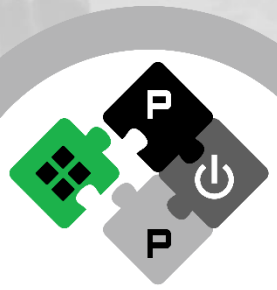


Open, Secure, Extreme Edge AI for the IoT

*An **Open Source Hardware** Success Story*

14.11.2019

Luca Benini[1,2]

[1]*Department of Electrical, Electronic and Information Engineering*

ETH zürich

[2]*Integrated Systems Laboratory*

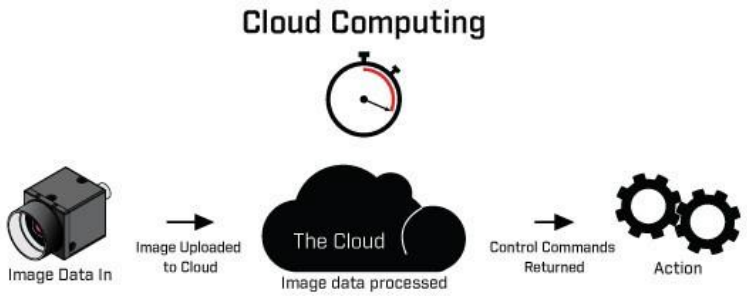
erc
European Research Council

European Commission
Horizon 2020
European Union funding
for Research & Innovation

nano-tera.ch

FNSNF
Fonds National Suisse
Schweizerischer Nationalfonds
Fondo Nazionale Svizzero
Swiss National Science Foundation

Idea in 2013: Cloud → Edge → Extreme Edge AI



#1 Customer Question on Amazon.com (out of 1,000+):

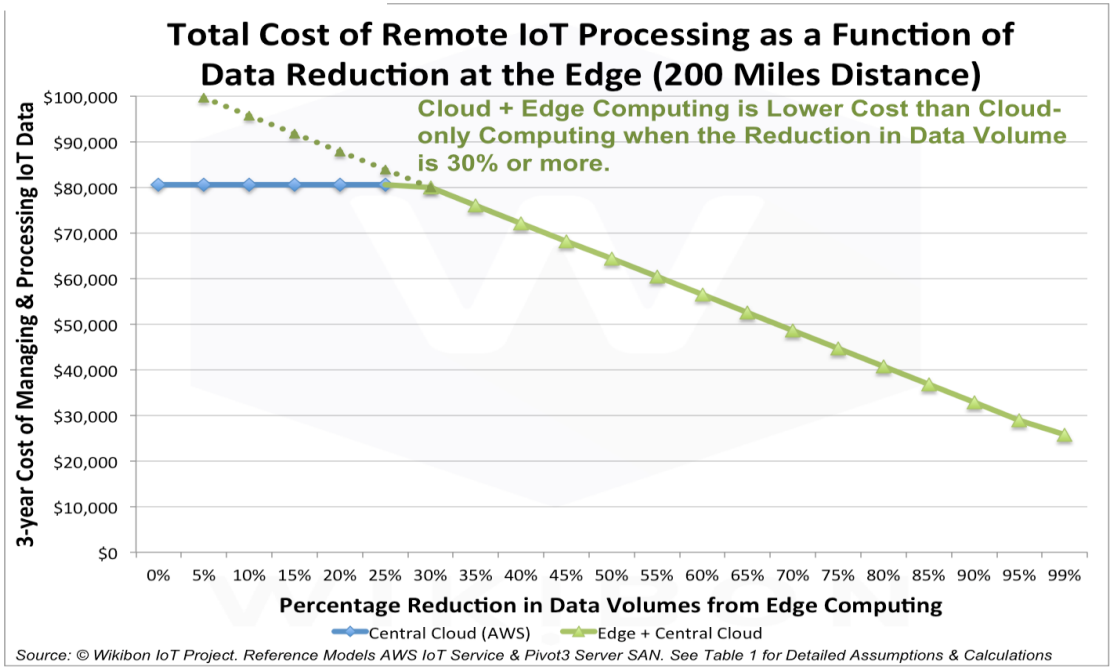
1. *I don't want any of my (private, personal) videos on any servers not in my control. Is this possible?*

#2 Customer Question on Amazon.com (out of 1,000+):

2. *How long does the battery charge last?*

Source: www.amazon.com/ask/questions/asin/B01M3VHG87/

Cost



Source: © Wikibon IoT Project. Reference Models AWS IoT Service & Pivot3 Server SAN. See Table 1 for Detailed Assumptions & Calculations

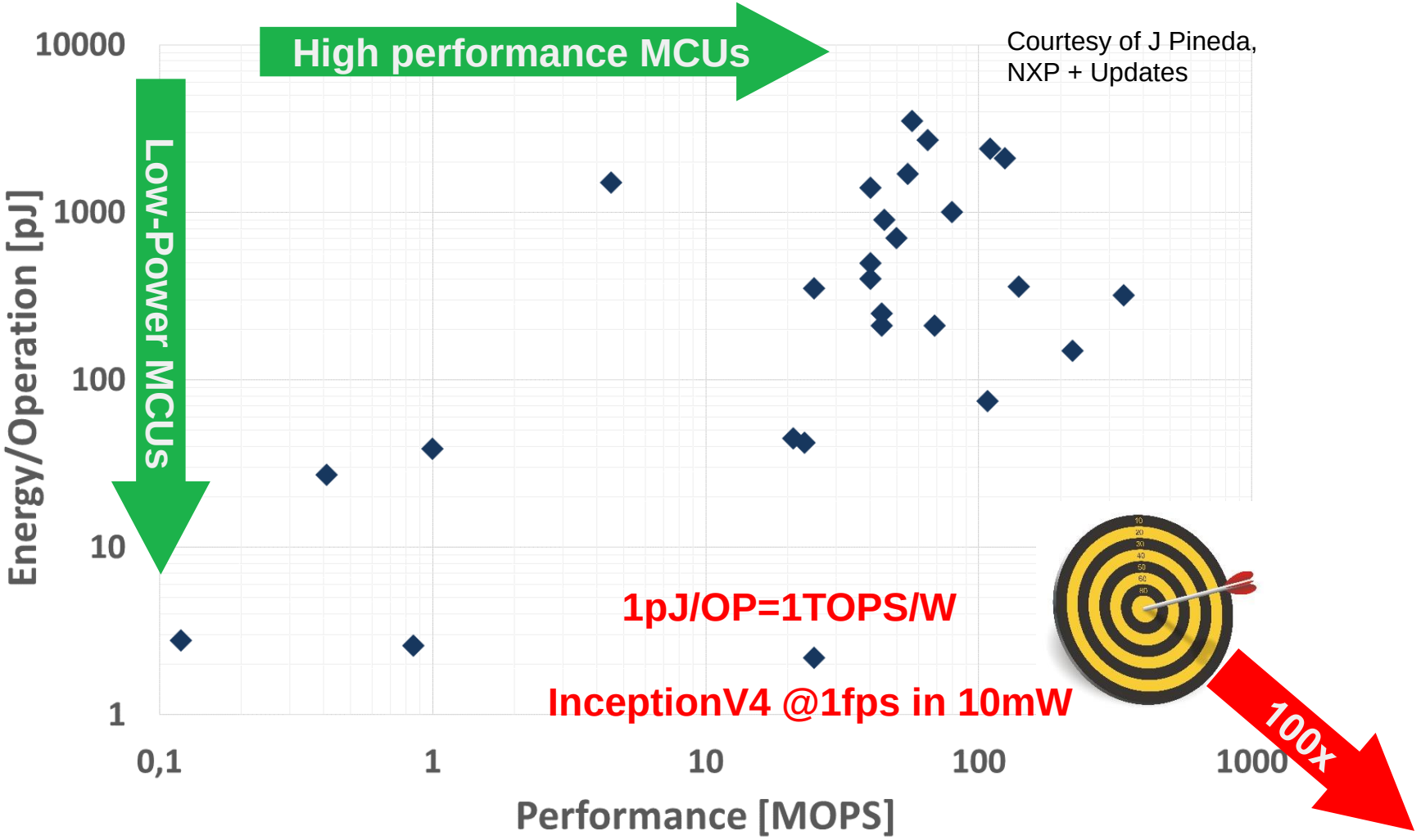
Extreme edge (i.e. on sensor) AI challenge

AI capabilities in the power envelope of an MCU: 100mW and below

PULP

Energy efficiency is THE Challenge

Current AI models need GOPS – Edge devices have mW power budget



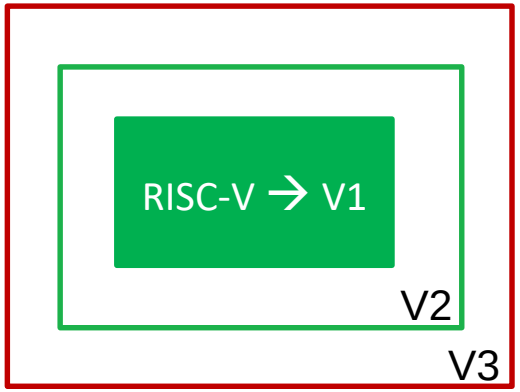
Cool… But, HOW??

Need a low power MCU that is good at AI →Bespoke ISA!

A modern, open, free ISA, extensible *by construction!*
Endorsed and Supported by 130+ Companies

V1 Baseline RISC-V RV32IMC
HW loops
V2 Post modified Load/Store
Mac

V3 SIMD 2/4 + DotProduct + Shuffling
Bit manipulation unit
Lightweight fixed point



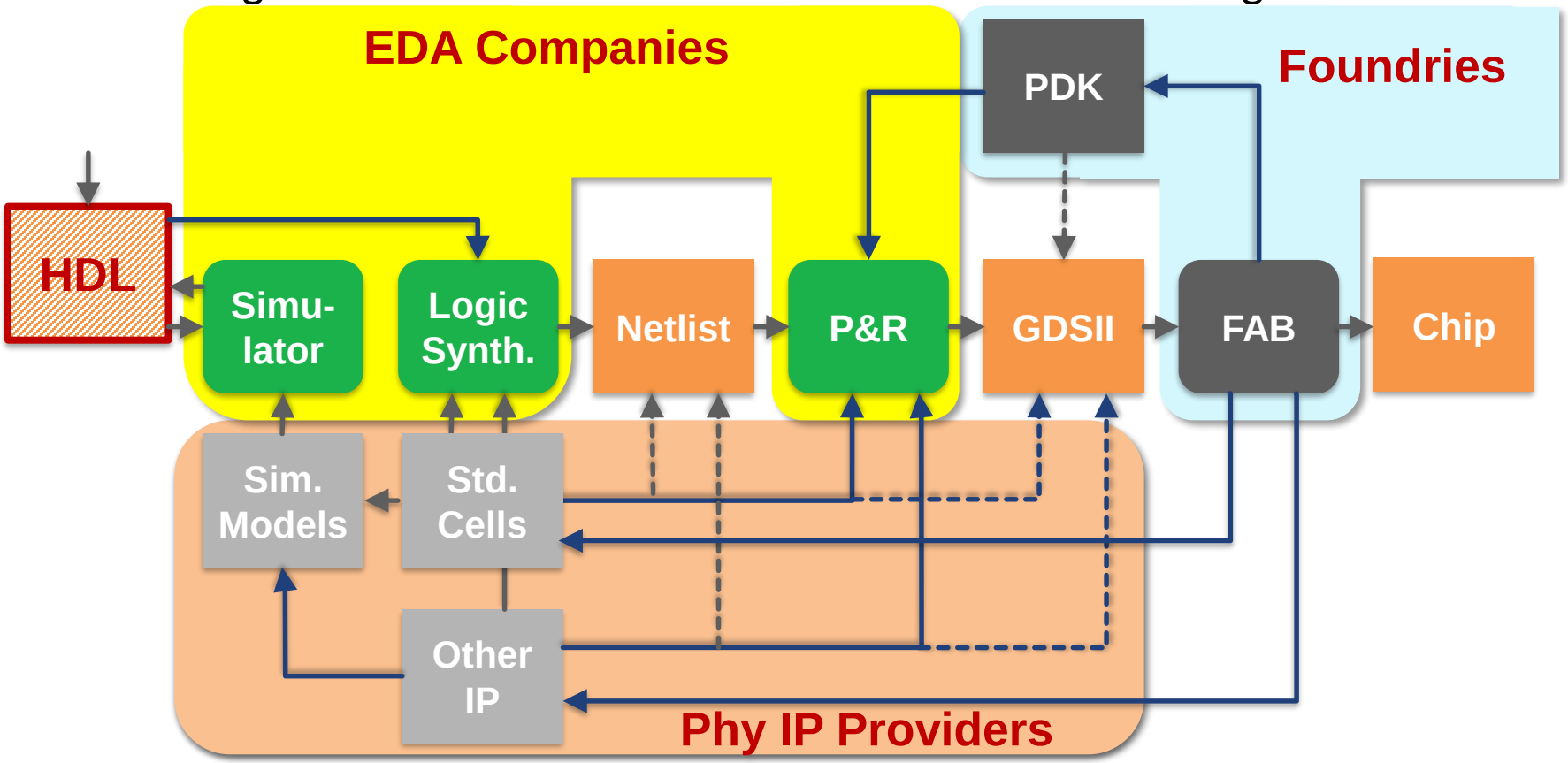
But… Low Energy comes from ISA+architecture+implementation

PULP: Ultra-Efficient Open (multi)-processor design based on Open ISA

M. Gautschi et al., "Near-Threshold RISC-V Core With DSP Extensions for Scalable IoT Endpoint Devices," in IEEE TVLSI, Oct. 2017.

Nice, but what exactly is “open” in Open HW?

- Only the first stage of the silicon productionpipeline can be open HW
  → **RTL source code** (in an HDL such as SystemVerilog)
- Later stages contain closed IP of various actors + tool licensing issues



Permissive, Copyright (e.g APACHE) License is Key for industrial adoption

PULP

Again Nice, but… Why Open Source HW?

- **For science** … fundamental "research infrastructure"
  - Community building: sharing of ideas, artefacts
  - Fantastic tool for dissemination (more citations ☺)!
  - Reduce "getting up to speed" overhead for partners
  - Enables fair and well controlled benchmarking
- **For Business** … it is truly disruptive
  - Reduces the NRE cost for silicon design
  - Faster innovation path for startups
  - New business models (for profit and non-for profit)
  - Helps exchange of information across NDA walls
  - Great for Marketing & Training
- **For society** …long term sustained benefits
  - More innovation, growth, jobs
  - Personalized silicon vision "Moore-for-all"
  - More Secure, safe, auditable HW

DARPA

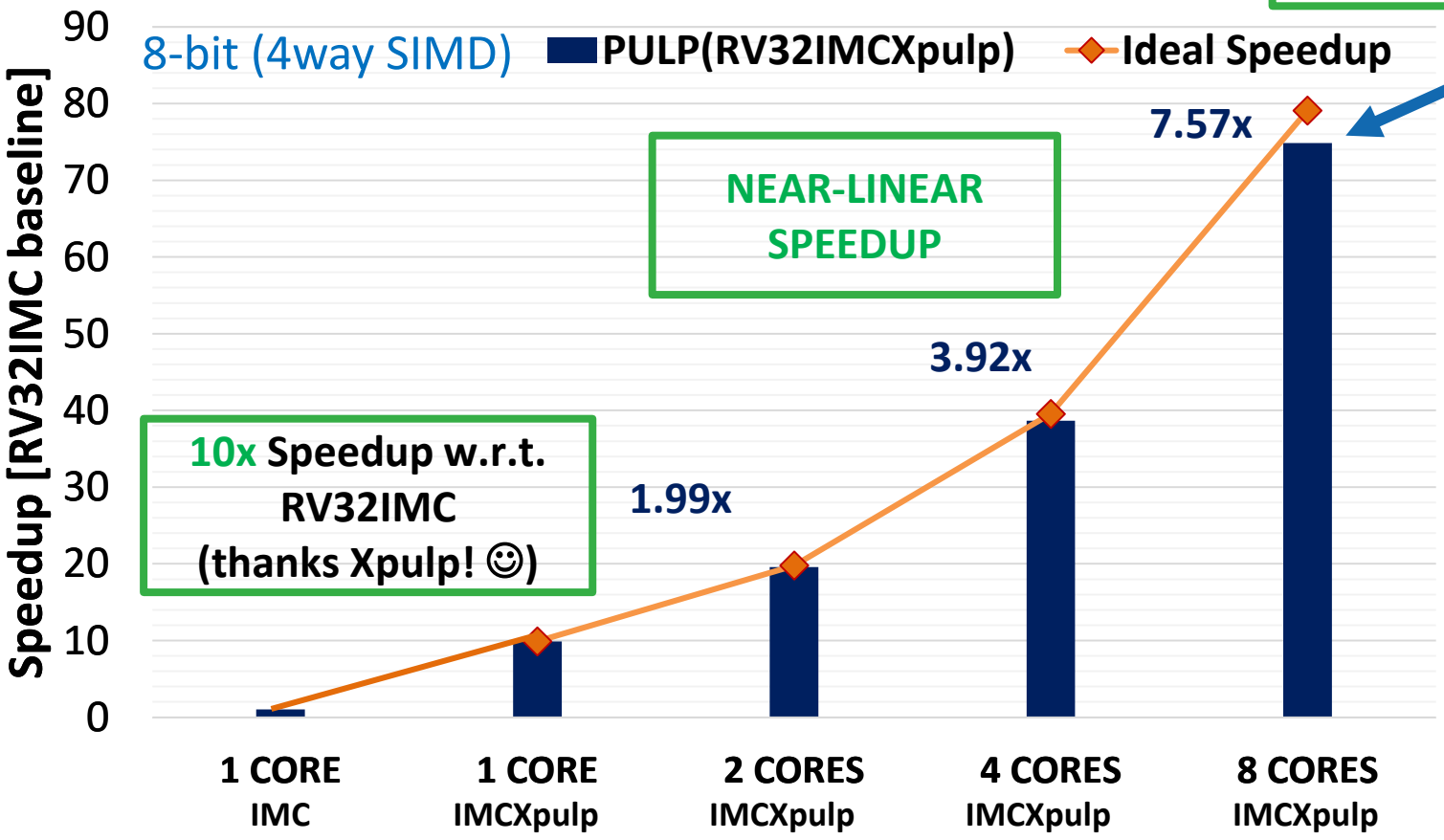
Posh Open Source Hardware (*POSH*): An open source System on Chip (SoC) design and verification eco-system that enables cost effective design of ultra-complex SoCs

USA's electronics resurgence intitiative

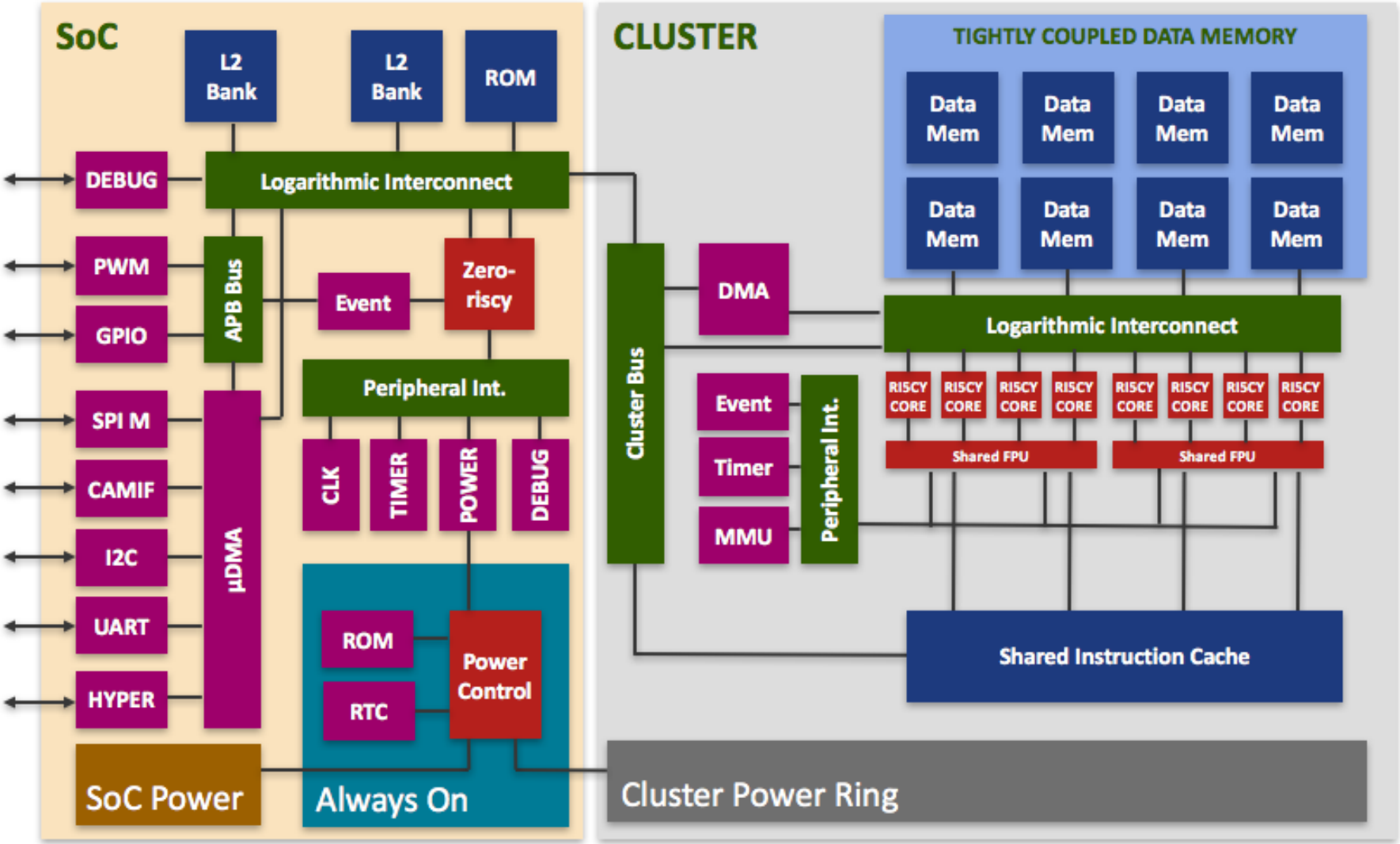
Results: RV32IMCXpulp vs RV32IMC

New instructions → 1.6 complexity increase

Overall Speedup of **75x**



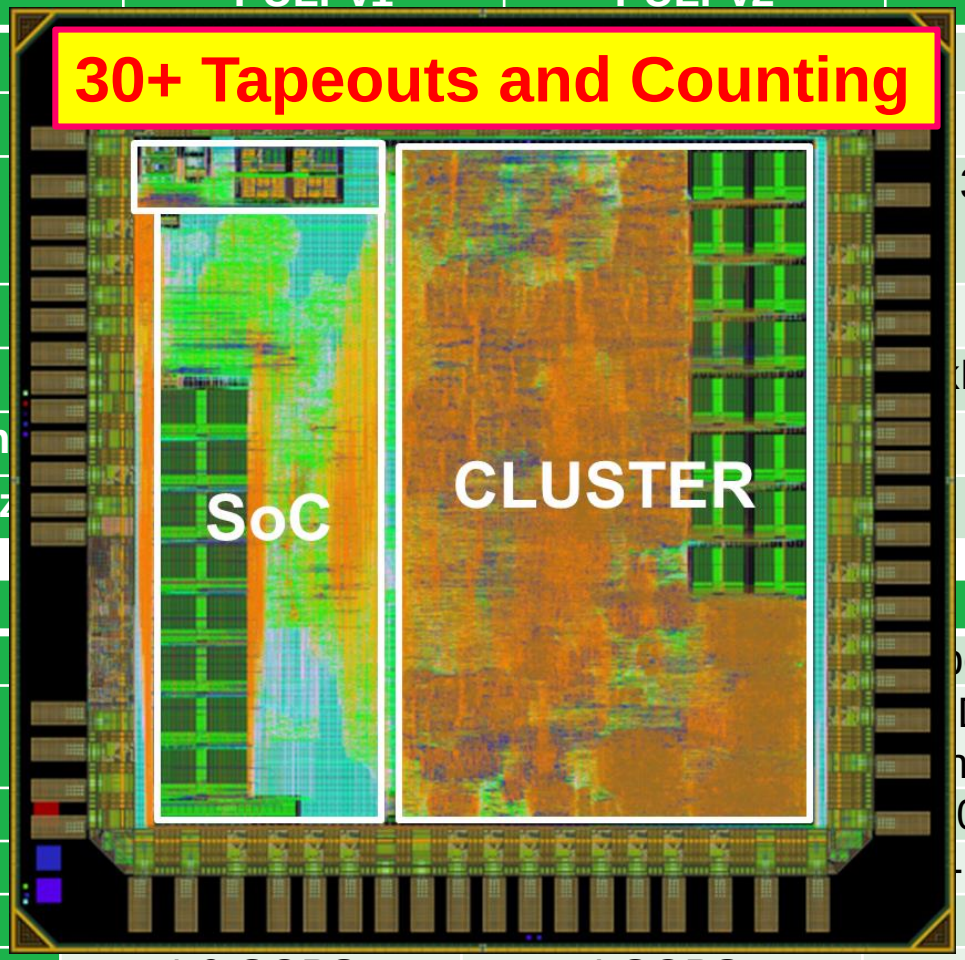
The Core is not enough → Open IoT Processor



PULP

Simulation is not enough → Silicon Proven Open HW

	PULPv1	PULPv2	PULPv3
# of cores			4
L2 memory			128 kB
TCDM			32kB SRAM 16kB SCM
DVFS			yes
I$			B SCM shared
DSP Extension			yes
HW Synchroniz			yes



			PULPv3
Status			post tape out
Technology			D-S nventional well
Voltage range			0.5V - 0.7V
BB range			-1.8V - 0.9V
Max freq.			200 MHz
Max perf.	1.9 GOPS	4 GOPS	1.8 GOPS
Peak en. eff.	60 GOPS/W	135 GOPS/W	385 GOPS/W

2.6pj/OP

life.augmented

PULP

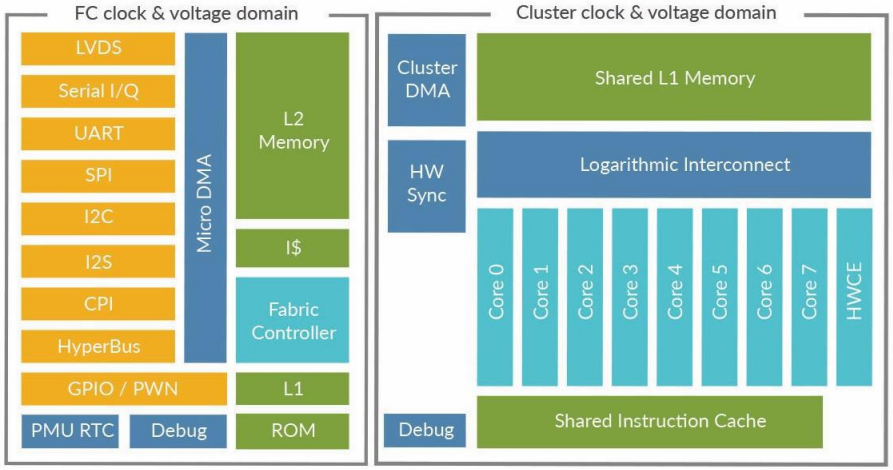
Prototype is not Enough → Open HW-based SoC Product

- Best in class Instruction Set Architecture (ISA)
- GWT Member of RiscV Foundation

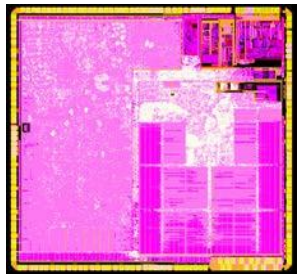
- Open Source Computing Platform created by ETHZ and UniBo
- Permissive license (solderpad)
- Multiple tape outs
- GWT contributes to PULP

- Innovating on Risc-V and PULP
- **GAP8: Proprietary balanced system solution (SOC) based on PULP open source elements plus GWT proprietary elements both on HW and SW/Tools side**

GAP8 is in mass production now

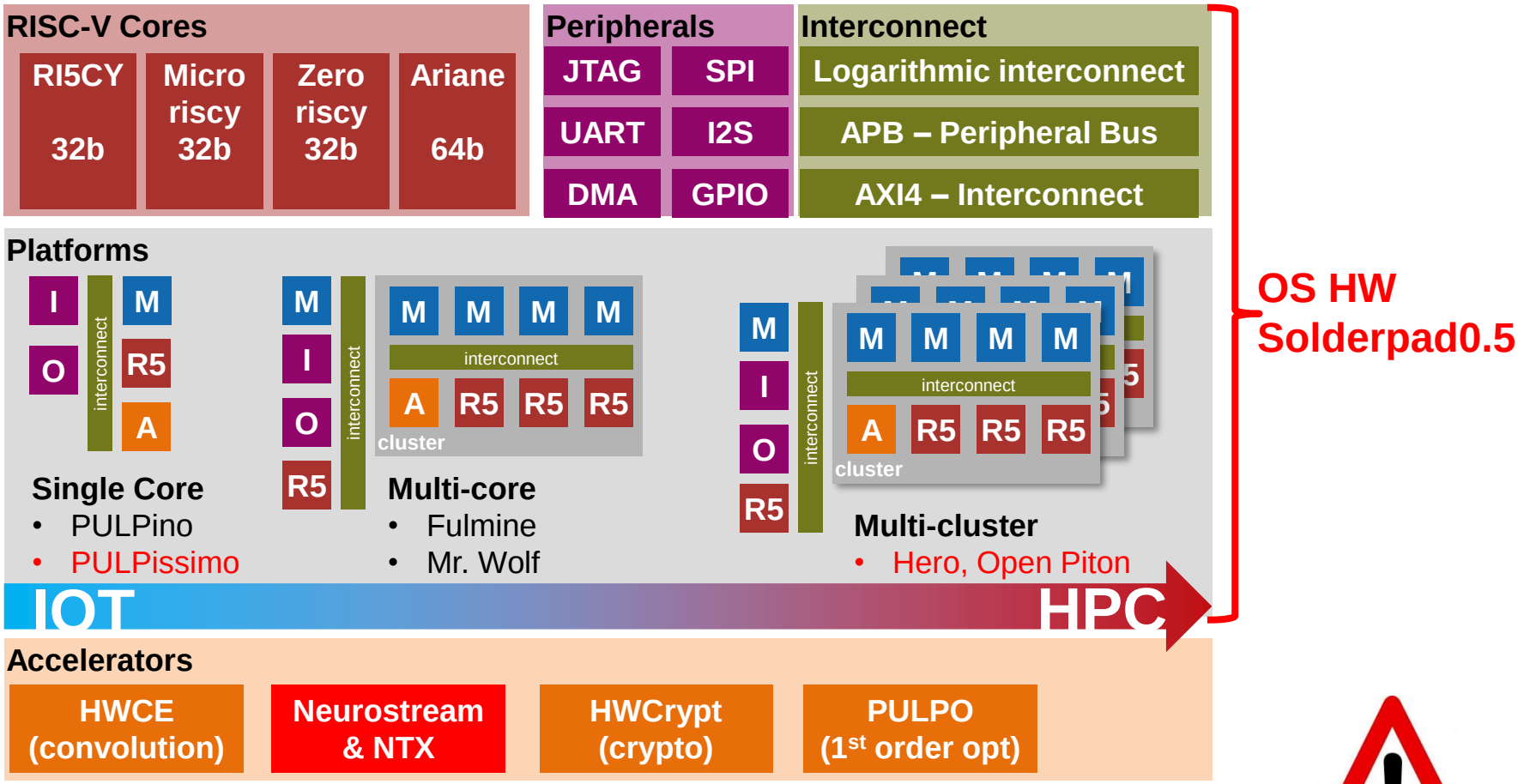


TSMC 55LP 1.0V to 1.2V
Max f: 133-250 MHz
Up to 12.8 GOPS (8bit)



Putting it all together: The Open PULP platform

https://www.pulp-platform.org/ https://github.com/pulp-platform



But this is way too much for a university (or two)!

Academic Open-Source → Industrial Open source

OPENHW GROUP™ PROVEN PROCESSOR IP

Rick O'Connor (OpenHW CEO, former RISC-V foundation director)

- **OpenHW Group** is a not-for-profit, global organization (EU,NA,Asia) driven by its members and individual contributors where HW and SW designers collaborate in the development of open-source cores, related IP, tools and SW such as the **CORE-V** Family of cores.
- **OpenHW Group** provides an infrastructure for hosting high quality open-source HW developments in line with industry best practices.

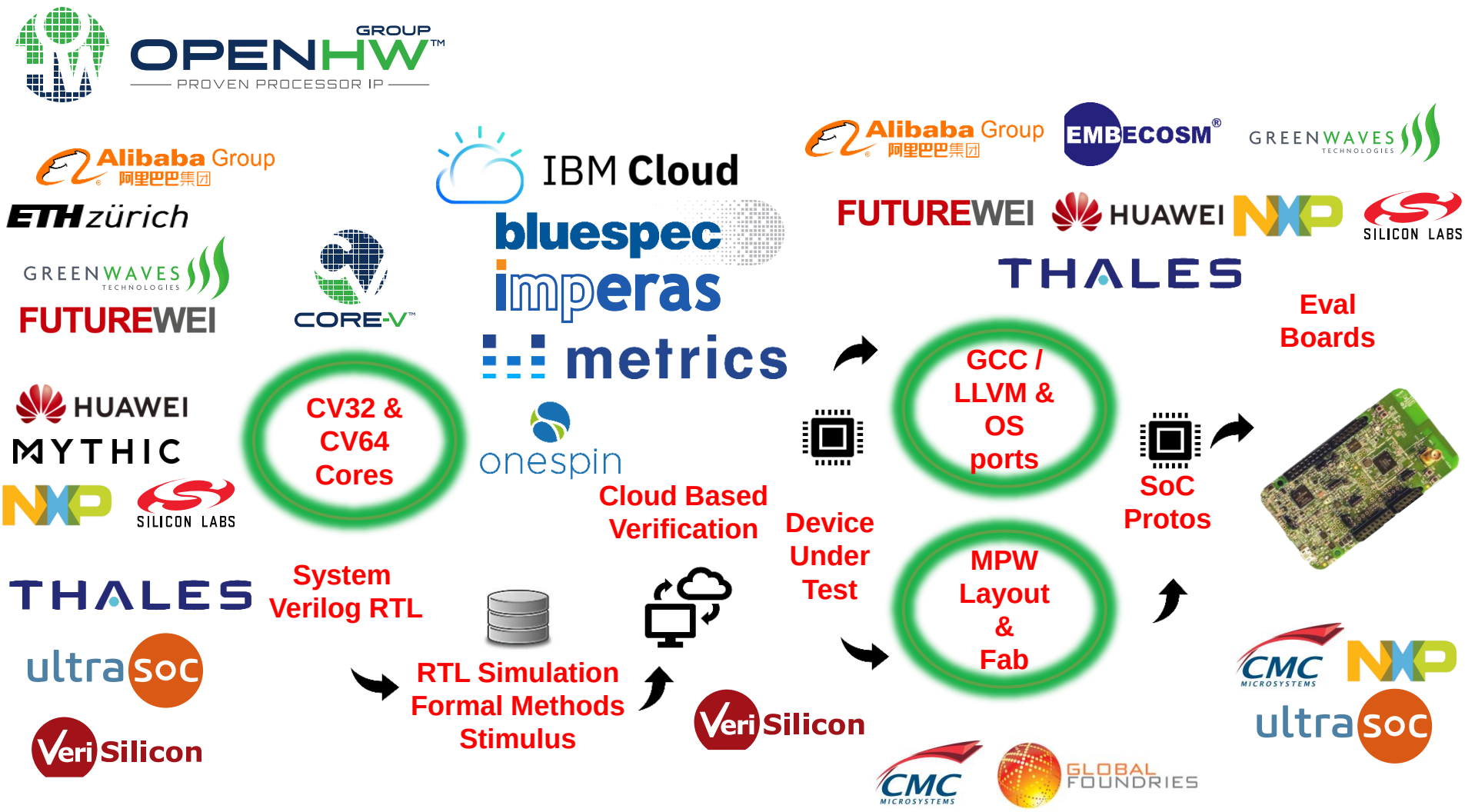
PULP Parallel Ultra Low Power

RI5CY, ARIANE

CORE-V™

Alibaba Group 阿里巴巴集团 · bluespec · BTA DESIGN SERVICES your project, our team · CMC · ETH zürich Integrated Systems Laboratory · EMBECOSM® · FUTUREWEI · GLOBAL FOUNDRIES · GREENWAVES TECHNOLOGIES · HUAWEI · imperas · metrics · MYTHIC · NXP · onespin · SILICON LABS · THALES · ultraSoC · VeriSilicon

OpenHW Group Ecosystem



A Vertical, Application-focused Open-Platform Approach

OpenTitan

More transparent, trustworthy, and secure RoT chip design

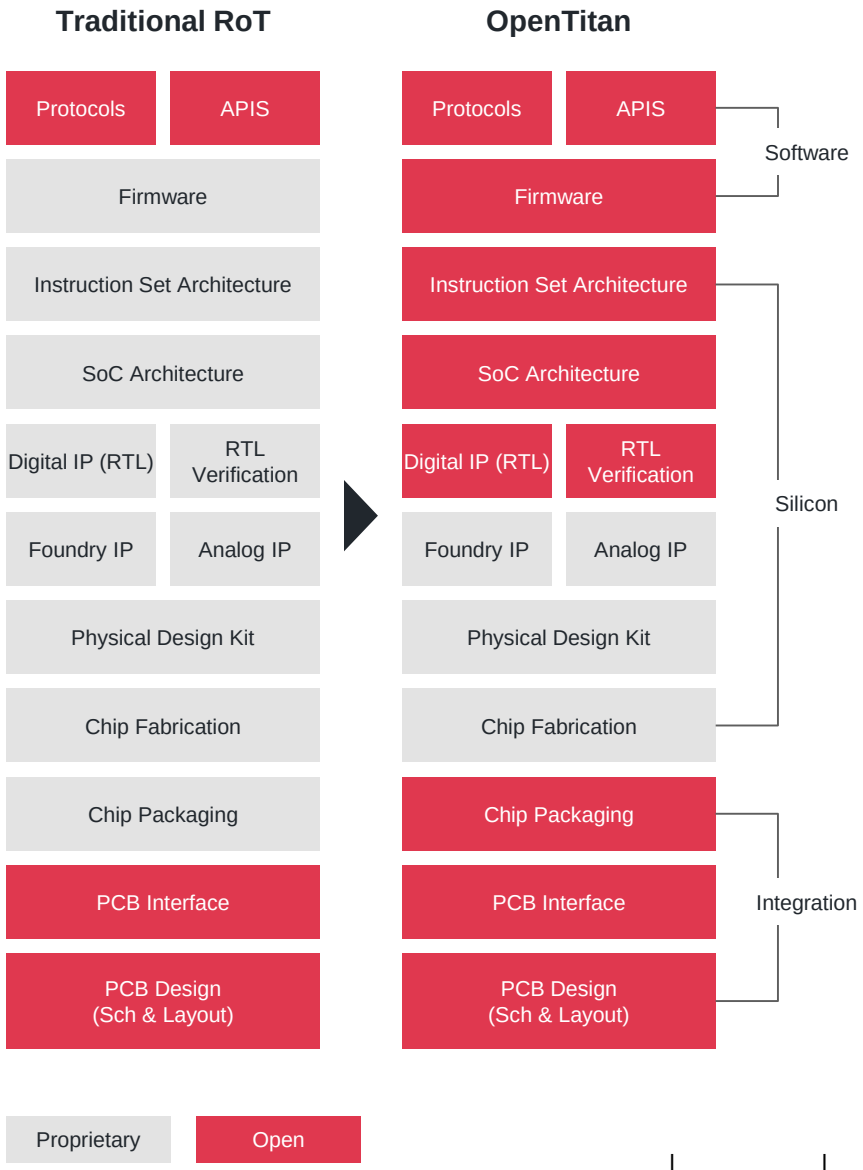
OpenTitan is the first open source silicon project building a transparent, high-quality reference design for silicon root of trust (RoT) chips.

Founding partners

Open HW enables a New Level of Openness in Security

Transparent:
Open implementation

- Transparency at the bottom; lower than any existing RoT solutions
- Transparency enables the community to proactively audit, evaluate, & improve the design
- Engineering: reference firmware, register-transfer level (RTL), design verification (DV), and integration guidelines



PULP

Feel the momentum!

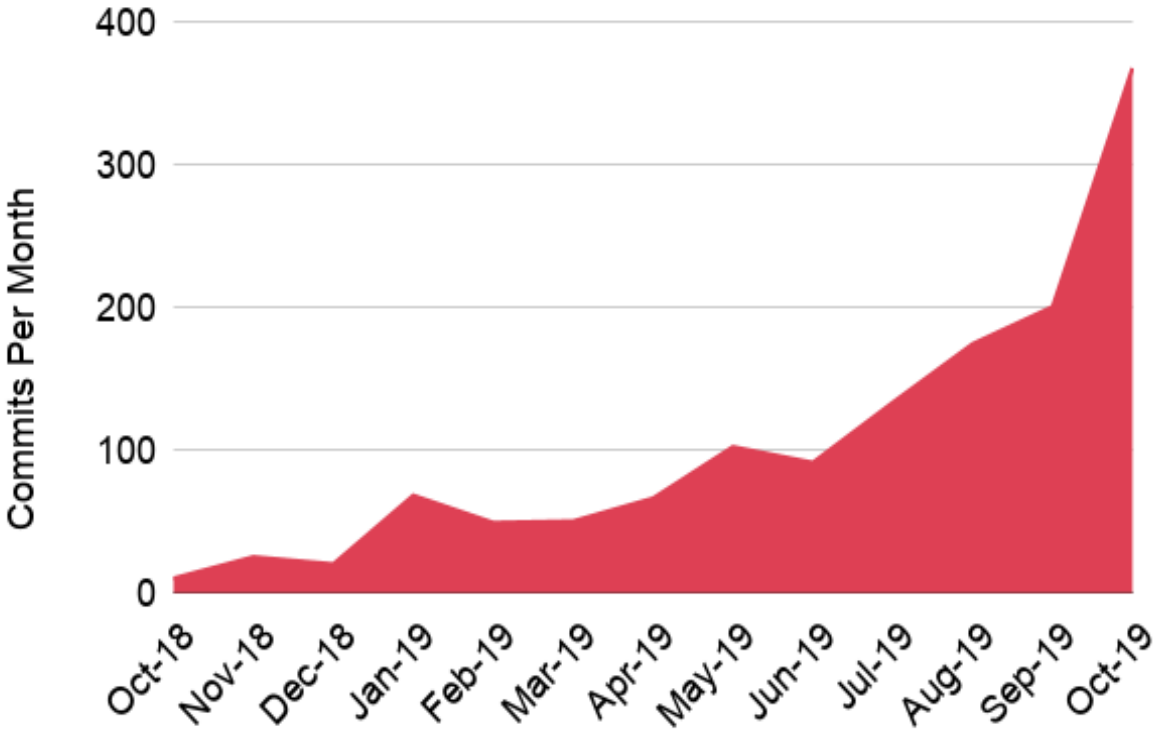
Ibex RISC-V core, flash interface, communications ports, cryptography accelerators, and more.

Vibrant repository

35+	Contributors
1300+	Contributions
470	GitHub Issues

PULP Parallel Ultra Low Power → lowRISC

Zero-Riscy → Ibex



NoT only IoT: the European Processor Initiative

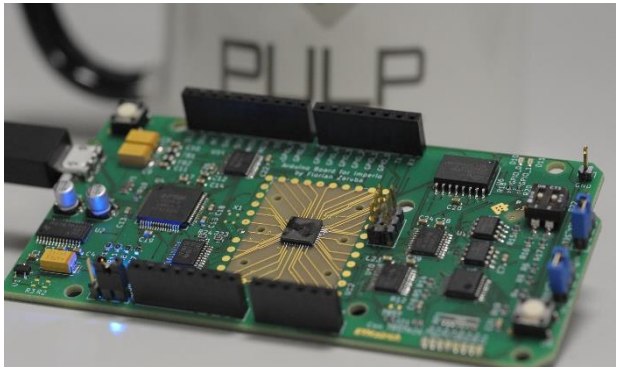
European Processor Initiative

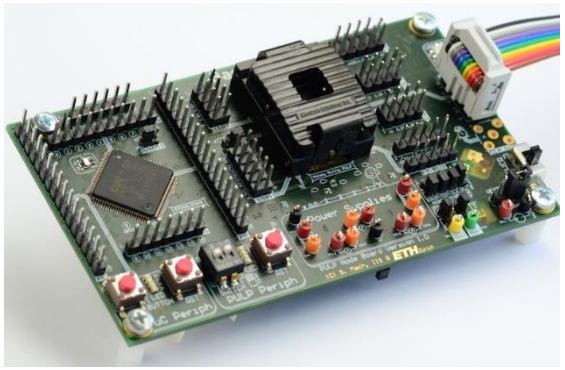
Europe Needs its own Processors

- Processors now control almost every aspect of our lives
- **Security** (back doors etc.)
- Possible **future restrictions on exports to EU** due to increasing protectionism
- **A competitive EU supply chain** for HPC technologies will create jobs and growth in Europe
- Sovereignty (data, economical, embargo)

- High Performance General Purpose Processor for HPC
- **High-performance RISC-V based accelerator**
- Computing platform for autonomous cars
- Will also target the AI, Big Data and other markets in order to be economically sustainable

PULP





www.pulp-platform.org

The fun is just beginning...