ETH zürich ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

PULP: 10 Years of Open Source Hardware

Integrated Systems Laboratory (ETH Zürich)

Frank K. Gürkaynak kgf@iis.ee.ethz.ch

@pulp_platform

pulp-platform.org

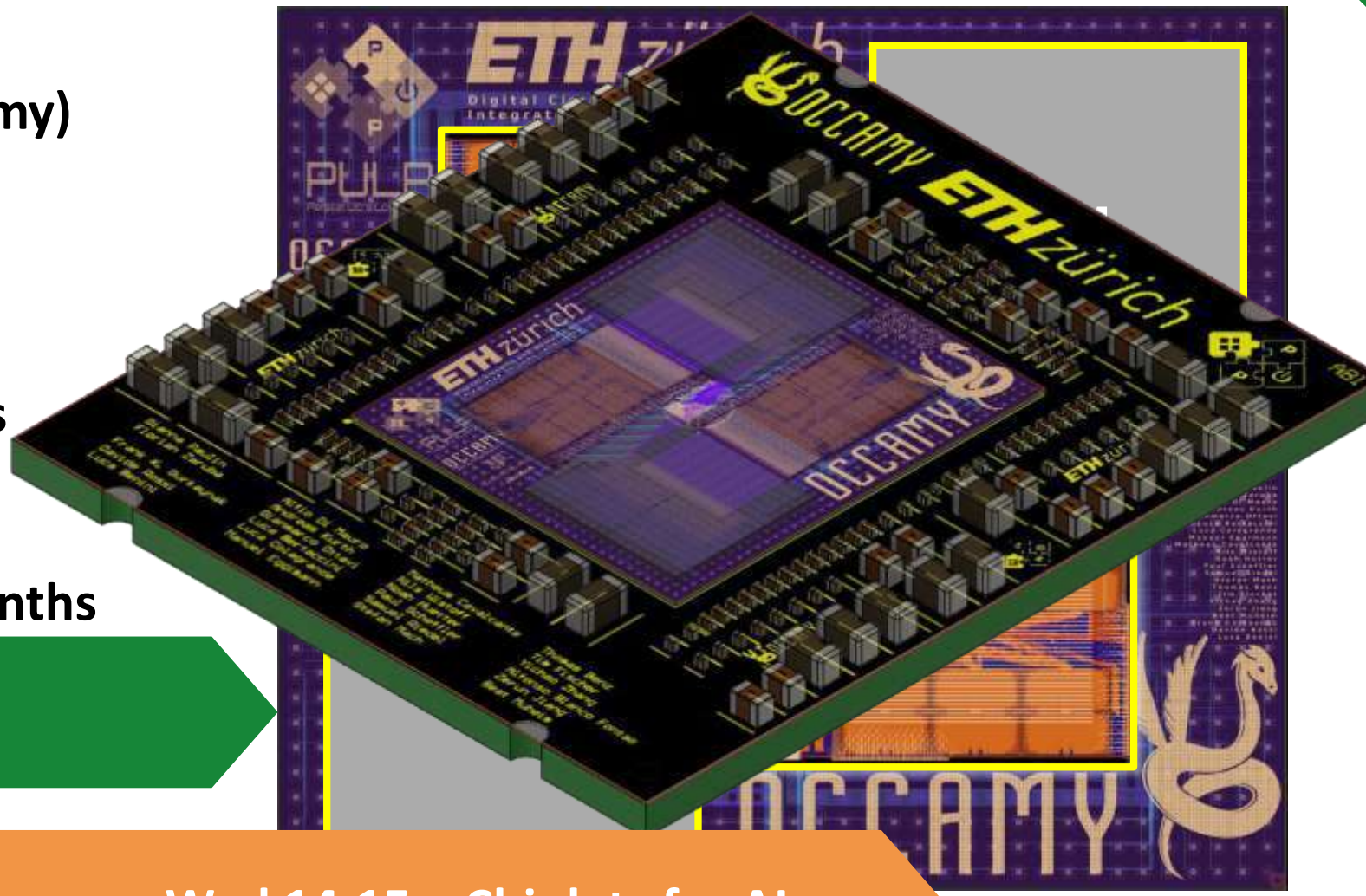
youtube.com/pulp_platform

PULP Platform

Open Source Hardware, the way it should be!

Our latest design Occamy: 0.75 TFLOP/s, 400+ cores

- **Chiplet based design**
- **2x Compute chiplets (Occamy)**
  - 216+1 RISC-V cores
  - GF12LPP
  - Running at 1 GHz
- **2x 16GByte HBM memories**
- **Silicon Interposer (Hedwig)**
- **Finished in less than 15 months**



How did we manage this?

- Taped out on 1st of July 2022

More on Occamy – talk by Gianna – Wed 14:15 – Chiplets for AI

Open source hardware is for us a necessity

- **Modern IC design like Occamy is complex and expensive**
  - We need partners to help and collaborate
  - We need support (IPs, donations) to realize designs

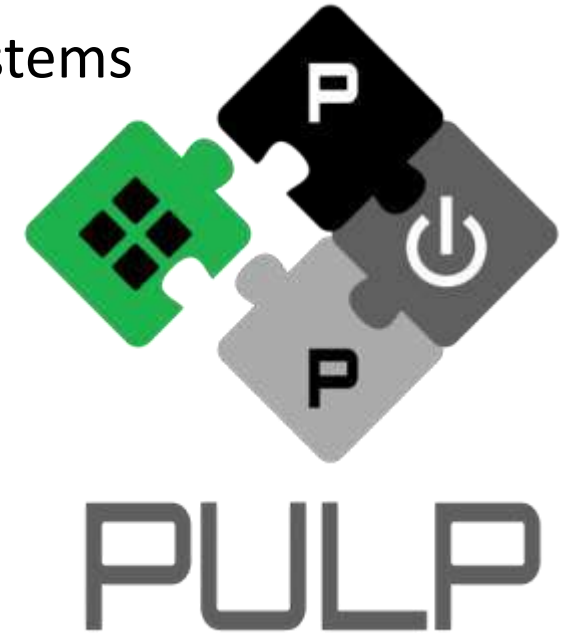
Partners that helped Occamy become reality

Open Source to the rescue

- **Makes it easy to collaborate with external partners, build teams (both industrial and academic)**
  - Less paperwork/NDAs to get started
  - Partners see/are aware of what we provide
- **What we do can be re-used (permissive licensing) by our partners**
- **Results can be more easily verified**

We started almost excatly 10 years ago (April 2013)

- **Investigating new computing architectures**
  - Efficient over a wide range from IoT applications to HPC systems
- **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**



Today the PULP team has grown to more than 70 people

- **Headed by Luca Benini**
- **Teams in both ETH Zürich and University of Bologna**

... and Bianca the Bear

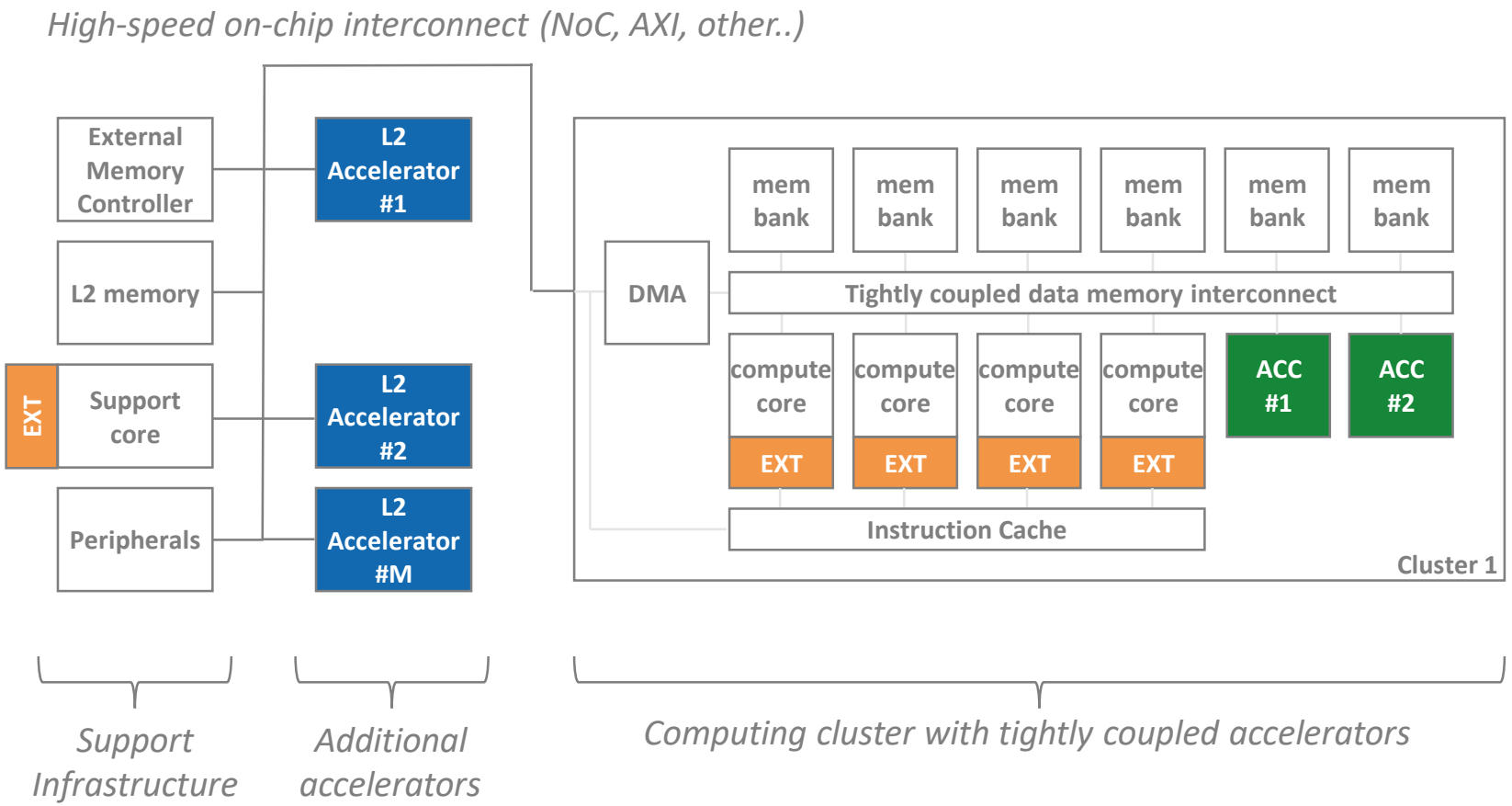
Our research focus: cluster-based many-core accelerators

Innovation factors

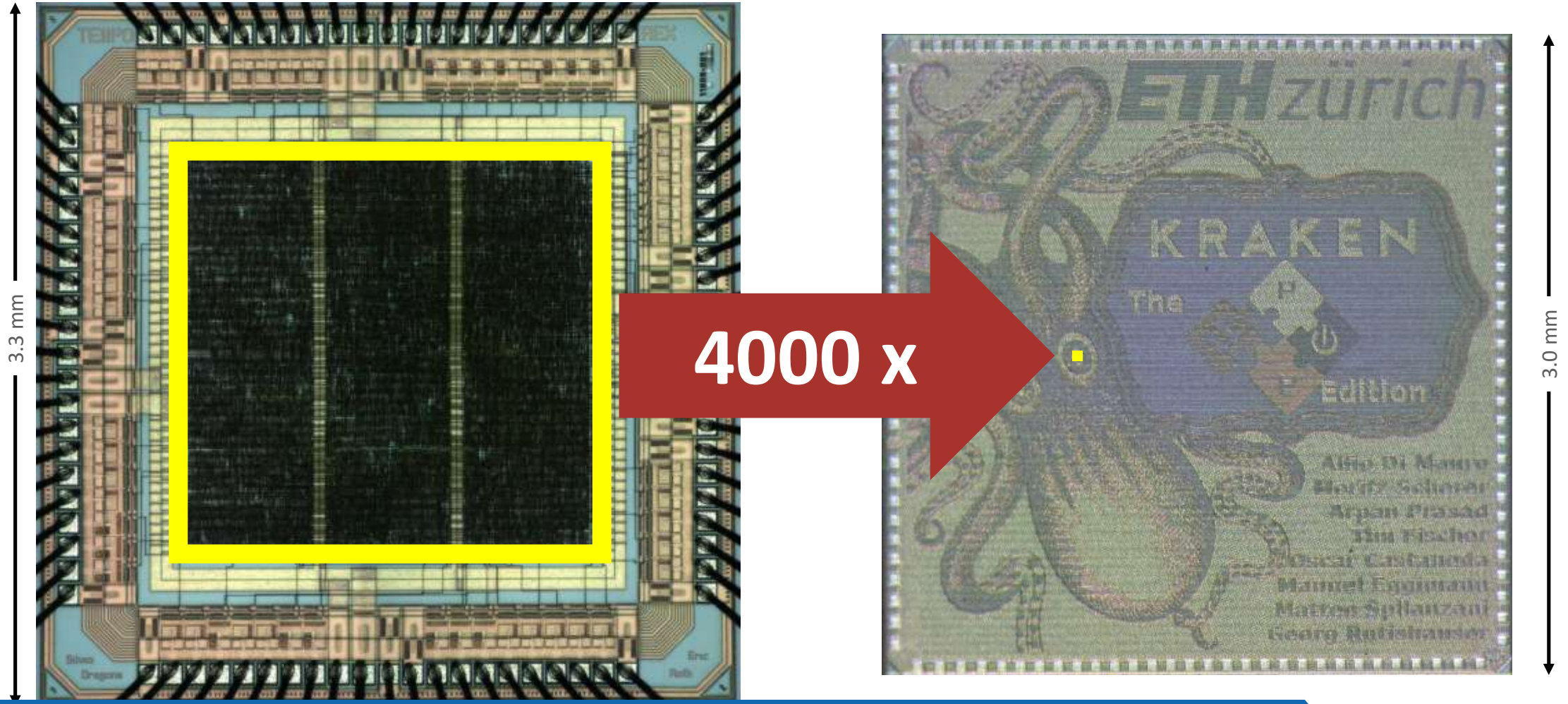
Extensions to processor cores
- Explore new extensions
- Efficient implementations

Shared-memory Accelerators
- Domain specific
- Local memory

Multiple computing clusters
- Communication
- Synchronization



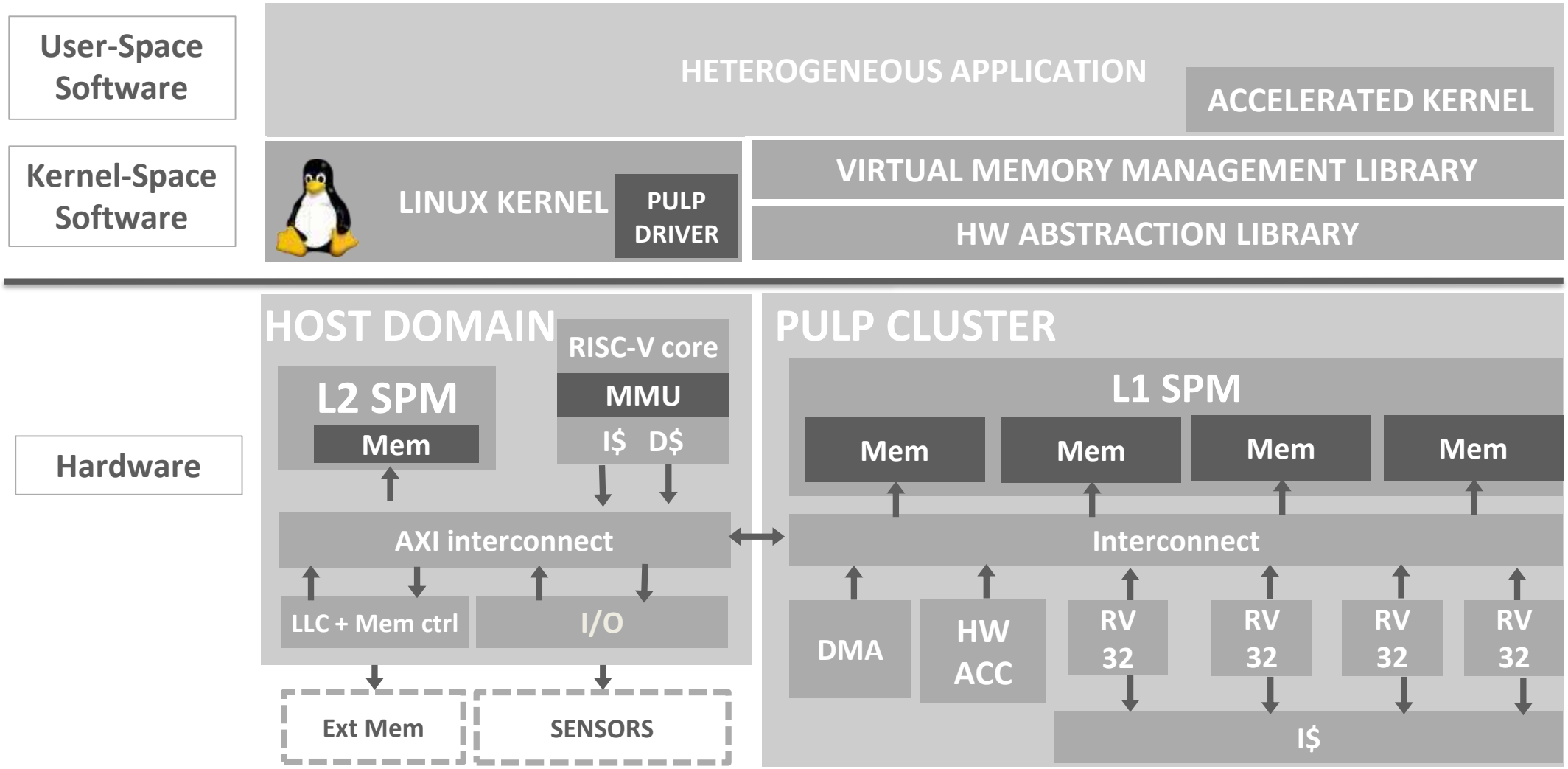
In the last 20 years IC Design has changed a lot



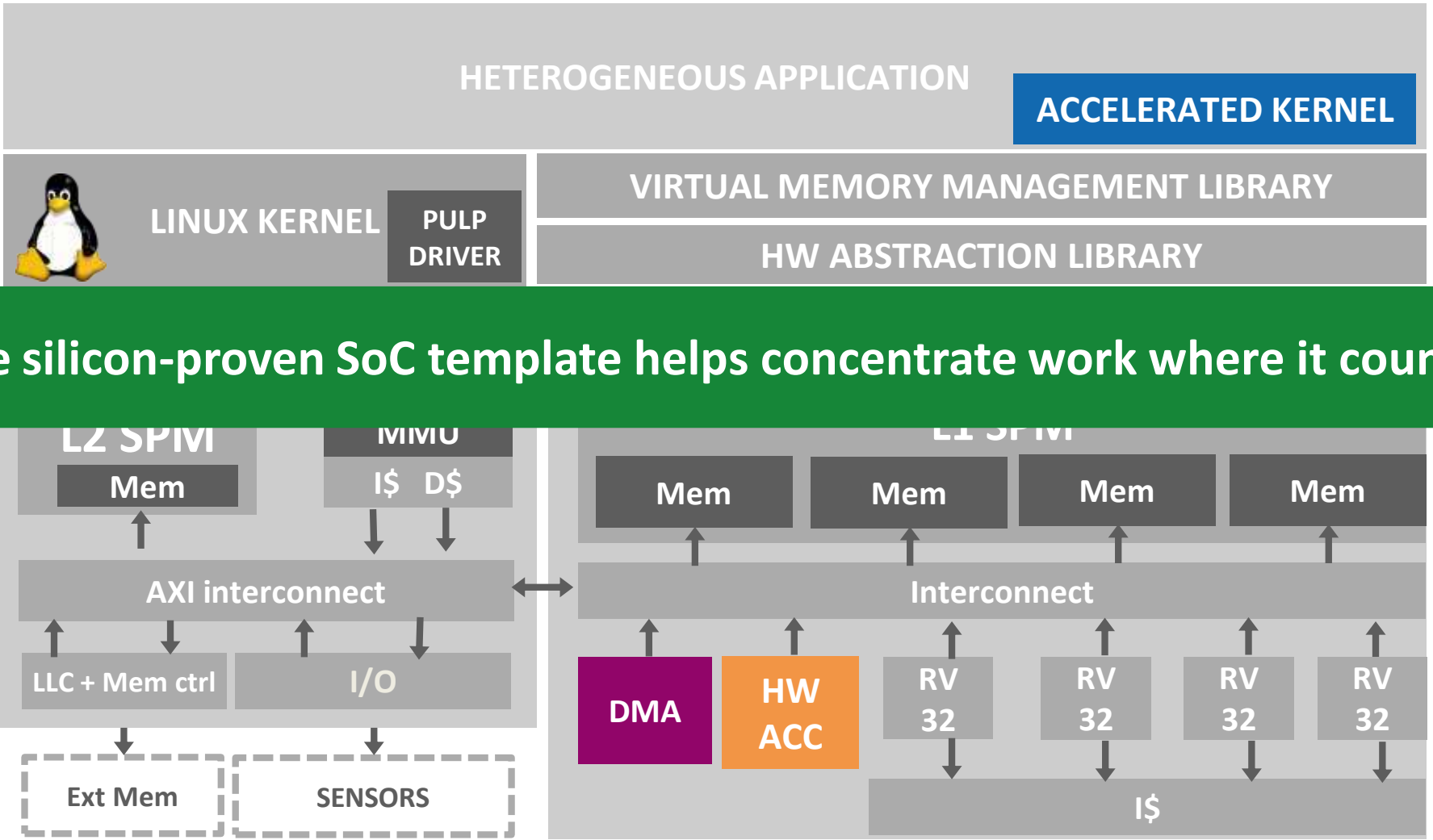
80 MGE

What used to be a complete chip is now a small part of a SoC !

There is so much that makes up a modern SoC



In a typical design, innovation is only in a limited scope



What PULP provides is a box of building blocks

RISC-V Cores

RI5CY	Ibex	Snitch	Ariane
32b	32b	32b	64b

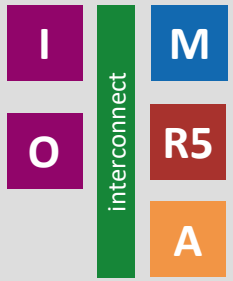
Peripherals

JTAG	SPI
UART	I2S
DMA	GPIO

Interconnect

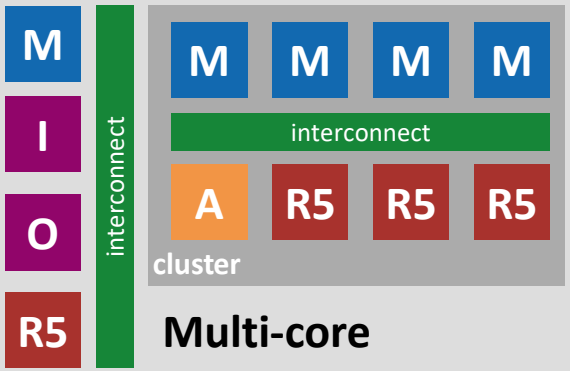
- Logarithmic interconnect
- APB – Peripheral Bus
- AXI4 – Interconnect

Platforms



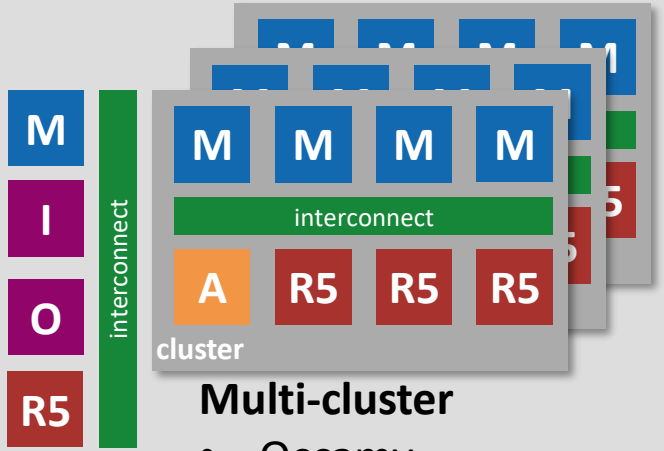
Single Core

- PULPino
- PULPissimo



Multi-core

- OpenPULP
- Mr. Wolf



Multi-cluster

- Occamy
- Hero

IOT → **HPC**

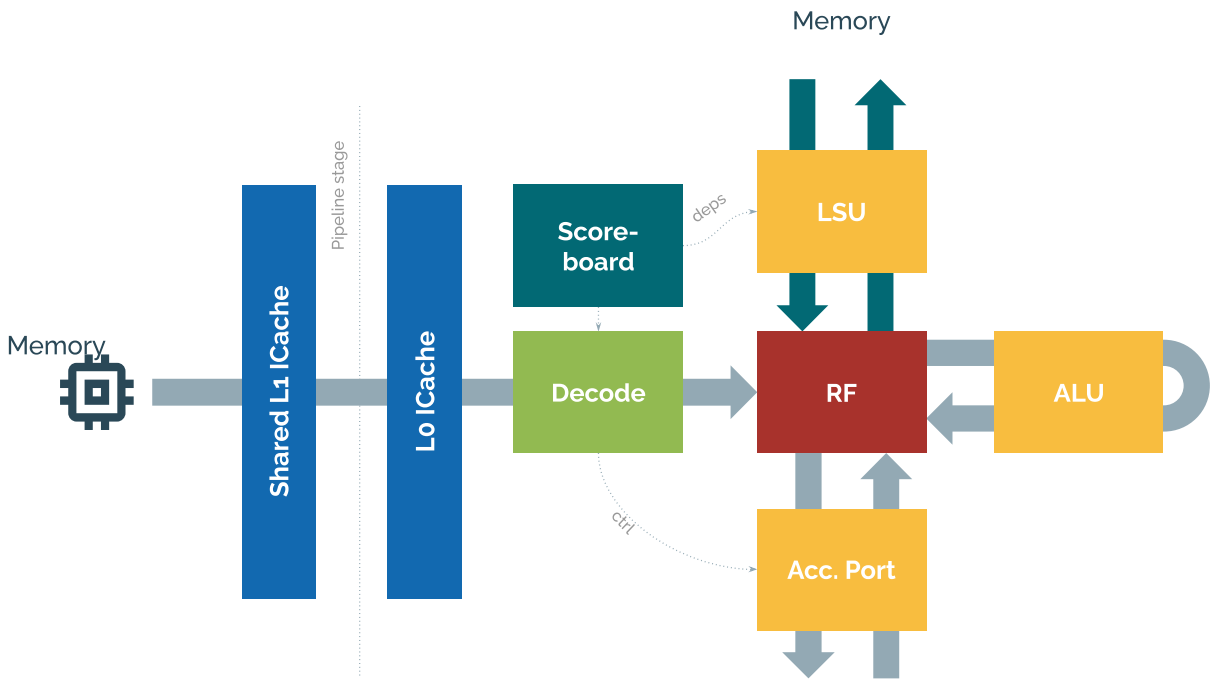
Accelerators

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

Why is RISC-V so special: Freedom to Explore and **Fail!**

- **The ISA provides a contract between HW and SW**
  - As long as you stick to the ISA, you can develop HW and SW independently
  - All RISC-V research in HW can continue to rely on growing SW ecosystem for RISC-V
- **RISC-V comes with plenty of options for extensions**
  - There are reserved encoding spaces for instruction set extensions
- **Being able to change everything gives great flexibility**
  - Do you want 33 registers, or a 48 bit accumulator.. No problem
  - You need to bring the SW support for your additions.

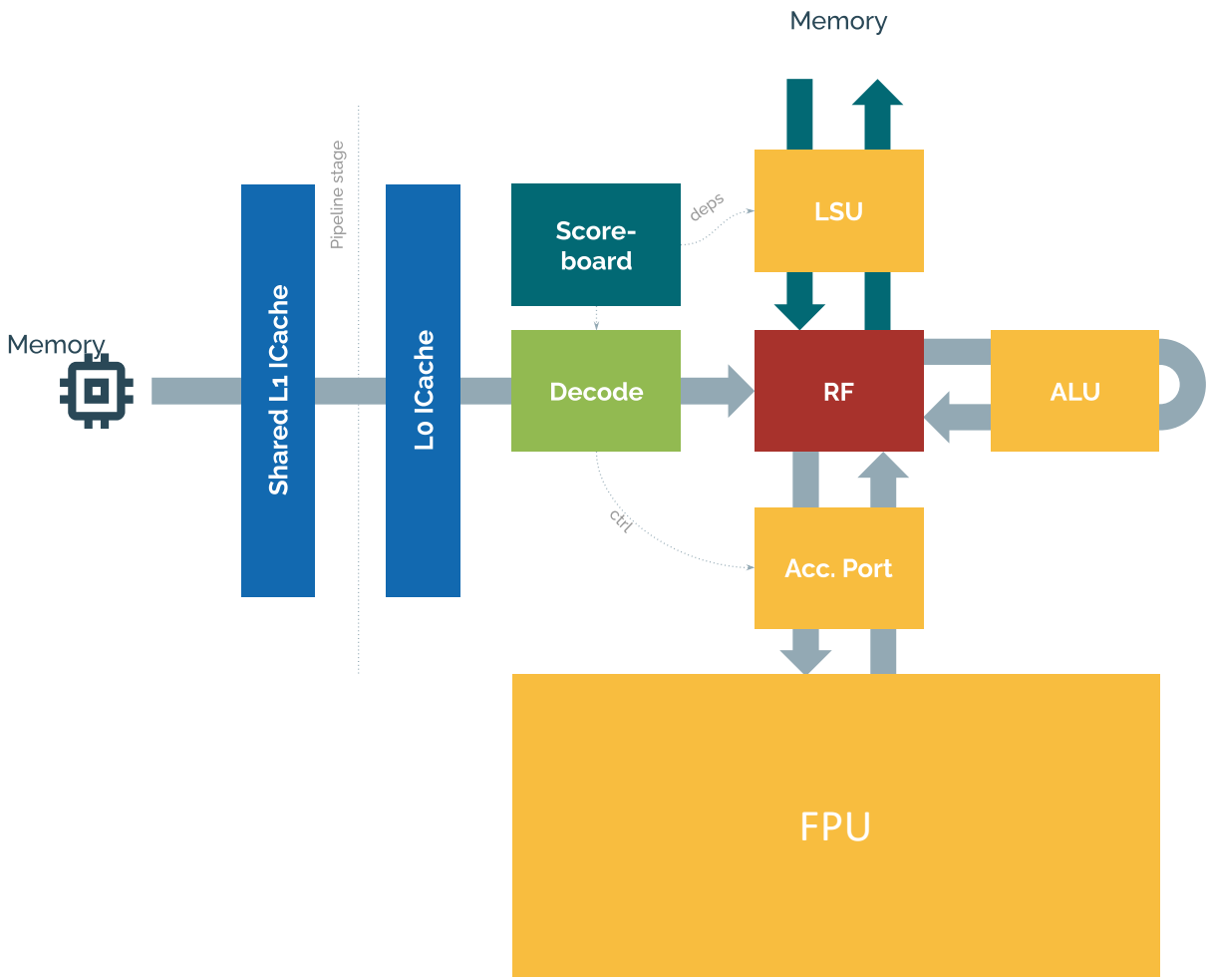
What if we had a tiny 32b core



Introducing SNITCH

- **Start with a simple RISC-V core**
- **Focus on key features:**
  - Lightweight microarchitecture
  - Extensibility: Performance through ISA extensions
  - Latency tolerant
  - Competitive frequency
- **Around 15-25 kGE**

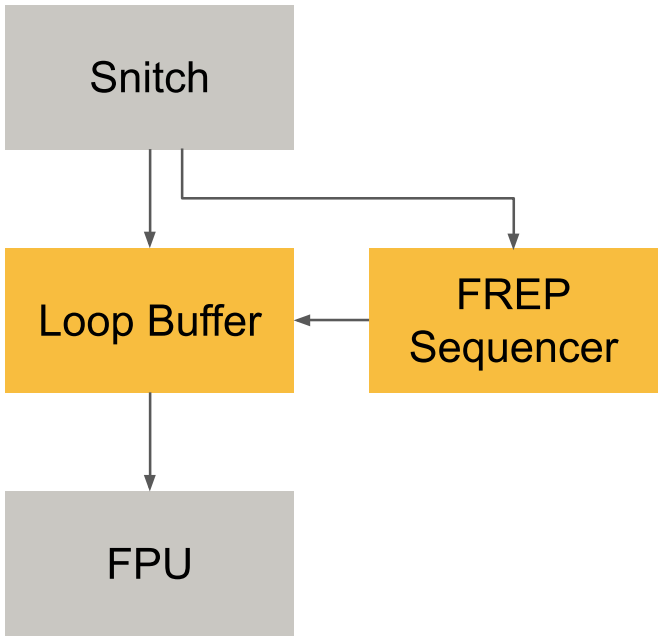
What if we had a tiny 32b core and add a big 64b FPU



Introducing SNITCH

- **Start with a simple RISC-V core**
- **Focus on key features:**
  - Lightweight microarchitecture
  - Extensibility: Performance through ISA extensions
  - Latency tolerant
  - Competitive frequency
- **Around 15-25 kGE**
- **Capable 64b FPU with many extensions**

What if we add a Floating-point Repetition Buffer? (FREP)



Remove control flow overhead

- **Programmable micro-loop buffer**
- **Sequencer steps through the buffer, independently of the FPU**
- **Integer core free to operate in parallel: Pseudo-dual issue**
- **High area- and energy-efficiency**

```
  mv     r0, zero                    frep  r1, 1
loop:                          →    loop:
  addi   r0, 1                         fmadd r2, ssr0, ssr1
  fmadd  r2, ssr0, ssr1
  bne    r0, r1, loop
```

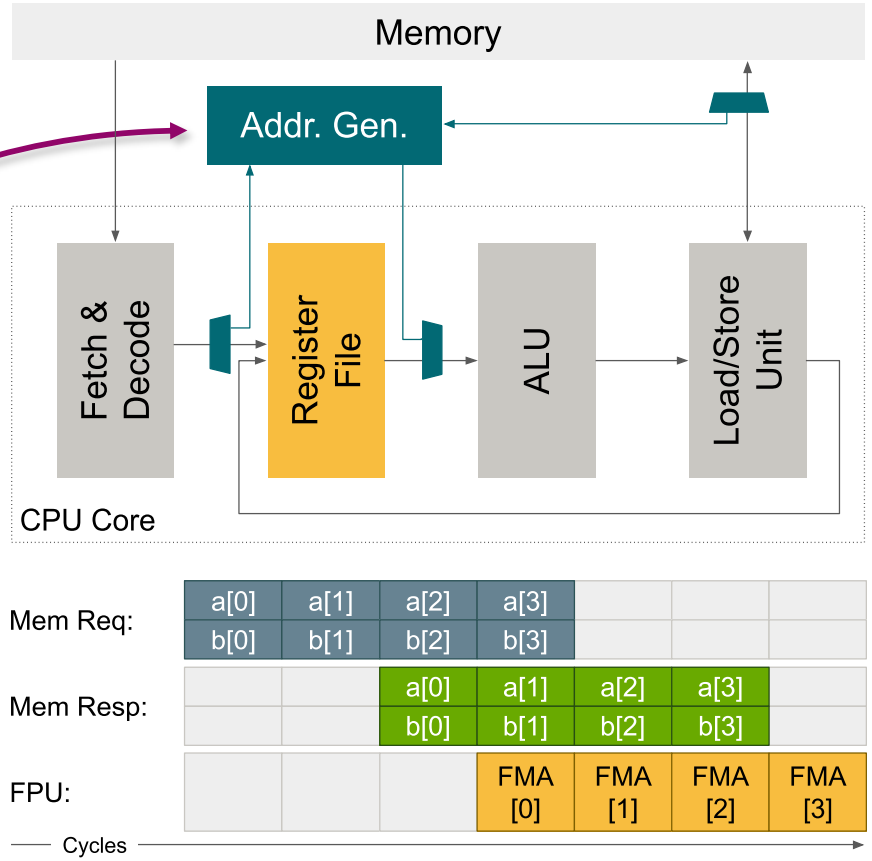
RISC-V

Allows custom instruction set extensions

What if we could stream data to from FPU directly? (SSR)

- **Intuition:**
  High FPU utilization ≈ high energy-efficiency
  - Idea: Turn register R/W into memory loads/stores.
  - Extension around the core's register file
  - Address generation hardware

```
loop:                          scfg 0, %[a], ldA
fld r0, %[a]                   scfg 1, %[b], ldB
fld r1, %[b]          ——→      loop:
fmadd r2, r0, r1               fmadd r2, ssr0, ssr1
```



- **Increase FPU/ALU utilization by ~3x up to 100%**
- **SSRs ≠ memory operands**
  - Perfect prefetching, latency-tolerant

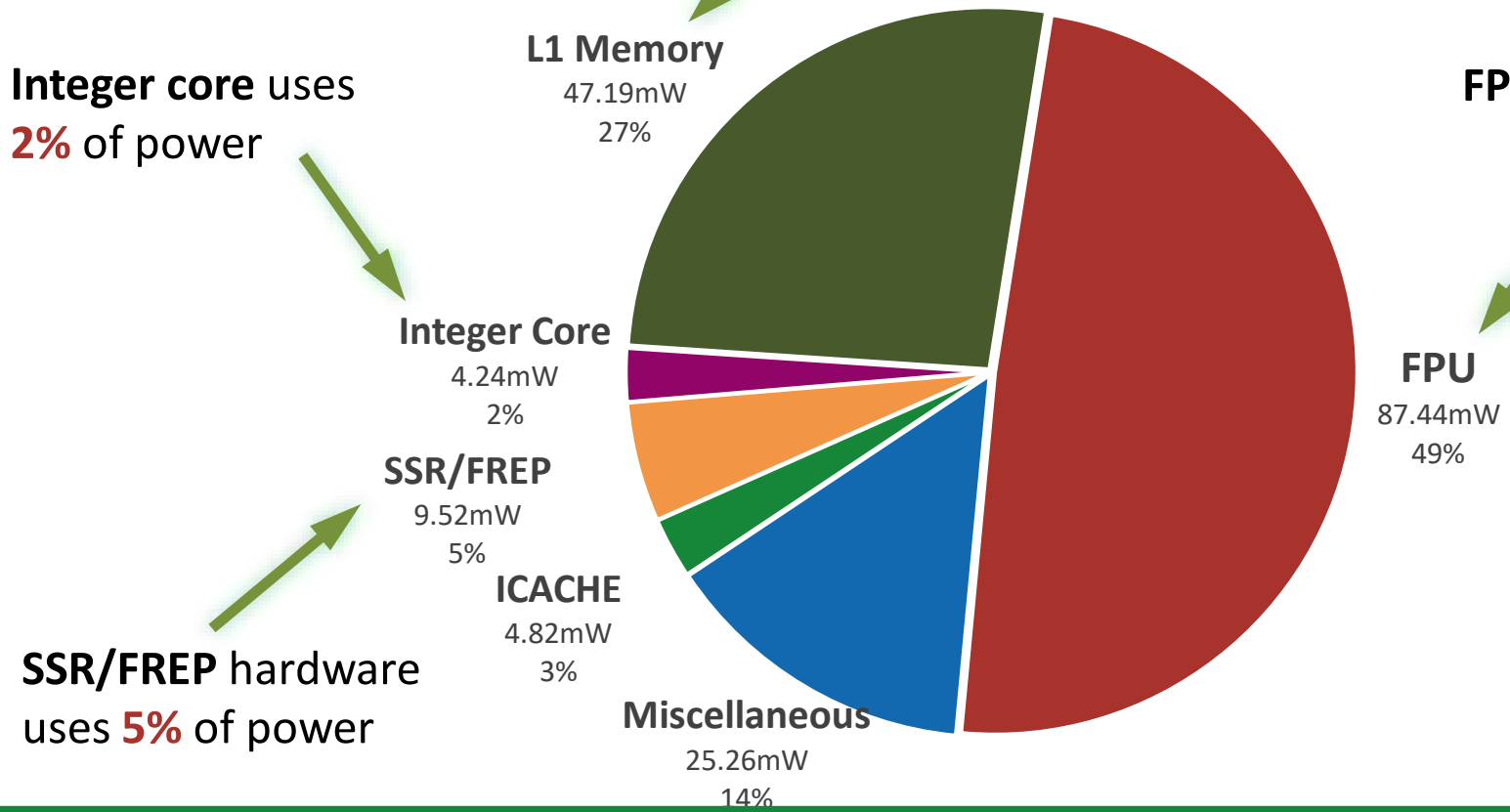
We have a processor that maximizes FPU efficiency

In an 8-core cluster

Inevitable to have local memory
(e.g., GPU/GPU L1 cache, vector register file)

Integer core uses
2% of power

FPU uses **50%** of power



SSR/FREP hardware
uses **5%** of power

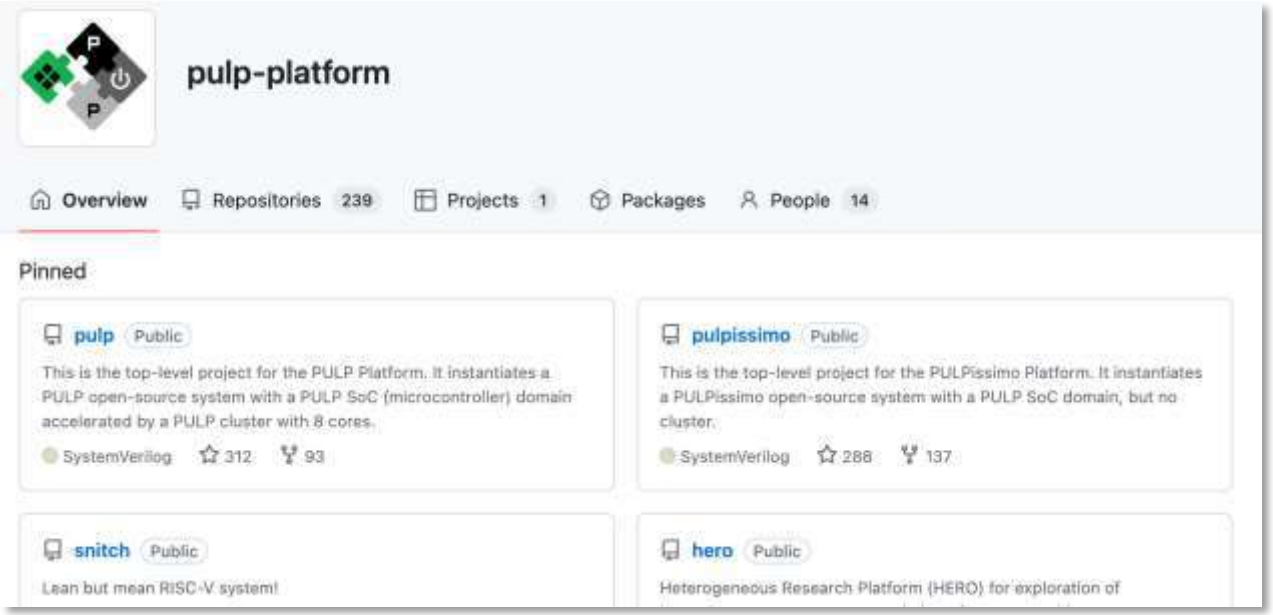
The flexibility of open ISA made it easy for us to explore such an approach

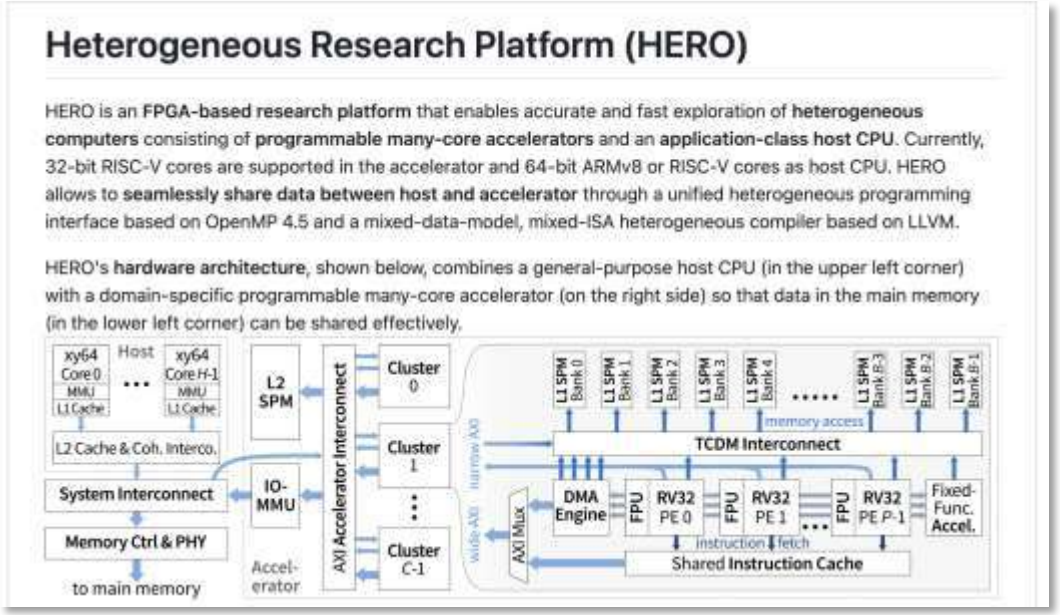
PULP uses a permissive open source license

- **All our development is on GitHub**
  - HDL source code, testbenches, software development kit, virtual platform

https://github.com/pulp-platform

- Allows anyone to use, change, and make products without restrictions.

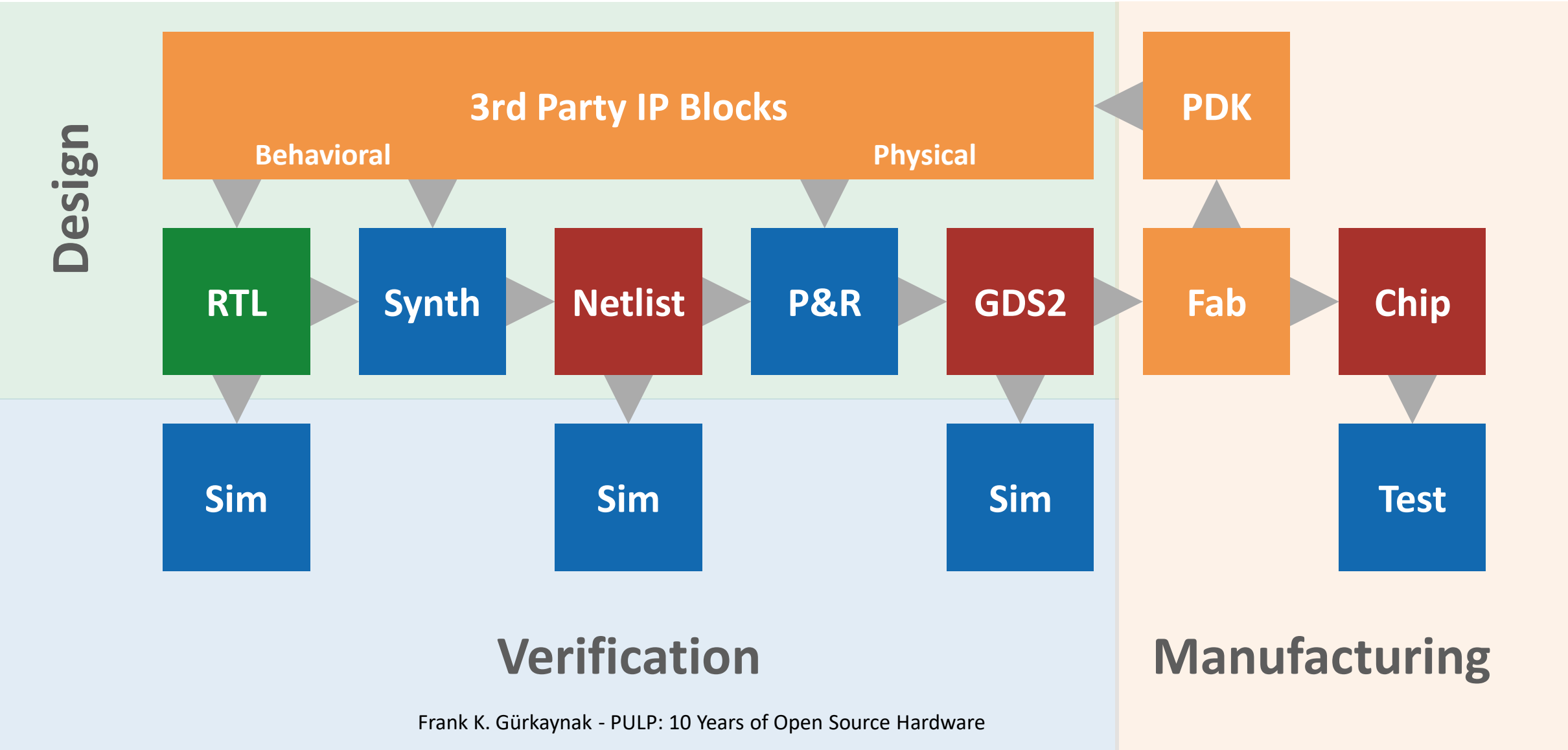




Open Source Hardware licensing still a critical issue

- **Two main flavors, divided opinion**
  - **Permissive** (Apache, MIT, BSD..): Favored by the industry, minimum obligations
  - **Reciprocal** (GPL, LGPL,..): Feared by industry
- **In theory, it should be possible to have reciprocal licensing for open hardware**
  - For example text of LGPL problematic for IC Design use.
  - Cern OHL (https://cern-ohl.web.cern.ch/), comes in many flavors (reciprocal, permissive)
  - Still more work needed, not many people understand issues of IC Design
  - Lawyers (in companies) prefer well-known licenses (less work for them).
- **PULP uses Solderpad (http://solderpad.org/licenses/)**
  - Permissive license based on Apache
  - Clarifications for hardware use added by Andrew Katz
  - Had no issues (so far) neither with academic nor industrial collaborations

At the moment most of our Open Hardware is still *only* RTL



State of Open Source for Hardware: Rapid Developments

TYPE	EXAMPLES	STATUS
Open Specifications	RISC-V	Established
Architectures	PULP	Quite mature
Implementations in RTL	Snitch, Hero	Many

State of Open Source for Hardware: Rapid Developments

TYPE	EXAMPLES	STATUS
Open Specifications	RISC-V	Established
Architectures	PULP	Quite mature
Implementations in RTL	Snitch, Hero	Many
Open source Hard IP	FLL, DDR PHY..	Very Limited

State of Open Source for Hardware: Rapid Developments

TYPE	EXAMPLES	STATUS
Open Specifications	RISC-V	Established
Architectures	PULP	Quite mature
Implementations in RTL	Snitch, Hero	Many
Open source Hard IP	FLL, DDR PHY..	Very Limited
Process Design Kits	Skywater 130nm	Just Started

State of Open Source for Hardware: Rapid Developments

TYPE	EXAMPLES	STATUS
Open Specifications	RISC-V	Established
Architectures	PULP	Quite mature
Implementations in RTL	Snitch, Hero	Many
Open source Hard IP	FLL, DDR PHY..	Very Limited
Process Design Kits	Skywater 130nm	On its way
Open Source Tools	Open Lane	Quite usable

What is PULP doing to 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**
  - Provide support
  - Develop industrial verification
  - Governance of open source repositories
- **Happy to receive this help from**
  - Open HW group (Ariane -> CVA6, RI5CY -> CV32E40P)
  - LowRISC (ZeroRiscy -> Ibex)
  - Others?

GROUP OPENHW PROVEN PROCESSOR IP

Academic open source → Industrial open source

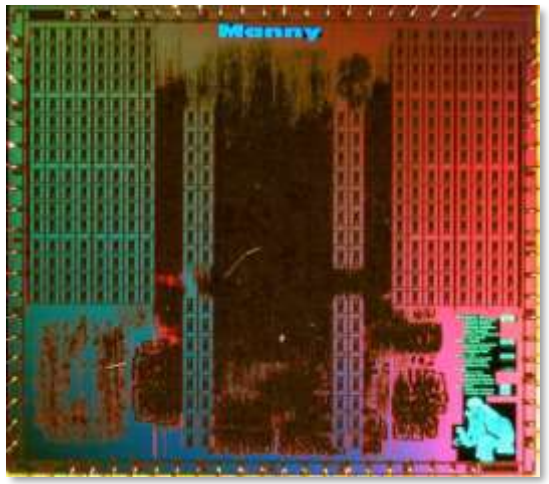
- OpenHW Group is a not-for-profit, global organization (EU,NA,Asia) driven by its members and individual contributors where HW and SW designers collaborate in the development of open-source cores, related IP, tools and SW such as the Core-V family of cores.
- OpenHW Group provides an infrastructure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

PULP has been used in all our publicly funded research

Swiss Funding (nano-tera)

ICYSoC

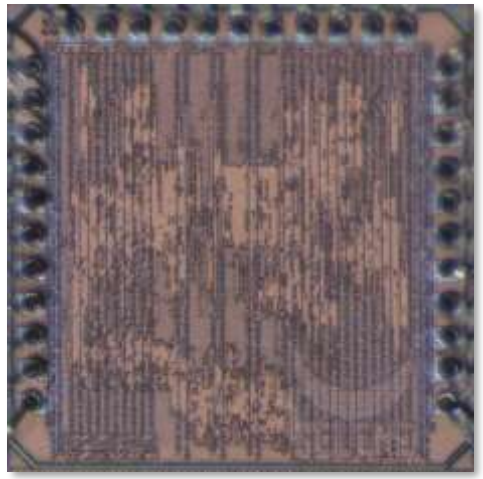
Near threshold computing



EU funded research projects

OPRECOMP

Approximate computing
Multi-precision arithmetic



European FPA

EPI / EUPilot

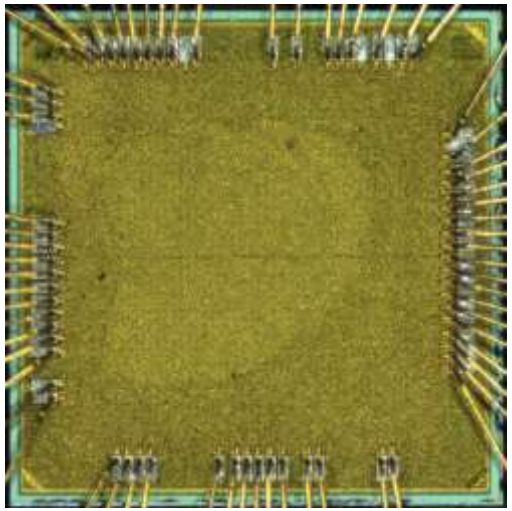
Snitch based accelerators

International projects

MITACS / Polara

Vector extensions
for RISC-V processors

Industrial Collaborations

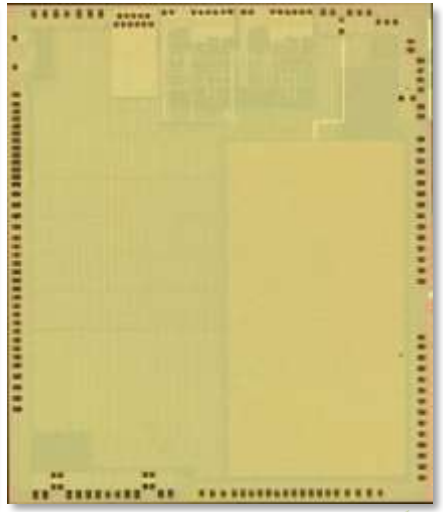


PULPv1,2,3 (ST28 FD)
Demonstrators of 28nm FD-SOI capabilities

Various publications ‘15 – ‘18

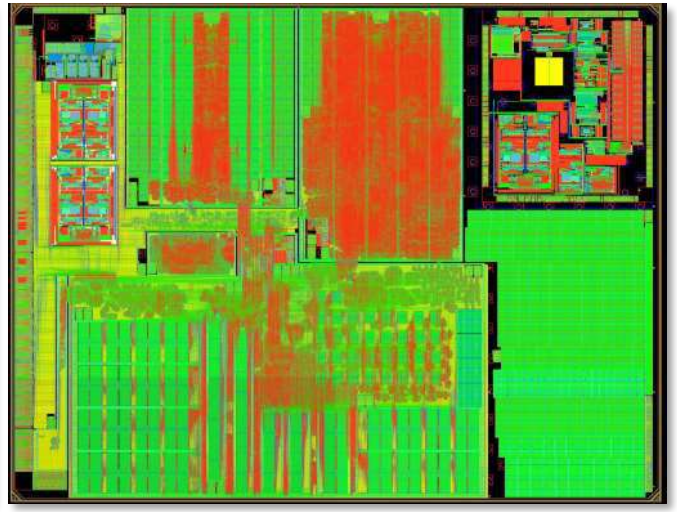
Arnold (GF22)
IoT SoC combining eFPGA with RISC-V core

Schiavone et al TVLSI ‘19



Vega (GF22)
IoT Processor with ML acceleration

Rossi et al ISSCC ‘21
Rossi et al JSSC ‘22

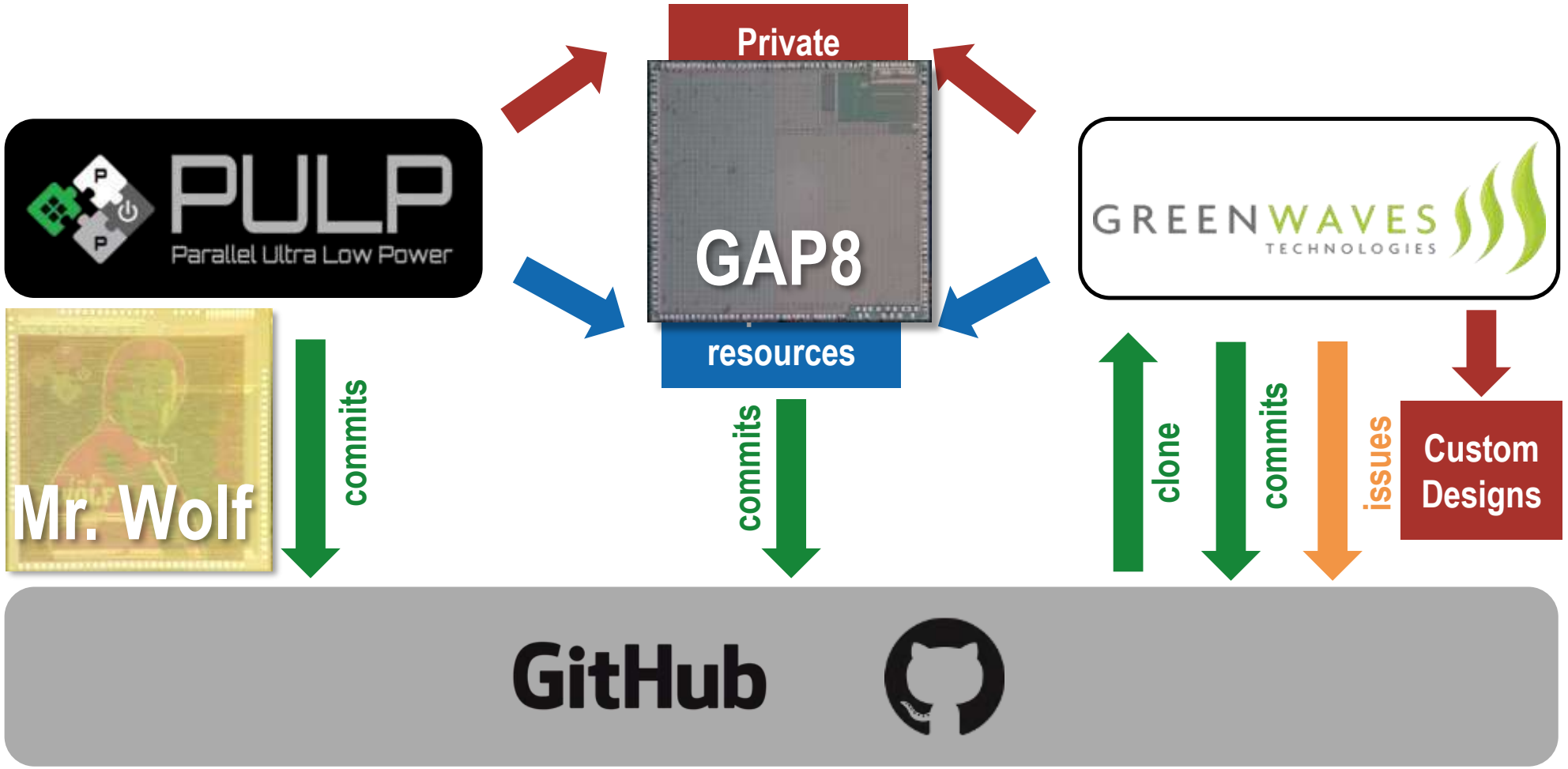


Marsellus (GF22)
IoT Processor with low power modes and AI Accelerators

Conti et al ISSCC ‘23

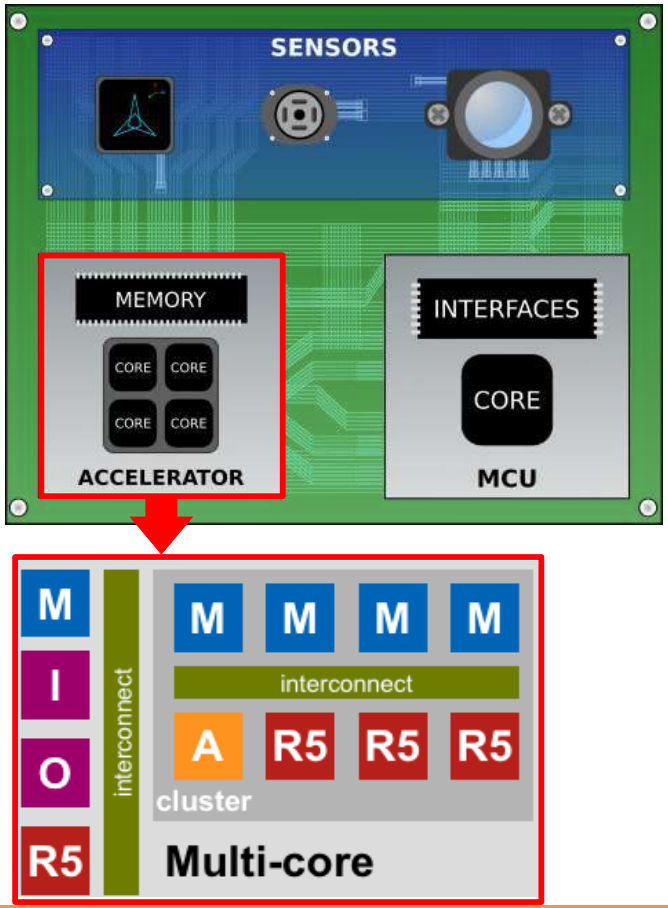
Currently working with Meta, Intel, GF, IHP, PragmatIC, IIT

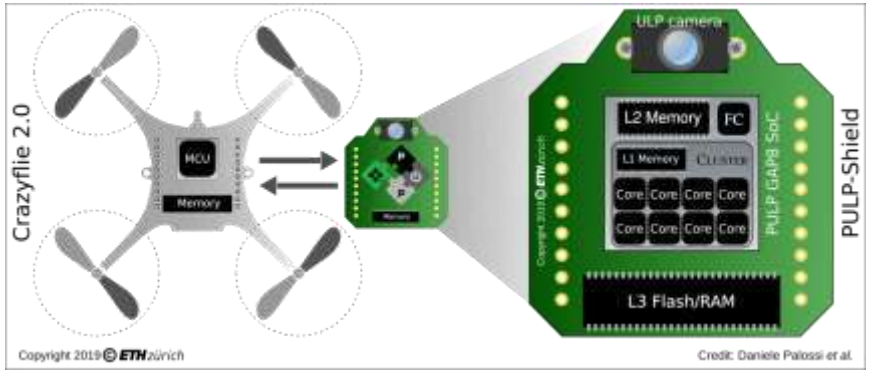
Open source collaboration scheme explained



The PULP-Based Commercial Nanodrone Platform

ULP heterogeneous model [1] → **PULP-Shield [2]** → **AI-deck**





- ~ 5 g – 30x28 mm
- PULP GAP8 SoC
- Off-chip DRAM/Flash
- QVGA ULP Camera

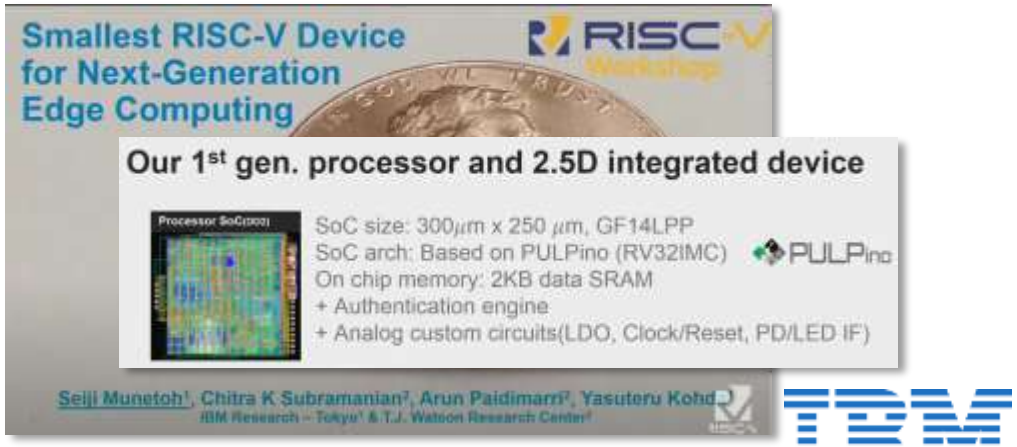
GREENWAVES TECHNOLOGIES

ALTIUM DESIGNER

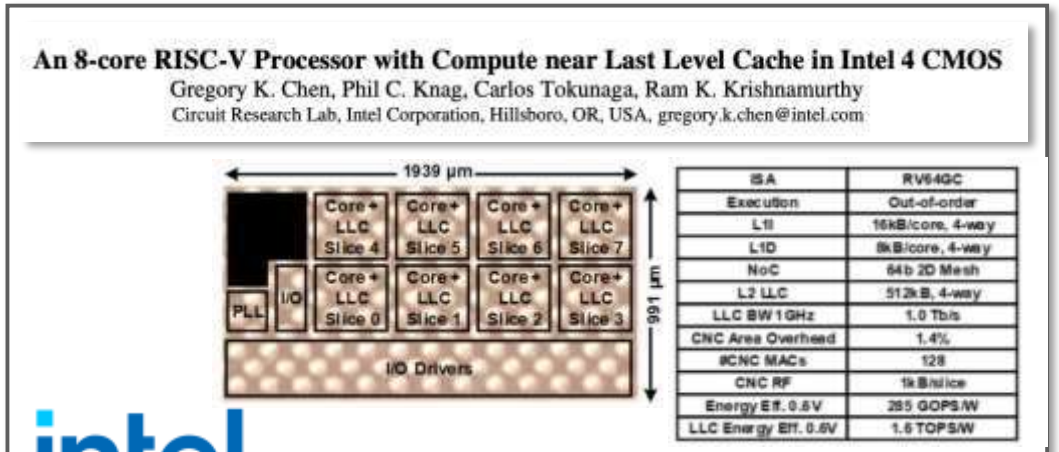
- ~ 8 g – 40x28 mm
- PULP GAP8 SoC
- Off-chip DRAM/Flash
- QVGA ULP Camera
- WiFi module

More on our work on Drones – talk by Lorenzo – Today 16:30

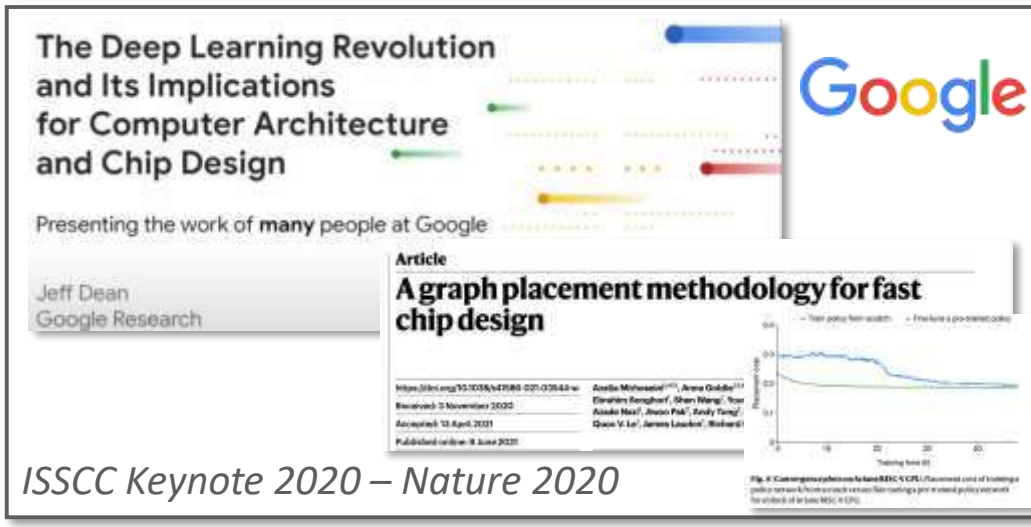
So proud to have supported others in their research



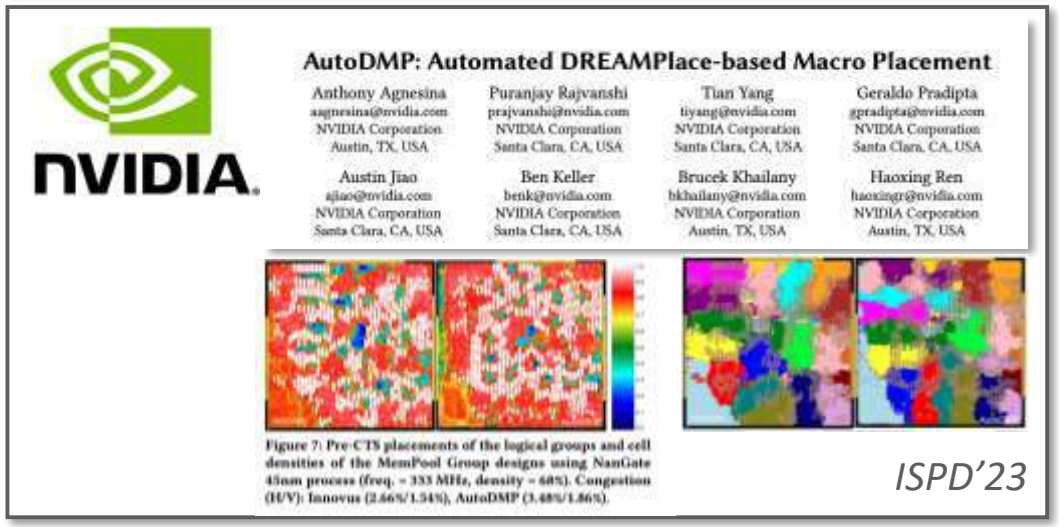
RISC-V week Barcelona 2018



VLSI Symposium 2022



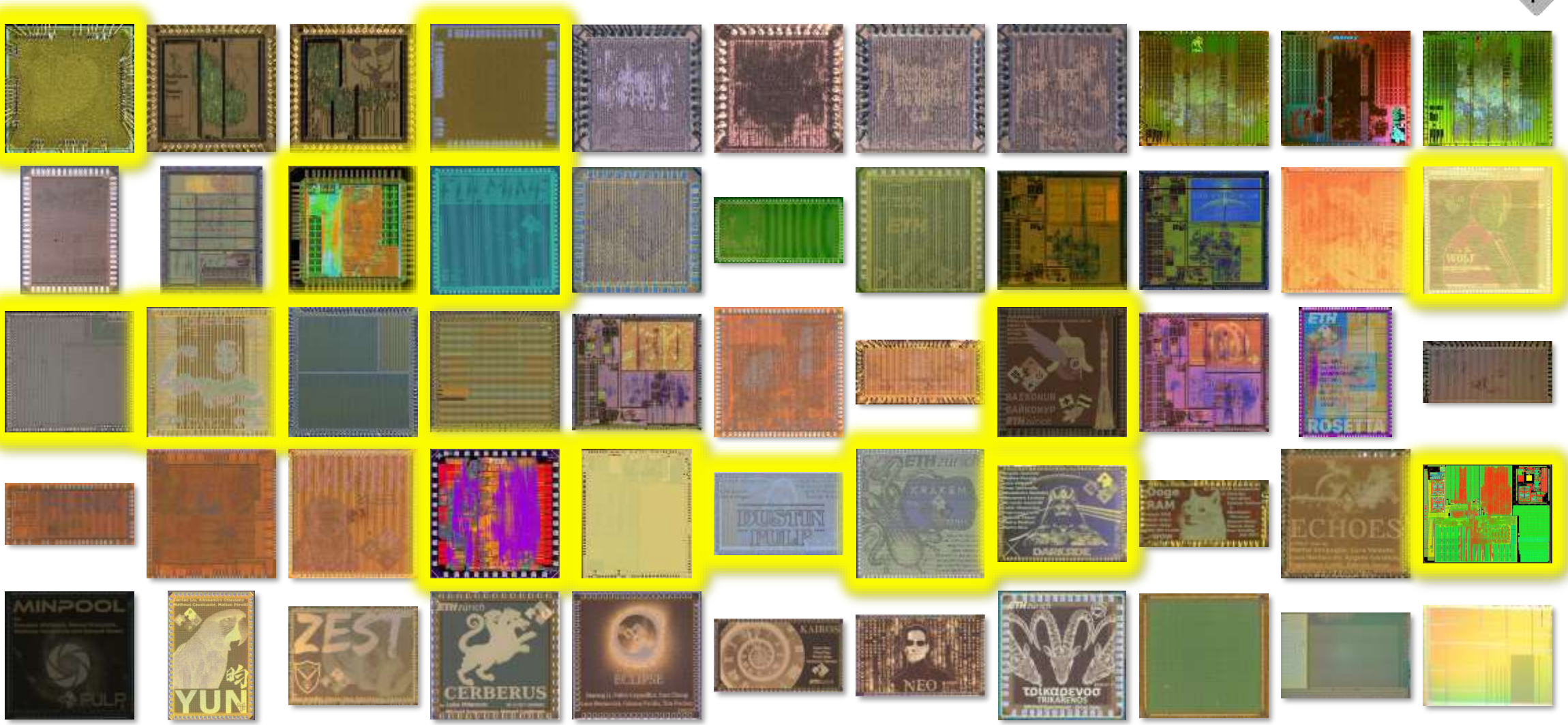
ISSCC Keynote 2020 – Nature 2020



ISPD'23

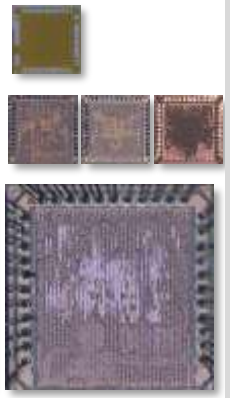
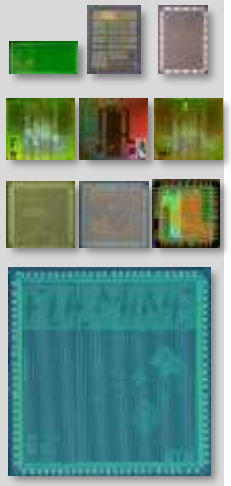
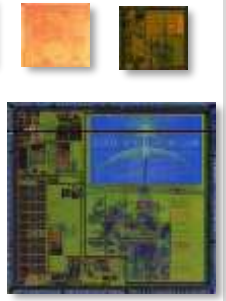
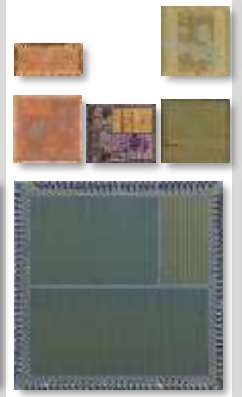
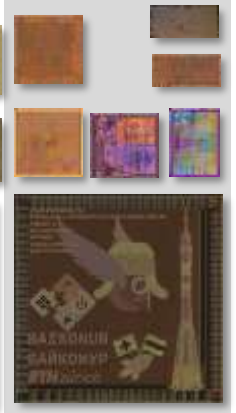
PULP chips until now

PULP chips until now – major publications



PULP chips until now – 64b/32b cores

PULP chips until now – 55 manufactured – 5 on the way

2013 (3)	2014 (5)	2015 (10)	2016 (3)	2017 (2)	2018 (6)	2019 (7)	2020 (3)	2021 (7)	2022 (7)
									
PULPv1 *STM 28FDSOI* Multi-core processor	**Diana** *UMC 65* 4-core system with approximate FPUs	**Fulmine** *UMC 65* 4-core system with ML and Crypto accelerators	**VivoSoC 2.001** *SMIC 130* Mixed signal system for biosignal acquisiton	**Mr. Wolf** *TSMC 40* 8+1 core IoT processor	**Poseidon** *GF 22FDX* 64bit RISC-V core, 32bit Microcontroller system, ML rator	**Baikonur** *GF 22FDX* Dual 64bit RISC-V core, 3x 8core snitch clusters, Body biasing test vehicle	**Dustin** *TSMC 65* IoT processor with 16 cores and QNN enhancements	**Kraken** *GF 22FDX* IoT processor with Spiking Neural and Ternary Inference Engines	**Occamy** *GF 12LPP* ML accelerator with 216 + 1 cores and HBM interface

Check http://asic.ethz.ch for all our chips

Coming soon from the PULP team

FlexIBEX

IoT processor with a twist a kHz range design

Pragmatic

Iguana

Going all the way in open source

ihp

Carfield

Cars (and cats) can also use a bit of PULP

intel

Occamy

Bringing up the beast

A bit of a public service announcement

PULP

Parallel Ultra Low Power

There is much more to come …

http://pulp-platform.org

@pulp_platform