This project has received funding from the Electronic Components and Systems for European Leadership Joint
Undertaking under grant agreement No 876925

AI for New Devices And Technologies at the Edge

D4.5 Design, test and evaluation of three digital ANN accelerators

Deliverable No.	D4.5	**Due Date**	*31-Oct-2023*
Type	Report	**Dissemination Level**	*Confidential*
Version	1.0	**Status**	Final
Description	This deliverable provides the details of the design, test and evaluation of three digital ANN Accelerators .		
Work Package	WP4 – Implementation of ICs and Platforms.		

Abstract (Published Summary)

ANDANTE's overarching goal is to leverage innovative hardware accelerators and related platforms for artificial neural networks (ANN) and spiking neural networks (SNN) as a basis for future products in the IoT Edge domain. IoT. The objective of WP4" Implementation of ICs and Platforms" is the development of the neuromorphic ASICs and platforms, based on the use cases and system requirements defined in WP1 "Use Case System Architectures Description and Application Requirements" and implemented in WP5 "Applications Integration, Validation and Evaluation". Moreover, WP4 ASIC developments will profit from the silicon technologies addressed in WP2 "New Memory Technologies for AI applications" as well as the HW building blocks and Foundation IPs defined by WP3 "AI Building Blocks, Methods and Tools".

This deliverable reports on the design and implementation of three digital Artificial Neural Networks '(ANN) accelerators, one ASIC and two SoCs, made in ANDANTE, for Edge AI applications:

- **ASIC 2.1 in 22 FDX designed by CEA**: It is a feature extractor circuit to address image classification, segmentation and detection applications. Three variations of this circuit were designed and fabricated, two of them using SRAM and another using SRAM and NVM memory (OxRAM). The two SRAM circuits were measured and integrated into the 4.1a platform to implement five use cases in different real-time application domains. The measurements obtained show impressive results compared to SoA. For example, HD image inference at 30 FPS only consumes 23.2 mW. Compared to SoA, the circuit is at least 9.2 times better than any research prototyping circuit. And compared to the commercial (COTS) product, it is 1000 times better in terms of energy per frame.
- **SoC 1.1 in 16 nm FinFet designed by ST-GNB**: The design of this circuit is based on the STM32 family of microcontrollers with a Neuro processor unit. Its architectural design was carried out at ANDANTE and its physical design in another research project. An important feature of this SoC is its ability to implement any AI model for Edge AI applications. However, the flexibility is paid for by higher power consumption (around a decade) compared to ASIC. This SoC 1.1 is operational and will be commercialized by the end of 2024. However, it is not used in any ANDANTE use cases because the availability of its first samples would arrive too late considering the schedule of the use cases implementation.
- **SoC 2.1 in 22 FDX designed by CSEM**: Visage 2 is an evolution of their previous circuit, Visage 1, with support for a larger NN model and dynamic bit precision. The HW AI design and accelerators support up to 200 GOPS while operating at 250 MHz. Additionally, this block achieved 10.9 TOPS/W when performing AI tasks with 8b accuracy. It should be noted that its design is not part of the Grand Accord. CSEM used ANDANTE synergy to carry out its design and physical implementation. Tape production is ready to start fabrication in the second quarter of 2024. Thus, only simulation results are reported.