the RISC-V community can grow using a flexible testing and development tool that encourages competition.



2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture

UC Berkeley

FireSim: FPGA-Accelerated Cycle-Exact Scale-Out System Simulation in the Public Cloud

Sagar Karandikar, Howard Mao, Donggyu Kim, David Biancolin, Alon Amid, Dayeol Lee, Nathan Pemberton, Emmanuel Amaro, Colin Schmidt, Aditya Chopra, Qijing Huang, Kyle Kovacs, Borivoje Nikolic, Randy Katz, Jonathan Bachrach, Krste Asanović

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
{sagark, zhemao, dgkim, biancolin, alonamid, dayeol, nathanp, amaro, colins, adichopra, qijing.huang, kylekovacs, bora, randy, jrb, krste}@eecs.berkeley.edu

Caffein-AI-tor

Double deep learning CNNs with [illegible] predictive machine learning, to bring you the optimal caffeine kick.

Advanced □ Full instructions provided 10 hours 716

Finally, Xilinx & building new open source communities…

Cloud Free Trials

Get started with Alveo accelerator card applications today on the Nimbix Cloud

Nimbix has partnered with Xilinx to provide developers and engineers a trial account that provides up to 100 hours of free time on the Nimbix Cloud using Xilinx Tools and Accelerators.

pynq.io/community

DAC2019 Design Contest

DAC 2019 | Las Vegas, NV | June 2-6

2019 System Design Contest

Sponsored by ACM SIGDA & DAC along with Premier Platform Sponsors: NVIDIA and Xilinx and Systems Sponsor: DJI

Show Your Talents in Machine Learning on Embedded/Hardware Platforms. Complimentary Registration Towards Grand Cash Prize!

Contest finalists will be invited to give demonstrations in the exhibition hall during the Design Automation Conference, June 3 - June 6, 2019 in Las Vegas, NV.

OpenHW Design Contest

XILINX Open Hardware 2019

Xilinx University Program FPGA and SOC University Design contest

- PhD
- Student
- PYNQ
- Compute Acceleration

Platforms

2019 Registration

Summary

Xilinx

Great for exploring and deploying inference

Xilinx Open Source

We're actively engaging with TVM and other communities

Email: graham.schelle@xilinx.com

Visit: Boulder, Colorado

Adaptable.

Intelligent.

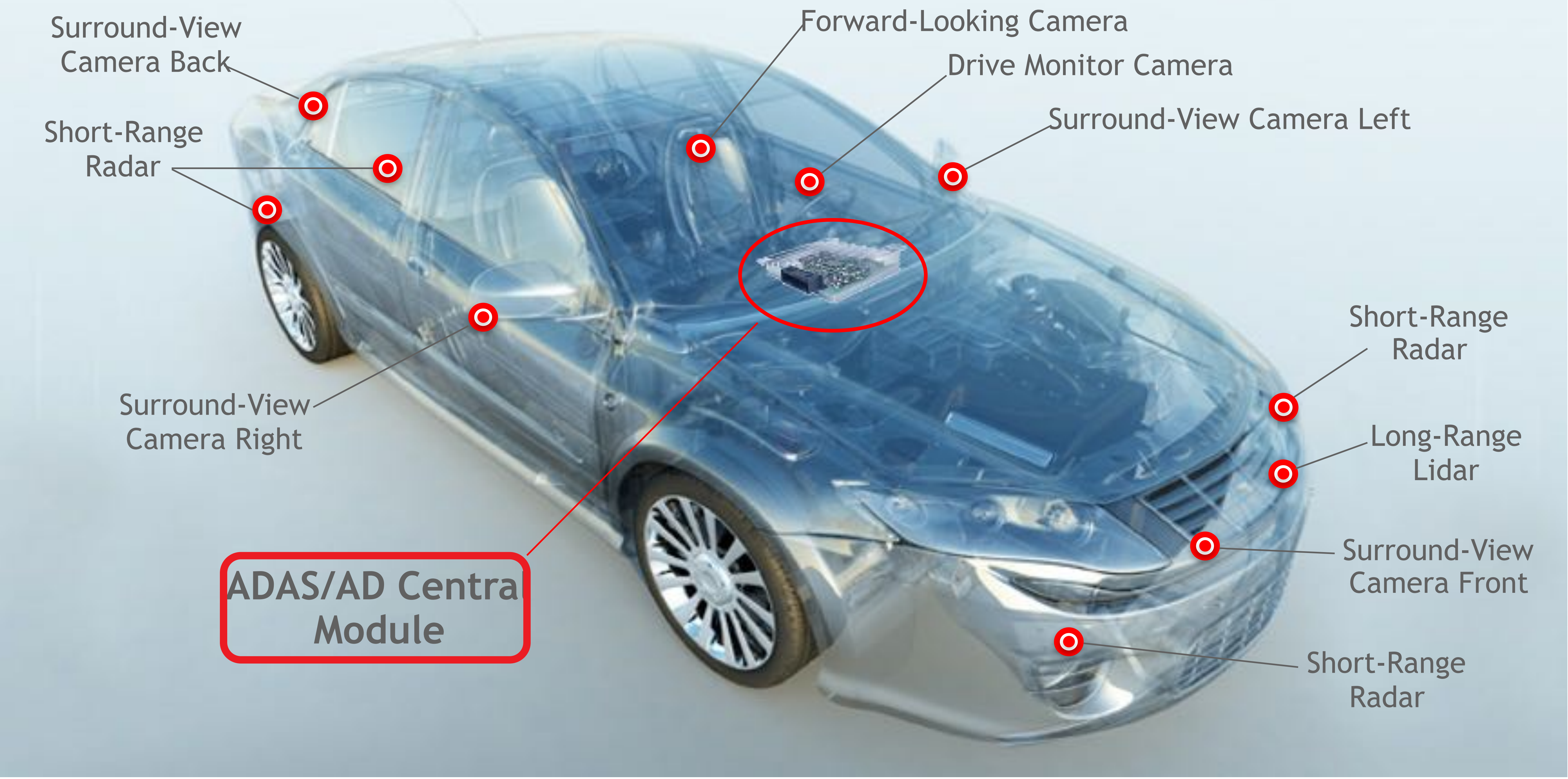
XILINX.

Edge to Cloud Inference – Automotive

Edge to Cloud Inference – Automotive

ADAS/AD Central Module

Edge to Cloud Inference – Automotive



Edge to Cloud Inference – Xilinx Platforms

PYNQ-Z1

ZCU104

Ultra96

ZCU102

Edge Devices

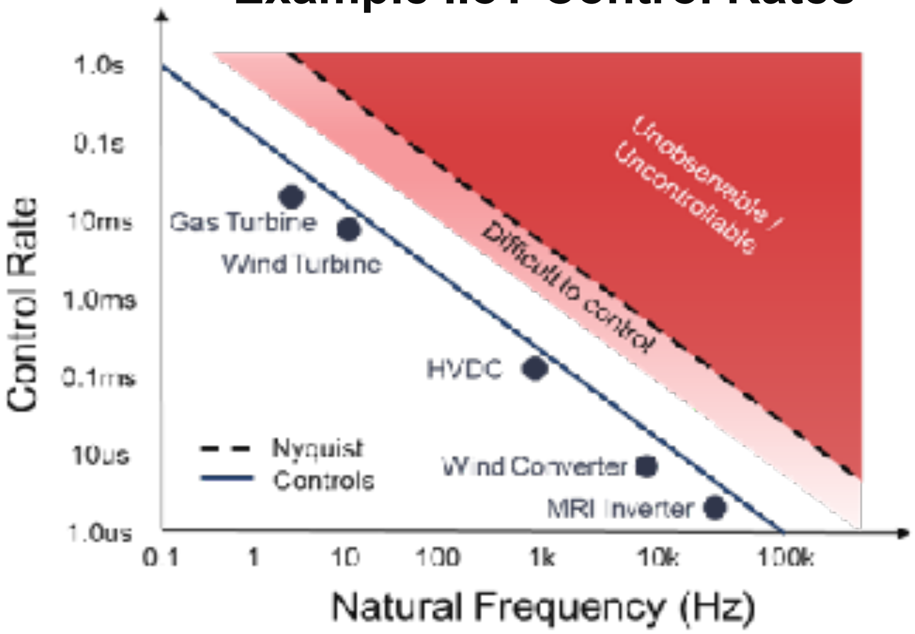
Custom I/O, ARM CPUs

Cloud Platforms

Power Efficient, PCIe, Networking

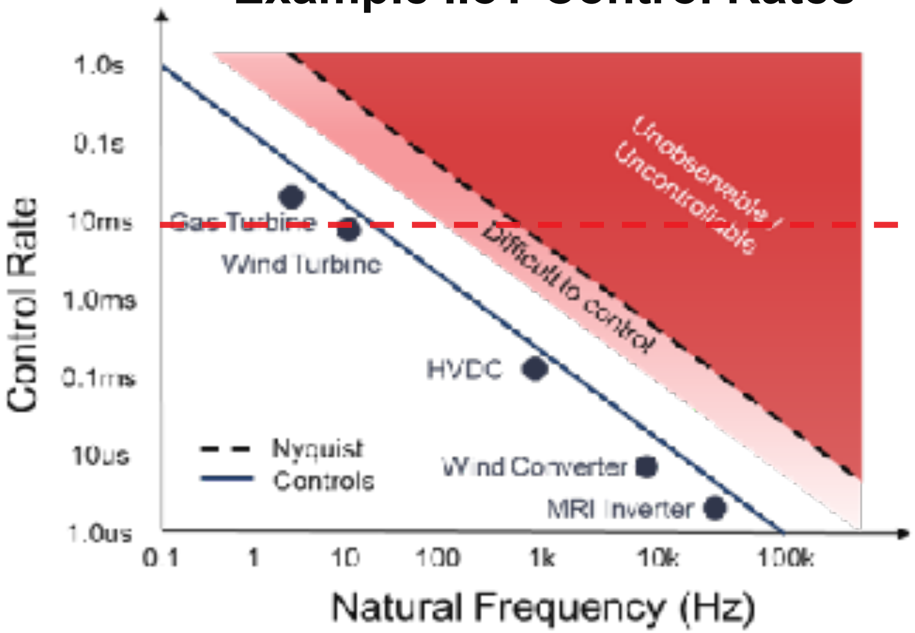
Edge to Cloud Inference – IIoT Latency/Data Example

Example IIoT Control Rates

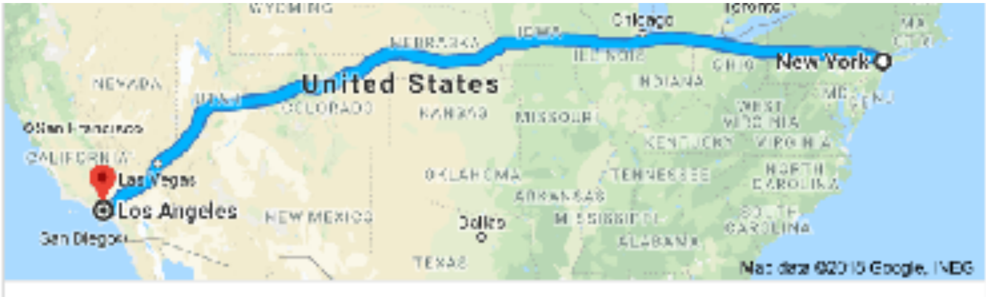


Edge to Cloud Inference – IIoT Latency/Data Example

Example IIoT Control Rates

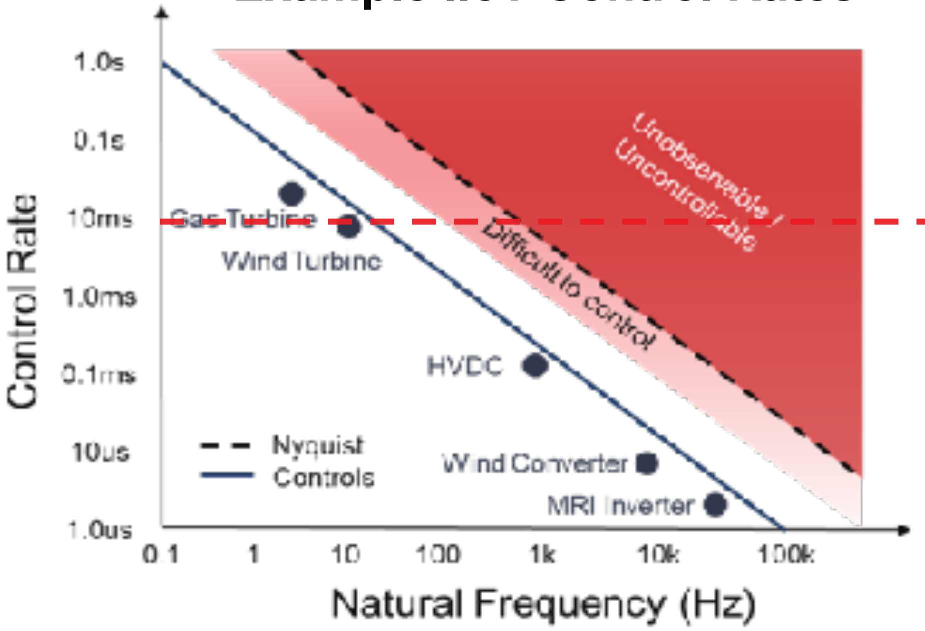


Distance NYC to LA: 2,800 miles
Speed of light: 186,000 miles/s
Round trip: 2*2800/186000 = 30ms
Required Control Rate = 10ms

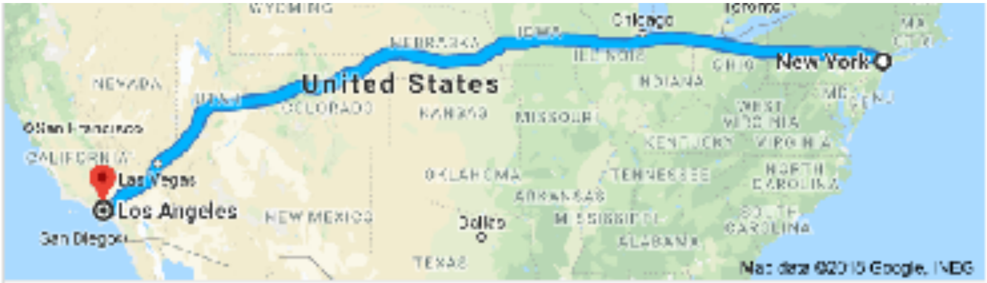


Edge to Cloud Inference – IIoT Latency/Data Example

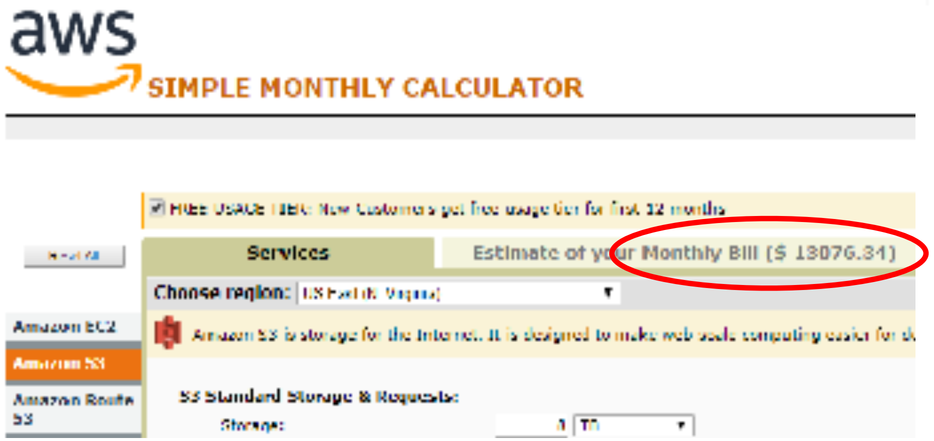
Example IIoT Control Rates



Distance NYC to LA: 2,800 miles
Speed of light: 186,000 miles/s
Round trip: 2*2800/186000 = 30ms
Required Control Rate = 10ms



E.g. Power Plant @ 8TB/Month



☑ FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services | Estimate of your Monthly Bill ($ 13076.31)

Choose region: US East (N. Virginia)

Amazon EC2
Amazon S3
Amazon Route 53

Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for d

S3 Standard Storage & Requests:
Storage: 1 TB