ANDANTE

ASIC 3.1b – Analogue Neural Network ASIC for the Extreme Edge

ANDANTE

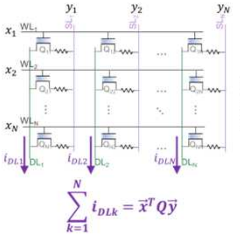
Fraunhofer IPMS

HEIMANN Sensor HEIMANN SENSOR GMBH

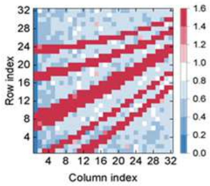
Overview

- Fraunhofer IPMS is exploring FeFET based Compute-In-memory accelerators
- For the exploration Fraunhofer IPMS developed the mixed-signal neural network ASIC 3.1b based on 28SLPe + technology with a FeFET memory array at its core
- Accelerator connected RISC-V coprocessor via DMA for flexibility
- An adapter board was developed to characterize the chip on wafer-level
- UC1.1: People Counting using infrared thermopile arrays to detect radiation pattern of persons

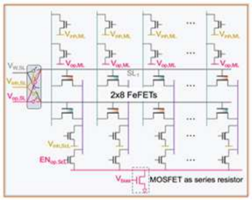
FeFET Compute-In-Memory



$$\sum_{k=1}^{N} i_{DLk} = \vec{x}^T Q \vec{y}$$



Pattern programming

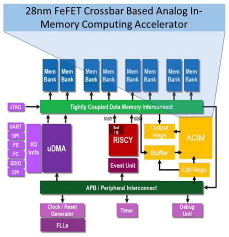


Segment structure

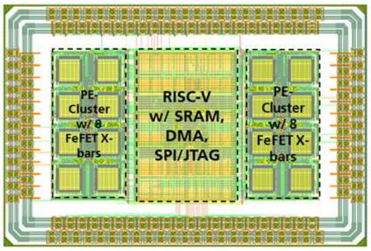
- FeFET based 1F1R concept implemented in low-mismatch, low-leakage segment structure
- MAC operation by Kirchhoff's law of current accumulation
- Passive array programming schemes implemented
- Pattern structures to be loaded
- Retention $>10^8$ s and endurance $>10^5$ tested

Architecture

- ASIC 3.1b: ANN ASIC for the extreme edge
- Integrated with a RISC-V for adaptability and controllability
- Mapping of various networks supported
- Flexible quantization of weights and activations
- I/O interfaces: SPI, UART, I2C, I2S
- Weight Memory: FeFET
- Instruction Memory: SRAM
- Accelerator controlled with custom DMA module
- Hierarchical accelerator design for scalability



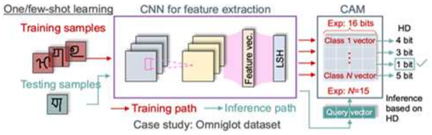
28nm FeFET Crossbar Based Analog In-Memory Computing Accelerator



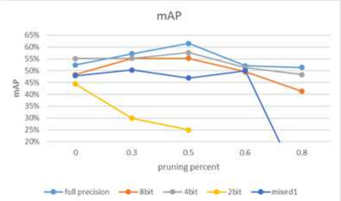
PE-Cluster w/ 8 FeFET X-bars
RISC-V w/ SRAM, DMA, SPI/JTAG
PE-Cluster w/ 8 FeFET X-bars

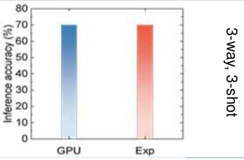
Results

KPI Name	Target	Simulated
Peak Throughput (1b x 1b)	N/A	452 GOPS/s
UC Accuracy	85 %	61.5 mAP
UC Inference Time	< 15 ms	10 ms
UC Power (w/o) IO	< 10 mW	~1 mW
UC Power Eff. (1b x 1b)	N/A	20 TOPS/W

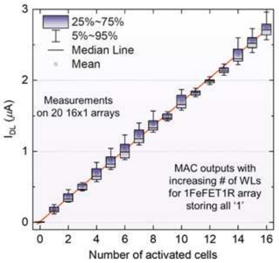


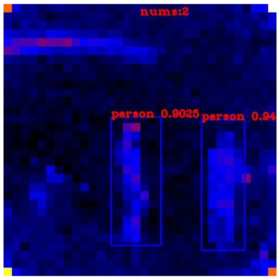
Memory augmented neural network (MANN) few-shot learning reference case





Validated MAC outputs





Impact

- Feasibility of concepts proved
- Next steps for commercialization:
  - Optimize ASIC implementation and transfer to smaller technology nodes
  - Looking into efficient integration of the concepts into sensor node circuits

Progress beyond SoA

- One of the first moves towards using FeFET for compute-in-memory
- Largest array demonstration
- Researching FeFET memory in the context of analogue NN and tinyML applications

Lessons learned

- Trade-off between flexibility regarding possible applications and resource limitation is challenging
- HW-SW co-design necessary for high performing designs

ANDANTE

AI for New Devices And Technologies at the Edge

andante-ai.edu   linkedin.com/company/andante-ai   mario.diaznava@st.com (coordinator)

ECSEL Joint Undertaking

"This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement № 876925. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Belgium, Germany, Netherlands, Portugal, Spain and Switzerland"

Scan Me
to visit website