A Fast Estimation of SRAM Failure Rate Using Probability Collectives

Fang Gong
Electrical Engineering Department, UCLA
http://www.ee.ucla.edu/~fang08

Collaborators: Sina Basir-Kazeruni, Lara Dolecek, Lei He
{fang08, sinabk, dolecek, lhe}@ee.ucla.edu

Outline

- Background
- Proposed Algorithm
- Experiments
- Extension Work

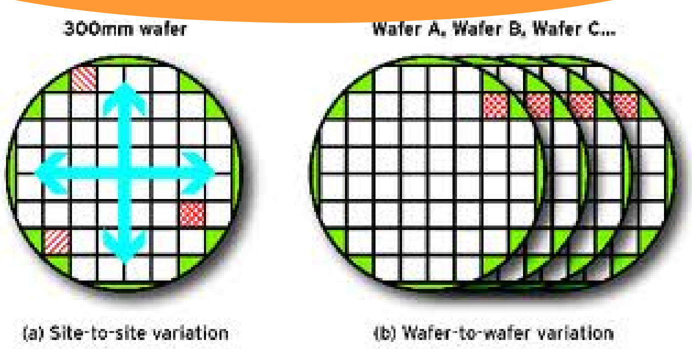
Background

Variations

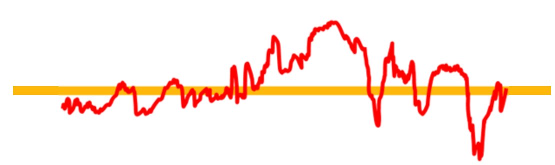
Static Variation

Dynamic Variation

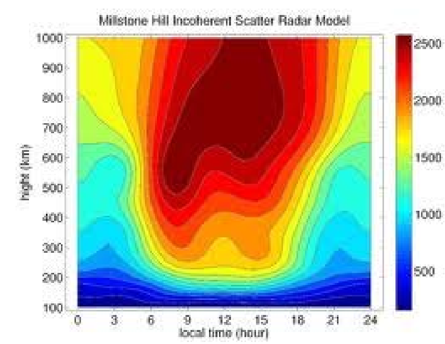
Process Variation



Voltage Variation



Temperature Variation



Rare Failure Event

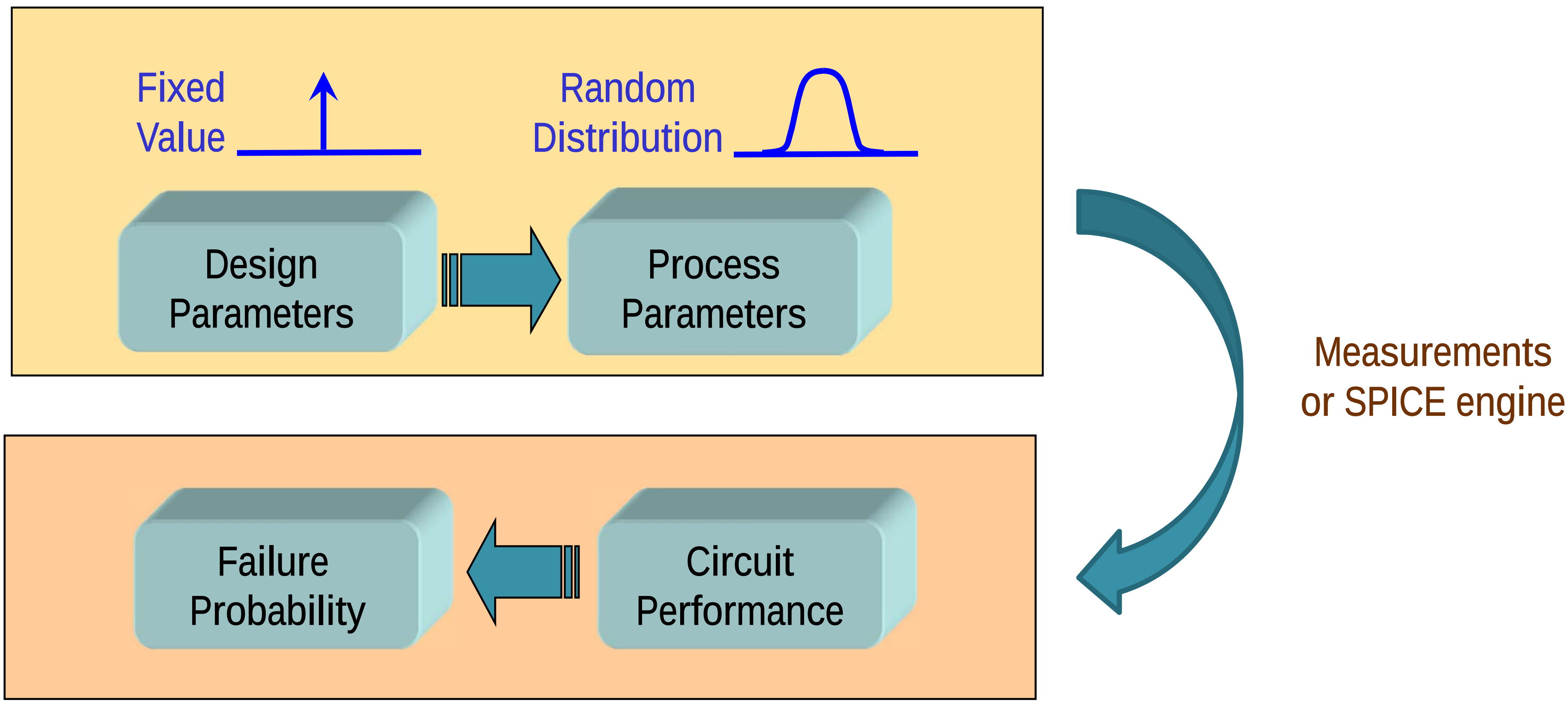
- Rare failure event exists in highly replicated circuits:
  - SRAM bit-cell, sense amplifier, delay chain and etc.
  - Repeat million times to achieve high capacity.
  - Process variation lead to statistical behavior of these circuits.

- Need extremely low failure probability:
  - Consider 1Mb SRAM array including 1 million bit-cells, and we desire 99% yield for the array*:
    - ➔99.999999% yield requirement for bit-cells.
  - Failure probability of SRAM bit-cell should < 1e-8!
  - Circuit failure becomes a "rare event".

* Source: Amith Singhee, IEEE transaction on Computer Aided Design (TCAD), Vol. 28, No. 8, August 2009

Problem Formulation

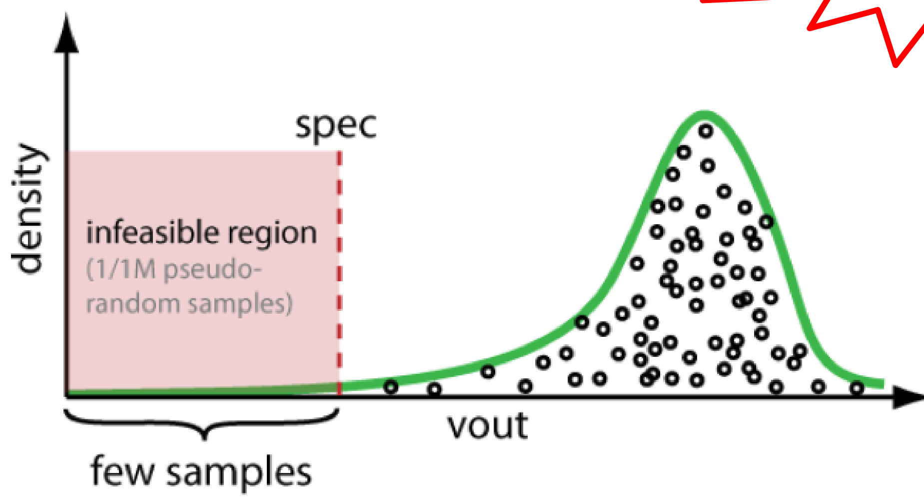
- Given Input:
  - Variable Parameters ➔ probabilistic distributions;
  - Performance constraints;
- Target Output:
  - Find the percentage of circuit samples that fail the performance constraints.



Monte Carlo Method for Rare Events

- Required time in order to achieve 1% relative error
- Assumes 1000 SPICE simulations per second!

Rare Event Probability	Simulation Runs	Time
1e-3	1e+7	16.7mins
1e-5	1e+9	1.2 days
1e-7	1e+11	116 days
1e-9	1e+13	31.7 years

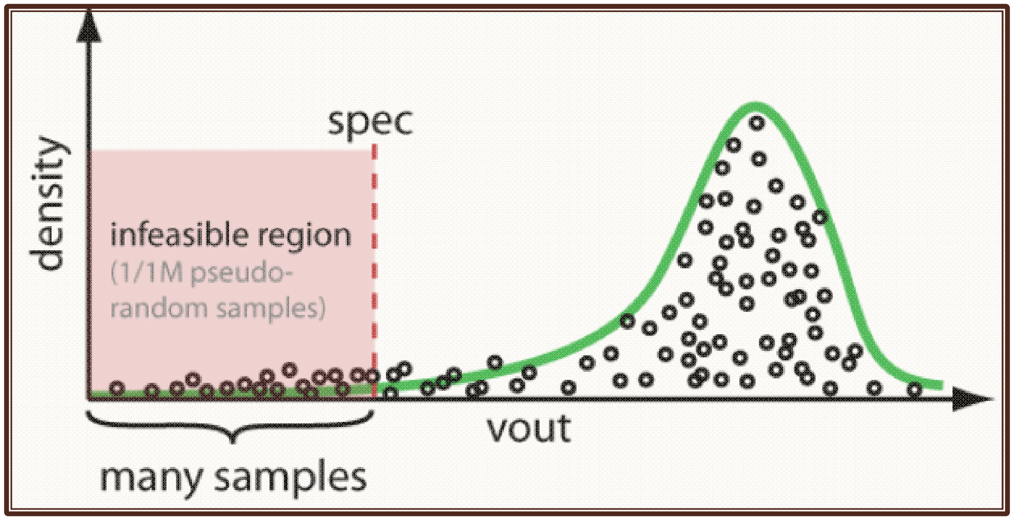


Monte Carlo method for rare events

(Courtesy: Solido Design Automation)

Importance Sampling

- Basic Idea:
  - Add more samples in the failure or infeasible region.



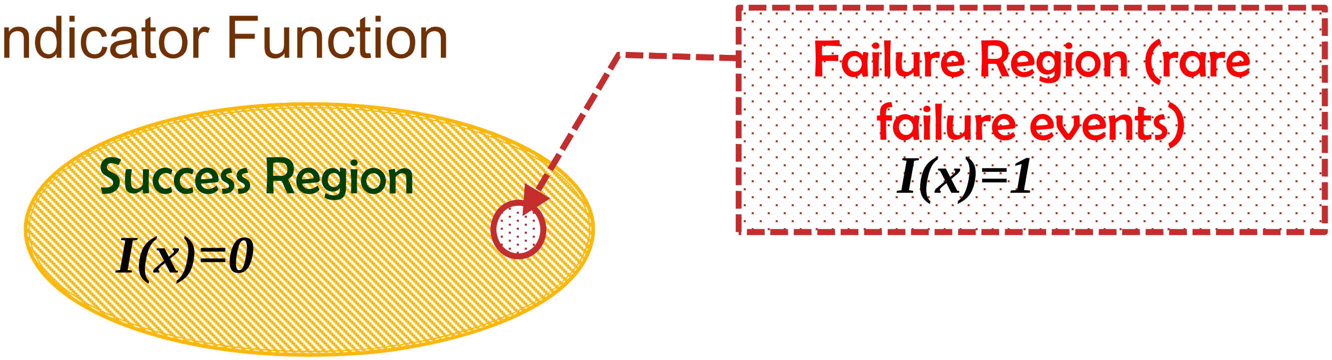
Importance Sampling Method*

- How to do so?
  - IS changes the sampling distribution so that rare events become “less-rare”.

*Courtesy: Solido Design Automation

Mathematic Formulation

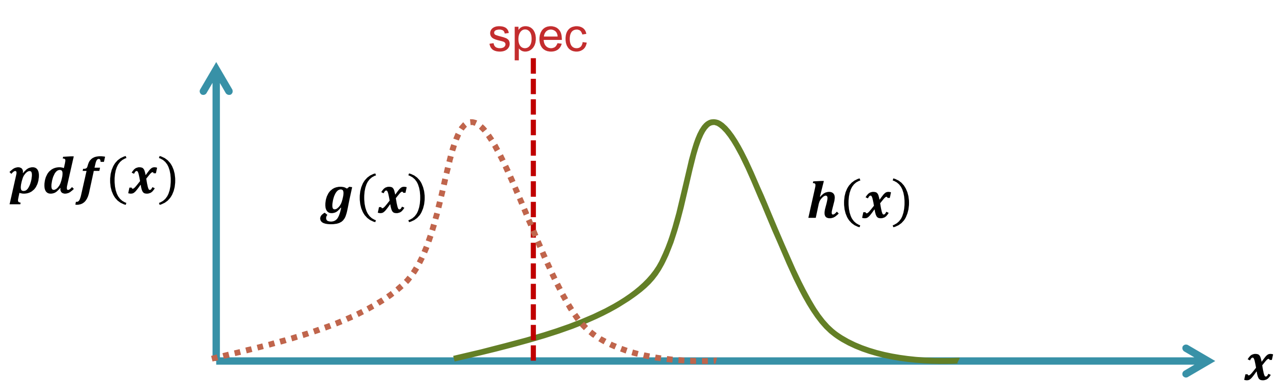
- Indicator Function



- Probability of rare failure events
  - Random 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 $\frac{h(x)}{g(x)}$



Key Problem of Importance Sampling

- Q: How to find the optimal ***g(x)*** as the new sampling distribution?

- A: It has been given in the literature but difficult to calculate:

$$g^{opt}(x) = \frac{I(x) \cdot h(x)}{prob(failure)}$$

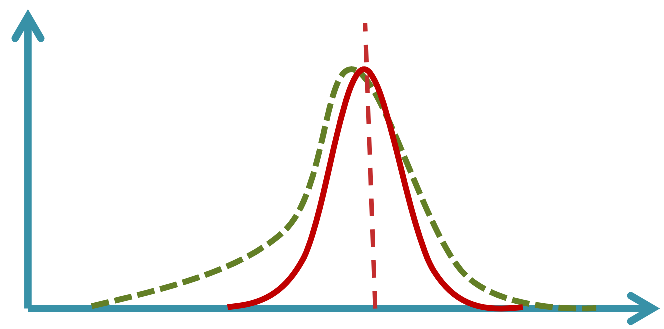
Outline

- Background
- Proposed Algorithm
- Experiments
- Extension Work

* Proposed algorithm is based on several techniques. Due to limited time, we only present the overall algorithm in this talk. More details can be found in the paper.

Basic Idea

- Find one parameterized distribution to approximate the theoretical optimal sampling distribution as close as possible.



- Modeling of process variations in SRAM cells:
  - VTH variations are typically modeled as independent Gaussian random variables;
  - Gaussian distribution can be easily parameterized by:
    - mean value → mean-shift : move towards failure region.
    - standard-deviation → sigma-change: concentrate more samples around the failure region.

parameterized Gaussian distribution ➔ the optimal sampling distribution.

Find the Optimal Solution

- Need to solve an optimization problem:
  - Minimize the distance between the parameterized distribution and the optimal sampling distribution.
- Three Questions:
  - What is the objective function?
    - e.g., how to define the distance?
  - How to select the initial solution of parameterized distributions?
  - Any analytic solution to this optimization problem?

Objective Function

- Kullback-Leibler (KL) Distance
  - Defined between any two distributions and measure how "close" they are. ➔ "distance"

$$D_{KL}(g^{opt}(x), h(x)) = E_{g^{opt}}\left[\log\left(\frac{g^{opt}(x)}{h(x)}\right)\right]$$

- Optimization problem based on KL distance

$$\min \ E_{g^{opt}}\left[\log\left(\frac{g^{opt}(x)}{h(x)}\right)\right] \Rightarrow \max \ E_h\left[I(x)\cdot\log\left(h(x)\right)\right]$$

- With the parameterized distribution, this problem becomes:

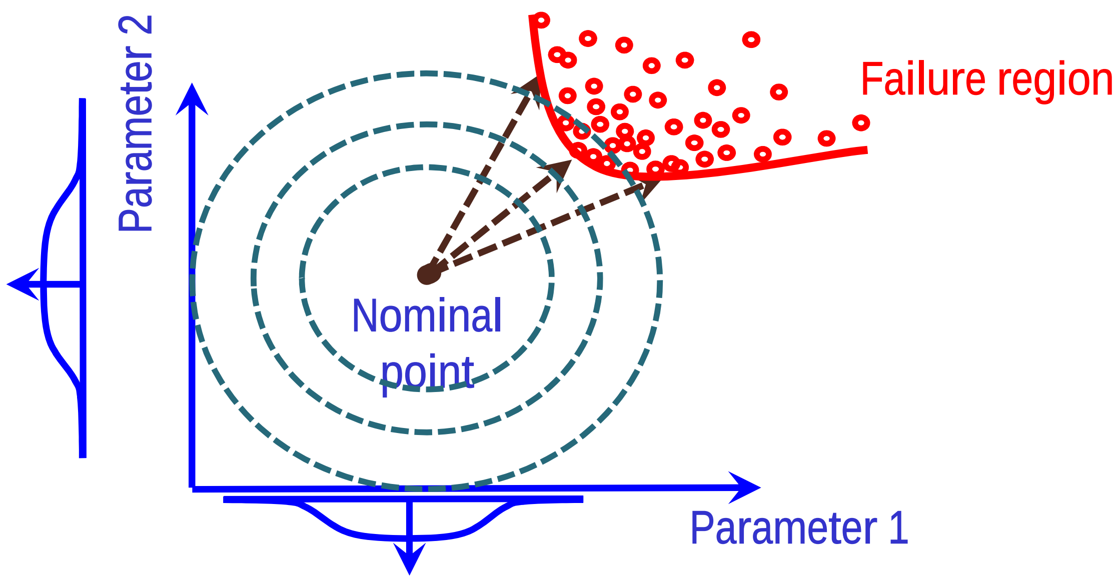
$$[\mu^*, \sigma^*] = \arg\max \ E_h\left[I(x)\cdot w(x,\mu,\sigma)\cdot\log\left(h(x,\mu,\sigma)\right)\right]$$

$$where \ \ w(x,\mu,\sigma) = \frac{h(x)}{h(x,\mu,\sigma)}$$

Initial Parameter Selection

- It is important to choose "initial solution" of mean and std-dev for each parameterized distribution.

- Find the initial parameter based on "norm minimization"
  - The point with "minimum L2-norm" is the most-likely location where the failure can happen.
  - The figure shows 2D case but the same technique applies to high-dim problems.



Analytic Optimization Solution

- Recall that the optimization problem is

$$[\mu^*, \sigma^*] = \arg\max \quad E_h\left[I(x)\cdot w(x,\mu,\sigma)\cdot \log\left(h(x,\mu,\sigma)\right)\right]$$

  - $\boldsymbol{E_h}$ cannot be evaluated directly and sampling method must be used:

$$[\mu^*, \sigma^*] = \arg\max \quad \frac{1}{N}\sum_{j=1}^{N}\left[I(x_j)\cdot w(x_j,\mu,\sigma)\cdot \log\left(h(x_j,\mu,\sigma)\right)\right]$$

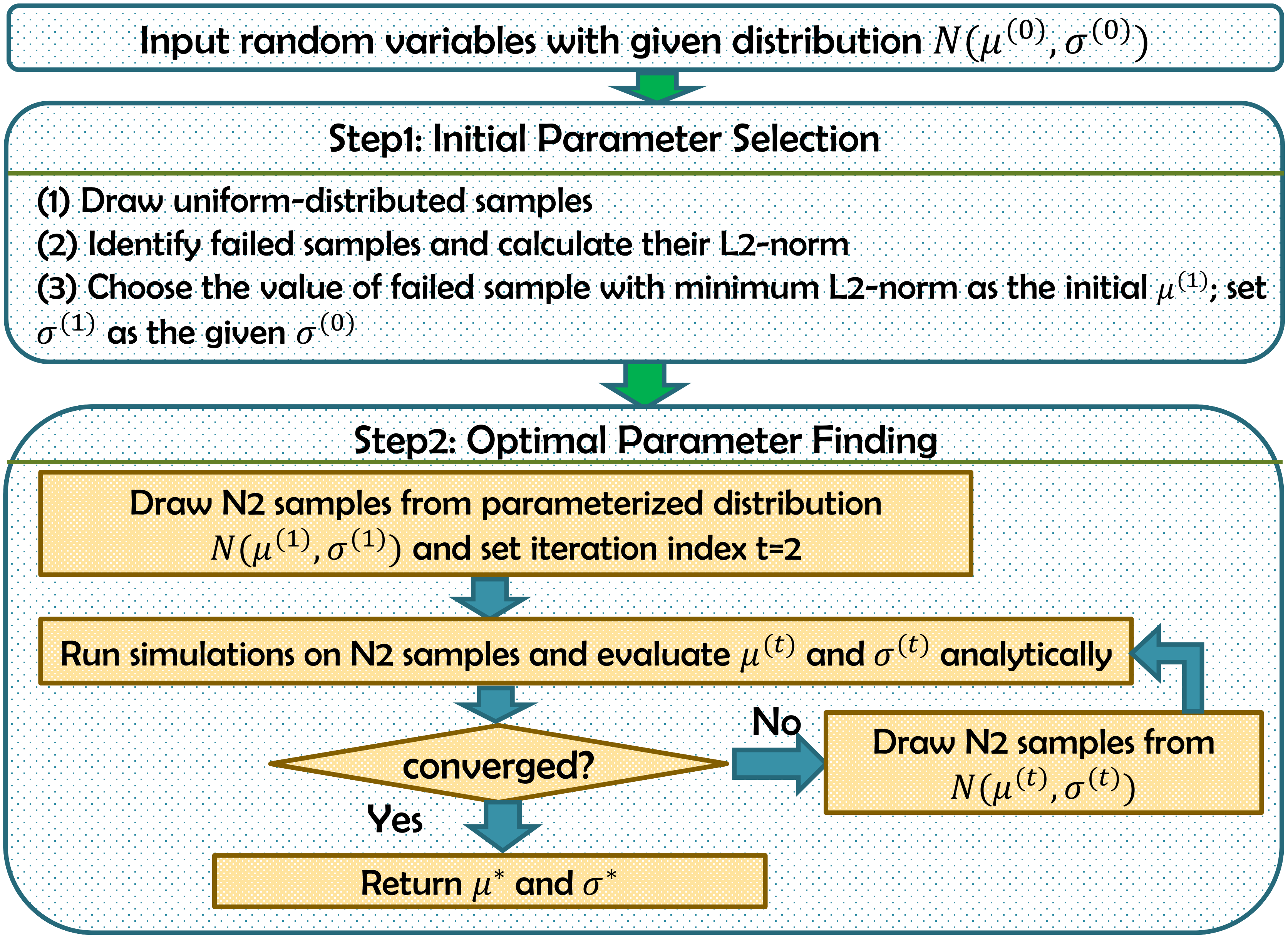
- Above problem can be solved analytically:
  - $h(x,\mu,\sigma)$ follows Gaussian distribution
  - The optimal solution of this problem can be solved by (e.g., mean):

$$\frac{\partial E_h\left[I(x)\cdot w(x,\mu,\sigma)\cdot \log\left(h(x,\mu,\sigma)\right)\right]}{\partial \mu} = 0$$

- Analytic Solution

$$\mu^{(t)} = \frac{\sum_{i=1}^{N} I(x_i)\cdot w(x_i,\mu^{(t-1)},\sigma^{(t-1)})\cdot x_i}{\sum_{i=1}^{N} I(x_i)\cdot w(x_i,\mu^{(t-1)},\sigma^{(t-1)})}; \quad \sigma^{(t)} = \sqrt{\frac{\sum_{i=1}^{N} I(x_i)\cdot w(x_i,\mu^{(t-1)},\sigma^{(t-1)})\cdot (x_i-\mu^{(t)})^2}{\sum_{i=1}^{N} I(x_i)\cdot w(x_i,\mu^{(t-1)},\sigma^{(t-1)})}}$$

Overall Algorithm



Overall Algorithm (cont.)

Step3: Failure Probability Estimation

Draw N3 samples from parameterized distribution $N(\mu^*, \sigma^*)$

Run simulations on N3 samples and evaluate indicator function $I(x)$

Solve for the failure probability p_r with sampled form as

$$p_r = \frac{1}{N_3} \sum_{i=1}^{N_3} I(x_i) \cdot w(x_i, \mu^*, \sigma^*)$$

where $w(x_i, \mu^*, \sigma^*) = {}^{h(x_i)} / {}_{h(x_i, \mu^*, \sigma^*)}$

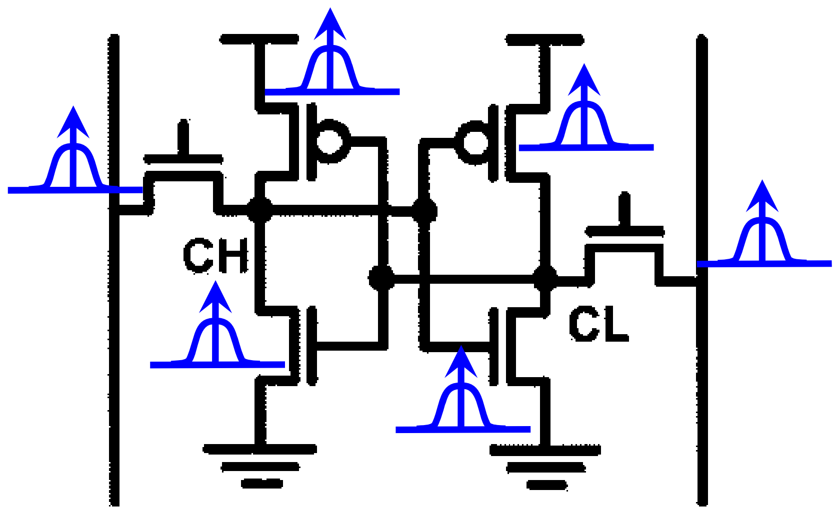
Return the failure probability estimation p_r

Outline

- Background
- Proposed Algorithm
- Experiments
- Extension Work

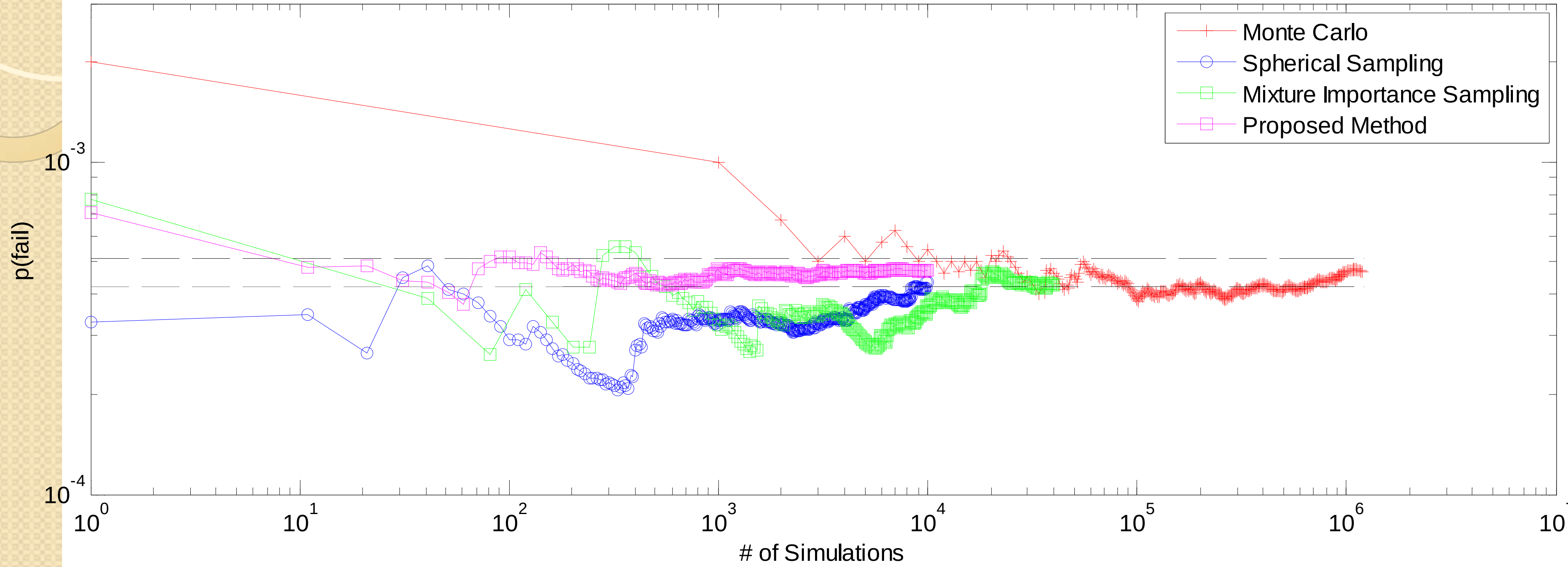
6-T SRAM bit-cell

- SRAM cell in 45nm process as an example:
  - Consider VTH of six MOSFETs as independent Gaussian random variables.
  - Std-dev of VTH is the 10% of nominal value.



- Performance Constraints:
  - Static Noise Margin (SNM) should be large than zero;
  - When SNM<=0, data retention failure happens ("rare events").

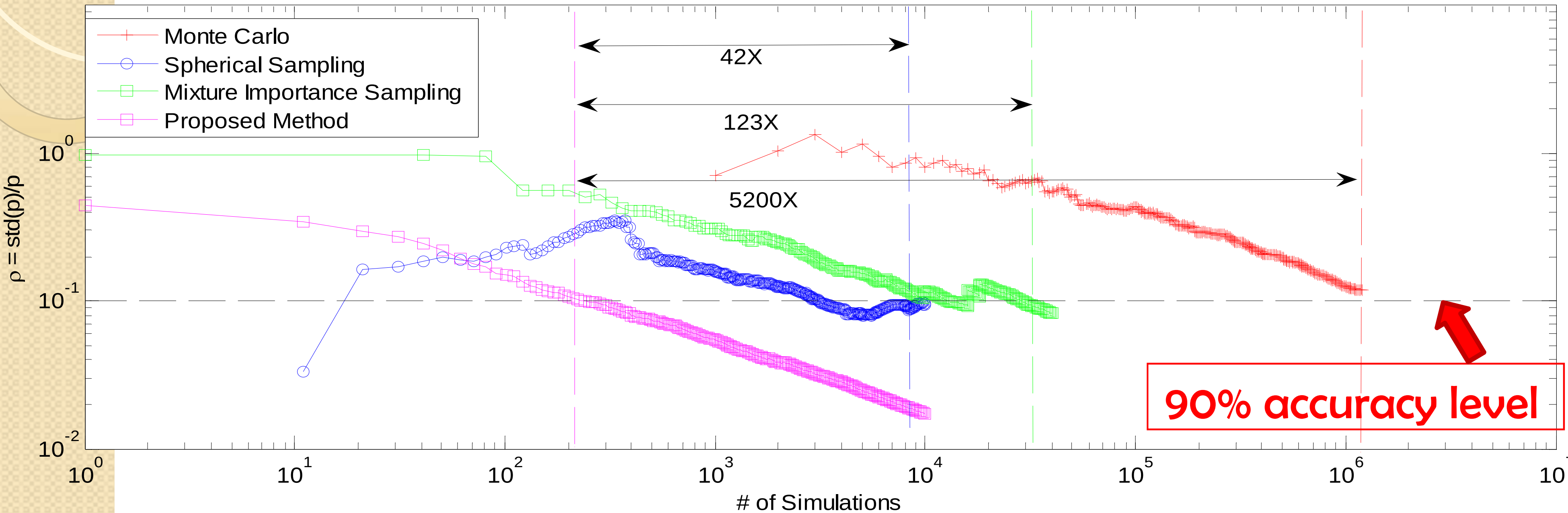
Accuracy Comparison (Vdd=300mV)
- Evolution of the failure rate estimation



- Failure rate estimations from all methods can match with MC;
- Proposed method starts with a close estimation to the final result;
- Importance Sampling is highly sensitive to the sampling distribution.

Efficiency Comparison (Vdd=300mV)

- Evolution of figure-of-merit (FOM)



- Figure-of-merit is used to quantify the error (lower is better):

$$\rho = \frac{\sigma_{prob\_fail}}{p_{fail}}$$

- Proposed method can improve accuracy with1E+4 samples as:

	MC	MixIS	SS	Proposed
Probability of failure	5.455E-4	3.681E-4	4.343E-4	4.699E-4
Accuracy	18.71%	88.53%	90.42%	98.2%

Outline

- Background
- Proposed Algorithm
- Experiments
- Extension Work

Problem of Importance Sampling

- Q: How does Importance Sampling behave for high-dimensional problems (e.g., tens or hundreds variables)?

- A: Unreliable or completely wrong!

  ➔ curse of dimensionality

- Reason: the degeneration of likelihood ratios

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

  - Some likelihood ratios become dominate (e.g., very large when compared with the rest)
  - The variance of likelihood ratios are very large.

Extension Work

- We develop an efficient approach to address rare events estimation in high dimension which is a fundamental problem in multiple disciplines.

- Our proposed method can reliably provide high accuracy:
  - tested problem with 54-dim and 108-dim;
  - probability estimation of rare events can always match with MC method;
  - It provides several order of magnitude speedup over MC while other IS-based methods are completely failed.

Study Case		Monte Carlo (MC)	Spherical Sampling	Proposed Method
108-dim	P(fail)	3.3723e-05	1.684	3.3439e-005
	Relative error	0.097354	0.09154	0.098039
	#run	5.7e+6 (1425X)	4.3e+4	4e+3 (1X)

- We can prove that IS method loses its upper bound of estimation in high-dimension, and estimation from our method can always be bounded.

Thank you!

Any further questions can be addressed to fang08@ee.ucla.edu