EPFL

EMBEDDED SYSTEMS LABORATORY

September 17th, 10-13hs @ ELG 123

Including: X-AGORA, talk by CERN EP SoC team, Apéro

X-AGORA IV + CERN

zoom

the X-HEEP eXchange space

EPFL - Embedded Systems Laboratory (ESL)

juan.sapriza@epfl.ch

New to the X-AGORA?

It’s a space where we can share our work around X-HEEP so we can better collaborate :)

ANYONE can do a **1-minute** presentation
(semester projects are super welcomed!)

EVERYONE has edit access
so you have to be careful with other people’s slides

Copy the X-ample slide and add your stuff
The moderators will move it to the proper place or contact you if needed.

Remember to add your contact information on the bottom of the slide!

Overview

10 am

- Introduction
- Tools and Support
- Silicon implementations
- Individual presentations
- Going forward...

11 am

- Presentation by CERN EP SoC team

12 am

- Closing remarks by Prof. David Atienza
- Apéro

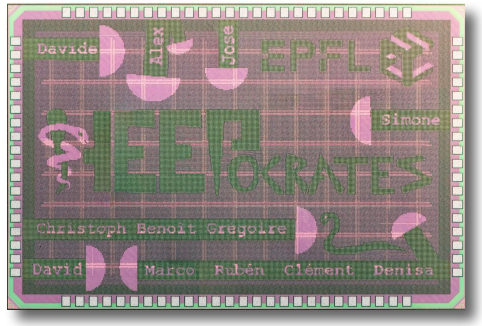
EPFL

Silicon Implementations (CHEEPs)

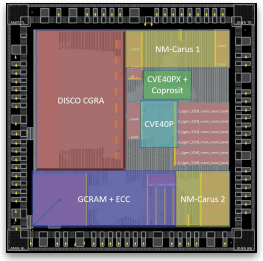
CHEEPs

EPFL

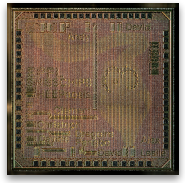
HEEPocrates



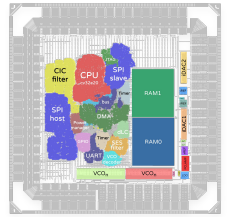
HEEPatia



HEEPnosis

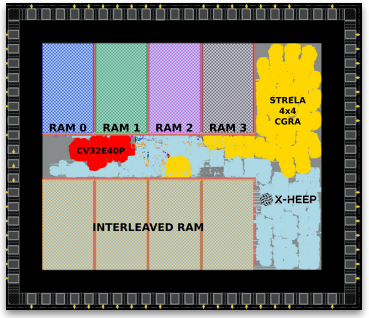


EEPiderWMis

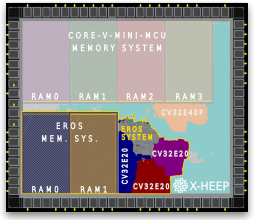


X-TRELA

CEI UPM

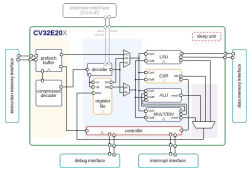


X-EROS



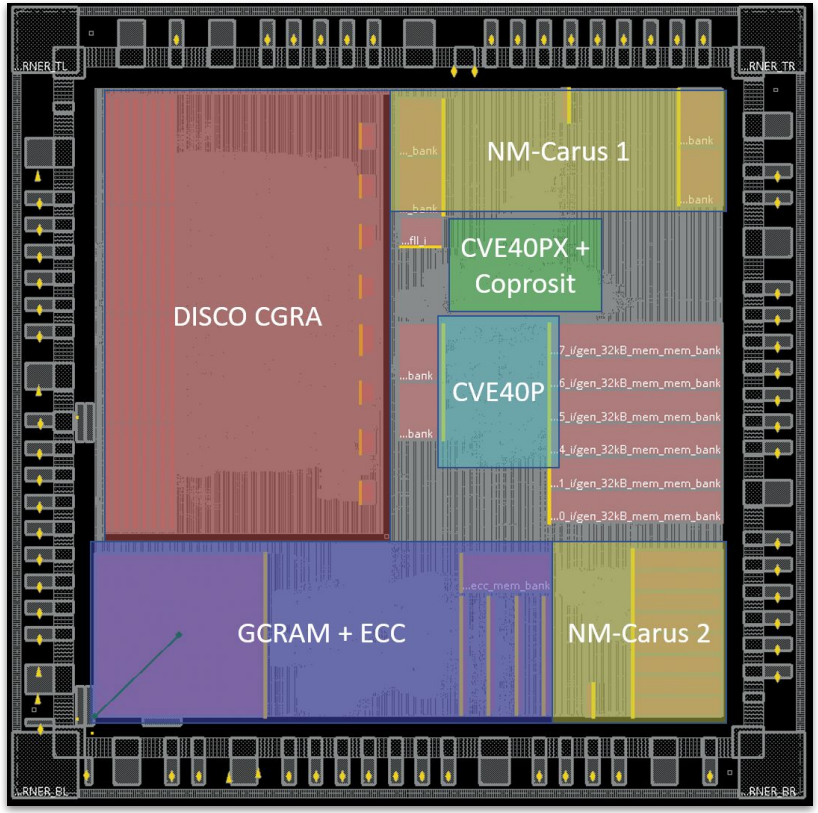
X-HEEP updates

verilator reads FLASH



cve2 + XIF

HEEPatia



4 mm²

@ TSCM16 - Q1 2025

Dual-core

- 2 x CV32e40P

Co-Processor

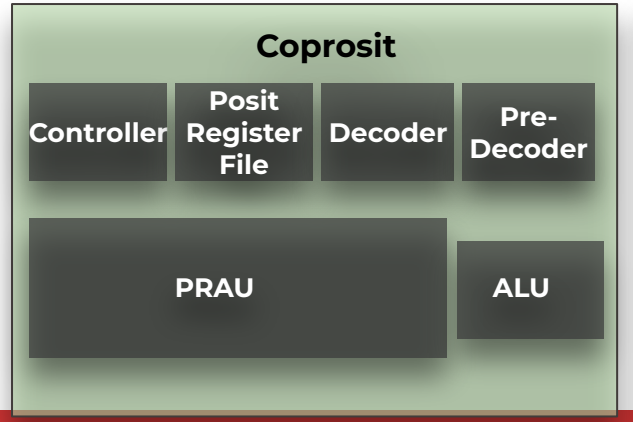
- RV32IMFCXpulpXposit

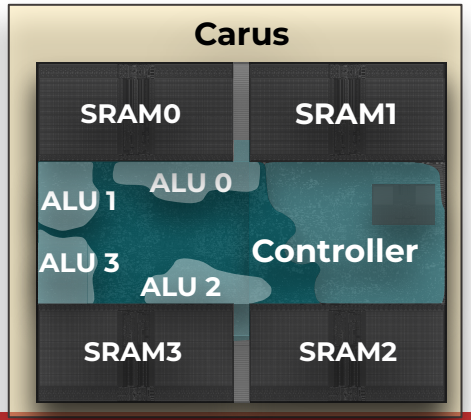
Accelerators for edgeAI

- Very Wide Register CGRA
- RISC-V-based NMC

Shared memory

- 224 kB SRAM
- Embedded DRAM (eDRAM)

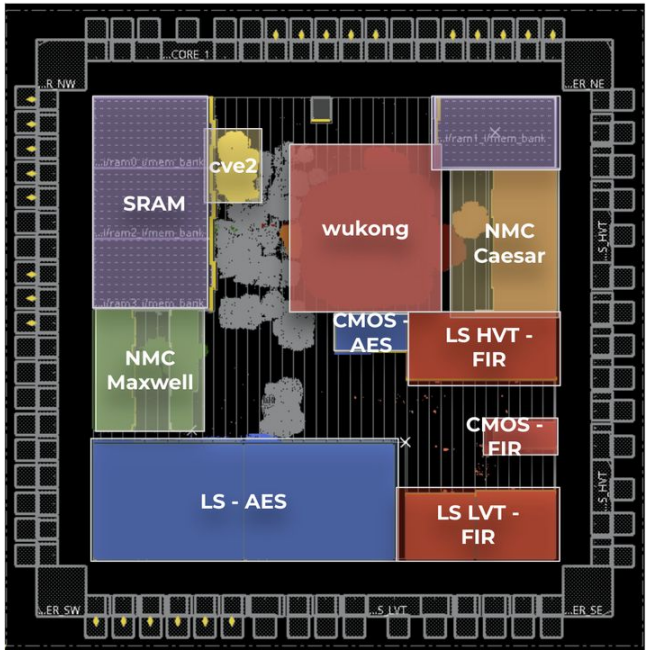


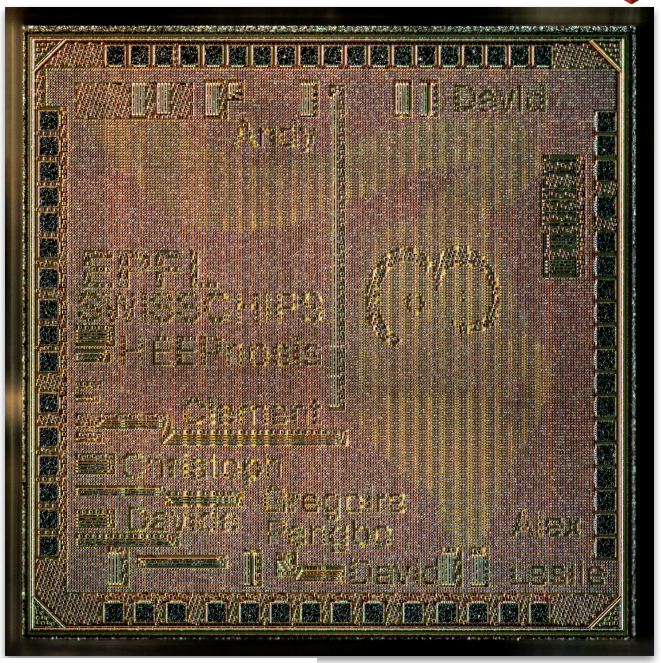


HEEPnosis

2 mm²

@ GF22FDX - Q1 2025





Single-core

- CV32e20

Shared Memory

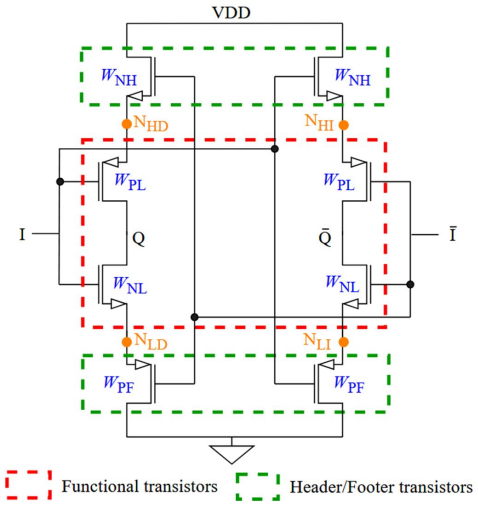
- 128 kB SRAM

Accelerators for edgeAI

- Maxwell (NMC)
- NM-Caesar (NMC)
- Wukong (Variable precision vector unit)

Ultra low power edgeAI

- ULP Leakage Suppression logic (FIR filters on different flavors for testing it)



X-TRELA chip

4.7 mm²

@ TSMC 65nm LP CMOS - Q1 2025

Single-core

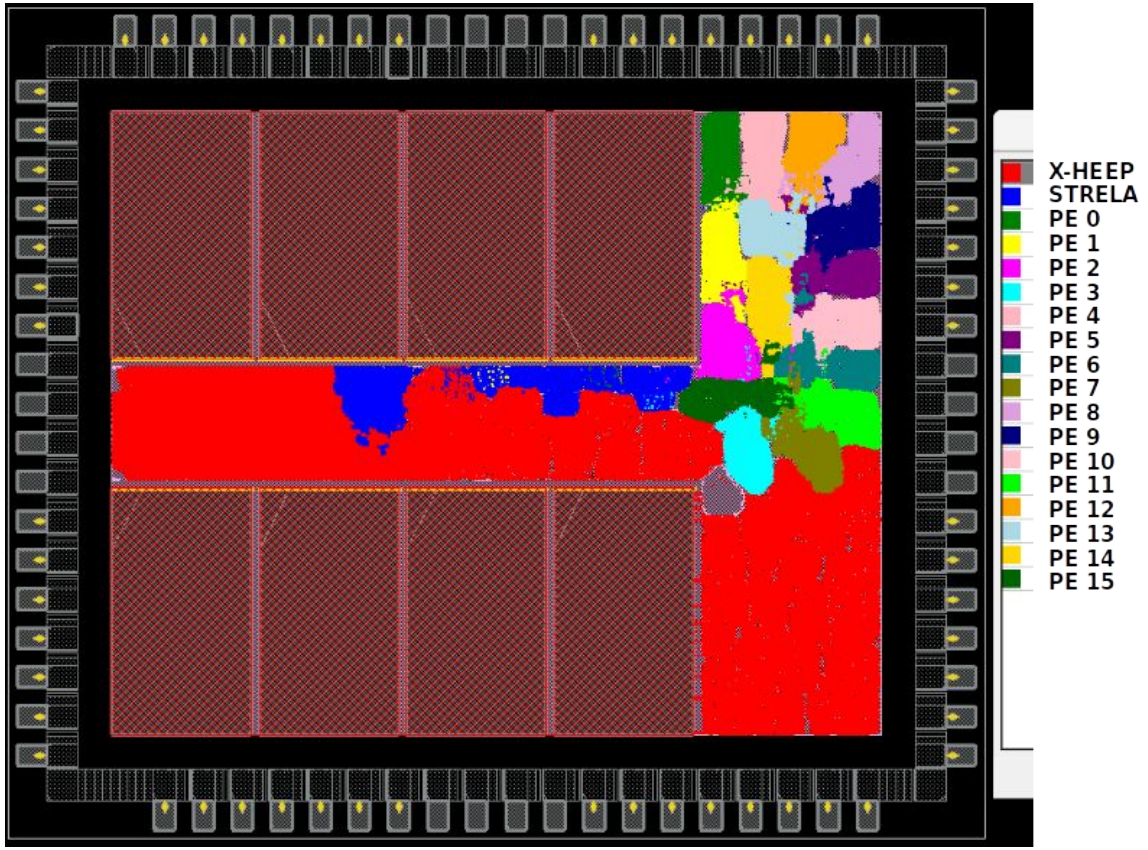
- CV32e40P

Accelerator

- 4x4 STRELA CGRA

Memory

- NtoM bus
- 256 kB SRAM (4 interleaved banks)



HEEPidermis [1]

2.25 mm²

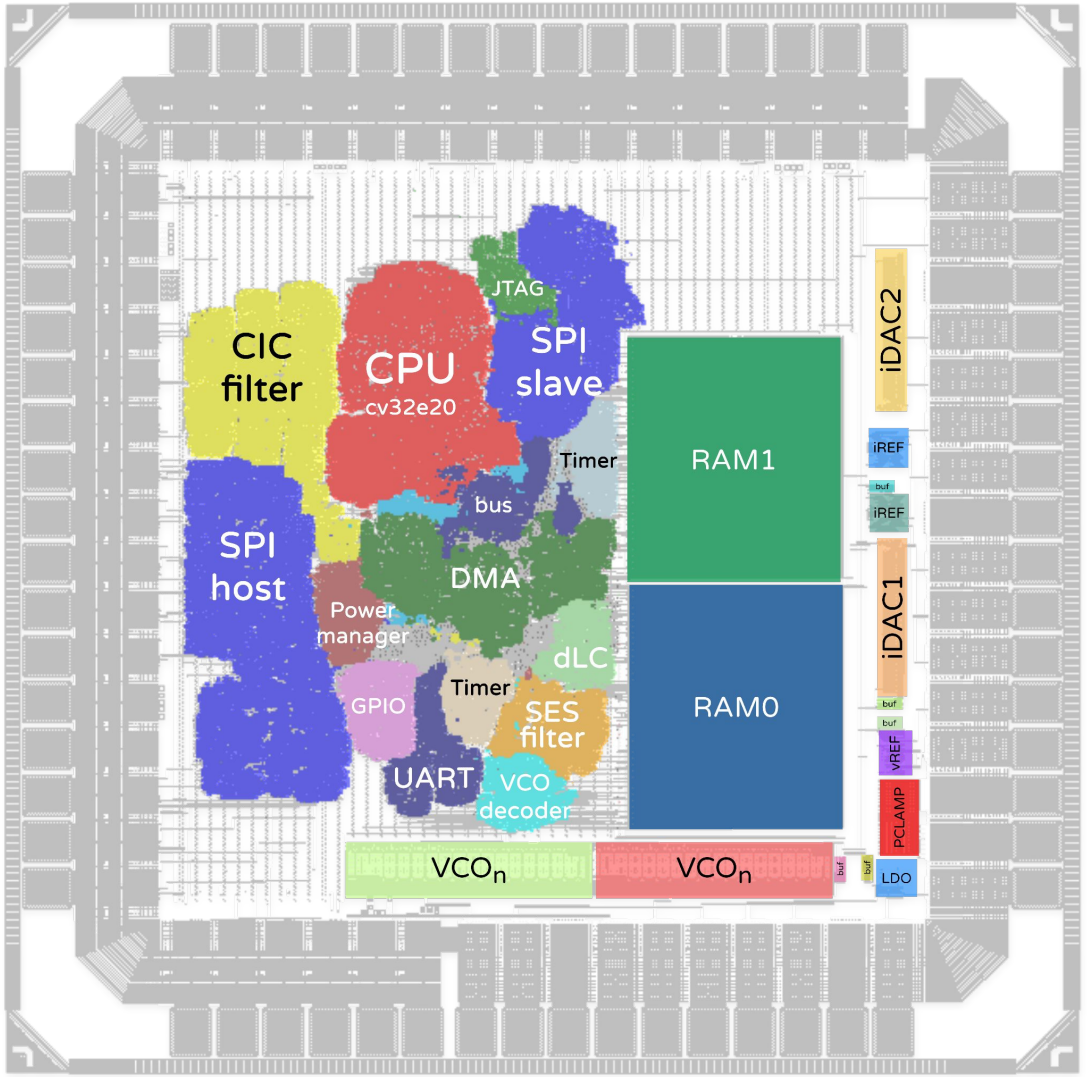
@ TSCM65 - Q2 2025

A versatile SoC for impedance measurement

- 2× current DAC
- 2× VCO-based ADC
- Interface with analog blocks as memory-mapped peripherals

- 1 mm² core area
- < 100 µW acquiring data (25 µW leakage)

- RTL and behavioral models are open source :)



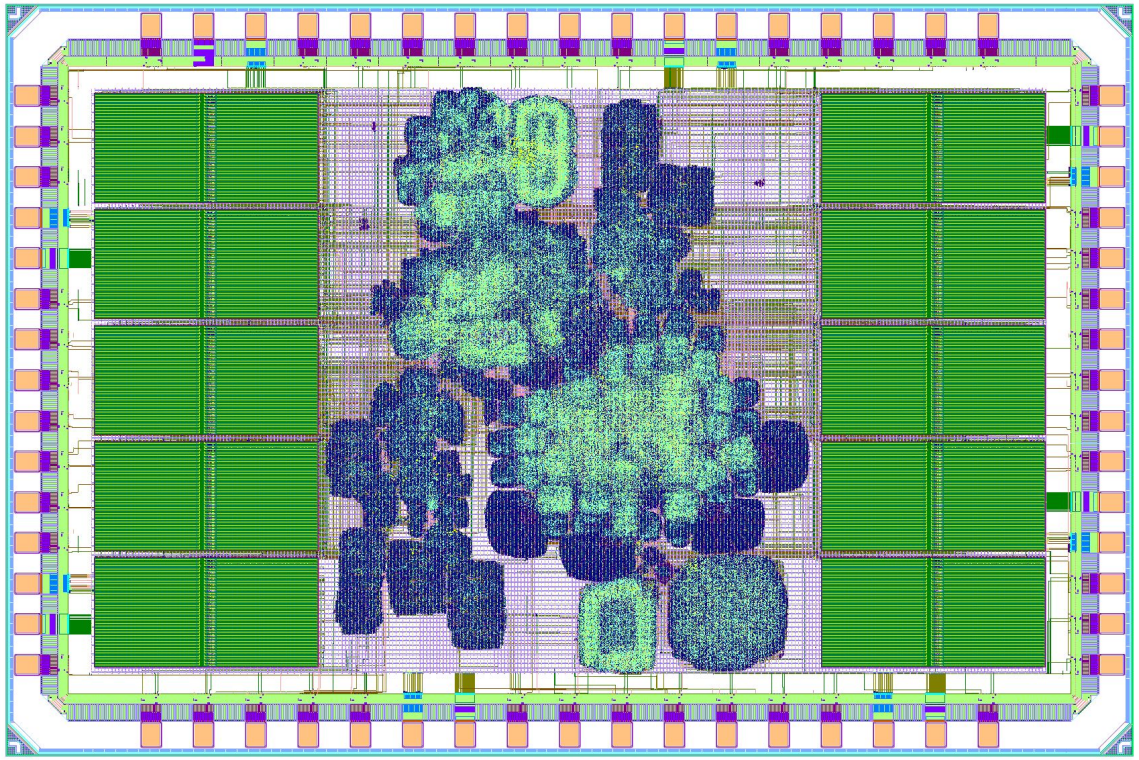
[1] Sapriza, J., Grassano, B., Naclerio, A., Quadri, F., Terzano, T., Mallasén, D., ... & Atienza, D. (2025). HEEPidermis: a versatile SoC for BioZ recording. arXiv preprint arXiv:2509.04528.

Upcoming CHEEPs: HEEPocraneos

- Implantable neural interfaces
- Integrate
  - High resolution ADC
  - Event-driven ADC
  - Energy harvesting
  - Wireless communication

- Q2 2026 - TSMC 65nm LP
- 1 mm^2
- 50 µW leakage + <200 µW acquisition ×16ch

Upcoming CHEEPs: POLHEEPO



- **SIZE:**
  2mm x 3mm @ TSMC 65nm

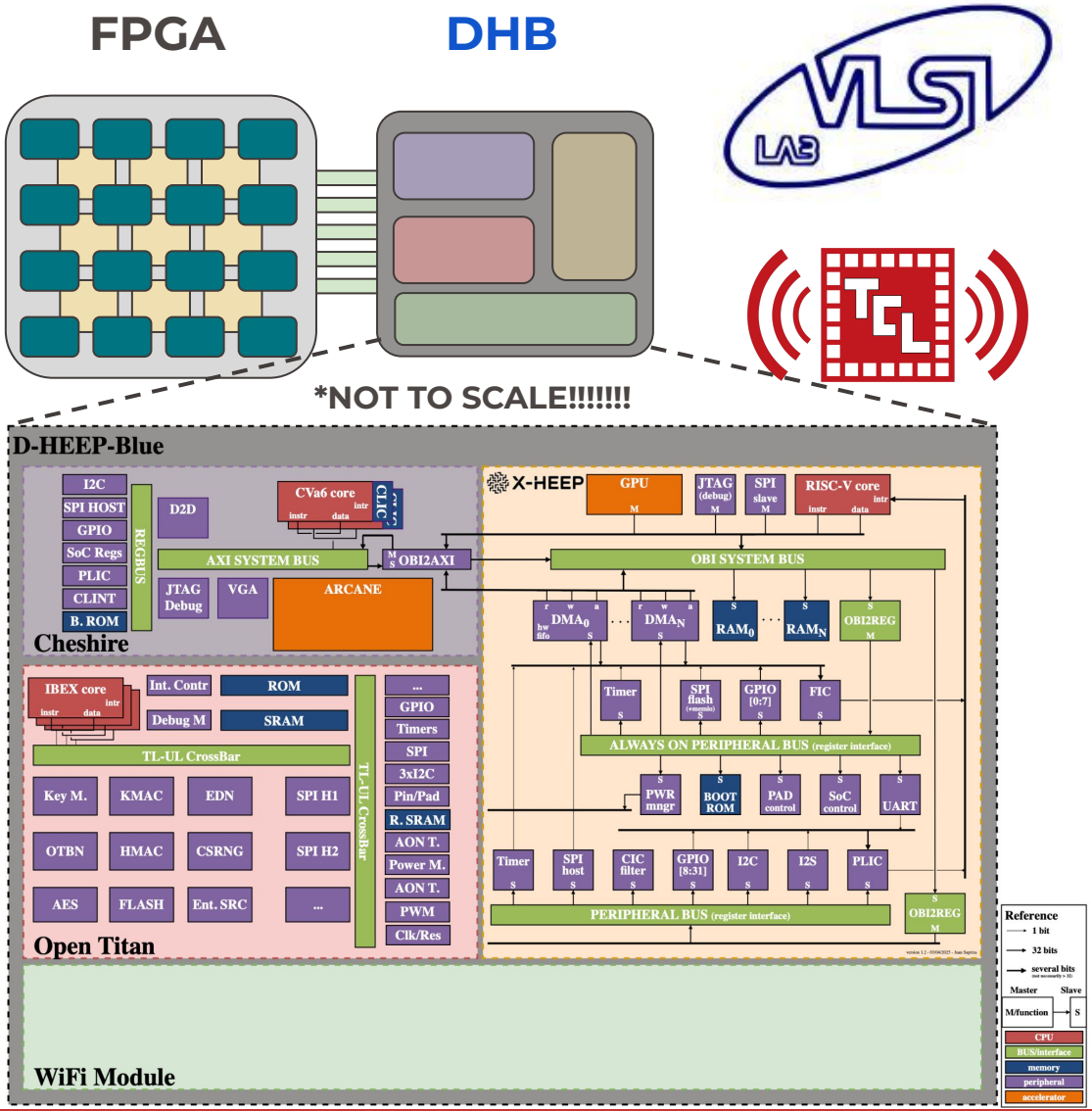
- **X-HEEP:**
- 8x32kiB interleaved banks
- 2x32kiB interleaved banks
- cv32e40p core (FPU + PULP extensions)
- 4 channel DMA (4MP + FIFO)

- **Targets:**
- **PQC with side-channel attacks resiliency**
- **AI/ML**

- **External subsystem:**
- **PUFfo**
- **KECCAK**
- **ASCON**
- **ROGUE**

Upcoming CHEEPs: D-HEEP Blue (DHB)

- **Architecture:**
  - **Application class:** Cheshire CVA6 (PULP)
  - **Compute:**
    - Tightly-coupled ARCANE NMC LLC
    - Real-time X-HEEP island with integrated e-GPU
  - **Reliability/Security:** OpenTitan island for secure boot, key management, and fault containment
- **FPGA pairing via reliable D2D:** high-bandwidth, low-latency link
- **WiFi Module:** OpenWiFi
- **Targets:** automotive, space, robotics, more
- **Value:** better perf/W, strong security, easy adaptability



Tools & Support

New editable diagram

Feel free to make your own

- Editable on the repo
- Create a “visual identity”
- You don’t need to start from scratch

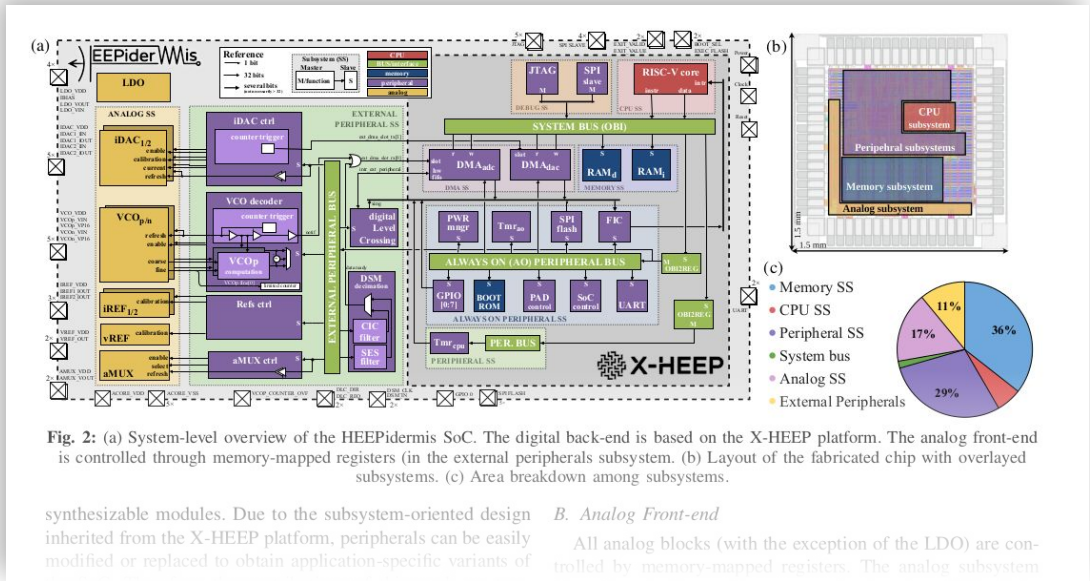
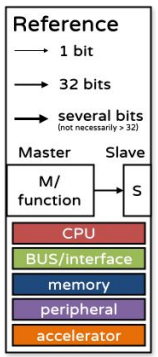


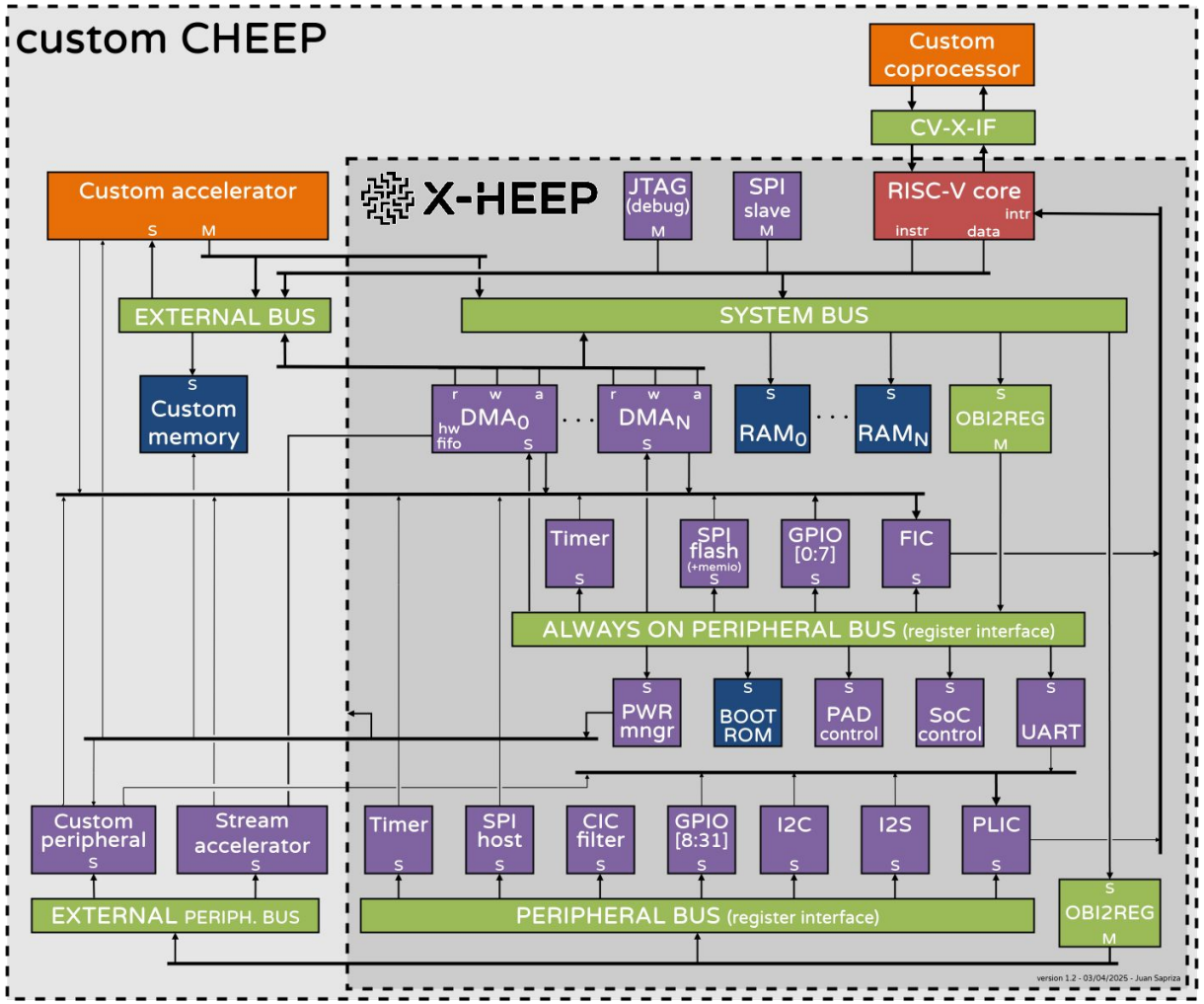
Fig. 2: (a) System-level overview of the HEEPidermis SoC. The digital back-end is based on the X-HEEP platform. The analog front-end is controlled through memory-mapped registers (in the external peripherals subsystem. (b) Layout of the fabricated chip with overlayed subsystems. (c) Area breakdown among subsystems.

synthesizable modules. Due to the subsystem-oriented design inherited from the X-HEEP platform, peripherals can be easily modified or replaced to obtain application-specific variants of

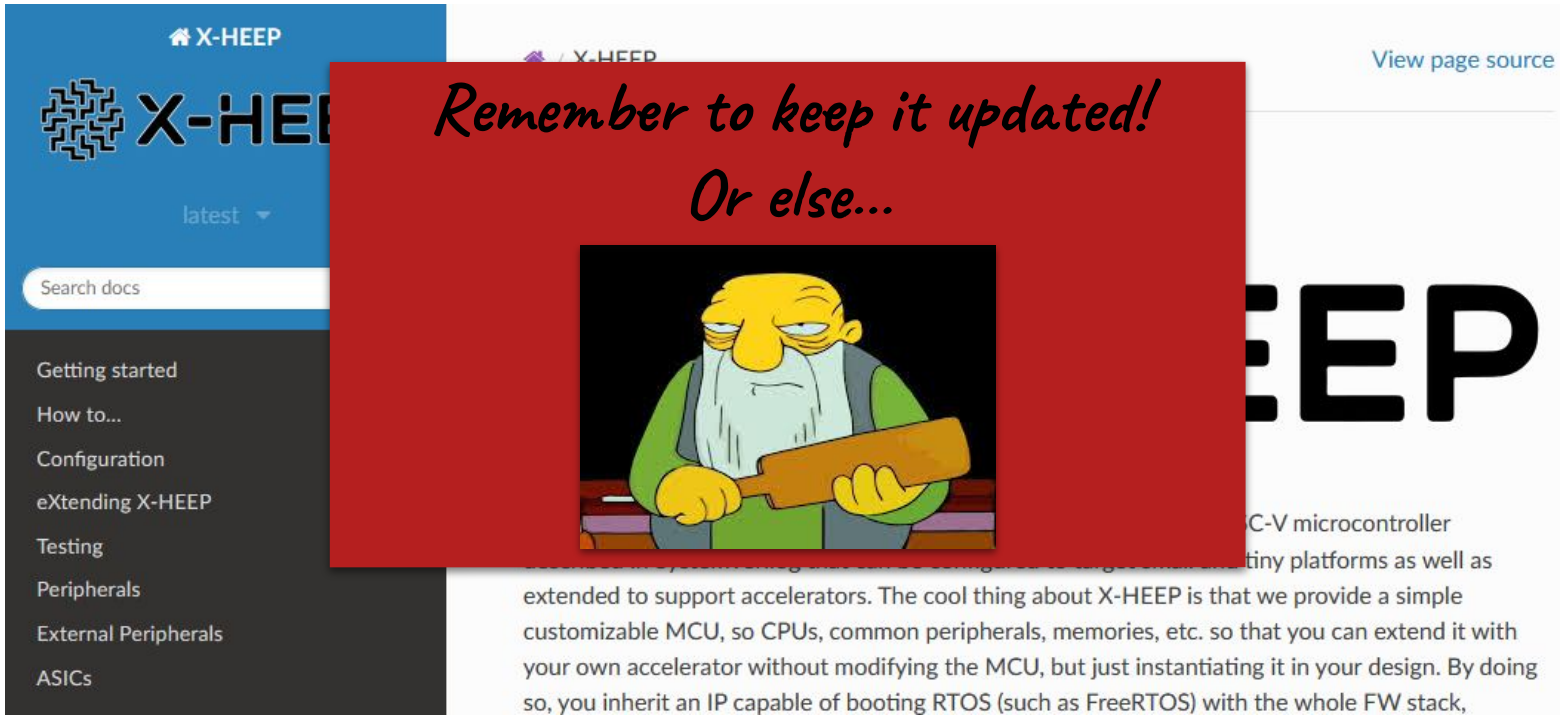
B. Analog Front-end

All analog blocks (with the exception of the LDO) are controlled by memory-mapped registers. The analog subsystem





Updated documentation



ReadTheDocs

- New “Getting started”
- Reorganized “How to…”
- Configuration (mcu-gen)
- …

`docs/source`
`&`
`How to… Update the documentation`

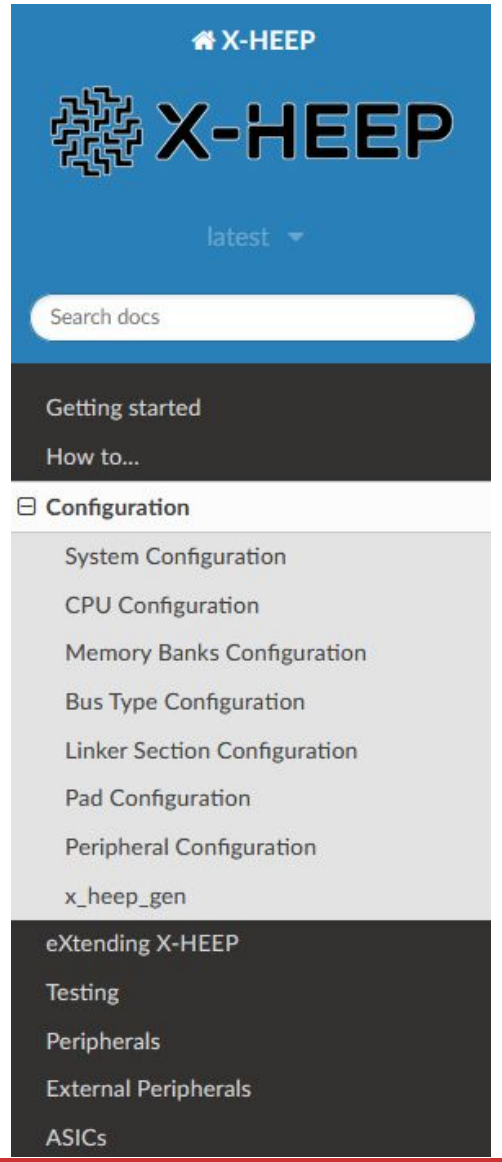
Do you have any local documentation or How to...?
Please let me know!

Changes to mcu-gen

Easier and more flexible configuration

- Supporting modelling the HW in Python in addition to HJSON
  - ✓ Using classes for HW abstraction (system, subdomains, peripherals...)
  - ✓ Validating configurations
  - ✓ Retrocompatible with HJSON
  - ✓ Easier to maintain in the templates (call to Python functions)

`util/x_heep_gen`



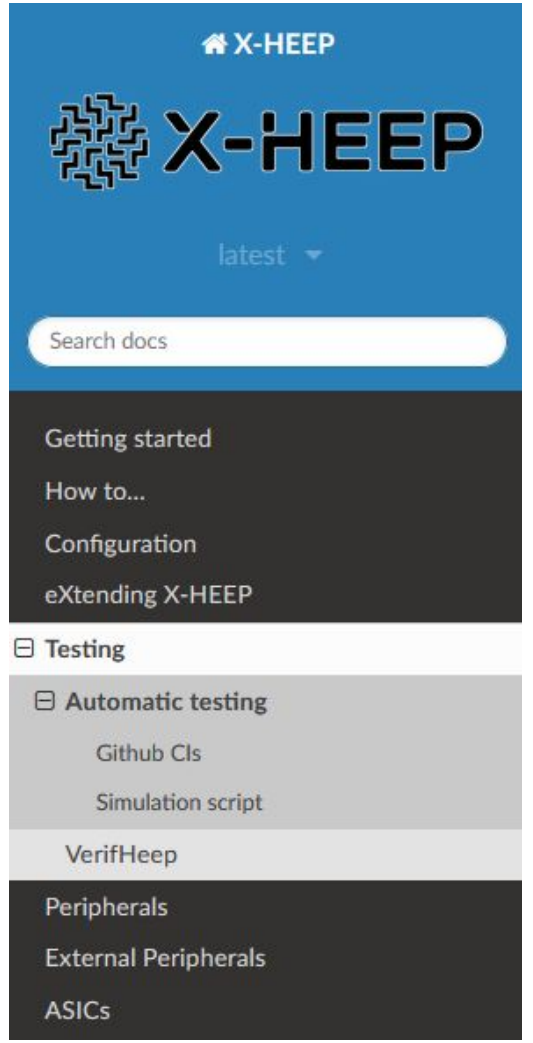
Changes to tests

Revamped application testing scripts

- Modular (Python) to include future tests
- Used in the CI
- Also for local quick debugging `make test`

`test/test_apps`

Would you like to make sure that things don't break?
Add to the tests!



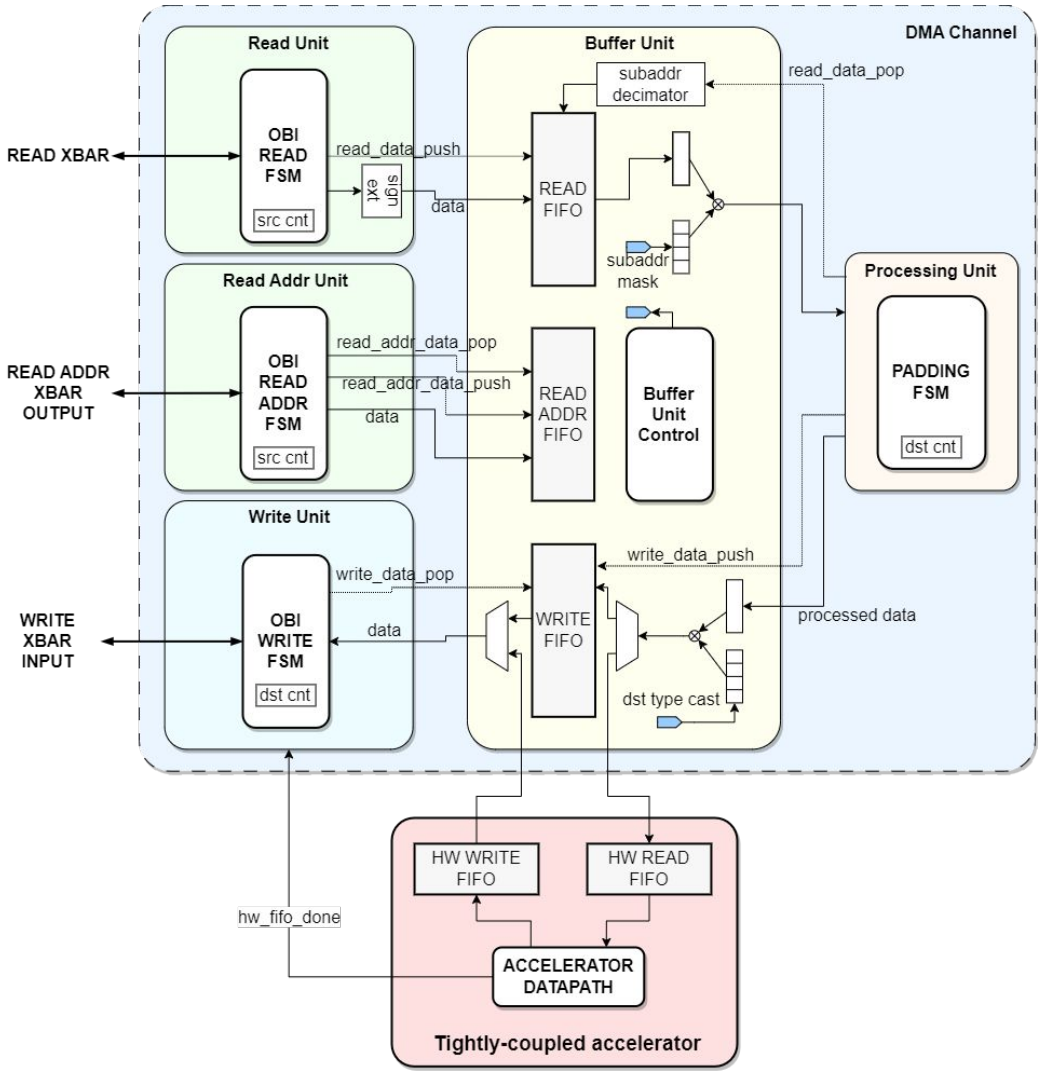
EPFL

Individual presentations

DMA-based Acceleration Made Easy
Tailoring the DMA unit for optimal trade-offs

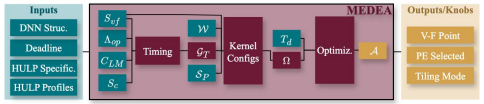
New features!

- FIFO interface for ***tightly-coupled streaming acceleration***
- Improved ***configurability***:
  - Address mode
  - Zero-padding
  - Sub-address mode
  - HW-FIFO mode



MEDEA

Manager for **E**nergy-efficient **D**NNs on h**E**terogeneous ULP **A**rchitectures



Inputs:

- AI model representation
- Application Deadline
- Platform Specs
  - Mem, Operating points, PEs, etc.
- Platform Profiles
  - Power & timing
  - Extrapolating capability

MEDEA's Core Logic:

- Goal: Minimize energy while respecting constraints
  - deadline, memory, etc.
- Defining as an optimization problem
  - MCKP → ILP

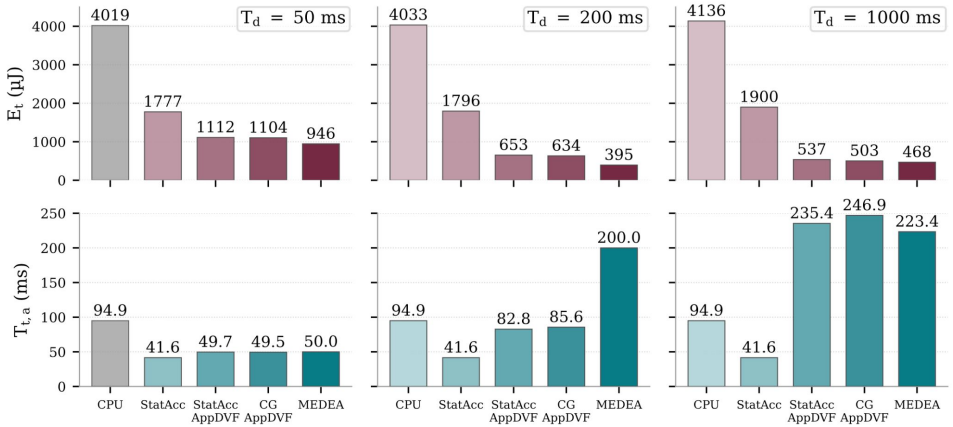
Knobs it Controls:

- Which PE to use?
- How fast to run it? (VF)
- How to manage memory? (Tiling strategy)

Evaluation Platform & Application:

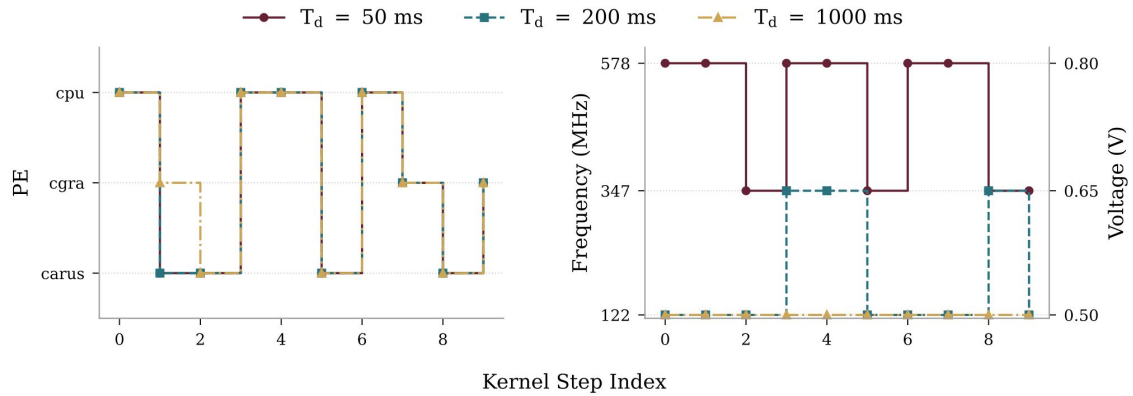
- HEEPtimize GF 22nm (OpenEdgeCGRA, NM-Carus , CV32E40P)
- Transformer for Seizure Detection

Timing and Energy compared to SToA



- **Energy reduction, upto 38%**
- **Utilizing the entire available time budget**

An snapshot of scheduling decisions

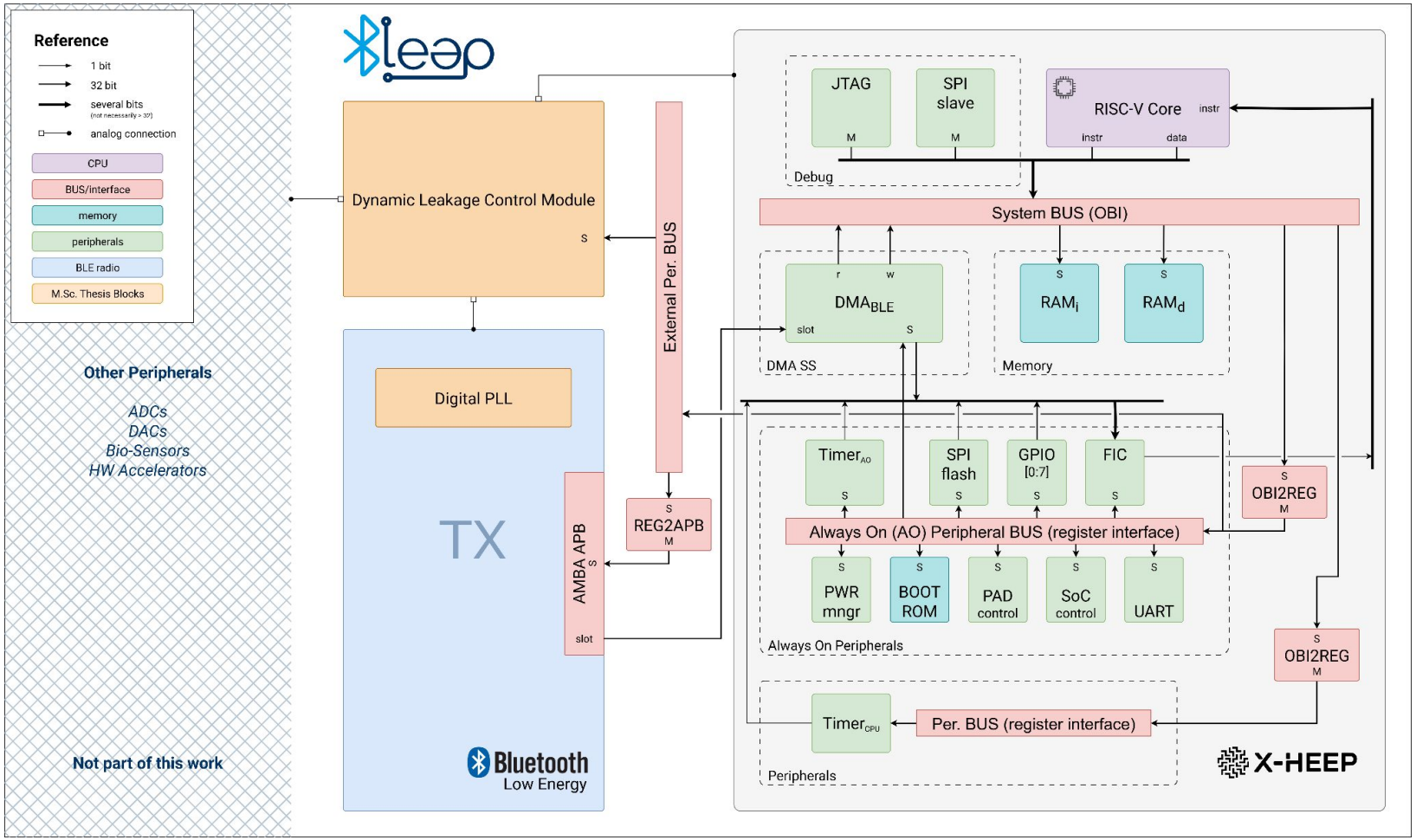


- **It is thanks to its fine-grained adaptive control**

[1] H. Taji, et al. "MEDEA: A Design-Time Multi-Objective Manager for Energy-Efficient DNN Inference on Heterogeneous Ultra-Low Power Platforms." arXiv preprint arXiv:2506.19067 (2025).

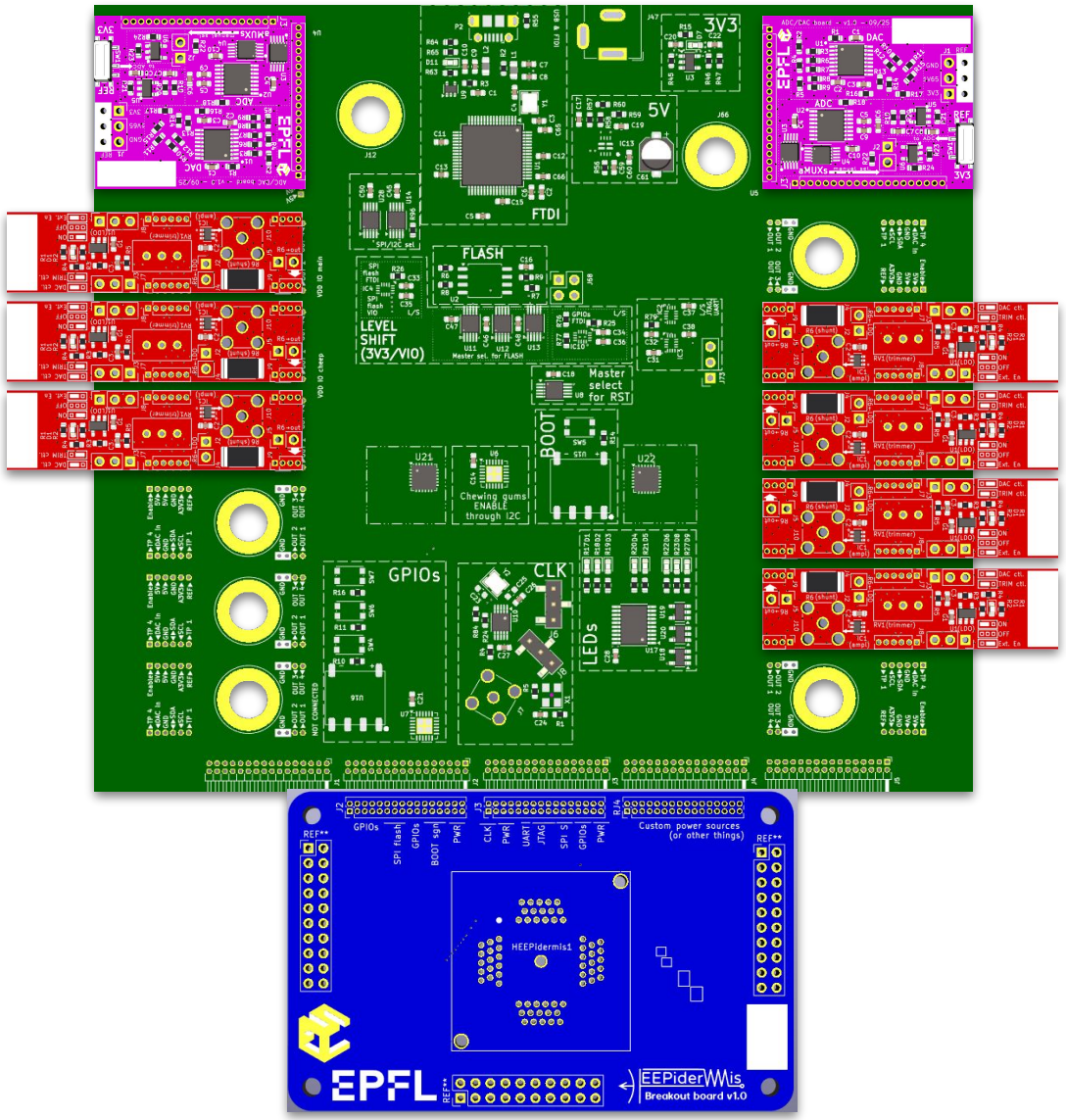
Overview

- Combine X-HEEP with a BLE transmitter
- Dynamic Leakage Control Module
  - Runtime power optimization via leakage tuning
  - Switch between performance/power modes on-the-fly
- Easy to integrate with other peripherals (e.g., Bio-Sensors, ML Accelerators, …)
- GF-22nm FDSOI

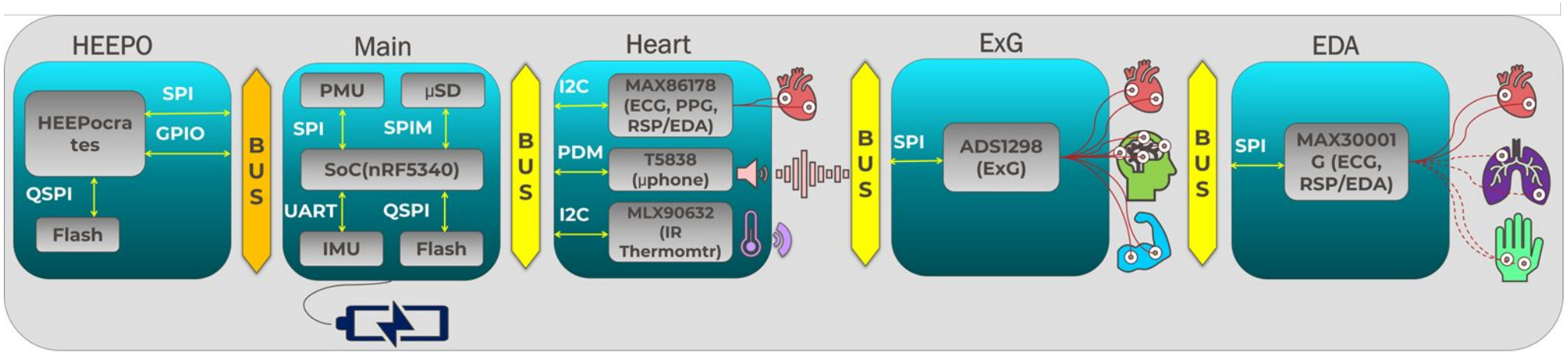


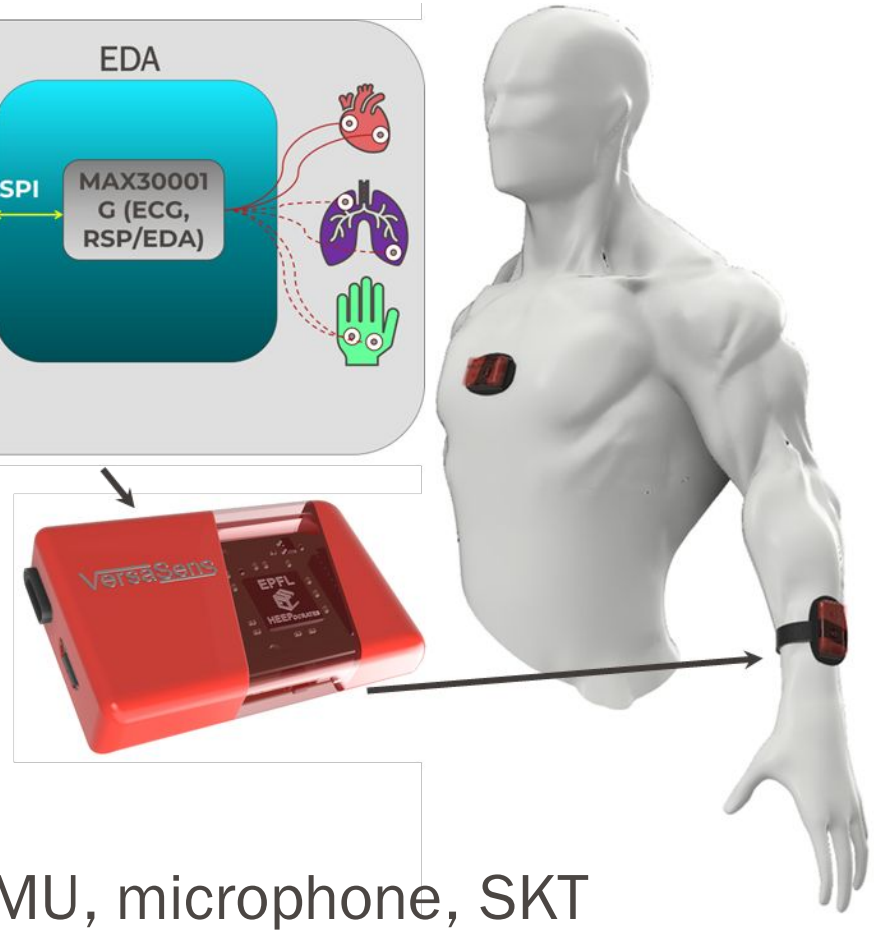
CHEEP-boards
Testing PCBs for CHEEPs

- **Main board**
  Programmer and interfaces
  (flash, JTAG, UART, GPIOs)
- **Power boards**
  Supply different voltages/currents
  You can create your own custom ones
- **ADC/DAC boards**
  Measure consumption and control
- **Breakout board**
  To connect your cheep to the desired sources
  You need to do this one
  You can add custom features here as well



VersaSens
A Platform for Hardware/Software Co-design





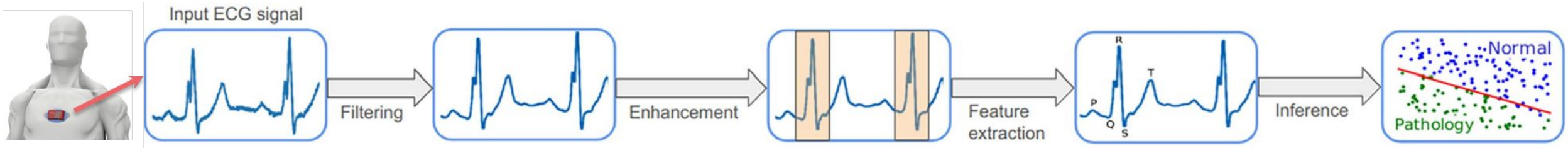
- **Compact, modular, configurable, extendable.**
- Operation: **idle, storing, streaming + storing**
- Real-time bio-signal acquisition and synchronization
- **Heepocrates:** Co-processor for real-time **inference**
- Modalities: biopotentials, bioimpedances, bio-optic, IMU, microphone, SKT
- Create your own smart wearable sensor

T. A. Najafi, et al., "VersaSens: An Extendable Multimodal Platform for Next-Generation Edge-AI Wearables", *IEEE TCASAI*, Sept. 2024

VersaSens
Heartbeat Classifier Application

- Modalities: ECG (ExG)
- Pipeline: Filtering, Enhancement, Feature Extraction, Classifier
- Processing:
  - HEEPO
  - Frequency 170MHz , 830mV
- Window (12s)

Parameter	Processing	Deep Sleep
Duration	22 ms	11978 ms
Power Consumption	8.68 mW	0.29 mW
Voltage	830 mV	830 mV
Frequency	170 MHz	32 KHz
Energy Consumption	0.19 mJ	3.47 mJ
Total Energy		**3.66 mJ**

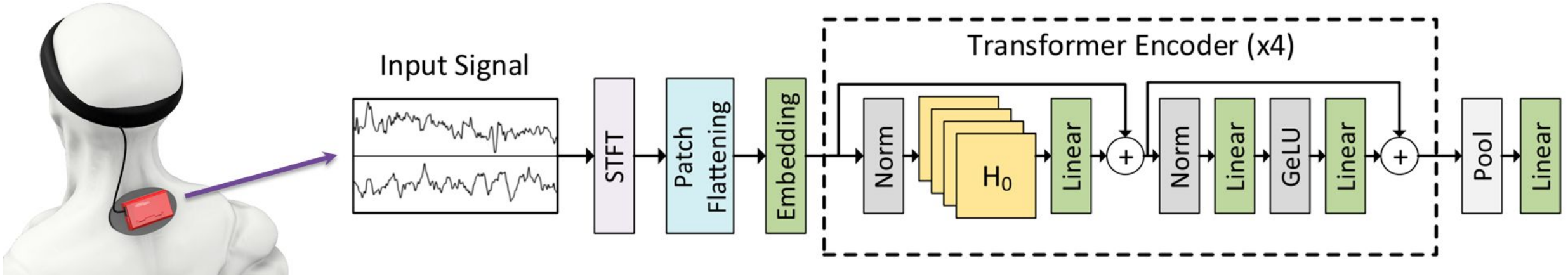


T. A. Najafi, et al., "VersaSens: An Extendable Multimodal Platform for Next-Generation Edge-AI Wearables", *IEEE TCASAI*, Sept. 2024

VersaSens
Seizure Detection Application

- Modalities: EEG (ExG)
- Models: VisionTransformer
- Processing:
  - HEEPO (w/ CGRA and w/o CGRA),
  - Frequency 160MHz , 830mV
- Window (12s)

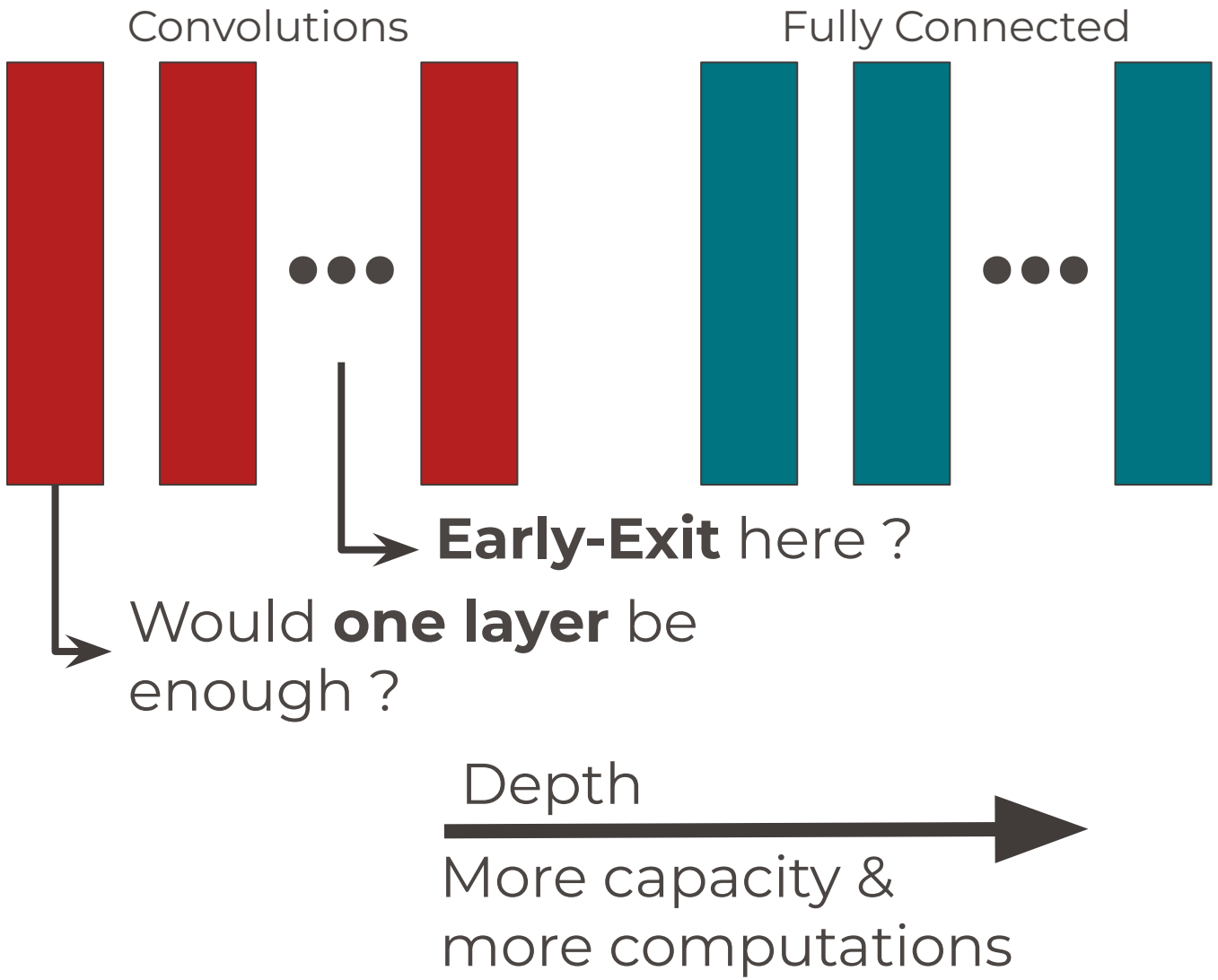
Parameter	With CGRA	Without CGRA	Deep Sleep
Processing time	53 ms	79 ms	11947 ms
Power Consumption	8.86 mW	8.83 mW	0.29 mW
Voltage	830 mV	830 mV	830 mV
Frequency	160 MHz	160 MHz	32 KHz
Energy Consumption	0.47 mJ	0.70 mJ	3.46 mJ
Total Energy	**3.93 mJ**	**4.16 mJ**	



T. A. Najafi, et al., "VersaSens: An Extendable Multimodal Platform for Next-Generation Edge-AI Wearables", *IEEE TCASAI*, Sept. 2024

Extremely tiny CNNs

with guarantees



You only need **one** convolution layer.

Then, you can **always*** **early exit.**

* Under conditions - let's discuss.

CGRA estimation tool[1]
Flexible framework for characterizing CGRA kernel execution

Instantaneous multi-level results

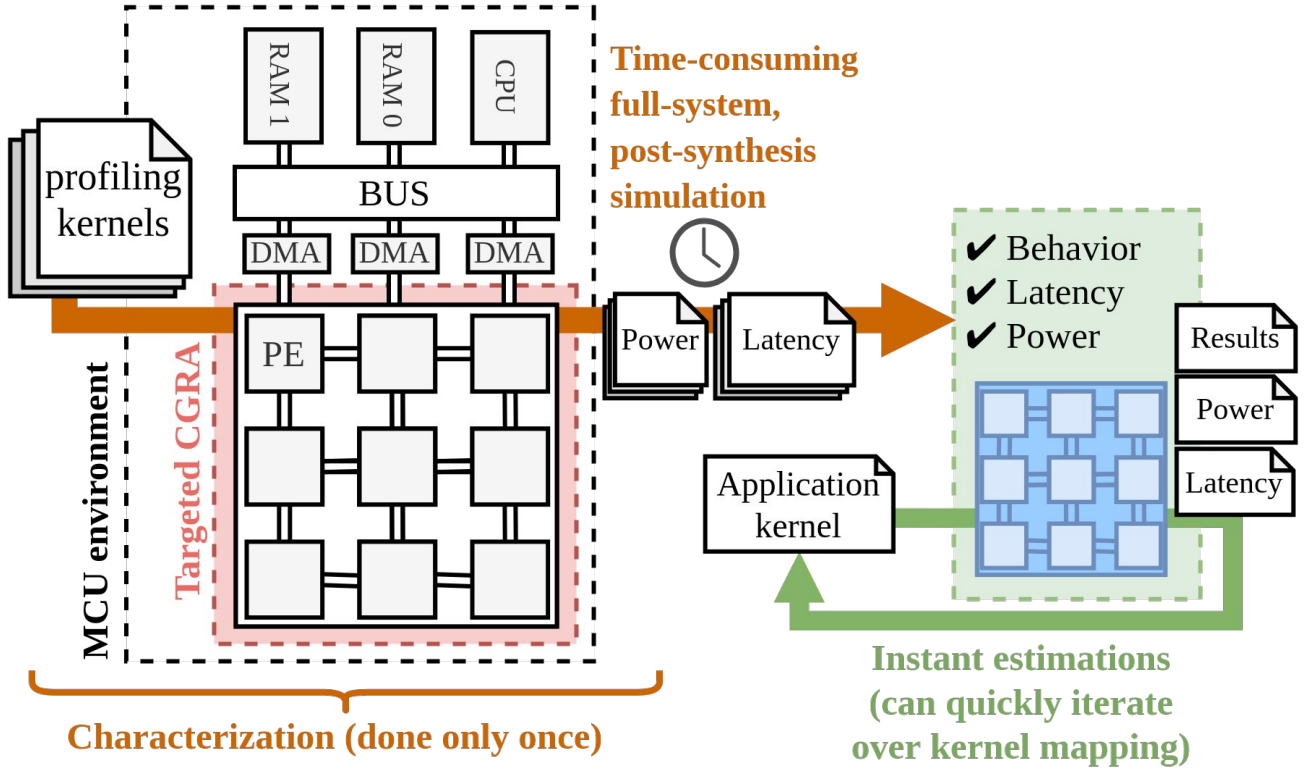
- P/E/L Estimation
- HW (PE, CGRA) and op (Instr., exec.)
- Quickly write and debug

Actionable insights for HW-SW co-design

- Avg. error: 22% (power), ~0% (latency)
- HW: diff. topologies
- SW: diff. mappings

Build profiling model

- Based on OpenEdgeCGRA[1]

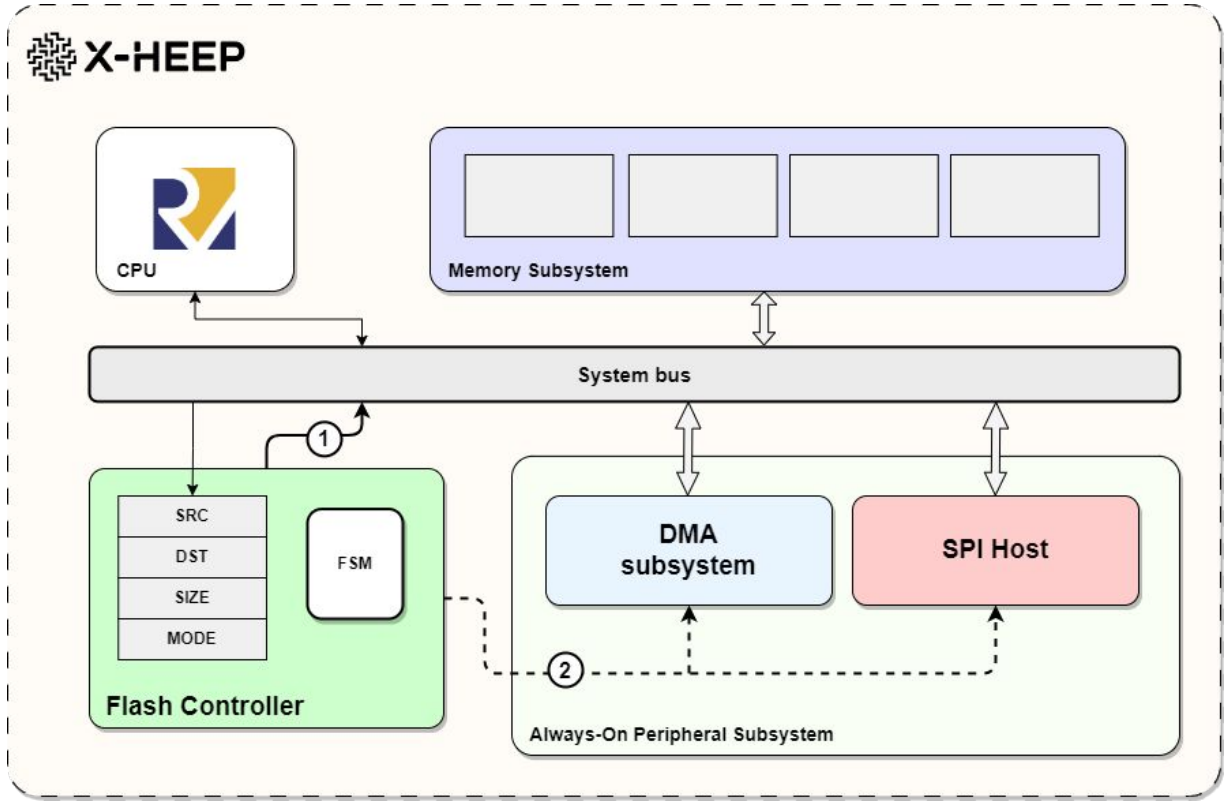


[1] Aspros, M. H., Sapriza, J., Ansaloni, G., & Atienza, D. (2025, May). A flexible framework for early power and timing comparison of time-multiplexed CGRA kernel executions. In Proceedings of the 22nd ACM International Conference on Computing Frontiers: Workshops and Special Sessions (pp. 62-65).

[2] R. Álvarez, B. Denkinger, J. Sapriza, J. Calero, G. Ansaloni, and D. Atienza Alonso. “An Open-Hardware Coarse-Grained Reconfigurable Array for Edge Computing”. Proceedings of the 20th ACM International Conference on Computing Frontiers (CF '23). Association for Computing Machinery, New York, NY, USA, 391–392. 2023

Automating Flash Operations

- SPI read and write operations to the Flash memory are *slow*
  - *High CPU utilization*
- **Automate SPI and DMA configuration** in HW
  - System bus transactions
  - Direct register configuration



MORPHEUS: a wake-up controller for x-heep

Context & Motivation

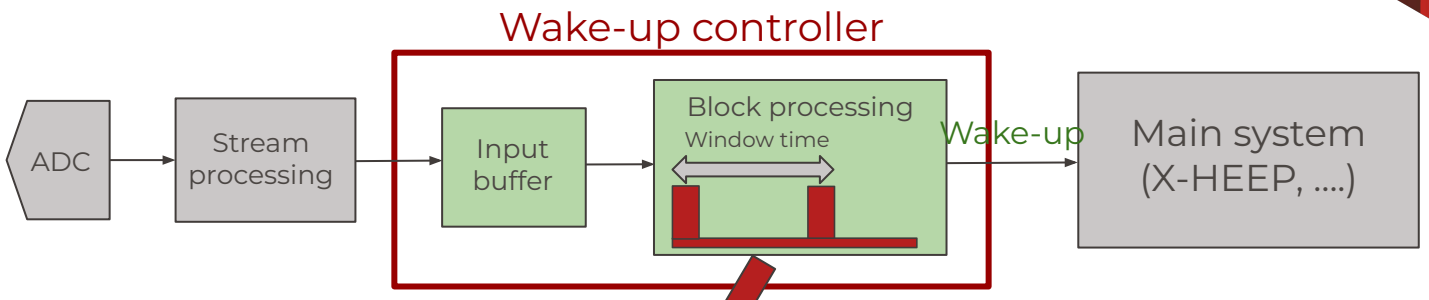
- Real-time monitoring app. (KWS, arrhythmia, seizure detection..)
- Edge AI → specific phases
- “Smart” controller to wake-up the system to process useful info

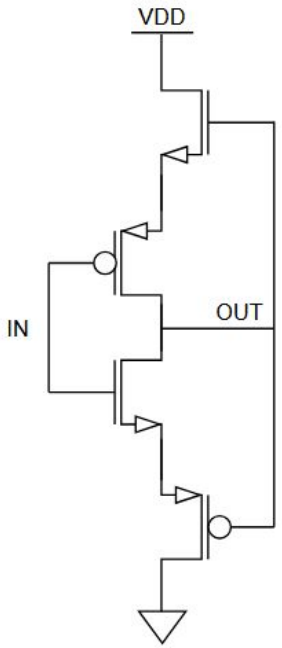
Specifications

- Accurate (Light ML)
- Energy efficient (NM-Carus)
- Low leakage (LL) AO memory (DLSL) → dominate overall energy
- Standard cell memory (SCM) for better voltage scaling

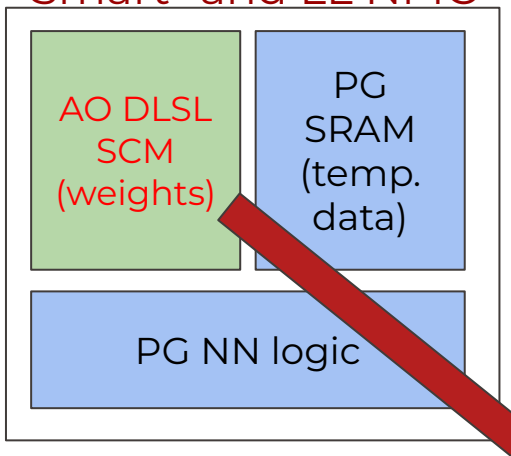
Estimations

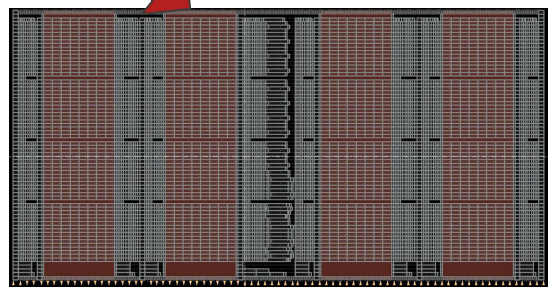
- 22x leakage reduction wrt. 64kB SRAM
- 22x area overhead wrt 64kB SRAM





DLSL Inverter

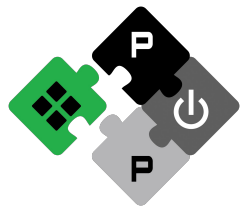


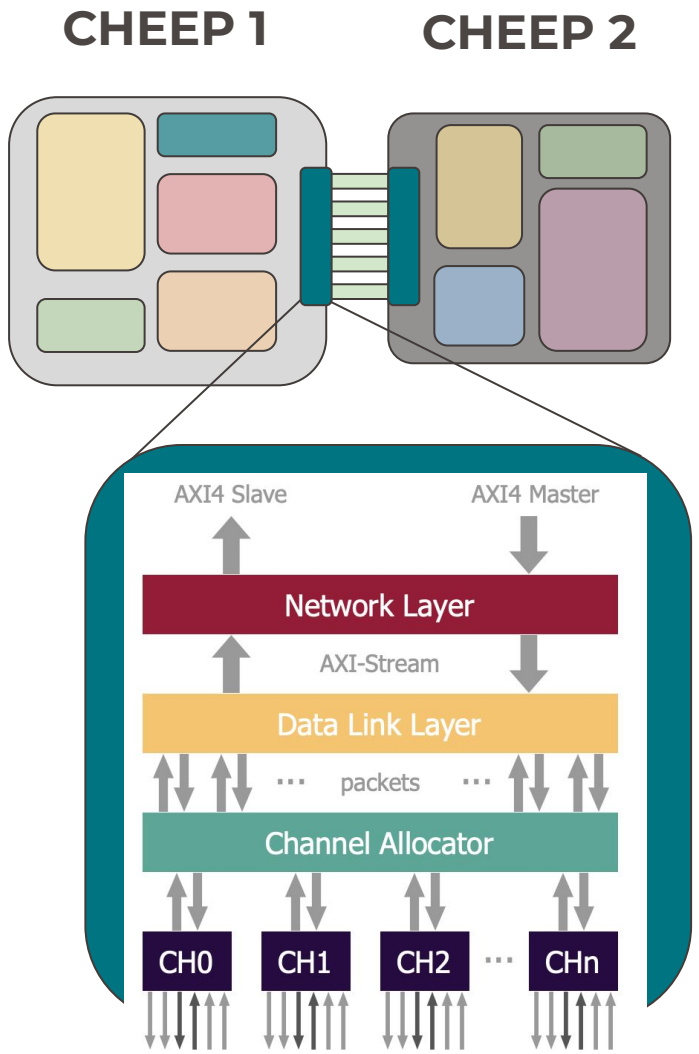


DLSL SCM (in-house compiler)

Die-to-Die Interconnect
X-HEEP as a Chiplet

- **D2D IP on PULP-based Serial Link**
- All-digital **DDR/SDR** link, **source-synchronous**
- Supports **AXI**, now extended with **OBI**
- **Customizable and scalable**
  - Dynamic payload options: multi-data per address or 1:1 transfers
  - Fits high-bandwidth die-to-die or lightweight preload use cases
- Ongoing **enhancements for optimized communication**





EPFL

Going forward...

Going forward...

Easier integration and eXtension

- Improved mcu-gen
- AMS front-ends
- Compatibility with more FPGAs

Scaling up

- NoC-SoC for more accelerators

Scaling down

- Low power logic & IPs
- Power gating and body-bias

Beyond the lab

- New translational applications
- Improving robustness & security

Conclusions

Template for everyone

- build your accelerator and inherit a RISC-V MCU

Lower Access Barrier

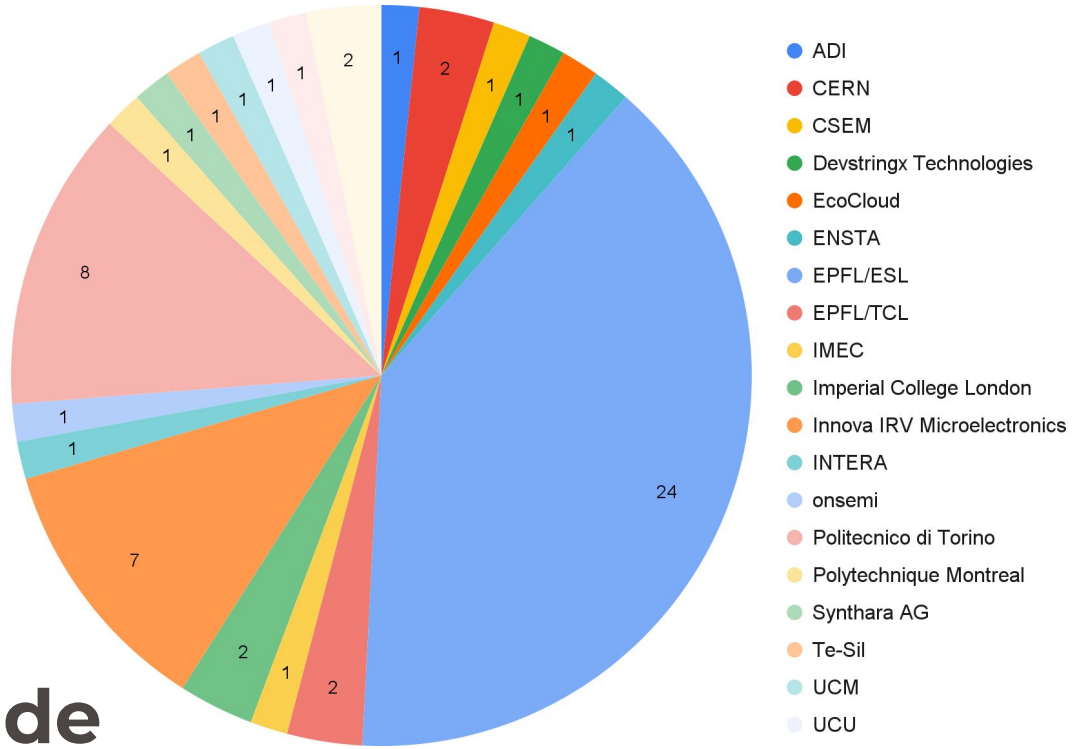
- new users building their own chips, education, demos, etc.
- universities and start-ups!

X-AGORA

- let's work all together as a big group!

Thanks for all the work you made

- external users are working with us making X-HEEP better



Links

X-HEEP

Read the docs

docker image

FEMU

HW

SW

EEPiderWMis

Fork us!

and star ⭐ if you like it