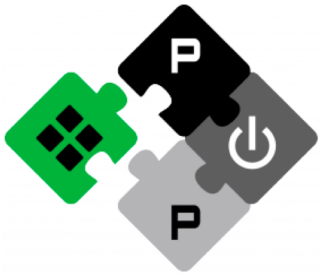
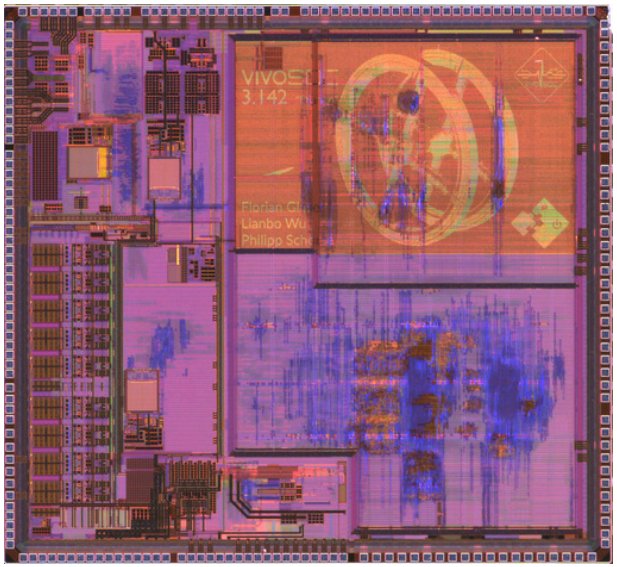


PULP PLATFORM
Open Source Hardware, the way it should be!

RISC-V for IoT, the PULP experience



Frank K. Gürkaynak <kgf@ee.ethz.ch>

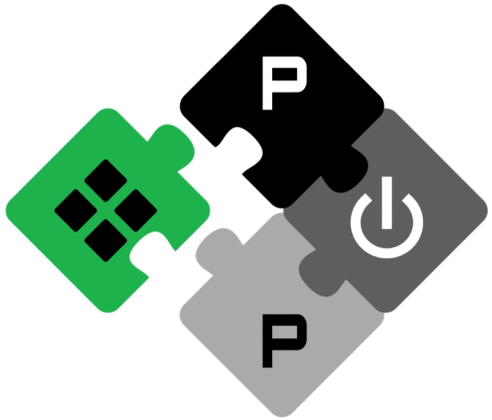
Digital Circuits and Systems Group
ETH Zürich

http://pulp-platform.org @pulp_platform

ETH zürich

How did we start in 2013?

- **We wanted to design energy efficient computing systems**
  - Equally efficient for IoT and HPC over a wide range
- **Key points**
  - Parallel processing
  - Near threshold computing
  - Efficient switching between operating modes
  - Making best use of technology
  - Heterogeneous acceleration
- **Parallel Ultra Low-Power (PULP) platform was born**

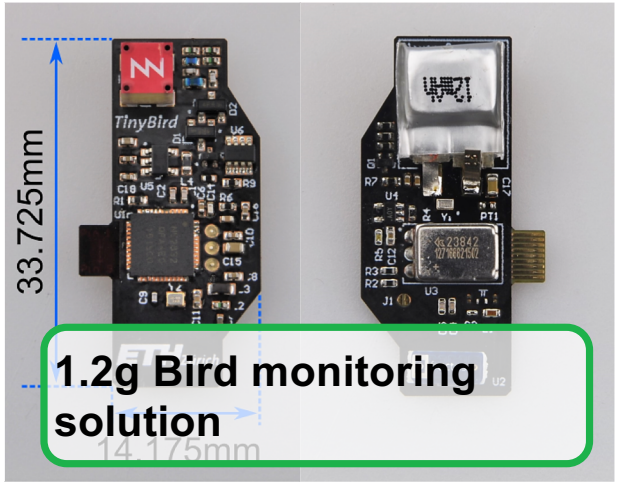


PULP

ETH zürich

IoT design @ Digital Circuits & Systems Group

64mW autonomous Nano Drone



1.2g Bird monitoring solution

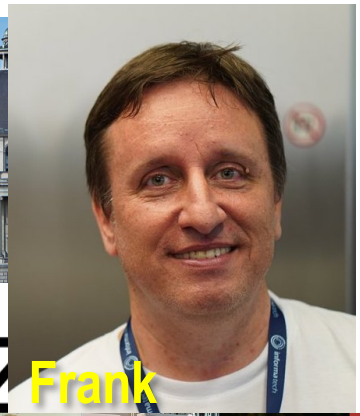
Sensor for Avalanche studies

Human touch powered sensor

LORA – GNSS localization, big & small

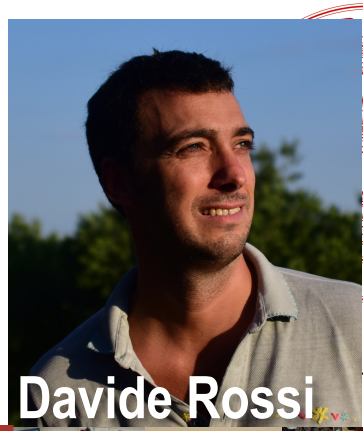
Wearable bio-signal processing systems

Who is behind PULP?



Frank

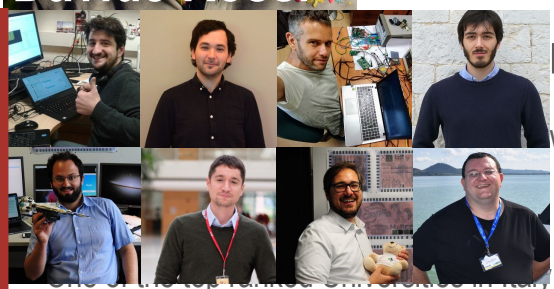
Prof. Luca Benini



Davide Rossi

In total about 60 people work on projects related to PULP in Zurich and Bologna

https://pulp-platform.org/team.html



- Chief Architect in STMicroelectronics (2009-2012)

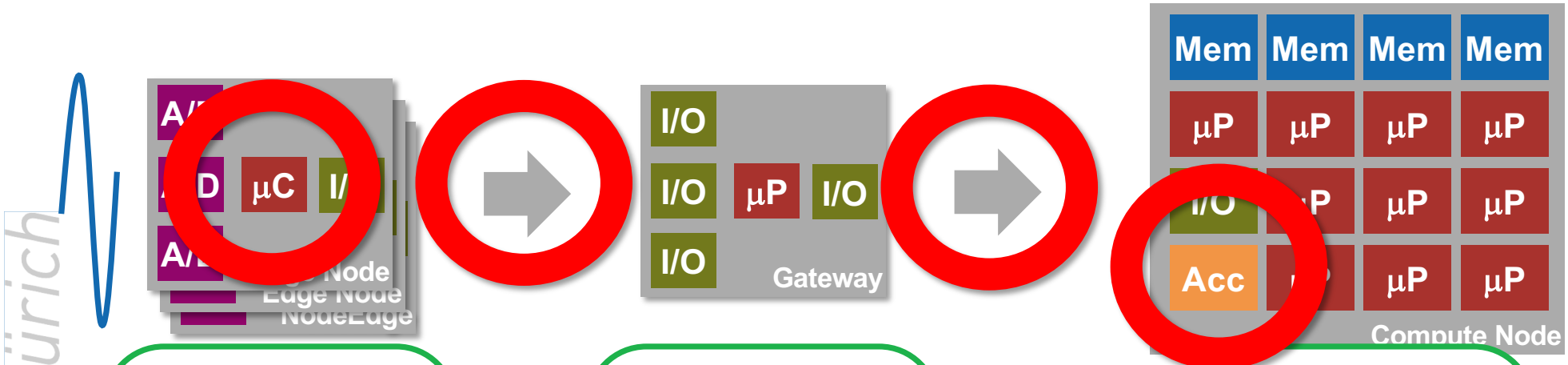
na

world

and Europe

Too much to do and not enough resources

SIGNAL → DATA → KNOWLEDGE



Energy efficient systems that can process more on edge nodes

Less data to transmit, large savings on energy needed for data transmissions

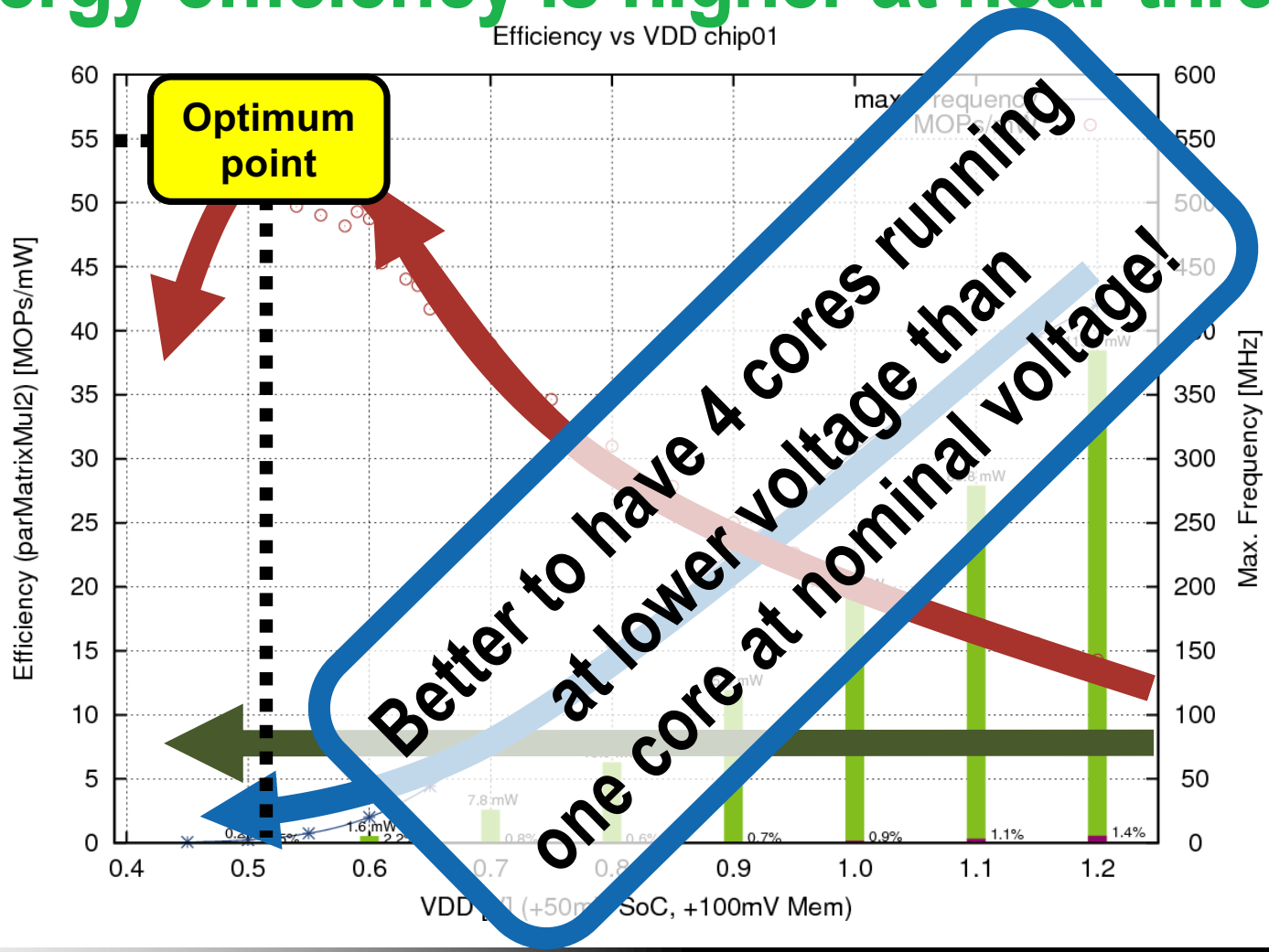
Accelerators for more efficient computing

PULP

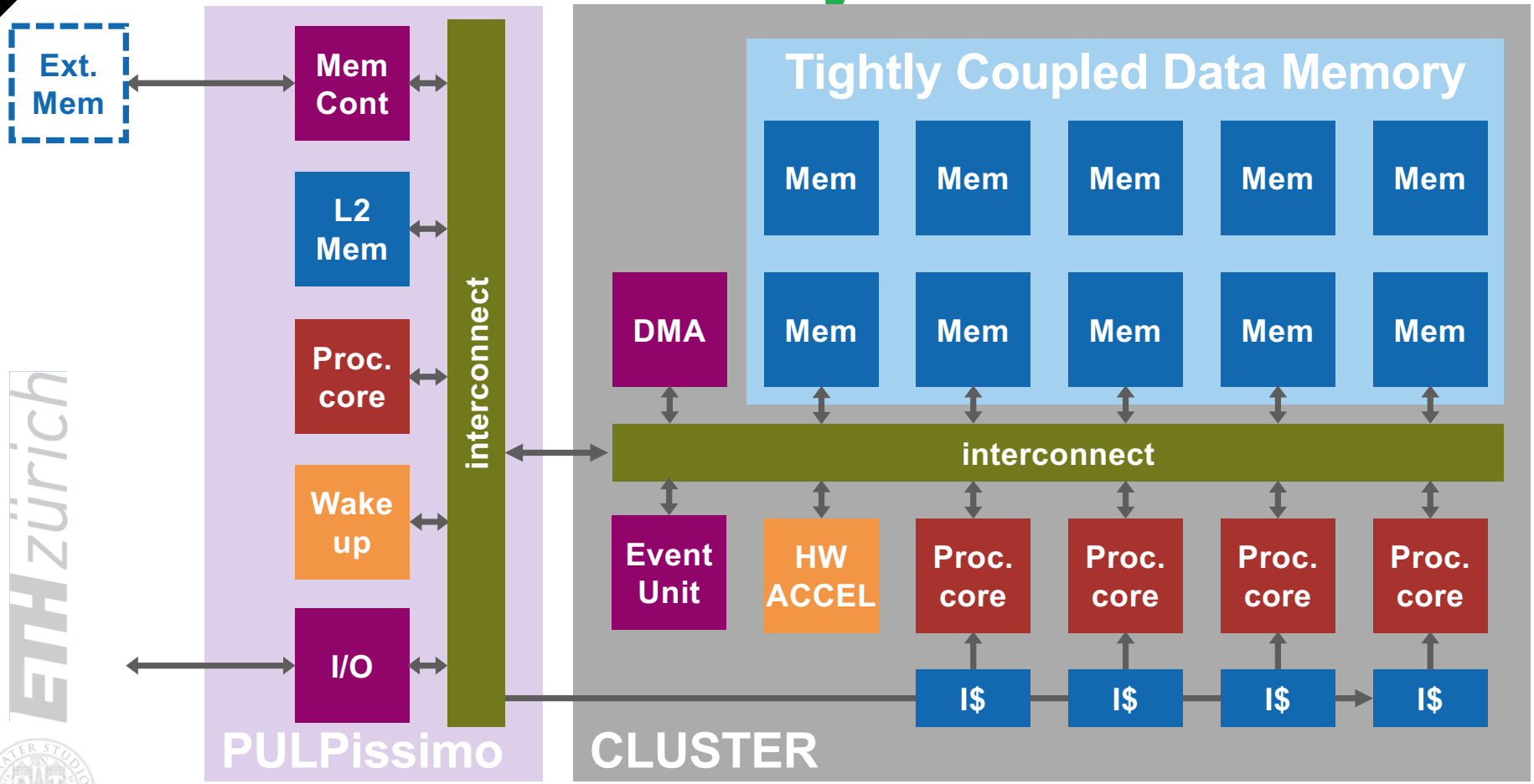
PULP

PULP

Energy efficiency is higher at near threshold

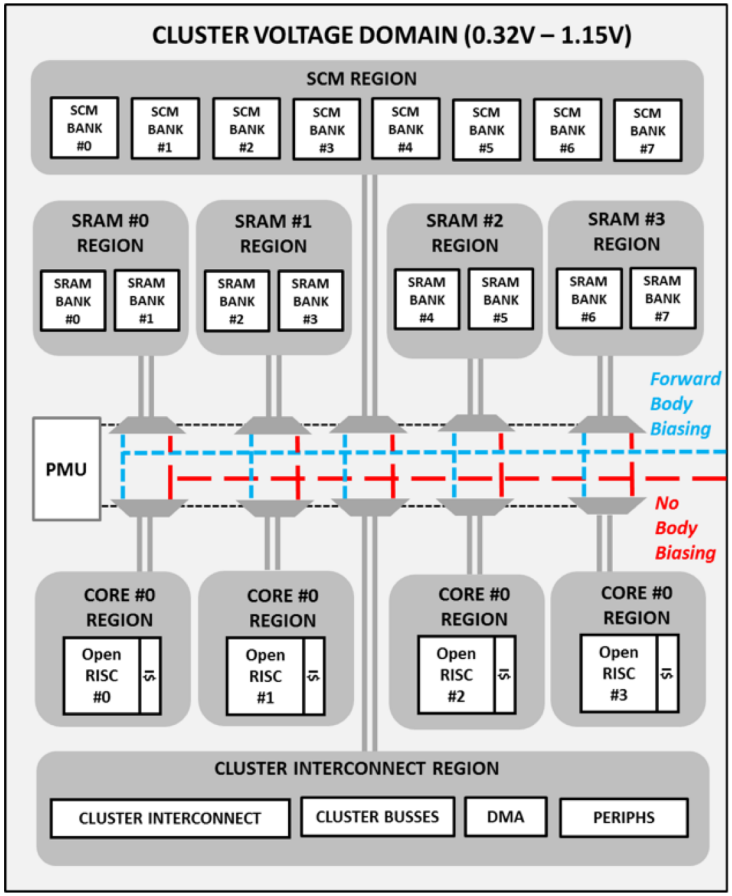


Cluster based PULP systems



ST28 FDSOI and GF22FDX designs with BB

- SoC partitioned in separate clock, power and body bias regions
- Cluster 1 Vdd, 10 BB regions
  - **Boost mode**: active + FBB
  - **Normal mode**: active + NO BB
  - **Idle mode**: clock gated + NO BB (in LVT) RBB (in RVT)
- SoC has 3 Vdd regions
  - Cluster, L2, Always on, IOs



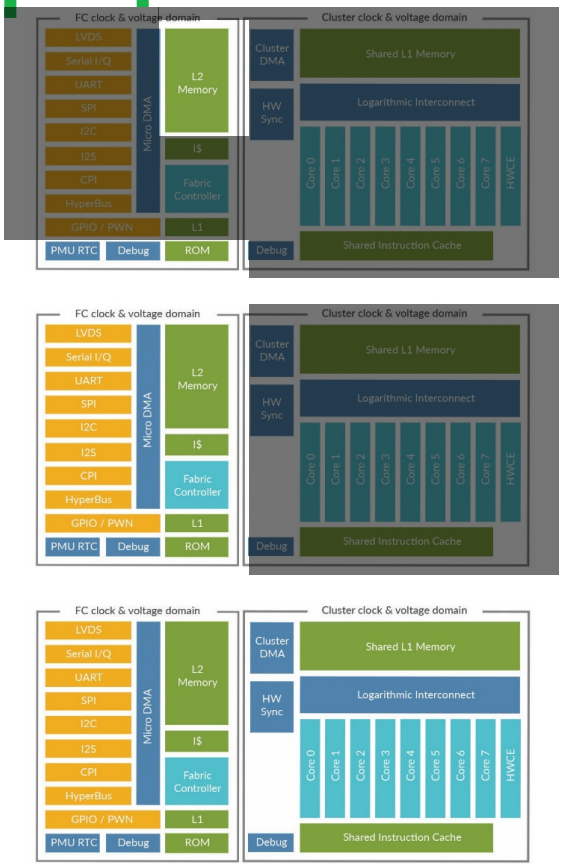
D. Rossi et. al., «A 60 GOPS/W, −1.8V to 0.9V body bias ULP cluster in 28nm UTBB FD-SOI technology», in Solid-State Electronics, 2016.

Scaling proportional to computing demand

Duty Cycling

Coarse Grain Classification

Full Blown Analysis



1 to 50 µW

MCU sleep mode

- ✓ Low quiescent LDO
- ✓ Real Time Clock 32kHz only
- ✓ L2 Memory partially retentive

0.5 to 5 mW

MCU active mode

- ✓ Embedded DC/DC, high current
- ✓ Voltage can dynamically change
- ✓ One clock gen active, frequency can dynamically change
- ✓ Systematic clock gating

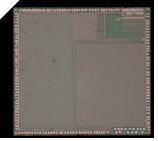
5 to 50 mW

MCU + Parallel processor active mode

- ✓ Embedded DC/DC, high current
- ✓ Voltage can dynamically change
- ✓ Two clock gen active, frequencies can dynamically change
- ✓ Systematic Clock Gating

Ultra fast switching time from one mode to another
Ultra fast voltage and frequency change time

→ Highly optimized system level power consumption



How PULP and RISC-V come together

- **Initially we did not want to design our own processors**
  - Wanted to use available processors (ARC, ARM..)
  - It proved difficult to design systems that we could share with our collaborators
- **Then we used OpenRISC cores (2013-2015)**
  - We had to completely redesign and optimize these cores
- **We moved to RISC-V starting in 2015**
  - Adapted the decoder of our optimized OpenRISC core
  - Make use of a growing SW development environment
  - ETH is one of the founding members of the RISC-V foundation

ETH zürich

Our research is not implementing RISC-V cores

- **We develop efficient programmable architectures**
  - Processor cores of various capabilities are required for that
- **We need efficient implementations of cores for our research**
  - To produce relevant results, our cores have to perform as good as other solutions
  - We ended up spending quite an effort to make sure they perform really well
- **Processor core alone is not enough**
  - You need peripherals, interconnect solutions, programming support...
- **PULP Platform provides us a playground for our research**
  - And we share it as open source

RISC-V cores developed by PULP team

32 bit

Low Cost Core

- **IBEX**
  - Zero-riscy
  - Micro-riscy
- **2 options**
  - RV32-ICM
  - RV32-EC

ARM Cortex-M0+

Maintained By LowRISC

Core with DSP support

- **RI5CY**
  - RV32-ICMXF
    - SIMD
    - HW loops
    - Post manipulation
    - Fixed point

ARM Cortex-M4F

Very Mature Core

Core for Streaming

- **Snitch**
  - RV32-IMAFD
  - Small int core
  - Big 64bit FPU
  - Extensions for streaming

novel

Brand New Core

64 bit

Linux capable Core

- **Ariane**
  - RV64-ICMAFD

ARM Cortex-A55

Frequent Updates

RI5CY ISA extensions for performance

```
for (i = 0; i < 100; i++)
    d[i] = a[i] + b[i];
```

Baseline

```
mv    x5, 0
mv    x4, 100
Lstart:
  lb     x2, 0(x10)
  lb     x3, 0(x11)
  addi   x10,x10, 1
  addi   x11,x11, 1
  add    x2, x3, x2
  sb     x2, 0(x12)
  addi   x4, x4, -1
  addi   x12,x12, 1
bne      x4, x5, Lstart
```

11 cycles/output

Auto-incr load/store

```
mv    x5, 0
mv    x4, 100
Lstart:
  lb     x2, 0(x10!)
  lb     x3, 0(x11!)
  addi   x4, x4, -1
  add    x2, x3, x2
  sb     x2, 0(x12!)
bne      x4, x5, Lstart
```

8 cycles/output

HW Loop

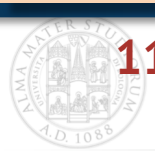
```
lp.setupi 100, Lend
  lb     x2, 0(x10!)
  lb     x3, 0(x11!)
  add    x2, x3, x2
Lend:  sb x2, 0(x12!)
```

5 cycles/output

Packed-SIMD

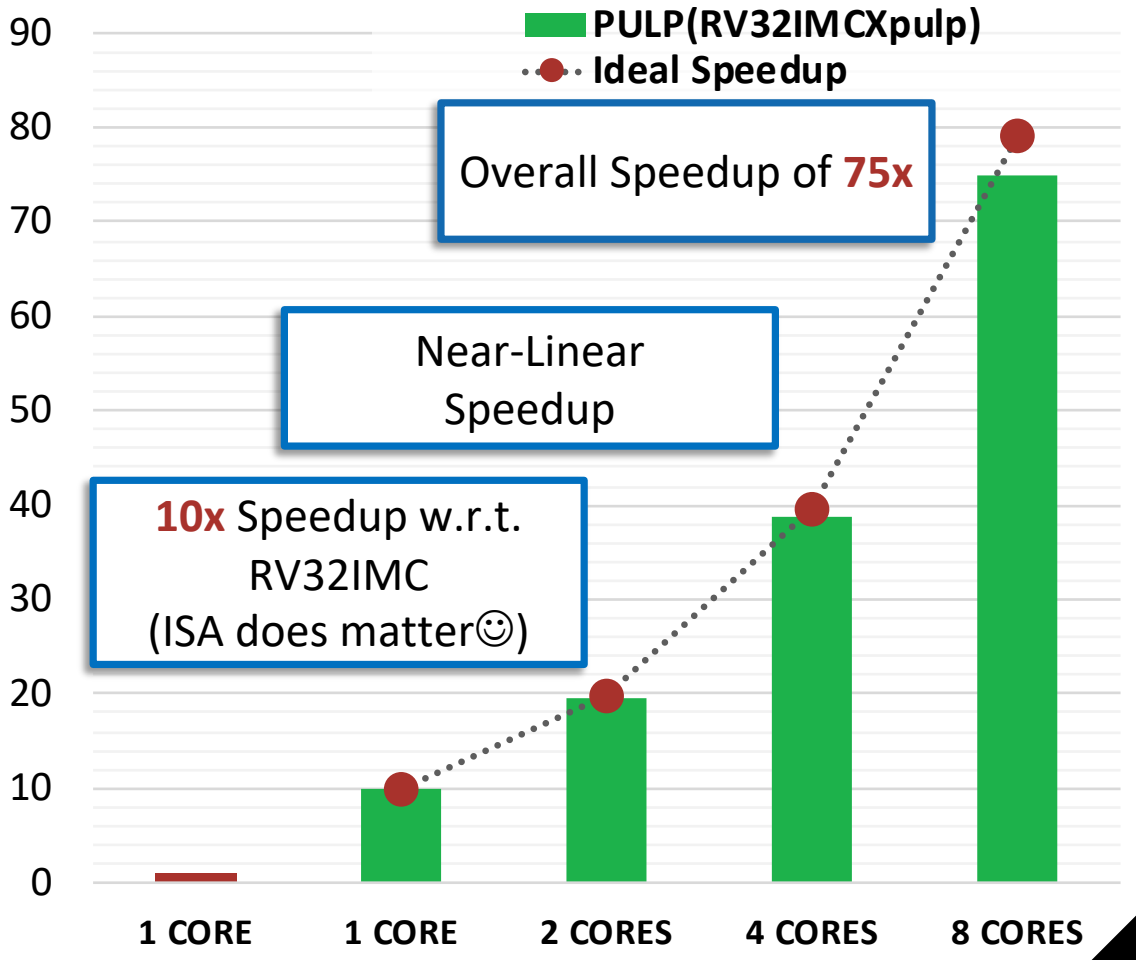
```
lp.setupi 25, Lend
  lw   x2, 0(x10!)
  lw   x3, 0(x11!)
  pv.add.b x2, x3, x2
Lend: sw x2, 0(x12!)
```

1,25 cycles/output

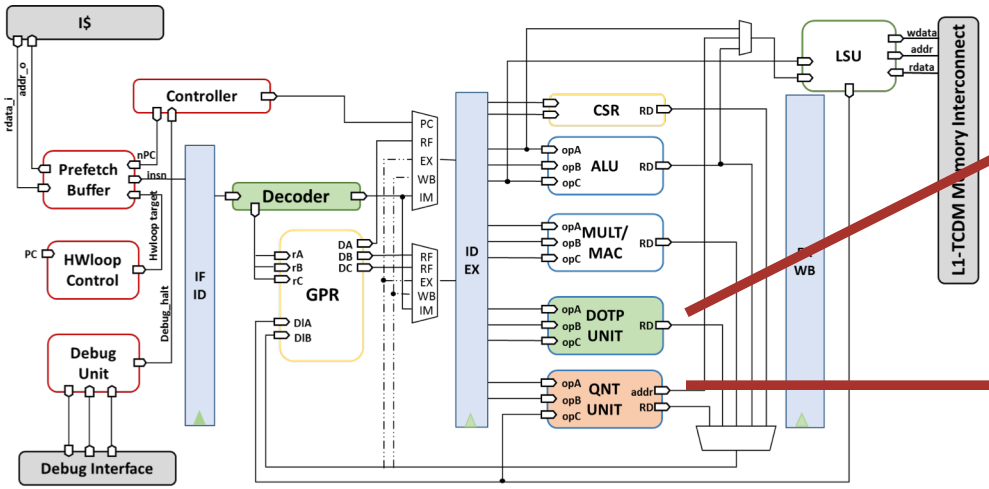


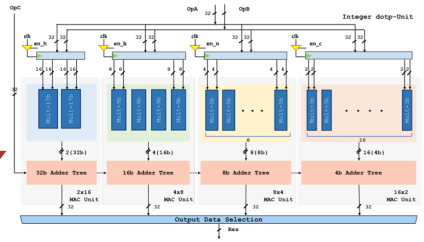
The extensions translate to real speed-ups

- 8-bit convolution
  - Open source DNN library
- 10x through xPULP
  - Extensions bring real speedup
- Near-linear speedup
  - Scales well for regular workloads.
- 75x overall gain

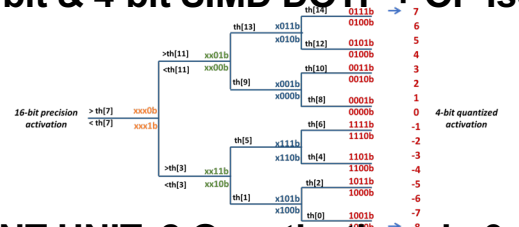


RISC-V ISA Extensions for extreme quantization



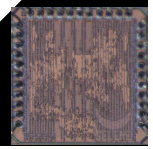


2-bit & 4-bit SIMD DOTP + OP Isolation



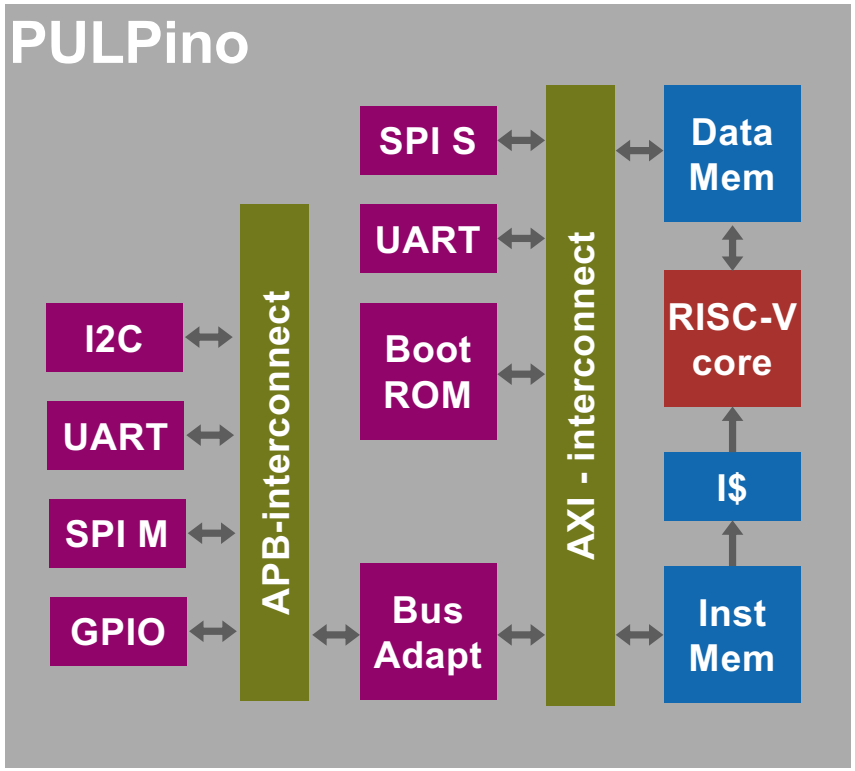
QNT UNIT: 2 Quantizations in 9 Cycles

- **Overheads (28nm FDX PULPissimo impl.):**
  - Area: ~11% (vs. Ri5CY)
  - Timing Overhead: negligible (integrated in PULPissimo)
  - 8-bit MatMul power overhead: 1.8% (integrated in PULPissimo)
  - GP-app power overhead: 3.5% (integrated in PULPissimo)



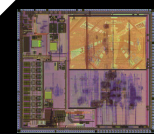
PULPino: Our first single core platform

- **Simple design**
  - Meant as a quick release
- **Separate data and inst. mem**
  - Makes it easy in HW
  - Not meant as a Harvard arch.
- **Can use all our 32bit cores**
  - RI5CY, Zero/Micro-Riscy (Ibex)
- **Peripherals from other projects**
  - Any AXI and APB peripherals could be used



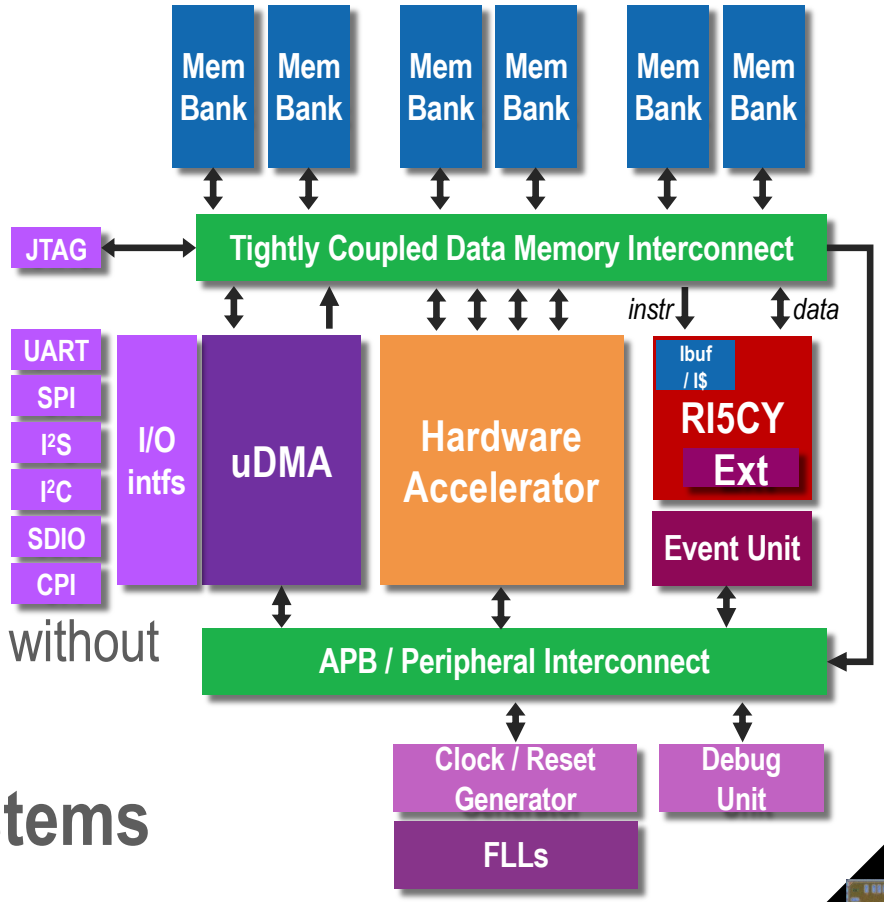
What kind of acceleration?

- *Standard peripheral talking over AXI/APB (standard)*
- *Instruction set extensions (already discussed)*
- **Shared functional units**
  - Amortizes expensive extensions (FPU/DIV) between multiple units
- **Shared memory accelerators**
  - Our bread and butter, PULPopen, NTX
- **Cluster as an accelerator**
  - HERO, BigPULP, etc

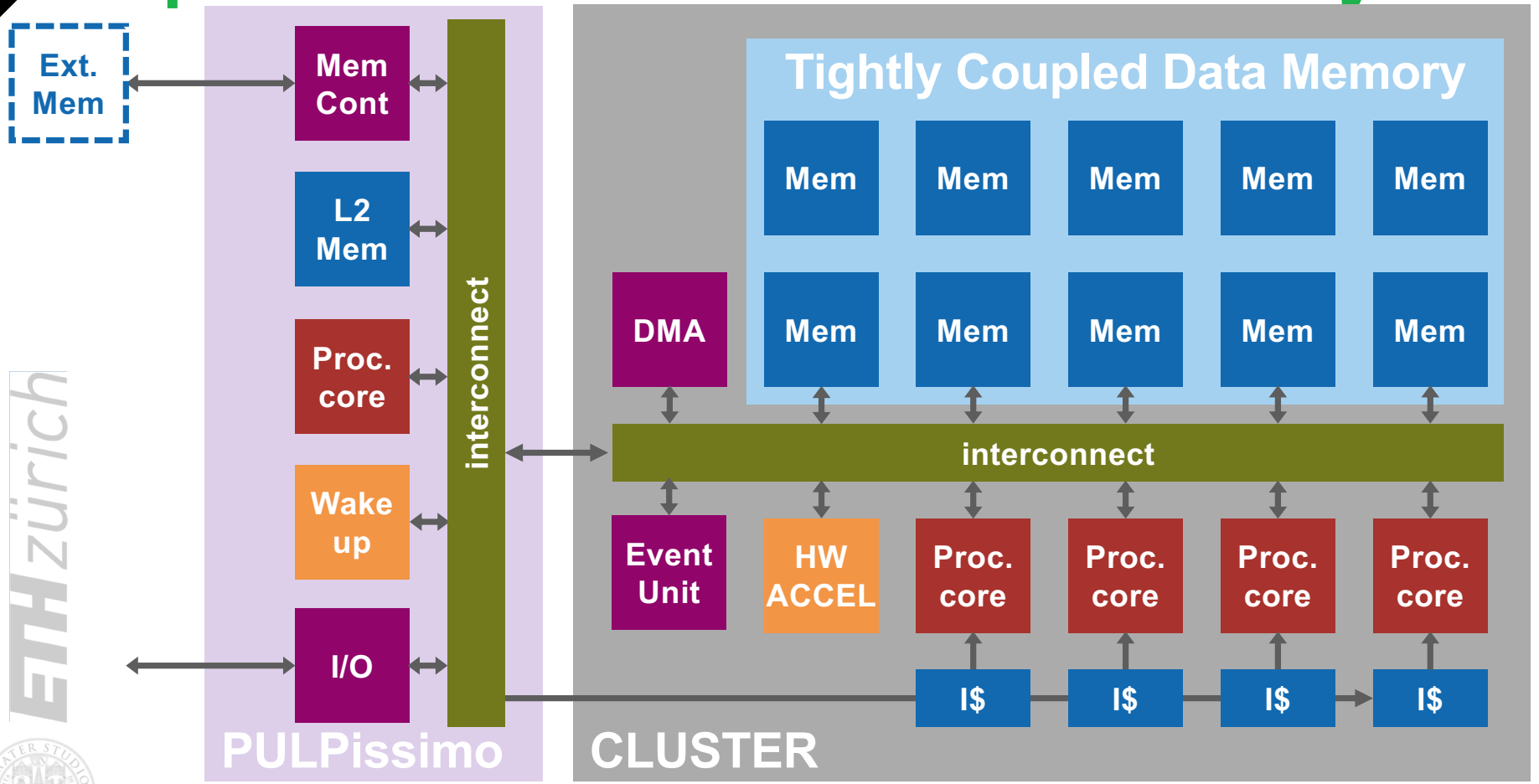


PULPissimo: The better single core platform

- **Shared memory**
  - Unified Data/Instruction Memory
- **Support for Accelerators**
  - Direct shared memory access
  - Programmed through APB bus
- **uDMA for I/O subsystem**
  - Can copy data directly from I/O to memory without involving the core
- **Used as controller in larger systems**

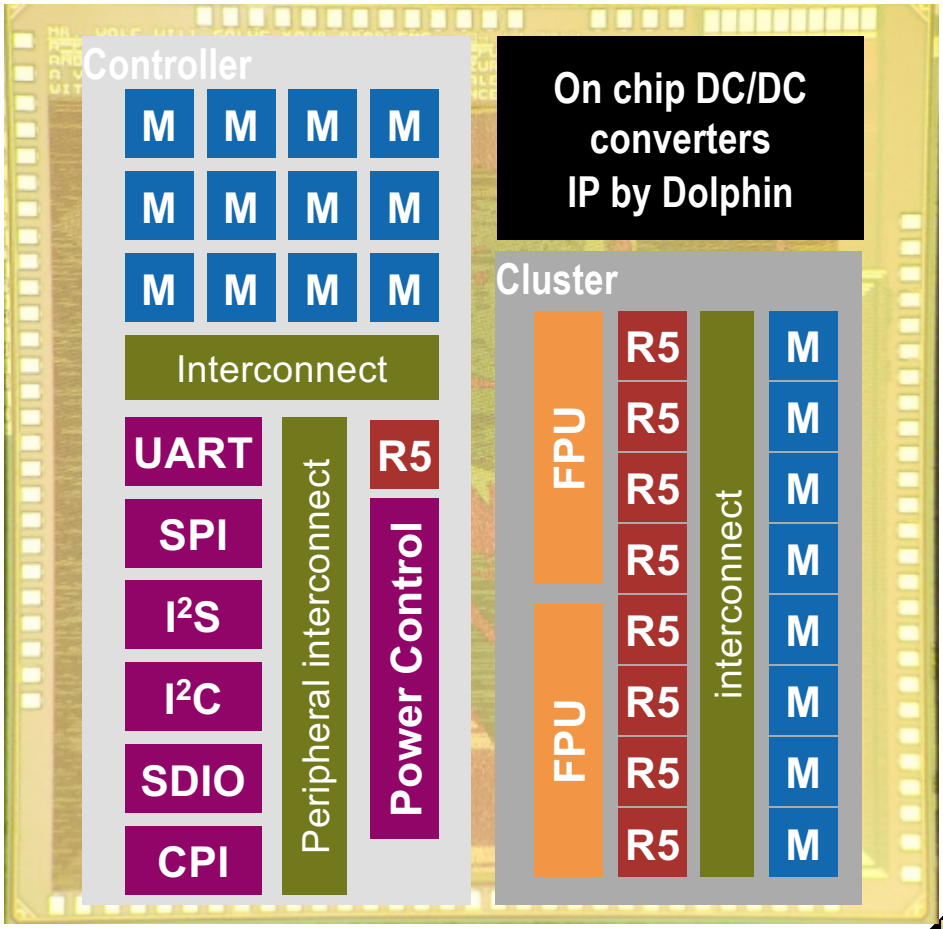


Open PULP: our main cluster based system



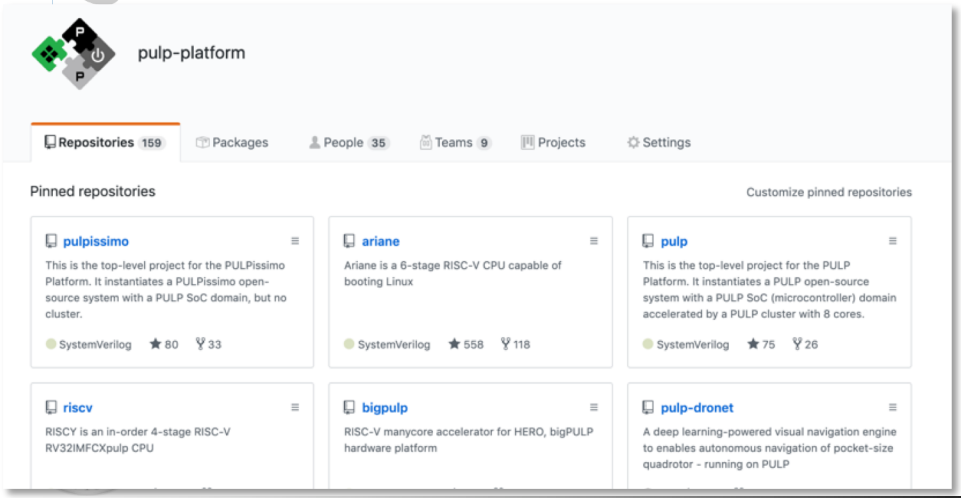
Mr. Wolf (TSMC 40): 8+1 core IoT Processor

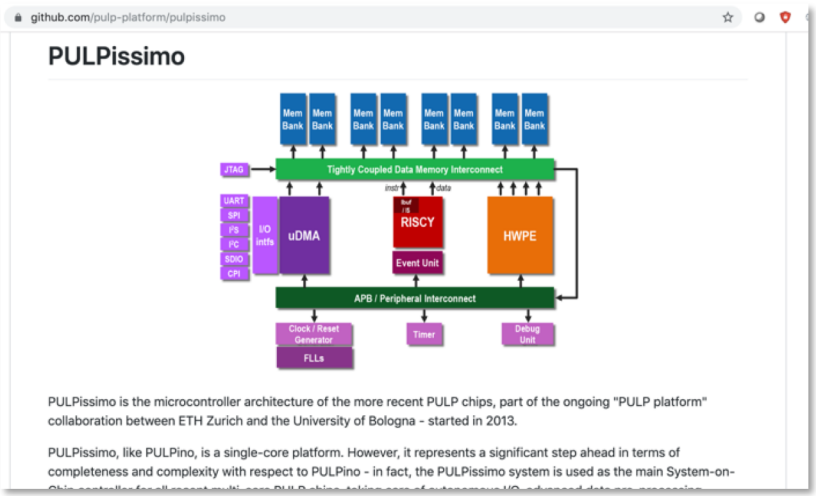
- One cluster with
  - 8 RISC-V cores
  - 2x shared FPU units
  - 64 kByte of TCDM
- One controller with
  - 512 kByte L2 RAM
  - Peripherals
- On chip voltage regulators
  - By Dolphin Integration

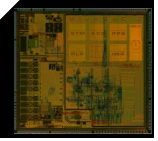


PULP uses a permissive open source license

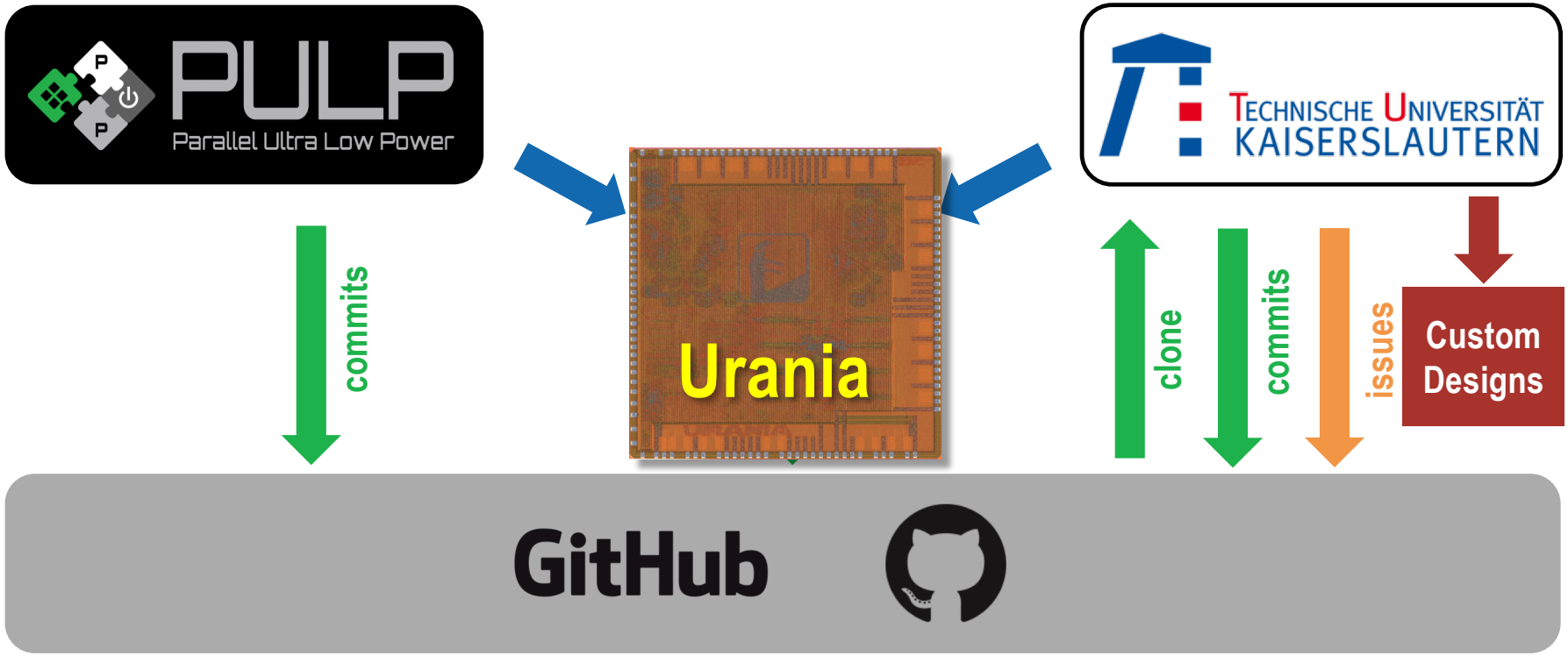
- All our development is on GitHub
  - HDL source code, testbenches, software development kit, virtual platform
  - https://github.com/pulp-platform
- PULP is released under the permissive Solderpad license
  - Allows anyone to use, change, and make products without restrictions.



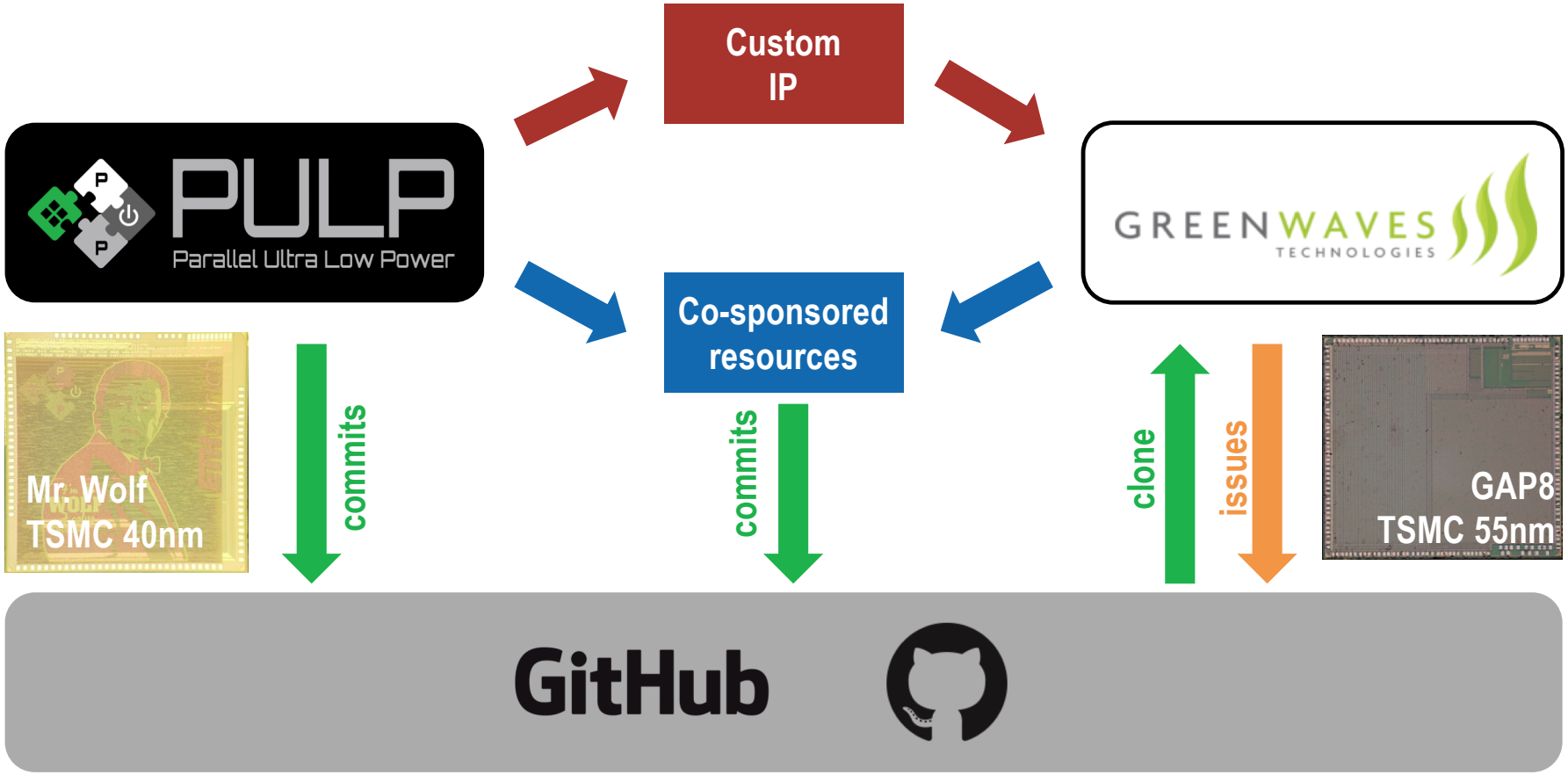




Open source collaboration scheme explained

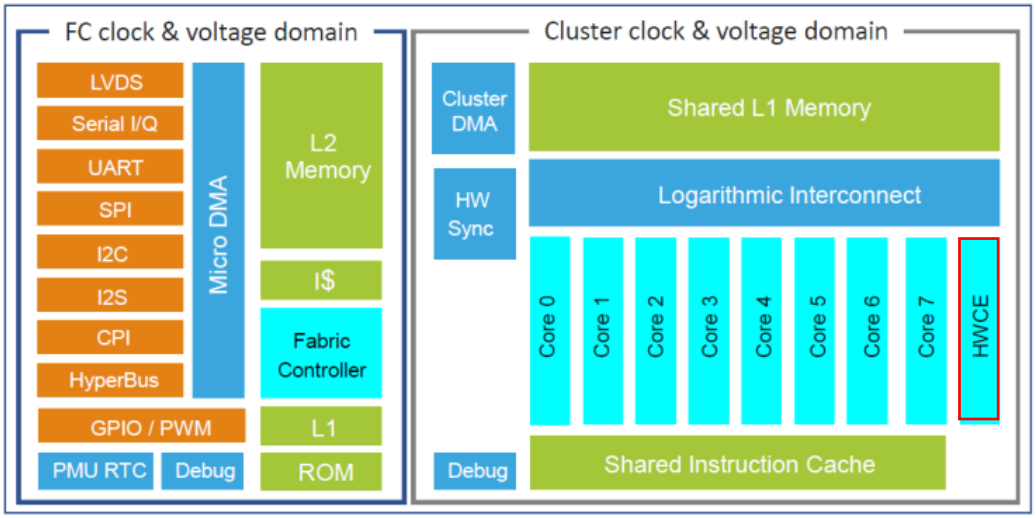


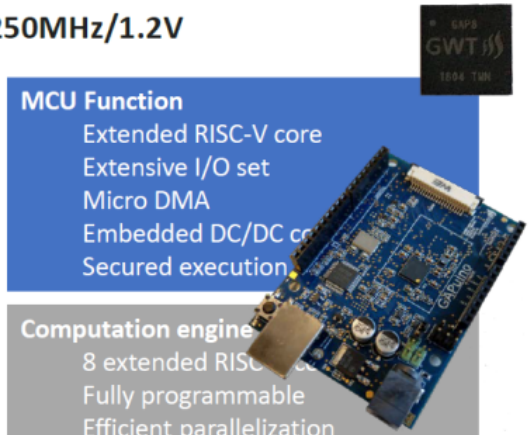
Open source collaboration scheme explained



PULP cluster+MCU+HWCE: GWT's GAP8

Two independent clock and voltage domains, from 0-133MHz/1V up to 0-250MHz/1.2V





What	Freq MHz	Exec Time ms	Cycles	Power mW
40nm Dual Issue MCU	216	99.1	21 400 000	60
GAP8 @1.0V	15.4	99.1	1 500 000	3.7
GAP8 @1.2V	175	8.7	1 500 000	70
GAP8 @1.0V w HWCE	4.7	99.1	460 000	0.8

11 X

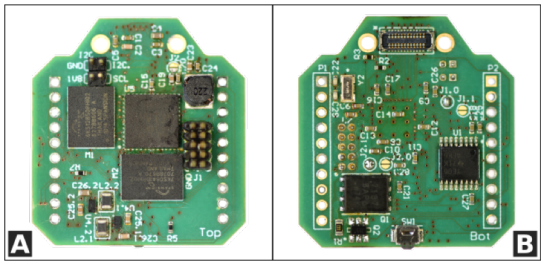
16 X

4x More efficiency at less than 10% area cost

Complete Application: DroNET on NanoDrone

Pluggable PCB: PULP-Shield

- **~5g**, 30×28mm
- GAP8 SoC
- 8 MB HDRAM
- 16 MB HFlash
- QVGA ULP HiMax camera
- Crazyflie 2.0 nano-drone (27g)



Only onboard computation for autonomous flight + obstacle avoidance
no human operator, no ad-hoc external signals, and no remote base-station!

PULP has released a large number of IPs

RISC-V Cores

RI5CY	Ibex	Snitch	Ariane + Ara
32b	32b	32b	64b

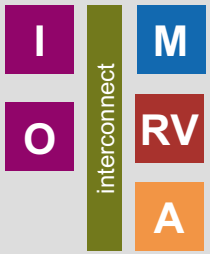
Peripherals

JTAG	SPI
UART	I2S
DMA	GPIO

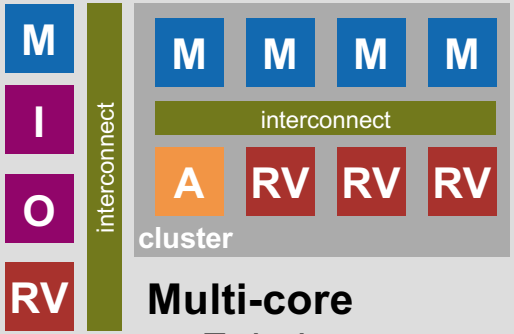
Interconnect

- Logarithmic interconnect
- APB – Peripheral Bus
- AXI4 – Interconnect

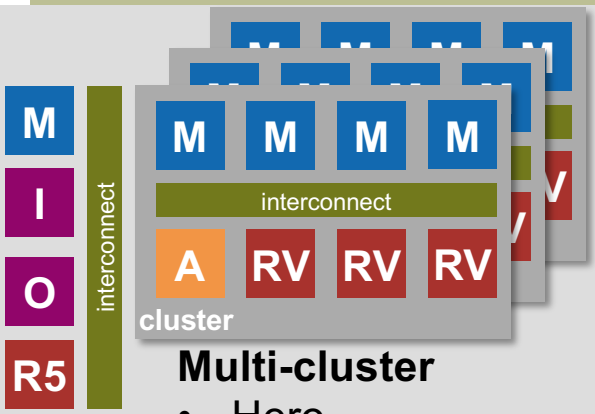
Platforms



Single Core
- PULPino
- PULPissimo



Multi-core
- Fulmine
- Mr. Wolf

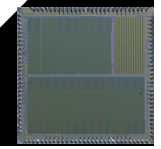


Multi-cluster
- Hero
- Open Piton

IOT → HPC

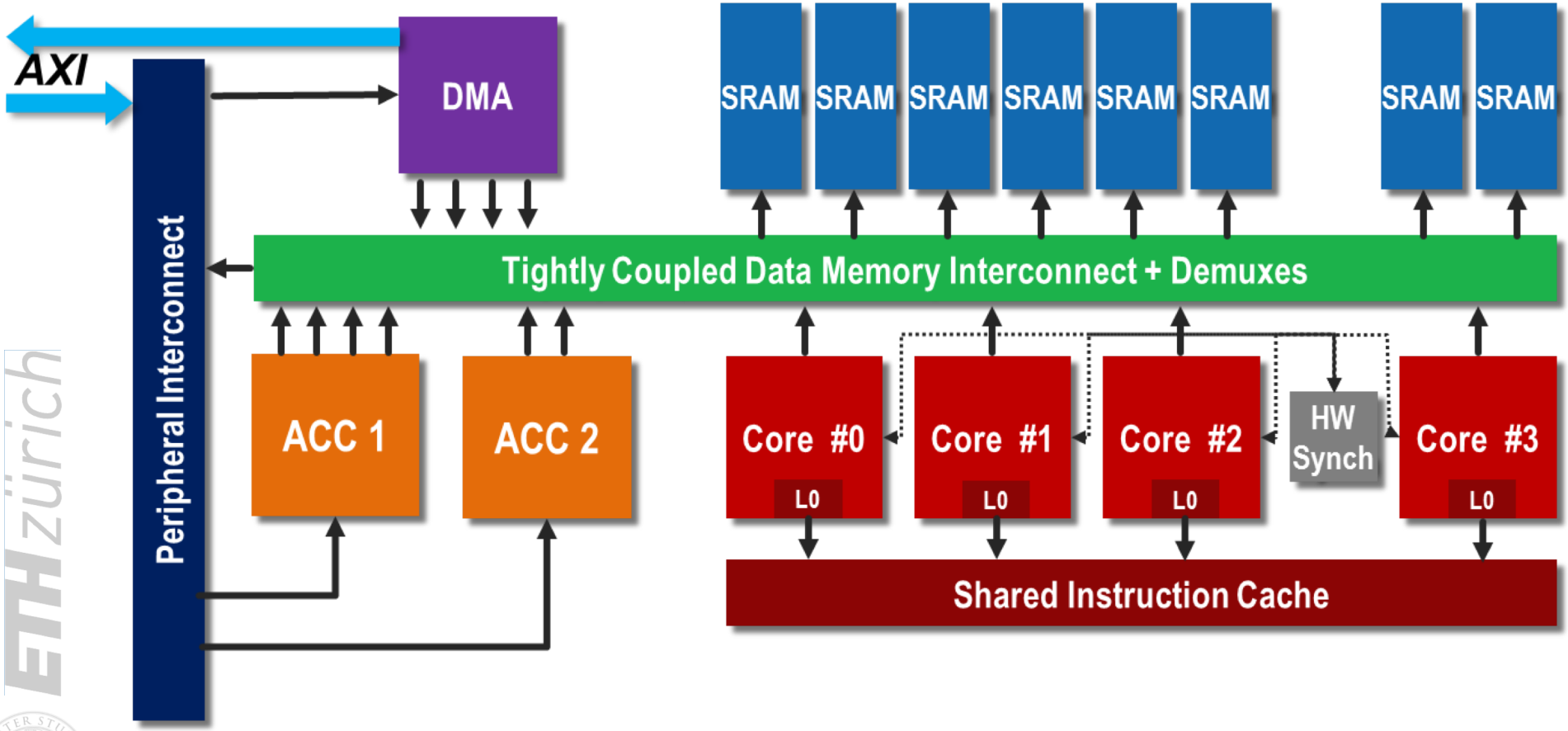
Accelerators

HWCE (convolution)	Neurostream (ML)	HWCrypt (crypto)	PULPO (1st ord. opt)



2018 Poseidon GF22FDX

Shared Memory Accelerators

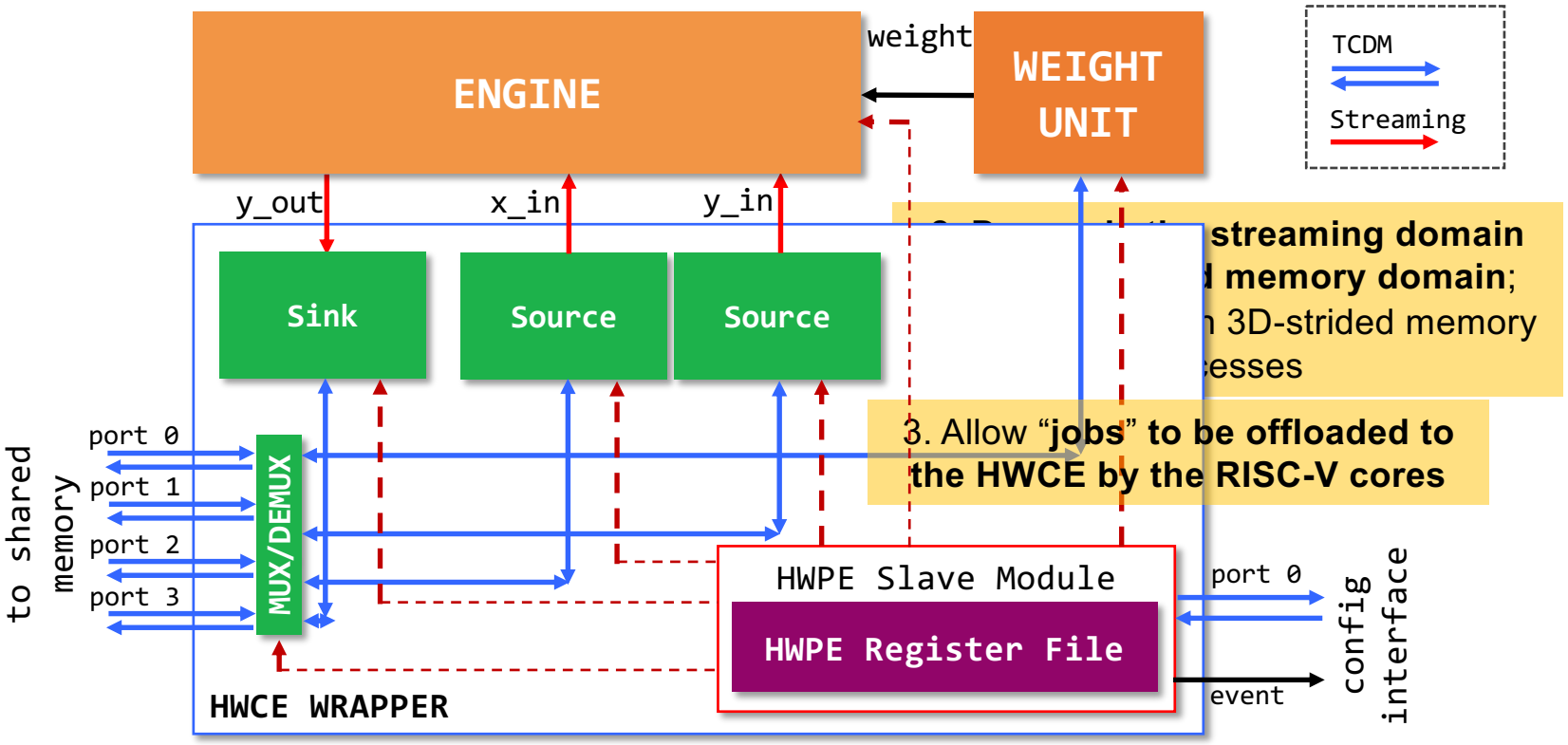


Accelerator: HW Convolution Engine

5. **Fine-grain clock gating** to minimize dynamic power

1. Perform convolve-accumulate in **streaming** fashion

4. **Weights** for each convolution filter are **stored privately**

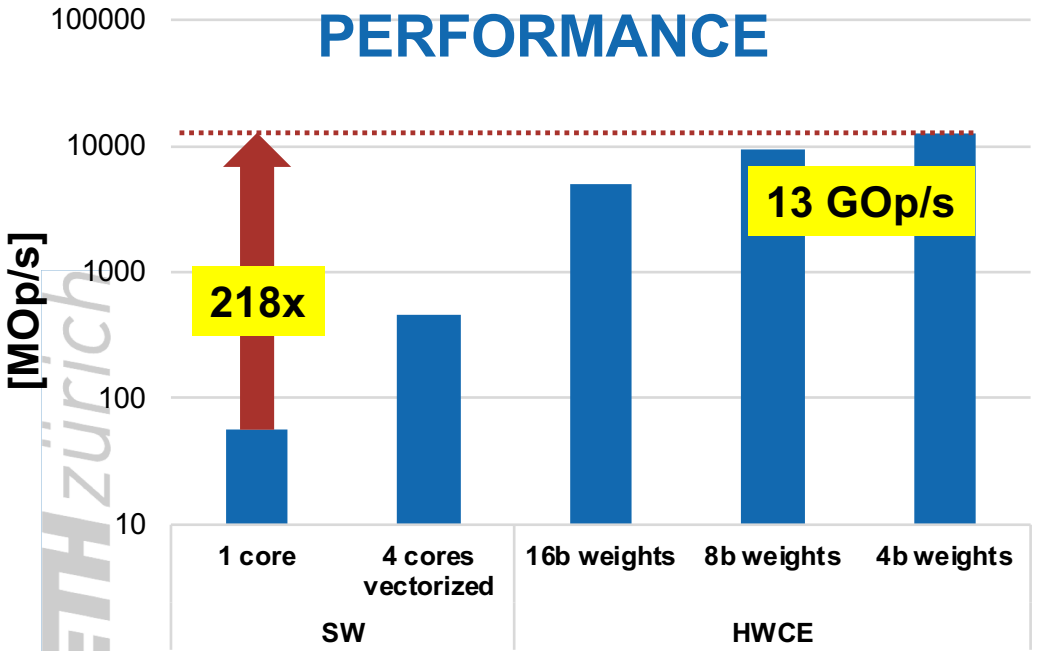


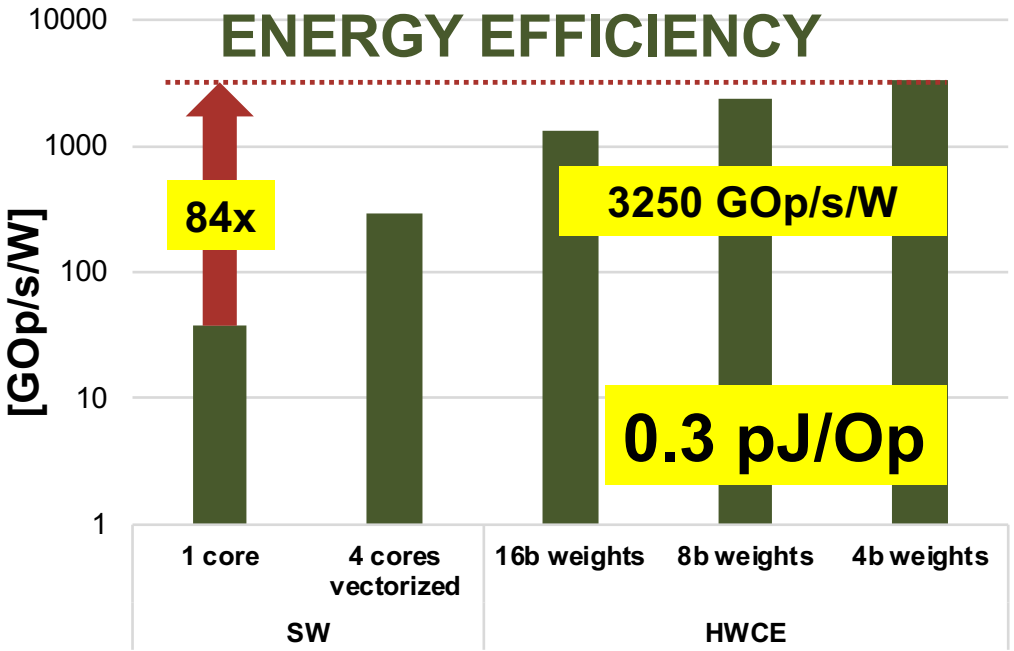
F. Conti and L. Benini, "A ultra-low-energy convolution engine for fast brain-inspired vision in multicore clusters," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015, pp. 683-688.

HW Convolution Engine Performance

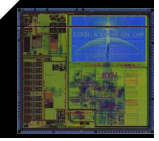
Cluster performance and energy efficiency on a 64x64 CNN layer (5x5 conv)

Scaled to ST FD-SOI 28nm @ Vdd=0.6V, f=115MHz

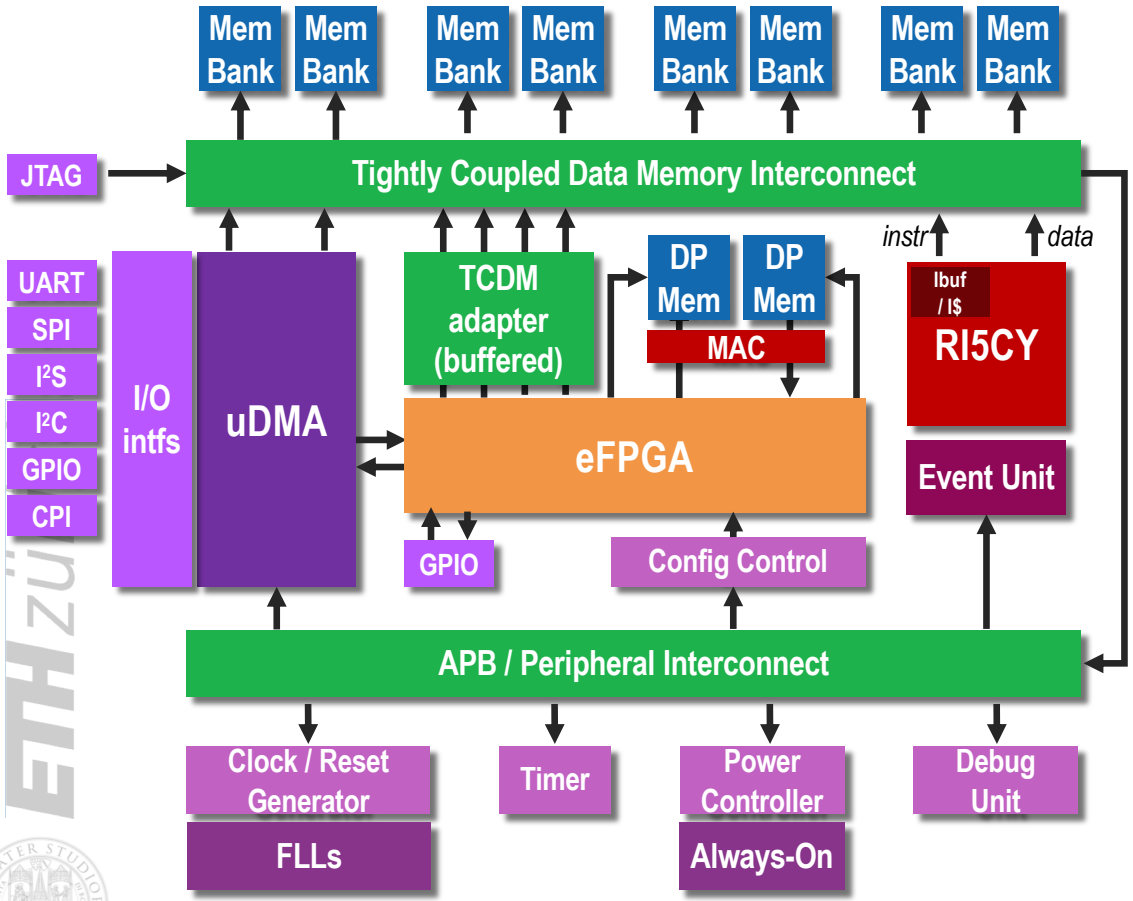




Now coming: HWCE 2.0 – improves scalability & flexibility @ 3TOPS/W

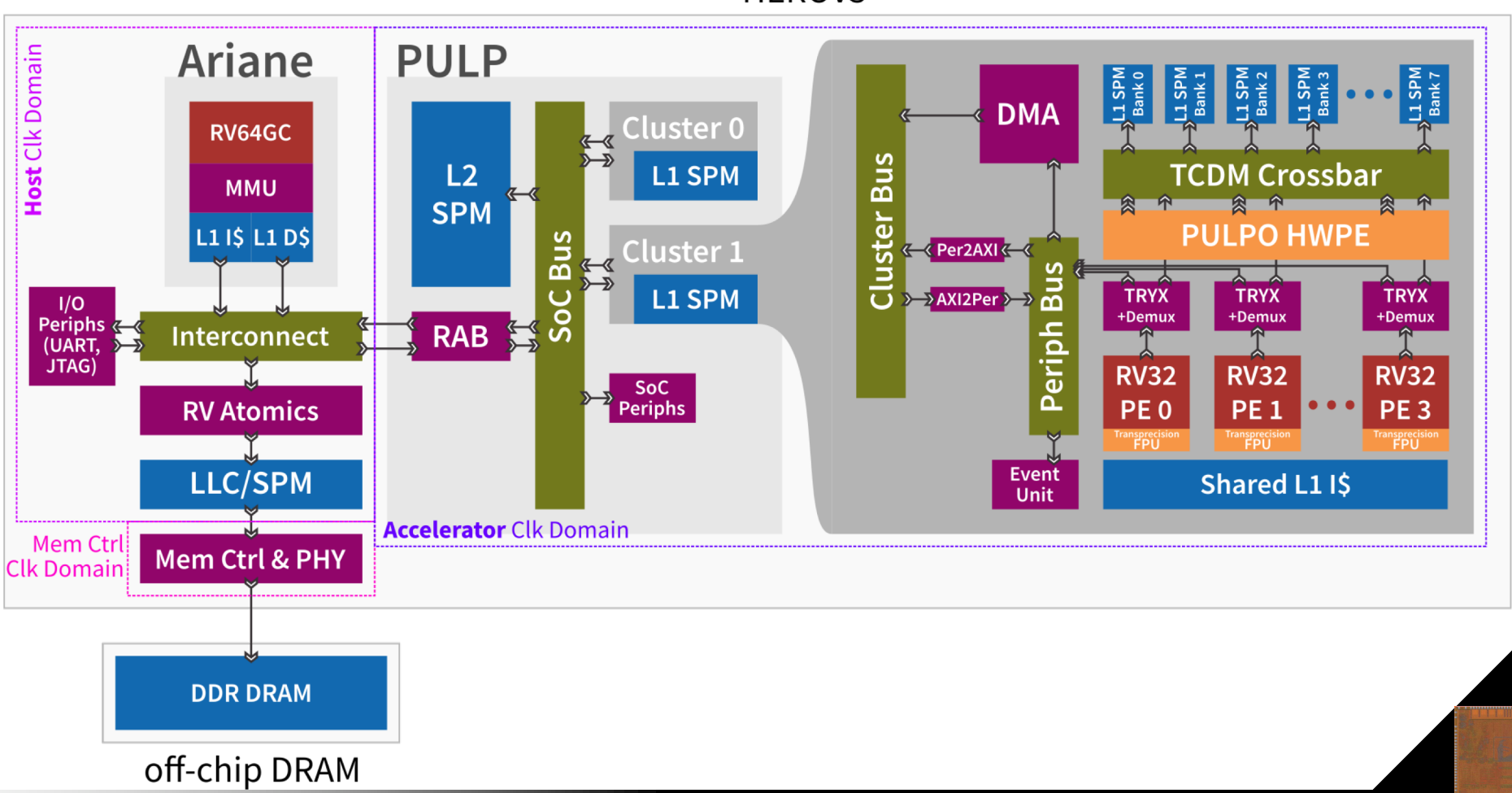


Arnold: a Collaboration with Quicklogic



- **Chip in 22nm FDX**
  - Combines e-FPGA (Quicklogic)
  - with PULPissimo (single core uC)
- **Multiple operation modes**
  - Configurable peripheral
  - Accelerator for core
  - Accelerator for independent I/O

Cluster as heterogenous accelerator (HERO)



FPGA Prototyping Platforms

Available:

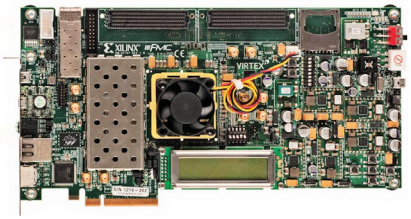
- **Digilent Genesys2**
  - $999 ($600 academic)
  - 1-2 cores at 66MHz

- **Xilinx VC707**
  - $3500
  - 1-4 cores at 60MHz
- **Digilent Nexys Video**
  - $500 ($250 academic)
  - 1 core at 30MHz

In progress:

- **Xilinx VCU118, BittWare XUPP3R**
  - $7000-8000
  - >100MHz

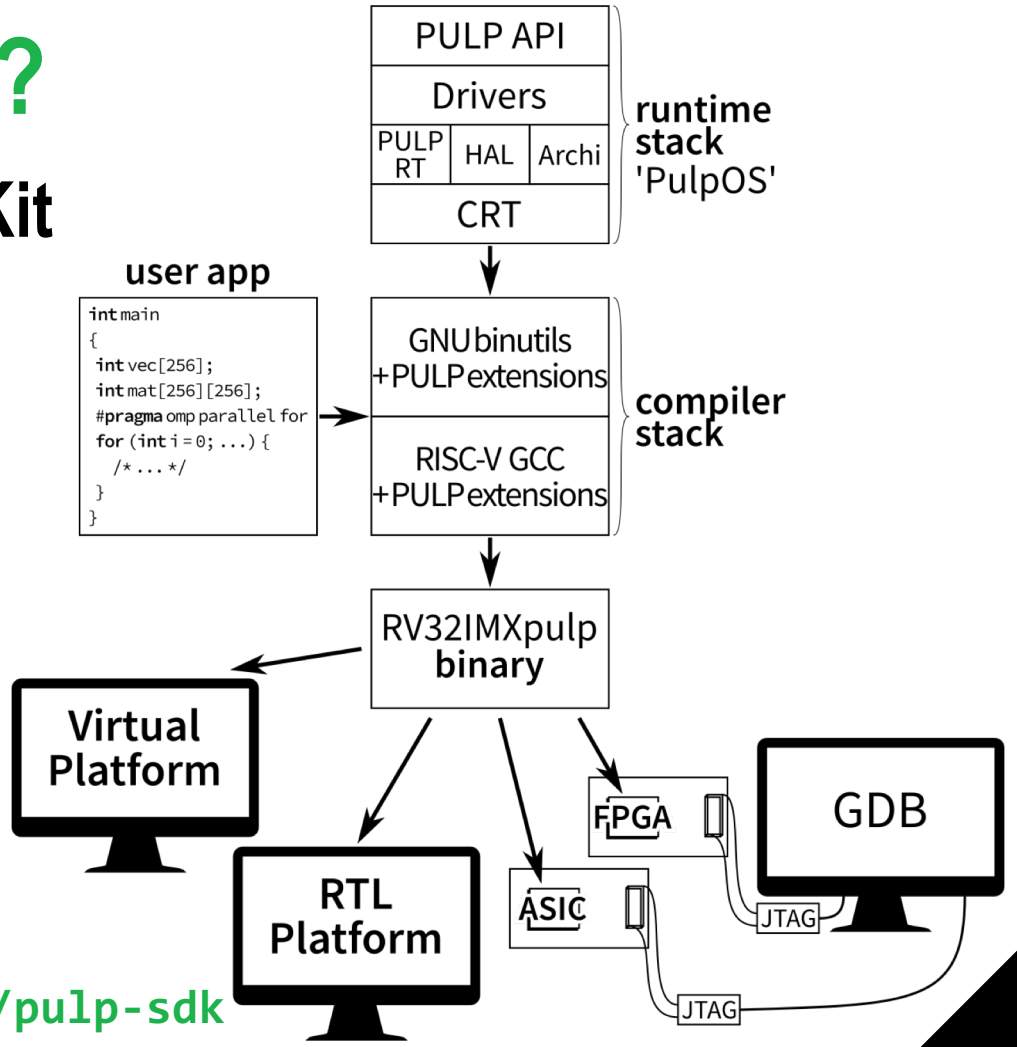
- **Amazon AWS F1**
  - Rent by the hour



How about SW support?

PULP Software Development Kit

- Package for compiling, running, debugging and profiling applications on PULP platforms
- Supports all recent and upcoming PULP chips: Mr.Wolf, GAP, Vega, ...
- Supports all targets: Virtual Platform, RTL platform, FPGA
- RISC-V GCC with support for PULP extensions
- Basic OpenMP support
- `https://github.com/pulp-platform/pulp-sdk`

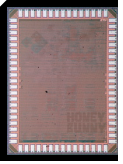


What is PULP doing for maintain our cores?

- **We (ETHZ and University of Bologna) are research groups**
  - Motivated to develop new architectures and systems
  - We needed efficient RISC-V cores (and peripherals) for our work
  - Not so good (or interested) in providing industrial level support for these cores
- **We need help to**

OPENHW GROUP
PROVEN PROCESSOR IP

lowRISC



Academic open source → Industrial open source

OPENHW GROUP — PROVEN PROCESSOR IP

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

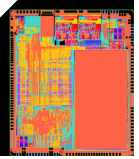
- **OpenHW Group** is a not-for-profit, global organization (EU,NA,Asia) driven by its members and individual contributors where HW and SW designers collaborate in the development of open-source cores, related IP, tools and SW such as the **Core-V** family of cores.
- **OpenHW Group** provides an infrastructure 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 MICROSYSTEMS · ETH zürich · EMBECOSM® · FUTUREWEI · GLOBAL FOUNDRIES · GREENWAVES TECHNOLOGIES · HUAWEI · imperas · metrics · MYTHIC · NXP · onespin · SILICON LABS · THALES · ultraSOC · VeriSilicon



A vertical, application-focused open approach

opentitan

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

ETH zürich

Giesecke+Devrient
Creating Confidence

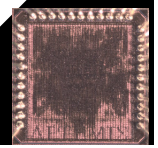
Google

lowRISC

nuvoTon

Western Digital

ETH zürich



Feel the momentum!

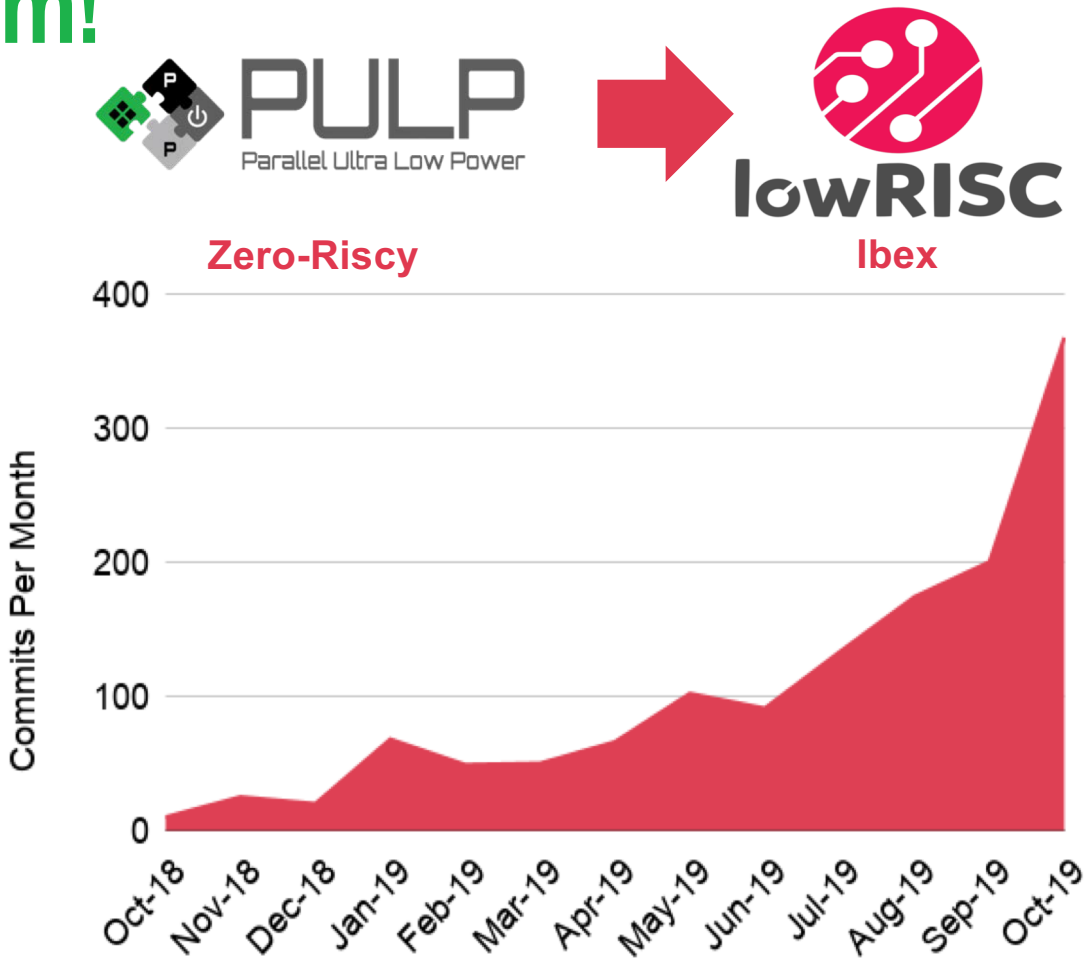
Ibex RISC-V core, flash interface, communications ports, cryptography accelerators, and more.

Vibrant repository

35+ Contributors

1300+ Contributions

470 GitHub Issues



PULP

Parallel Ultra Low Power

Luca Benini, Davide Rossi, Andrea Borghesi, Michele Magno, Simone Benatti, Francesco Conti, Francesco Beneventi, Daniele Palossi, Giuseppe Tagliavini, Antonio Pullini, Germain Haugou, Lukas Cavigelli, Manuele Rusci, Florian Glaser, Renzo Andri, Fabio Montagna, Bjoern Forsberg, Pasquale Davide Schiavone, Alfio Di Mauro, Victor Javier Kartsch Morinigo, Tommaso Polonelli, Fabian Schuiki, Stefan Mach, Andreas Kurth, Florian Zaruba, Manuel Eggimann, Philipp Mayer, Marco Guermandi, Xiaying Wang, Michael Hersche, Robert Balas, Antonio Mastrandrea, Matheus Cavalcante, Angelo Garofalo, Alessio Burrello, Gianna Paulin, Georg Rutishauser, Andrea Cossettini, Luca Bertaccini, Maxim Mattheeuws, Samuel Riedel, Sergei Vostrikov, Vlad Niculescu, Frank K. Gurkaynak,
and many more that we forgot to mention

http://pulp-platform.org @pulp_platform

FOSSistanbul will bring together, enthusiasts, members of industry and academia that are working on open source hardware design, in a lively and attractive city.

With keynotes by: Luca Benini, Nele Mentens, Onur Mutlu

Register for FREE

https://fossi-foundation.org/fossistanbul/

FOSSIstanbul
13-15 March 2020