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

ECSEL Joint Undertaking
Electronic Components and Systems for European Leadership

ANDANTE

AI for New Devices And Technologies at the Edge

D3.4 Architecture exploration results

Deliverable No.	D3.3	**Due Date**	*30-Nov-2021*
Type	Report	**Dissemination Level**	*Confidential*
Version	1.0	**Status**	Final
Description	Architecture exploration results		
Work Package	WP3 – AI building blocks, Methods and Tools.		

Abstract (Published Summary)

Abstract (Published Summary)

The overarching goal of ANDANTE is to explore neuromorphic technologies and architectures in the context of a wide range of different use cases. Hence, many different neuromorphic ASICs, FPGAs and platforms are newly developed or existing ones are improved and optimized.

The objective of WP3 “AI Building Blocks, Methods and Tools” is to develop the building blocks, Foundation IPs, tools, algorithms, and architectures required to realize new neuromorphic hardware to be created and/ir improved by ANDANTE consortium. These building blocks will be part of the design of the ASICs and FPGAs, which will be integrated in the platforms to validate their features and performances through 13 application demonstrators developed in WP5.

This deliverable describes the outcomes of Task T3.2 “Architecture exploration”. It explores and specifies various architectures for the planned ASICs, FPGAs, and their associated platforms. Each ASIC/FPGA architecture depends on the target application and the chosen technological platform. Moreover, the performances of the ASIC/FPGA architectures will also depend on the approaches, methods, and tools used for the architecture design exploration, evaluation, and optimization.

In this deliverable, the partners of Task 3.2 present in detail the SNN/ANN architecture exploration for different applications:

SynSense gives an overview on its *exploration of various approaches for examining audio source localization from microphone arrays*. Based on the investigation results, SynSense specified its own approach including a spiking neuronal network for the data processing. This network serves as a base architecture for its spiking neuromorphic network accelerator.

CEA presents the *exploration results of the architecture design space of its convolutional neural network accelerator*, NeuroCorgi, for the efficient processing of image-based data. CEA also describes in detail how its design flow and associated tools developed in Task 3.1 are used to perform such exploration.

FhG presents the design *exploration results for the architecture of the analog neural network accelerator for voice detection*, under development in ANDANTE. The work is focused on investigating the FeFET memory impact on different network architectures. Furthermore, the setup for the architecture exploration as well as some first simulation results are described too.

IFAG presents the *exploration of multiple parallelized and distributed architectural patterns for its spiking neuronal network* hardware for the simultaneous processing of multiple data streams *in applications such as person tracking in big rooms* that is developed in ANDANTE. Furthermore, IFAG developed for the investigation of analog neuronal network (aNN) architectures a specialized tool based on automated machine learning. IFAG also discusses the tool, the investigation process, the process results, and the investigation conclusion of the patterns that will enable the aNN hardware to execute tinyML applications.

GML presents the work on the *architecture and design flow of its GrAI accelerator for object recognition and path planning. Specifically, the exploration of SparNets*, the neuron flow compute model and the compute architecture of the GrAI Core v3 are discussed in detail.

IMEC-NL introduces the *neuromorphic platform ‘SENeCA’ based on a RISC-V inspired processor for acceleration of bio-inspired spiking neuronal networks for extreme edge applications*. With this platform and the simulation tool ‘SEN-Sim’, IMEC-NL explored potential architecture configurations of an PilotNet that was transformed to spiking neuronal network and the impact of different architectures on the performance of SENeCA.

TUD provides details about the *SpiNNaker 2 and e-prop implementation created in ANDANTE for improving embedded learning*. Furthermore, they investigate the e-prop implementation and perform a profiling of the memory usage and compute time in the context of the SpiNNaker 2. The results of the investigation and profiling were used to explore various parallelization strategies. From the exploration insight about improvements, the architecture of the SpiNNaker is derived.

TRT presents the results of a hybrid accelerator having spiking neurons as well as formal digital neurons under development in ANDANTE. In this work, *TRT explored layer types, network configurations, spike coding and crossing location for drones' applications*.