EFFICIENT IMAGE PROCESSING VIA MEMRISTIVE-BASED APPROXIMATE IN-MEMORY COMPUTING

By Fabian Seiler, Nima TaheriNejad (presenter)

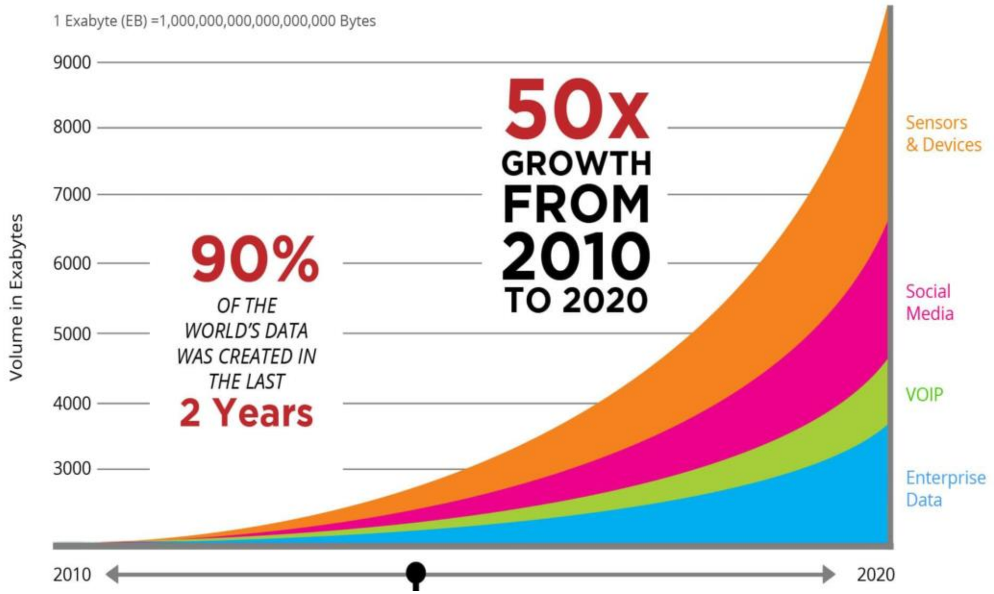
- fabian.seiler@student.tuwien.ac.at
- nima.taherinejad@ziti.uni-heidelberg.de

CONTENT

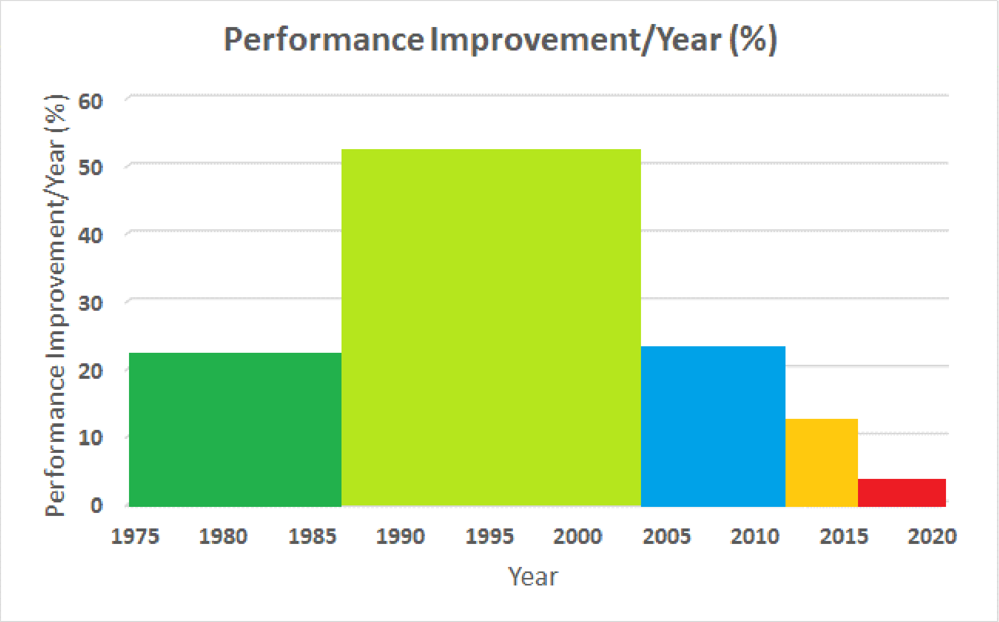
COMPUTERS ARE EVERYWHERE

CHALLENGES

Big Data Era



Performance Improvement in GPP



ENERGY FOOT PRINT OF ICT

N. Jones, The information factories, Nature 2018

Data centers

- 2021: 200 TWh
  - Half of Iran's annual energy consumption

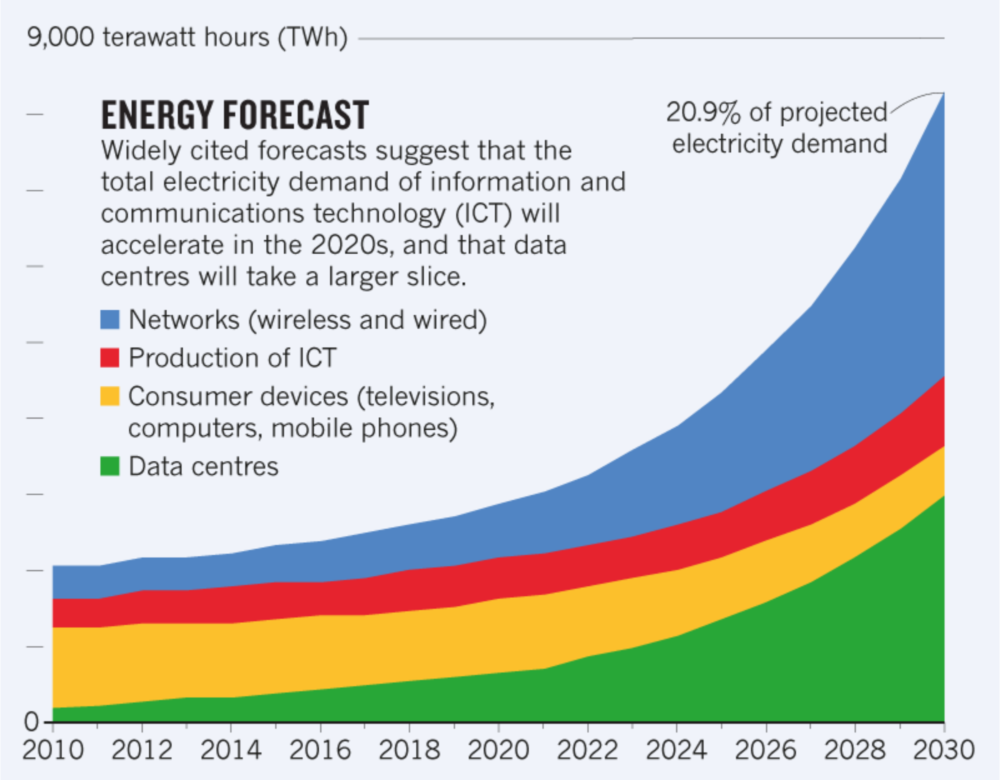
Carbon footprint

- On par with aviation industry

By 2040

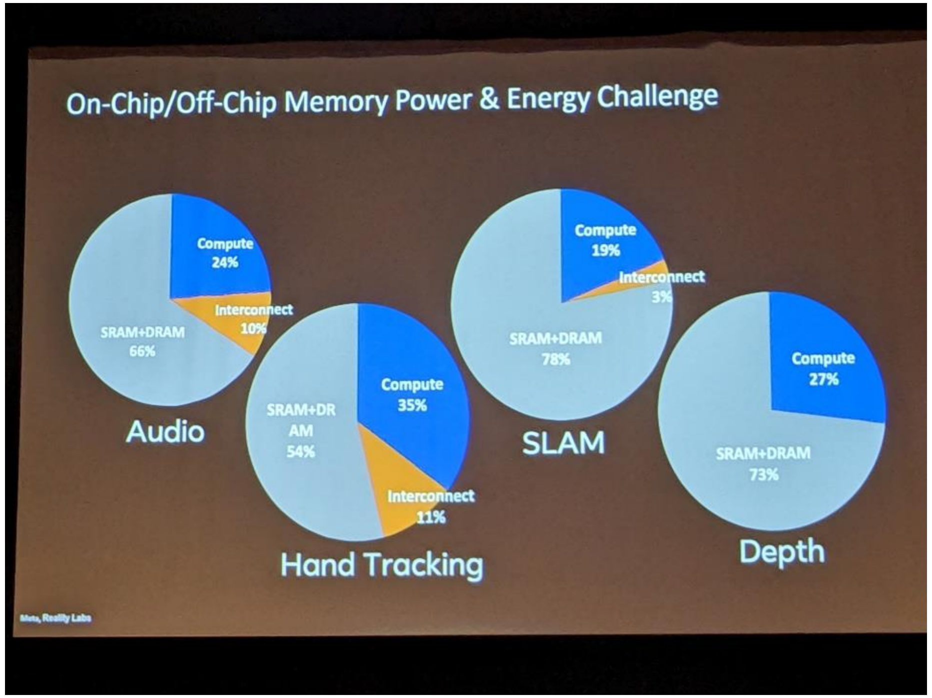
- Surpass energy production capacity

Ometov & Nurmi, Towards approximate computing for achieving energy vs. accuracy trade-offs, DATE 2022



MEMORY ACCESS COST

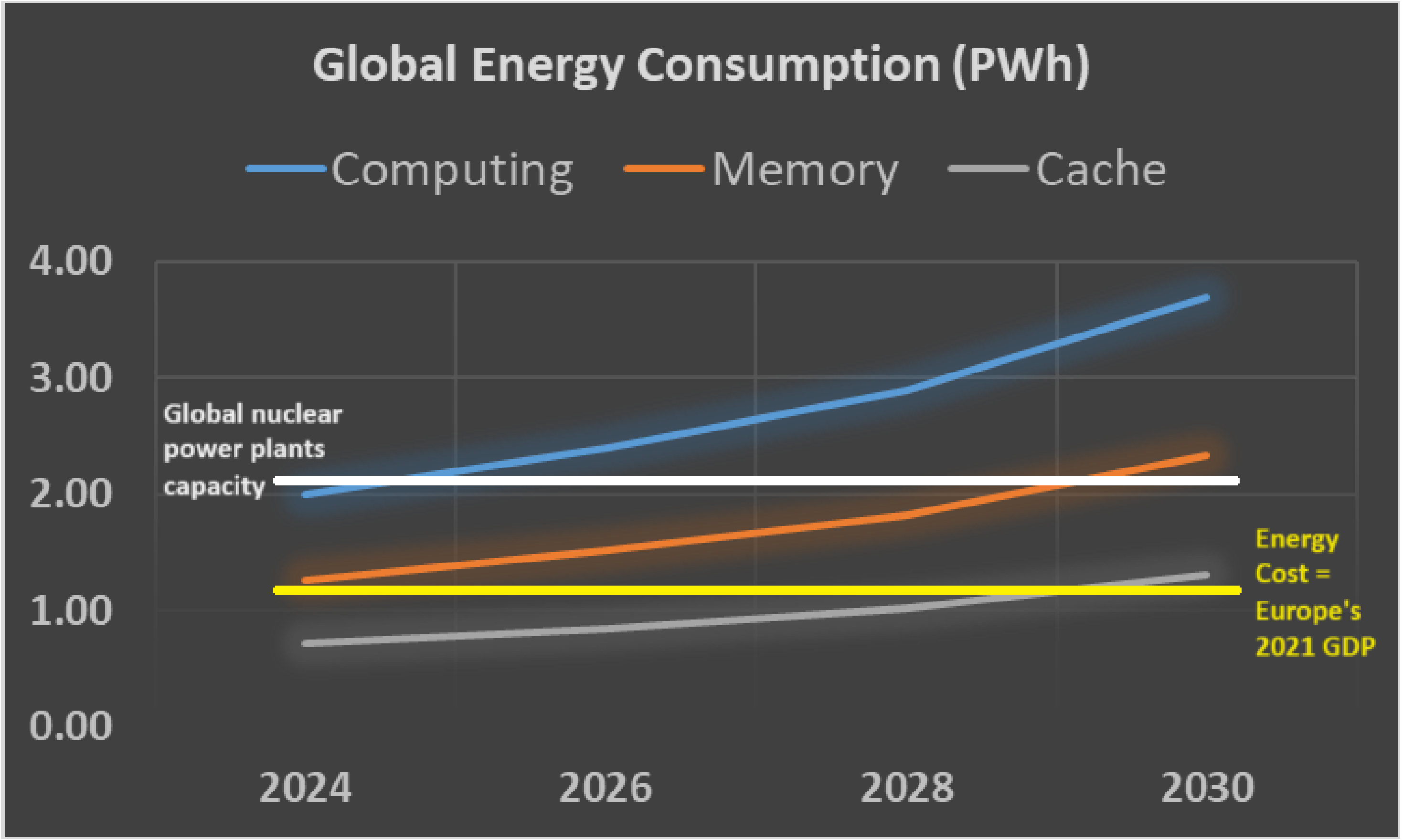
Meta (Facebook)



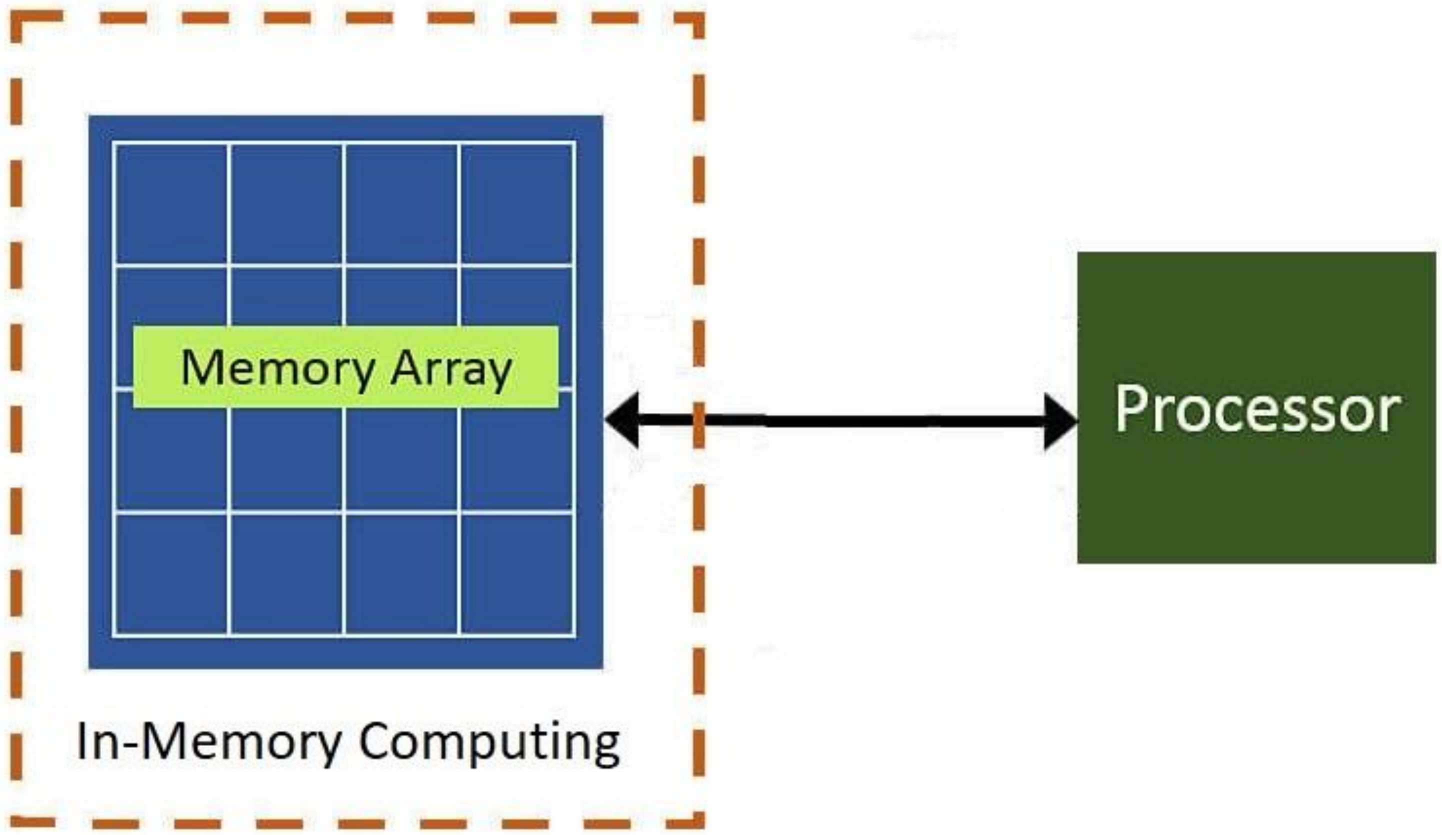
Google

- ~63%
  - A. Boroumand,et al., Google Workloads for Consumer Devices: Mitigating Data Movement Bottlenecks. New York, NY, USA: Association for Computing Machinery, 2018, p. 316–331.
- Let's take this as an indicative number

FOOTPRINT

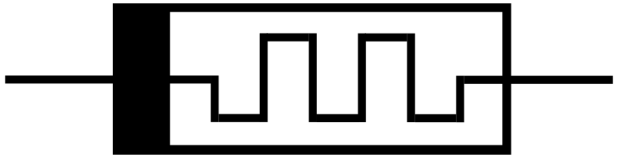

Global Energy Consumption (PWh)
Computing
Memory
Cache
4.00
3.00
2.00
1.00
0.00
2024
2026
2028
2030
Global nuclear power plants capacity
Energy Cost = Europe's 2021 GDP

PERFORMANCE BOTTLENECK & SOLUTION

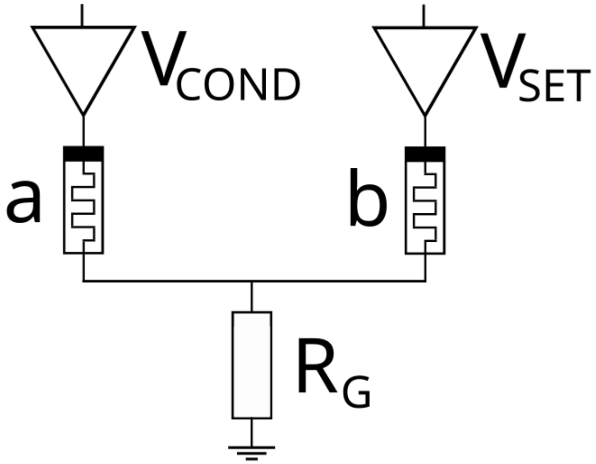


MEMRISTORS - IMPLY

- Memristor
  - Memory + Resistor
  - Resistance represents logical states
- IMPLY
  - Stateful Logic
  - Complete Logic Set with NOT



[1] Memristor – Circuit symbol



[2] IMPLY Gate

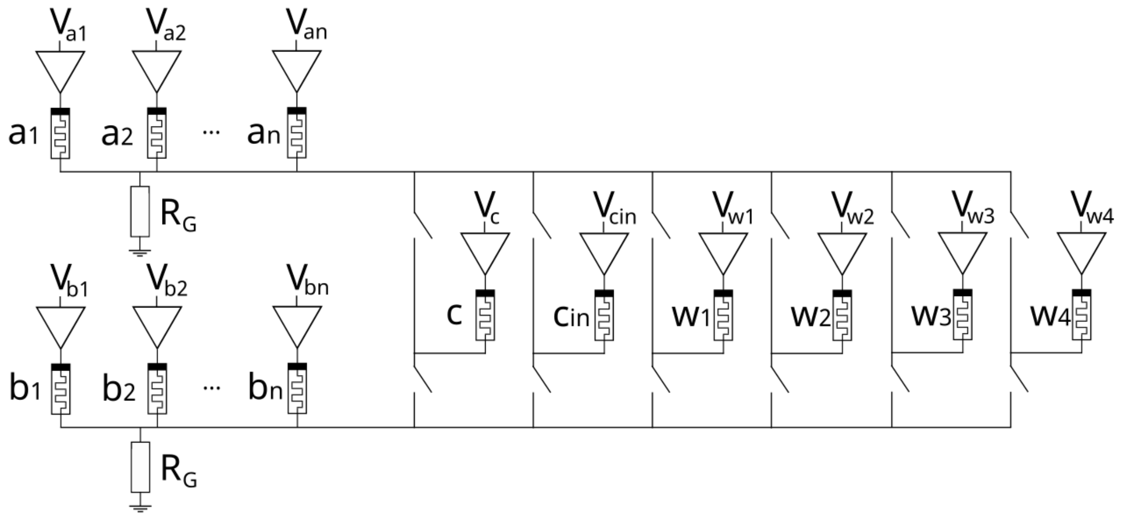
a	b	$b' = a \rightarrow b$
0	0	1
0	1	1
1	0	0
1	1	1

[2] IMPLY - Logical operation

[1] Memristor - the missing circuit element, L. Chua. 1971, IEEE Transactions on Circuit Theory
[2] Memristive switches enable stateful logic operations via material implication, J. Borghetti et al. 2010, Nature

SEMI-SERIAL TOPOLOGY

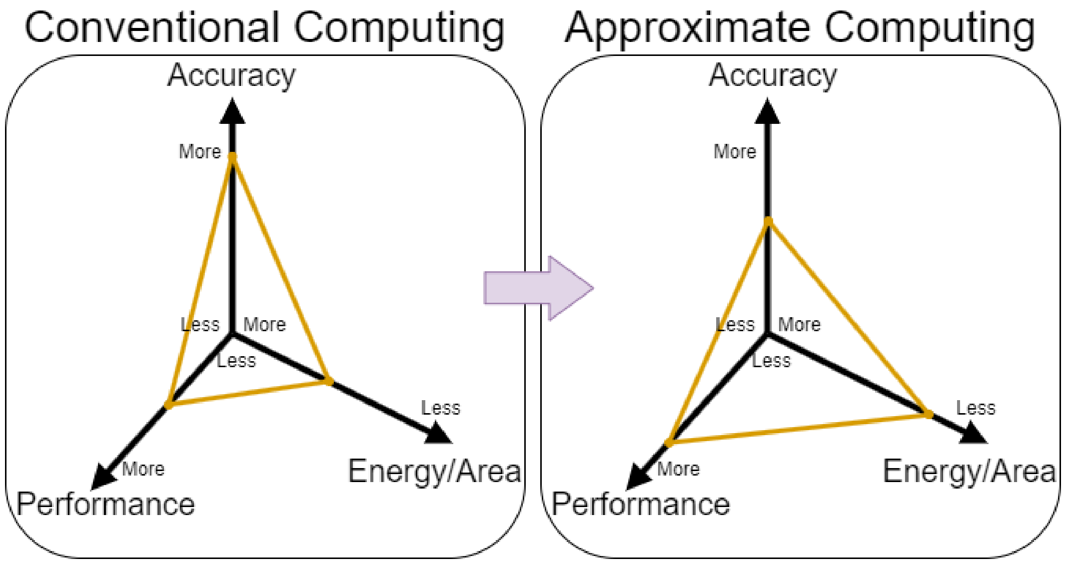
- Combination of topologies
- Two computing sections
- Shared work memristors
- Efficient trade-off



Semi-serial topology, TaheriNejad et al. 2019, IEEE NEWCAS

APPROXIMATE COMPUTING

- Improved speed & energy
- Reduced accuracy
- Redefined truth table
- Evaluation with error metrics
- Error-resilient applications
  - Image Processing
  - Machine Learning

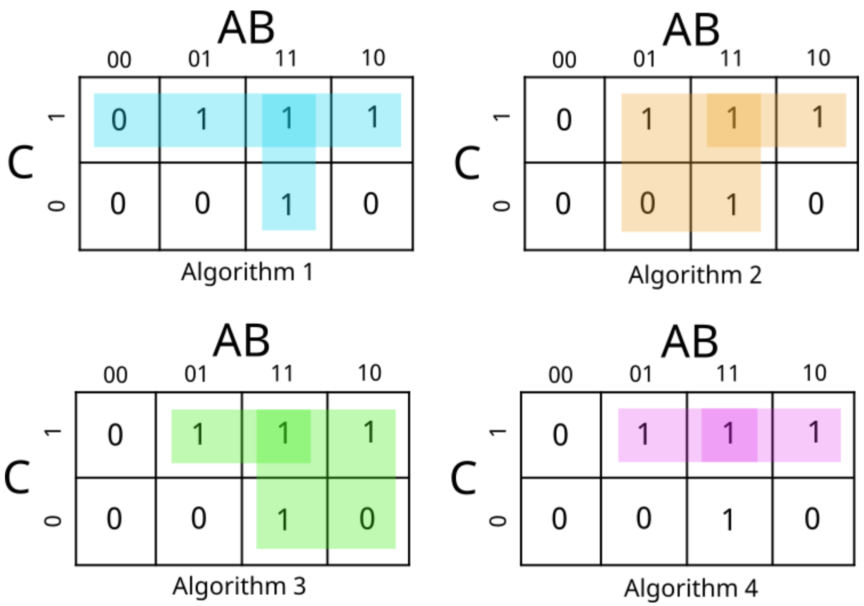


METHODOLOGY

- Exact Full Adder Truth Table
- Error at $Cout$
- Approximation $Sum \approx \overline{Cout}$

A	B	Cin	Sum	Cout
0	0	0	~~0~~ 1	0
0	0	1	~~1~~ 0	~~0~~ 1
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	~~1~~ 0	1

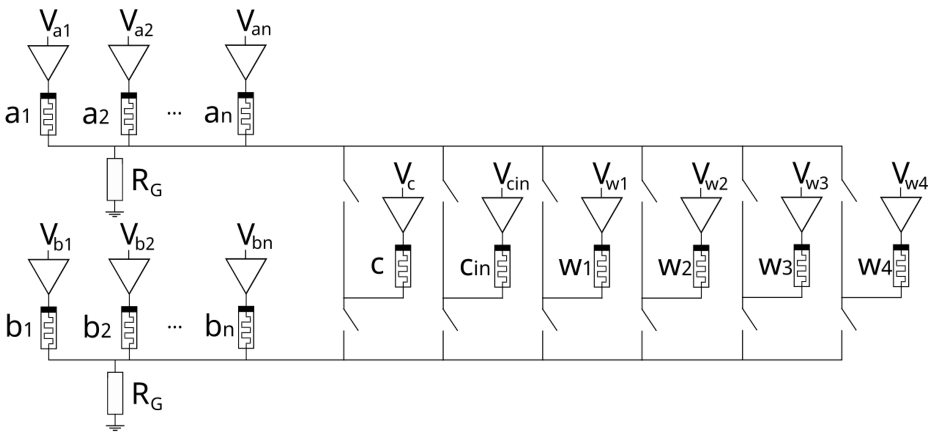
Truth table of exact full-adder



KVD of Cout for presented algorithms

REALIZATION WITH IMPLY-LOGIC

- Semi-Serial Topology
- Parallelization
- Reset of work memristors
- Cout saved in c-memristor
- Sum saved in a-memristor



Semi-serial topology, TaheriNejad et al. 2019, IEEE NEWCAS

$$Cout = ac + b = \overline{b} \rightarrow \overline{(a \rightarrow \overline{c})}$$

$$Sum = \overline{ac + b} = \overline{\overline{b} \rightarrow \overline{(a \rightarrow \overline{c})}}$$

Logic of algorithm 2

Steps	Section 1	Section 2	Equivalent Logic
-		$w_1 = w_2 = 0$	$False(w_1, w_2)$
1	$w_1' = c \rightarrow w_1$	$w_2' = b \rightarrow w_2$	$w_1 = \overline{c}, w_2 = \overline{b}$
2	$w_1'' = a \rightarrow w_1'$	$c = 0$	$w_1 = a \rightarrow \overline{c}, False(c)$
3	$c' = w_1'' \rightarrow c$		$c = \overline{a \rightarrow \overline{c}}$
4	$a = 0$	$c'' = w_2' \rightarrow c'$	$False(a), c = \overline{b} \rightarrow \overline{(a \rightarrow \overline{c})} = Cout$
5	$a' = c'' \rightarrow a$	$w_1 = w_2 = 0$	$a = \overline{\overline{b} \rightarrow \overline{(a \rightarrow \overline{c})}} = Sum, False(w_1, w_2)$

Approximated algorithm 2

OVERVIEW OF THE ALGORITHMS

- Same error rate
- Different places for error
- Algorithm 1 superior for 1-bit
- Only 2/3 work memristors

A	B	Cin	Sum	Cout
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Errors

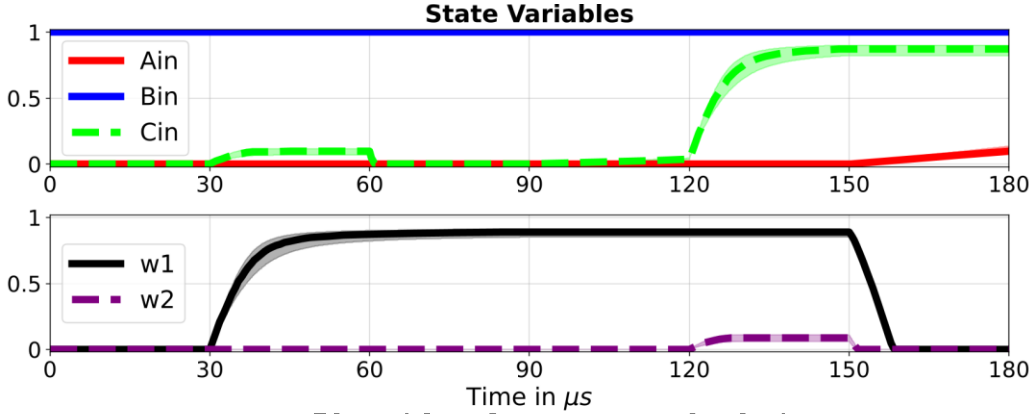
Truth table of exact full adder

Algorithm	ER(C_{out})	ER(Sum)	No. of steps	No. of memristors
Algorithm 1	1/8	3/8	$4n + 1$	$2n + 2$
Algorithm 2	1/8	3/8	$5n + 1$	$2n + 3$
Algorithm 3	1/8	3/8	$5n + 1$	$2n + 3$
Algorithm 4	1/8	3/8	$5n + 1$	$2n + 3$

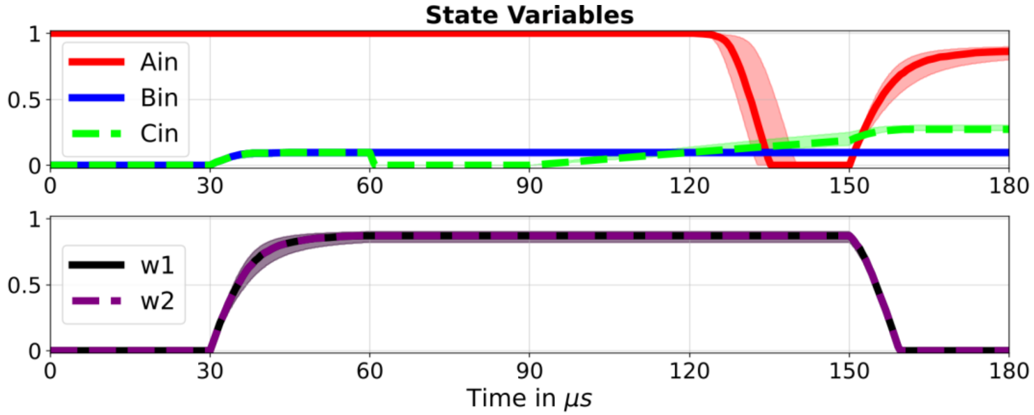
CIRCUIT-LEVEL SIMULATION

- VTEAM-Model
  - Fitted to measurements
- ATOMIC Pipeline
  - Stateful Logic Validation
  - Automated SPICE Simulations
  - Deviation Experiments

Get ATOMIC access here: https://github.com/fabianseiler/ATOMIC



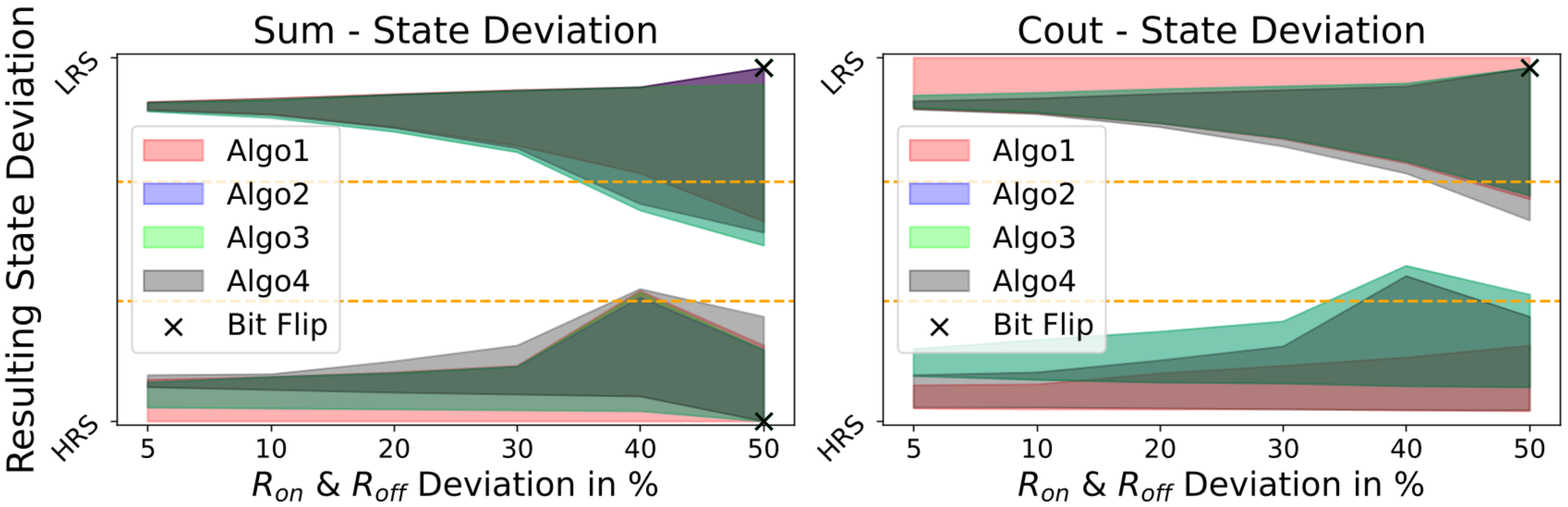
Algorithm 2: correct calculation



Algorithm 2: incorrect calculation

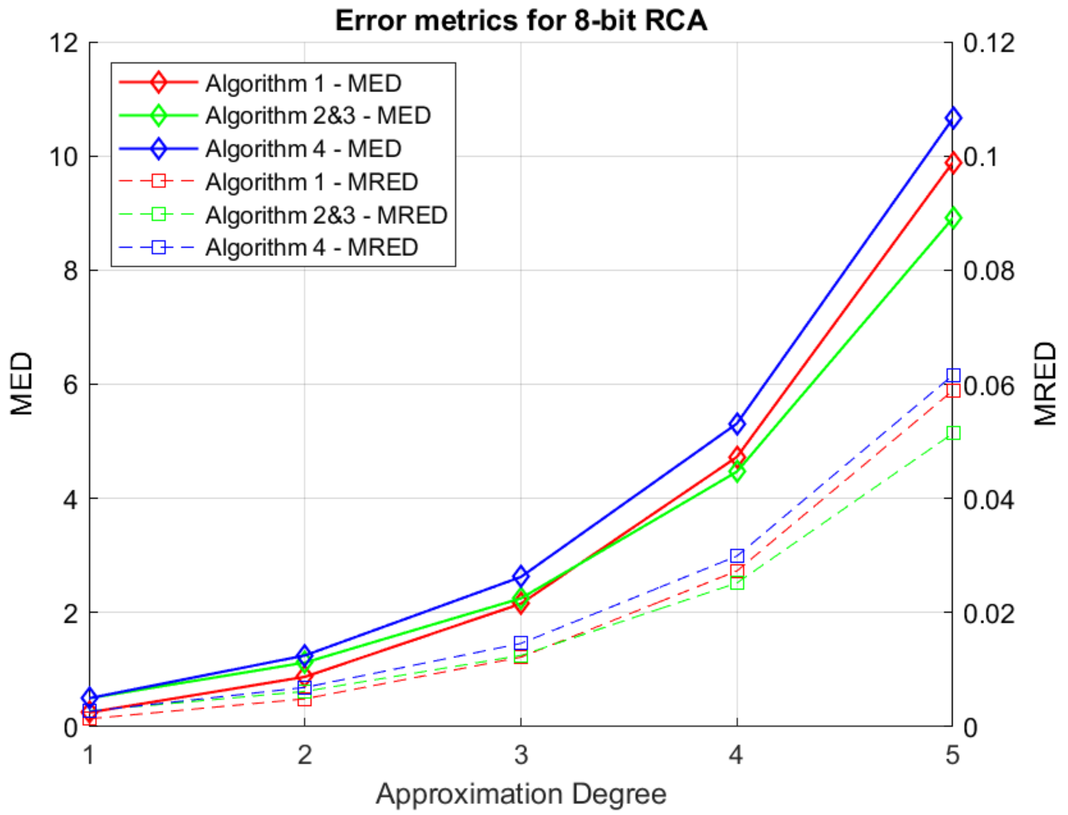
DEVIATION EXPERIMENTS

- Non-ideal behavior
- Deviation of resistive States R_{on} & R_{off}
- Correct up to $\pm 30\%$



ERROR METRICS

- Accuracy evaluation
- MED, NMED, MRED
- 8/16/32-bit



CIRCUIT-LEVEL COMPARISON

- Exact Semi-Serial Adder (8-bit)
  - 6-38% less energy
  - 5-35% fewer steps
  - Same area usage
- SoA Serial Ax Adder (8-bit)
  - 5-29% less energy
  - 45-54% fewer steps
  - 3 memristors & 12 switches more
  - Better accuracy with algorithms 2&3

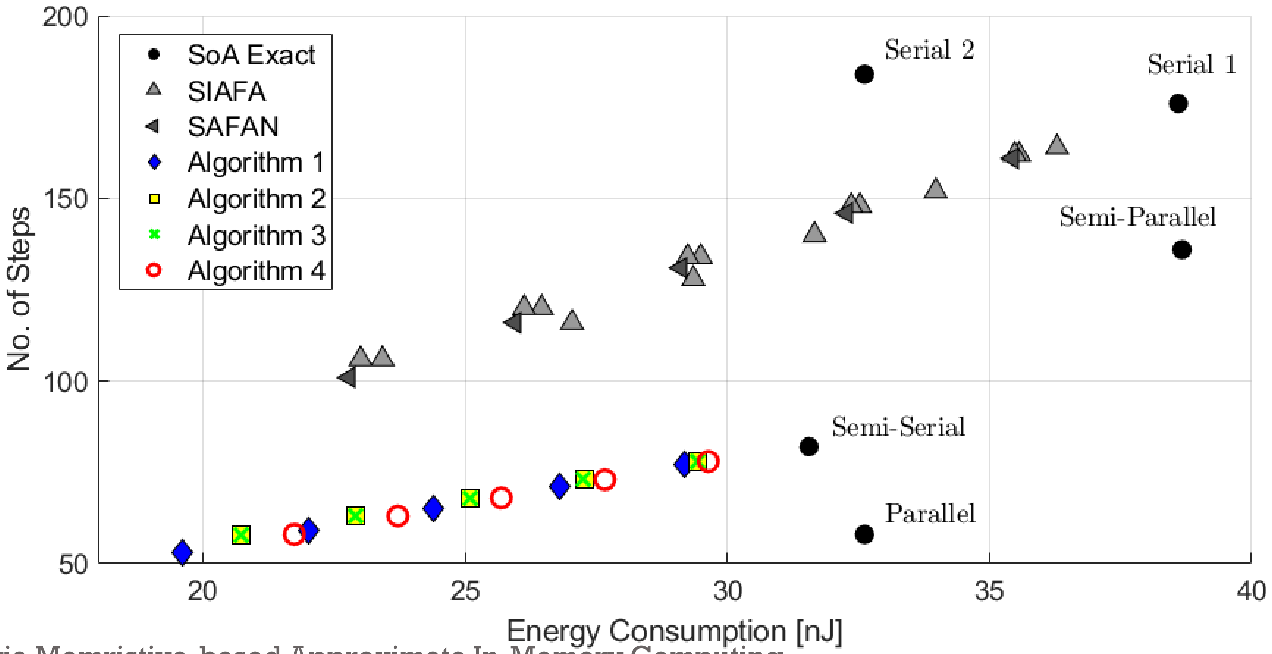
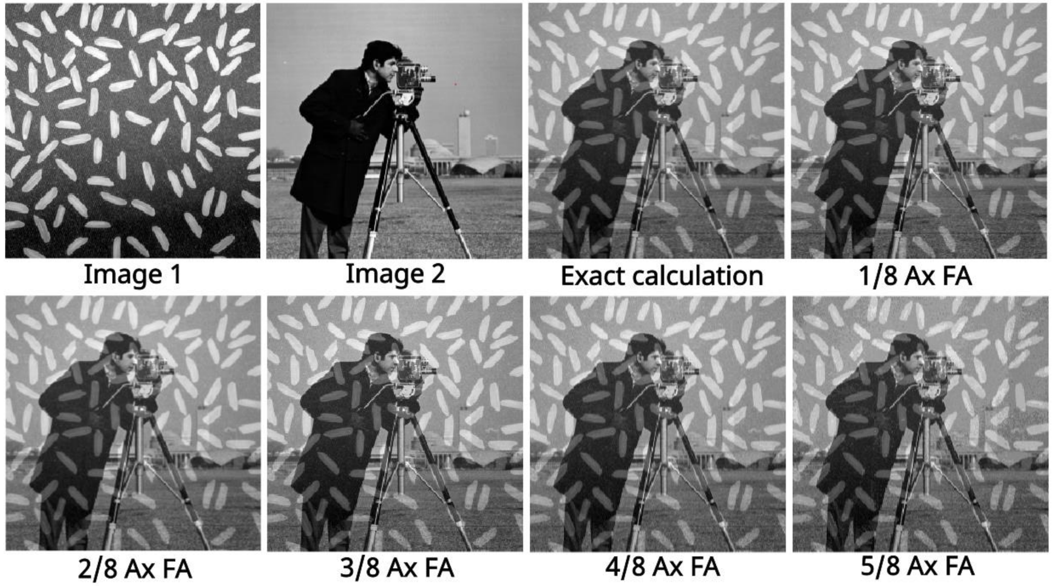
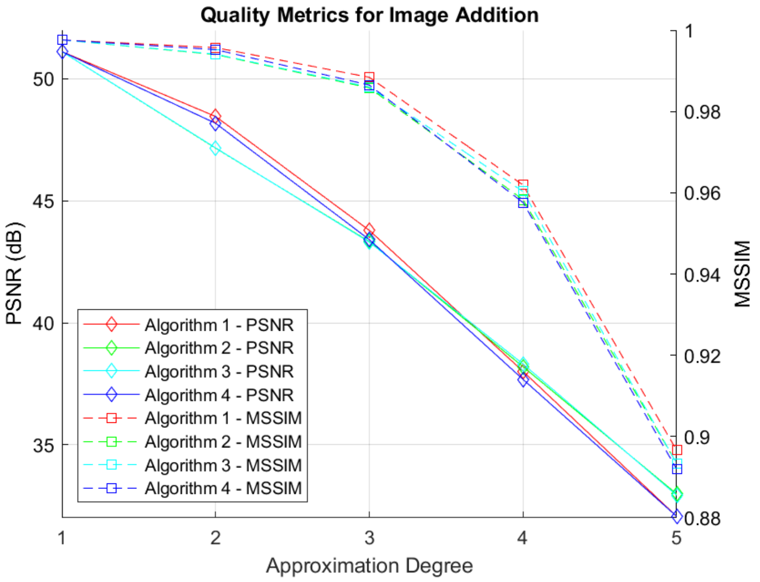


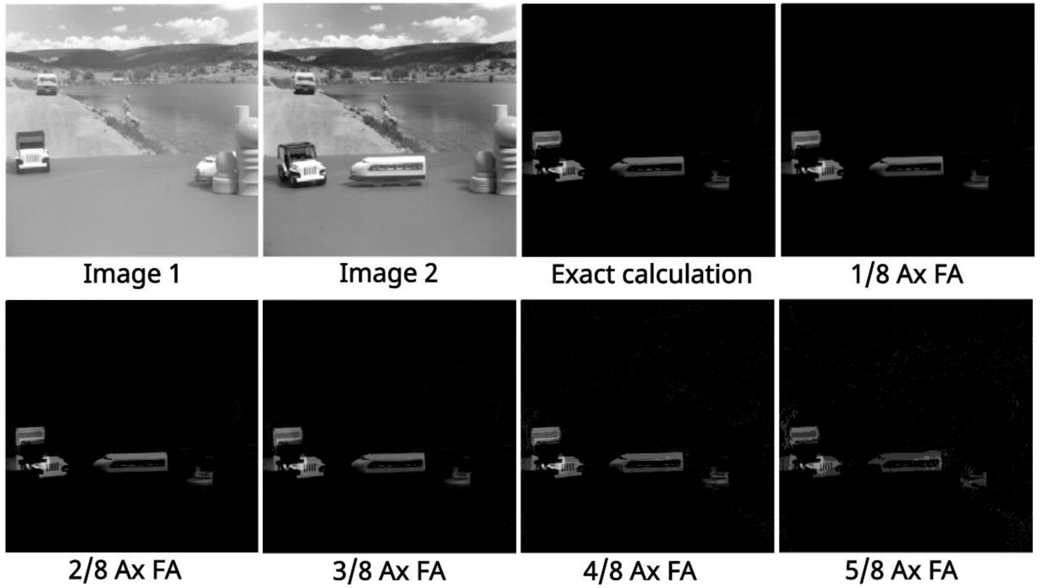
IMAGE ADDITION

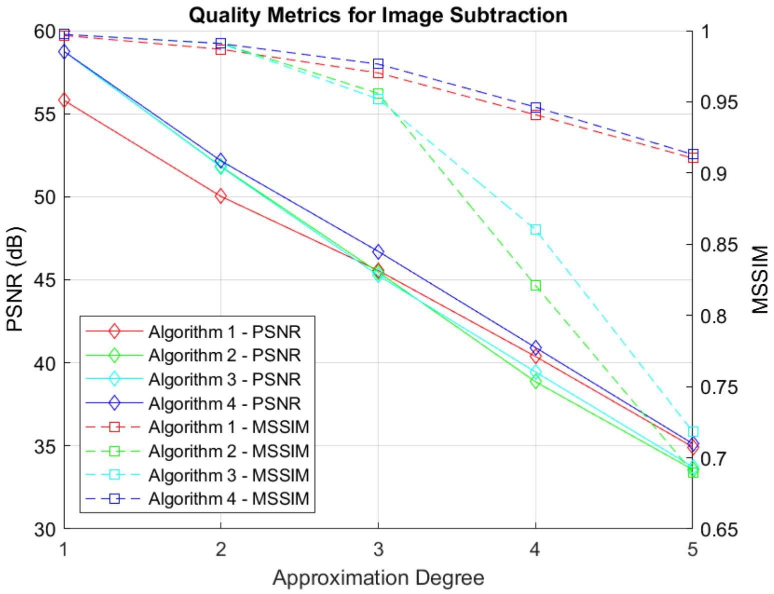




Algorithm 3: Addition with different approximation degrees

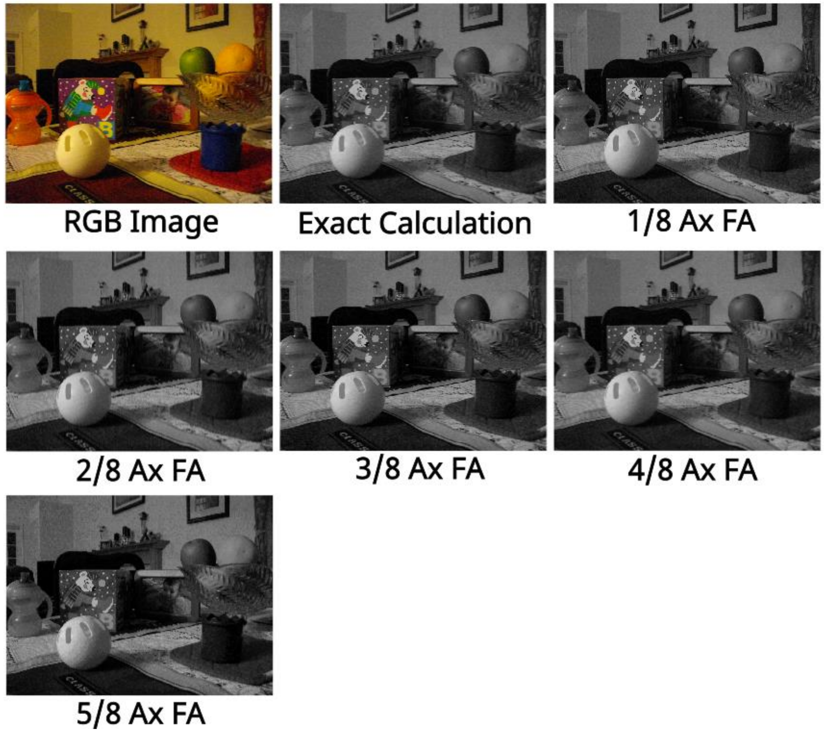
IMAGE SUBTRACTION

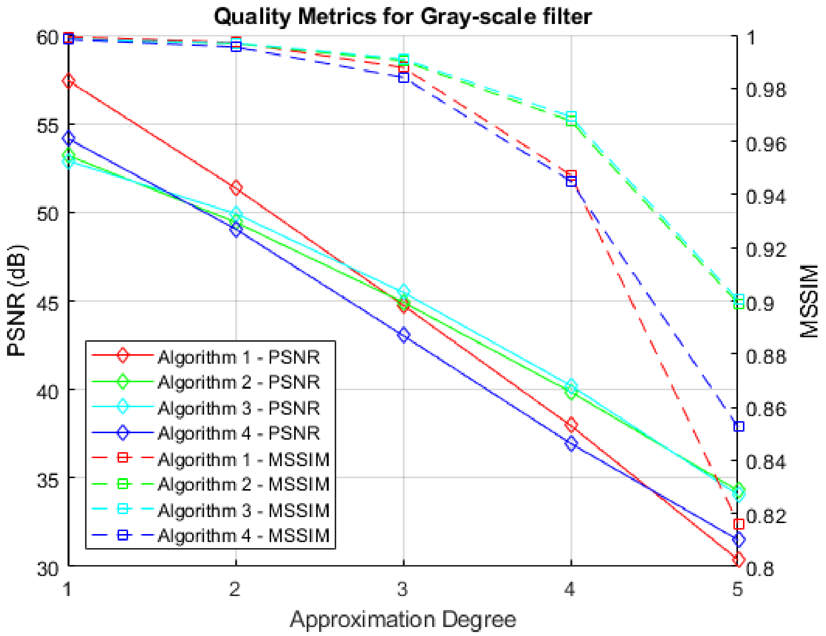




Algorithm 3: Subtraction with different approximation degrees

GRAY-SCALE FILTER

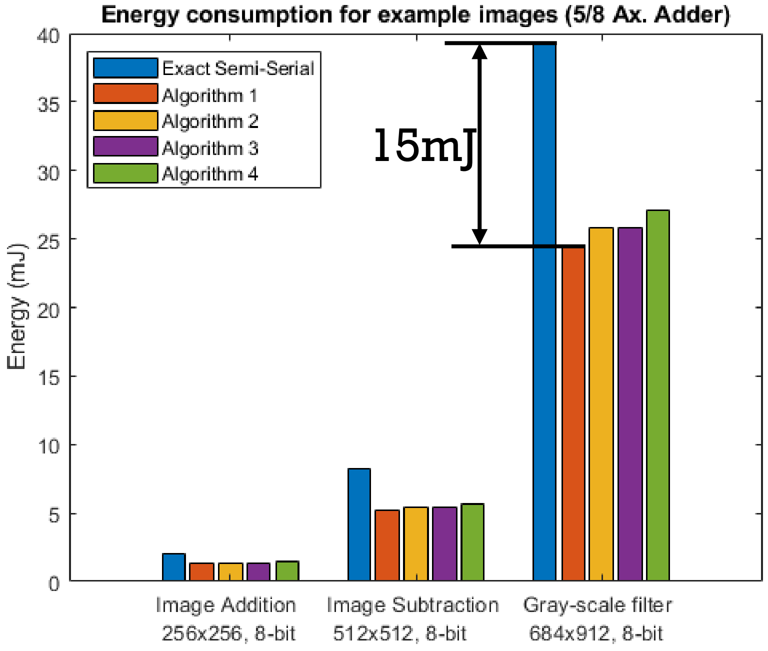


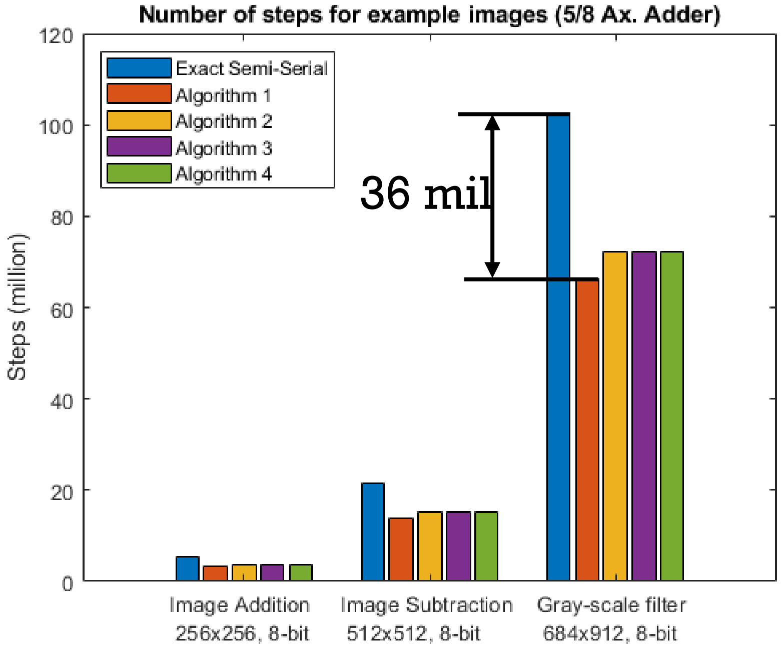


Algorithm 3: Gray-scale filter with different approximation degrees

APPLICATION LEVEL COMPARISON

- Energy and steps reduction at application level

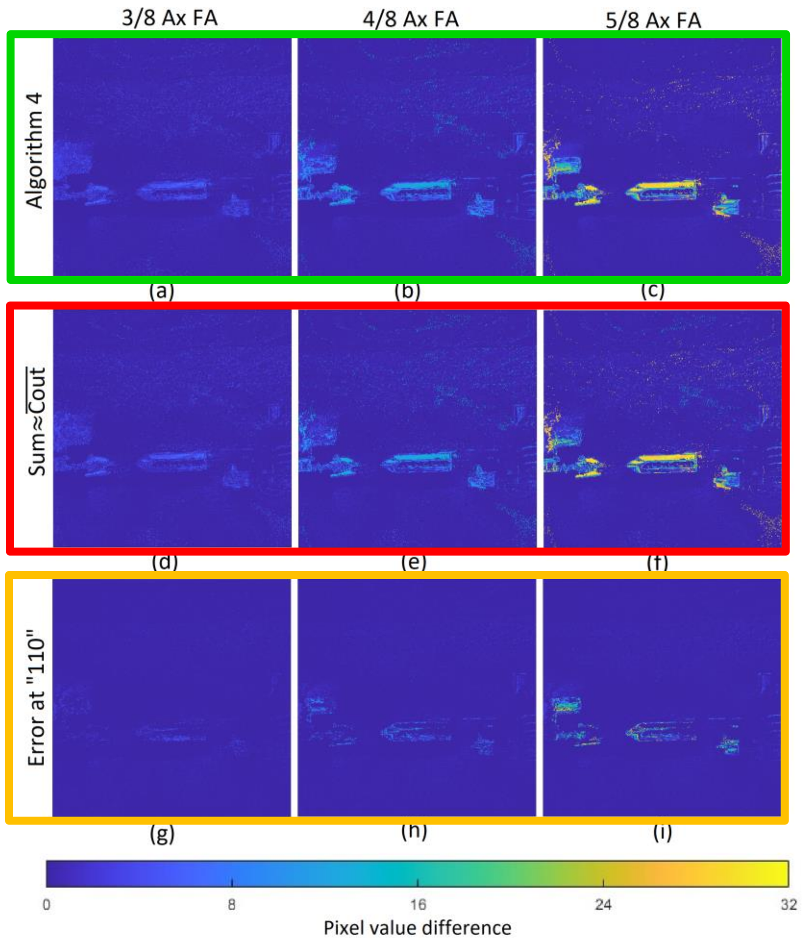




ERROR EMERGENCE

- Application Specific
- Dependent on Algorithm
- Error Placement

A	B	Cin	Sum	Cout
0	0	0	~~0~~ 1	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	~~0~~ 1	~~1~~ 0
1	1	1	~~1~~ 0	1



CONCLUSIONS AND OUTLOOK

- Conclusion:
  - We combined IMC and AxC
  - Fast and efficient algorithms
  - Advantages of the semi-serial topology
  - Gains at application level
- Outlook:
  - In-depth Analysis of Error placements
  - Application-based Approximations
  - Application in machine learning

ECLECTX

25

THANK YOU FOR YOUR TIME AND ATTENTION!