neurSoC — A multiprocessor System-on-Chip with in-memory neural processing unit

COORDINATOR: STMICROELECTRONICS

13 PARTNERS
9 COUNTRIES

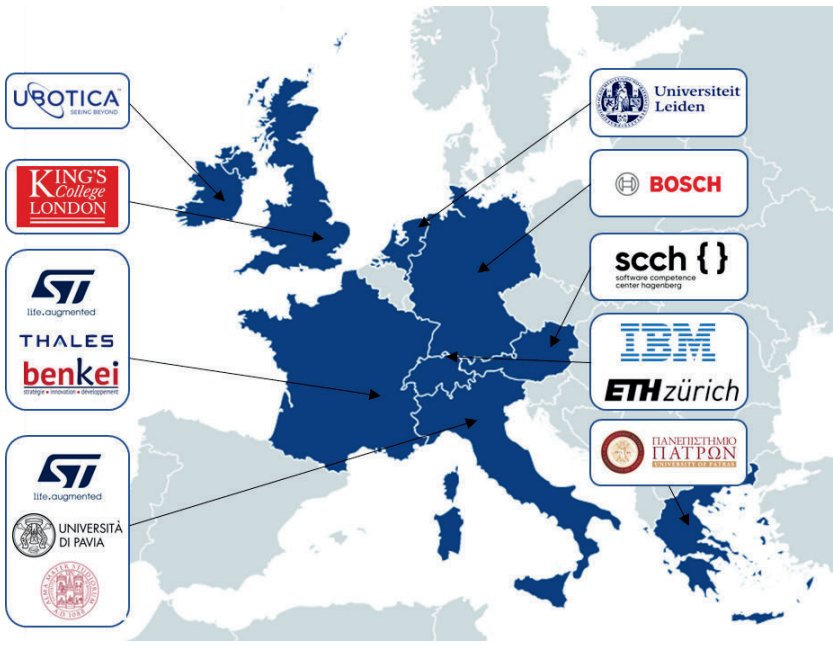
01/09/2022 to 28/02/2026

€ 10.6M

CONTEXT

The **explosive growth of artificial intelligence** and its movement to the edge and end devices have prompted significant research on highly energy efficient and low-latency non-von Neumann computing paradigms such as In-Memory Computing (IMC). Those properties correspond to needs expressed for **industry** (predictive maintenance, robotics), **medical and healthcare**, **mobility** (autonomous driving, vehicle electrification, smart bikes-trains-airplanes), **urban management** (smart cities, IoT nodes), **consumer and personal electronics** (smartphones, watches, glasses, wearables, smart home) and **robotics** (drones, object detection and tracking).

CONSORTIUM



TECHNOLOGY

Despite the prominence of new memory technologies like SRAM and FLASH, **Phase Change Memory** remains a solid and reliable technology, able to achieve mass volume production and cost. NeuroSoC chosen approach is to use an **In-Memory Neural Processing Unit comprising IMC tiles based on ST's highly dense Phase-Change Memory** [1]. This IMNPU adopts the mixed-signal IMC paradigm that IBM recently validated in silicon [2]. Digital computational cores and memory will be integrated to allow for end-to-end inference of industry-relevant models.

TARGETED IMPACTS

The NeuroSoC architecture is expected to enhance the energy efficiency and compute density (TOPS/W/mm^2) by **more than 100x** when compared to conventional general-purpose systems.

NeuroSoC will focus on a **scalable architecture** to target different needs for a variety of products.

In the end, NeuroSoC will deliver a **fully integrated solution with toolchain support** for development and optimization.

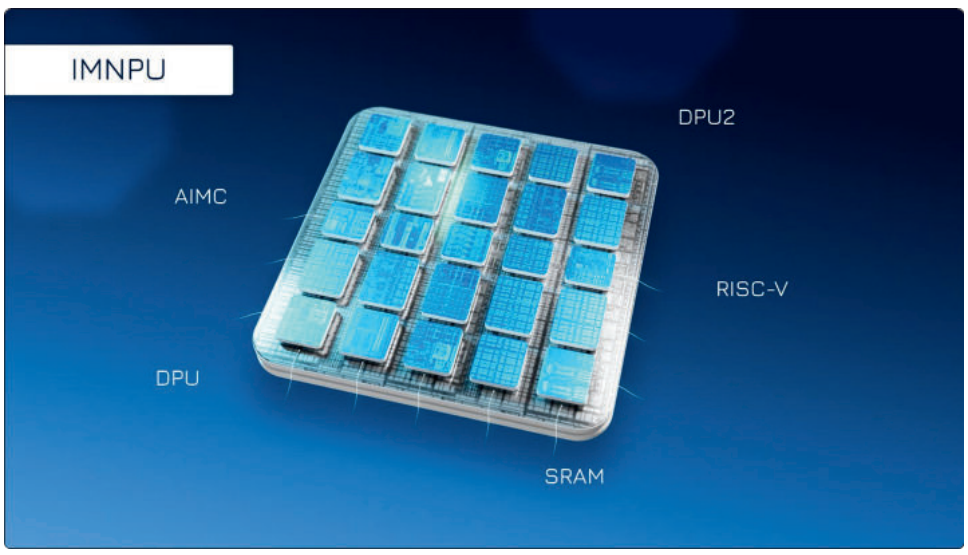
INNOVATION

NeuroSoC consortium is going for a **hybrid NPU approach,** using not only **IMC in PCM tiles** but **combining it with other technologies** (digital processing units, SRAM, RISC-V) to take advantage of the individual characteristics (compute precision, weight storage capabilities, flexibility etc.). The architecture of the IMNPU draws inspiration from research conducted within the IBM Research AI HW Center [2, 3].

The IMNPU will bring forward the following innovations:

- Architecture and design of IMNPU units,
- Tight integration with the processor,
- Software toolchain support,
- AI algorithms to fully exploit the architecture,
- Embedded cryptography.

IN-MEMORY NEURAL PROCESSING UNIT



Funded by the European Union

UK Research and Innovation

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

REFERENCES

[1] F. Arnaud et al., "High Density Embedded PCM Cell in 28nm FDSOI Technology for Automotive Micro-Controller Applications," IEDM 2020

[2] M. Le Gallo et al., "A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference," Nat. Electron. 2023

[3] S. Jain et al., "A Heterogeneous and Programmable Compute-In-Memory Accelerator Architecture for Analog-AI Using Dense 2-D Mesh," TVLSI 2023