

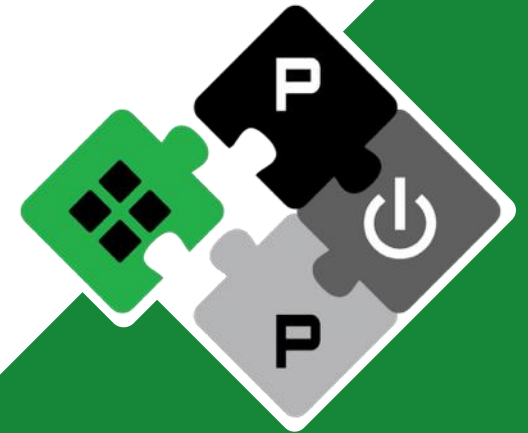
# Open-source SoC design using PULP

Austrochip 2025, Linz

**Frank K. Gürkaynak** kgf@iis.ee.ethz.ch

## **PULP Platform**

Open Source Hardware, the way it should be!



@pulp\_platform 

pulp-platform.org 

youtube.com/pulp\_platform 

# Team of 100 people in ETH Zürich – University of Bologna



- Research on open-source energy-efficient computing





# In 12 years PULP team has designed more than 60 chips



**RISC-V and open-source hardware have been instrumental in our success**



Including complex chiplet based designs like this one

# OCCAMY

432 RISC-V cores

Chiplets

GF12nm

1GHz

You can find all our designs in our **Chip Gallery** <http://asic.ethz.ch>

# What I want to share today



## Why open source?

- Managing complexity, Collaboration, Exploitation, Education

## A sample of our projects

- Croc, Cheshire, Kraken, Occamy, Picobello

## What is next for open-source hardware

- Challenges and opportunities

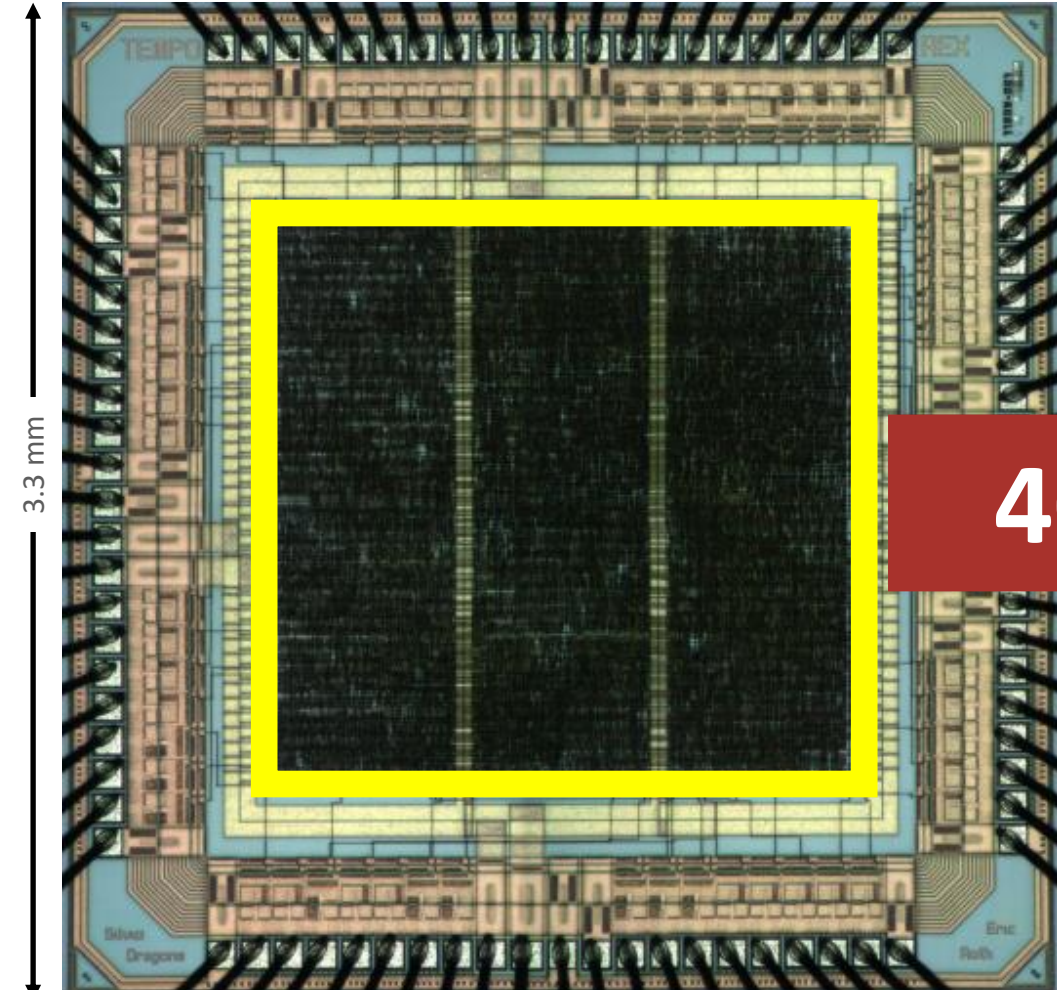


## Why open source?

- Managing complexity
- Collaboration
- Exploitation
- Education



# In the last 20 years IC Design has changed a lot



4000 x



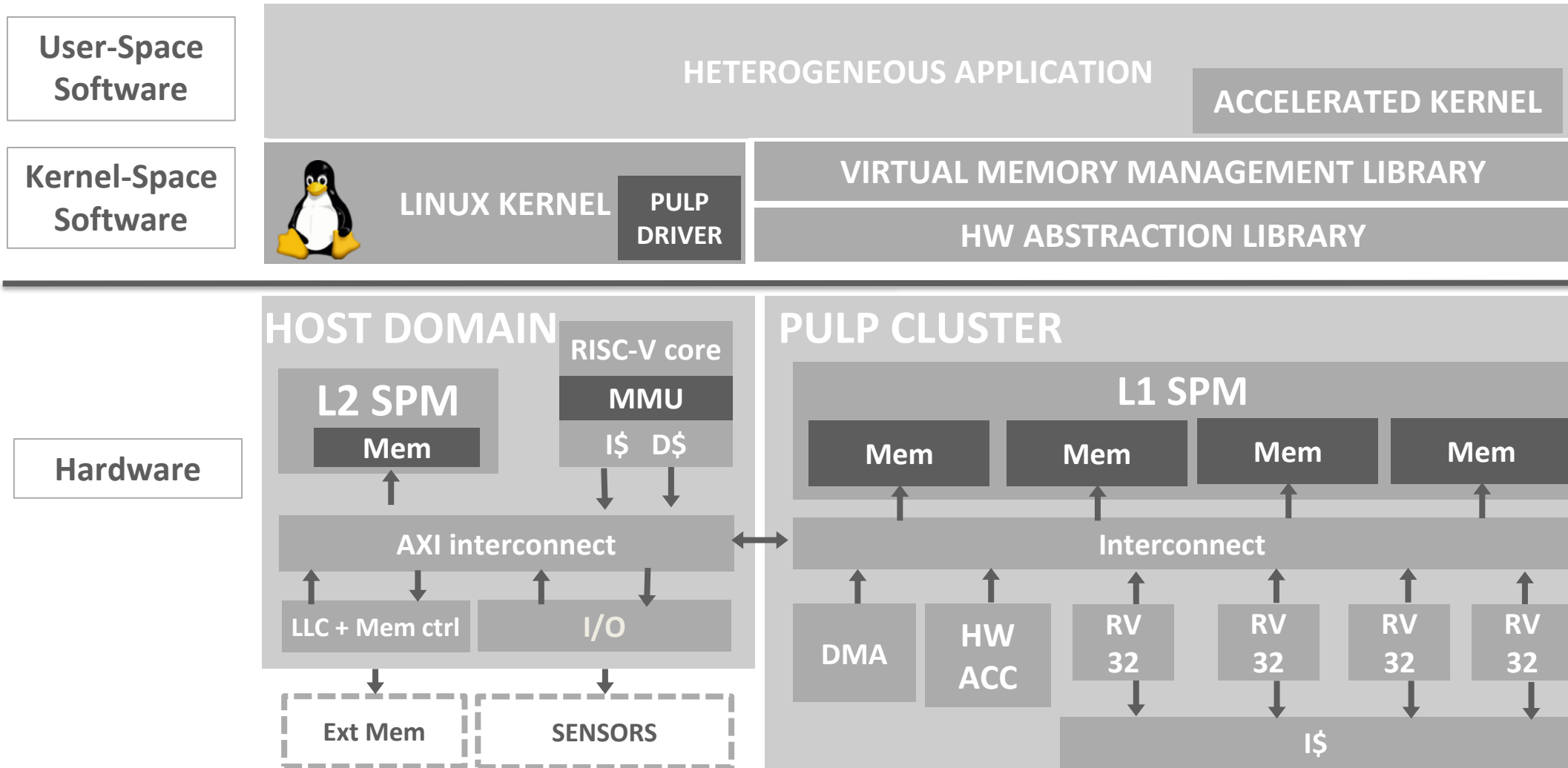
What used to be a complete chip is now a small part of a SoC !

80 MGE

chip 2025, Linz

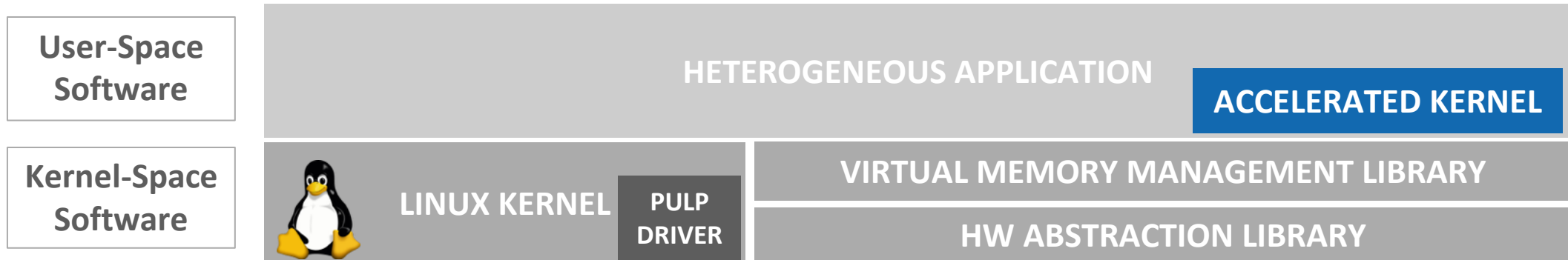
7

# There is so much that makes up a modern SoC

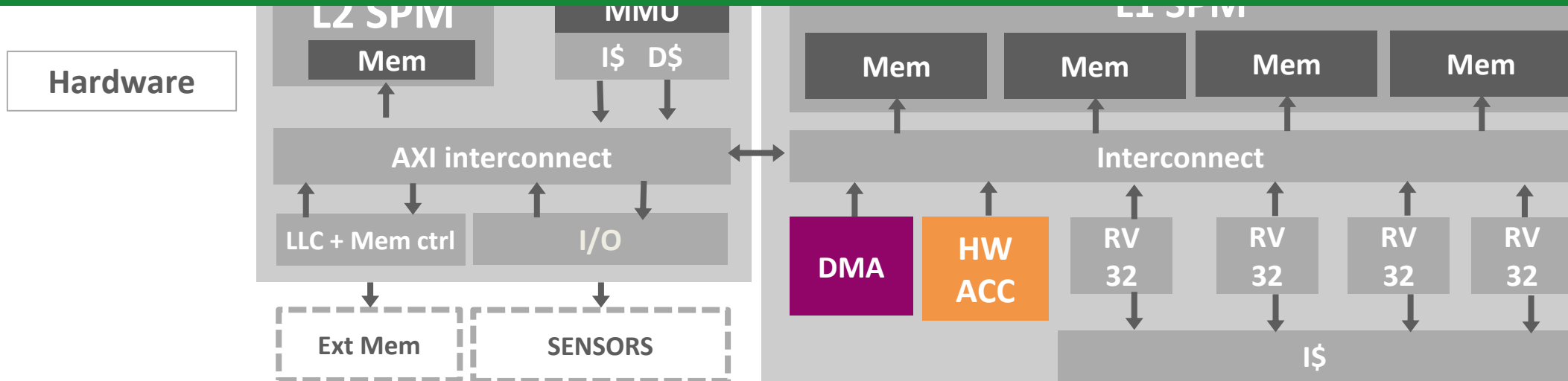




# In a typical design, innovation is only in a limited scope



Open-source silicon-proven IPs helps concentrate work where it counts



# All of our designs are open-source hardware



- All our development is on GitHub using a permissive license
  - HDL source code, testbenches, software development kit, virtual platform

<https://github.com/pulp-platform>



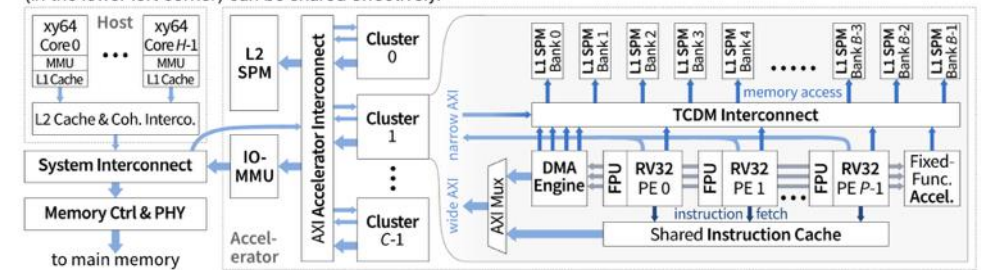
- Allows anyone to use, change, and make products without restrictions.

The screenshot shows the GitHub profile for 'pulp-platform'. It includes a repository overview with 239 repositories, 1 project, and 14 people. The 'Pinned' section features four repositories: 'pulp' (Public), 'pulpissimo' (Public), 'snitch' (Public), and 'hero' (Public). Each repository has a brief description and statistics like stars and forks.

## Heterogeneous Research Platform (HERO)

HERO is an FPGA-based research platform that enables accurate and fast exploration of heterogeneous computers consisting of programmable many-core accelerators and an application-class host CPU. Currently, 32-bit RISC-V cores are supported in the accelerator and 64-bit ARMv8 or RISC-V cores as host CPU. HERO allows to seamlessly share data between host and accelerator through a unified heterogeneous programming interface based on OpenMP 4.5 and a mixed-data-model, mixed-ISA heterogeneous compiler based on LLVM.

HERO's hardware architecture, shown below, combines a general-purpose host CPU (in the upper left corner) with a domain-specific programmable many-core accelerator (on the right side) so that data in the main memory (in the lower left corner) can be shared effectively.





# What PULP provides is a box of building blocks



## RISC-V Cores and Vector Units

RI5CY <i>CV32E</i>	Zero R <i>lbex</i>	Snitch	Spatz	Ariane <i>CVA6</i>	ARA
RV32	RV32	RV32	RVV	RV64	RVV

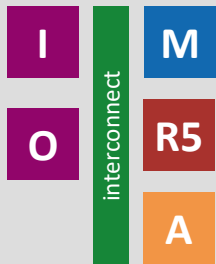
## Peripherals

JTAG	SPI
UART	I2S
DMA	GPIO

## Interconnects

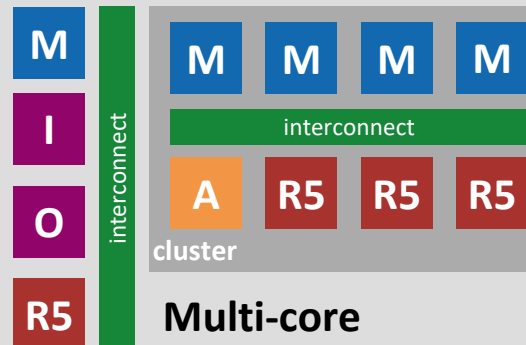
LIC	HCI
APB	FlooNoC
AXI4	

## Platforms



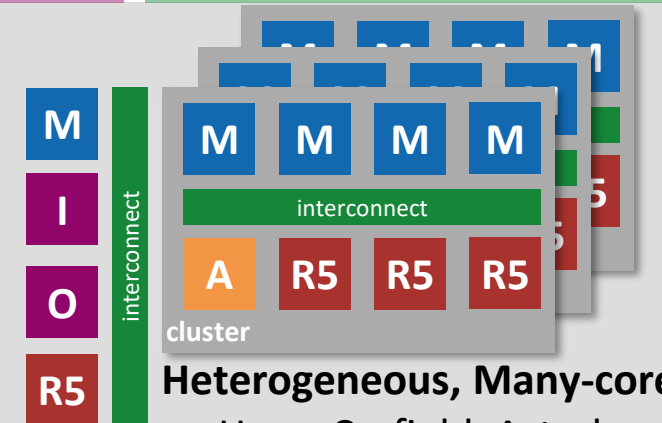
### Single core

- Croc, PULPissimo
- Cheshire



### Multi-core

- OpenPULP
- ControlPULP



### Heterogeneous, Many-core

- Hero, Carfield, Astral
- Occamy, Mempool

# IOT

# HPC

## Accelerators and ISA extensions

XpulpNN,  
XpulpTNN

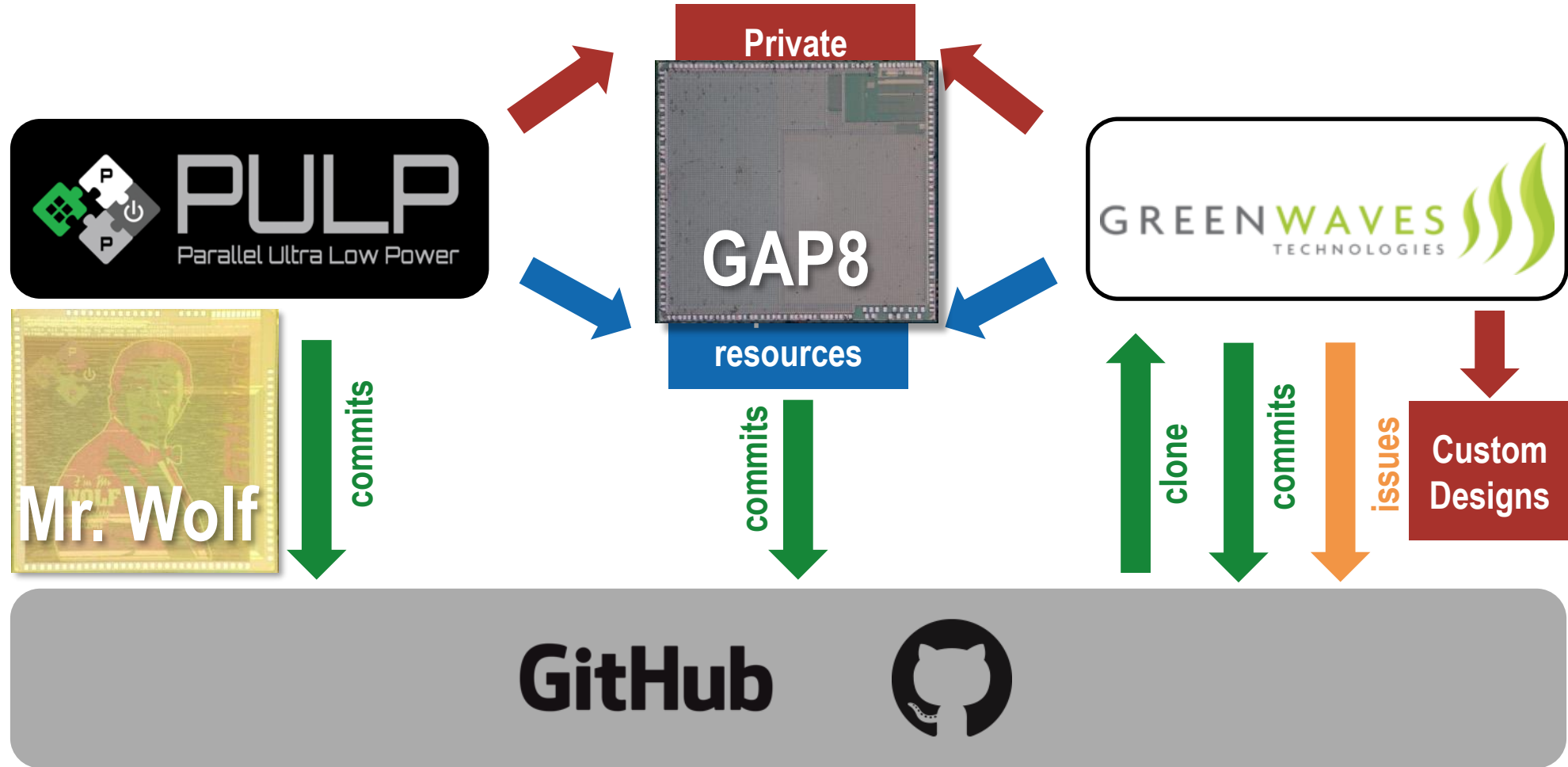
ITA  
(Transformers)

RBE, NEUREKA  
(QNNs)

FFT  
(DSP)

REDMULE  
(FP-Tensor)

# How does PULP collaborate with 3<sup>rd</sup> parties?





# Diverse set of open source based industry collaborations



GF22 (2018)

## Arnold

*eFPGA coupled with a RISC-V microcontroller.*

*In one year from agreement to actual tapeout*

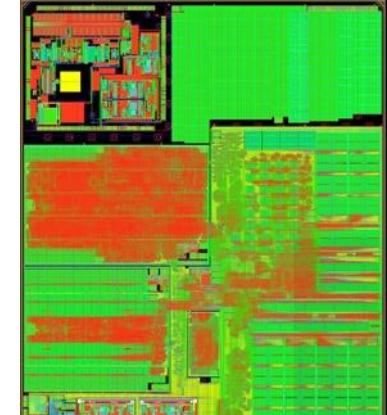


GF22 (2022)

## Marsellus

*Heterogeneous IoT processor  
With Aggressive voltage scaling*

DOLPHIN

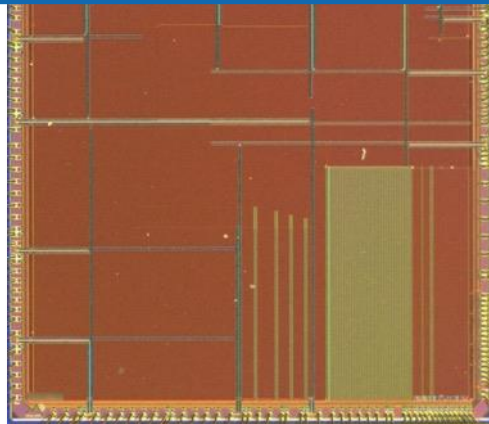


Permissive open-source licensing key to our industrial relationships

## Siracusa

*SoC for Extended Reality  
visual processing*

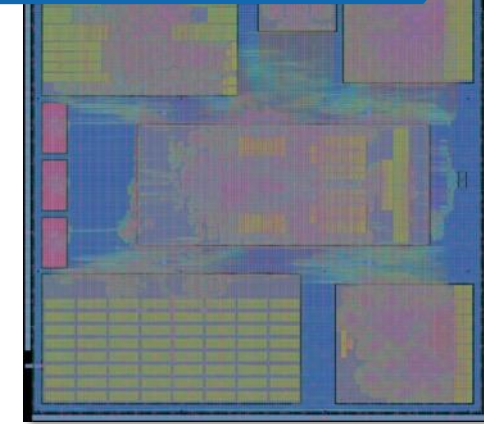
 Meta



## Carfield

*Open-Research platform for  
safety, resilient and  
time-predictable systems*

 intel®



# And many continue to use our work for their research



**Smallest RISC-V Device for Next-Generation Edge Computing**

**RISC-V Workshop**

**Our 1<sup>st</sup> gen. processor and 2.5D integrated device**

Processor SoC (002)

SoC size: 300  $\mu\text{m}$  x 250  $\mu\text{m}$ , GF14LPP  
 SoC arch: Based on PULPino (RV32IMC) + PULPino  
 On chip memory: 2KB data SRAM  
 + Authentication engine  
 + Analog custom circuits (LDO, Clock/Reset, PD/LED IF)

Seiji Munetoh<sup>1</sup>, Chitra K Subramanian<sup>2</sup>, Arun Paidimarri<sup>2</sup>, Yasuteru Kohda<sup>2</sup>  
 IBM Research – Tokyo<sup>1</sup> & T.J. Watson Research Center<sup>2</sup>

**IBM**

*RISC-V week Barcelona 2018*

**An 8-core RISC-V Processor with Compute near Last Level Cache in Intel 4 CMOS**

Gregory K. Chen, Phil C. Knag, Carlos Tokunaga, Ram K. Krishnamurthy  
 Circuit Research Lab, Intel Corporation, Hillsboro, OR, USA, gregory.k.chen@intel.com

ISA	RV64GC
Execution	Out-of-order
L1I	16kB/core, 4-way
L1D	8kB/core, 4-way
NoC	64b 2D Mesh
L2 LLC	512kB, 4-way
LLC BW	1.0 Tb/s
CNC Area Overhead	1.4%
#CNC MACs	128
CNC RF	1kB/slice
Energy Eff. 0.6V	285 GOPS/W
LLC Energy Eff. 0.6V	1.6 TOPS/W

**intel**

*VLSI Symposium 2022*

**The Deep Learning Revolution and Its Implications for Computer Architecture and Chip Design**

Presenting the work of **many** people at Google

**Google**

**AutoDMP: Automated DREAMPlace-based Macro Placement**

Anthony Agnesina  
 aagnesina@nvidia.com  
 NVIDIA Corporation  
 Austin, TX, USA

Puranjay Rajvanshi  
 prajvanshi@nvidia.com  
 NVIDIA Corporation  
 Santa Clara, CA, USA

Tian Yang  
 tiyang@nvidia.com  
 NVIDIA Corporation  
 Santa Clara, CA, USA

Geraldo Pradipta  
 gpradipta@nvidia.com  
 NVIDIA Corporation  
 Santa Clara, CA, USA

Austin Jiao  
 ajiao@nvidia.com  
 NVIDIA Corporation  
 Santa Clara, CA, USA

Ben Keller  
 benk@nvidia.com  
 NVIDIA Corporation  
 Santa Clara, CA, USA

Bruce Khailany  
 bkhailany@nvidia.com  
 NVIDIA Corporation  
 Austin, TX, USA

Haoxing Ren  
 haoxingr@nvidia.com  
 NVIDIA Corporation  
 Austin, TX, USA

Some smaller companies you might have heard of ☺

Accepted: 13 April 2021  
 Published online: 9 June 2021

Quoc V. Le, James Laudon, Richard...

**ISSCC Keynote 2020 – Nature 2020**

Fig. 4: Convergence plots on Ariane RISC-V CPU. Placement cost of training policy network from scratch versus fine-tuning a pre-trained policy network for a block of Ariane RISC-V CPU.

Figure 7: Pre-CTS placements of the logical groups and cell densities of the MemPool Group designs using NanGate 45nm process (freq. ~ 333 MHz, density ~ 68%). Congestion (H/V): Innovus (2.66%/1.54%), AutoDMP (3.48%/1.86%).

**ISPD'23**



# We rely more and more on open-source HW for teaching



- **Open source RTL is standard for all student projects**
  - Our research projects have been based on open source RTL anyway
  - **Legal aspects:** In CH we need the *permission* of students to make their code openly available
- **Open source EDA and PDK allows exercises to be transferred anywhere**
  - We can share our exercises with others
  - No need to rely on costly infrastructure to provide students a place to run exercises

Thanks to the great IIC-OSIC tools from JKU

and the open-source PDK from IHP

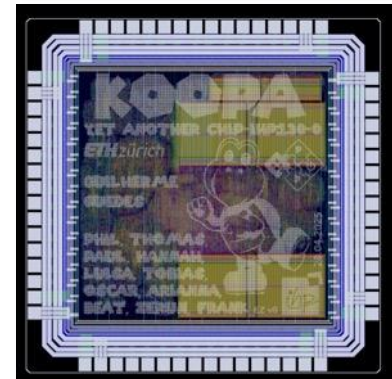
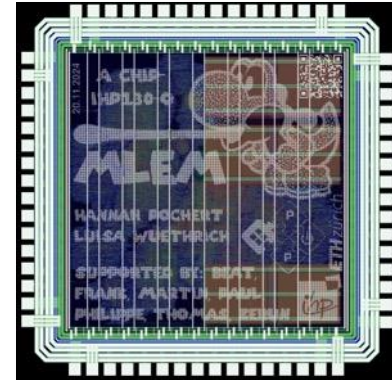
# At ETHZ, IC Design teaching now uses open source HW



- **In Spring 2025, our IC Design course switched to (mostly) open source**
  - Using IHP 130, Yosys and OpenROAD
    - Parts for backannotated simulation, test pattern generation, DRC/LVS, still use proprietary tools
    - Will be gradually replaced by open tools

<https://vlsi.ethz.ch>

- **Project based grading**
  - Students (in groups of two) will have to modify the Croc reference design
  - Best five designs will be taped-out
- **72 students enrolled**
  - Projects finished in summer
  - Taped-out 5 designs in IHP130

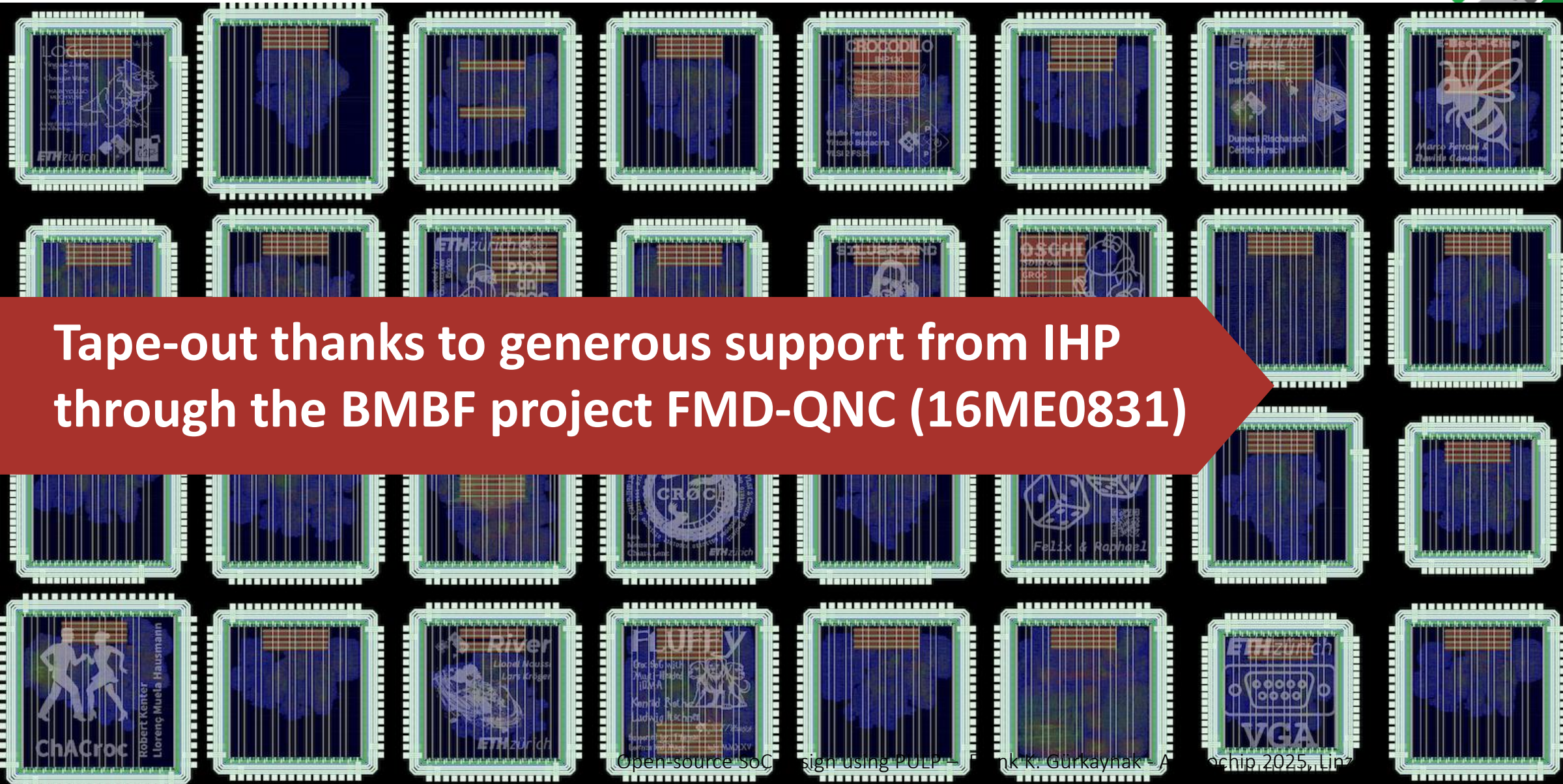




# And the students delivered: 33 valid designs, 5 taped-out



Tape-out thanks to generous support from IHP  
through the BMBF project FMD-QNC (16ME0831)







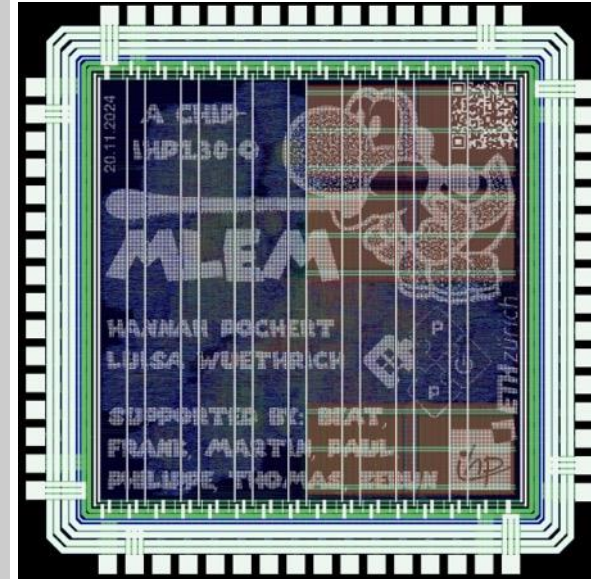
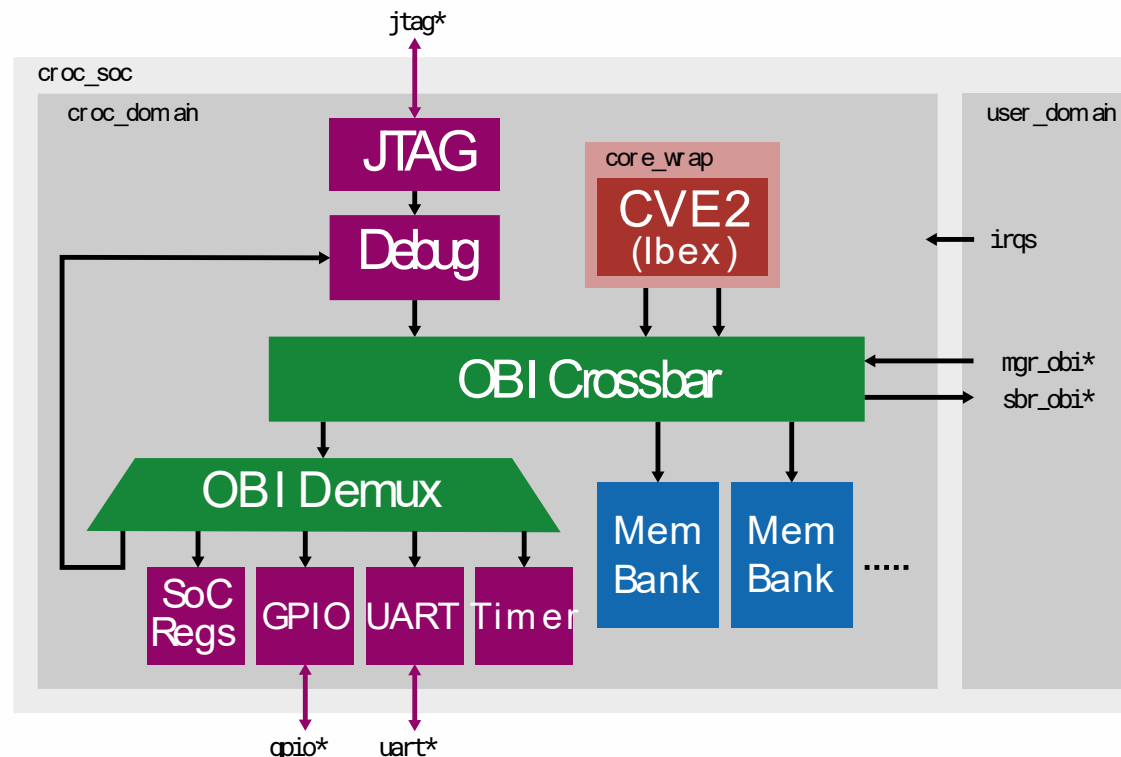
## A sample of our projects

- Croc
- Cheshire
- Kraken
- Occamy
- Picobello

# Croc our end-to-end open SoC for teaching and training



- **Scalable ULP design**
- **32-bit RISC-V Core**
  - Complete SoC
  - Simple “Raspberry Pi”
- **Rich Peripherals**
- **Ready for Acceleration**
  - Digital-only interface
- **Silicon-proven**
  - Tapeouts with open & commercial EDA



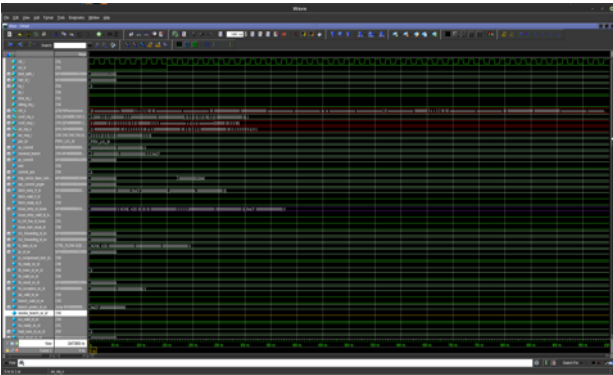
[github.com/pulp-platform/croc](https://github.com/pulp-platform/croc) 

# CVA6 (aka Ariane): The standard 64bit core

- Open-source 64-bit application-class RISC-V processor
- Boots Linux
- Developed by PULP team at ETH Zürich (as “Ariane”)
- Now owned and maintained by OpenHW Group
- Widely used in academia and industry



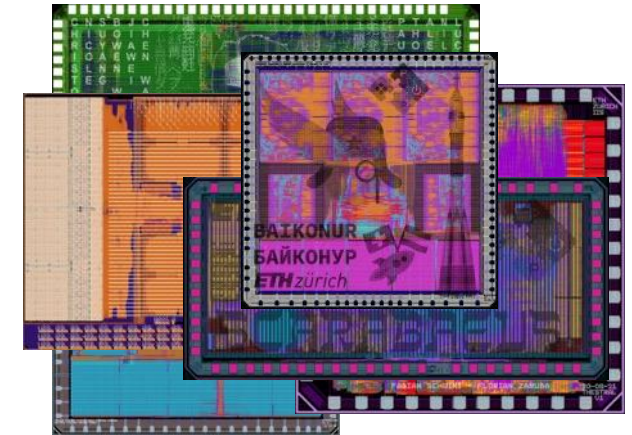
RTL Simulation



FPGA Emulation



ASIC



<https://github.com/openhwgroup/cva6>





# CVA6: Many optional features added over time

- XLEN 32/64 bits (RV32/RV64)
- WB/WT/HPDcache L1 data cache
- Embedded/application class (MMU, U/S mode)
- Vector extension (V)
- Hypervisor extension (H)
- Code-size extensions (Zcb, Zcmp)
- Bit-manip extensions (Zba, Zbb, Zbc, Zbkb Zbkx, Zbs)
- Scalar crypto extensions (Zknd, Zkne, Zknh)
- ECC support
  - ACE support
- Fast interrupts (CLIC)
  - ...



**ETH** zürich



**THALES**



Universidade do Minho



# Cheshire SoC: A SoC around CVA6

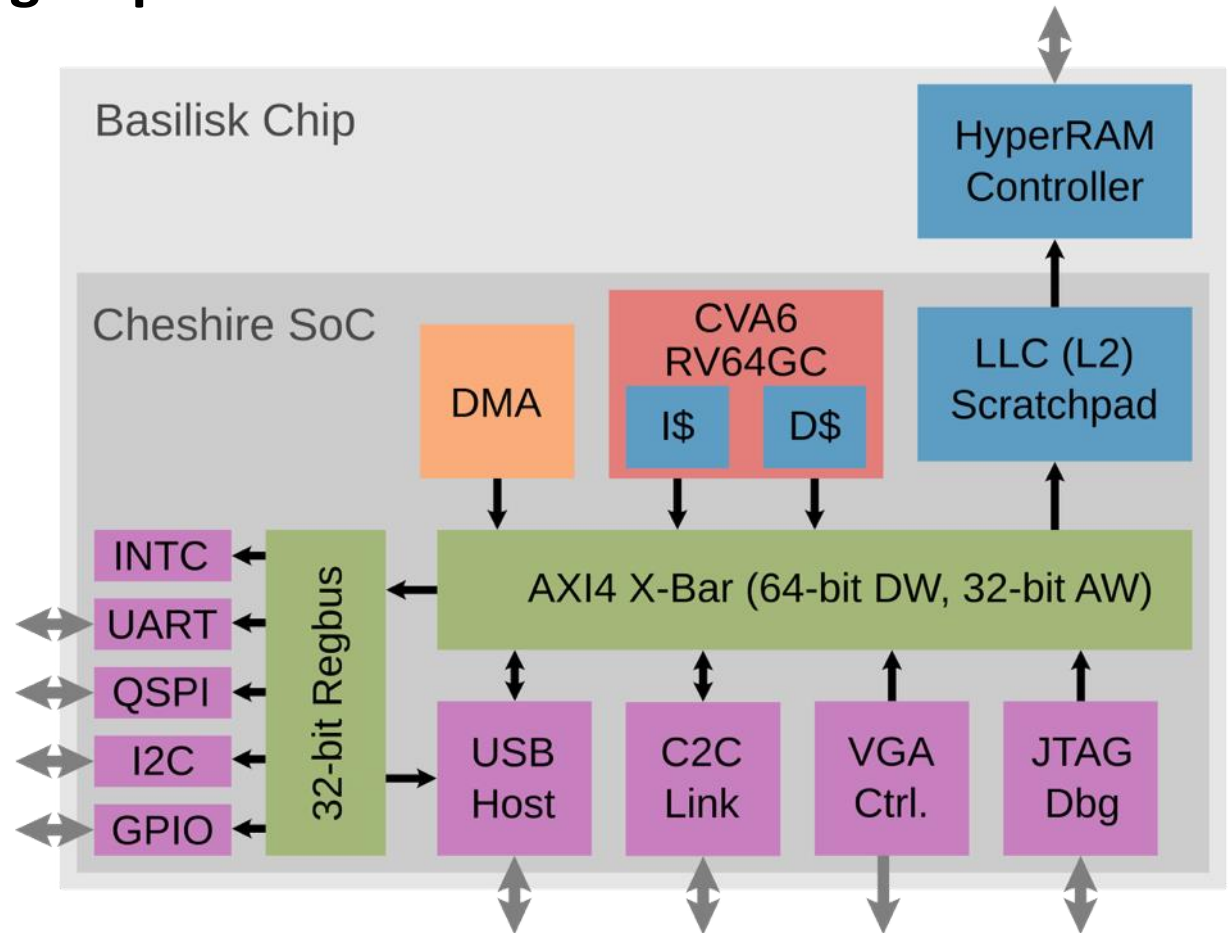


- **Permissive licensed RTL from various groups**

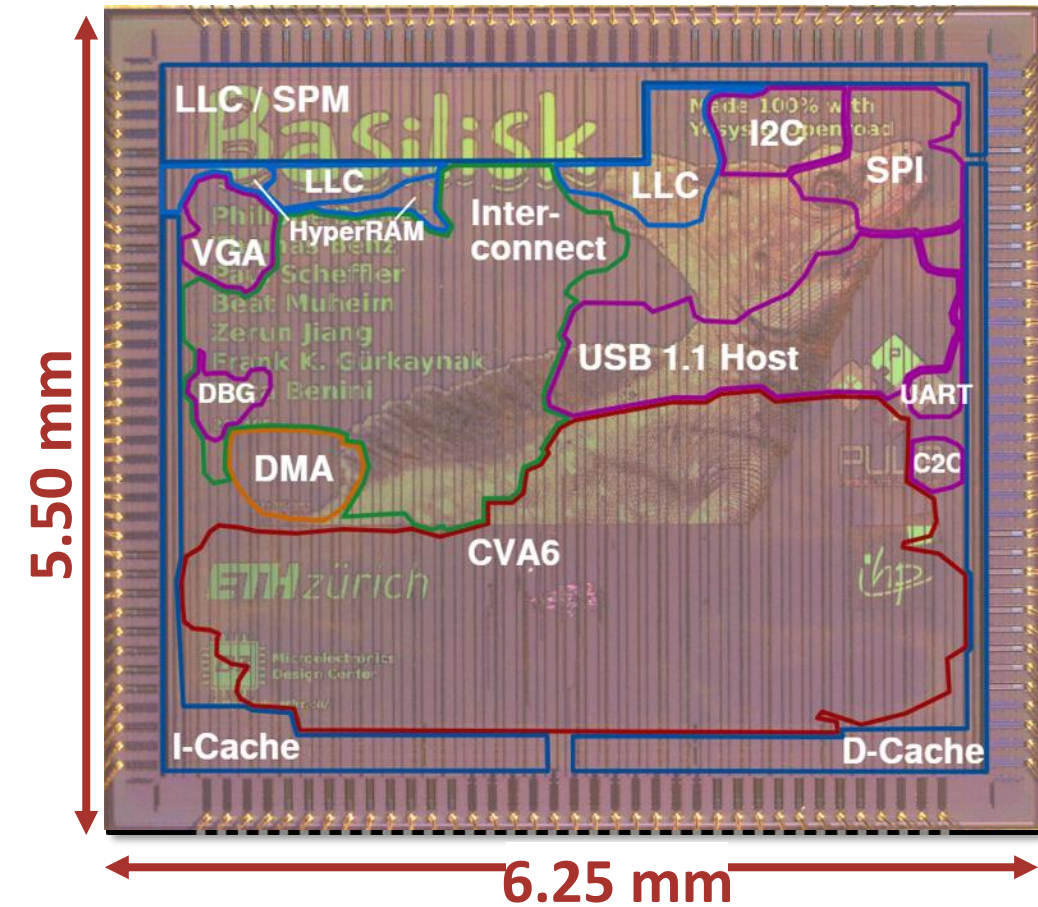
- PULP Platform (Cheshire, AXI, Hyperbus...)
- OpenTitan (SPI, I2C)
- OpenHW Group (CVA6 core)
- SpinalHDL (USB)

- **Architecture**

- **RV64GC**-compliant CVA6 core
- **Cheshire SoC** provides vital peripherals
- 4-way 16KiB L1-D and L1-I cache
- 4-way 64KiB LLC / Scratchpad memory
- **Hyperbus DRAM controller** achieving transfer speeds up to 124MB/s



# Meet Basilisk: Open RTL, Open EDA, Open PDK



- **Designed in IHP 130nm OpenPDK**
  - **34mm<sup>2</sup>** (6.25mm x 5.50mm)
  - **~5× larger** than previous end-to-end OS designs
  - 2.7 MGE total, 1.14MGE logic
  - 24 SRAM macros (114 KiB)
  - 62MHz at nominal voltage (1.2V)
- **RV64GC design runs Linux**
- **Active collaboration with**

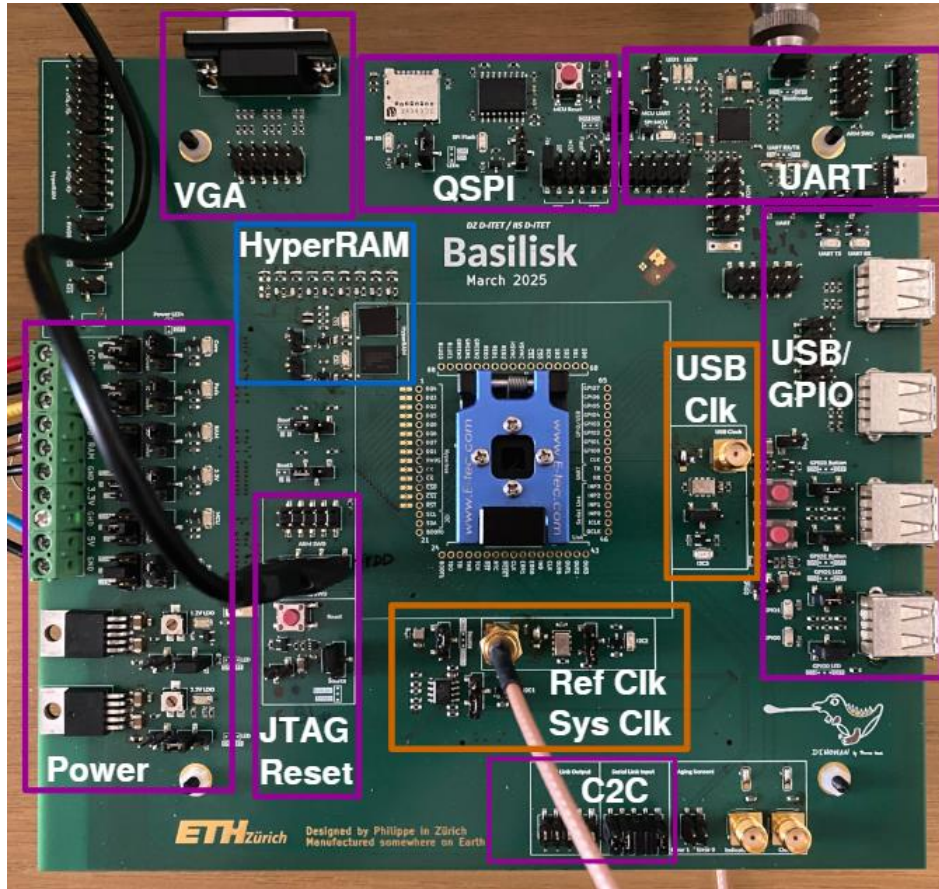


[github.com/pulp-platform/cheshire-ihp130-o](https://github.com/pulp-platform/cheshire-ihp130-o)





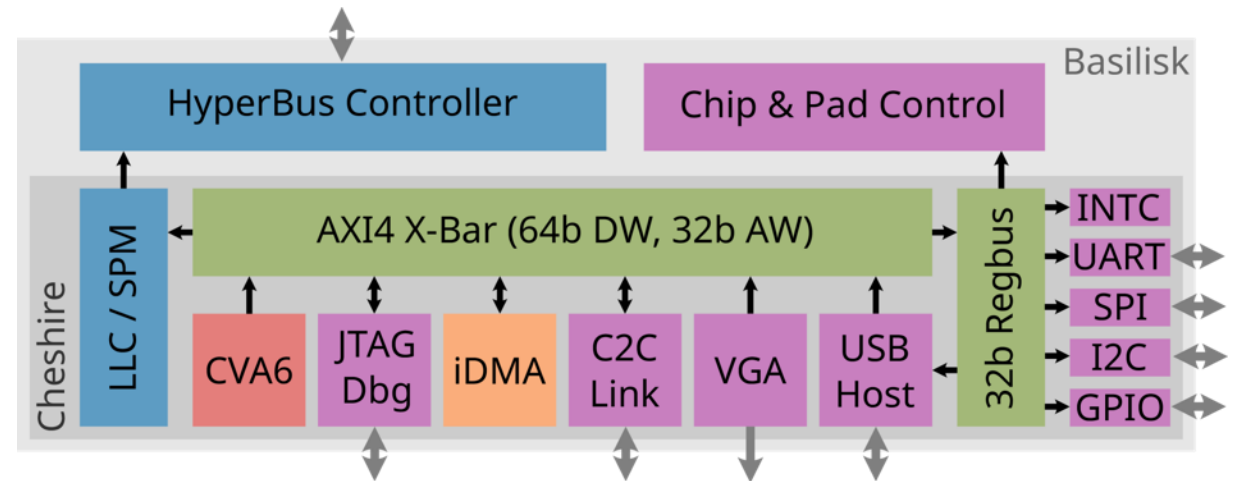
# Basilisk is a complete Linux-capable SoC



- 64-bit RISC-V core
- Rich peripherals:
  - HyperRAM controller @154MB/s
  - C2C AXI-Link @77MB/s
- Automatic boot via scratchpad



Watch the video



[arxiv.org/pdf/2505.10060](https://arxiv.org/pdf/2505.10060)

# Our research focus: cluster-based many-core accelerators



## Multiple Scales of acceleration

### Extensions to processor cores

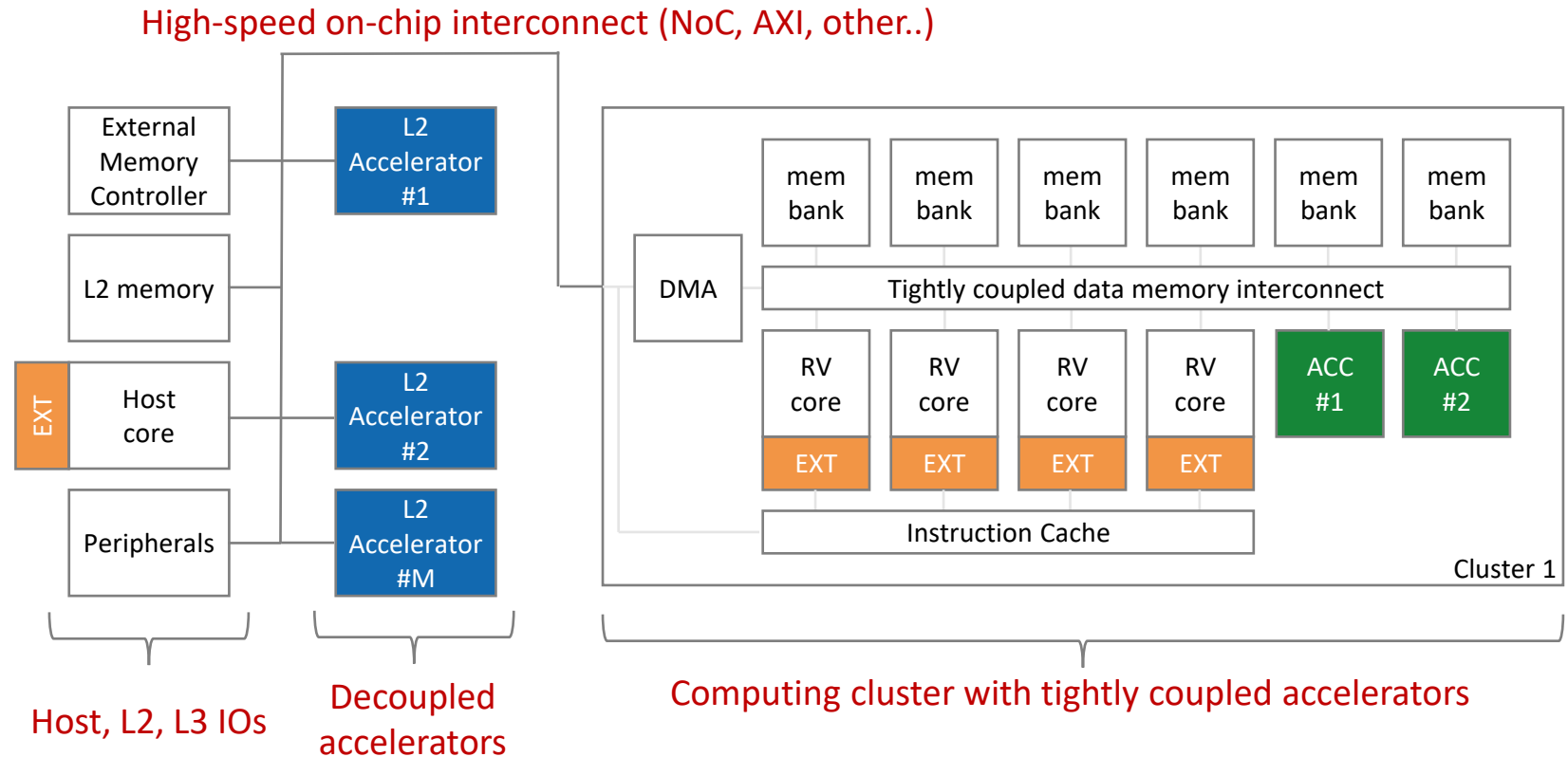
- Explore new extensions
- Efficient implementations

### Shared-memory Accelerators

- Domain specific
- Local memory

### Multiple Decoupled Accelerators

- Communication
- Synchronization



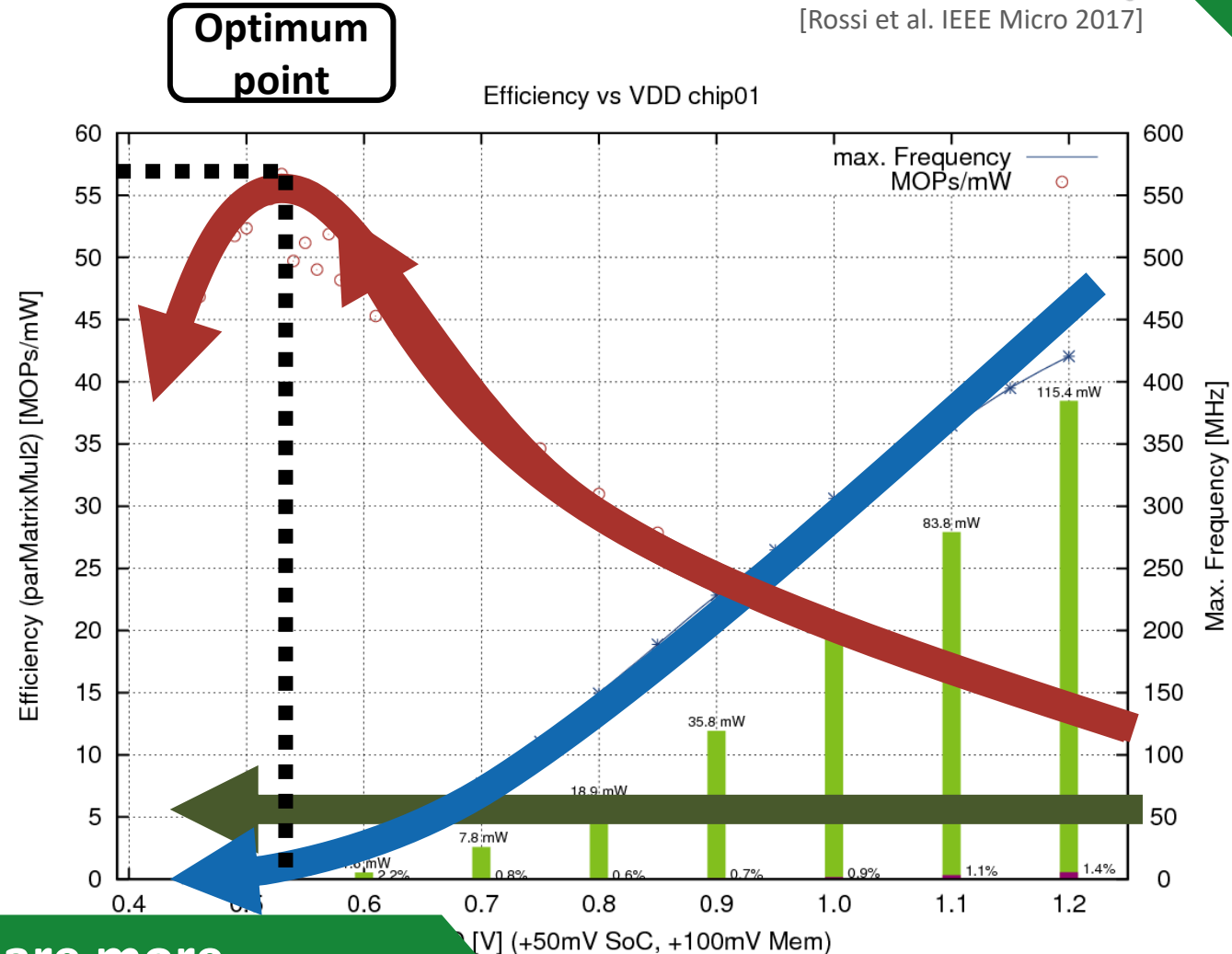
**RISC-V is a key enabler → max agility, enabling SW build-up, without vendor lock-in**

# 2013: Perf. + Efficiency + Flexibility ← Parallelism



- As **VDD** decreases, **operating speed** decreases as well.
- However **efficiency** increases → more work done per Joule
  - Until leakage effects start to dominate
- **More units in parallel**
  - Get performance up (if you can keep them busy)
  - Energy efficiency stays high!

[Rossi et al. IEEE Micro 2017]



N cores running at moderate f, low Vdd are more energy efficient than a single core at Nxf, high Vdd



# PULP Paradigm: Multiple Cores (2-16)



**efficient DSPs (CV32E40P)**  
simple in-order 4-stage  
pipeline with RISC-V ISA  
+ DSP extensions (xPULP)

**RISC-V  
core**

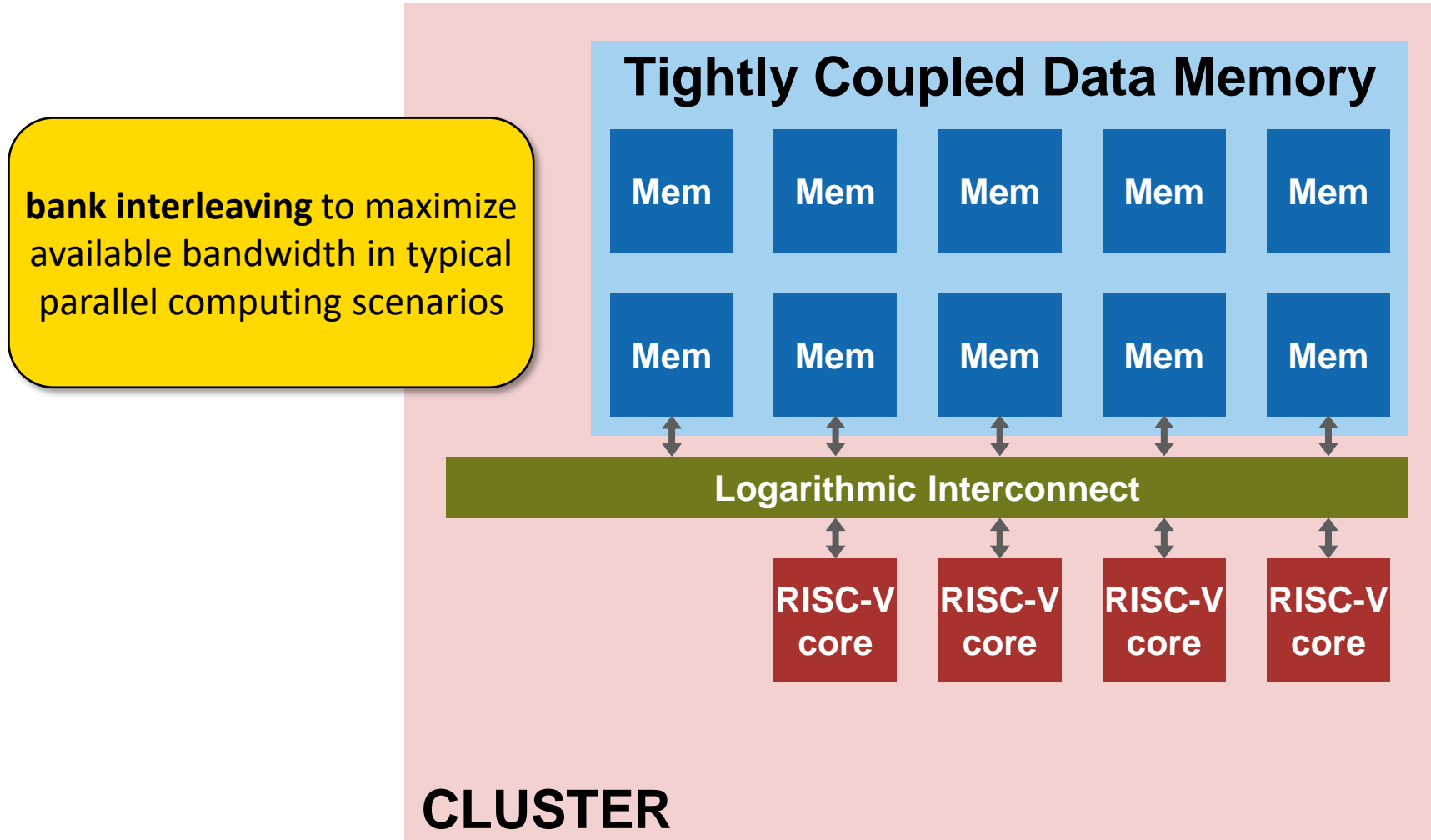
**RISC-V  
core**

**RISC-V  
core**

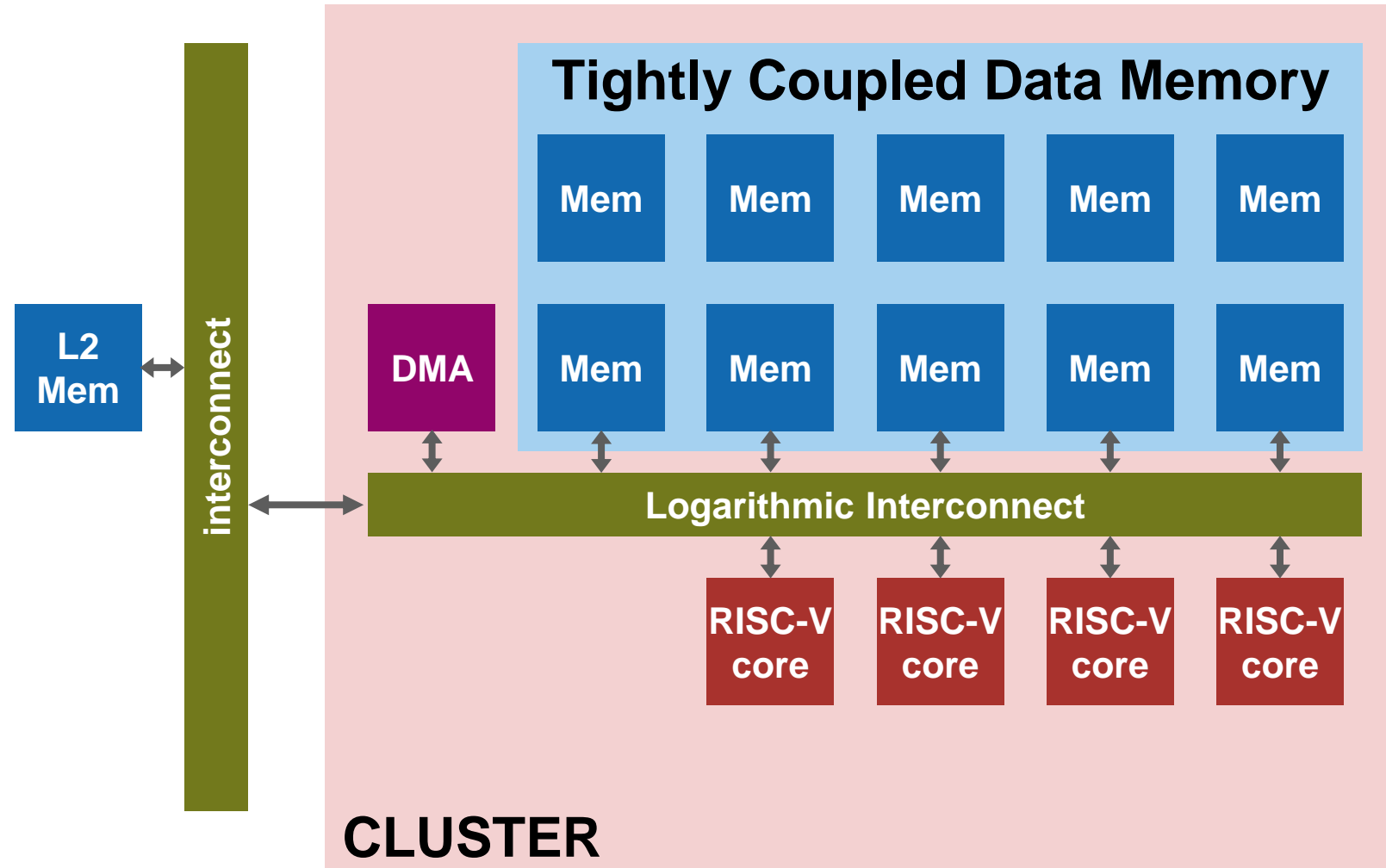
**RISC-V  
core**

**CLUSTER**

# PULP Paradigm: Low-Latency Shared TCDM

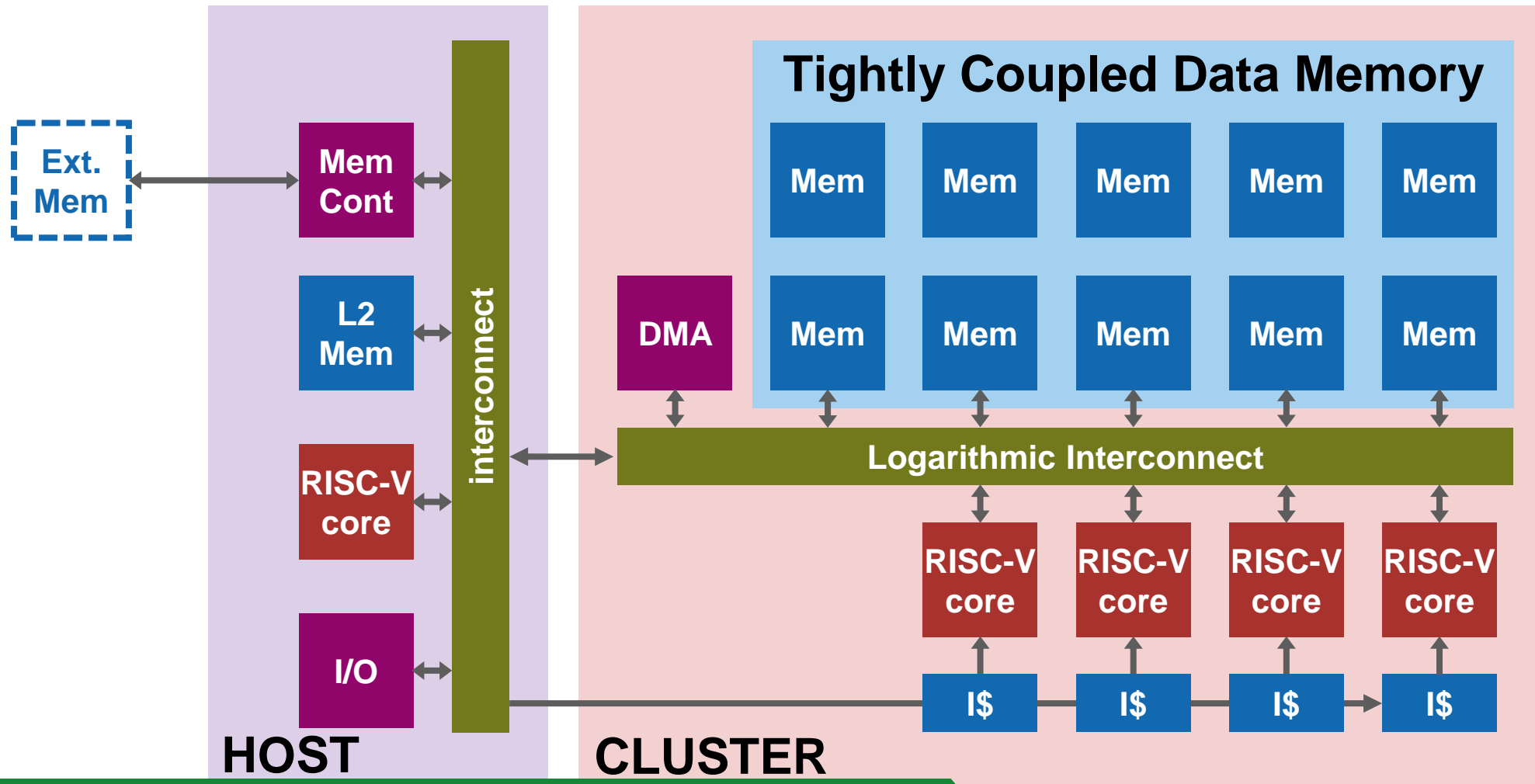


# PULP Paradigm: DMA and I\$ to talk with ext. memory



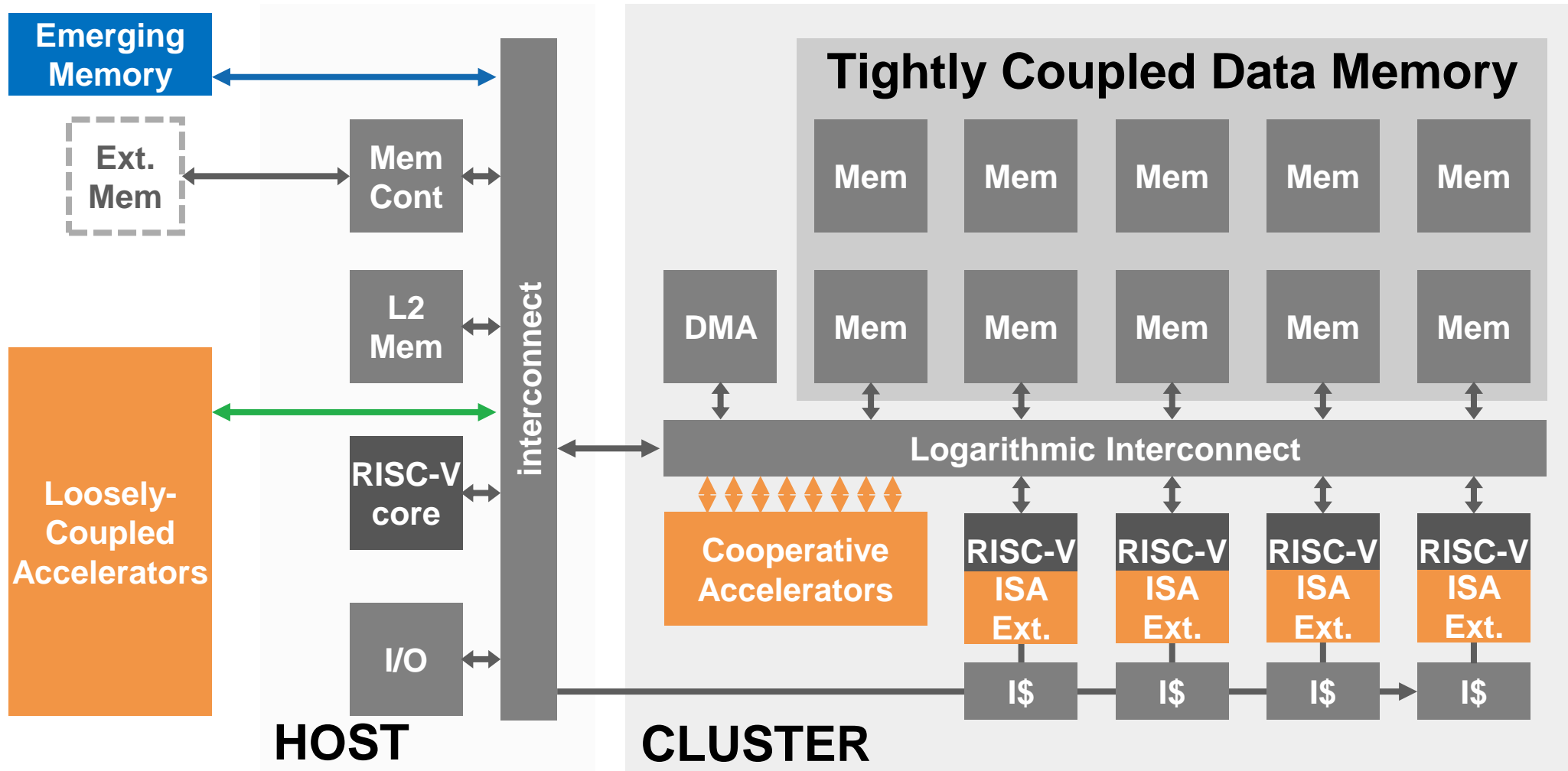


# PULP Paradigm: The cluster accelerates a host system



[github.com/pulp-platform/pulp](https://github.com/pulp-platform/pulp)

# PULP is a template for heterogeneous parallel SoCs

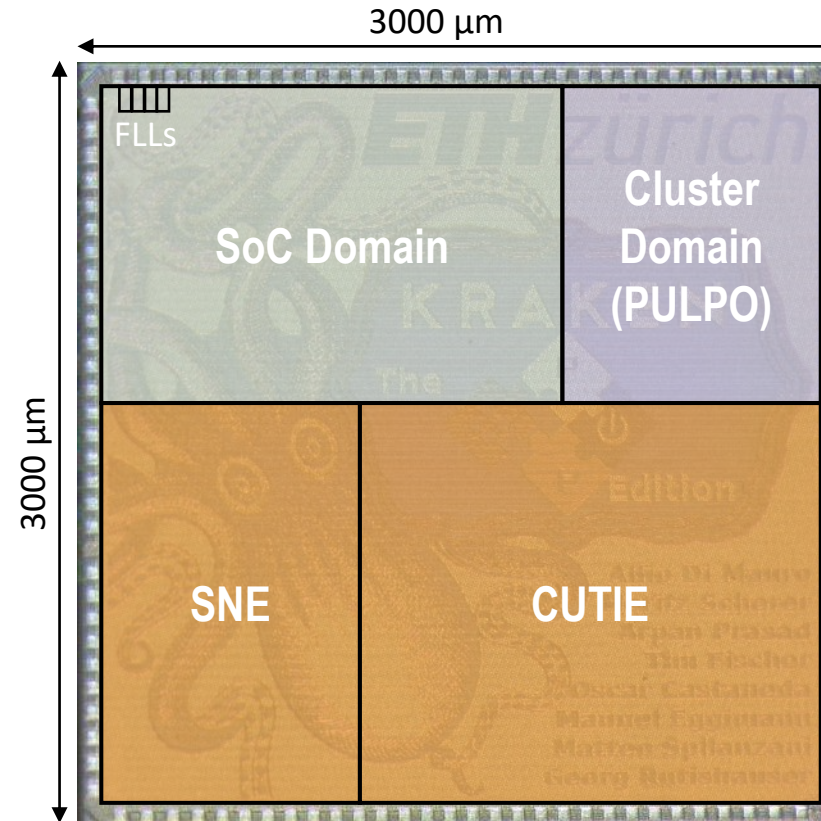


# Kraken: Multiple Heterogeneous Accelerators



The *Kraken*: an “Extreme Edge” Brain

- **RISC-V Cluster**  
(8 Cores + 1)
- **CUTIE**  
Dense ternary neural network accelerator
- **SNE**  
Energy-proportional spiking neural network accelerator



Technology	22 nm FDSOI
Chip Area	9 mm <sup>2</sup>
SRAM SoC	1 MB
SRAM Cluster	128 KB
VDD range	0.55 V - 0.8 V
Cluster Freq	~370MHz
SNE Freq	~250MHz
CUTIE Freq	~140MHz

[Di Mauro HotChips22]





# Specialization in perspective



Using 22FDX tech, NT@0.6V, High utilization, minimal IO & overhead

Energy-Efficient RV Core → **20pJ (8bit)**



ISA-based 10-20x → **1pJ (4bit)**



Configurable DP 10-20x → **100fJ (4bit)**



Highly specialized DP 100x → **1fJ (ternary)**



**XPULP**

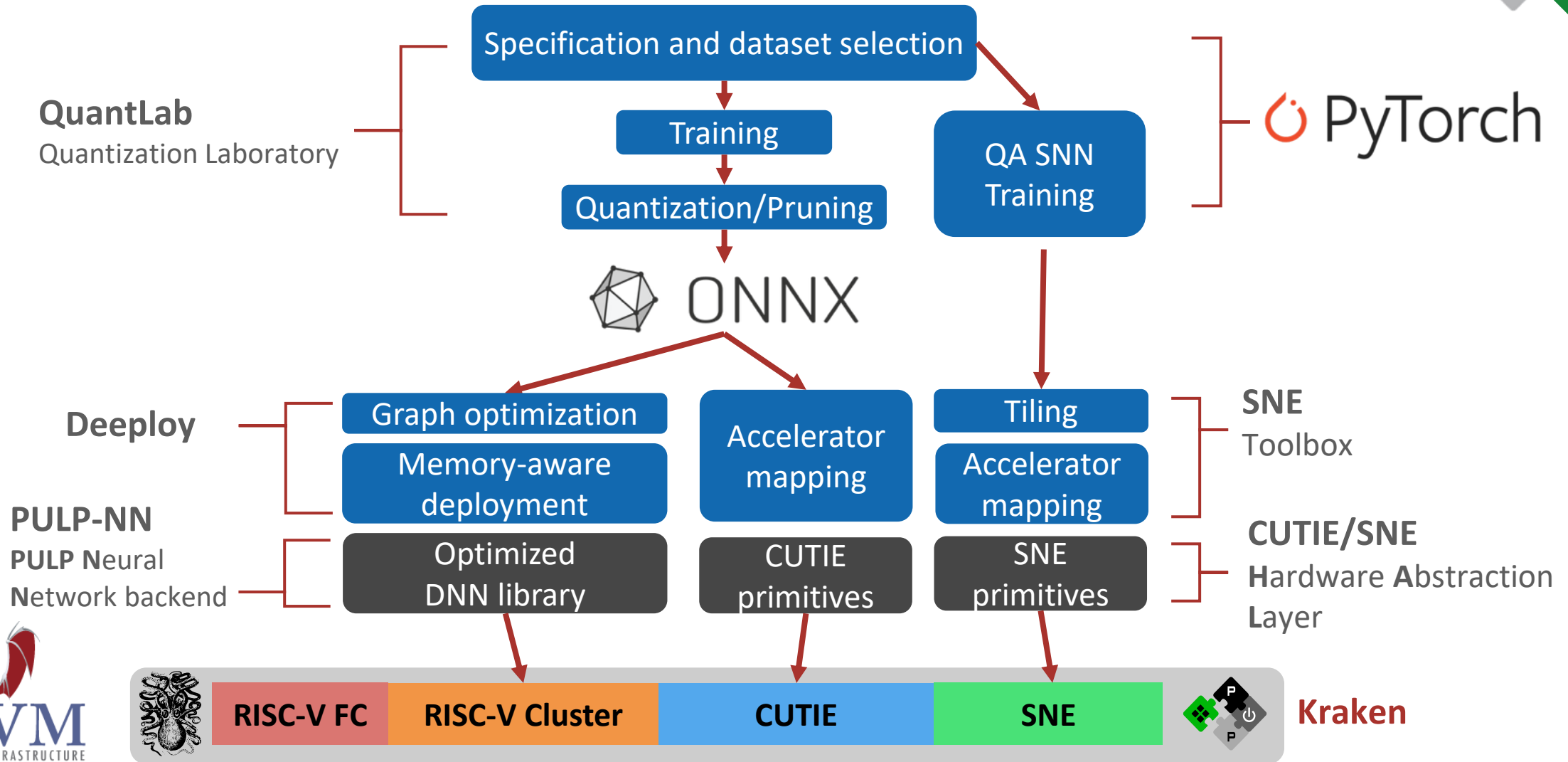


**RBE**

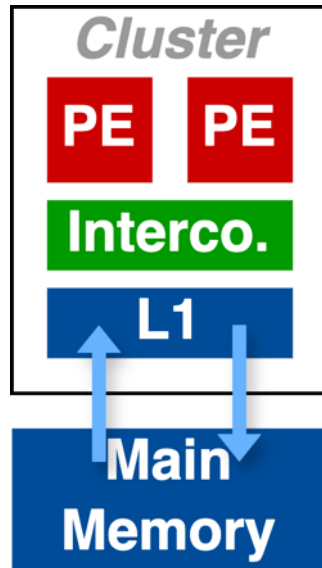


**CUTIE, SNN**

# Fully Open-Source Deployment Flow!



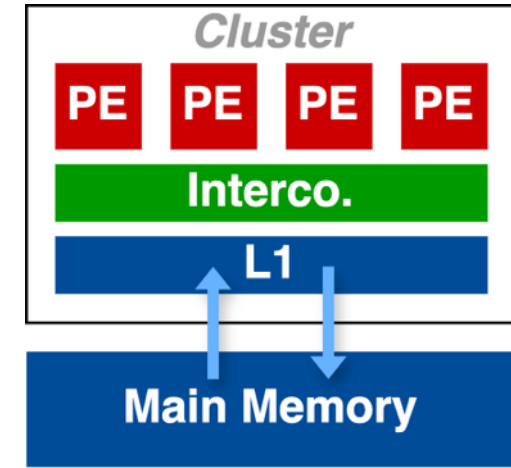
# Where do we go from here: *Scale-Up* vs. *Scale-Out*



**Cluster:** A group of **PEs** tightly coupled with a shared **L1 memory** through low-latency **interconnect**

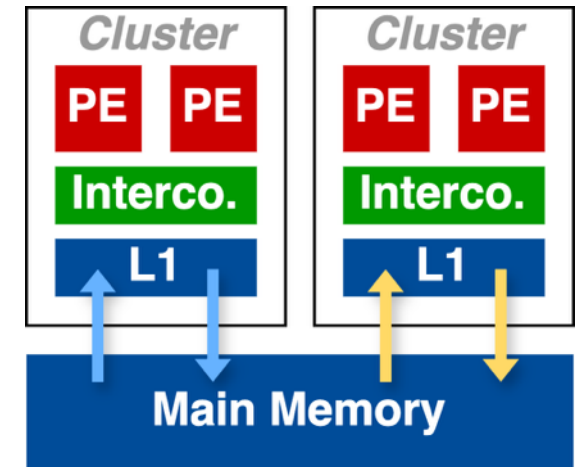
## *Scale-up*

Increase the number of PEs in a cluster



## *Scale-out*

Increase the number of parallel Clusters

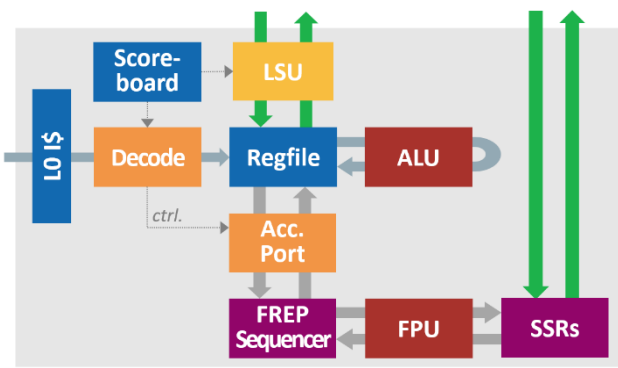




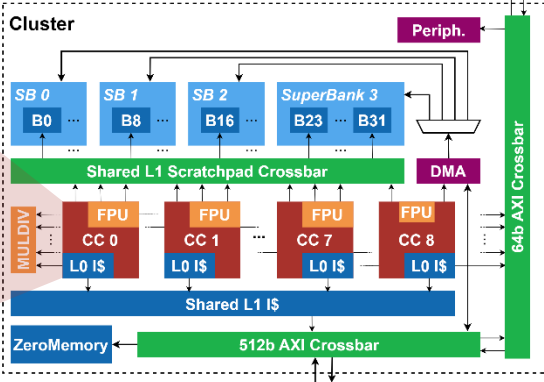
# Achieving Scale through Hierarchical Design



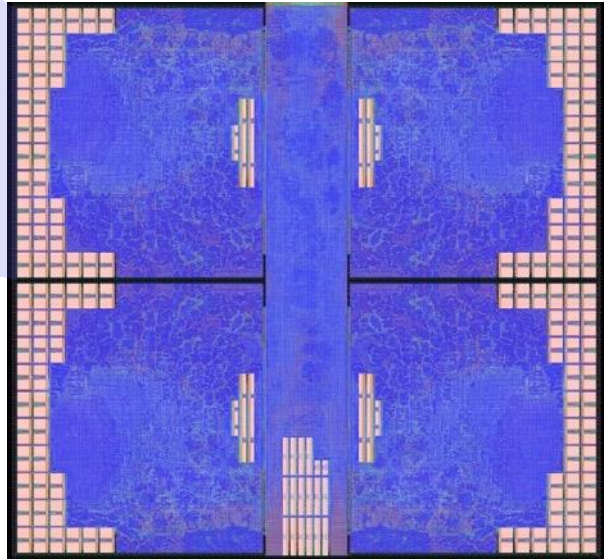
Snitch Core



Snitch Cluster



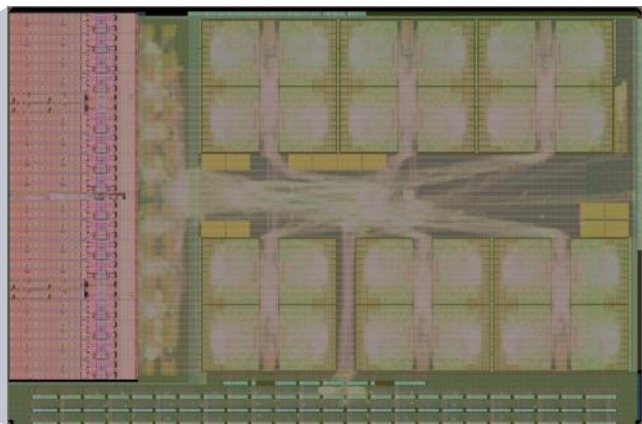
Occamy Group



Occamy System



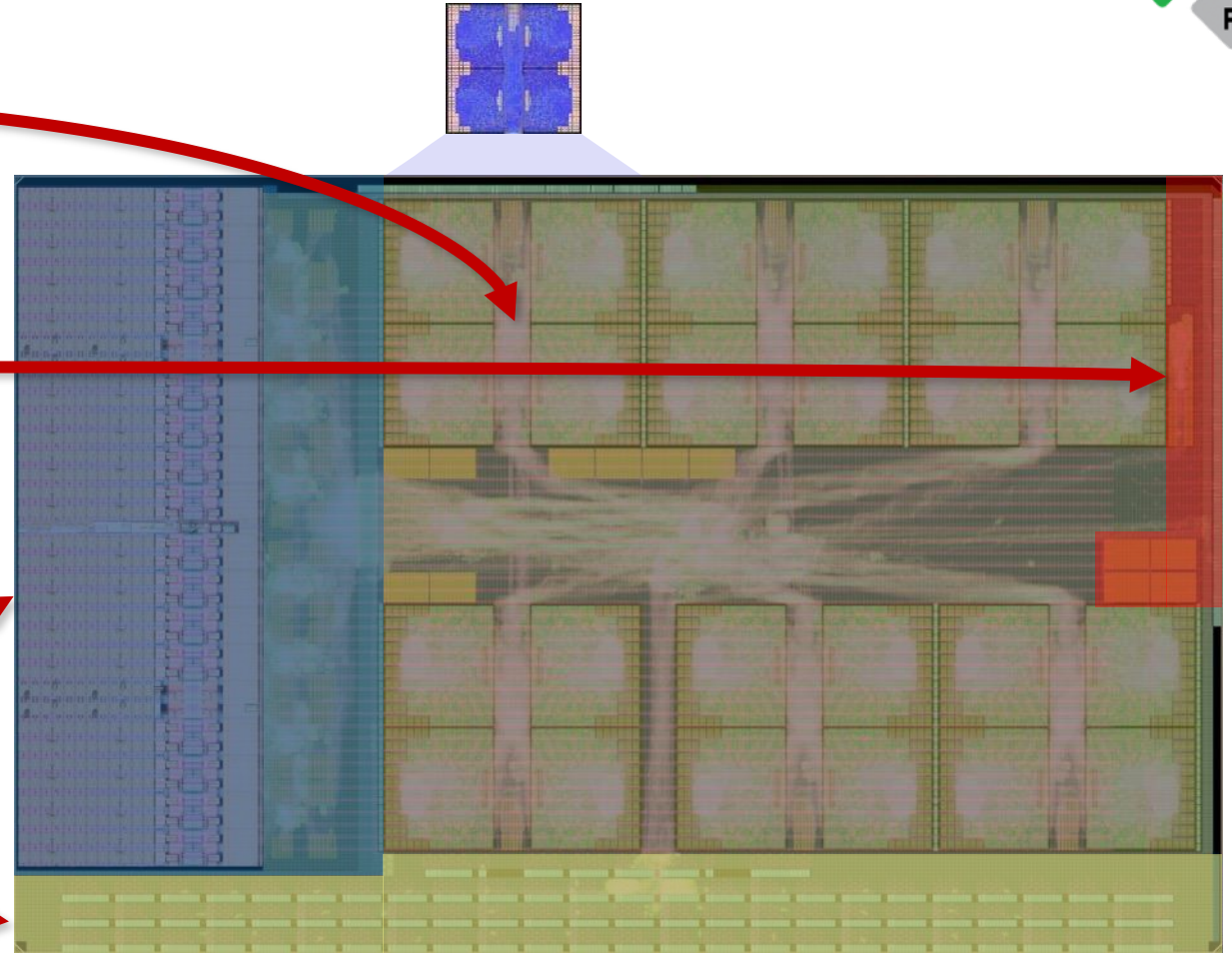
Occamy Chiplet



# Occamy Chiplet: Six Groups with HBM and D2D Link

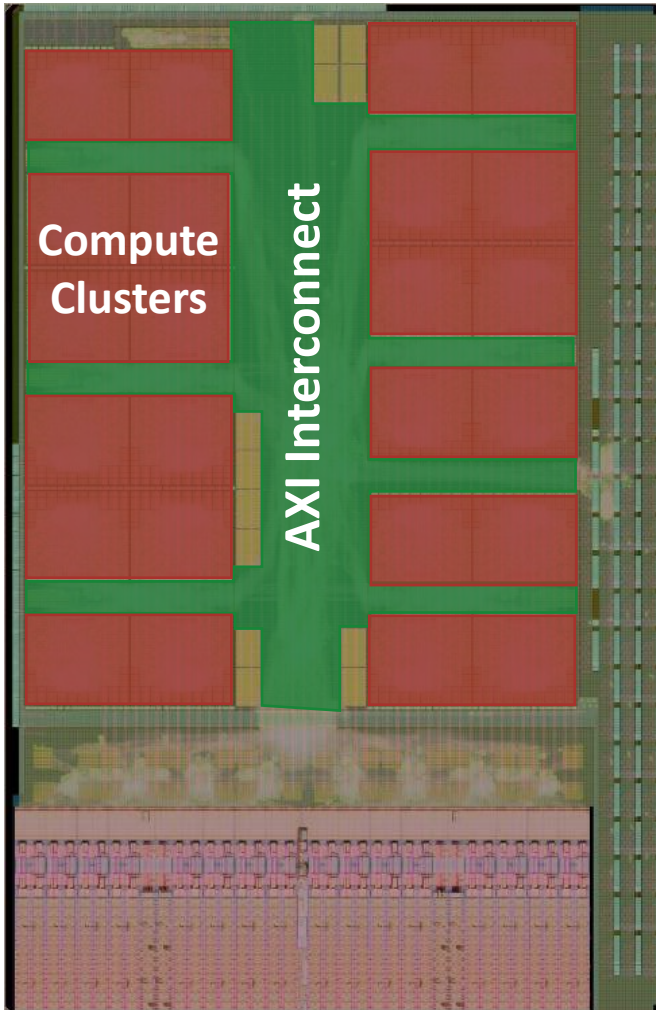


- **6 fully connected groups**
  - 24 clusters, 216 cores total
  - 512b data, 64b for messages interconnect
- **Autonomous 64b host domain**
  - CVA6 RV64GC Linux-capable core
  - Rich peripherals (SPI, I2C, UART...)
- **16 GiB, 410 GB/s HBM2E**
  - Optional page-level interleaving
- **12.8 GiB/s die-to-die link**
  - Fully digital and fault-tolerant



[github.com/pulp-platform/snitch\\_cluster](https://github.com/pulp-platform/snitch_cluster) 

# Addressing interconnect scalability



- **AXI interconnect was very challenging for PD**
  - AXI has severe scalability issues
  - Top-level Xbar had to be split up
  - Still, interconnect takes up almost **40%\***
- **Working on NoC solution, *FlooNoC***
  - Fully AXI4 compatible
  - Solves AXI4 **scalability issues**
  - Designed with awareness of physical design
  - **Wide & physical** channels



[github.com/pulp-platform/FlooNoC](https://github.com/pulp-platform/FlooNoC)

*\*HBM & C2C excluded*

Open-source SoC design using PULP – Frank K. Gürkaynak - Austrochip 2025, Linz

38

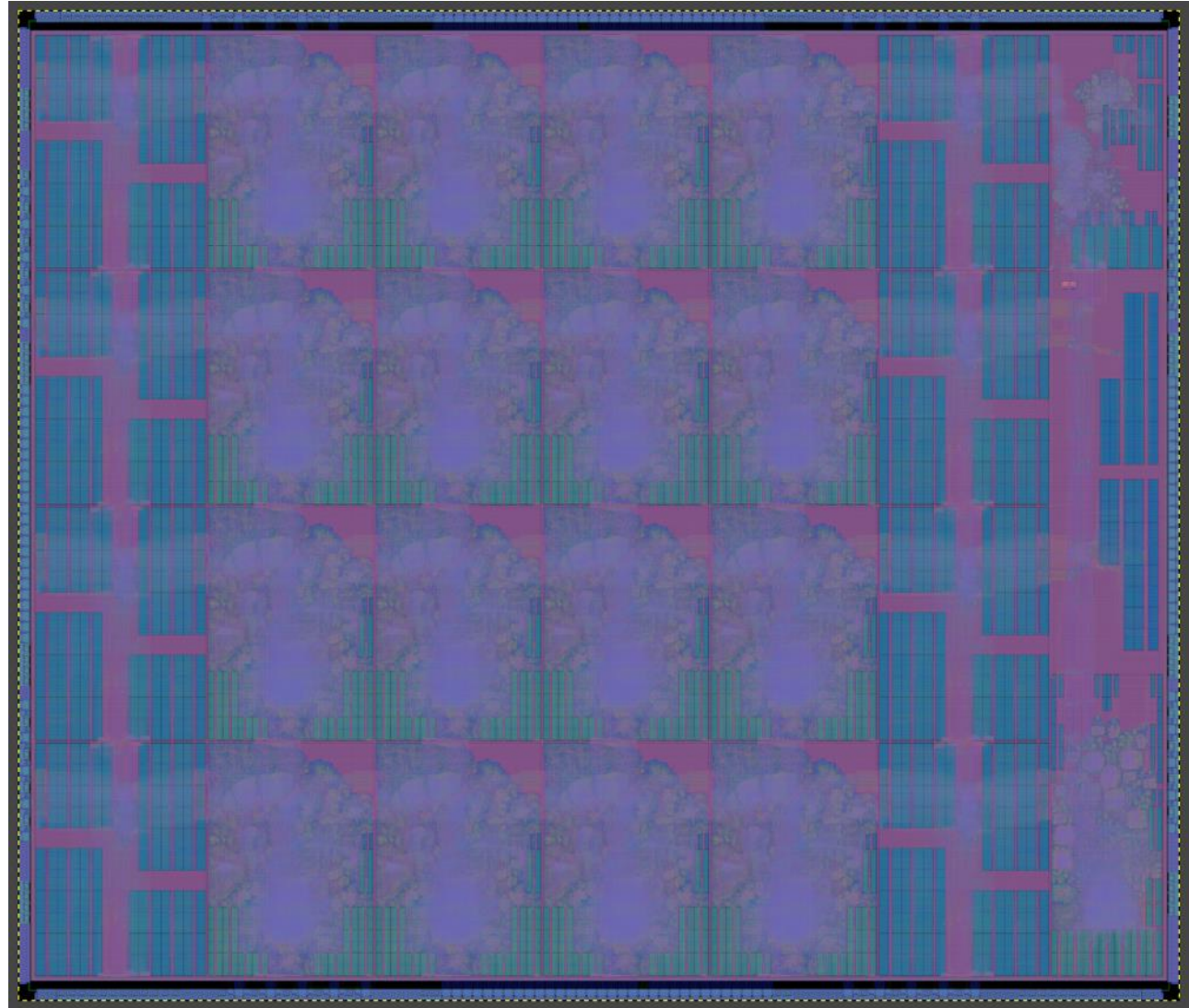


# Picobello Next generation many-core architecture in 7nm



- **64bit host**
  - Linux capable CVA6
- **16 clusters**
  - 9x Snitch (RV32) cores with Sparse SSRs
  - MatrixMul
- **8 MB L2 memory**
- **FlooNoc interconnect**
- **Taped out in Aug 2025**

THE **EUPILOT**





# What is next for open-source hardware

- Challenges
- Opportunities

# End-to-end Open-Source IC Design is possible today!



**Design:** from PULP

[github.com/pulp-platform](https://github.com/pulp-platform)



**Tools:** from Johannes Kepler University (JKU)

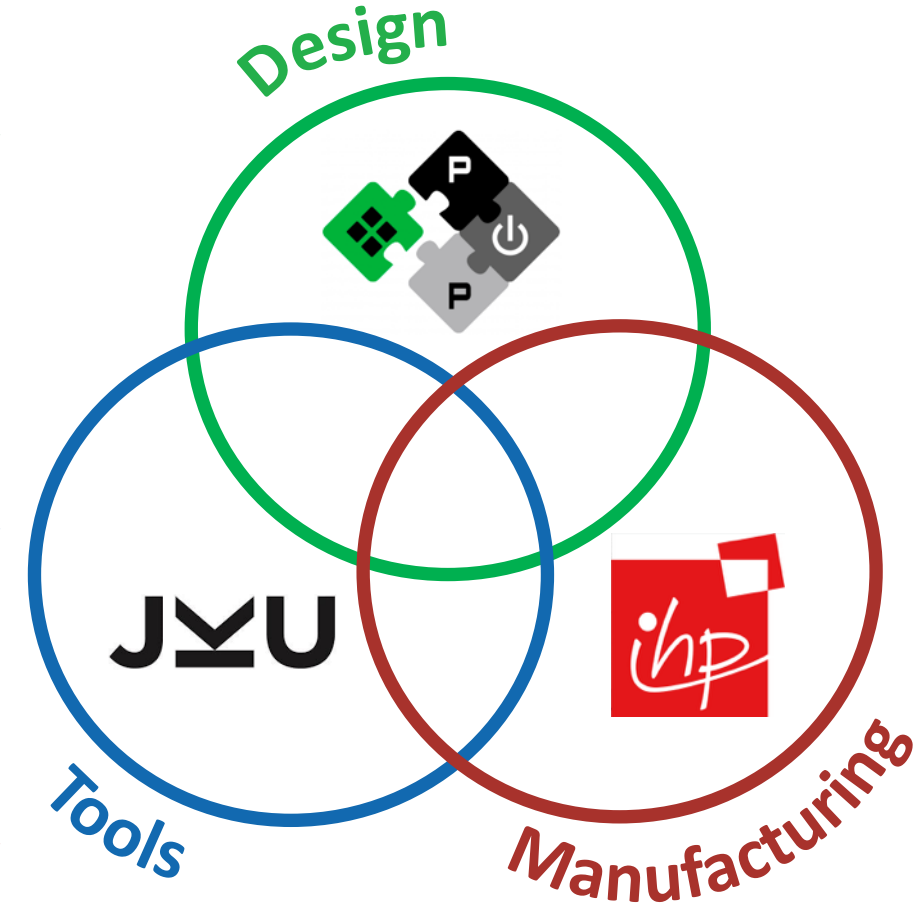
Reliable VM with large collection of open-source tools

[github.com/iic-jku/IIC-OSIC-TOOLS](https://github.com/iic-jku/IIC-OSIC-TOOLS)

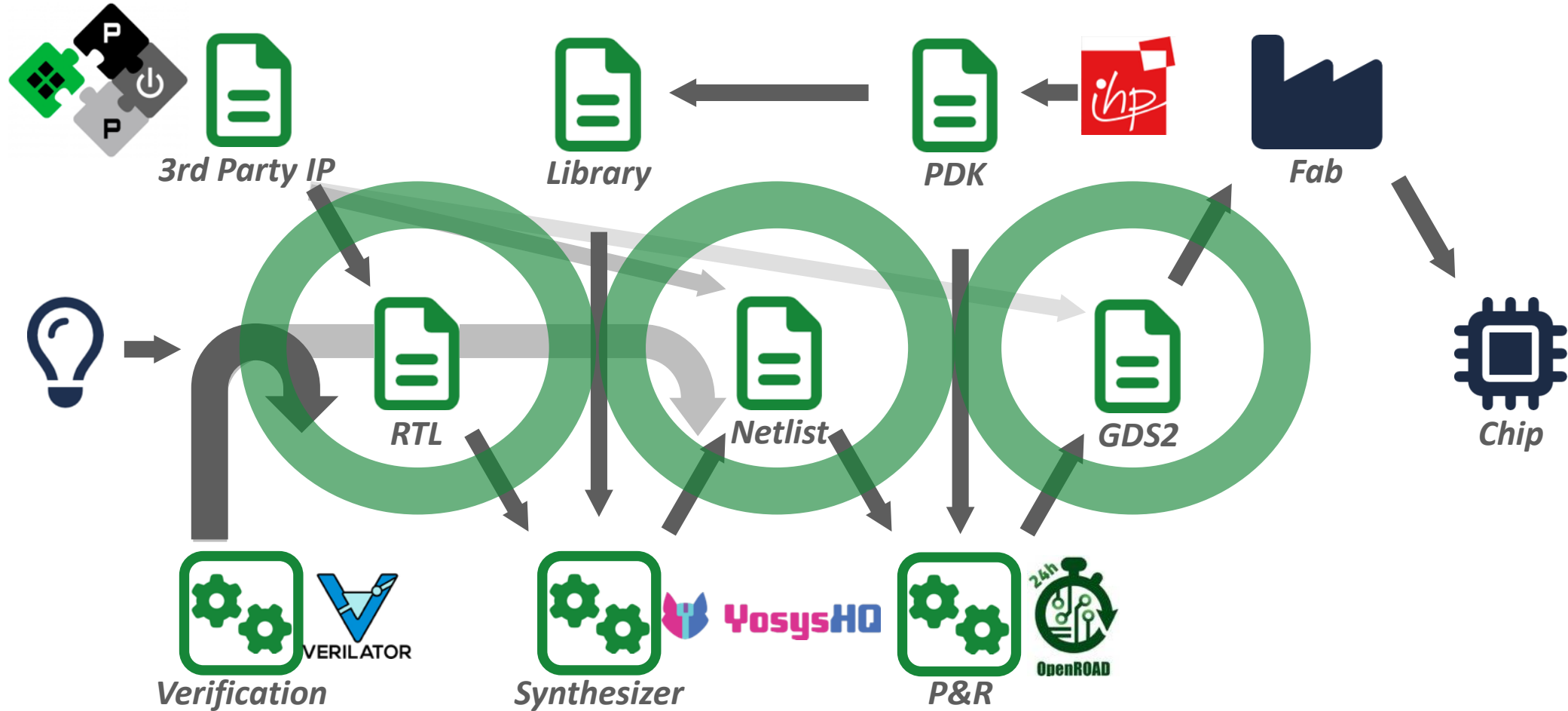


**Manufacturing:** IHP130nm

[github.com/IHP-GmbH/IHP-Open-PDK](https://github.com/IHP-GmbH/IHP-Open-PDK)



# End-to-end Open-Source allows sharing of design data

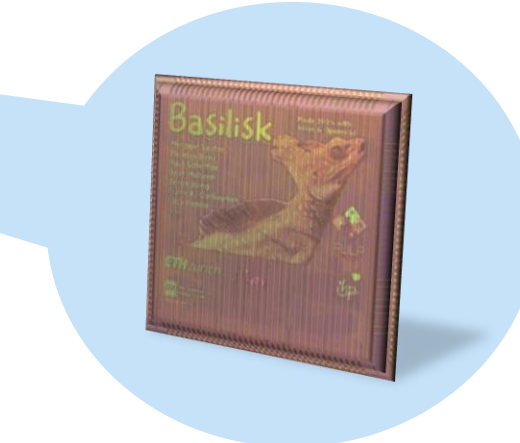
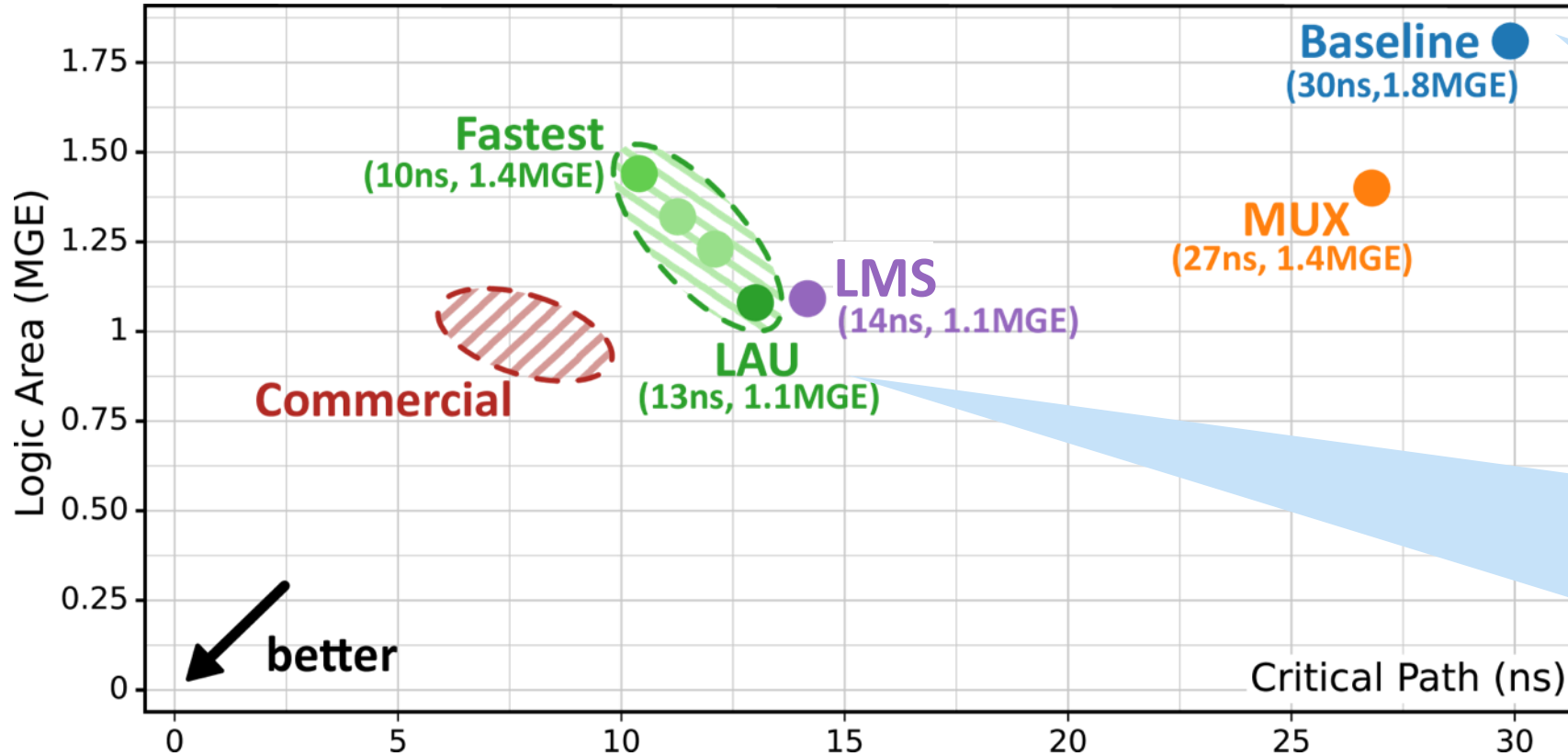




# Closing the PPA gap to commercial EDA



6

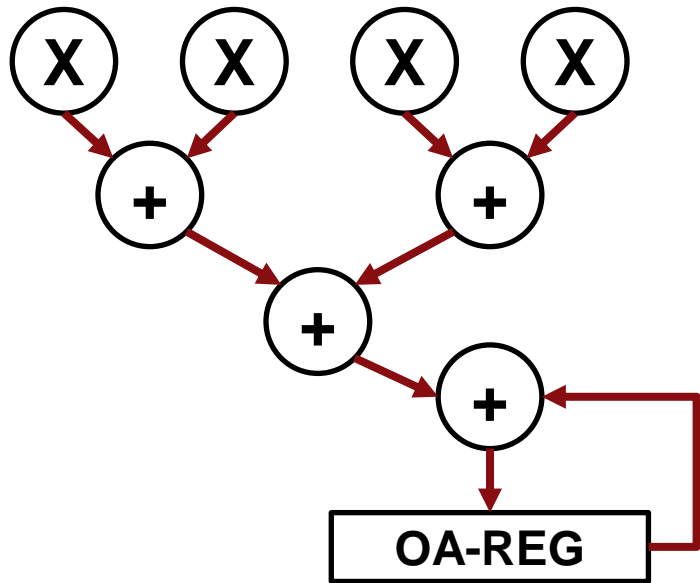


Yosys-slang full Sysverilog Frontend: @ <6sec runtime (from minutes)

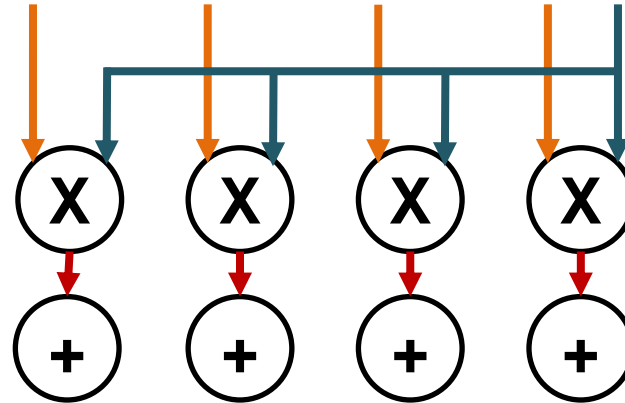
Yosys synthesis: 1.1 MGE (1.6×) @ 77 MHz (2.3×), 2.5× less runtime, 2.9× less RAM

OpenROAD P&R: tuning -12% die area, +10% core utilization

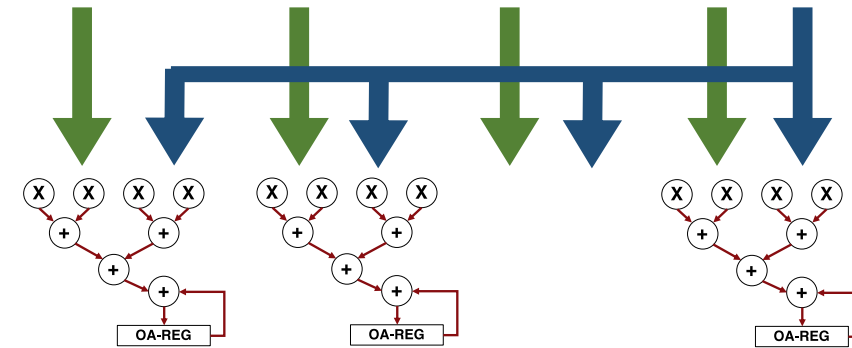
# Yes but why? Specialization + EDA multiplicative effect!



Inner Product



Outer Product



Mixed

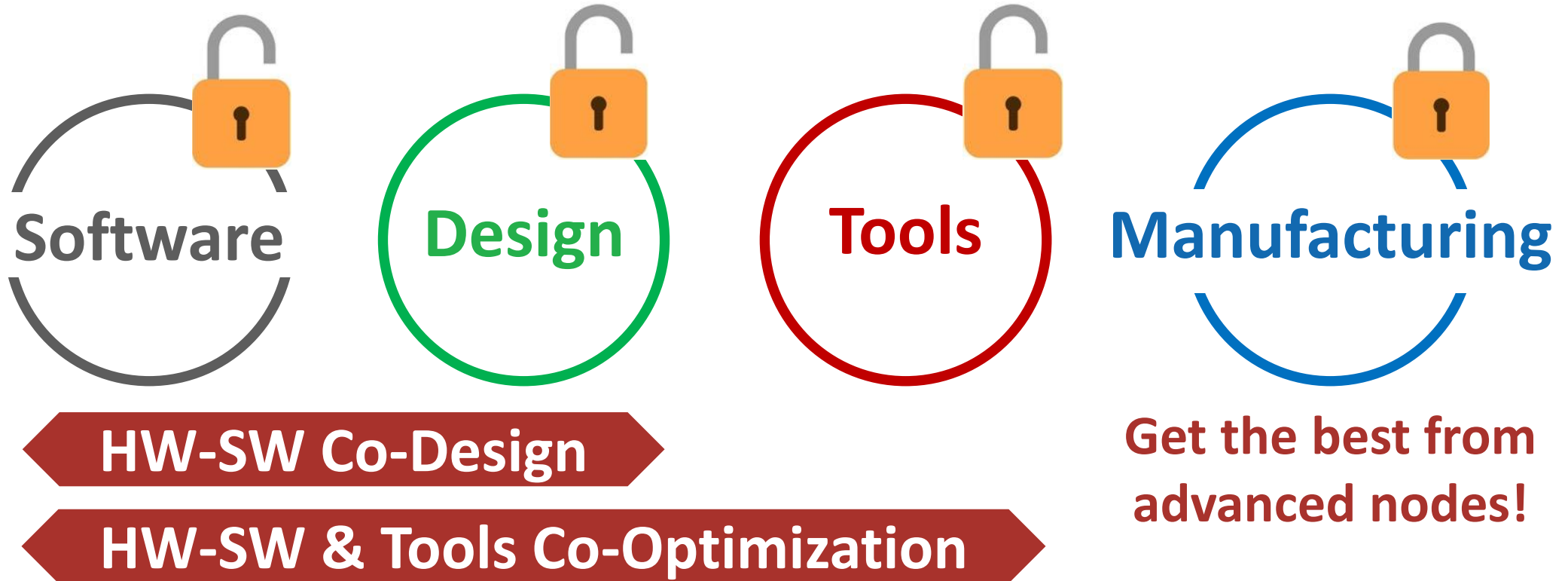
Precision tuning – OP/Mem tuning - deep arithmetic optimization – operand network tuning...

**Co-Specialize SW, HW, EDA & Technology is the frontier**

# Open EDAs Heterogeneous Chips in Advanced CMOS?



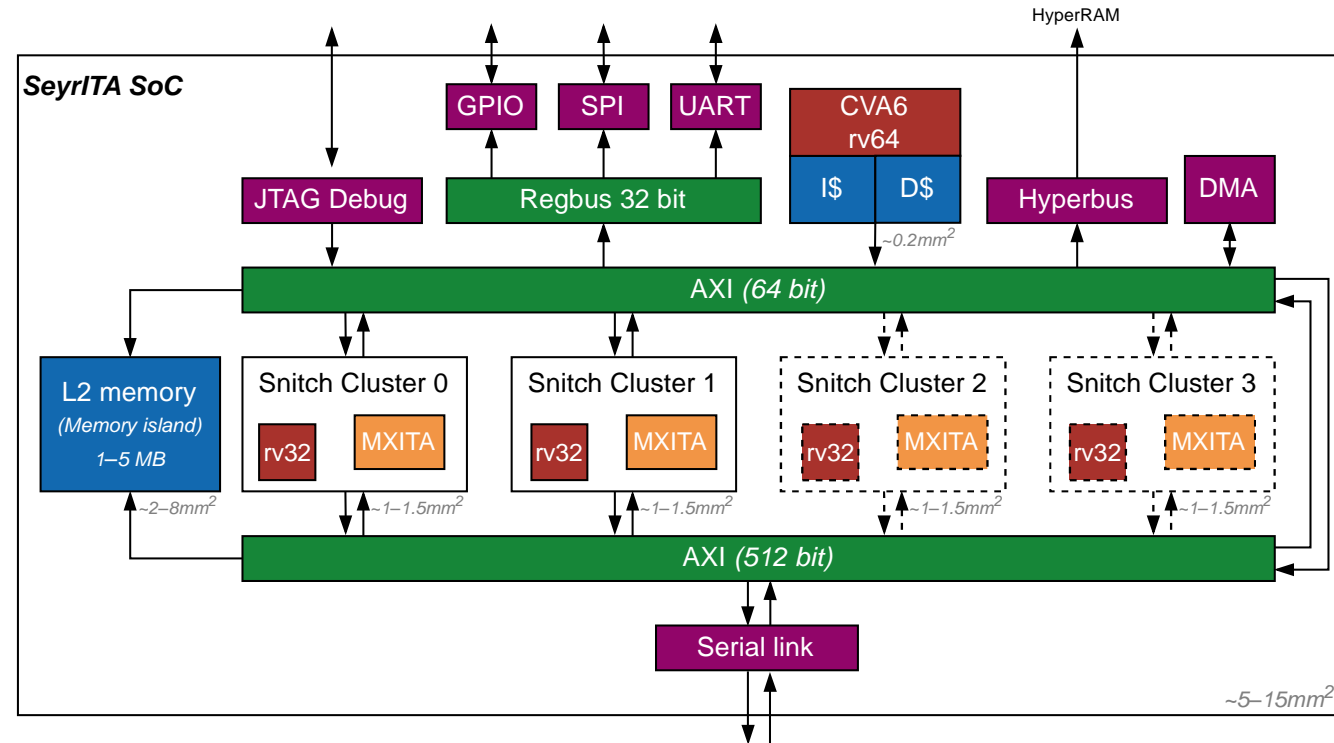
**Extreme Performance + Energy Efficiency is required!**



# On the Horizon: SeyrITA – GF22 with Open-Source Tools

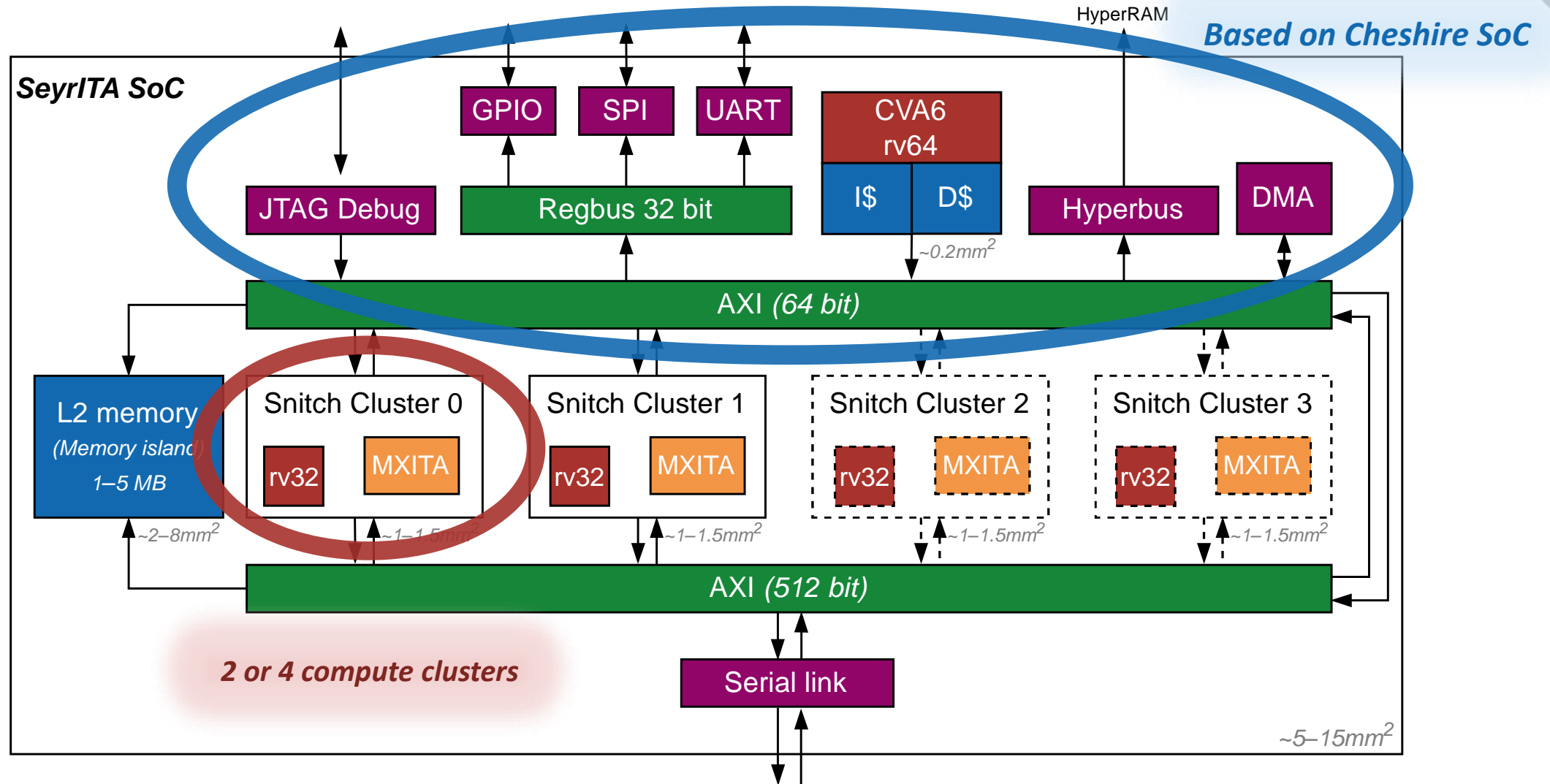


- RISC-V Linux host platform
- Transformer accelerator
  - Targeting BERT, mobileBERT, DEiT-T
  - Leveraging microscaling quantization
  - MXINT and MXFP32 formats
- **10x larger!**
  - 20-40MGE (SeyrITA) vs 2-3MGE (Basilisk)
- **500MHz target frequency**
- **1-2TFLOP/s**





# On the Horizon: SeyrITA – Top Level



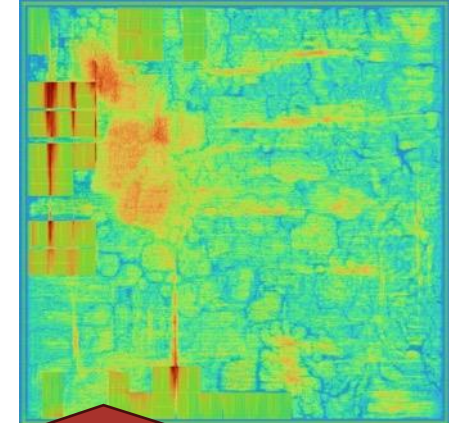
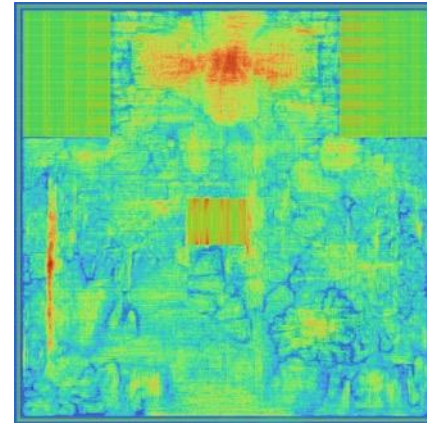
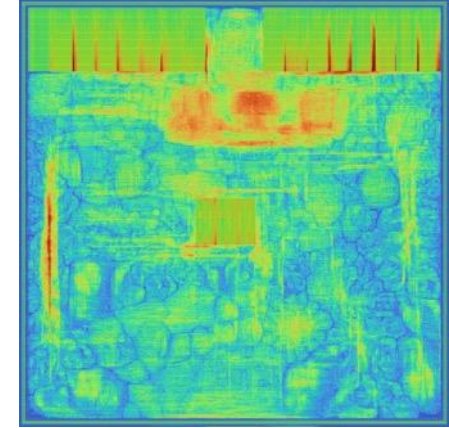
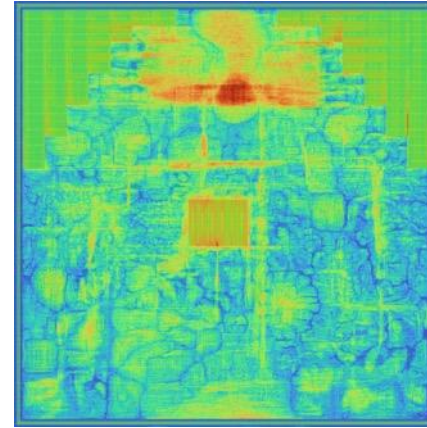
# On the Horizon: SeyrlTA – Working on the Tapeout



- **Demonstrate** a large **22nm** tapeout with open-source tools
- **Improve tools** and close the **performance gap**
- Identify and **implement missing features** along the way
- **Active Collaboration with**



**YosysHQ**



Snitch cluster floorplan exploration

# Open PDKs are key to success of open source HW



- **We need more open PDKs**

- Something in the 65nm – 28nm range would be a **game changer**
- Drafted an open letter to raise awareness with 300+ signatures



<https://open-source-chips.eu/>

- **One possible avenue open Proxy PDKs for a selected technology**

- A PDK that is *not the same* as the original PDK, developed independently and openly.
- But designs made using this proxy PDK could be *translated* to be manufactured using the original PDK through a service center.
- **Important:** We need to ensure that the technology provider is not *'against'* this idea
  - better yet is supportive
- This will be tricky, but we think it is doable.

# Next steps: ODE4EC is an opportunity



- **ODE4EC proposal**

- Three sub projects: Digital, Analog, Productivity
- HORIZON-JU-CHIPS-2025-IA-EDA-two-stage proposal,
  - PO stage passed,
  - FPP submitted 19<sup>th</sup> of September !

- **Austria has major parts in the proposal**

- JKU, Yosys are vital parts of the open source HW developments
- We need national co-funding in Austria to make all parts of the proposal supported

- **If accepted funding will start April/May 2026**

- IHP, JKU, ETHZ will have major parts
- And work together with other partners





# Finally, let's not forget the Austrian colleagues of PULP



*From Florian Zaruba's talk on 5<sup>th</sup> RISC-V Symposium*



**Philipp**



**Michael**



**Florian**



# Our WWW page contains a wide collection of talks/papers

PULP Platform Resources Projects About FAQ Privacy Policy Contact



## PULP Platform

Open hardware, the way it should be!

Most of our talks: <https://pulp-platform.org/conferences.html>

RISCY CV32E	Zero R lbex	Snitch	Spatz	Ariane CVA6	ARA	JTAG	SPI	LIC	HCI
RV32	RV32	RV32	RVV	RV64	RVV	UART	I2S	APB	FlooNoC
						DMA	GPIO	AXI4	

And most of our papers: <https://pulp-platform.org/publications.html>

<ul style="list-style-type: none"><li>PULPino, PULPissimo</li><li>Cheshire</li></ul>	<ul style="list-style-type: none"><li>OpenPULP</li><li>ControlPULP</li></ul>	<ul style="list-style-type: none"><li>Hero, Carfield, Astral</li><li>Occamy, Mempoal</li></ul>
<b>IOT</b>		<b>HPC</b>
Accelerators and ISA extensions		
XpulpNN, XpulpTNN	ITA (Transformers)	RBE, NEUREKA (QNNs)
		FFT (DSP)
		REDMULE (FP-Tensor)

Latency 3D-SDR won the Best PhD Forum Award at VLSI-SoC 2024 in Tangier.

**28 August 2024**

Luca Benini received the 2024 TCMM Open Source Hardware Contribution Award for his work on the PULP platform. The award was presented at Hot Chips 2024.



# The future of open-source HW is bright

