A Fast and Provably Bounded Failure Analysis of Memory Circuits in High Dimensions

Wei Wu[1], Fang Gong[1], Gengsheng Chen[2], Lei He[1,2]

[1] Electrical Engineering Department, UCLA

http://eda.ee.ucla.edu

[2] School of Microelectronics, Fudan University

Outline

- **Preliminary of High Sigma Analysis and Existing Approaches**
- The Proposed Approach
- Experiment Results
- Conclusions and Future Work

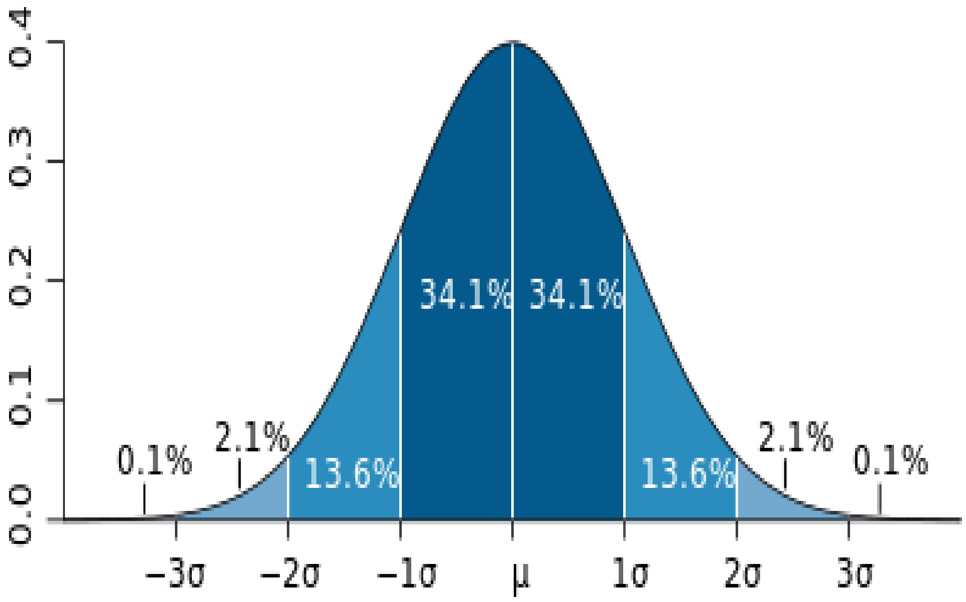
Why Stochastic Modeling and Analysis?

- Ongoing scaling trends
  - Shrinking devices ➔ larger process variations
  - More duplicated circuits: memory, IO, multi-core ➔ higher robustness over variations
- Stochastic modeling and analysis helps to debug circuits in the pre-silicon phase, and enhances yield rate

High Sigma Analysis

- High sigma for analog and custom circuits (IO, memory control, PLL)

n	$F(\mu+n\sigma) - F(\mu-n\sigma)$
1	0.682 689 492 137
2	0.954 499 736 104
3	0.997 300 203 937
4	0.999 936 657 516
5	0.999 999 426 697
6	0.999 999 998 027



*source: normal distribution on Wikipedia

Existing Methods and Limitations

- MC simulation:
  - *time-consuming*
- Traditional Importance Sampling methods
  - *inaccurate* and *unreliable* at high dimension
- Statistical Blockade[1]:
  - Existing classifier is *not robust*
- Other approaches: probability collectives[2], quick yield[3] only work on low dimension problem.

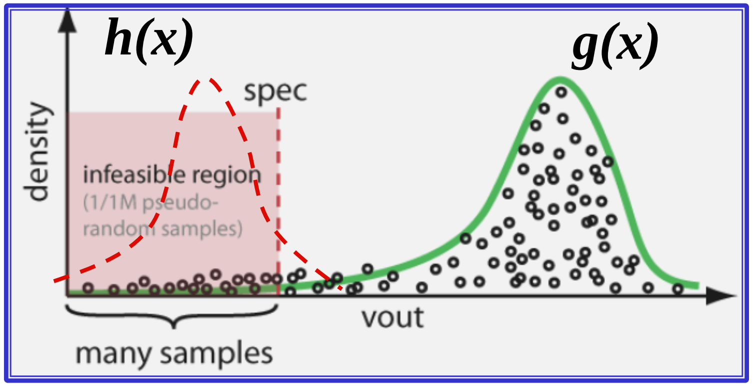
[1] Singhee, A.; Rutenbar, R.A.; , “Statistical Blockade: A Novel Method for Very Fast Monte Carlo Simulation of Rare Circuit Events, and its Application", DATE, 2007.

[2] F. Gong, S. Basir-Kazeruni, L. Dolecek, L. He. “A fast estimation of SRAM failure rate using probability collectives”, ISPD, 2012.

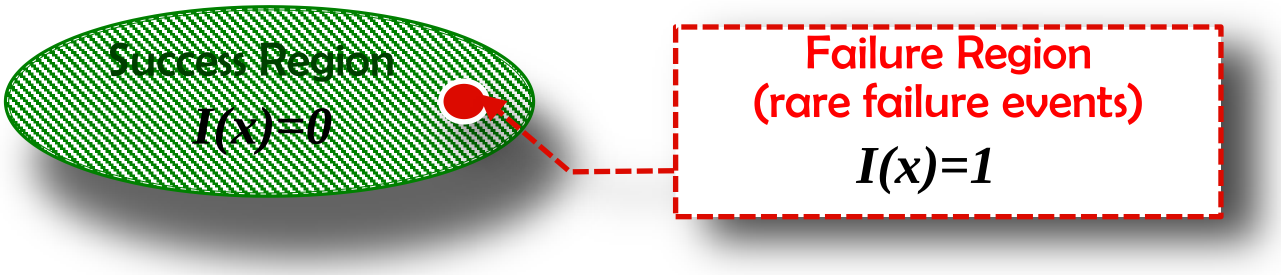
[3] F. Gong, H. Yu, Y. Shi, D. Kim, J. Ren, L. He. “QuickYield: an efficient global-search based parametric yield estimation with performance constraints”, DAC, 2010.

Basic Idea in Importance Sampling

- Importance Sampling
  - Shift sampling distribution towards the failure region.



- Indicator Function



- Probability of rare failure events
  - variable x and its PDF $h(x)$

$$prob(failure) = \int I(x) \cdot h(x)dx = \int I(x) \cdot \frac{h(x)}{g(x)} \cdot g(x)dx$$

  - Likelihood ratio or weights for each sample of x is $h(x)/g(x)$, which is unbounded when $g(x)$ becomes very small under high dimension

Outline

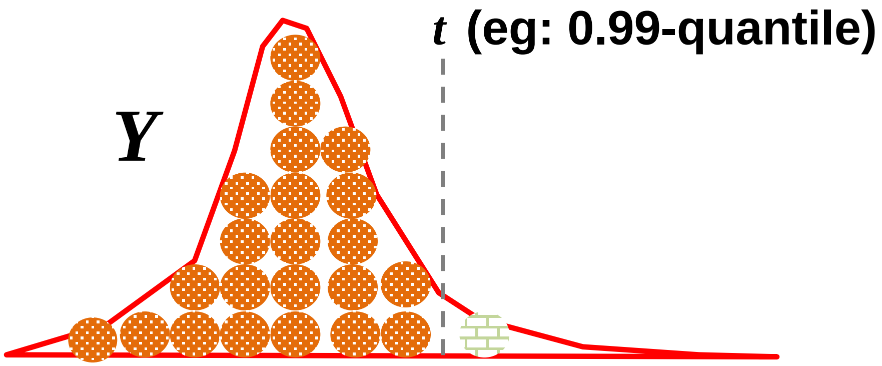
- Preliminary of High Sigma Analysis and Existing Approaches
- **The Proposed Approach**
- Experiment Results
- Conclusions and Future Work

Overview of the Proposed Algorithm

- Three stage algorithm:
  - Build a region **R (eg. 0.99 quantile)**, {Y|Y≥*t*}, which is not so rare, and evaluate the probability of this region, P(Y≥*t*) with MC

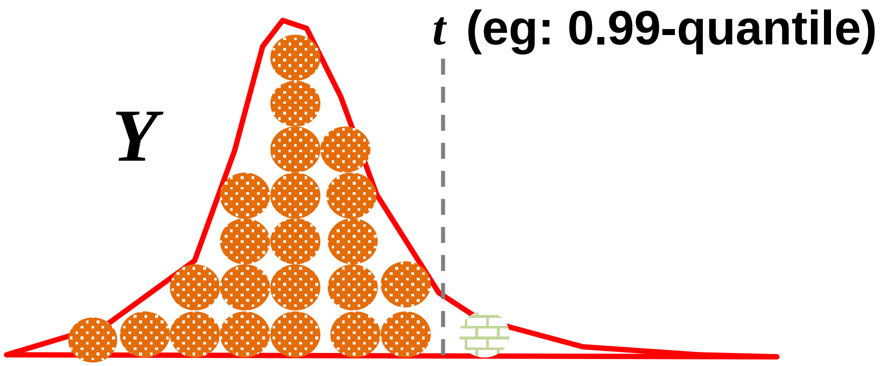
Overview of the Proposed Algorithm

- Three stage algorithm:
  - Build a region **R (eg. 0.99 quantile)**, {Y|Y≥*t*}, which is not so rare, and evaluate the probability of this region, P(Y≥*t*) with MC



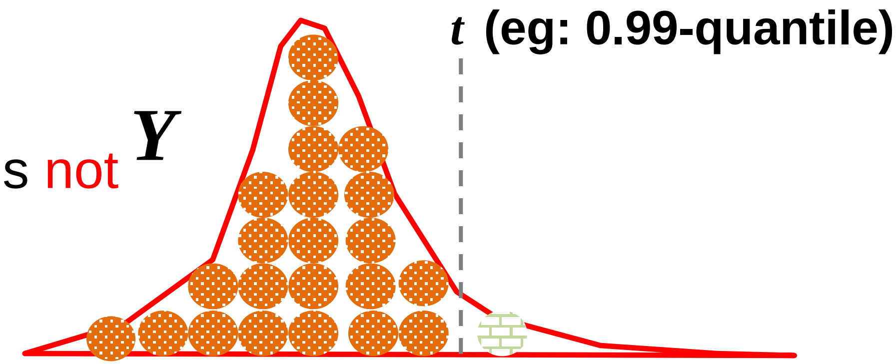
Overview of the Proposed Algorithm

- Three stage algorithm:
  - Build a region **R (eg. 0.99 quantile)**, {Y|Y≥t}, which is not so rare, and evaluate the probability of this region, P(Y≥t) with MC
  - Generate a new distribution Y_t covers **R** and estimate the conditional failure probability: P(Y ≥t_c|Y≥t).

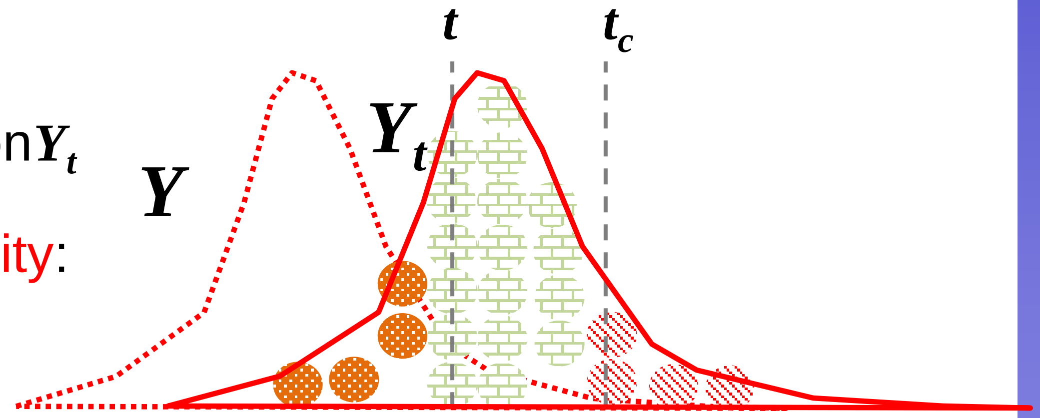


Overview of the Proposed Algorithm

- Three stage algorithm:
  - Build a region **R (eg. 0.99 quantile)**, {Y|Y≥*t*}, which is not so rare, and evaluate the probability of this region, P(Y≥*t*) with MC

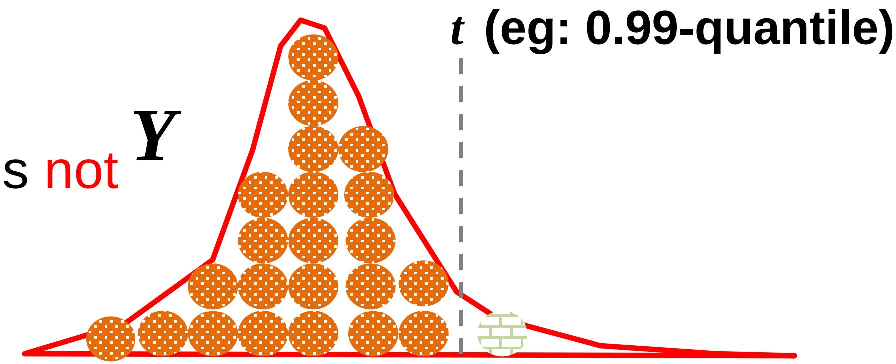


  - Generate a new distribution Y_t covers **R** and estimate the conditional failure probability: $P(Y \geq t_c | Y \geq t)$.

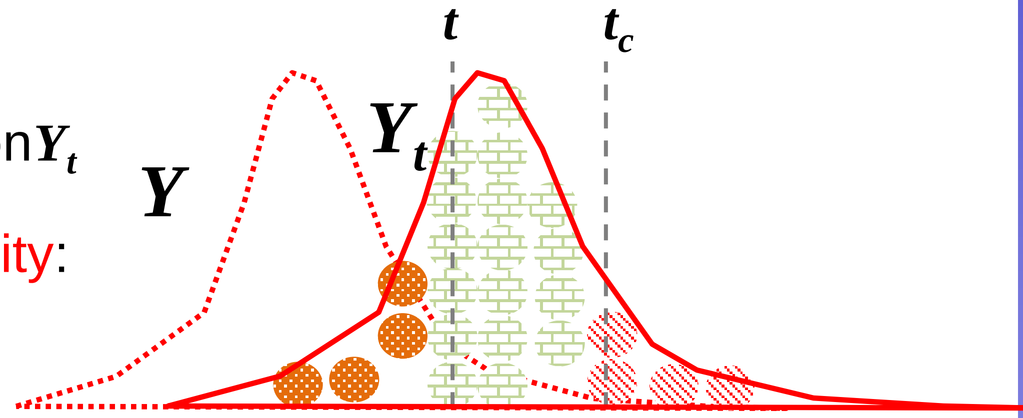


Overview of the Proposed Algorithm

- Three stage algorithm:
  - Build a region **R (eg. 0.99 quantile)**, {Y|Y≥*t*}, which is not so rare, and evaluate the probability of this region, P(Y≥*t*) with MC



  - Generate a new distribution Y_t covers **R** and estimate the conditional failure probability: $P(Y \geq t_c | Y \geq t)$.



  - Failure Probability:

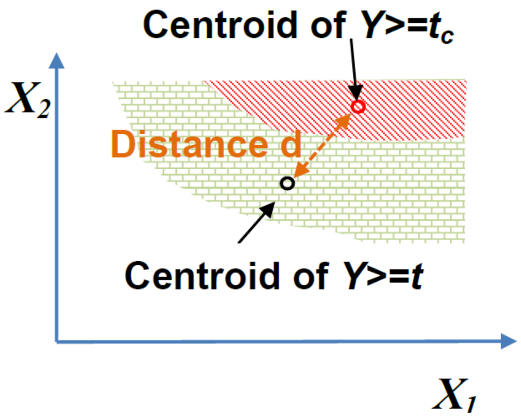
    $P(Y \geq t_c) = P(Y \geq t) * P(Y \geq t_c | Y \geq t)$

Stage2: Choosing Mean and Sigma for Y_t

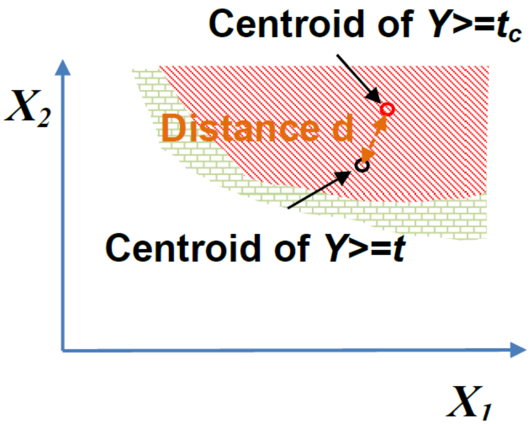
- Stage 2: Generate a new distribution Y_t covers **R** and estimate the conditional failure probability: $P(Y \geq t_c | Y \geq t)$.
  - *mean-shift*: move towards the region with more potential failure.
    e.g. we move the mean to the centroid of **R** in this work
  - *sigma-change*: reshape to dominate the "rare-event" region.

    $$\sigma = max(d, \sigma(Y_t))$$

    to make sure the entire failure region can be properly covered



d is larger than std-dev of f(x)



d is smaller than std-dev of f(x)

Stage3: Evaluation of Conditional Probability

- Failure Probability: $P(Y \geq t_c) = P(Y \geq t) * P(Y \geq t_c | Y \geq t)$

- Conditional Probability is calculated as:

$$P(Y \geq t_c \mid Y \geq t) = \frac{P(Y \geq t_c, Y \geq t)}{P(Y \geq t)} = \frac{P(Y \geq t_c)}{P(Y \geq t)} = \frac{\sum_{i=1}^{N} w(x_i) \cdot I_{\{Y \geq t_c\}}(x_i)}{\sum_{i=1}^{N} w(x_i) \cdot I_{\{Y \geq t\}}(x_i)}$$

$$w(x_i) = \frac{h(x_i)}{g(x_i)}; \quad I_{\{Y \geq t\}}(x_i) = \begin{cases} 0 & if \quad Y(x_i) \notin \{Y \mid Y \geq t\} \\ 1 & if \quad Y(x_i) \in \{Y \mid Y \geq t\} \end{cases}$$

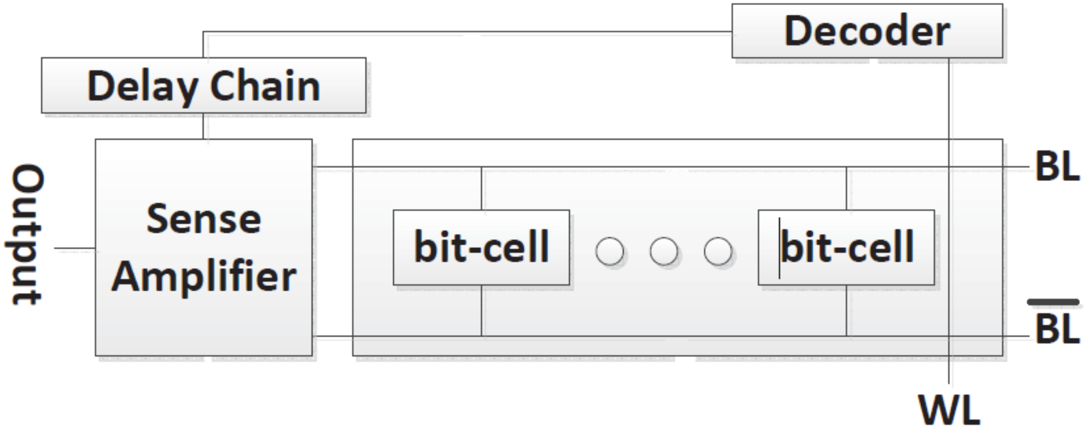
- Boundedness analysis:
  - Upper bound of estimations from classic importance sampling approaches ∞!
  - The estimations of the proposed algorithm are always bounded.

Outline

- Preliminary of High Sigma Analysis and Existing Approaches
- The Proposed Approach
- **Experiment Results**
- Conclusions and Future Work

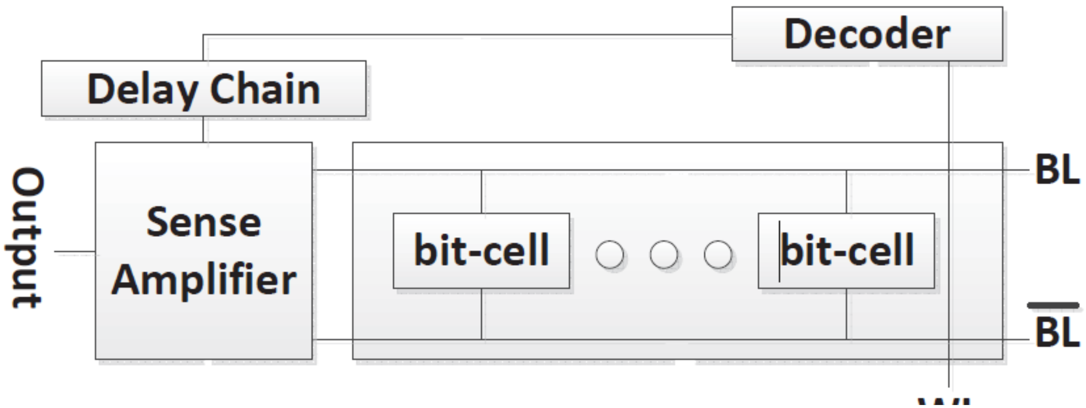
High-Sigma Analysis on a SRAM circuit

- Functional Diagram on an SRAM circuit

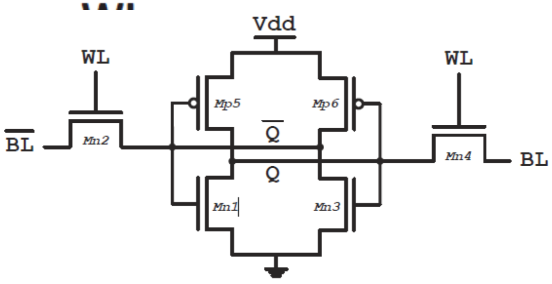


High-Sigma Analysis on a SRAM circuit

- Functional Diagram on an SRAM circuit

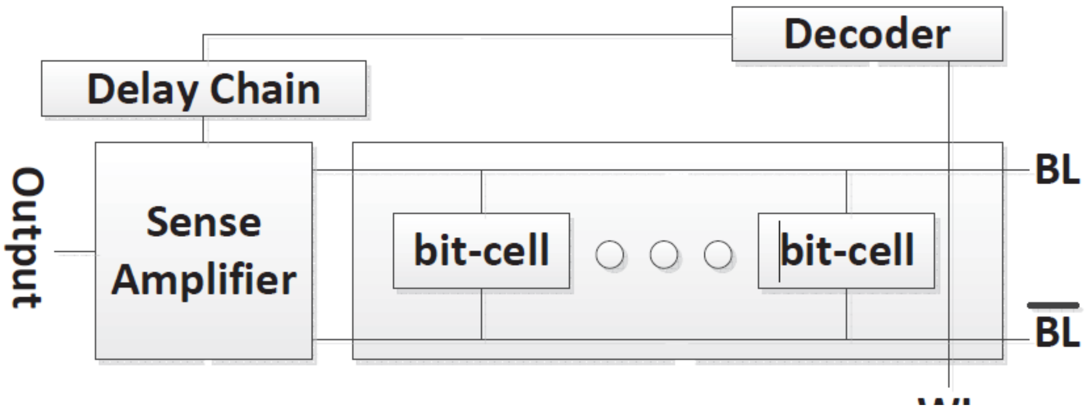


- Test Cases
  - Bit-cell(54 variables, effectively 36 variables as *Mp5* and *Mp6* are OFF)
    - Consider timing failure as the “rare event” of interest.

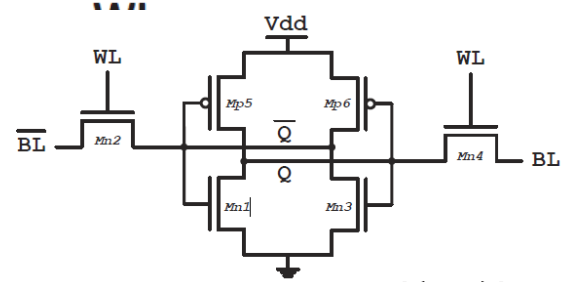


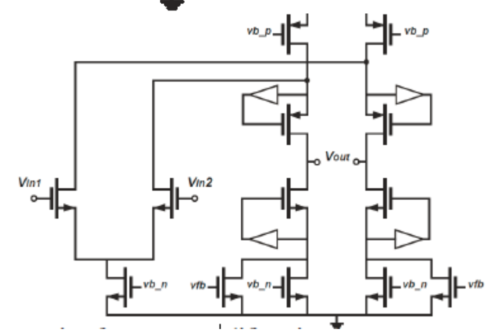
High-Sigma Analysis on a SRAM circuit

- Functional Diagram on an SRAM circuit



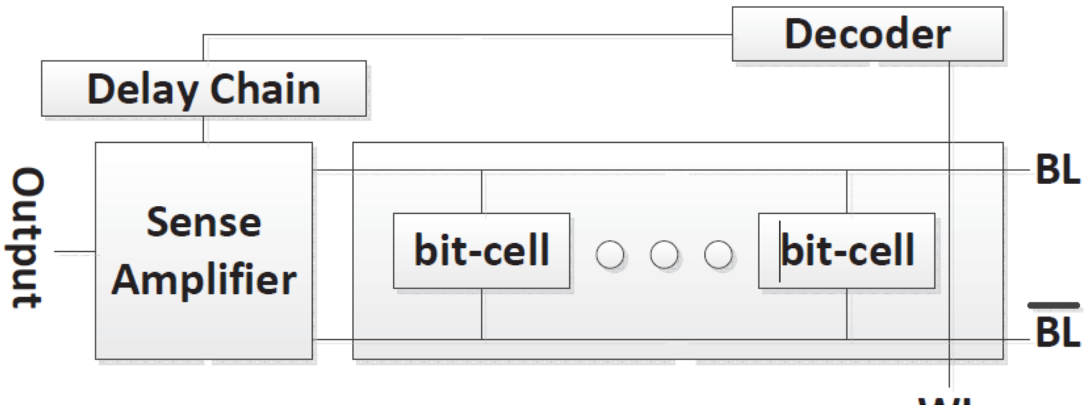
- Test Cases
  - Bit-cell(54 variables, effectively 36 variables as *Mp5* and *Mp6* are OFF)
    - Consider timing failure as the "rare event" of interest.
  - Sense Amplifier (117 variables, with 90 independent variables)
    - Evaluate the gain of the amplifier.
  - Build on 45nm technology node



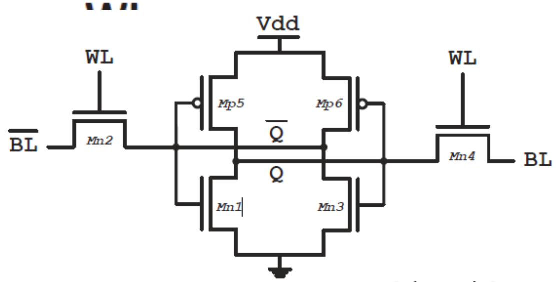


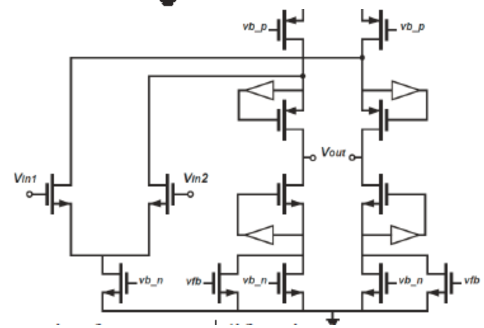
High-Sigma Analysis on a SRAM circuit

- Functional Diagram on an SRAM circuit



- Test Cases
  - Bit-cell(54 variables, effectively 36 variables as *Mp5* and *Mp6* are OFF)
    - Consider timing failure as the “rare event” of interest.
  - Sense Amplifier (117 variables, with 90 independent variables)
    - Evaluate the gain of the amplifier.
  - Build on 45nm technology node





SRAM bit-cell circuit

- Experiment results with 90% confidence level on target accuracy:

	MC	SS	SB	HDIS
Failure rate	2.413E-05 (0%)	28415E-05 (+17.7%)	2.7248E-05 (+12.9%)	2.4949E-05 (+3.39%)
# of simulations (x1000)	4600 (1150X)	20 (5X)	816 (204X)	4 (1X)

MC: Monte Carlo, **SS**: Spherical Sampling, **SB**: Statistical Blockade, **HDIS**: the proposed high-dimensional importance sampling

Runtime of 1000 simulations: ~ 5 mins.

SRAM bit-cell circuit

- Experiment results with 90% confidence level on target accuracy:

	MC	SS	SB	HDIS
Failure rate	2.413E-05 (0%)	28415E-05 (+17.7%)	2.7248E-05 (+12.9%)	2.4949E-05 (+3.39%)
# of simulations (x1000)	4600 (1150X)	20 (5X)	816 (204X)	4 (1X)

- HDIS is the most accurate approach with 3.39% failure rate estimation error.

MC: Monte Carlo, **SS**: Spherical Sampling, **SB**: Statistical Blockade, **HDIS**: the proposed high-dimensional importance sampling

Runtime of 1000 simulations: ~ 5 mins.

SRAM bit-cell circuit

- Experiment results with 90% confidence level on target accuracy:

	MC	SS	SB	HDIS
Failure rate	2.413E-05 (0%)	28415E-05 (+17.7%)	2.7248E-05 (+12.9%)	2.4949E-05 (+3.39%)
# of simulations (x1000)	4600 (1150X)	20 (5X)	816 (204X)	4 (1X)

- HDIS is the most accurate approach with 3.39% failure rate estimation error.
  - The performance of SS is acceptable because that's not actually a real high dimensional circuit. (only part of the transistors operations during the SRAM reading)

MC: Monte Carlo, **SS**: Spherical Sampling, **SB**: Statistical Blockade, **HDIS**: the proposed high-dimensional importance sampling

Runtime of 1000 simulations: ~ 5 mins.

SRAM bit-cell circuit

- Experiment results with 90% confidence level on target accuracy:

	MC	SS	SB	HDIS
Failure rate	2.413E-05 (0%)	28415E-05 (+17.7%)	2.7248E-05 (+12.9%)	2.4949E-05 (+3.39%)
# of simulations (x1000)	4600 (1150X)	20 (5X)	816 (204X)	4 (1X)

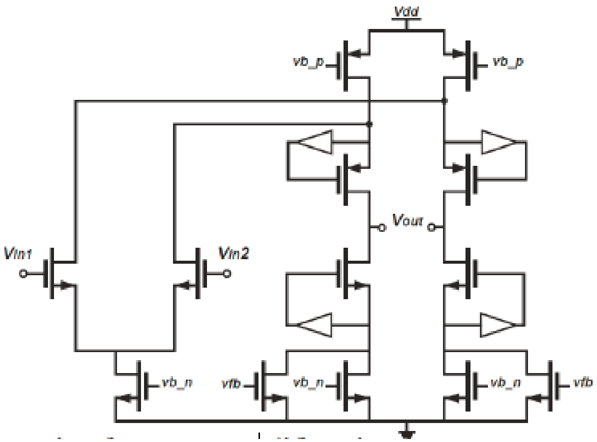
- HDIS is the most accurate approach with 3.39% failure rate estimation error.
  - The performance of SS is acceptable because that's not actually a real high dimensional circuit. (only part of the transistors operations during the SRAM reading)
- It is also the most efficient one with 1150X speedup on MC method

MC: Monte Carlo, **SS**: Spherical Sampling, **SB**: Statistical Blockade, **HDIS**: the proposed high-dimensional importance sampling

Runtime of 1000 simulations: ~ 5 mins.

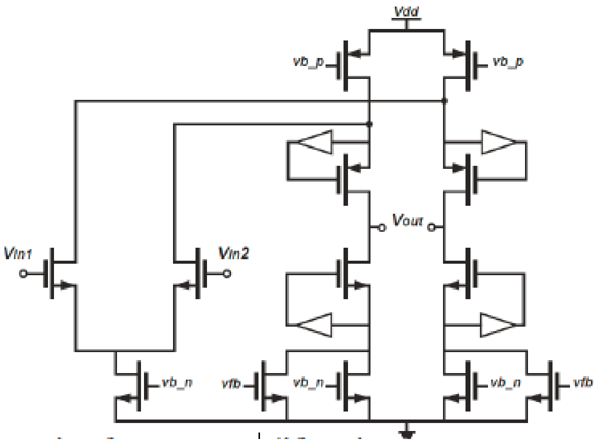
Sense Amplifier circuit

- A circuit with larger number of process variables
- Failure probability

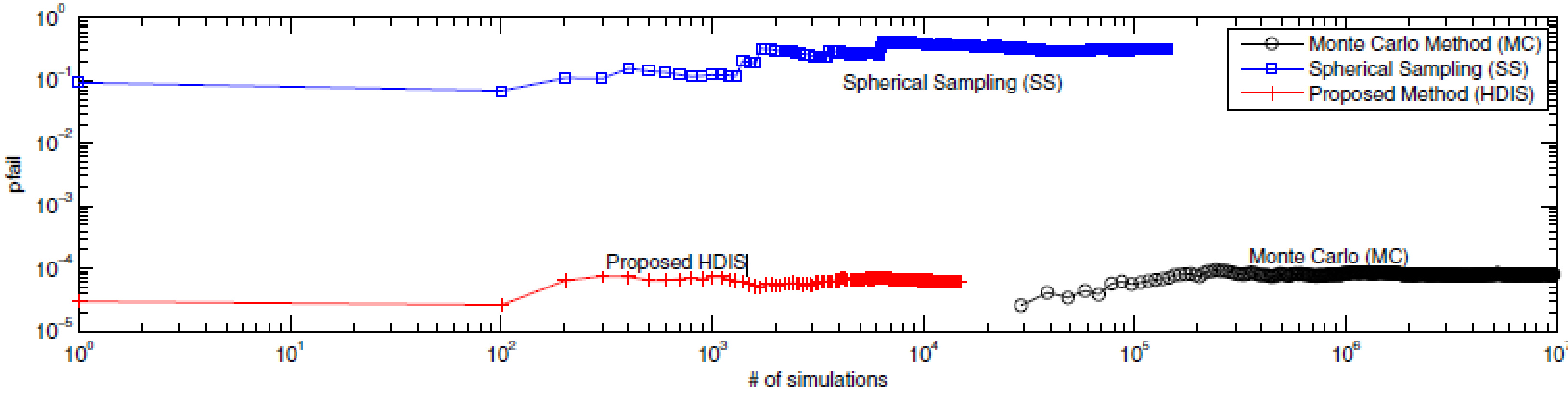


Sense Amplifier circuit

- A circuit with larger number of process variables

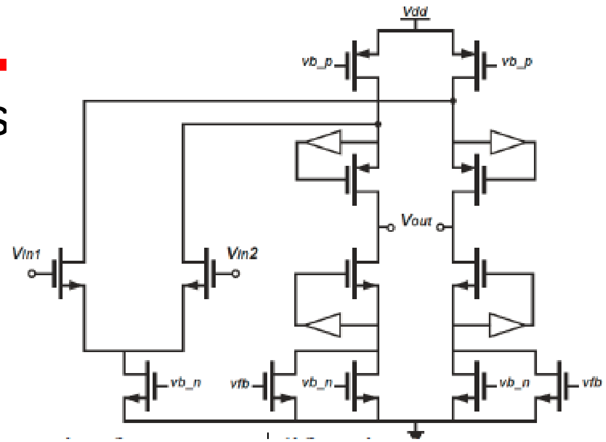


- Failure probability

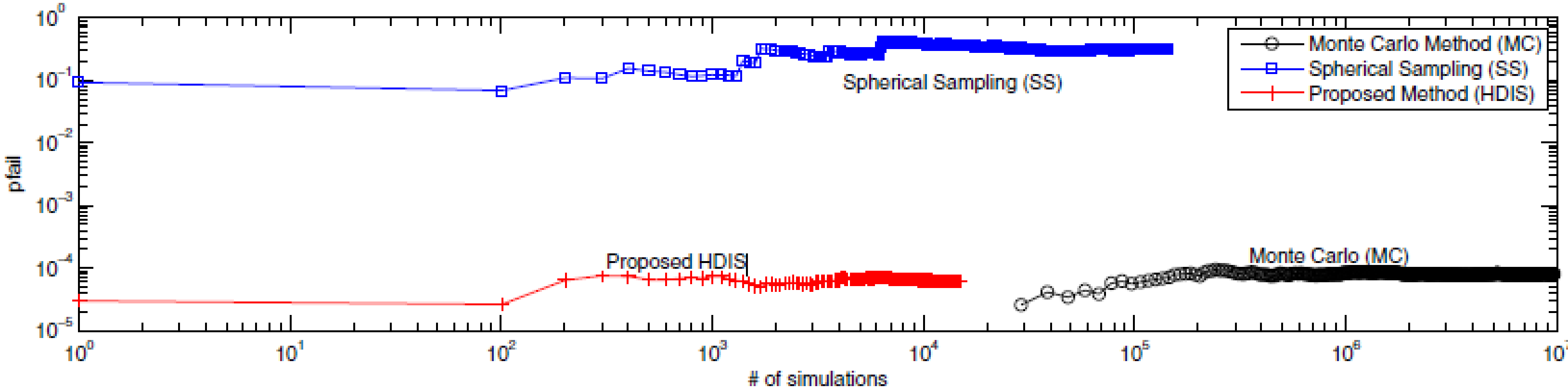


Sense Amplifier circuit

- A circuit with larger number of process variables



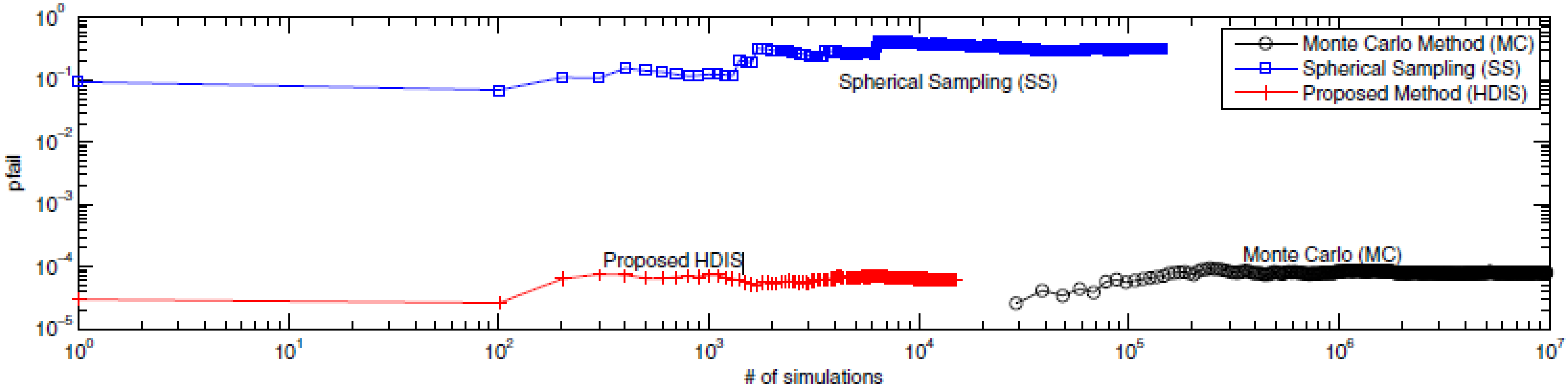
- Failure probability



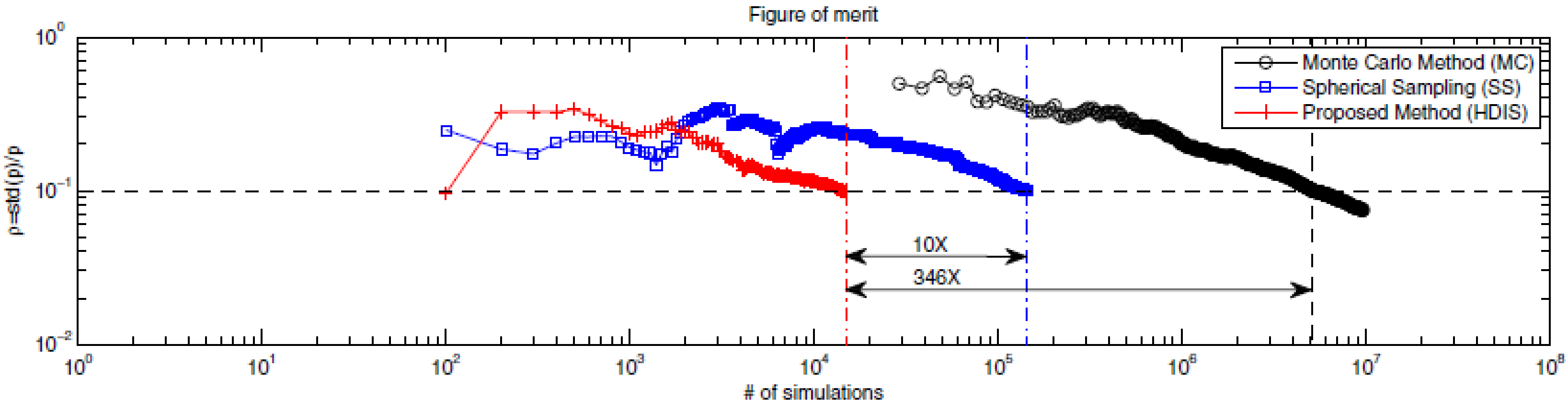
  - The classifier in Statistical Blockade (SB) is not blocking any samples. So the efficiency of SB is degraded to the same as MC.
  - The Spherical sampling is converging to a totally wrong failure rate.

Sense Amplifier circuit

- Failure probability



- Figure of Merit (demonstrate the fast converging rate of HDIS)



Sense Amplifier circuit

- Evaluation on different failure probabilities:

Target failure probability		Monte Carlo (MC)	Spherical Sampling (SS)	Proposed Method (HDIS)
8e-3 (2.6 sigma)	prob:(failure)	8.136e-4	0.2603	7.861e-3 (3.4%)
	#sim. runs	4.800e+4 (24X)	16000 (8X)	2000
8e-4 (3.3 sigma)	prob:(failure)	8.044e-4	0.2541	8.787e-4 (9.2%)
	#sim. runs	4.750e+5 (36X)	8.330e+4 (6.4X)	1.300e+4
8e-5 (3.96 sigma)	prob:(failure)	8.089e-5	0.3103	8.186e-5 (1.2%)
	#sim. runs	5.156e+6 (346X)	1.430e+5 (10X)	1.500e+4

- The accuracy of HDIS agrees with MC on different failure probabilities.
- The efficiency is also consistent under these three cases.

Conclusions and Future Work

- We have proposed a failure probability analysis algorithm, where the failure probability is proved to be always bounded.
- Experiments demonstrated up to 1150X speedup over MC and less than 10% estimation error, while other approaches failed to capture the correct failure rate.
- The proposed algorithm uses mean-shifting, which may be invalid for multiple failure regions. This will be fixed in the future.

Q&A

Thank you!

Address comments to lhe@ee.ucla.edu

Source of process variations

- 9 variables to model the variations in one CMOS transistor

Variable Name	σ/μ	Unit
Flat-band Voltage (V_{fb})	0.1	V
Gate Oxide Thickness (t_{ox})	0.05	m
Mobility (μ_0)	0.1	(m^2/Vs)
Doping concentration at depletion (N_{dep})	0.1	(cm^{-3})
Channel-length offset (ΔL)	0.05	m
Channel-width offset (ΔW)	0.05	m
Source/drain sheet resistance (R_{sh})	0.1	(Ohm/mm^2)
Source-gate overlap unit capacitance (C_{gso})	0.1	(F/m)
Drain-gate overlap unit capacitance (C_{gdo})	0.1	(F/m)