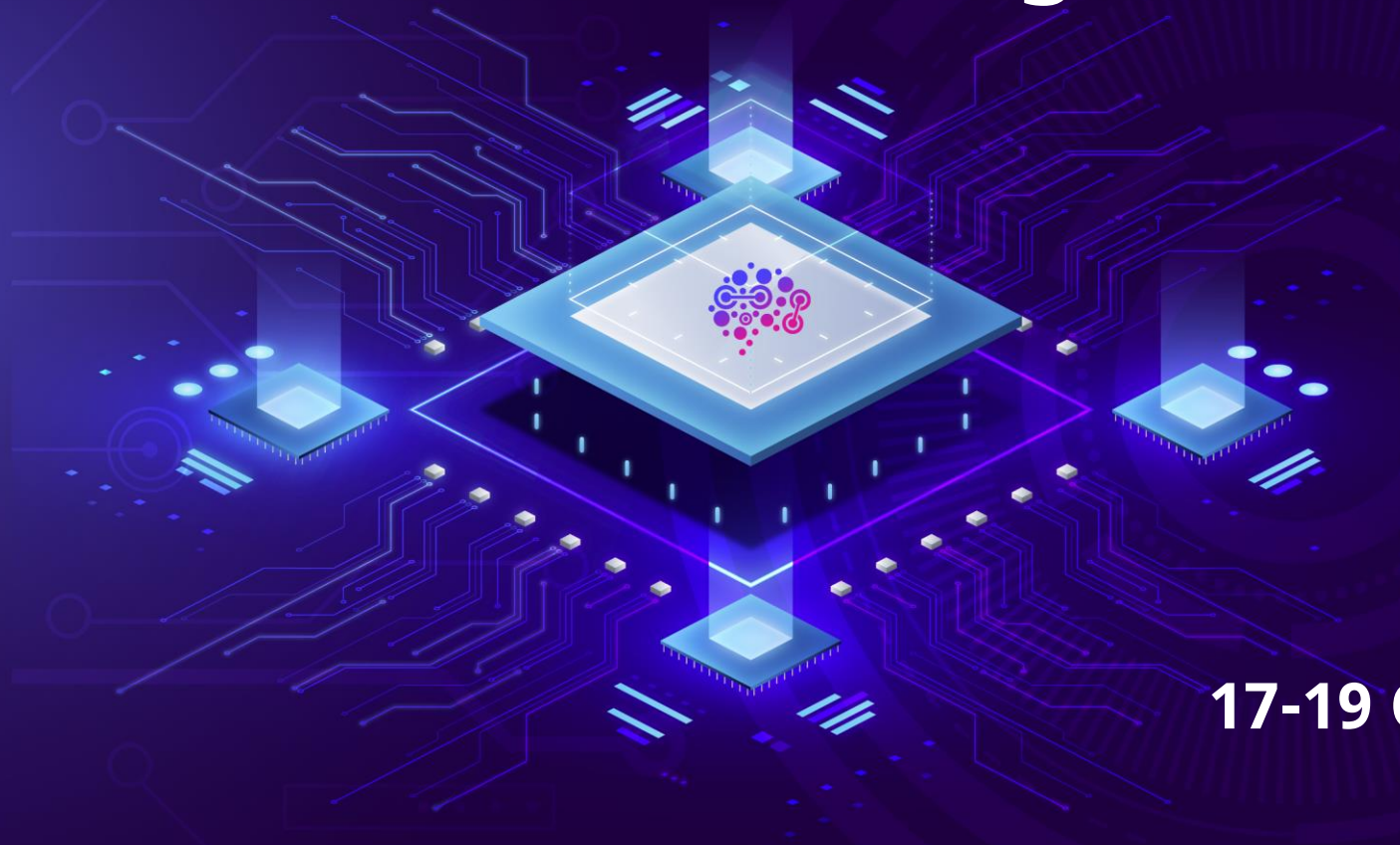
European Conference on EDGE AI Technologies and Applications - EEAI



Advancing emerging edge AI technologies and driving next generation intelligent applications.

17-19 October 2023, Athens, Greece

European Conference on EDGE AI Technologies and Applications - EEAI

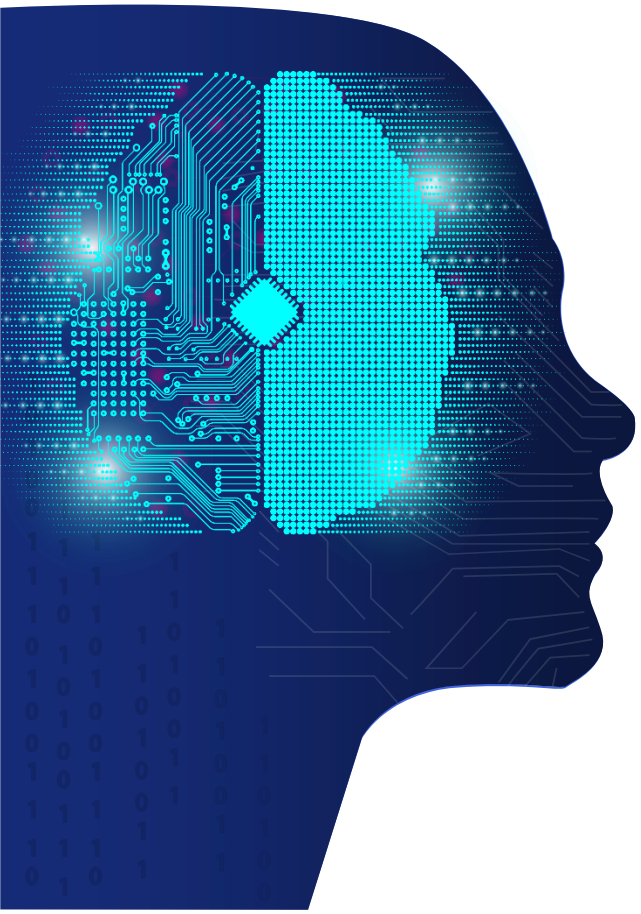
A FeFET-based non-volatile memory and AI accelerator macro

Marko Noack, Ferroelectric Memory GmbH, Germany

17-19 October 2023 Athens, Greece

Copyright © 2023

Presentation Outline



- Introduction
- Goal
- Methodology
- Design and Implementation
- Results
- Future Research
- Conclusion

Copyright © 2023

Introduction

FMC The Ferroelectric Memory Company

- Edge AI offers
  - Reduced latency
  - Independency of network connections
  - Scalability
  - Enhanced Security
- But we need solutions with
  - Low power
  - Low cost
  - Non-volatile storage of network weights
- Our goal
  - Demonstrate viability of FeFET memory in Edge AI systems

Copyright © 2023

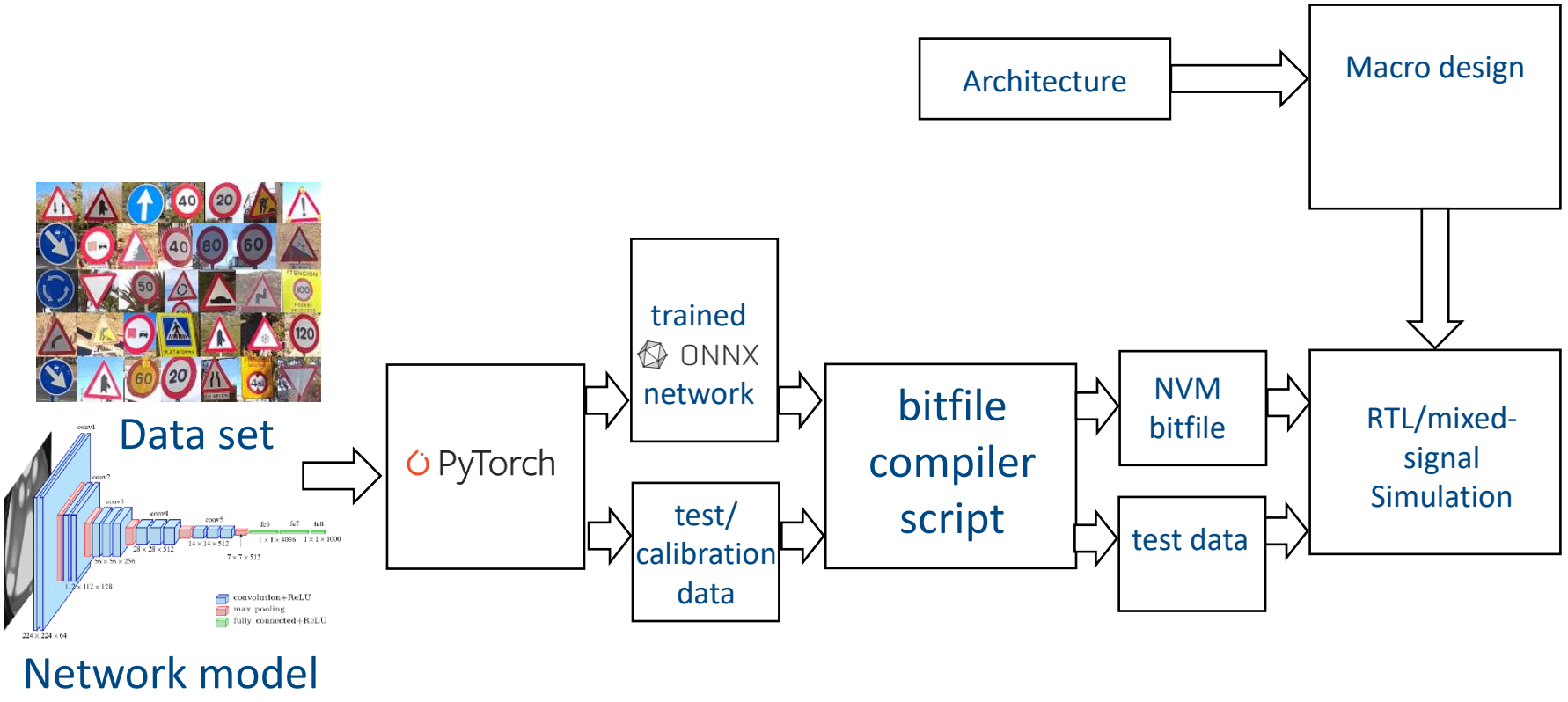
Goal

- Development of a FeFET-based NVM macro
  - Capable of data storage, but also with
  - Built-in accelerator circuit for CNNs
  - Maximize speed and utilization
  - Maintain Flexibility for different kinds of network models

Copyright © 2023

Methodology

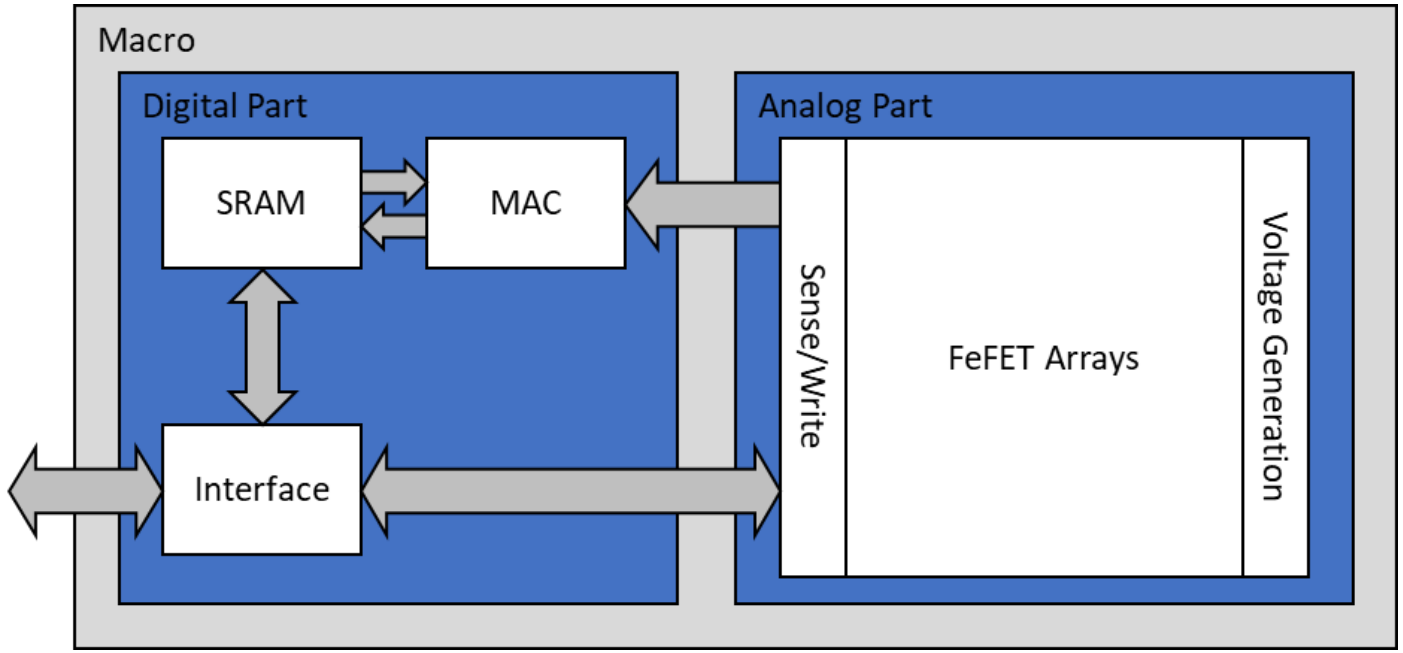
- Implement macro design
- Train network model with data set
- Generate NVM bitfile and testdata from trained network and test data



Copyright © 2023

Design and Implementation

- Design split in full-custom analog and RTL-based digital part
- Analog part contains:
  - memory arrays
  - charge pumps
  - Sense amps
- Digital part contains
  - Interface
  - SRAM to store IFMs/OFMs
  - MAC unit
  - State machines

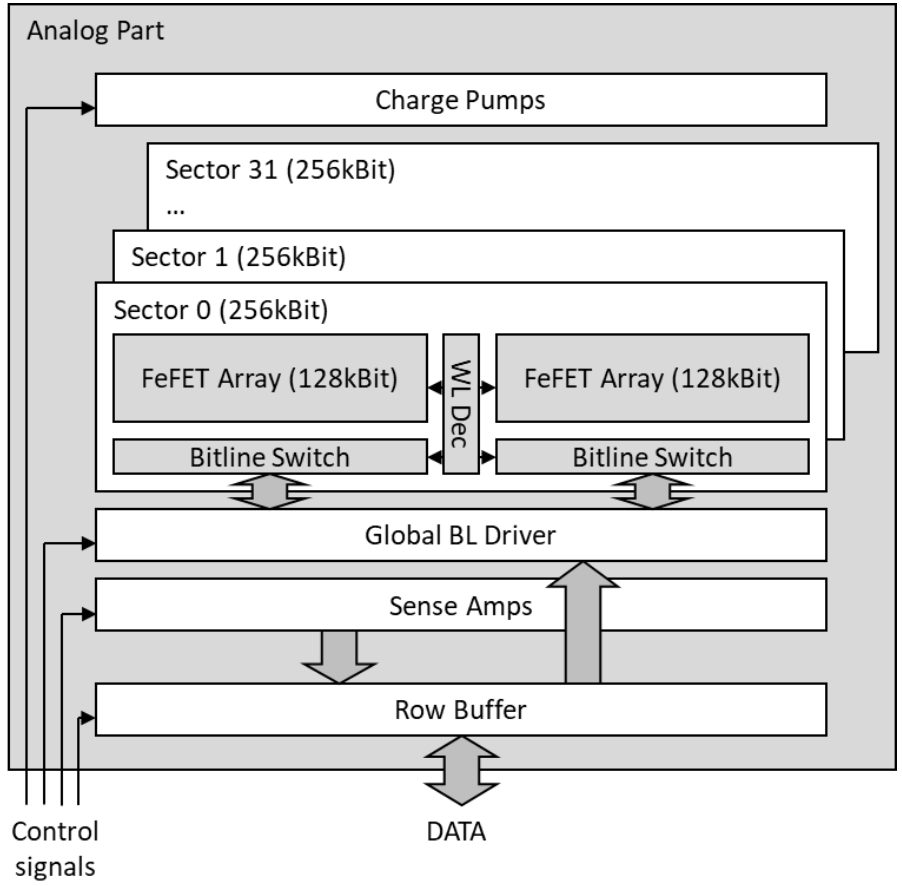


Copyright © 2023

FMC The Ferroelectric Memory Company

Analog Part

- FeFET Arrays
  - 8 Mbit organized in 32 sectors
  - Stores commands, configuration settings and kernel parameters
  - 1024 bits read at a time and multiplexed out with row buffer while next sensing operation can already be performed



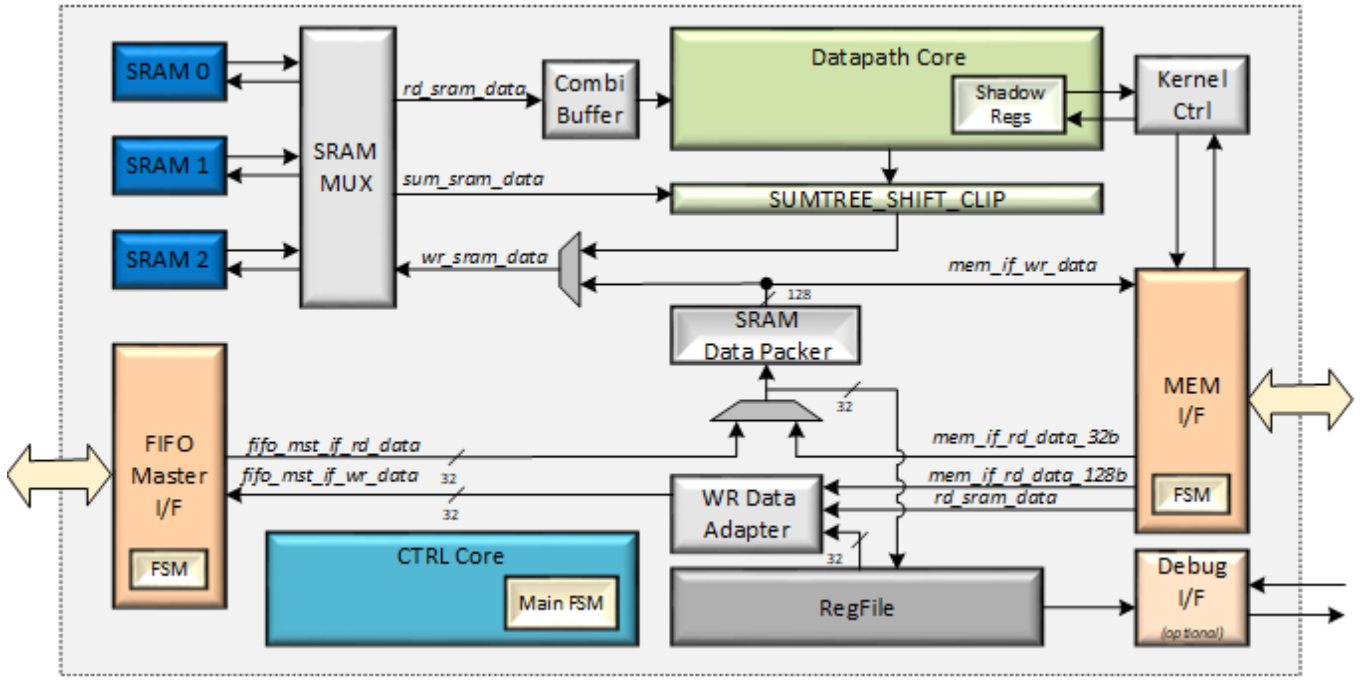
Copyright © 2023

Design and Implementation

FMC The Ferroelectric Memory Company

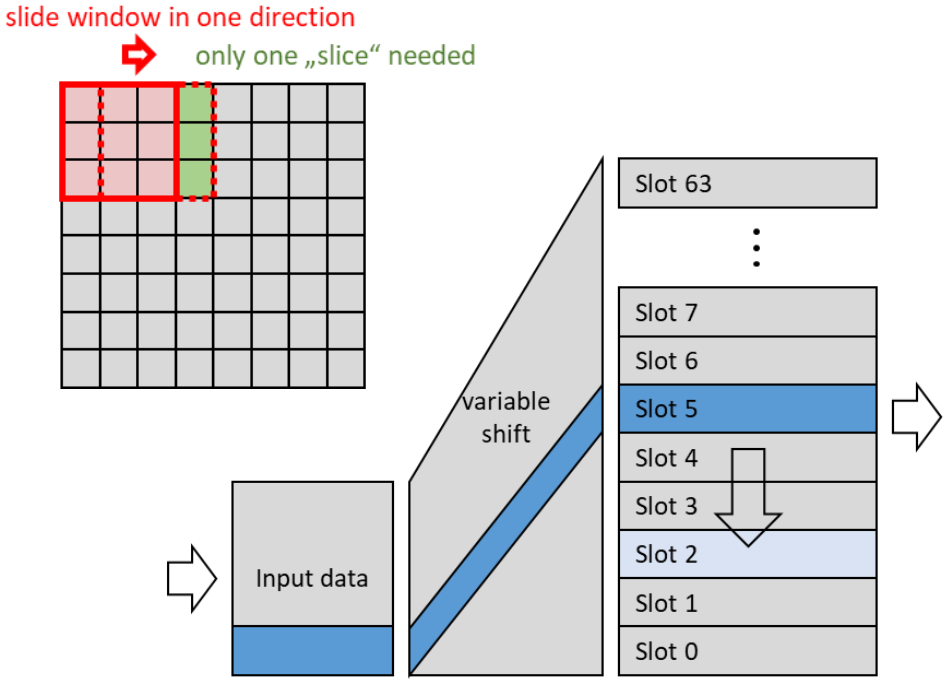
Digital Part

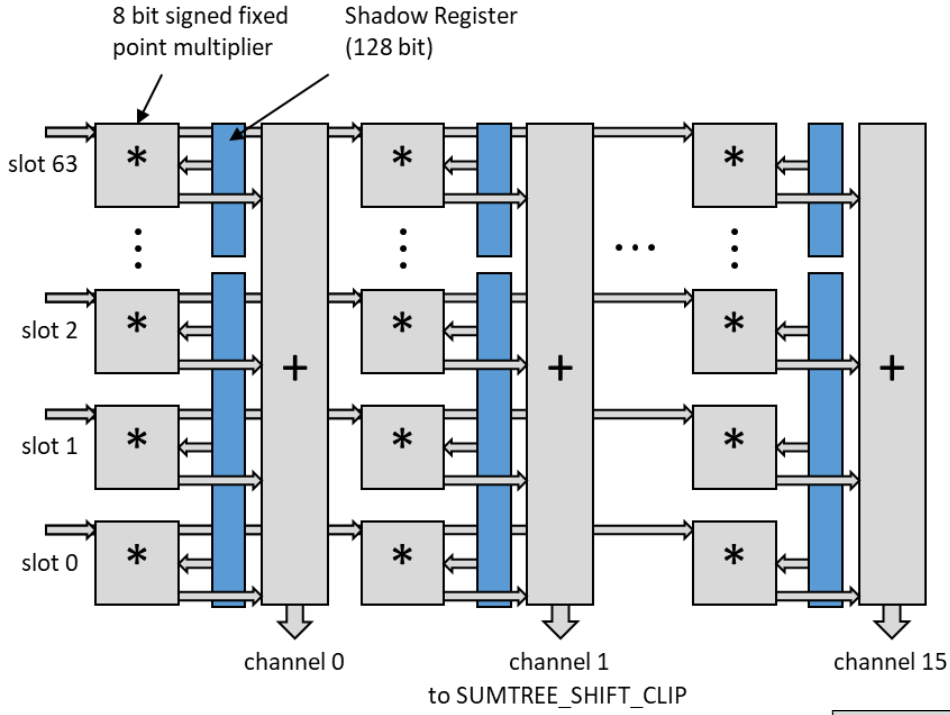
- Input feature map is streamed in to one of the SRAMs via FIFO interface
- Convolution/MaxPool
- Operations performed layer by layer
- Layers can be split into parts, partial sums can be added through second SRAM
- Output feature map stored in SRAM → streamed out via FIFO interface

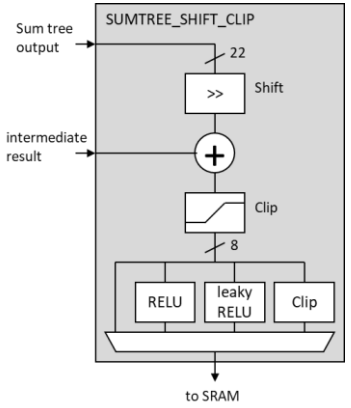


Copyright © 2023

Design and Implementation







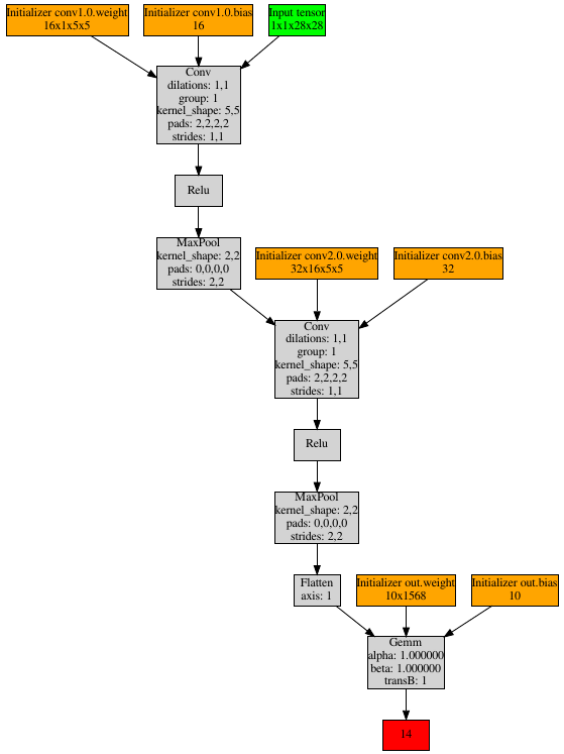
- For convolution operation a sliding window over the IFM is generated
- Only one input „slice“ is needed per clock cycle
- Input buffer arranges IFM data and sends it to multiplier array
- IFM data shared by row
- Each multiplier stores kernel weight in shadow register
- 16 colums/channels generates OFM data which fits again in SRAM width
- Clip/Relu function can be applied to OFM directly

Copyright © 2023

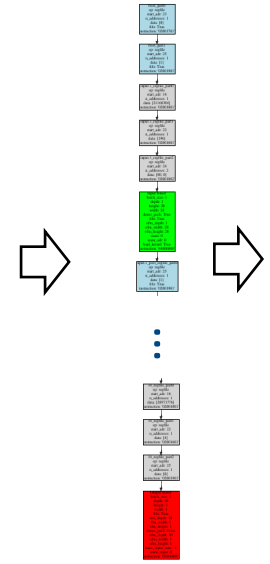
Simulation Results

- 100ns FeFET read spead (1024 bits)
- Multiplexed to 16 x 128-bit words @ 100MHz
- Up to 0.1 TOPs (signed 8-bit fixed-point multiplications)

MNIST used as input and training data set



network graph



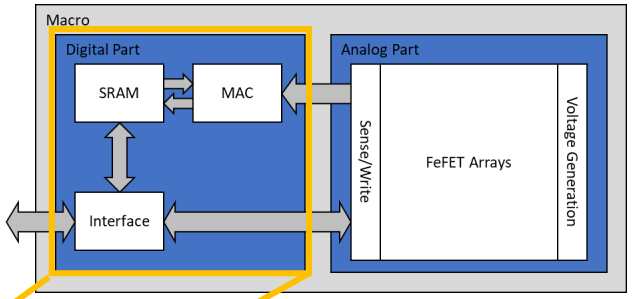
translated to list of operations

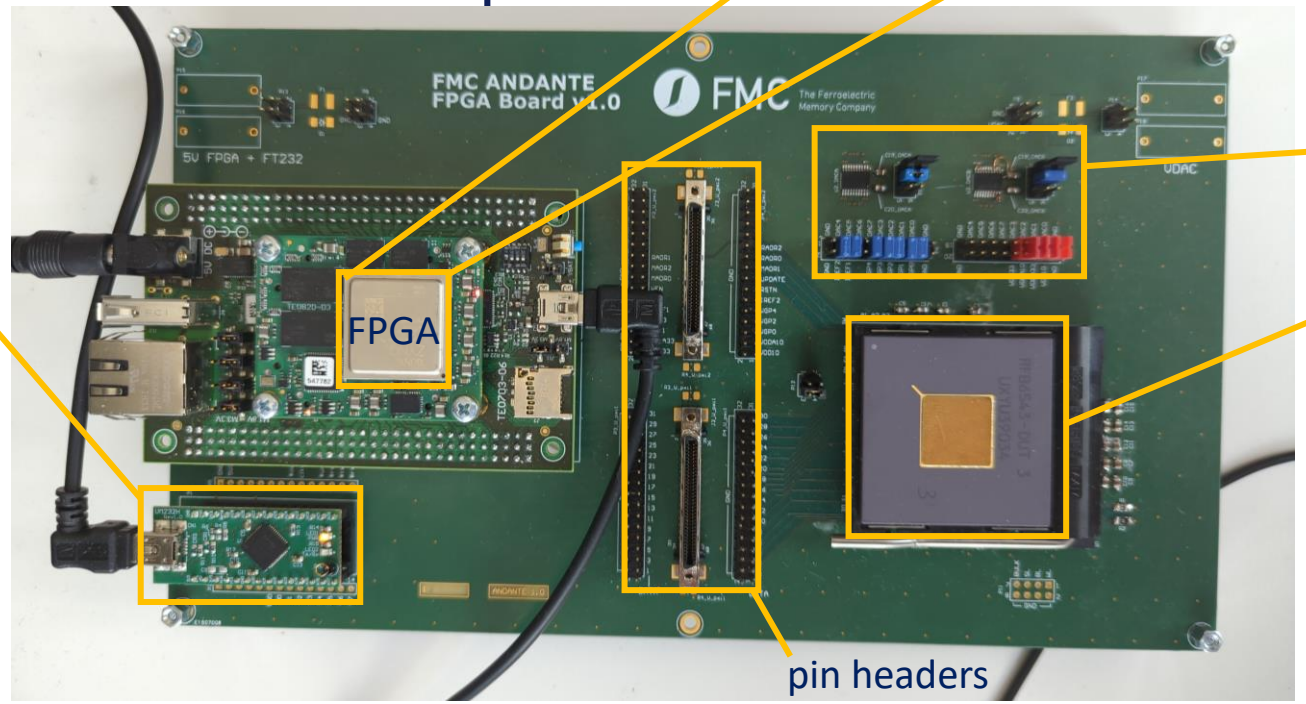
Simulation verified correct classification results

Copyright © 2023

Experimental Results

- Digital part has been implemented on a Xilinx FPGA for testing
- FeFET memory data comes from an external test chip





FTDI USB to FIFO bridge

FeFET chip power supply

FeFET test chip

pin headers

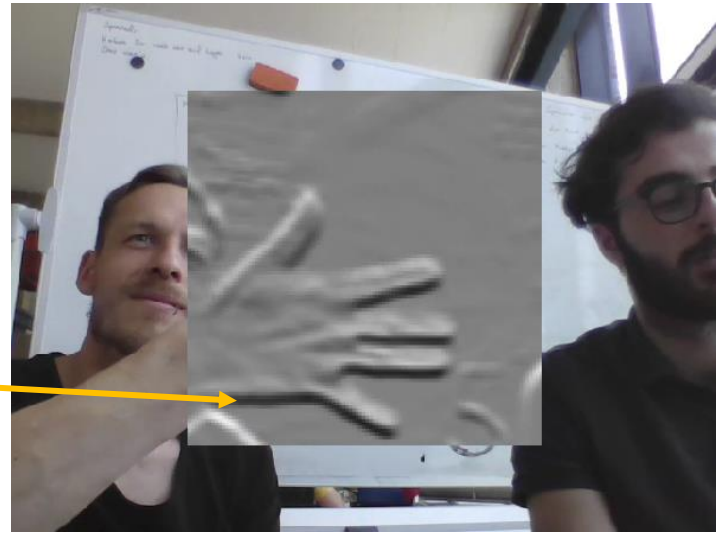
Copyright © 2023

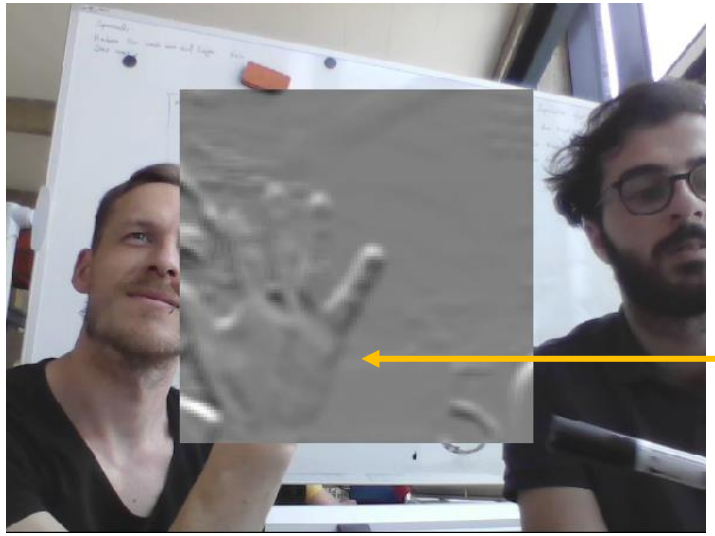
Experimental Results

FMC The Ferroelectric Memory Company

- Limitation in FPGA setup allowed for small network only
- Example: Sobel edge detection filter to detect horizontal features

strong horizontal features





no horizontal features

Copyright © 2023

Future Research

FMC The Ferroelectric Memory Company

- Looking for future cooperations
- Explore possible optimizations

Copyright © 2023

Conclusion

FMC The Ferroelectric Memory Company

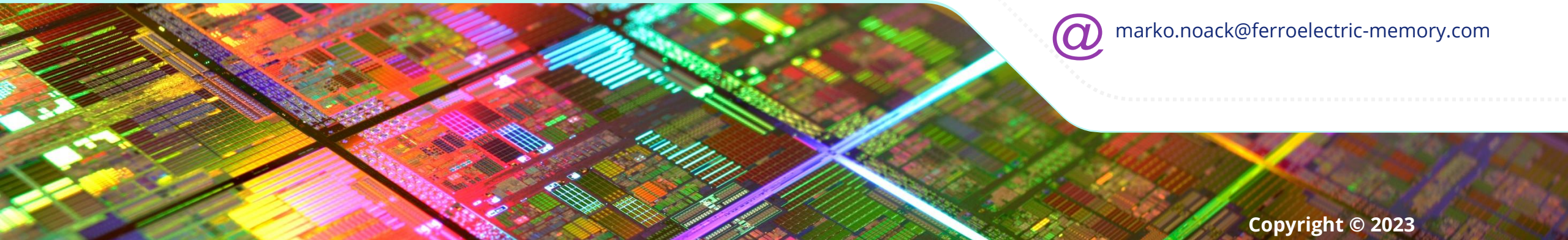
- FeFET-based NVM very suitable for Edge AI applications
- Low cost (low mask count, easy CMOS integration)
- With the presented architecture, high throughputs can be achieved while maintaining high flexibility

Copyright © 2023

EDGE AI

Thank You

For your attention



@ marko.noack@ferroelectric-memory.com

Copyright © 2023