Prospects of the Nonvolatile FPGA and Its application to Edge-AI Accelerators

Daisuke Suzuki

Adaptive Systems Laboratory
E-mail: daisuke@u-aizu.ac.jp

Short Biography

Japanese Name: 鈴木　大輔

English Name: Daisuke Suzuki

Birthplace: Koriyama city, Fukushima

Ph. D: Engineering in Tohoku Univ.

Academic society: IEEE, IEICE, IPSJ

Research interests:
Nonvolatile logic circuit, nonvolatile FPGA, and their application to AI accelerators

Working experience: (Actually, I was at Hanyu & Natsui laboratory, Tohoku Univ.)

Research Associate - Center for Spintronics Integrated Systems, Tohoku University (2010 - 2014).

Assistant Professor - Center for Innovative Integrated Electronic Systems, Tohoku University (2014-2015).

Assistant Professor - Frontier Research Institute for Interdisciplinary Sciences, Tohoku University (2015-2020).

Associate Professor - Computer Engineering Division, the University of Aizu (2020-).

Contents

1. Introduction

2. NV-FPGA and NV-LUT Circuit

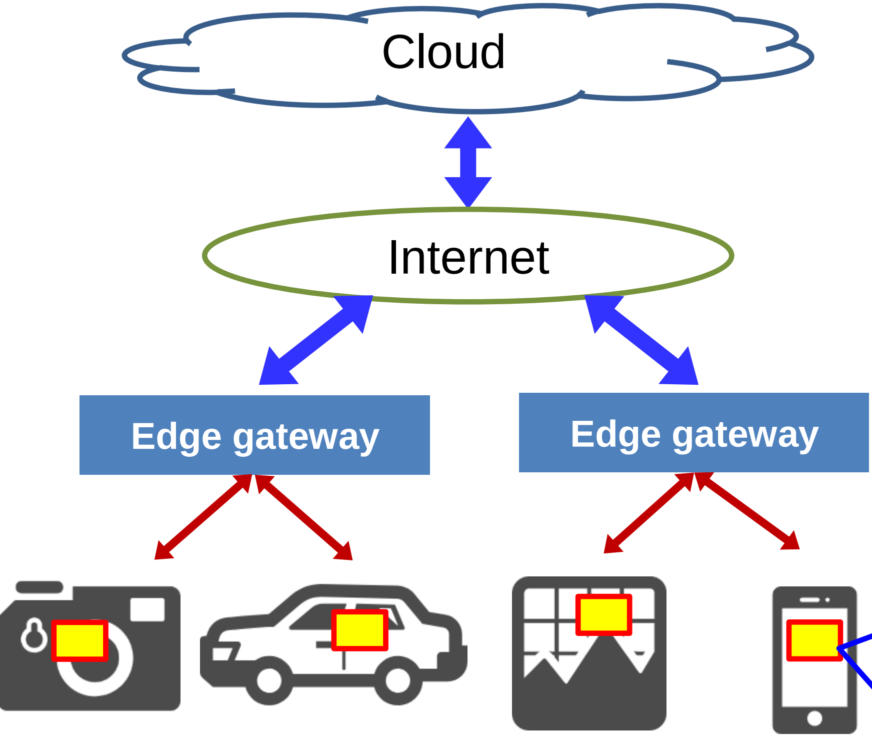
3. Research Plan at UoA

4. Conclusion

Background

■ Internet of Things (IoT)

Applications: Healthcare, sports, agriculture, smart house/city, automated driving, etc…)



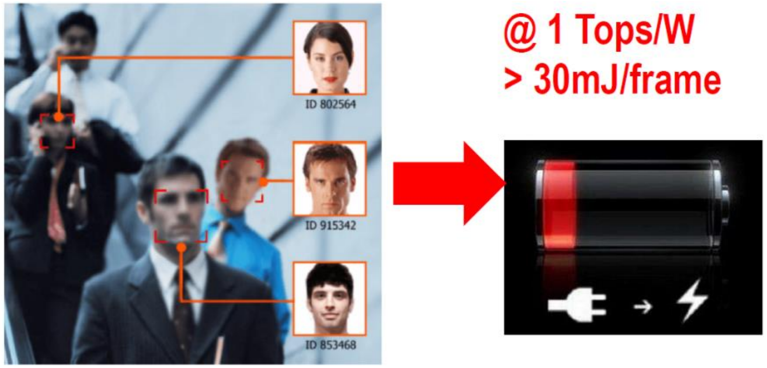
- >90B connections in 2025 [1]
- > $2.7 trillion economic impact [2]

[1] https://www.idcjapan.co.jp/Press/Current/20180813Apr.html

[2] M. Mohammadi, et al., IEEE Communication Surveys & Tutorials, vol. 20, no. 4, pp. 2923-2960, 2018.

- Not cloud-centric data processing, but distributed data processing
- Recognition (Image, sound, etc.)

Artificial intelligence (AI) on edge



@ 1 Tops/W
> 30mJ/frame

Image recognition

*B.Moons et al. ISSCC 2017

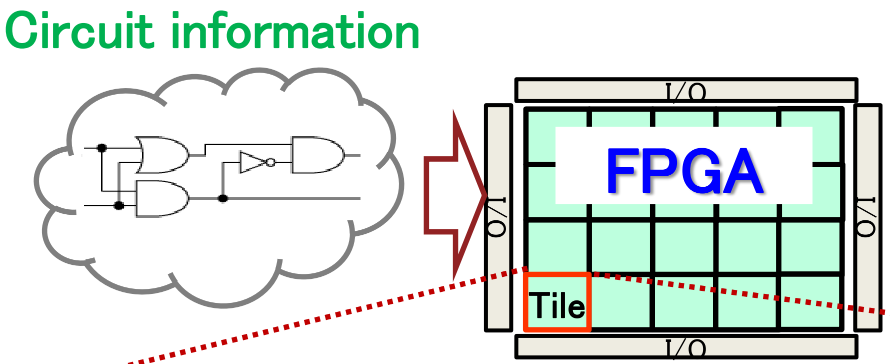
1200mAh-1.5V Battery
< 2 Hour Operation*

[3] J. Yang et al., ISSCC 2019

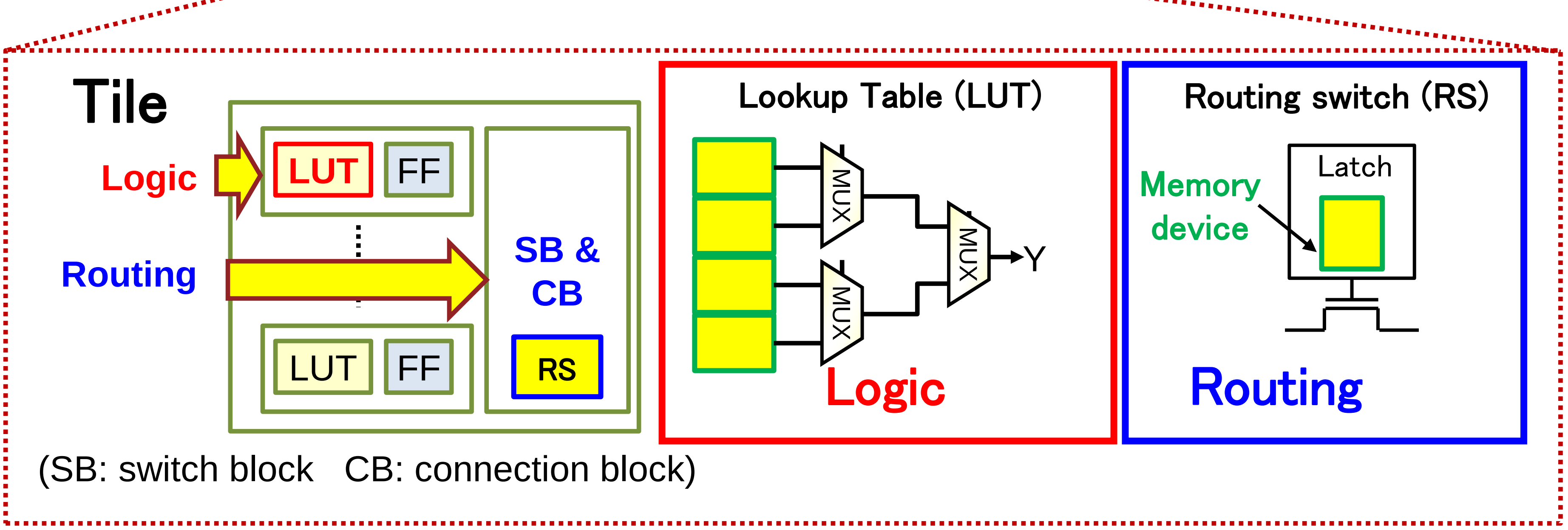
☹ Strongly limited power supply

Low-power, energy-efficient edge-AI hardware is required.

Field-Programmable Gate Array (FPGA)



Users can implement various digital circuits

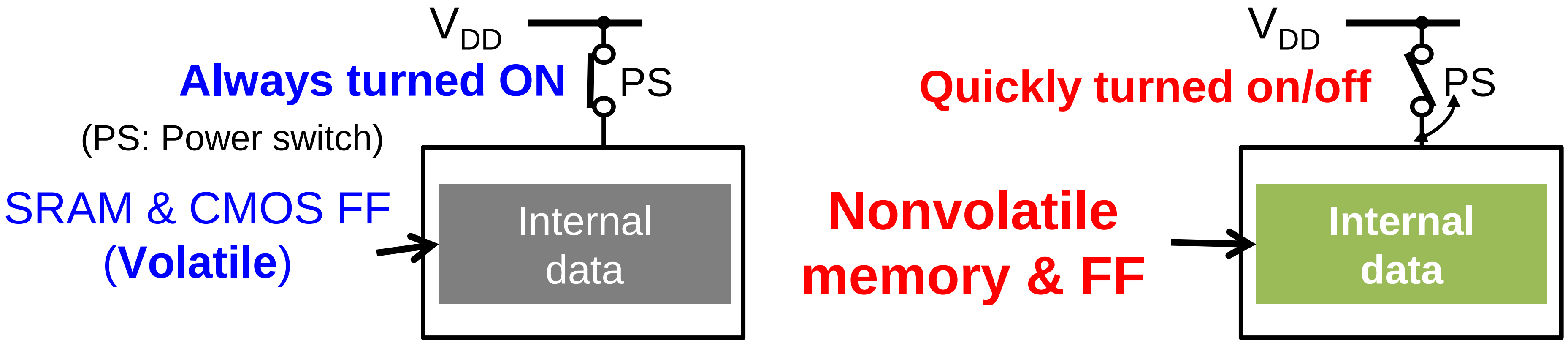


(SB: switch block CB: connection block)

Merit: Short design time, flexibility, low design cost

Demerit: Large amount of standby power consumption

Nonvolatile FPGA (NV-FPGA)



Always turned ON

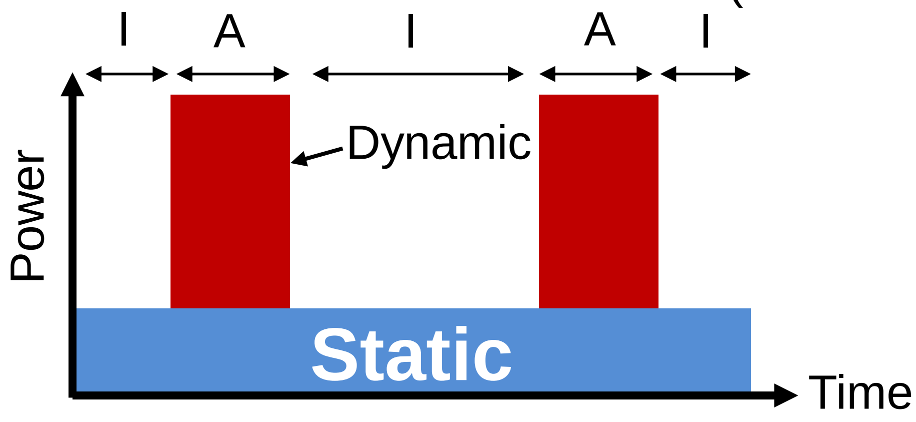
(PS: Power switch)

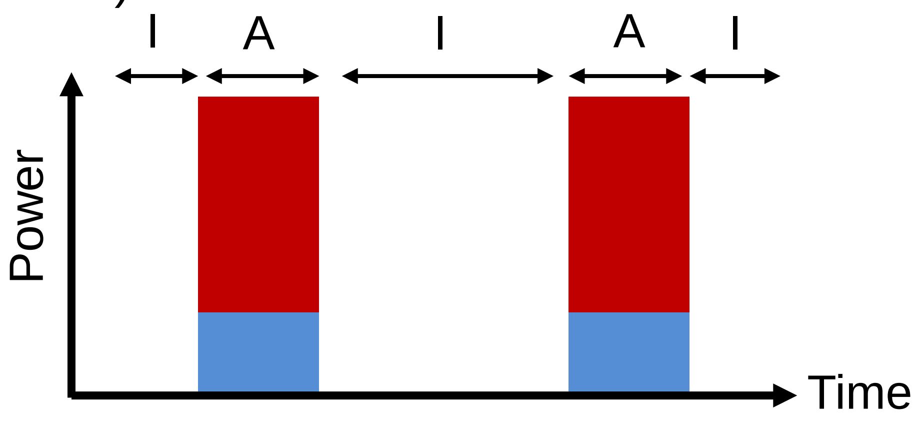
SRAM & CMOS FF
(Volatile)

Quickly turned on/off

Nonvolatile memory & FF

(A: Active I: Idle)





CMOS-only FPGA(volatile)

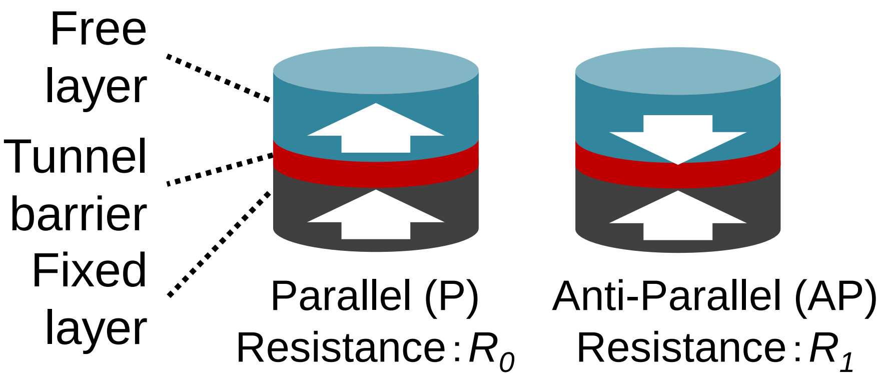
☹ Always on to keep data
→ High standby power

Nonvolatile FPGA

☺ Instant on/off w/o losing data
→ Ultra-low standby power

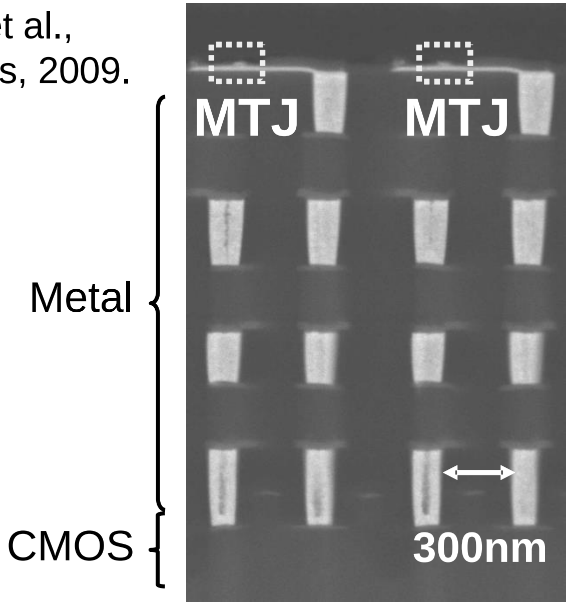
NV-FPGA -> Suitable for IoT/AI device

Magnetic Tunnel Junction(MTJ) Device

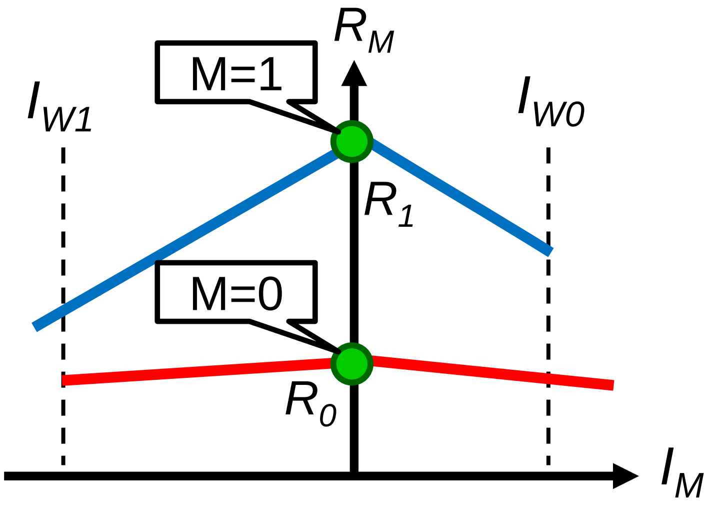


Variable resistor

D. Suzuki, et al., VLSI Circuits, 2009.



Cross-sectional SEM image



R-I Characteristic

One possible candidate for nonvolatile storage element

Summary of Research Roadmap

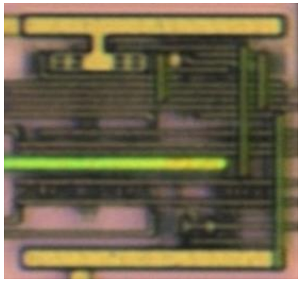
2009 2010 2014 2019 2020 Year

FIRST Program
(Prof. Ohno, Tohoku Univ.)

ImPACT Program
(Prof. Sahashi, Tohoku Univ.)

[Circuit level]

2-input NV-LUT circuit chip
VLSIC 2009



Variation resilient 6-input NV-LUT circuit
MMM2011/JAP2012

Differential-pair-based NV-FF
SSDM2011/JJAP2012

Compact NV-LUT circuit
SSDM2012/ JJAP2013

Compact multi-context NV-LUT circuit
SSDM2014/ JJAP2015

Low-power shift function for NV-LUT circuit
SSDM2017/ JJAP2018

Compact fracturable NV-LUT circuit
SSDM2018/ JJAP 2019

Data shifting in BCNN accelerator
SSDM2020

Compact NV-FF
ELEX2014

Low-power NV-FF
MMM2014/ JAP2015

Compact/ low-power NV-FF
SSDM2015/ JJAP2016

Shift controller for NV-LUT circuit
SSDM2019/ JJAP2020

Compact routing switch
MMM 2013/ JAP2014

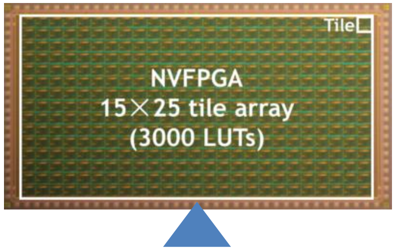
Low-energy MTJ write circuit
Intermag2014/ TMAG2014

[Architecture level]

1st NV-FPGA test chip

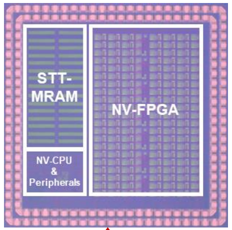
ELEX 2013

2nd NV-FPGA test chip



VLSIC 2015

NV-FPGA embedded NV-MCU



ISSCC 2019

Design automation of NV-FPGA
ISMVL2020 (TBP)

Contents

1. Introduction

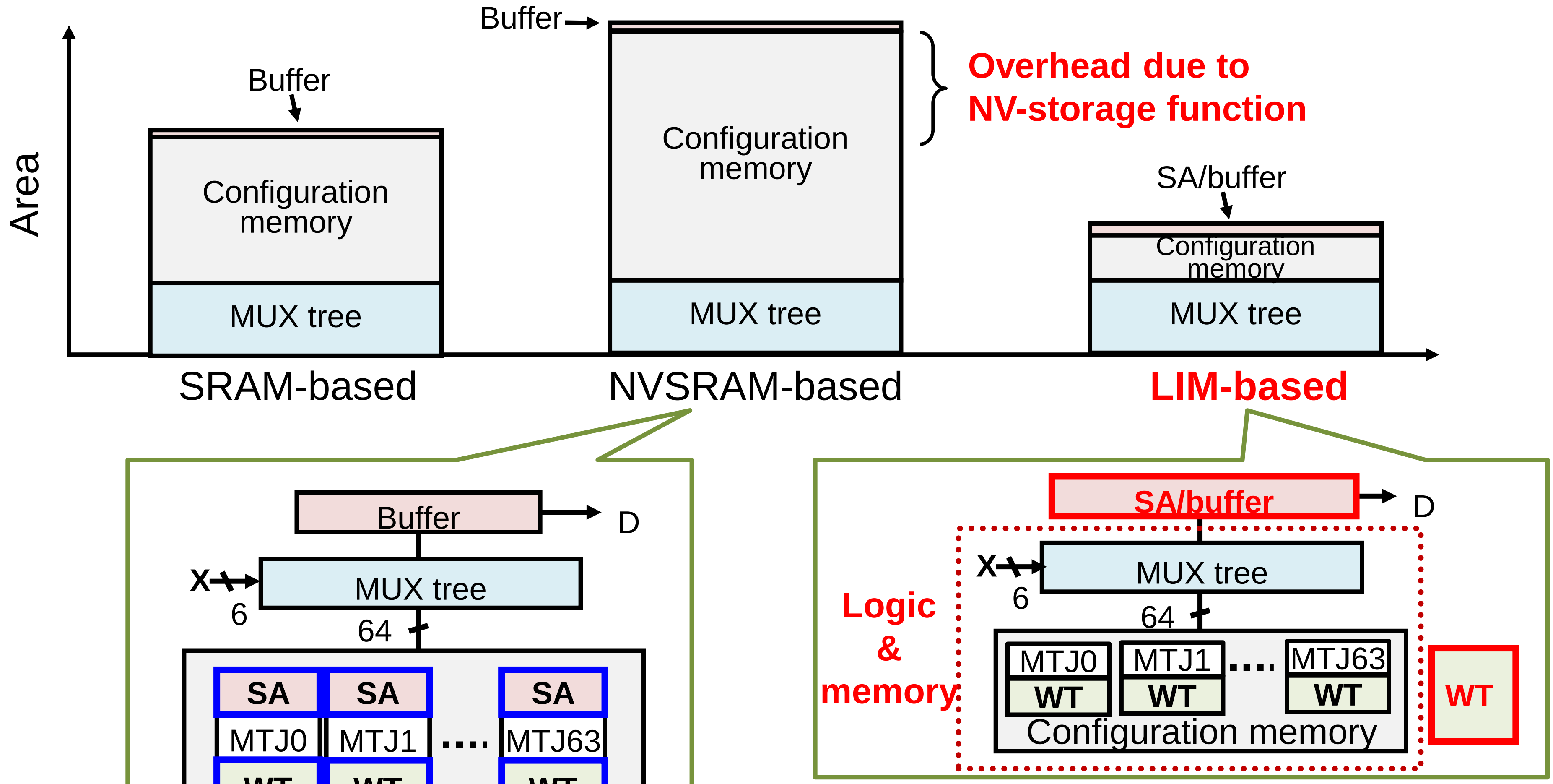
2. NV-FPGA and NV-LUT Circuit

- LIM-based LUT Circuit
- Only-Once-Write Shifting
- NV-FPGA-Embedded MCU

3. Research Plan at UoA

4. Conclusion

Logic-In-Memory (LIM) Structure



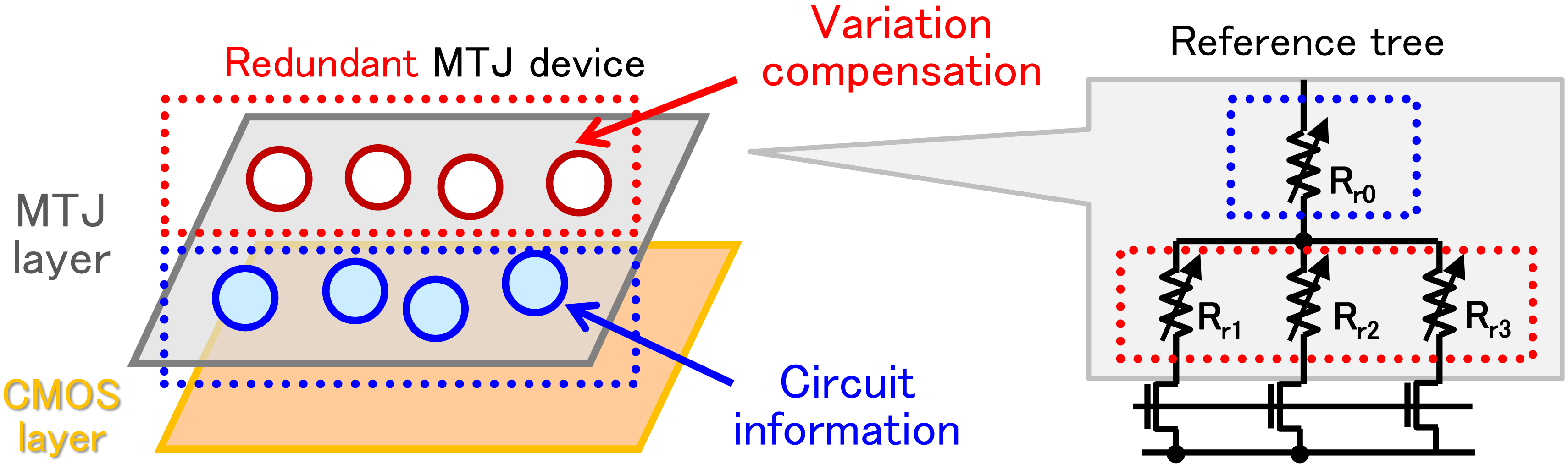
☺ Logic and memory functions are compactly merged.

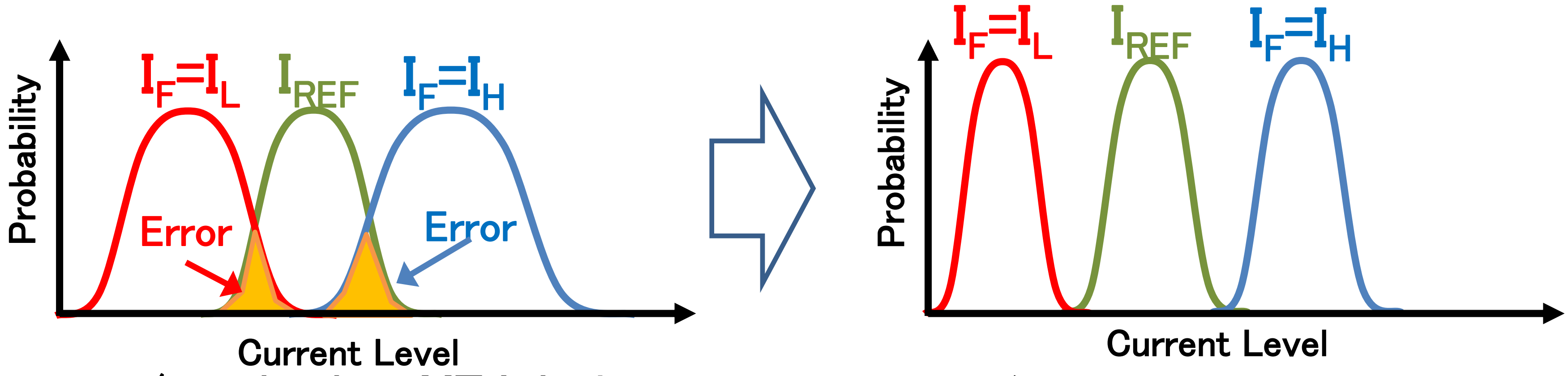
SA: Sense amplifier WT: Write transistor

Compact circuitry by sharing circuit components.

Use of Redundant MTJ Devices

☹ Process variation affects current levels I_L (logic 0), I_H (logic 1), and I_{REF} (reference current).





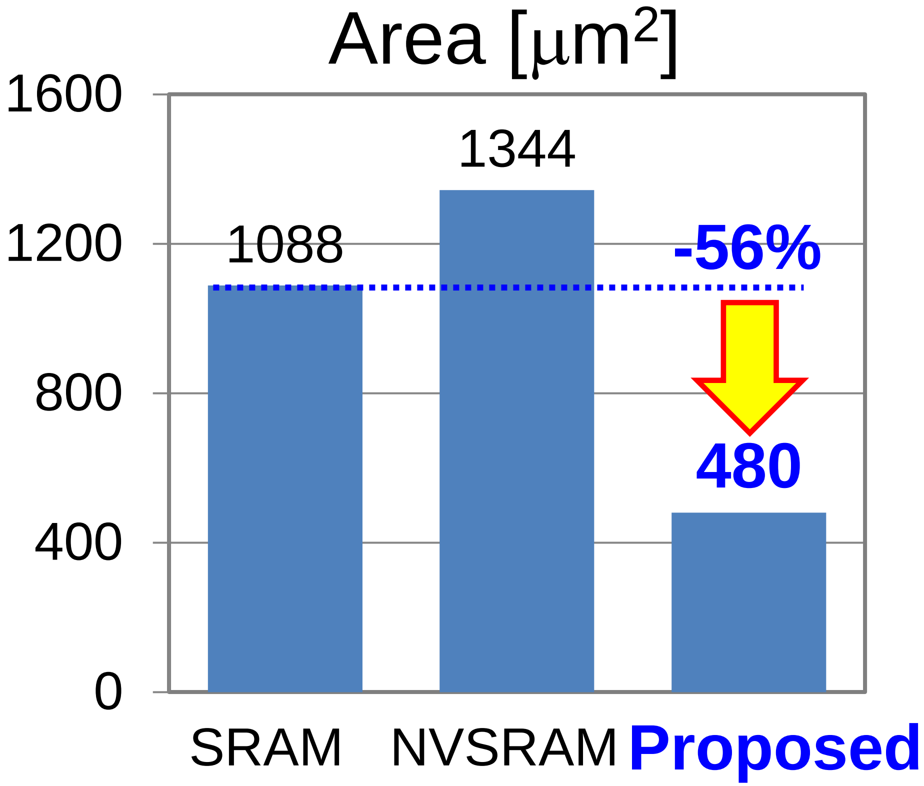
w/o redundant MTJ device

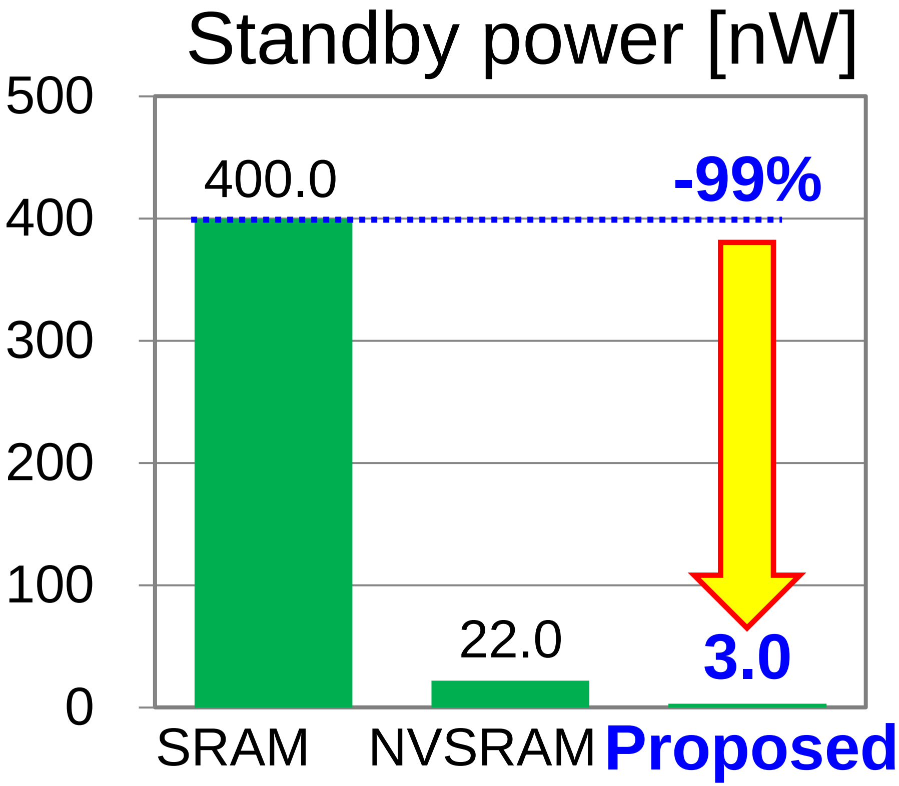
w/ redundant MTJ device

Variation resilient circuitry without area overhead

Performance Comparison of 6-input LUT Circuit

D. Suzuki, et al., VLSI Circuits, 2015.





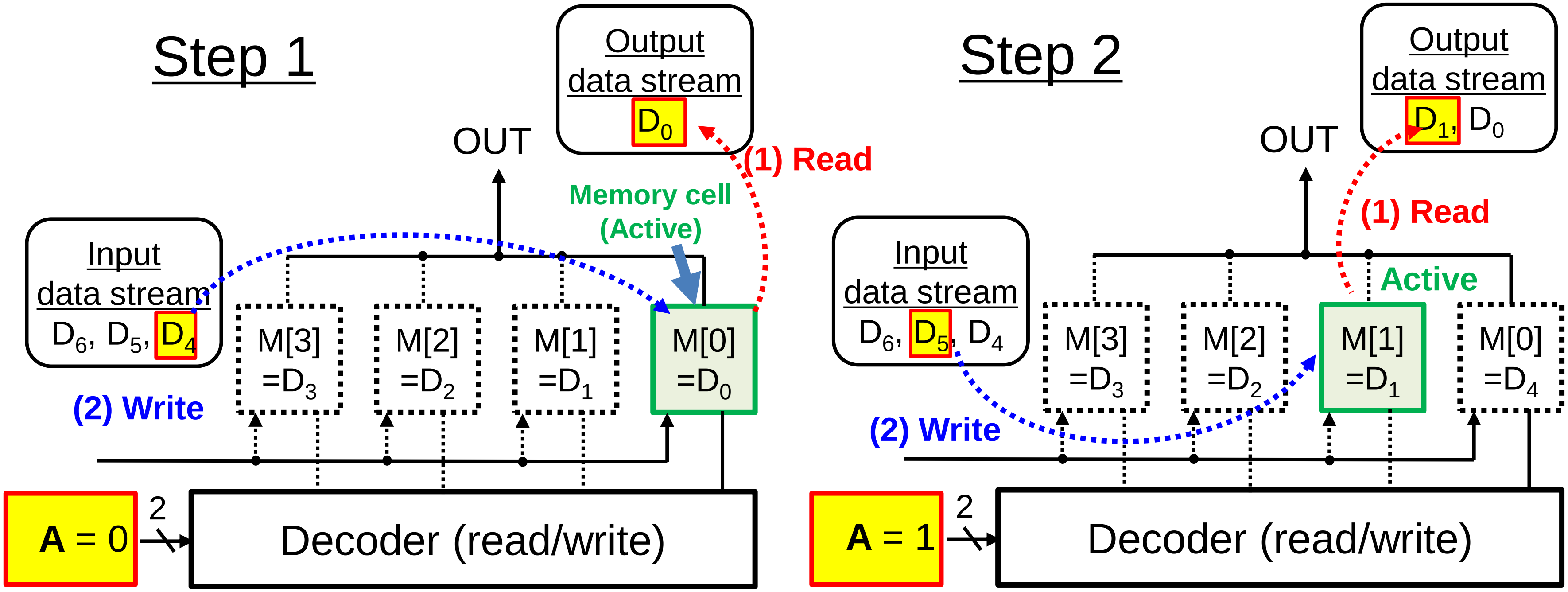
(NVSRAM: Nonvolatile SRAM)

Area and standby power reduction by LIM structure.

Only-Once-Write Shifting [1]

[1] D. Suzuki et al., Jpn. J. Appl. Phys., **57,** 04FE09 (2018).

Data-shit function -> Key function of the **LUT circuit**



(A: Address)

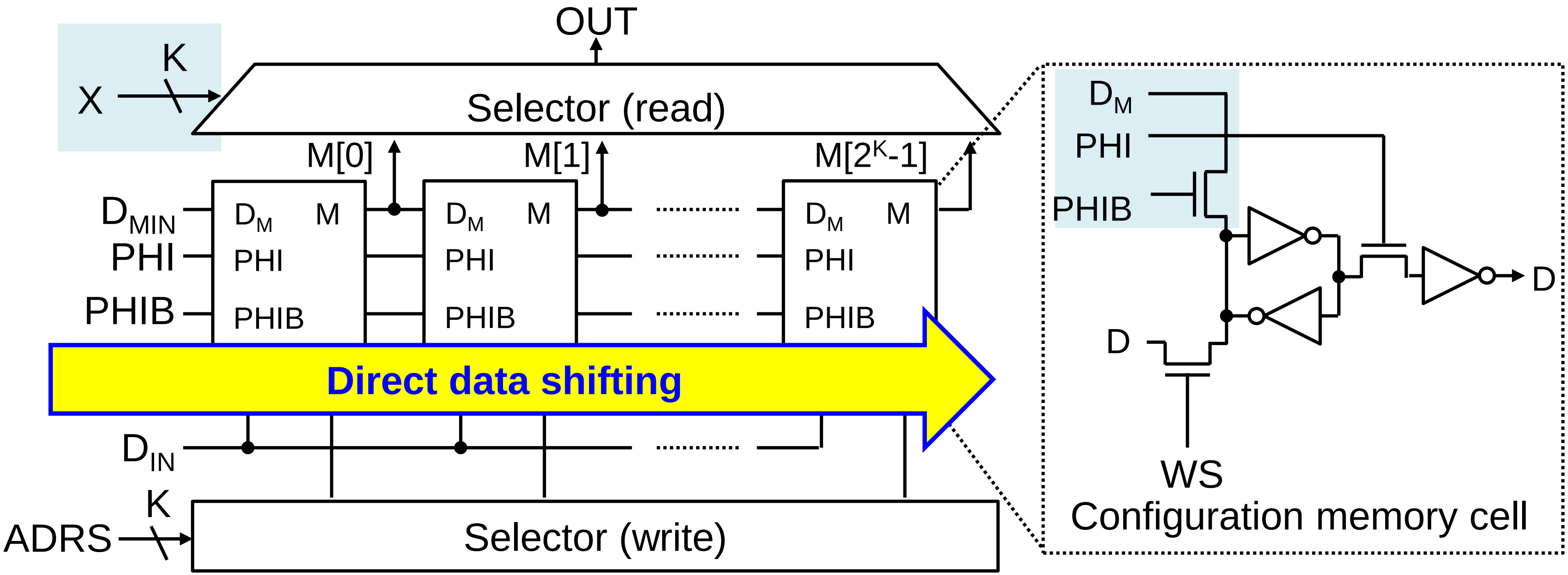
Updated

☺ Number of write access per cycle is minimized to one.

Write power reduction by minimizing # of write access

Conventional Data-Shift Function

■ SRAM-based LUT circuit

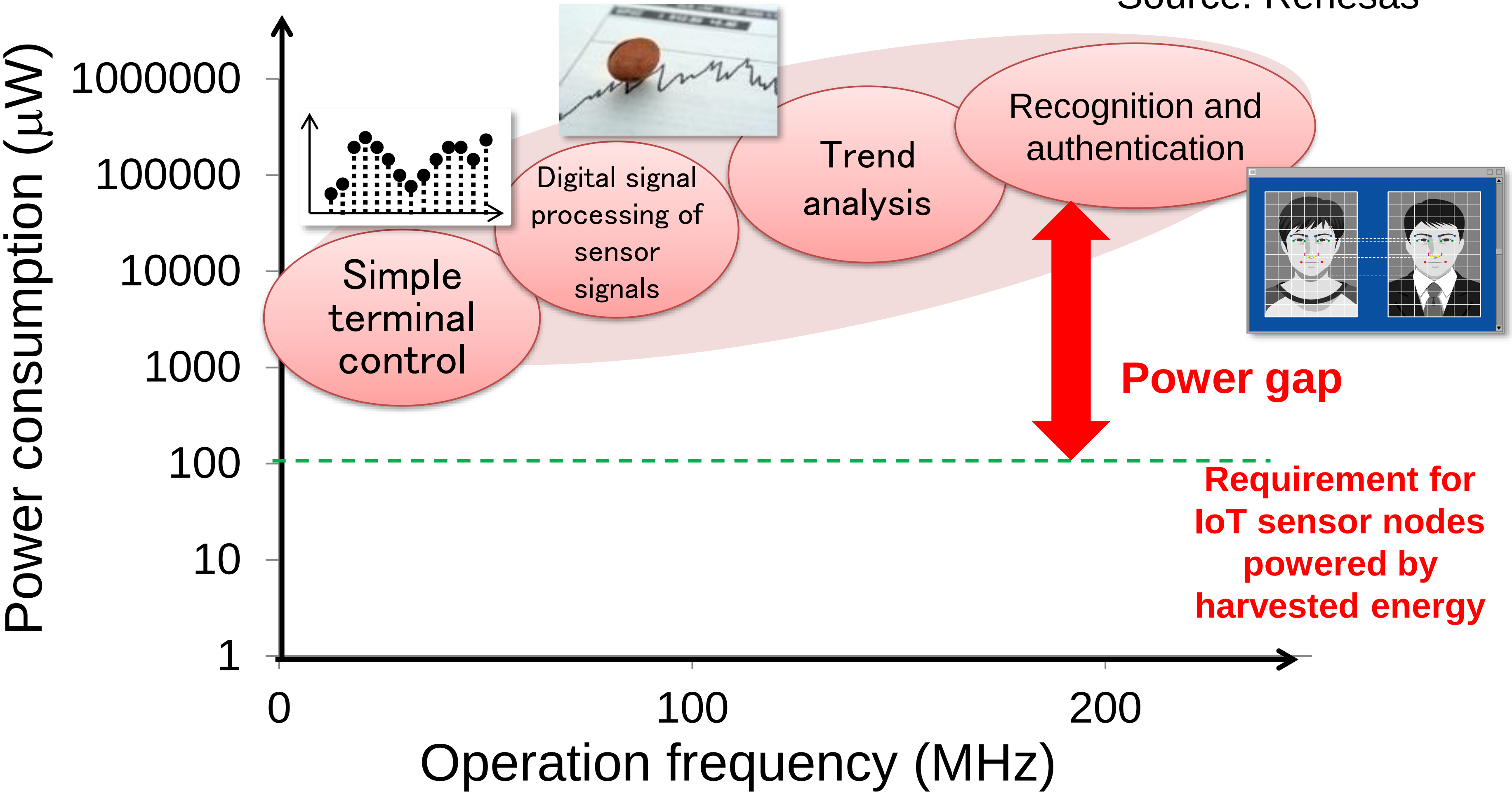


	Conventional	Proposed
# of write access per cycle (K-input LUT)	2^K	**1**

Proposed method is further fewer write access.
-> Low-write-power consumption

Microcontroller Unit (MCU) for Sensor Node

Source: Renesas



Ultra-low-power/high-performance MCU is required for intelligent sensor node app.

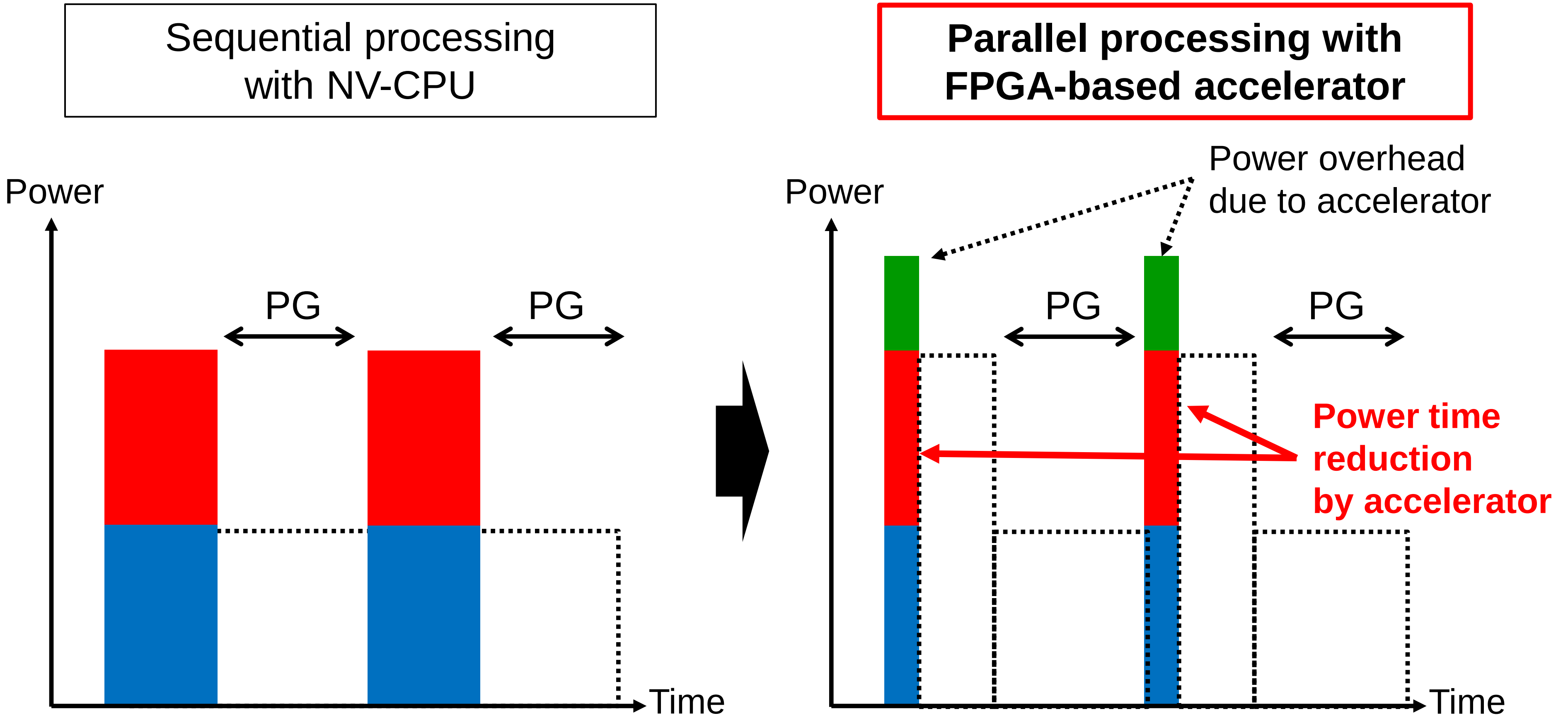
NV-FPGA Accelerated NV-MCU

[Concept]

Replace sequential processing by CPU with **parallel processing by FPGA**

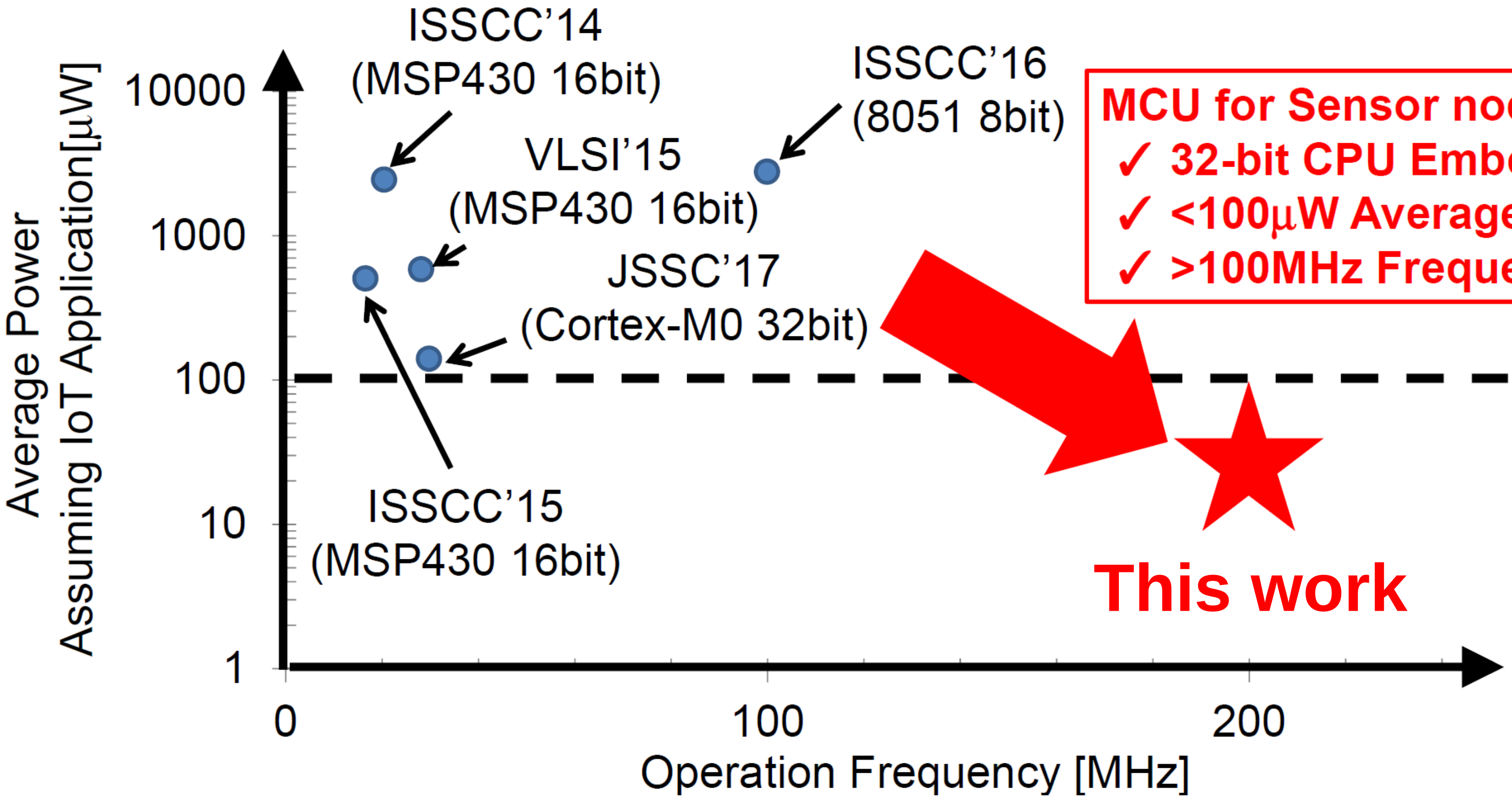
→ reduce processing time and increase the amount of standby state

→ further improve energy efficiency


Sequential processing with NV-CPU
Parallel processing with FPGA-based accelerator
Power
PG
PG
Time
Power
Power overhead due to accelerator
PG
PG
Power time reduction by accelerator
Time

Total power reduction by PG & FPGA-based acceleration

Comparison of Past Works



MCU for Sensor node with
✓ 32-bit CPU Embedded
✓ <100µW Average Power
✓ >100MHz Frequency

47.14µW Operation at 200MHz is achieved.

Contents

1. Introduction

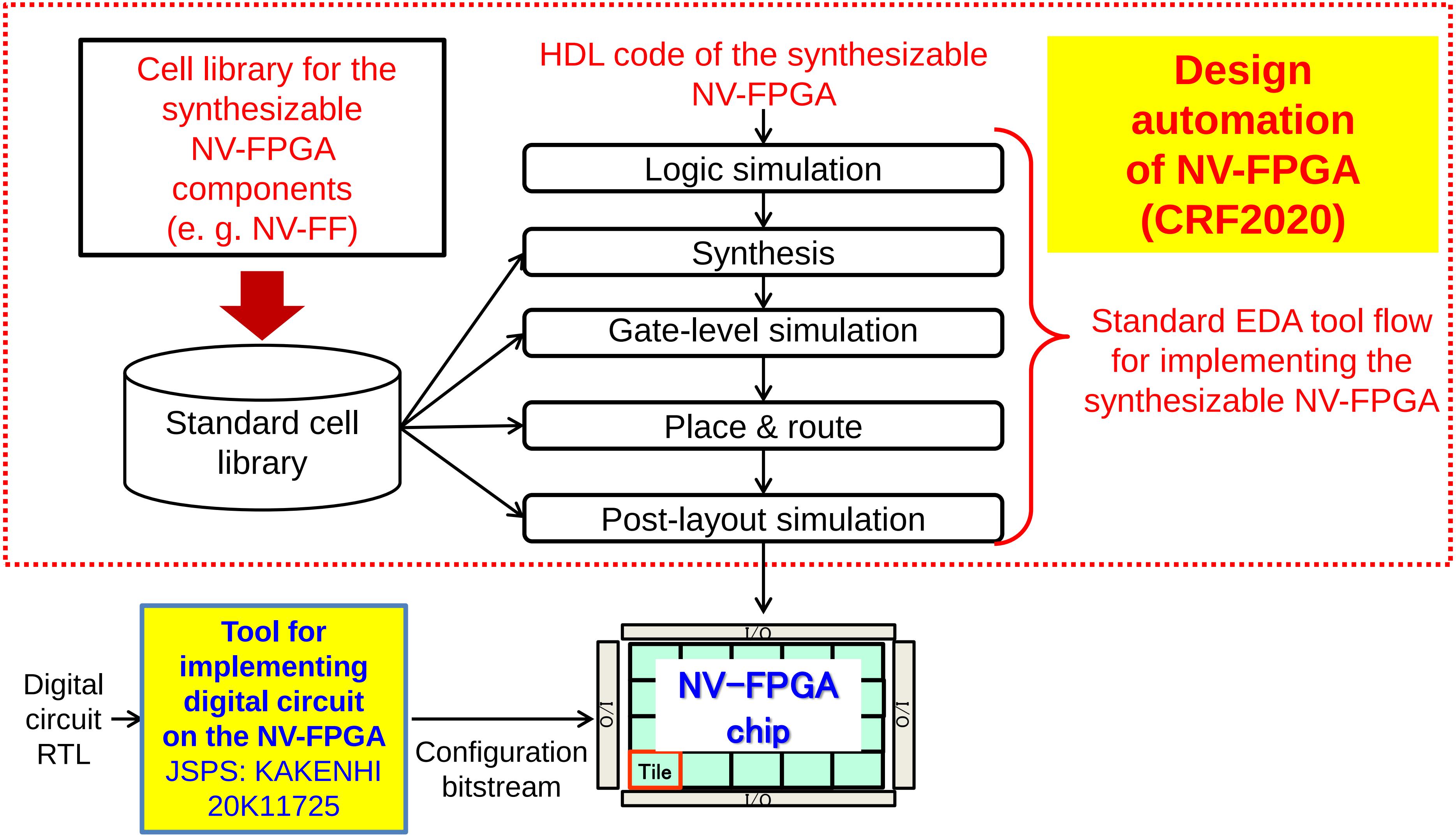
2. NV-FPGA and NV-LUT Circuit

3. Research Plan at UoA

4. Conclusion

(1) Establish EDA Tool Flow for NV-FPGA

☹ In current situation, almost of NV-FPGA design is manual.
-> Establish design automation flow of the NV-FPGA



Cell library for the synthesizable NV-FPGA components (e. g. NV-FF)

Standard cell library

HDL code of the synthesizable NV-FPGA

Logic simulation

Synthesis

Gate-level simulation

Place & route

Post-layout simulation

Design automation of NV-FPGA (CRF2020)

Standard EDA tool flow for implementing the synthesizable NV-FPGA

Digital circuit RTL

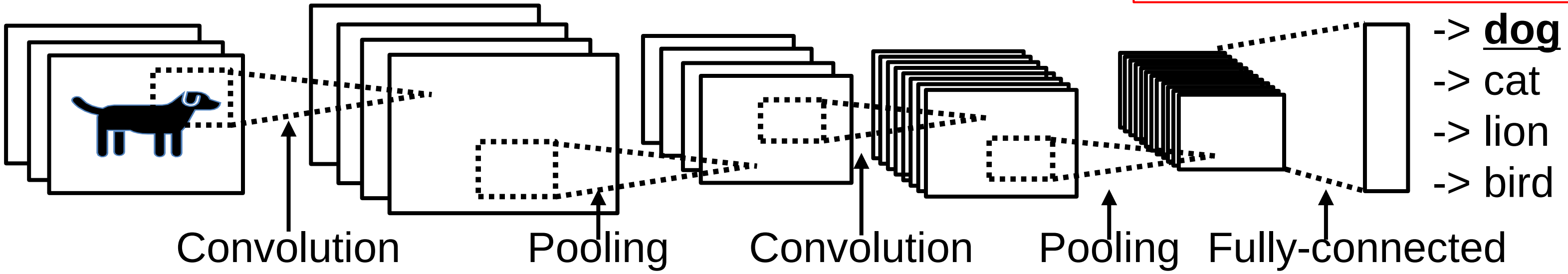
Tool for implementing digital circuit on the NV-FPGA
JSPS: KAKENHI 20K11725

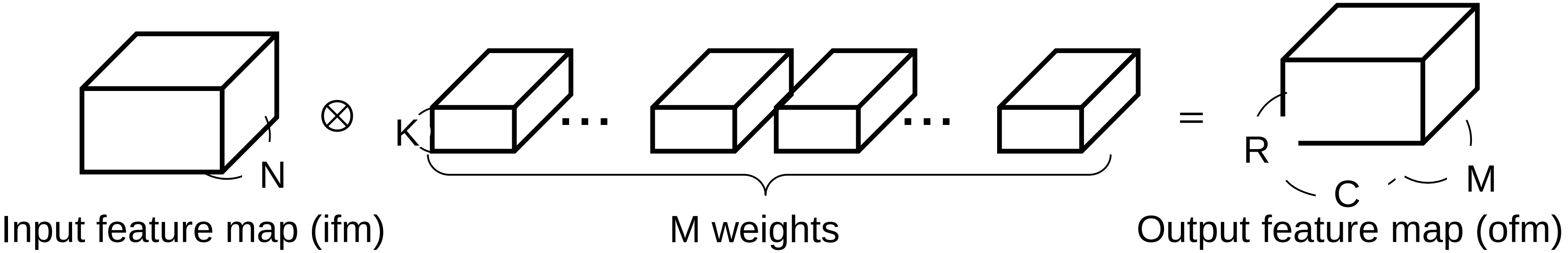
Configuration bitstream

(2) Design NV-FPGA-based AI Accelerator

Example: Binary Convolutional Neural Network (BCNN) [1]

[1] M. Courbariaux et al.,
arXiv:1602.02830, 2016.





```
for (to=0; to<M; to++)
   for (row=0; row<R; row++)
      for (col=0; col<C; col++)
         for (ti=0; ti<N; ti++)
            for (i=0; j<K; i++)
               for (j=0; j<K; j++)
                  ofm[to][row][col] += weight[to][ti][row+i][col+j] * ifm[ti][row+i][col+j];
```

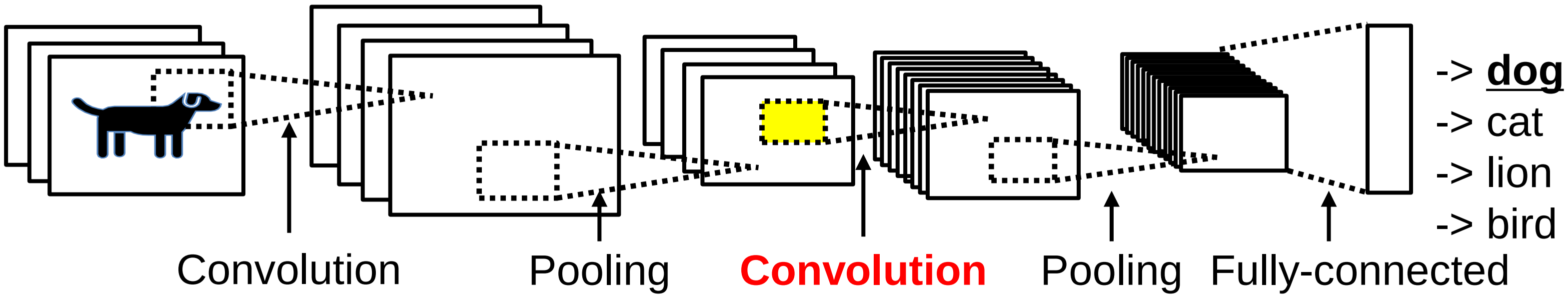
Binary information

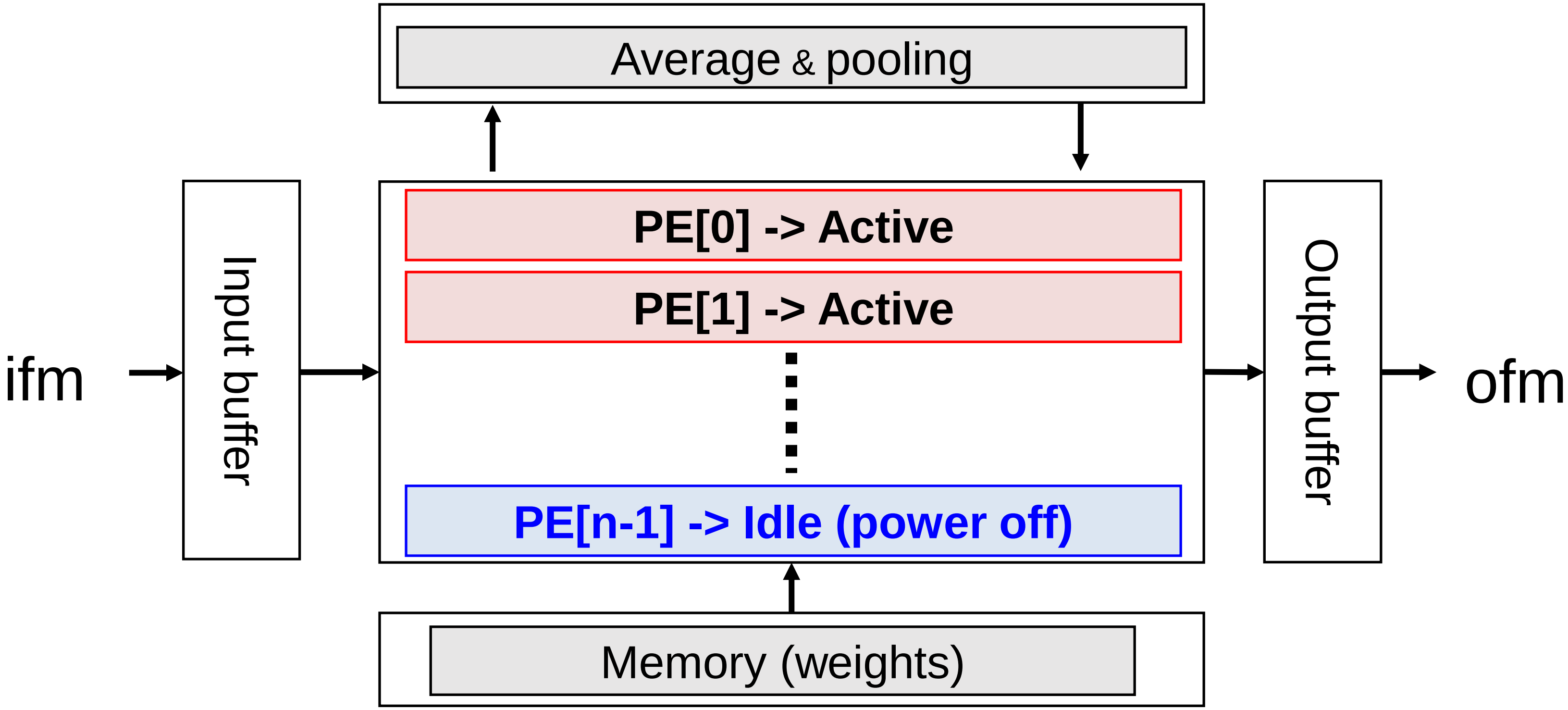
SUM -> Bit-count

Multiplication -> XNOR

Massively parallel computing is required.

NV-FPGA-Based BCNN Accelerator





Massively parallel architecture with no wasted standby power consumption

Ongoing Research Collaborations

[with Prof. Ben]
Competitive Research Fund 2020 in UoA,
`` Development of an Energy-efficient Heterogeneous Spiking Neuro-inspired System for Deep Neural Networks.”

[with Prof. Saito]
IoT/AI Device Cluster

Contents

1. Introduction

2. NV-FPGA and NV-LUT Circuit

3. Research Plan at UoA

4. Conclusion

Conclusion

NV-FPGA and NV-LUT Circuit

- Compact & variation resilient circuity by using LIM structure
- Low-power data shifting by using only-once-write shifting
- Energy-efficient NV-MCU chip by embedding NV-FPGA

Research Plan

- Establish EDA Tool Flow for NV-FPGA
- Design NV-FPGA-based AI Accelerators
- Collaborations with UoA members (Prof. Ben, Prof. Saito, etc.)