Normalized Stability: A Cross-Level Design Metric for Early Termination in Stochastic Computing

Di Wu, Ruokai Yin and Joshua San Miguel
University of Wisconsin-Madison

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

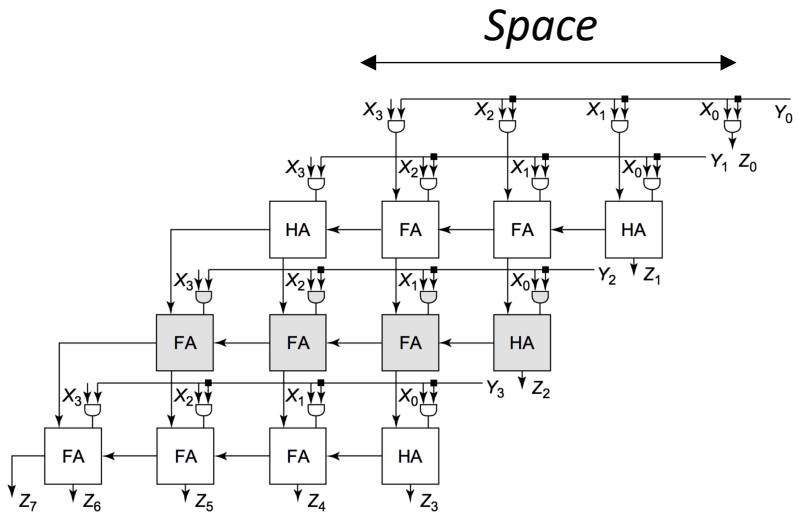
Outline

- ❑ Primer on stochastic computing
- ❑ Early termination for stochastic computing
- ❑ Metric-based characterization for early termination
- ❑ Evaluation of the metric on applications

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

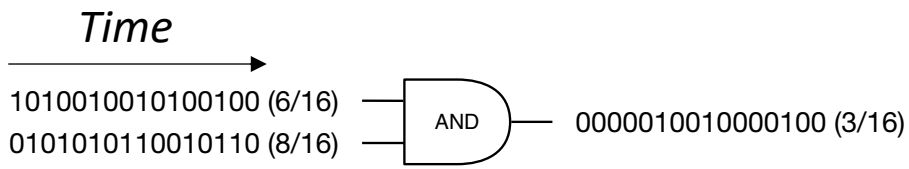
Stochastic Computing – Comparison

- Binary Computing
  - Varying significance in parallel data bits.
  - Spatial domain with complex logic.



Binary multiplier (4-bit data in 1 cycle)

- Stochastic Computing (SC)
  - Equal significance in serial data bits.
  - Temporal domain with simple logic.



SC multiplier (4-bit data in 16 cycles)

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

SC Data

- Bit stream with rate coding
  - Valued by the frequency of 1s in the bit stream

A=0.5	1	1	0	1	0	1	1	1	0	0	1	0	0	1	0	0
B=0.5	0	0	1	0	0	1	1	1	0	1	0	1	0	1	0	1
C=0.75	1	0	1	1	0	1	1	1	1	1	0	1	0	1	1	1

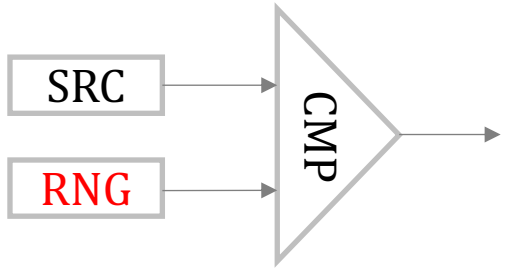
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

SC Data

- Bit stream with rate coding
  - Valued by the frequency of 1s in the bit stream

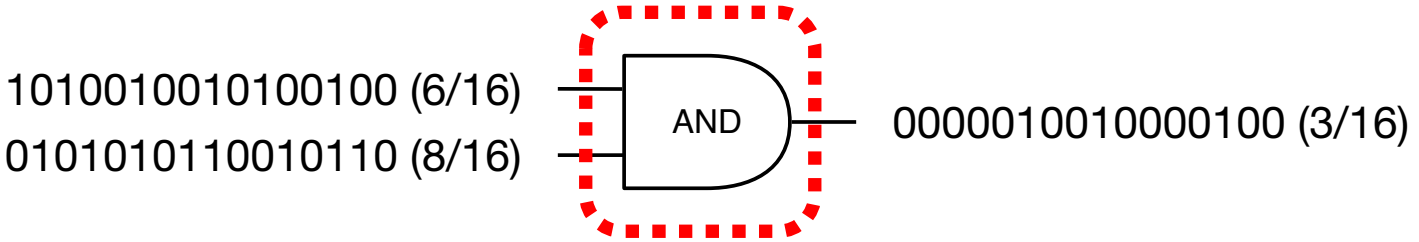
A=0.5	1	1	0	1	0	1	1	1	0	0	1	0	0	1	0	0
B=0.5	0	0	1	0	0	1	1	1	0	1	0	1	0	1	0	1
C=0.75	1	0	1	1	0	1	1	1	1	1	0	1	0	1	1	1

  - Generated by comparing source data with random numbers



Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

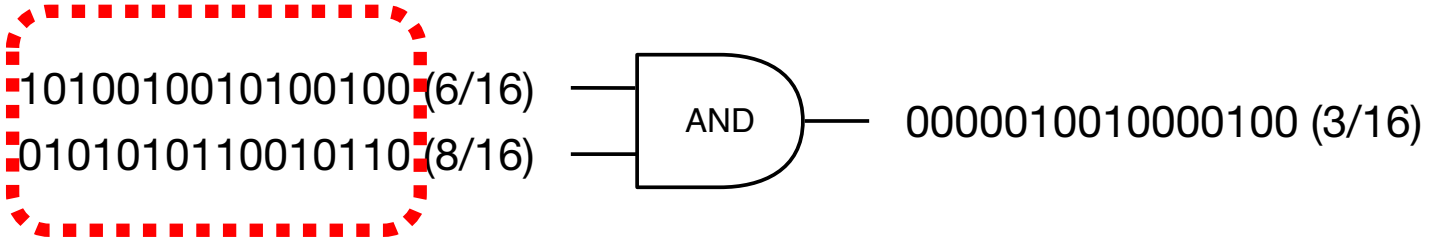
Pros & Cons of SC



AND gate as multiplier

- Pros: computing unit
  - Low area and power
  - High parallelism

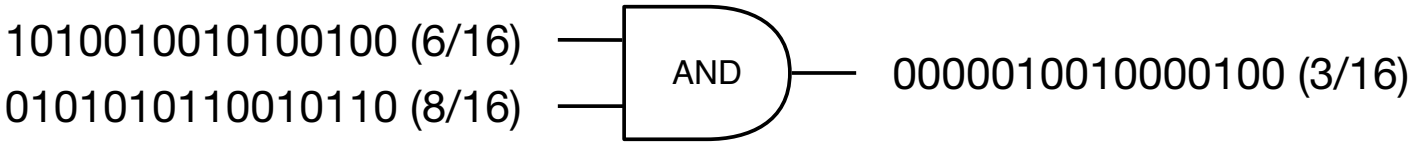
Pros & Cons of SC



n-bit data require 2^n cycles, i.e., exponentially increasing latency

- Pros
  - Low area and power
  - High parallelism
- Cons: data representation
  - Inaccuracy due to randomness
  - Long latency

Pros & Cons of SC

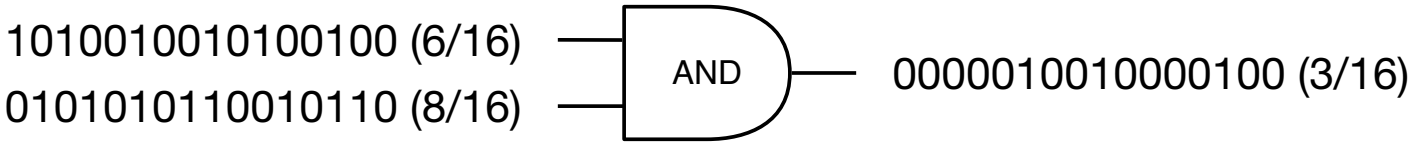


- Pros
  - Low area and power
  - High parallelism
- Cons
  - Inaccuracy due to randomness
  - Long latency

Joint effect:
Undetermined energy efficiency

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Pros & Cons of SC



- Pros
  - Low area and power
  - High parallelism
- Cons
  - Inaccuracy due to randomness
  - Long latency

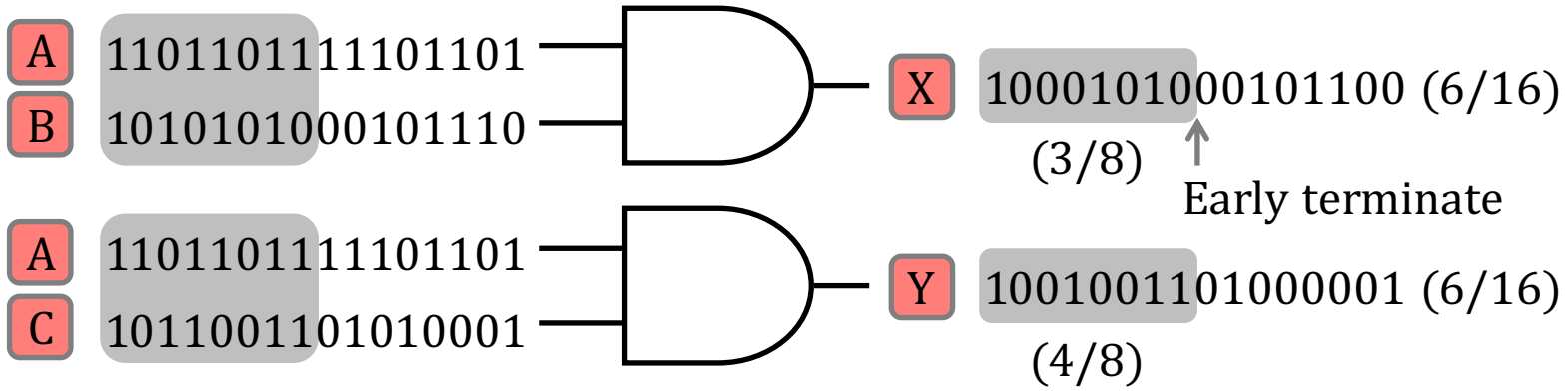
Joint effect:
Undetermined energy efficiency

Improve energy efficiency with early termination

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Early Termination

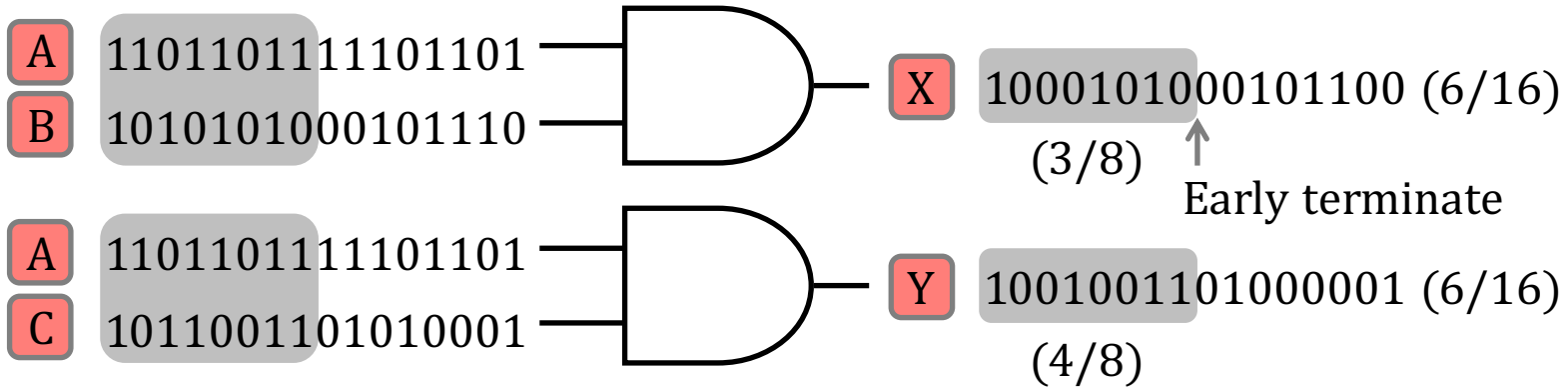
- Early termination (ET) enables high energy efficiency



  - X can be terminated without error using half cycles

Early Termination

- Early termination (ET) enables high energy efficiency



  - X can be terminated without error using half cycles

- New metric for characterizing early termination
  - Normalized stability

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Normalized Stability

- Capability
  - Unit level:
    - Identify the competitiveness for ET
  - Application level:
    - Explore design space for ET
    - Predict the timing for ET

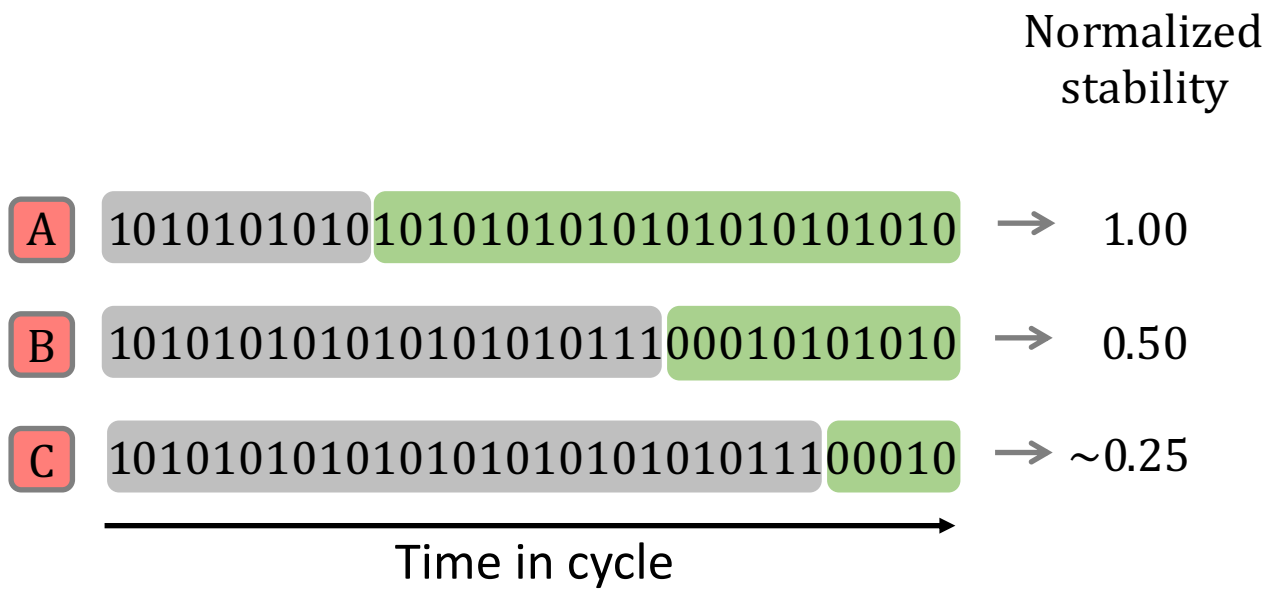
Normalized Stability

- Definition
  - How long a bit stream has been stable within a pre-defined accuracy budget, normalized to the maximum achievable stable duration

Normalized Stability

- Definition
  - How long a bit stream has been stable within **a pre-defined accuracy budget**, normalized to the maximum achievable stable duration

5% error, adapted from approximate computing

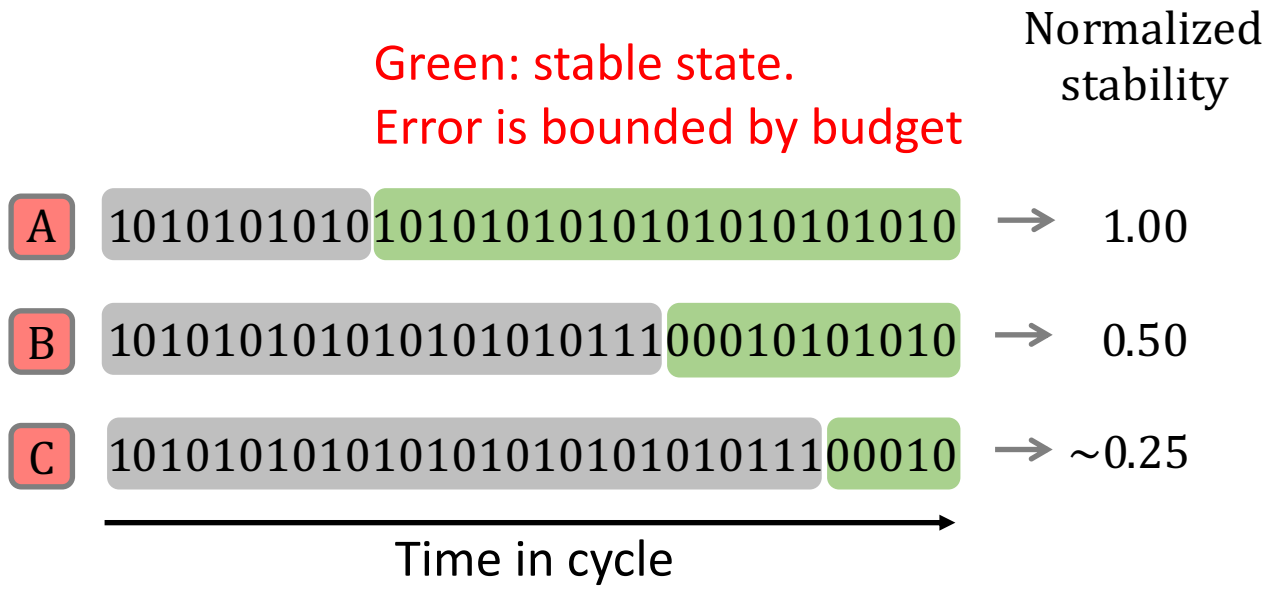


Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Normalized Stability

- Definition
  - How long a bit stream has been **stable** within a pre-defined accuracy budget, normalized to the maximum achievable stable duration

5% error, adapted from approximate computing



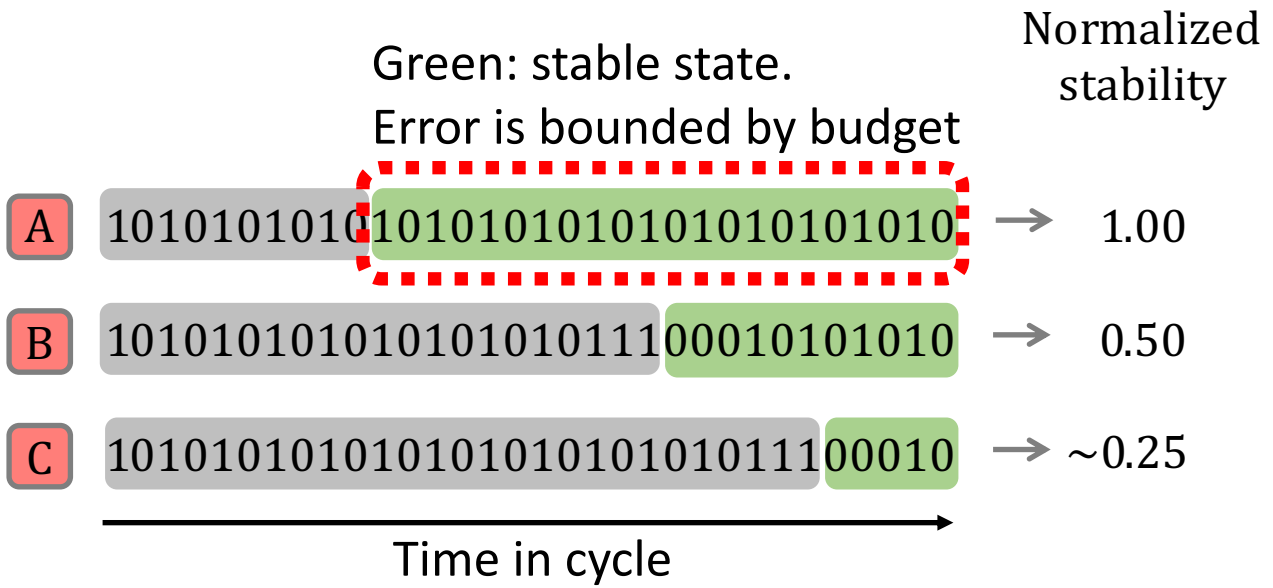
When entering stable state, early termination can be enabled

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Normalized Stability

- Definition
  - How long a bit stream has been stable within a pre-defined accuracy budget, normalized to the **maximum achievable stable duration**

5% error, adapted from approximate computing

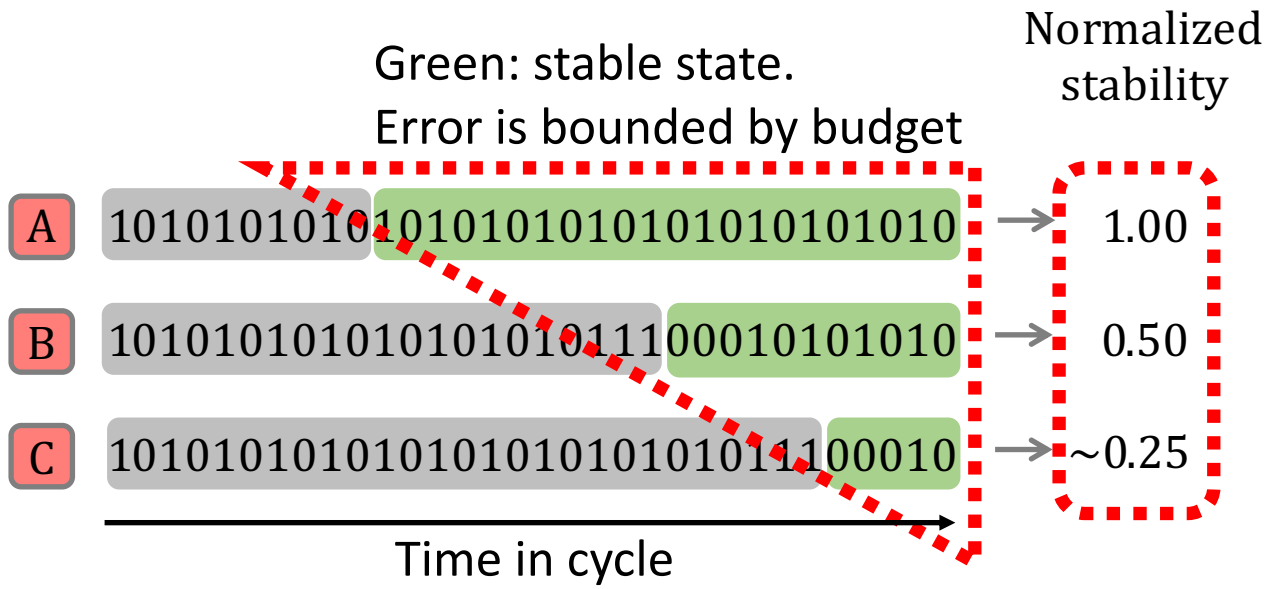


Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Normalized Stability

- Definition
  - How long a bit stream has been stable within a pre-defined accuracy budget, **normalized** to the maximum achievable stable duration

5% error, adapted from approximate computing



A higher value indicates earlier termination

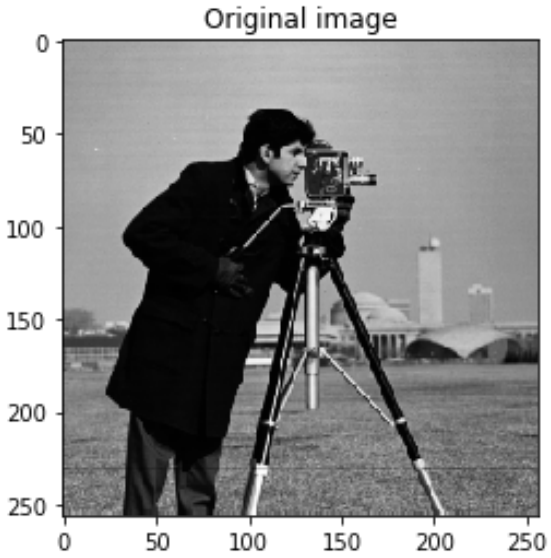
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

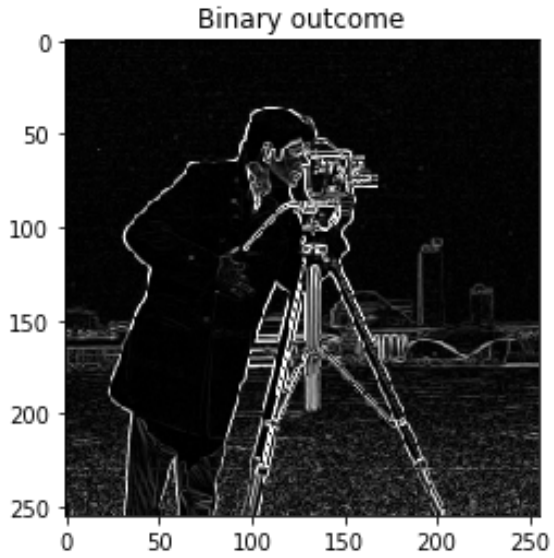
Evaluate Normalized Stability

- Capability
  - Unit level:
    - Identify the competitiveness for ET
  - Application level:
    - Explore design space for ET
    - Predict the timing for ET

Evaluate Normalized Stability

- Target application
  - SC edge detection

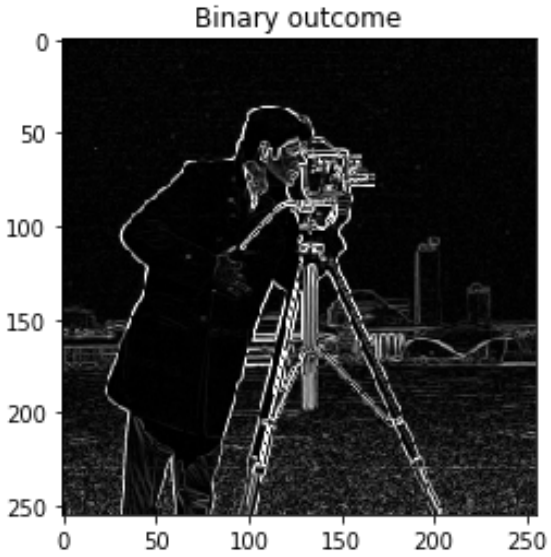


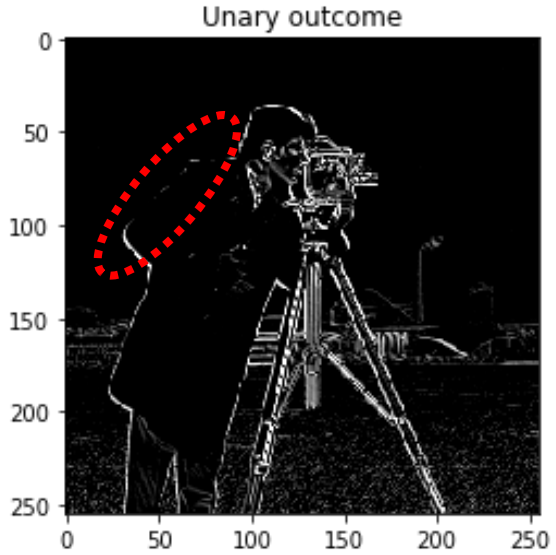


Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

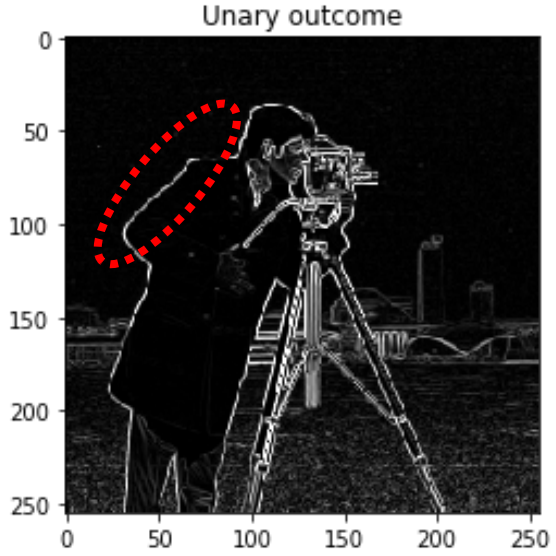
Unit-level Evaluation

- Final results with varying units





Combinational adder



Counter-based adder

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Unit-level Evaluation

- Flux normalized stability (NS)
  - Output-to-input NS for SC units
  - A higher value indicates the ability for earlier termination

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Unit-level Evaluation

- Flux normalized stability (NS)
  - Output-to-input NS for SC units
  - A higher value indicates the ability for earlier termination

Assume input NS = 0.5

Op.	Error budget	Combinational	Counter-based
Add	5%	0.92	0.98
	10%	0.97	0.99

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Unit-level Evaluation

- Flux normalized stability (NS)
  - Output-to-input NS for SC units
  - A higher value indicates the ability for earlier termination
    - Better SC units yield higher flux NS

Assume input NS = 0.5

Op.	Error budget	Combinational	Counter-based
Add	5%	0.92	0.98
	10%	0.97	0.99

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Unit-level Evaluation

- Flux normalized stability (NS)
  - Output-to-input NS for SC units
  - A higher value indicates the ability for earlier termination
    - Better SC units yield higher flux NS
    - Increase error tolerance

Assume input NS = 0.5

Op.	Error budget	Combinational	Counter-based
Add	5%	0.92	0.98
	10%	0.97	0.99

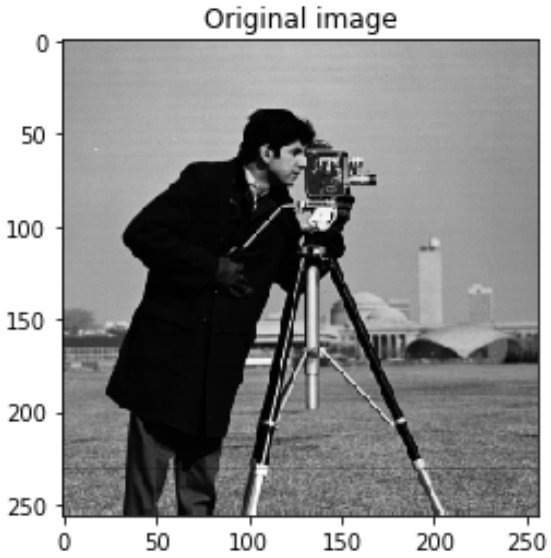
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

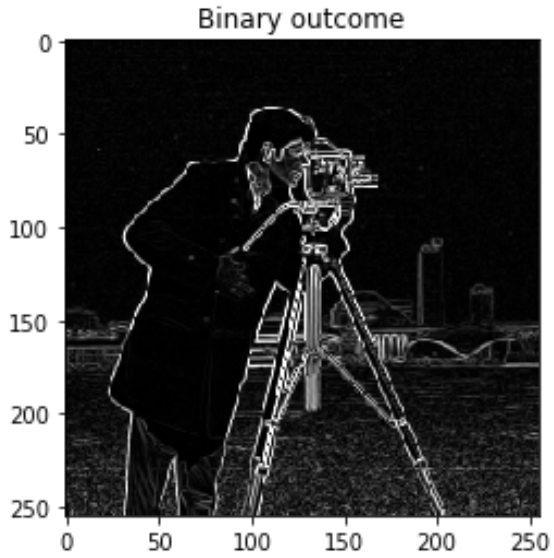
Evaluate Normalized Stability

- Capability
  - Unit level:
    - Identify the competitiveness for ET
  - Application level:
    - Explore design space for ET
    - Predict the timing for ET

Evaluate Normalized Stability

- Target application
  - SC edge detection





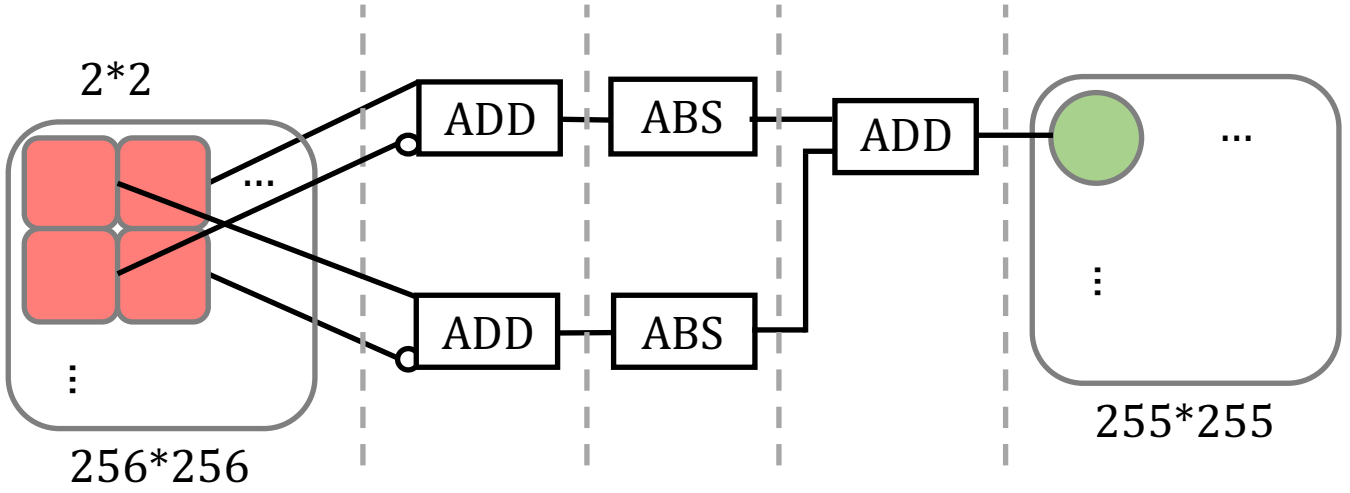
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

- Higher flux NS enables earlier termination

Adder	Error Budget	NS				ET Cycle
Comb.	5%	0.97	0.95	0.20	0.18	838
CNT		0.97	0.97	0.97	0.97	34

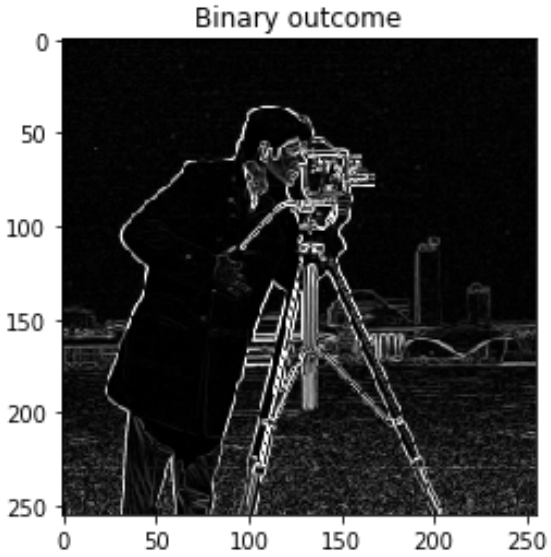
Use better SC units

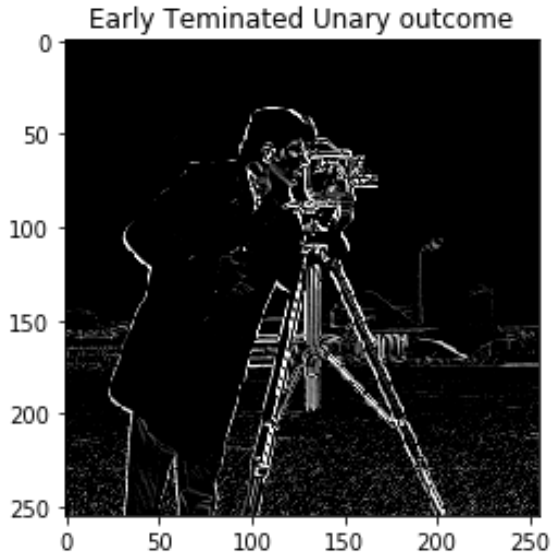


Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

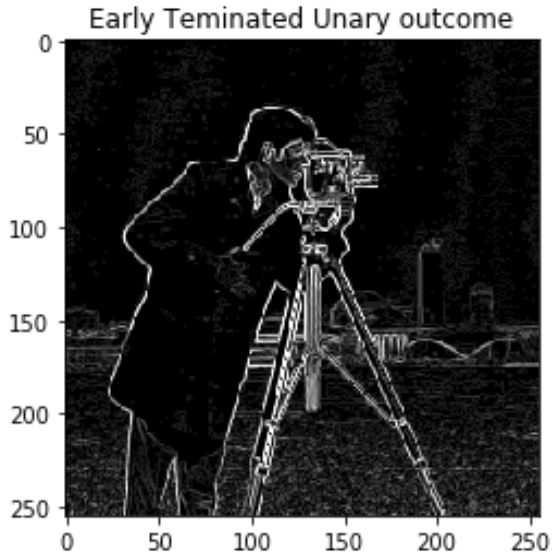
- Better SC units





Combinational adder

Error tolerance 5%

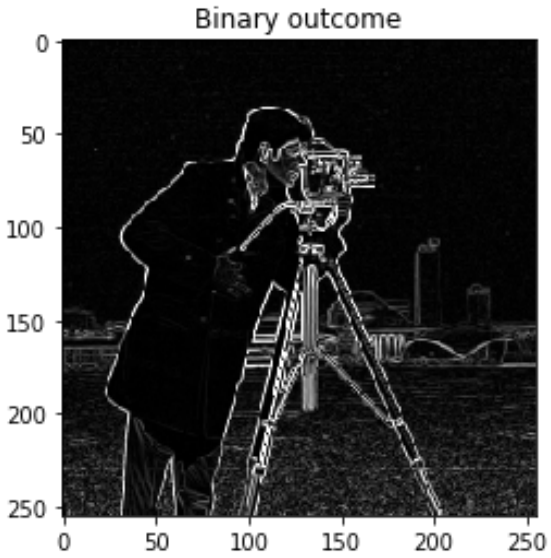


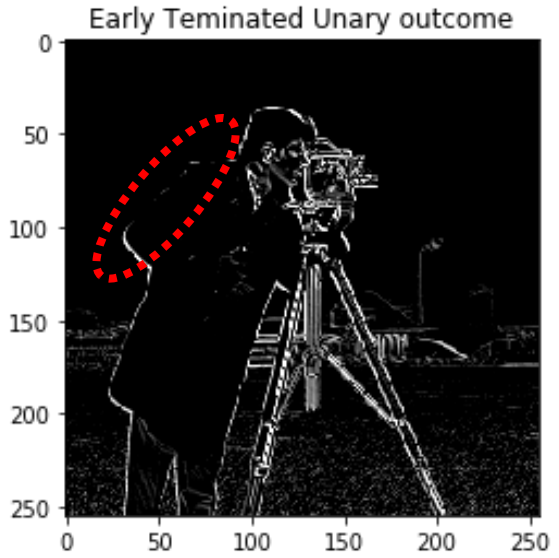
Counter-based adder

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

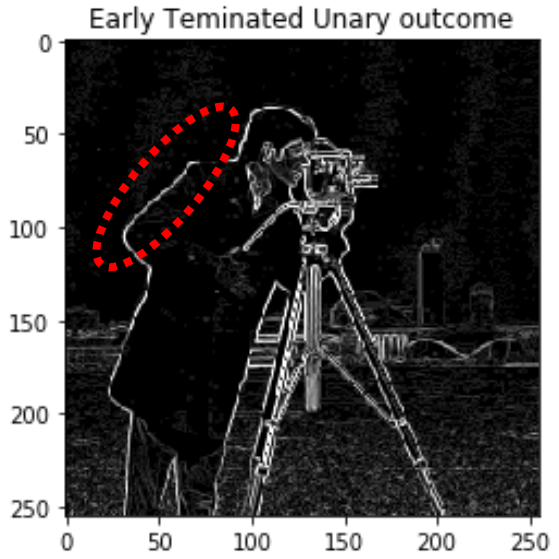
- Better SC units





Combinational adder

Error tolerance 5%



Counter-based adder

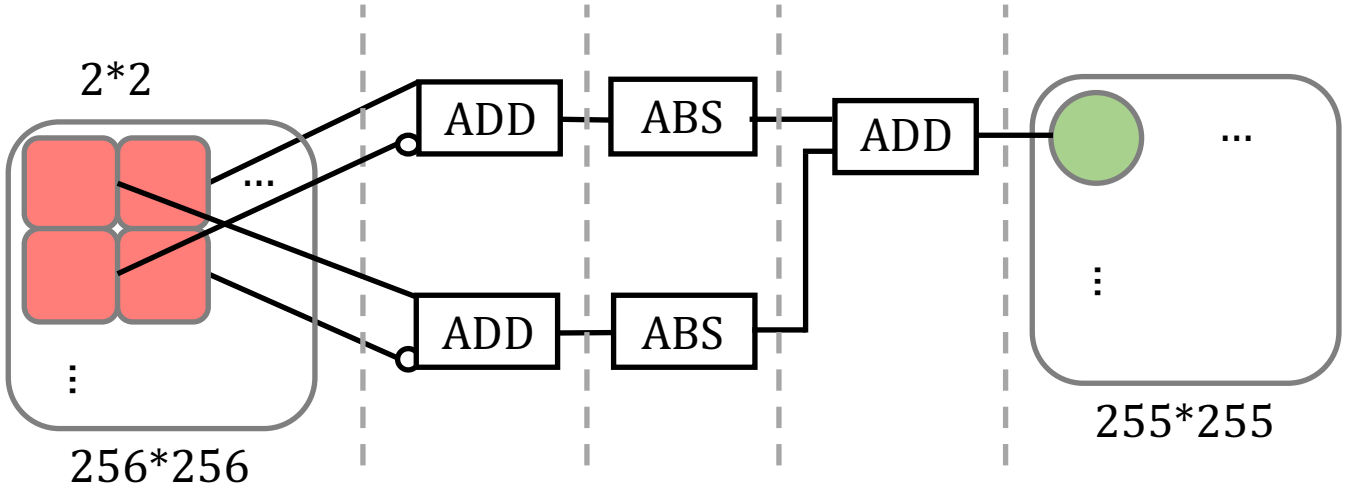
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

- Higher flux NS enables earlier termination

Adder	Error Budget	NS				ET Cycle
Comb.	5%	0.97	0.95	0.20	0.18	838
	10%	0.99	0.98	0.32	0.31	702

Increase error tolerance.

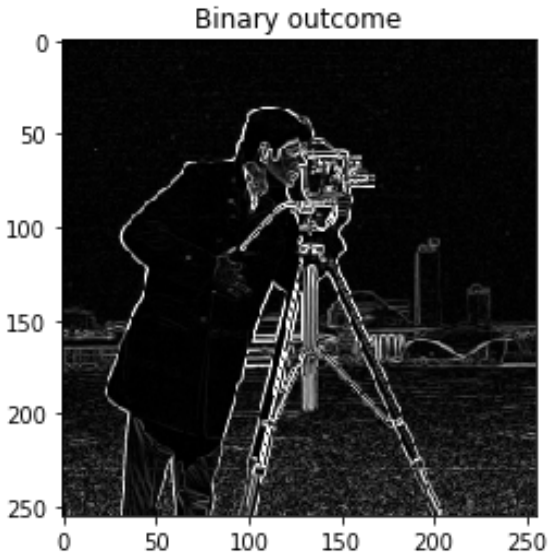


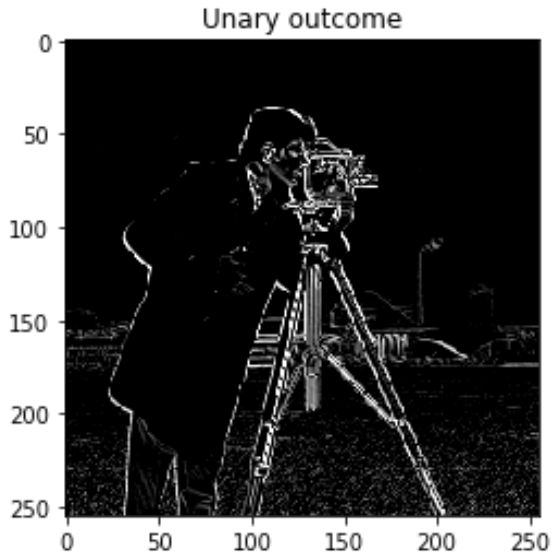
https://github.com/diwu1990/UnarySim/blob/master/sw/test/metric/test_metric_normstability_edge_detect_detail.ipynb

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

- Higher error tolerance

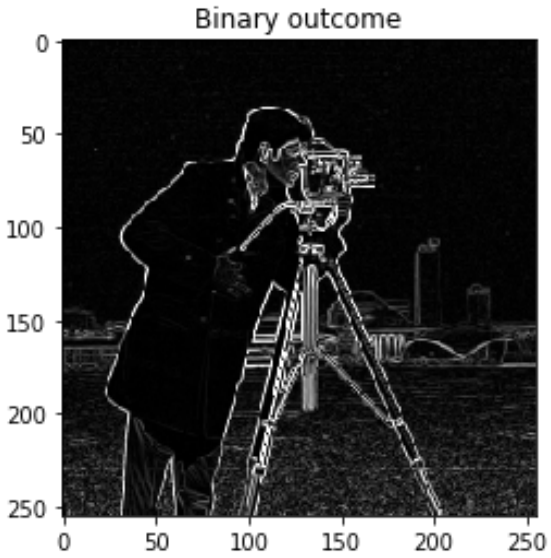


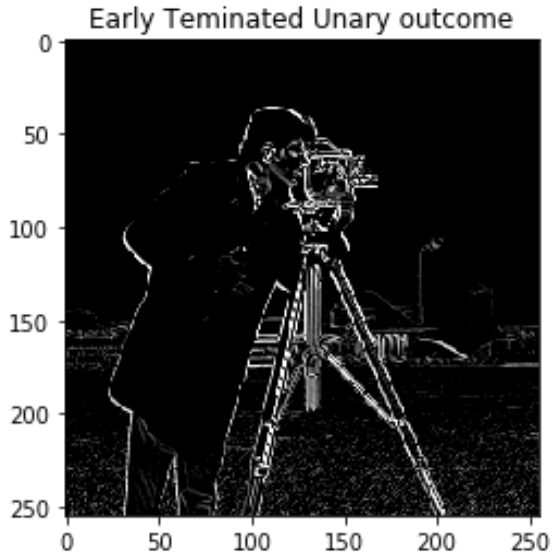


Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

- Higher error tolerance

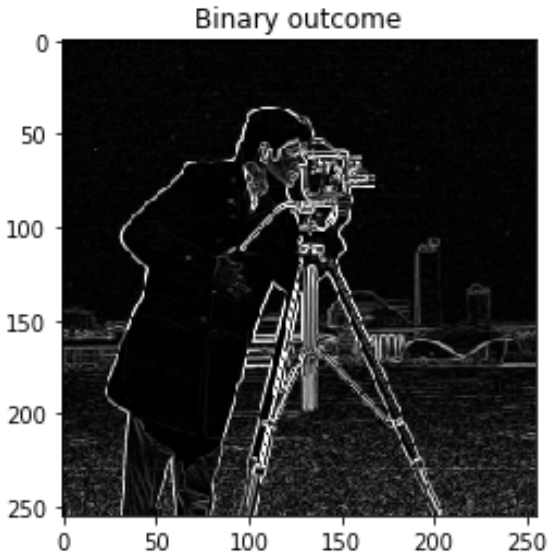


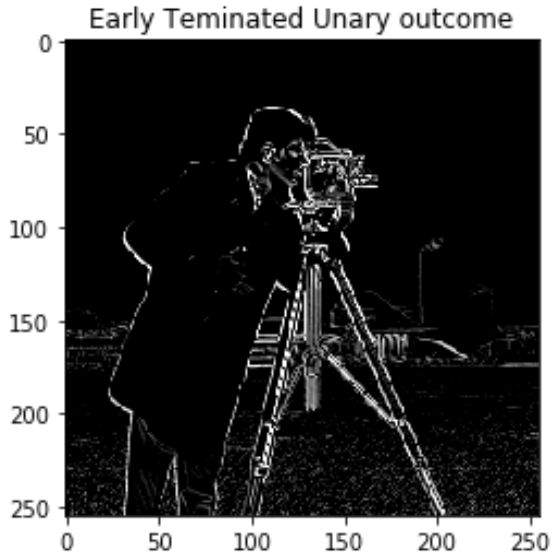


Error tolerance 5%

App-level Evaluation

- Higher error tolerance



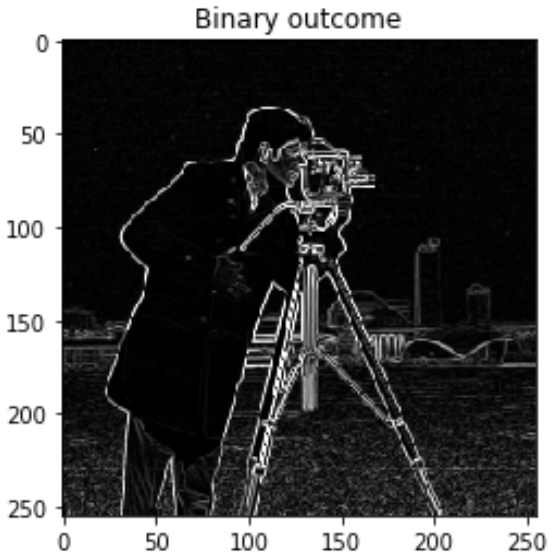


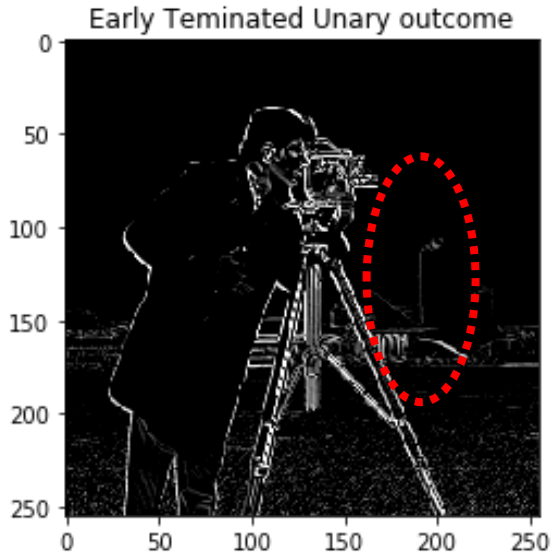
Error tolerance 10%

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

- Higher error tolerance





Error tolerance 10%

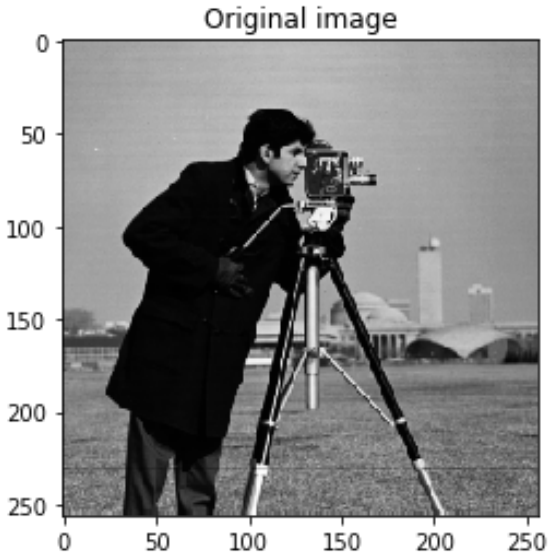
Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

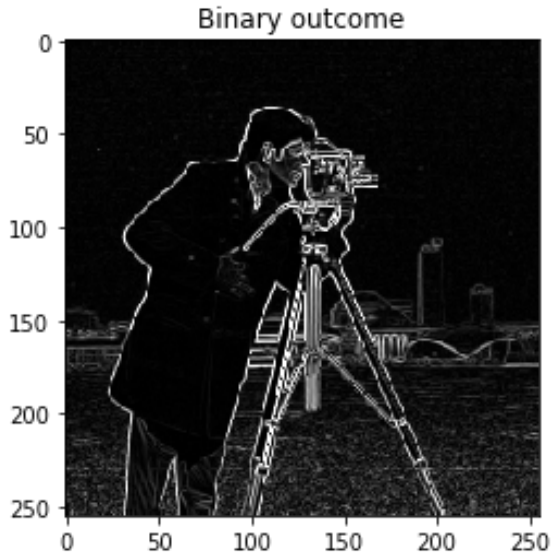
Evaluate Normalized Stability

- Capability
  - Unit level:
    - Identify the competitiveness for ET
  - Application level:
    - Explore design space for ET
    - Predict the timing for ET

Evaluate Normalized Stability

- Target application
  - SC edge detection

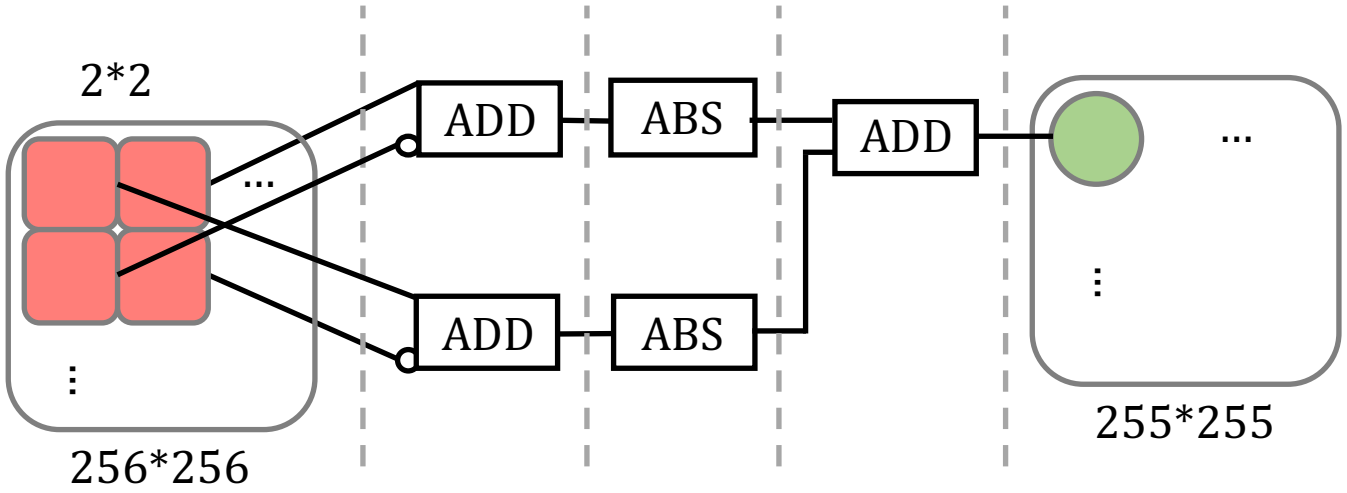




Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

App-level Evaluation

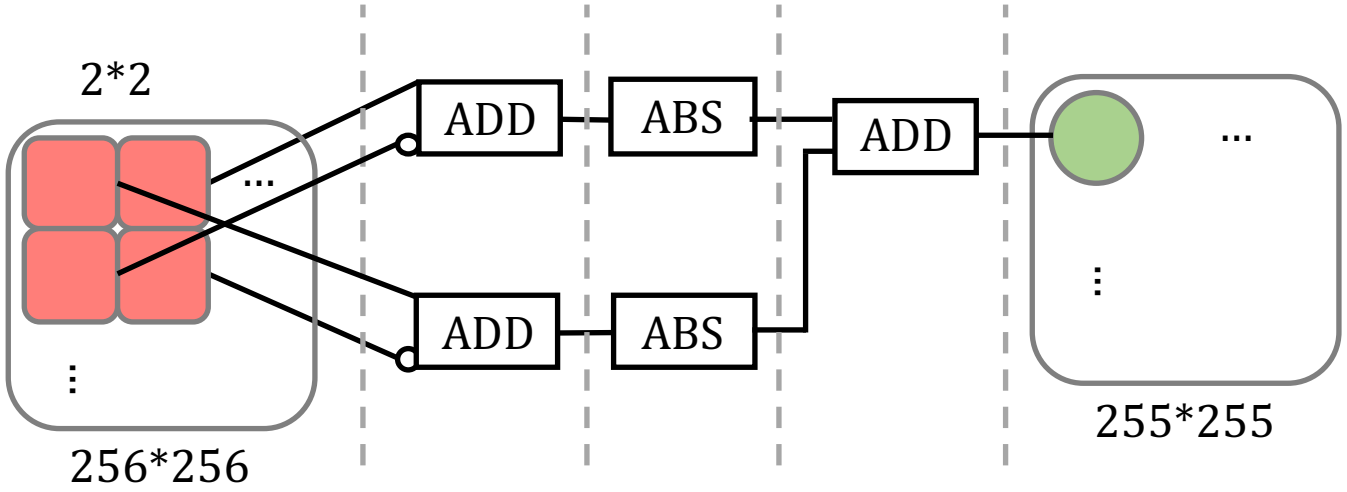
- Prediction of early termination



App-level Evaluation

- Prediction of early termination

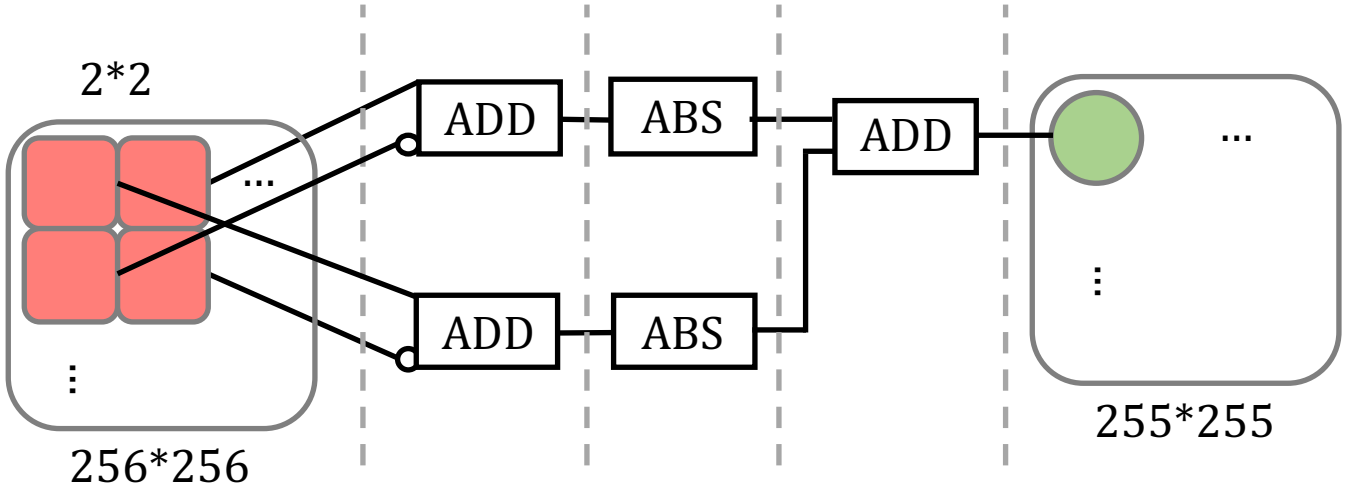
Phase	Output NS	Early termination cycle	Output Error
Train			5%



App-level Evaluation

- Prediction of early termination

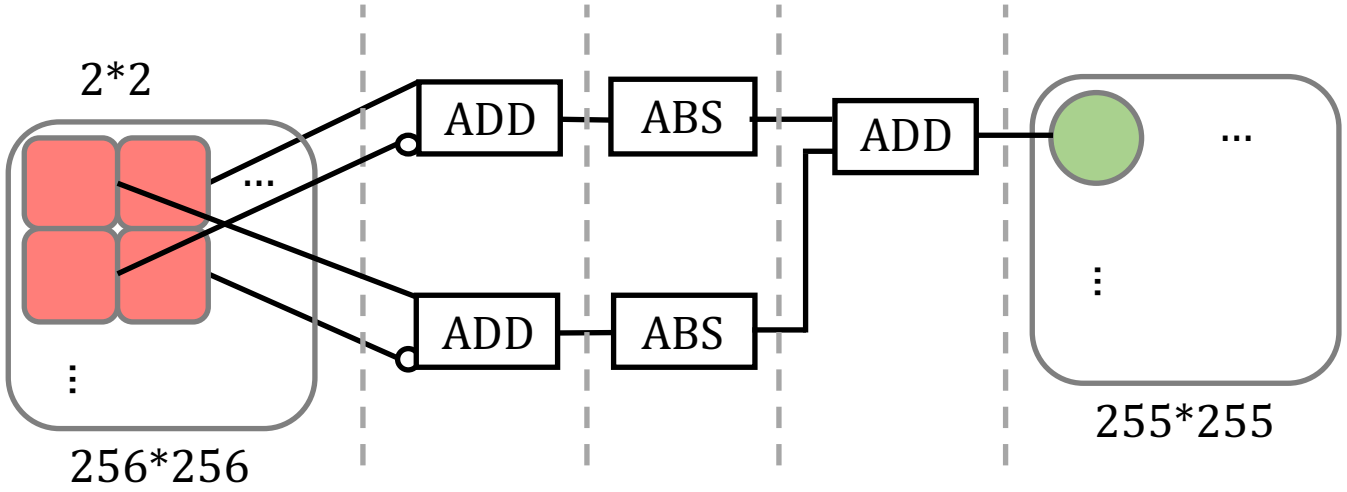
Phase	Output NS	Early termination cycle	Output Error
Train	0.82		5%



App-level Evaluation

- Prediction of early termination

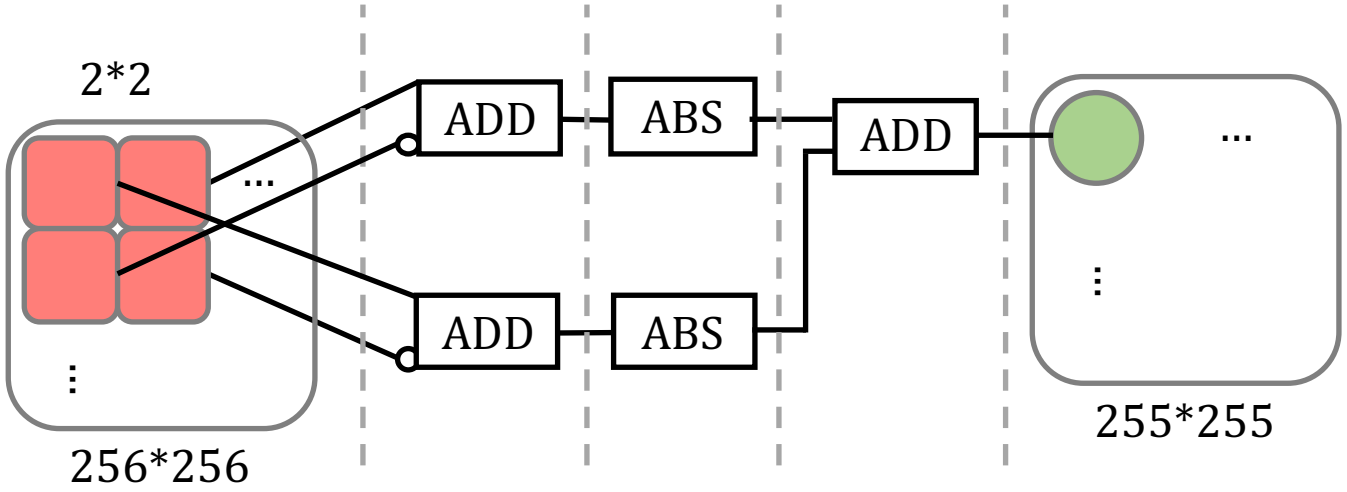
Phase	Output NS	Early termination cycle	Output Error
Train	0.82	184	5%



App-level Evaluation

- Prediction of early termination

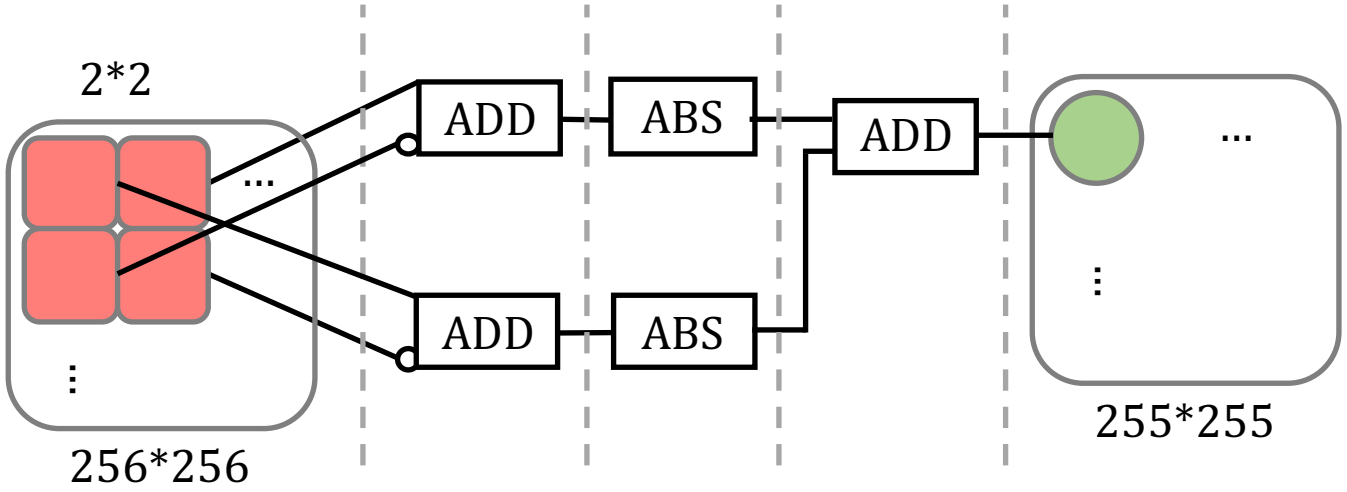
Phase	Output NS	Early termination cycle	Output Error
Train	0.82	184	5%
Test		184	



App-level Evaluation

- Prediction of early termination

Phase	Output NS	Early termination cycle	Output Error
Train	0.82	184	5%
Test		184	2.5%

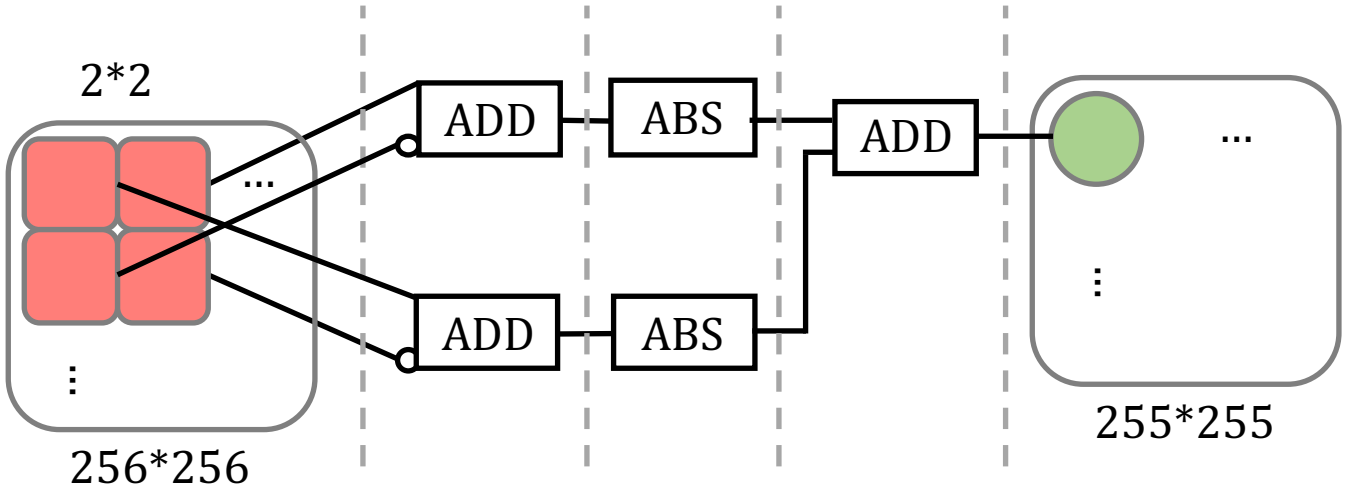


App-level Evaluation

- Prediction of early termination

Phase	Output NS	Early termination cycle	Output Error
Train	0.82	184	5%
Test		184	2.5%

<5%



Implementation

- UnarySim
  - A PyTorch-based simulator for stochastic computing
    - Stream
    - Kernel
    - Metric

- Stability metrics
  - Embedded in UnarySim as a metric component

Di Wu, et al., “uGEMM: Unary Computing Architecture for GEMM Applications”, *ISCA* 2020.

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON

Thank you!
Q & A

Di Wu, Ruokai Yin and Joshua San Miguel
University of Wisconsin-Madison

Department of Electrical and Computer Engineering
UNIVERSITY OF WISCONSIN–MADISON