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

Seven stories from seven years of PULP project

Luca Benini <lbenini@iis.ee.ethz.ch>
Frank Gurkaynak <kgf@ee.ethz.ch>

ETH zürich

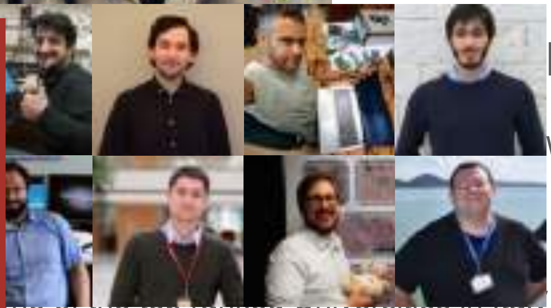
ALMA MATER STUDIORUM A.D. 1088

http://pulp-platform.org @pulp_platform https://www.youtube.com/pulp_platform

The PULP project in a nutshell

- **Started in 2013, after my P2012 experience in STM**
- **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

Who is behind PULP?



na

world

and

urope

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

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

Why Open Source Hardware

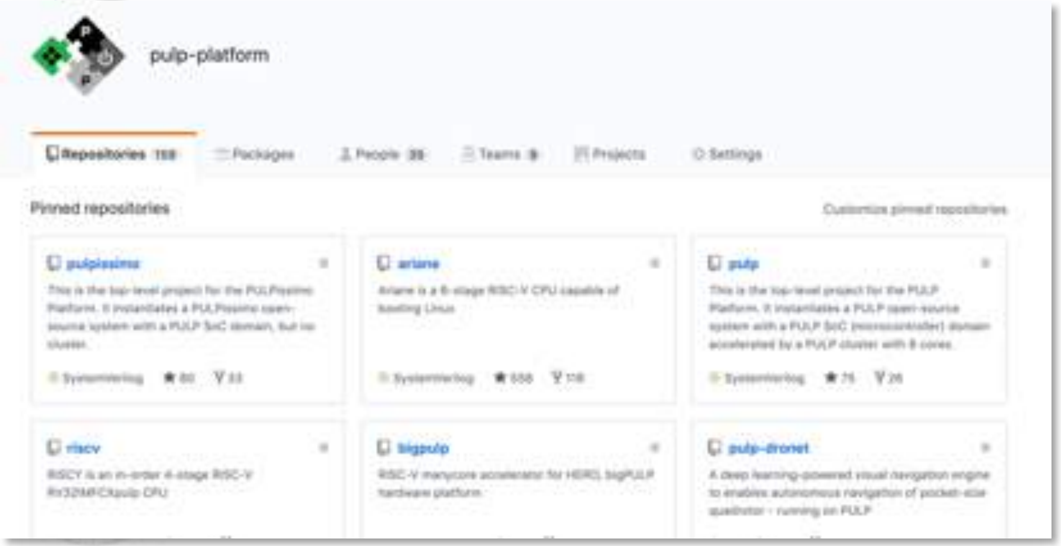
- It is a necessity
  - We can not afford to make everything ourselves, we need to collaborate
  - Makes it possible to work together quickly
  - Your results are more trustworthy, anybody can verify it!
- It works
  - We have actually more projects, and more funding due to open source activities
  - We were able to start many interesting and fruitful collaborations
- It helps others as well
  - Many companies, universities, individuals are using pieces of PULP, even commercially

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.



Pinned repositories

pulpissimo
This is the top-level project for the PULPissimo Platform. It instantiates a PULPissimo open-source system with a PULP SoC domain, but no cluster.
SystemVerilog ★ 80 ⑂ 33

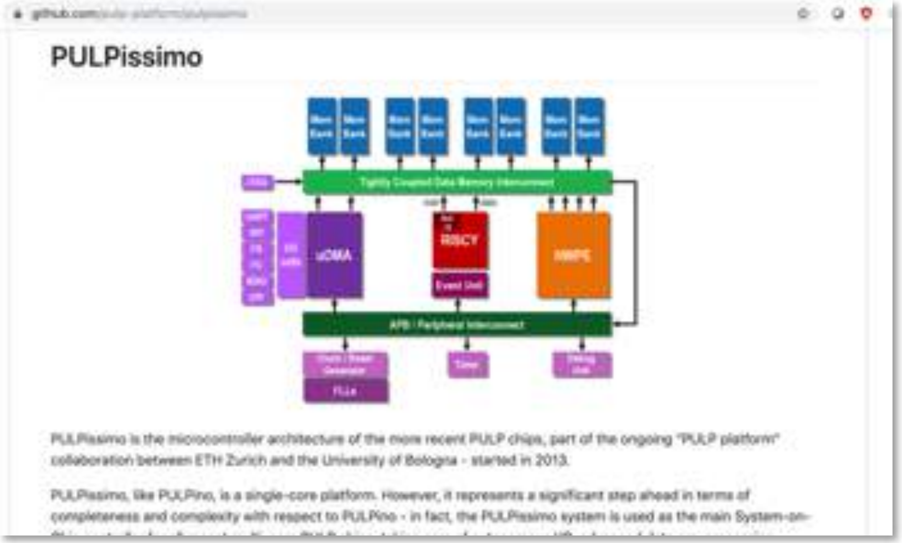
ariane
Ariane is a 6-stage RISC-V CPU capable of booting Linux
SystemVerilog ★ 558 ⑂ 116

pulp
This is the top-level project for the PULP Platform. It instantiates a PULP open-source system with a PULP SoC (microcontroller) domain accelerated by a PULP cluster with 8 cores.
SystemVerilog ★ 75 ⑂ 26

riscv
RISCY is an in-order 4-stage RISC-V RV32IMFCXpulp CPU

bigpulp
RISC-V manycore accelerator for HERO, bigPULP hardware platform

pulp-dronet
A deep learning-powered visual navigation engine to enables autonomous navigation of pocket-size quadrotor - running on PULP



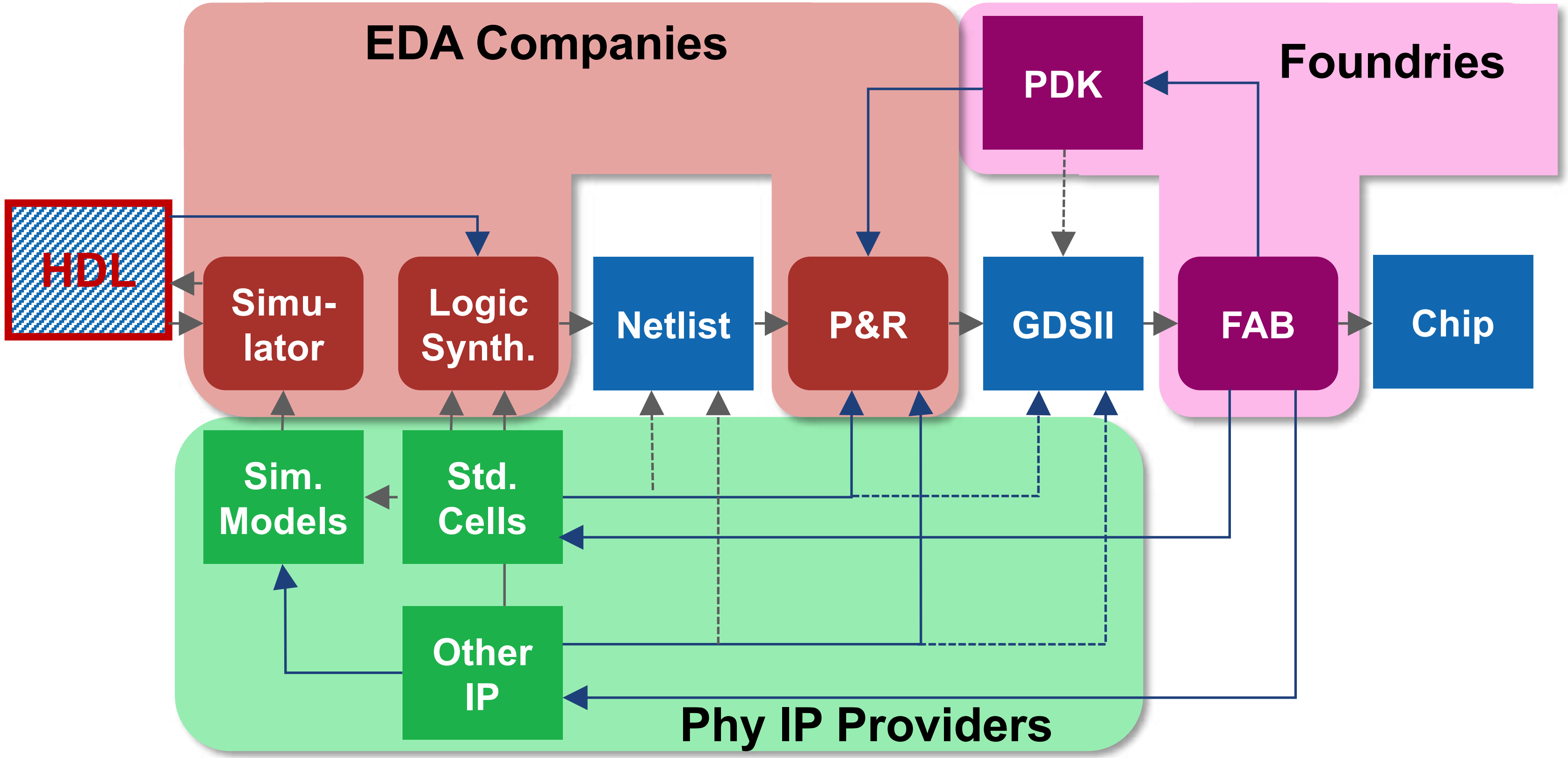
PULPissimo

PULPissimo is the microcontroller architecture of the more recent PULP chips, part of the ongoing "PULP platform" collaboration between ETH Zurich and the University of Bologna - started in 2013.

PULPissimo, like PULPino, is a single-core platform. However, it represents a significant step ahead in terms of completeness and complexity with respect to PULPino - in fact, the PULPissimo system is used as the main System-on-

Nice, but what exactly is “open” in OSCHW?

- Only the first stage of the silicon production pipeline
  → **RTL source code** (***permissive****, e.g. Apache is key for industrial adoption)
- Later stages contain closed IP of various actors → not open source by default



Cadence license for academic usage forbids ***permissive*** open sourcing of designs made with CDNS tools unless a ***reciprocal**** license is used

“See: https://cern-ohl.web.cern.ch/ (CERN-OHL-S, -W, -P)

PULP has released a large number of IPs

RISC-V Cores

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

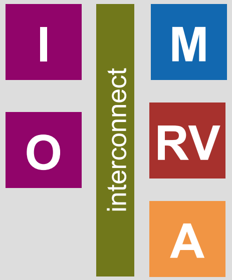
Peripherals

- JTAG
- SPI
- UART
- I2S
- DMA
- GPIO

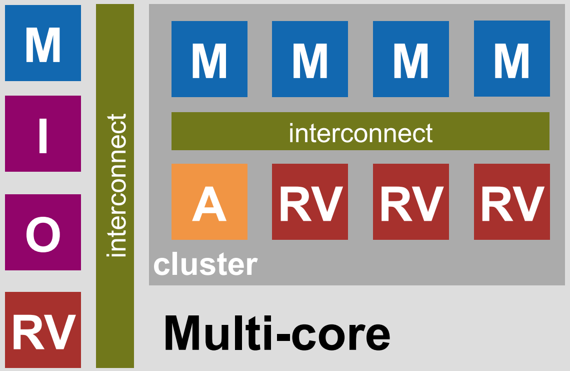
Interconnect

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

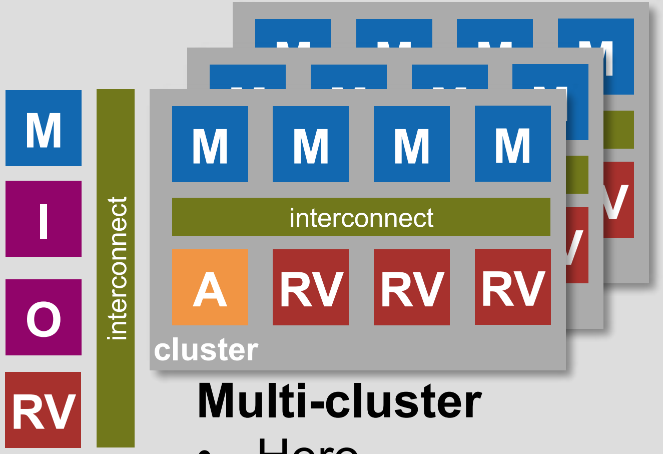
Platforms



Single Core
- PULPino
- PULPissimo



Multi-core
- PULP-open



Multi-cluster
- Hero
- Manticore

IOT → HPC

Accelerators

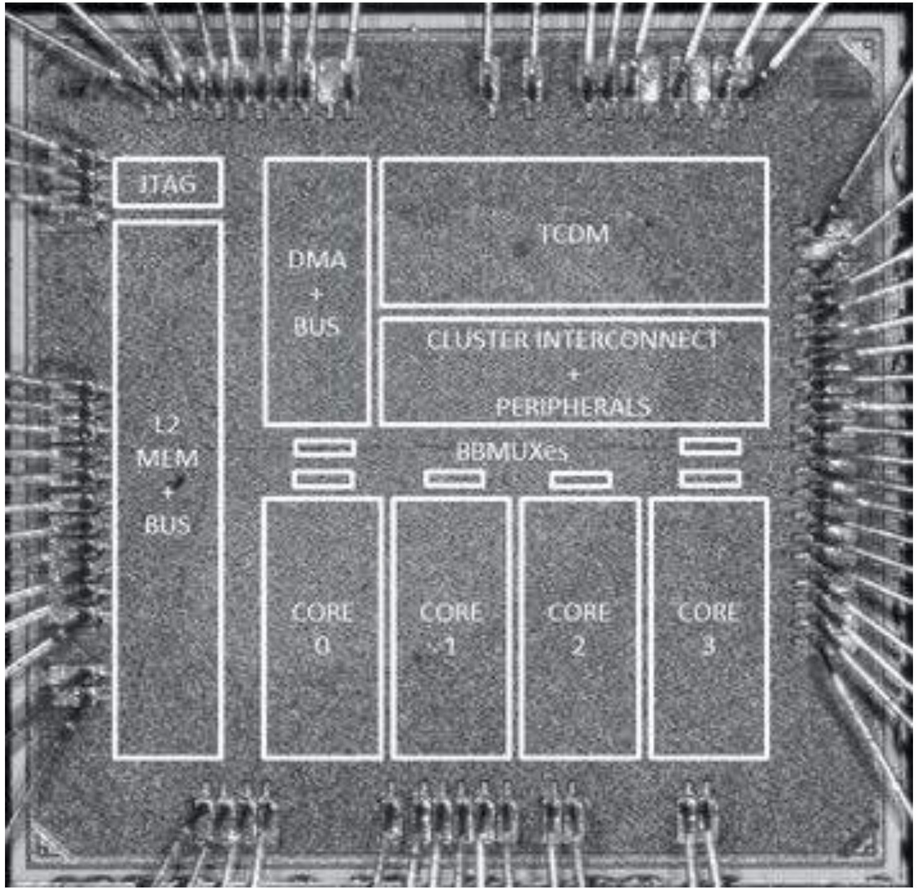
- HWCE (convolution)
- Neurostream (ML)
- RBE
- Hypno

How open source HW shaped our work

- I have chosen seven (out of 40) chips we had as part of PULP
  - Tried to pick from different times, different uses and different technologies
- Each chip has its own story
  - I will concentrate mainly on the open source aspects
- In addition to their technical results, each chip taught us
  - Collaboration models
  - What works what does not
- Most of what I talk is available as open source
  - We will briefly talk about what can and can not be open sourced as well

#1 - Pulpv1 (2013) – The first chip

- **Our first complete PULP chip**
  - 4x OpenRISC cores
  - STM 28FDSOI technology (RBB)
  - Explores body-biasing
- **Collaboration with STM (France)**
  - They needed a complete system demo (more than ring oscillators)
  - Demo for technology capabilities
- **Meant for an IC tester**
  - Almost no I/Os



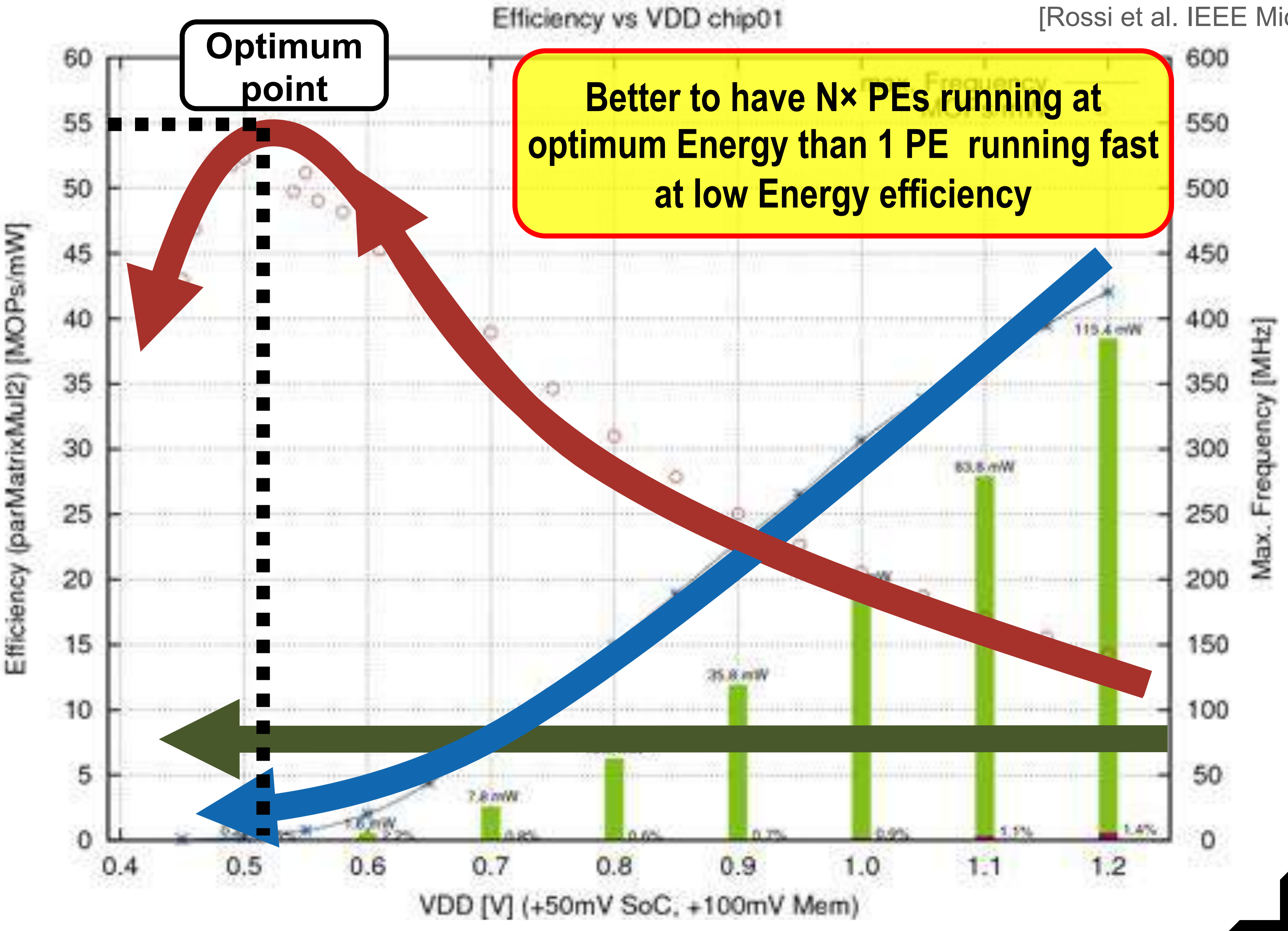
Davide Rossi, Antonio Pullini, Igor Loi, Michael Gautschi, Frank K. Gurkaynak, Andrea Bartolini, Philippe Flatresse, Luca Benini, "A 60 GOPS/W, -1.8 V to 0.9 V body bias ULP cluster in 28 nm UTBB FD-SOI technology", Journal of Solid-State Electronics, Volume 117, March 2016, Pages 170-184, DOI: 10.1016/j.sse.2015.11.015

Parallel, NT: a Marriage Made in Heaven

- As VDD decreases, operating speed decreases
- However efficiency increases→ more work done per Joule
- Until leakage effects start to dominate
- Put more units in parallel to get performance up and keep them busy with a parallel workload

Workloads like ML are massively parallel and scale very well (P/S ↑ with NN size)

[Rossi et al. IEEE Micro 2017]



Optimum point

Better to have N× PEs running at optimum Energy than 1 PE running fast at low Energy efficiency

First steps to open source, how to start?

- **At this time nothing was released**
  - We were 100% sure it would become open source
  - But we had no idea how
    - What can we open source, and what not
    - We work for ETH Zurich, we have to ask their permission
  - We also did not have much idea about licensing
- **We need support of industry**
  - This project was supported by ST Microelectronics
    - They would not support a project where they can not use our work ‘freely’
  - Permissive licenses are the only way
    - Even though purists consider it not ‘free’ enough

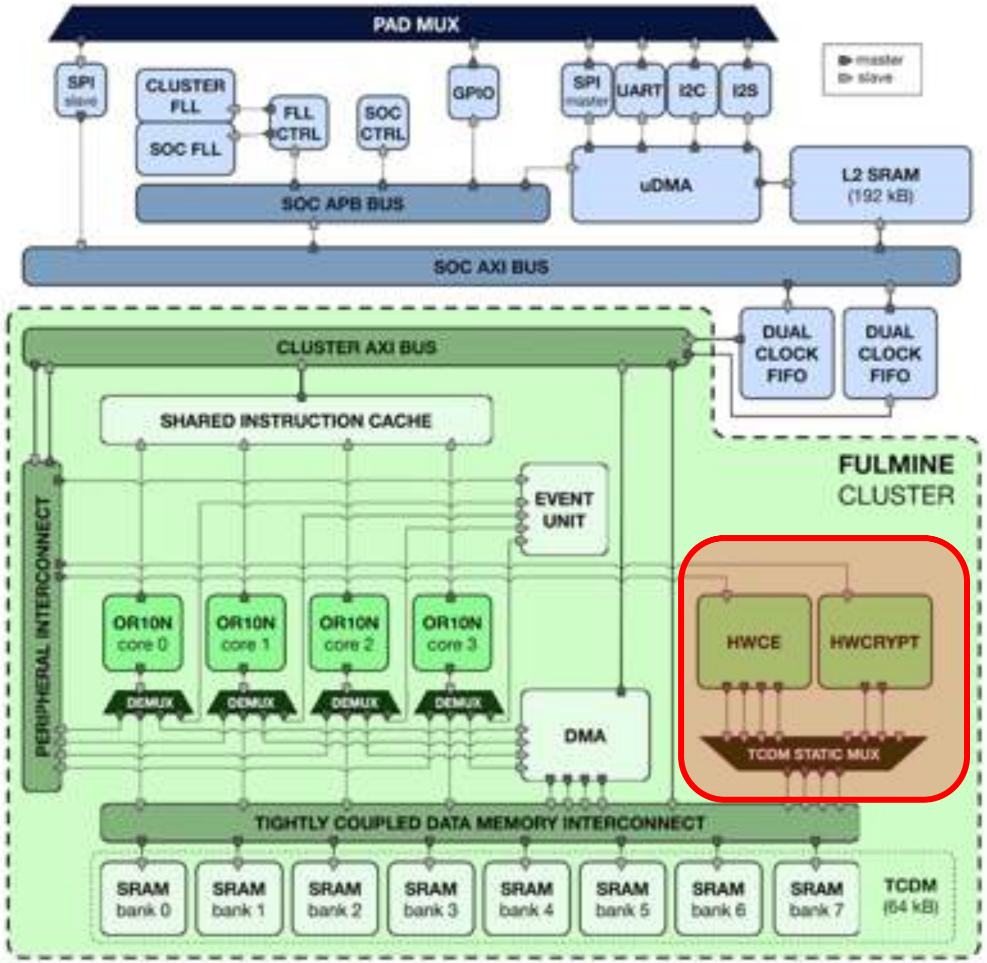
#2 – Fulmine (2015) – The award winning one

- UMC65
- Novelty – HW Accelerators!
  - 4x OpenRISC cores (still not RISC-V)
  - 2x HW accelerators
    - HW – Crypt (together with TU-Graz)
    - HW – Convolution Engine
- Meant as a chip for boards
  - Not only on a tester for characterization
  - Followed Mia Wallace, Honey Bunny
  - Paved the way for next wave of chips

IEEE Circuits and Systems, Darlington Award for Best Paper in 2020

Francesco Conti, Robert Schilling, Davide Schiavone, Antonio Pullini, Davide Rossi, Frank K. Gurkaynak, Michael Muehlberghuber, Michael Gautschi, Igor Loi, Germain Haugou, Stefan Mangard, Luca Benini, "An IoT Endpoint System-on-Chip for Secure and Energy-Efficient Near-Sensor Analytics", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol: 64, Issue: 9, Sept. 2017, pp 2481 - 2494, DOI: 10.1109/TCSI.2017.2698019

We have a base to work on and expand



- **Much more than a core**
  - Peripherals (SPI, UART, I2C, I2S)
  - DMA, Busses, event unit
- **First chip with accelerators**
  - 0-copy connection to the memory
  - Allows independent systems (HWCrypt/HWCE) to be added easily.
- **Still not openly released**
  - Using our OpenRISC core (3rd gen)

First open source release comes at this time

- **PULPino was the first release (February 2016)**
  - Used the SoC infrastructure and peripherals
  - Much simpler: single core, separate data, instruction memories
- **It is still the most popular release (name recognition wise)**
  - We have much more advanced releases, but PULPino is much better known
  - Your **first release** will end up **carrying a lot of weight**
- **Used SolderPad as a license**
  - Our friends at LowRISC suggested this license
  - Additions to Apache to clarify hardware related issues
  - We **still use the same license**

RISC-V is a game changer

Nice ISA design, patent troll safe, extensible, huge momentum

It's the Software, stupid!

- Toolchains → GCC, LLVM
- System tools → Emulators: QEMU, TinyEMU, Spike, Renode
  Bootloaders: Coreboot, U-boot, BBL, OpenSBI
  BINUTILS, GDB, OpenOCD, Glibc, Musl, Newlib
- Language Runtimes → C, C++, Fortran, GO, Rust, Java, Ocaml,
- Operating Systems → Linux: Fedora, OpenSUSE, Gentoo, OpenEmbedded/Yocto, Buildroot, OpenWRT, FreeBSD
  FreeRTOS, Zephyr, RTEMS, Xv6, HelenOS

https://github.com/riscv/riscv-software-list

#3 - Mr. Wolf (2017) – The application chip

- **TSMC40 LP**
- **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**

Antonio Pullini, Davide Rossi, Igor Loi, Alfio Di Mauro, Luca Benini, "Mr.Wolf: A 1 GFLOP/s Energy-Proportional Parallel Ultra Low Power SoC for IoT Edge Processing", In Proc. European Solid State Circuits Conference (ESSCIRC) 2018, 3-6 Sep 2018, Dresden, DOI: 10.1109/ESSCIRC.2018.8494247

PULP-NN on Xpulp: The Power of ISA Extension

8-bit Convolution

HW Loop

LD/ST with post increment

8-bit SIMD sdotp

RV32IMC

N

```
addi   a0,a0,1
addi   t1,t1,1
addi   t3,t3,1
addi   t4,t4,1
lbu    a7,-1(a0)
lbu    a6,-1(t4)
lbu    a5,-1(t3)
lbu    t5,-1(t1)
mul    s1,a7,a6
mul    a7,a7,a5
add    s0,s0,s1
mul    a6,a6,t5
add    t0,t0,a7
mul    a5,a5,t5
add    t2,t2,a6
add    t6,t6,a5
bne    s5,a0,1c000bc
```

RV32IMCXpulp

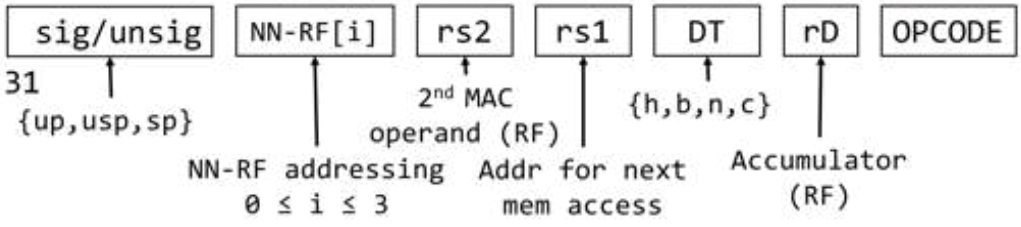
N/4

```
lp.setup
p.lw  w1, 4(a0!)
p.lw  w2, 4(a1!)
p.lw   x1, 4(a2!)
p.lw   x2, 4(a3!)
pv.sdotsp.b   s1, w1, x1
pv.sdotsp.b   s2, w1, x2
pv.sdotsp.b   s3, w2, x1
pv.sdotsp.b   s4, w2, x2
end
```

can we remove?

Yes! DOTP+LW

```
pv.cusdot{up,usp,sp}.{h,b,n,c}.{NN-RF[i]} rD,rs1,rs2
```



sig/unsig | NN-RF[i] | rs2 | rs1 | DT | rD | OPCODE

31

{up,usp,sp}

NN-RF addressing $0 \leq i \leq 3$

2nd MAC operand (RF)

Addr for next mem access

{h,b,n,c}

Accumulator (RF)

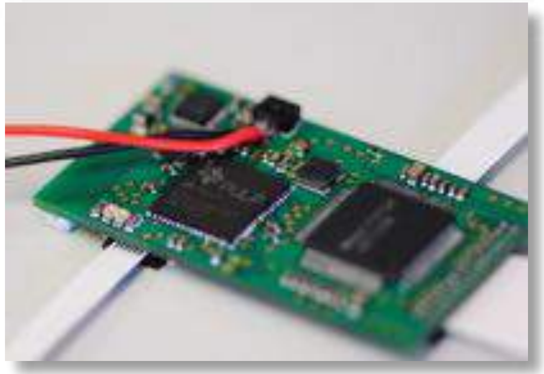
9x less instructions than RV32IMC

Extra 1.7x cycle reduction at “1-2%” area cost (~400GEs)

1.6x Area, 1.5Power, 15x Speed → 10x Energy Efficiency!

Mr. Wolf has been used in multiple systems

- **Designed as an application processor**
  - We still build boards with it
  - Despite only 200 manufactured
- **Widespread industrial use:**
  - Dolphin IP was validated on this chip
  - Greenwaves GAP8 is based on the open source release OpenPULP
  - BitCraze AI Deck is related



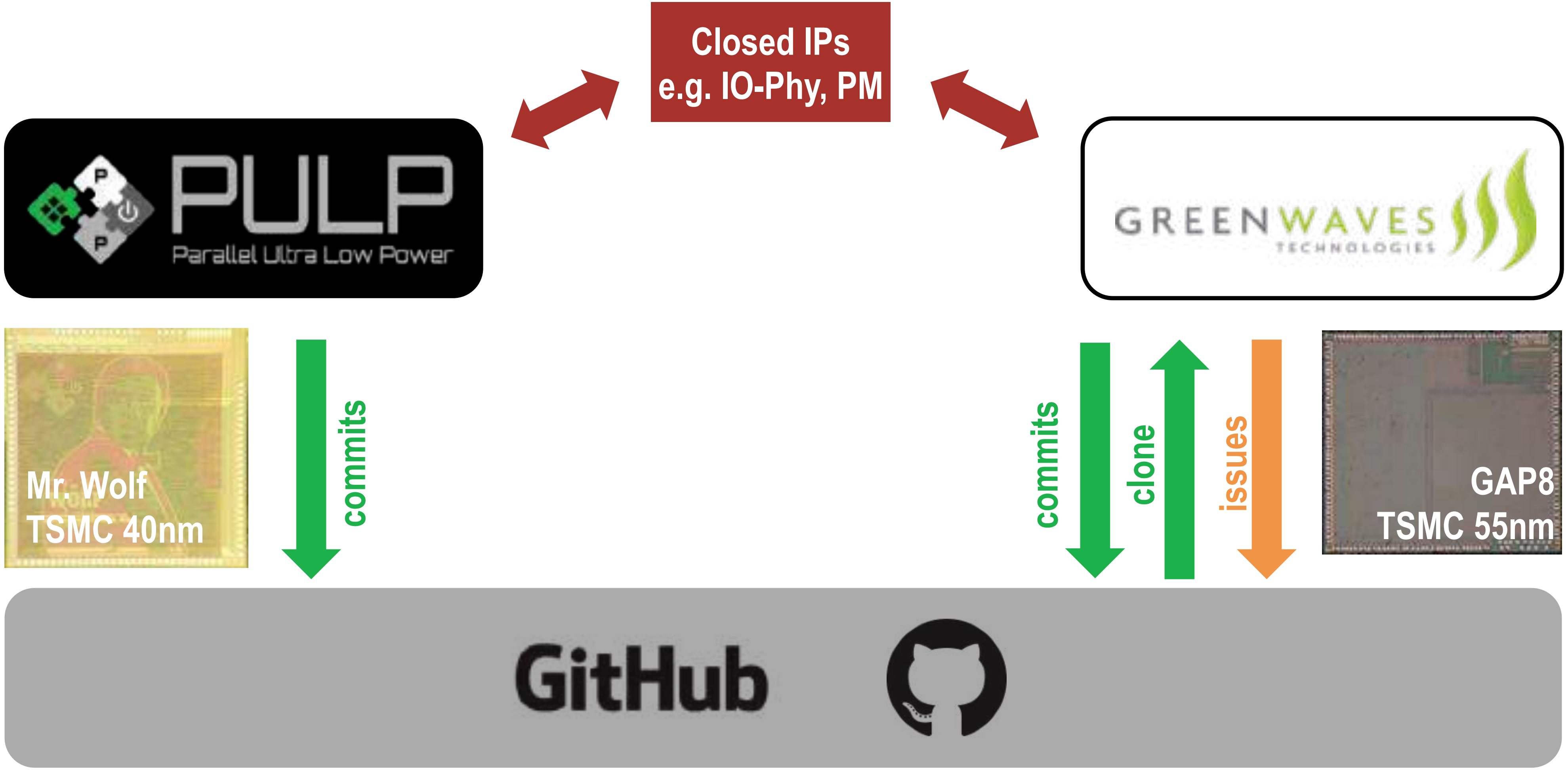
GREENWAVES TECHNOLOGIES

bitcraze

What a difference two years make

- **With Mr. Wolf, most of what we have is open sourced**
  - This is a **complex IoT processor**, not like the much simpler PULPino
  - 8 + 1 cores, FPUs, shared accelerators, multiple power down modes.
- **The cores are now RISC-V**
  - Supports RV32IMCF and custom extensions (xPULP)
- **Interesting collaboration with Dolphin Integration (SOITEC)**
  - They have their IP demonstrated on an complex design, they can freely share
  - We get to use industrial IP in our chip
- **Still many parts can still not be open source**
  - FLL, analog macros, I/O cells, memory cuts (affects performance), P&R scripts

Open source collaboration scheme explained



Closed IPs
e.g. IO-Phy, PM

Mr. Wolf
TSMC 40nm

commits

commits

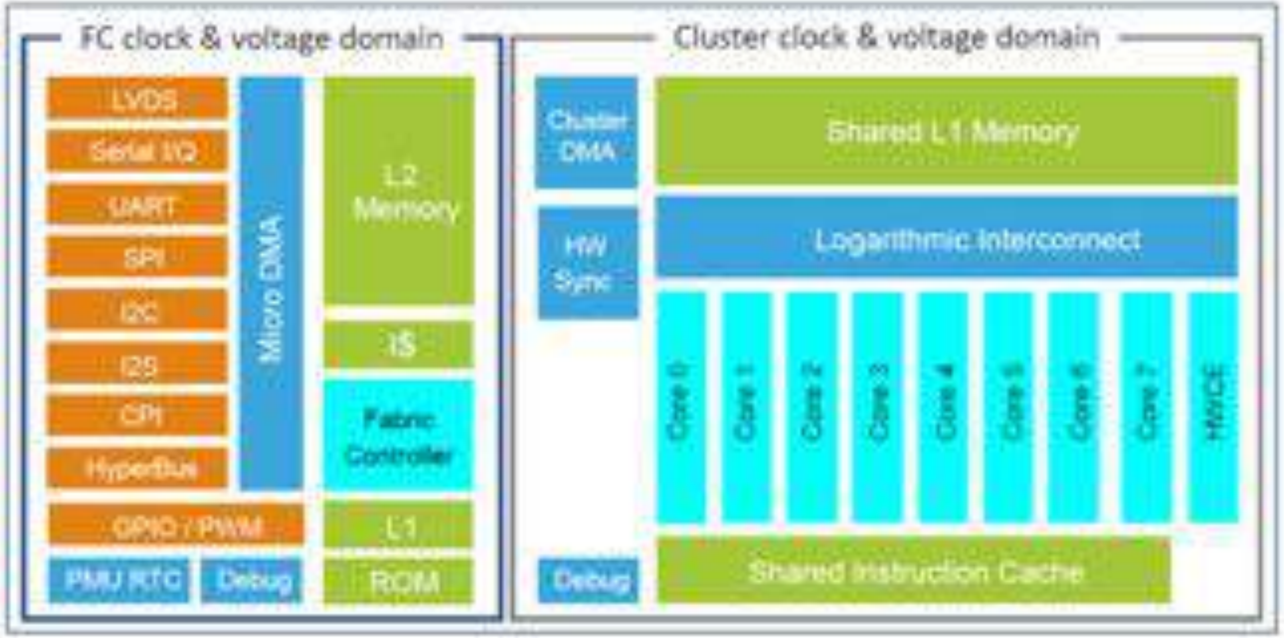
clone

issues

GAP8
TSMC 55nm

Successful product development: GWT's GAP8

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



MCU Function

- Extended RISC-V core
- Extensive I/O set
- Micro DMA
- Embedded DC/DC c
- Secured execution

Computation engine

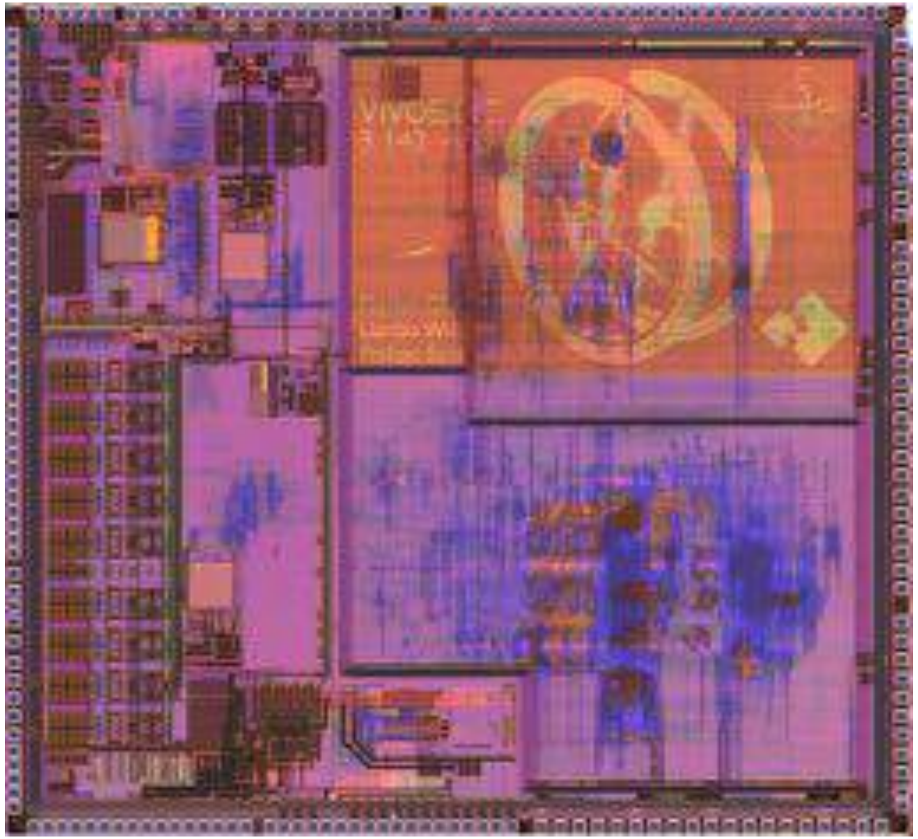
- 8 extended RIS
- Fully programmable
- Efficient parallelization
- Shared instruction cache
- Multi channel DMA
- HW synchronization
- HW convolution Engine

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 17 5	99.1 11 X	1 500 000	3.7 16 X
GAP8 @1.2V	17.5	8.7	1 500 000	70
GAP8 @1.0V w HWCE	4.7	99.1	460 000	0.8

GREENWAVES TECHNOLOGIES

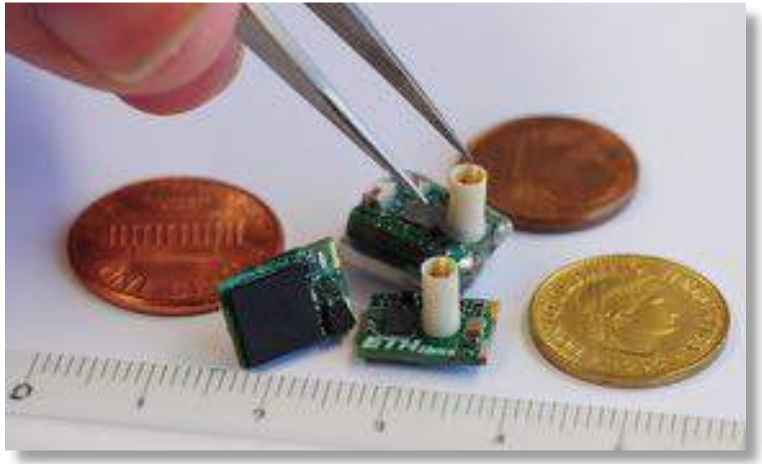
#4 - VivoSoC 3.142 (2019) – Analog and Digital

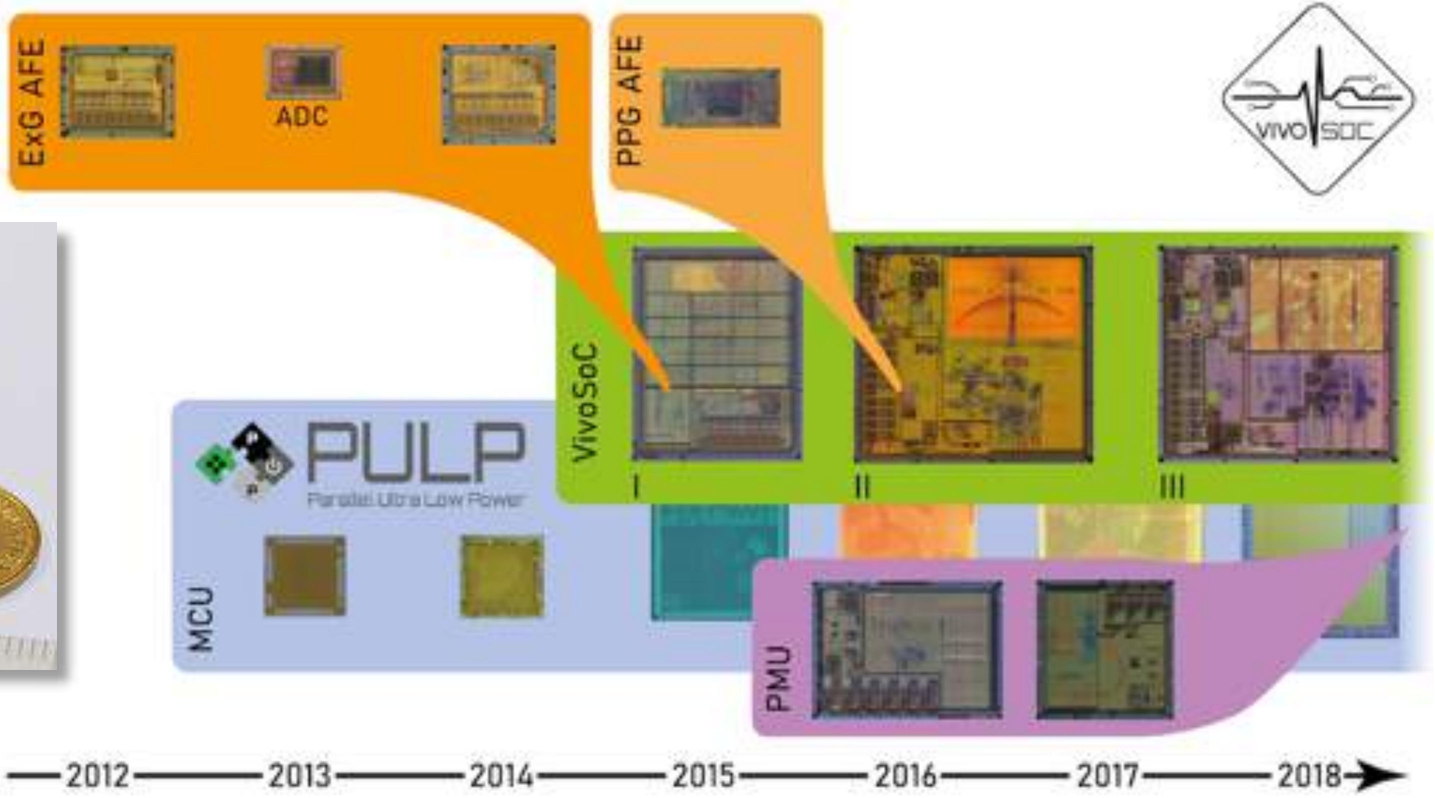
- **Actually 4+ VivoSoCs since 2015**
- **SMIC 130/110 technology**
  - Many Analog IPs
    - ExG interfaces, A/D converters
    - Pulse Oximetry
    - Neuro stimulators
- **PULP cluster for post processing**
  - 4x RISC-V cores
  - Digital interfaces
  - DMA transfer from analog block to digital



Philipp Schoenle, Florian Glaser, Thomas Burger, Giovanni Rovere, Luca Benini, Qiuting Huang, "A Multi-Sensor and Parallel Processing SoC for Miniaturized Medical Instrumentation", IEEE Journal of Solid-State Circuits PP issue:99, pp 1-12, DOI: 10.1109/JSSC.2018.2815653

PULP allows us to co-operate with everyone





- **Collaboration between Prof. Benini and Prof. Huang**
  - Permissive licensing allows collaboration even if the result is **not** open source

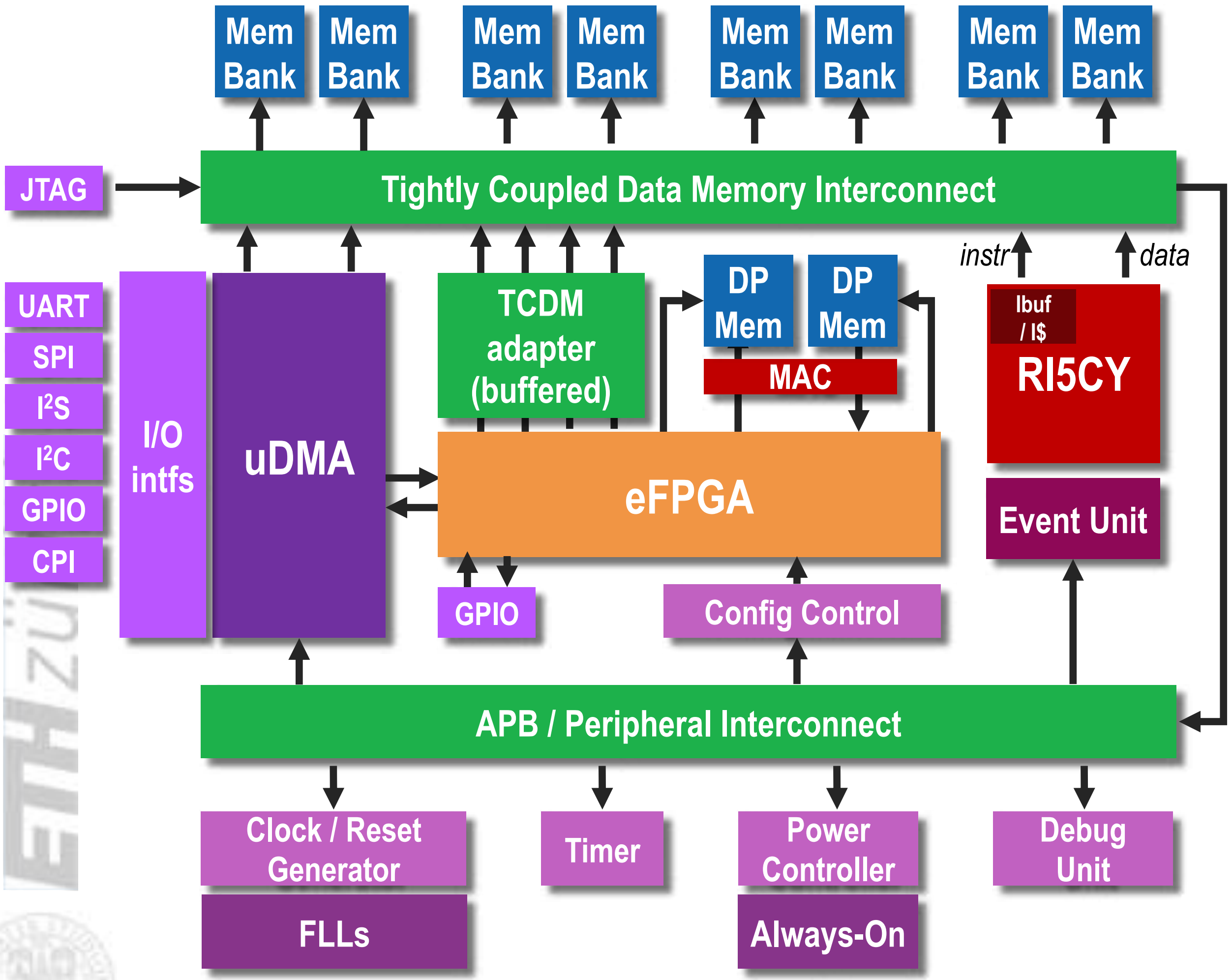
#5 - Arnold (2018) – Fastest collaboration

- **GF22nm**
  - RISC-V microcontroller with eFPGA
  - Based around PULPissimo
- **Collaboration with Quicklogic**
  - Met at GTC 2017 by coincidence
  - In one year chip was taped out
  - Only possible because of open source nature
- **Quicklogic is going open source**
  - They announced June 2020 the Quicklogic Open Reconfigurable Computing

    https://www.quicklogic.com/QORC/

Davide Schiavone, Davide Rossi, Alfio Di Mauro, Frank Gurkaynak, Timothy Saxe, Mao Wang, Ket Chong Yap, Luca Benini, "Arnold: an eFPGA-Augmented RISC-V SoC for Flexible and Low-Power IoT End-Nodes", arXiv: 2006.14256

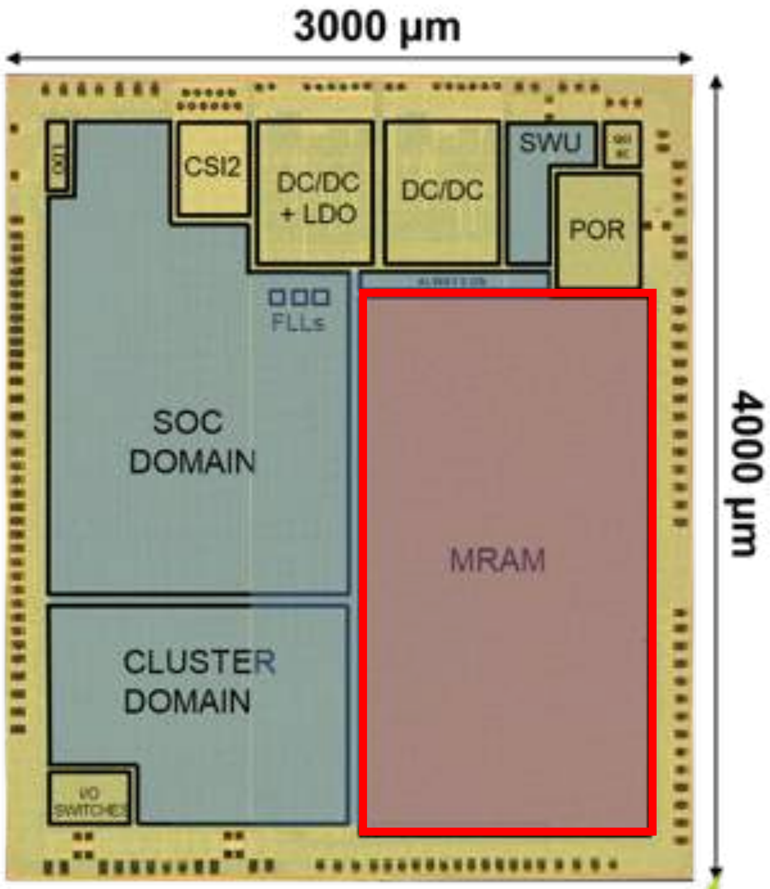
PULPissimo: very good platform for extensions



- eFPGA added as accel.
  - Easy plug and play
  - Configuration over APB
  - Additional ALU and memory
  - Uses the same memory
- Multiple operation modes
  - Configurable peripheral
  - Accelerator for core
  - Accelerator for independent I/O

#6 – VEGA (2020): Next-Generation IoT Processor

- RISC-V cluster (8cores +1)
- **Multi-precision HWCE(4b/8b/16b)** for NN acceleration (MAC engine)
- **Cognitive unit for autonomous wake-up** from retentive sleep (high-dimensional computing)
- **Fully-on chip DNN inference with 4MB MRAM**



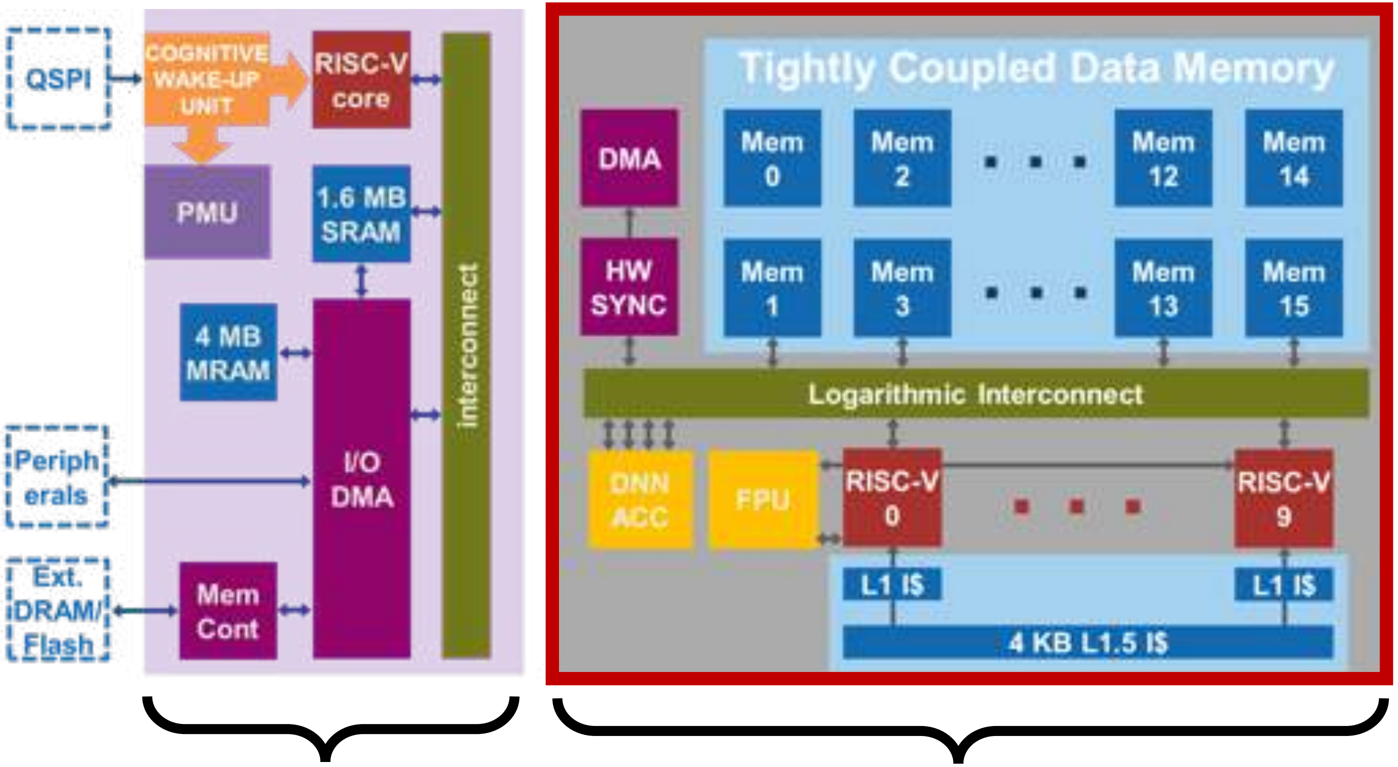
In cooperation with GREENWAVES TECHNOLOGIES

Technology	22nm FDSOI
Chip Area	12mm^2
SRAM	1.7 MB
MRAM	4 MB
VDD range	0.5V - 0.8V
VBB range	0V - 1.1V
Fr. Range	32 kHz - 450 MHz
Pow. Range	1.7 µW - 49.4 mW

[Rossi et al. ISSCC21]

All together in VEGA: Open Processors & Accelerators

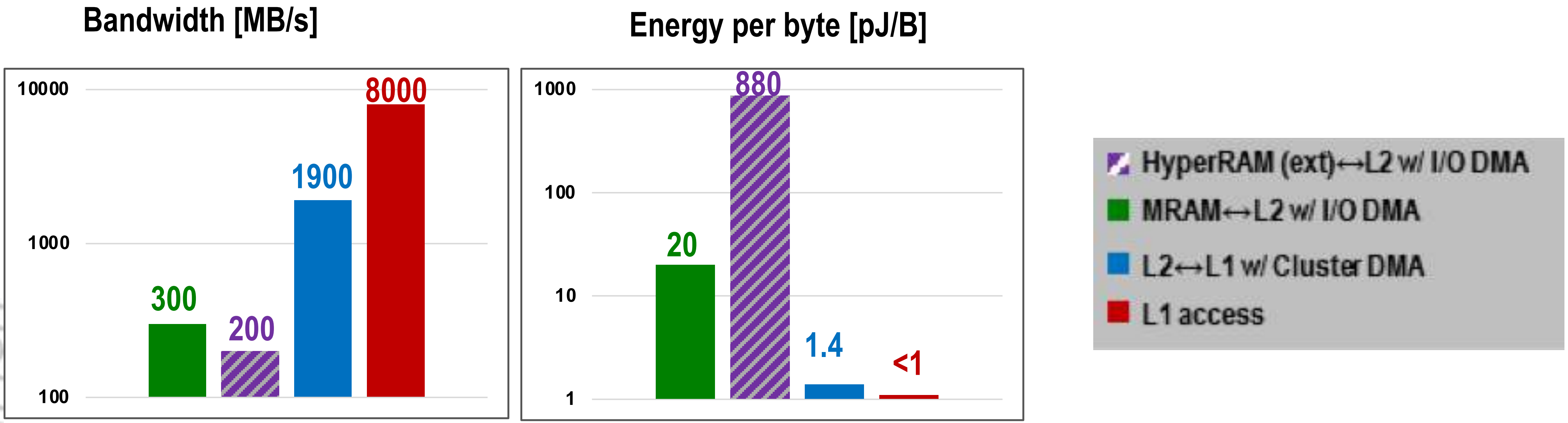
- **RISC-V cluster** (8cores +1) **614GOPS/W @ 7.6GOPS** (8bit DNNs), **79GFLOPS/W @ 1GFLOP** (32bit FP appl)
- **RBE**: (4b/8b/16b) 3×3×3 MACs with normalization / activation: **32.2GOPS** and **1.3TOPS/W** (8bit)
- **Hypnos: 1.7µW** cognitive unit for autonomous wake-up from retentive sleep mode

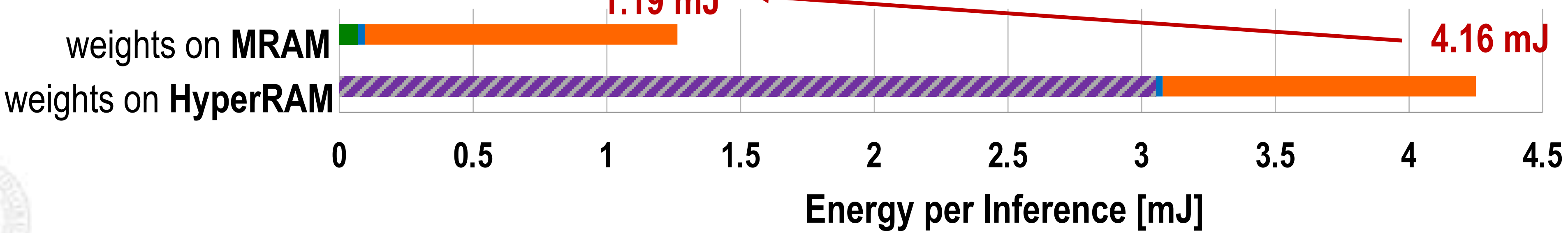


SoC: SoA MCU

Parallel SW prog. accelerator

Full DNN Energy (MobileNetV2)



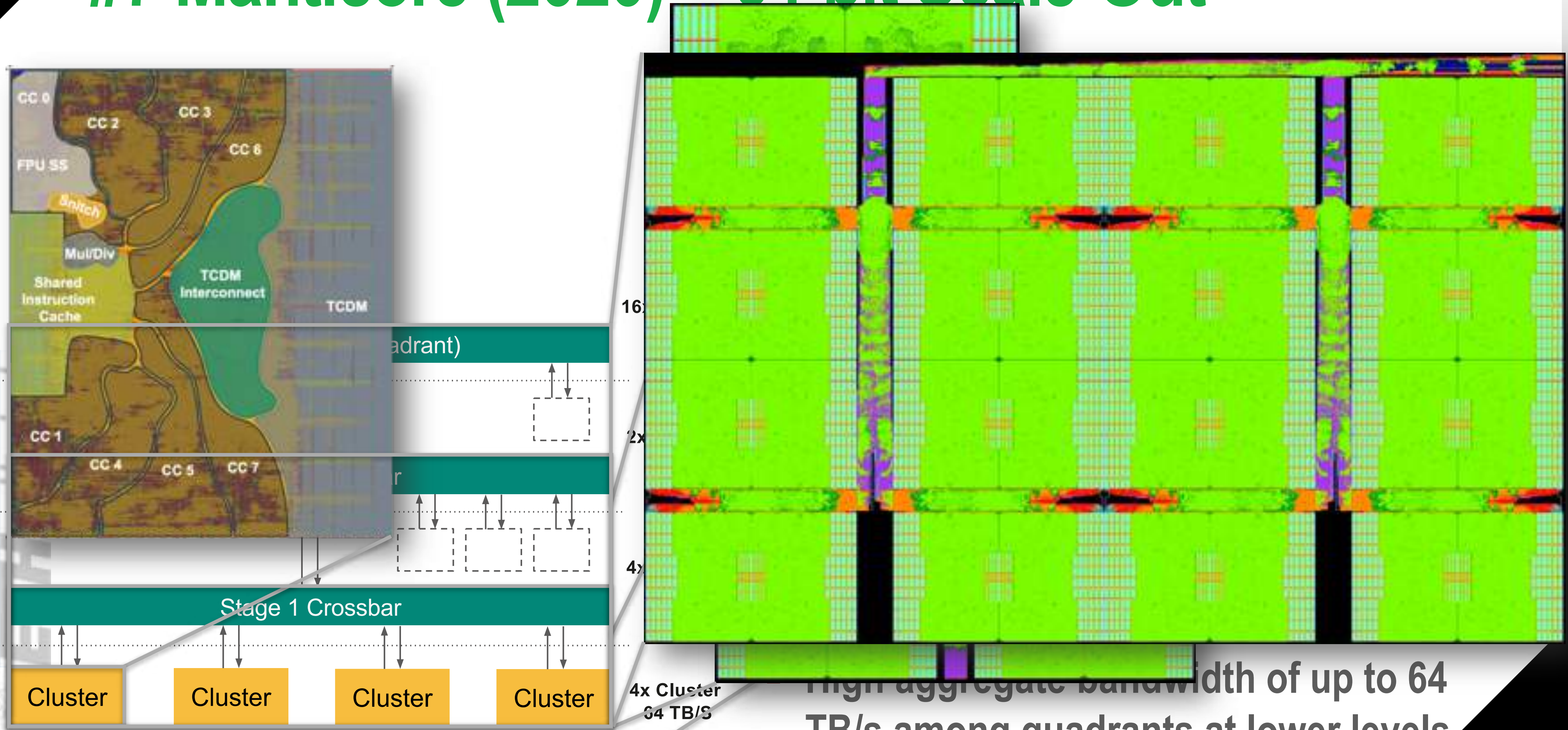


end-to-end on-chip computation
1.19 mJ

3.5x less energy

4.16 mJ

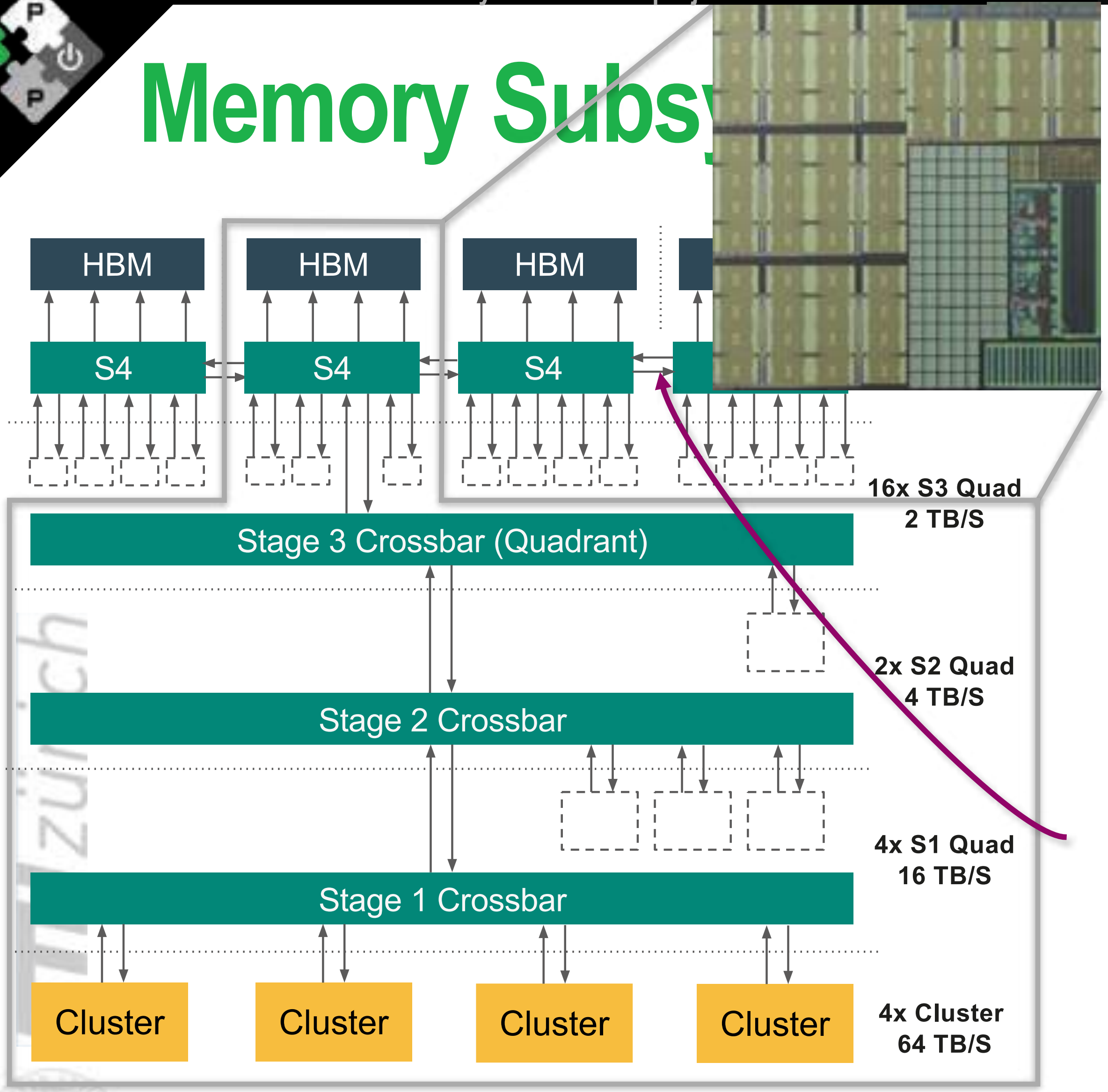
#7 Manticore (2020) – 64-bit Scale-Out



High aggregate bandwidth of up to 64 TB/s among quadrants at lower levels

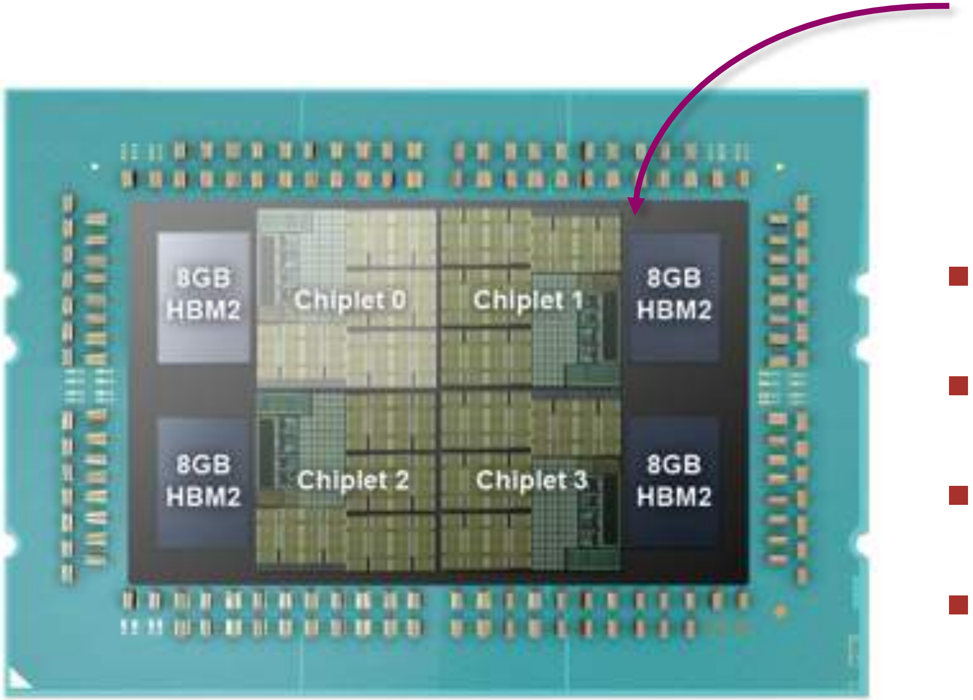
[Zaruba et al. HOT-CHIPS20]

Memory Subsy



- **HBM: Matching the bandwidth with the memory interface**
- **Block-wise DMA accesses:**
  - High bus utilization ≈ high energy-efficiency
  - Multi-dimensional blocks with Snitch DMA
  - Good fit for parallel interfaces (HBM/HBI)
  - Latency tolerance through double buffering
  - Different from GPU memory hierarchy
- **Multi-chiplet design**
- **HBI: Scaling across dies (NUMA)**

Manticore Multi-Chip Concept



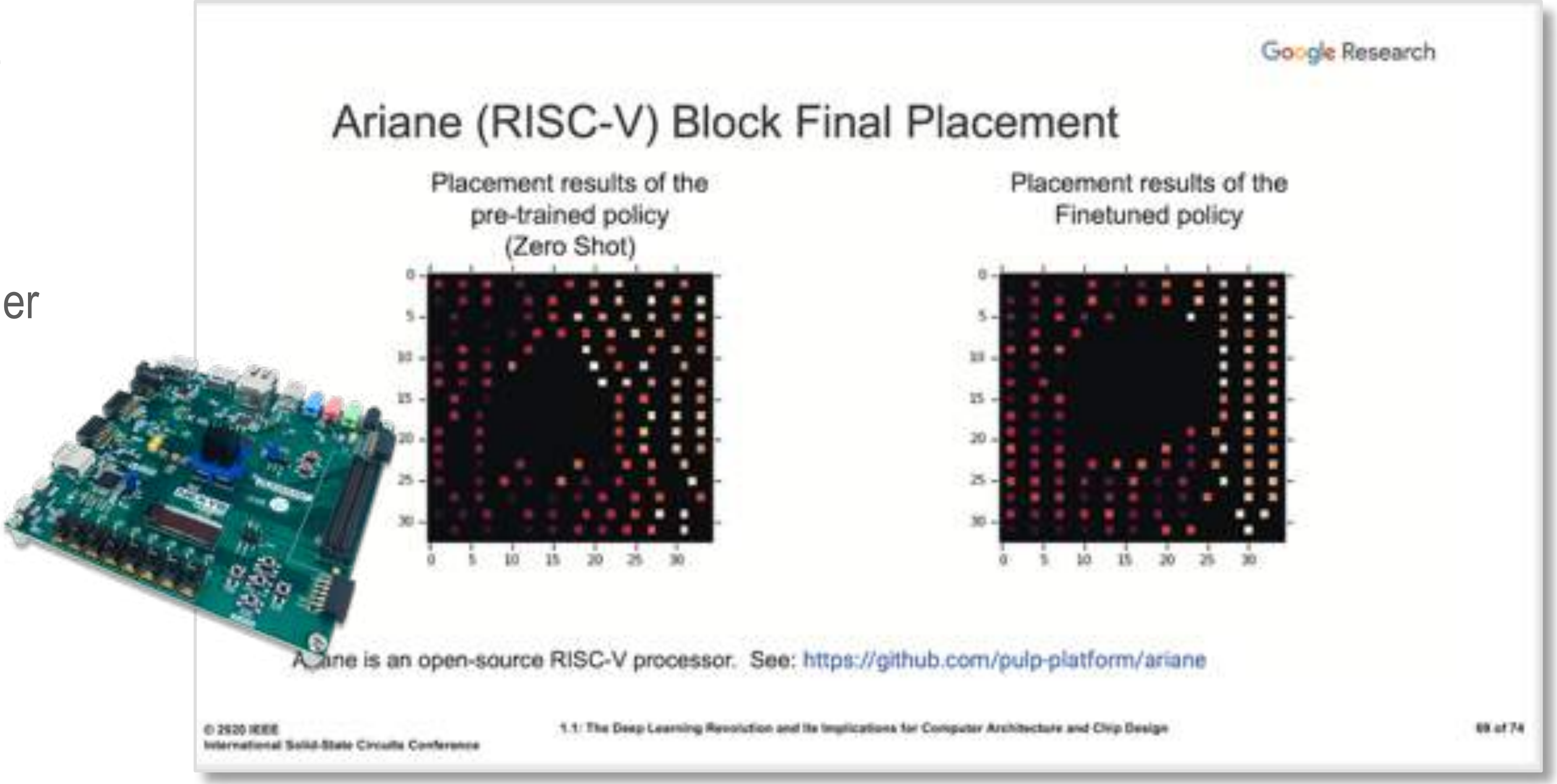
- Four chiplets and 8GB HBM2 on an interposer
  - Interposer enables **high-bandwidth**, **energy-efficient** parallel interfaces
- High D2D bandwidth
- High die to HBM bandwidth
- Total 4096 Snitch cores, peak > 8 Tdpflop/s
- Four Ariane ”manager” cores

Outperforms SoA, **open** building-blocks, foundation of **next generation** high-performance computing systems!

Manticore

Manticore System Controller: Ariane

- Linux capable RV64GC core
  - Very Popular
  - Single-issue in-order
- FPGA port
  - Xilinx Genesys
- Used in many projects:
  - EPI
  - Hensoldt Cyber (Mig-V)
  - OpenPiton (Princeton)



Ariane (RISC-V) Block Final Placement

Placement results of the pre-trained policy (Zero Shot)

Placement results of the Finetuned policy

A ane is an open-source RISC-V processor. See: https://github.com/pulp-platform/ariane

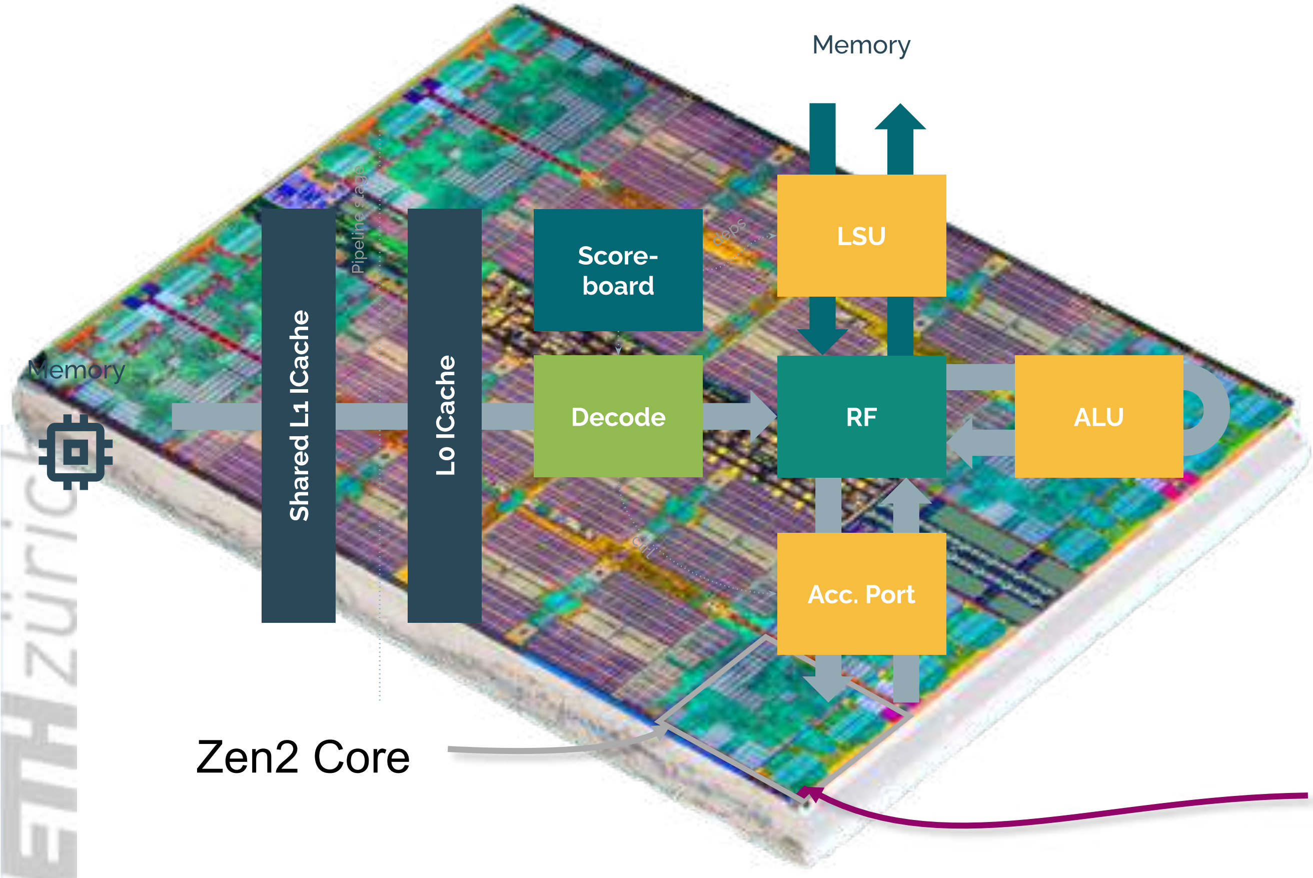
© 2020 IEEE
International Solid-State Circuits Conference

1.1: The Deep Learning Revolution and its Implications for Computer Architecture and Chip Design

69 of 74

Slide from keynote speech from ISSCC 2020 by Jeff Dean of Google

Snitch: Tiny Control Core



- **Feeds the FPU**
  - **Tiny**, **simple**, and **lightweight control** core
  - **Competitive frequency**
  - **Latency-tolerant – non-blocking with scoreboard**
  - **Throughput-oriented extensions: FREP, SSR, pseudo-dual issue**
- **Around 10-20 kGE (DP-FPU 100kGE!)**

Graduating our cores for a better future

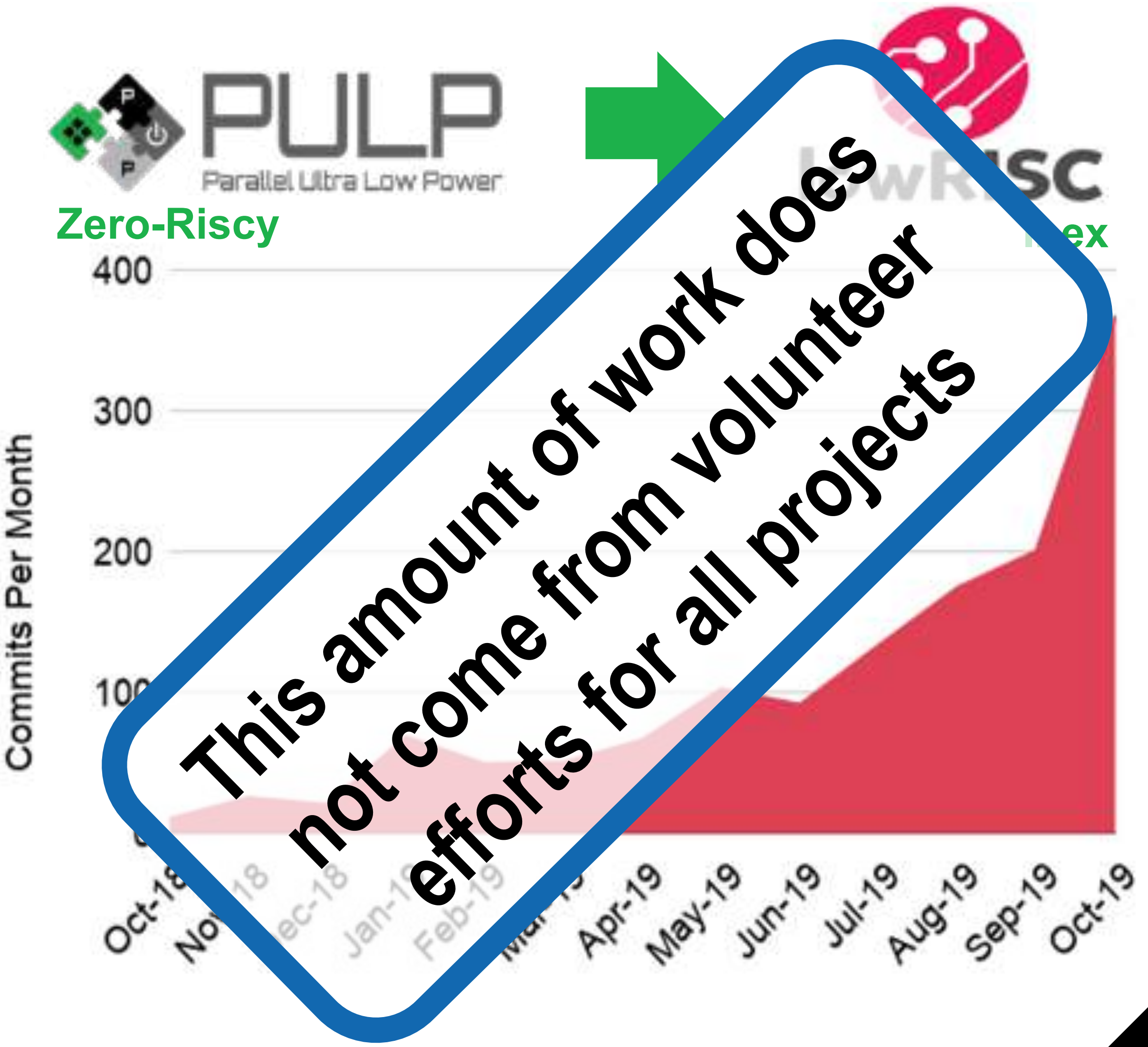
- Several of our open source cores are maintained by others
  - Zero-riscy became Ibex and is maintained by LowRISC
  - RI5CY became CV32E40P and is maintained by OpenHW group
  - Ariane (recently) became CVA6 and is maintained by OpenHW group
- This is an excellent opportunity for us
  - These groups have funds to support much needed but tedious work
    - Documentation, verification, user support
- Also means that we have done a good job ☺
- And creates opportunities for our graduates
  - At the moment Pirmin, Davide S., Florian, Gianmarco are involved

LowRISC and OpenHW are essential for us

Example

Although Zero-Riscy was open source since 2016, work on it really picked up when LowRISC took over.

35+	Contributors
1300+	Contributions
470	GitHub Issues



This amount of work does not come from volunteer efforts for all projects

Open HW group

OPENHW GROUP™
PROVEN PROCESSOR IP

- **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.
- Many partners
  - Many more resources
  - Also to do the work that would normally not done like verification, documentation

Multi-B$ companies using (or planning) our cores in their products!

Alibaba Group, ASHLING, axiomise, bluespec, Barcelona Supercomputing Center, BTA Design Services, CMC Microsystems, DATUM, ECSPEC, em microelectronic, EMBECOSM, ETH zürich, ETS, FUTUREWEI, GLOBALFOUNDRIES, GREENWAVES, HENSOLDT, HUAWEI, IAR SYSTEMS, imperas, metrics, Mitacs, NVIDIA, NXP, onespin, OPERSYS, Polytechnique Montréal, PRAESUM Communications, SECURE THINGZ, SILICON LABS, Symbiotic EDA, THALES, ultrasoc, Alma Mater Studiorum Università di Bologna, uOttawa, University of Toronto, The University of Utah, verifAI, VeriSilicon

We benefit from our open source activities

- **Science**
  - Community building, sharing ideas
  - Reduce “getting up to speed” overhead
  - Work on things that make a difference
  - Fair benchmarking
- **Business**
  - Reduce NRE costs for silicon
  - Faster innovation paths for startups
  - New business models
  - Helps exchange ideas across NDA walls
- **Society**
  - More innovation, growth, jobs
  - Bridges the gap between groups, allows more people to contribute
  - More secure, safe auditable HW

http://asic.ethz.ch

40 SoCs & counting

Thanks!

Luca Benini, Davide Rossi, Andrea Borghesi, Michele Magno, Simone Benatti, Francesco Conti, Francesco Beneventi, Daniele Palossi, Giuseppe Tagliavini, Antonio Pullini, Germain Haugou, Manuele Rusci, Florian Glaser, 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, Hanna Mueller, Matteo Perotti, Nils Wistoff, Luca Bertaccini, Thorir Ingulfsson, Thomas Benz, Paul Scheffler, Alessio Burello, Moritz Scherer, Matteo Spallanzani, Andrea Bartolini, Frank K. Gurkaynak,

and many more that we forgot to mention

http://pulp-platform.org   @pulp_platform