Adaptive Clustering and Sampling for High-Dimensional and Multi-Failure-Region SRAM Yield Analysis

Xiao Shi[1,2], Hao Yan[3], Jinxin Wang[3], Xiaofen Xu[3] ,
Fengyuan Liu[3], Lei He[1,2], Longxing Shi[3]

April 17, 2019

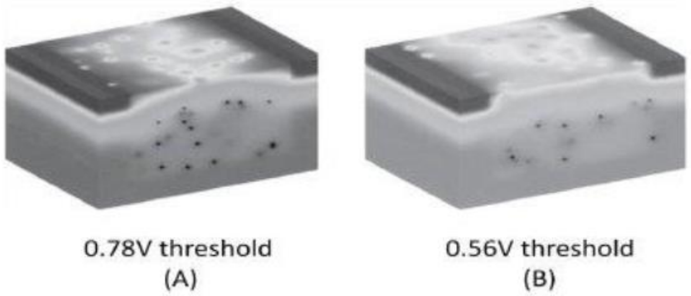
1 State Key Lab of ASIC & System, Microelectronics Dept., Fudan University, China
2 Electrical and Computer Engineering Dept.,University of California, Los Angeles, CA,USA
3 Southeast University, China

Outline

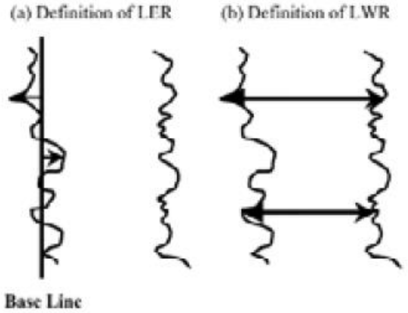
- Preliminary of High Sigma Analysis and Existing Approaches
- The Proposed Approach
- Experiment Results
- Summary

Statistical Circuit Simulation

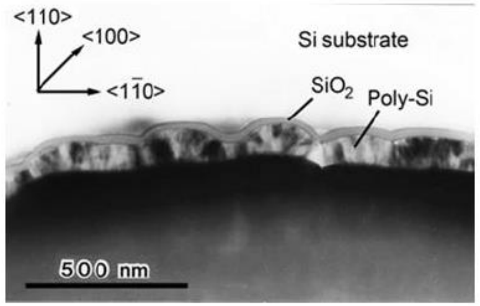
- Process Variation
  - First mentioned by William Shockley in his analysis of P-N junction breakdown[S61] in 1961
  - Revisited in 2000s for long channel devices[JSSC03, JSSC05]
  - Getting more attention at sub-100nm[IBM07, INTEL08]
- Sources of Process Variation



Random dopant fluctuations: both transistor has the same number of dopants (170)
- Courtesy of Deepak Sharma, Freescale Semiconductors



Line-edge and Line-width Roughness (LER and LWR)
-Courtesy of Technology and Manufacturing Group, Intel Corporation



Gate oxide thickness:
-Courtesy of Professor Hideo Sunami at Hiroshima University

- Statistical Circuit Simulation helps to debug circuits in the pre-silicon phase to improve yield rate

[S61] Shockley, W., "Problems related to p-n junctions in silicon." Solid-State Electronics, Volume 2, January 1961, pp. 35–67.
[JSSC03] Drennan, P. G., and C. C. McAndrew. "Understanding MOSFET Mismatch for Analog Design." IEEE Journal of Solid-State Circuits 38, no. 3 (March 2003): 450–56.
[JSSC05] Kinget, P. R. "Device Mismatch and Tradeoffs in the Design of Analog Circuits." IEEE Journal of Solid-State Circuits 40, no. 6 (June 2005): 1212–24.
[IBM07] Agarwal, Kanak, and Sani Nassif. "Characterizing process variation in nanometer CMOS." Proceedings of the 44th annual Design Automation Conference. ACM, 2007.
[Intel08] Kuhn, K., Kenyon, C., Kornfeld, A., Liu, M., Maheshwari, A., Shih, W. K., ... & Zawadzki, K. (2008). Managing Process Variation in Intel's 45nm CMOS Technology. Intel Technology Journal, 12(2).

High Sigma Analysis

- High sigma (rare event) tail is difficult to achieve with Monte Carlo method
  - # of simulations required to capture 100 failed samples

STANDARD DEVIATION OF THE MEAN

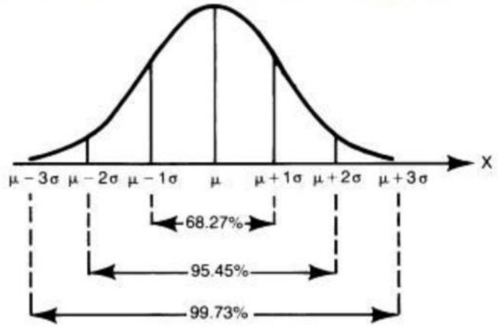


Figure 2

Percent	99.73%	99%	95.45%	95%	90%	80%	68.27%
No. of ± σ's	3.00	2.58	2.00	1.96	1.645	1.28	1.00

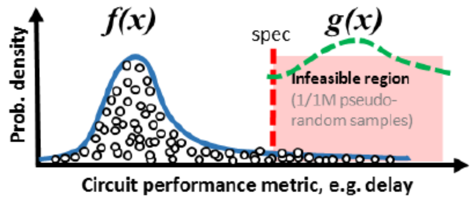
Sigma	Probability	# of Simulations[1]
1	0.15866	700
2	0.02275	4,400
3	0.00135	74,100
4	3.17E-05	3,157,500
5	2.87E-07	348,855,600

- High sigma analysis is critical for highly-duplicated circuits
  - Memory cells (up to 4-6 sigma), IO and analog circuits (3-4 sigma) [1]
- How to efficiently and accurately estimate P_{fail}(yield rate) on high sigma tail?

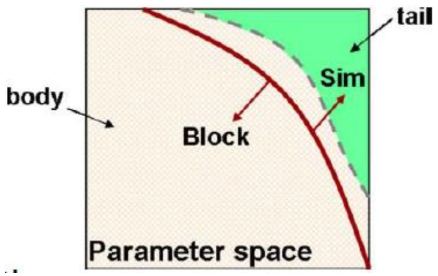
1 Cite from Solido Design Automation whitepaper

Existing Methods and Limitations

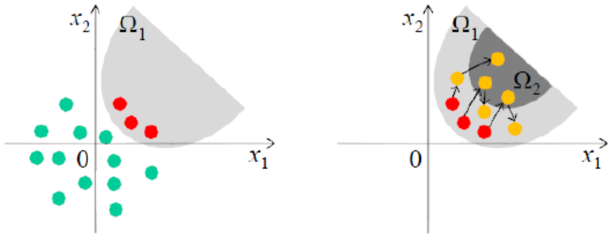
- Draw more samples in the tail
- Importance Sampling[DAC06]
  - Shift the sample distribution to more "important" region
  - Curse of dimensionality[Berkeley08, Stanford09]



- Classification based methods[TCAD09]
  - Filter out unlikely-to-fail samples using classifier
  - Classifiers perform poorly at high dimensional with limited number of training samples.



- Markov Chain Monte Carlo[ICCAD14]
  - It is difficult to cover the failure regions using a few chains of samples



[DAC06] R. Kanj, R. Joshi, and S. Nassif. "Mixture Importance Sampling and Its Application to the Analysis of SRAM Designs in the Presence of Rare Failure Events." DAC, 2006
[Berkeley08] Bengtsson, T., P. Bickel, and B. Li. "Curse-of-Dimensionality Revisited: Collapse of the Particle Filter in Very Large Scale Systems." *Probability and Statistics: Essays in Honor of David A. Freedman*2 (2008): 316–34.
[Stanford09] Rubinstein, R.Y., and P.W. Glynn. "How to Deal with the Curse of Dimensionality of Likelihood Ratios in Monte Carlo Simulation." *Stochastic Models*25, no. 4 (2009): 547–68.
[TCAD09] Singhee, A., and R. Rutenbar. "Statistical Blockade: Very Fast Statistical Simulation and Modeling of Rare Circuit Events and Its Application to Memory Design." TCAD, 2009
[ICCAD14] Sun, Shupeng, and Xin Li. "Fast Statistical Analysis of Rare Circuit Failure Events via Subset Simulation in High-Dimensional Variation Space." ICCAD 2014

Outline

- Preliminary of High Sigma Analysis and Existing Approaches
- The Proposed Approach
- Experiment Results
- Summary

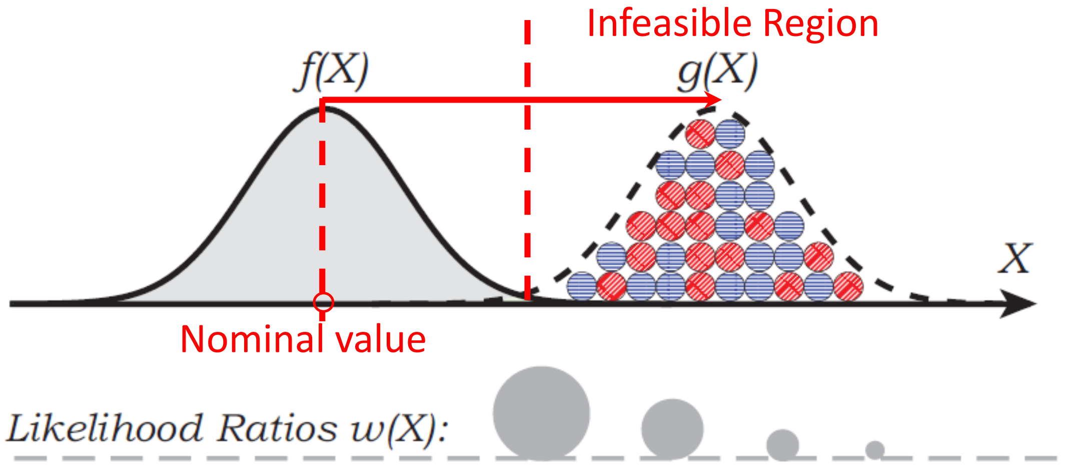
Importance Sampling

- Shift the sample distribution to more "important" region
  - $P_{fail} = \int I(X) \cdot f(X)dX$

$$= \int I(X) \cdot \boxed{\frac{f(X)}{g(X)}} \cdot g(X)dX$$

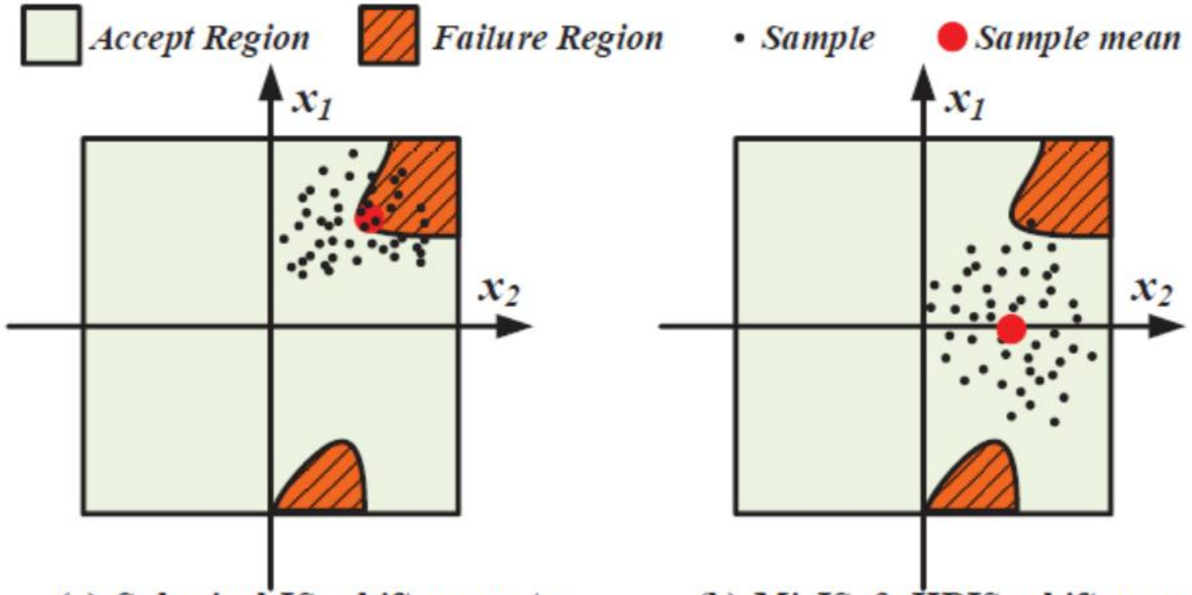
$$= \int I(X) \cdot w(X) \cdot g(X)dX$$

- $I(X)$ is the indicator function
- $g_{opt}(X) = \frac{I(X)f(X)}{P_{fail}}$
  - Smallest variance
  - Infeasible in analytical form



Challenges - Optimal Sampling PDF $g_{opt}(x)$

- How to generate target distribution $g(x)$ that can capture more important failure samples?
  - Mean-shift methods fail at multi-failure-region cases
  - More desirable to approximate the failure region

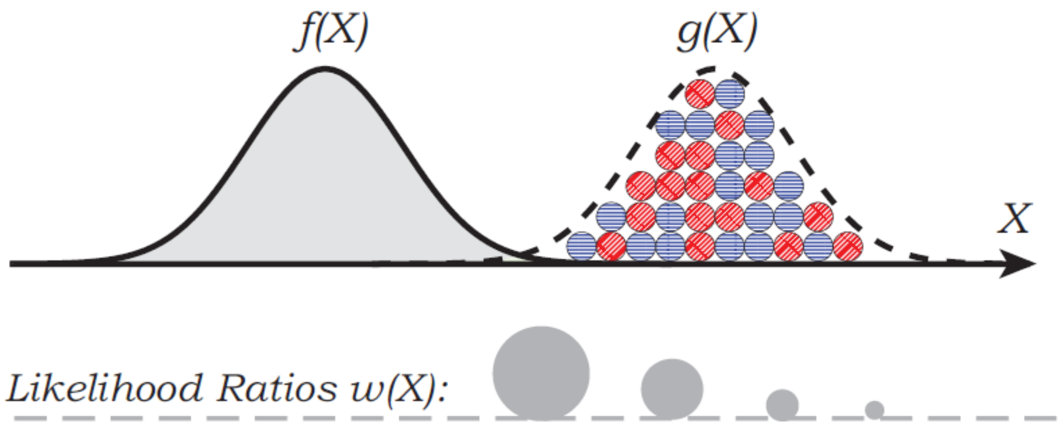


(a) Spherical IS: shift mean to the min-norm failed sample

(b) MixIS & HDIS: shift mean to the centroid of failed samples

Challenges - Weight Instability

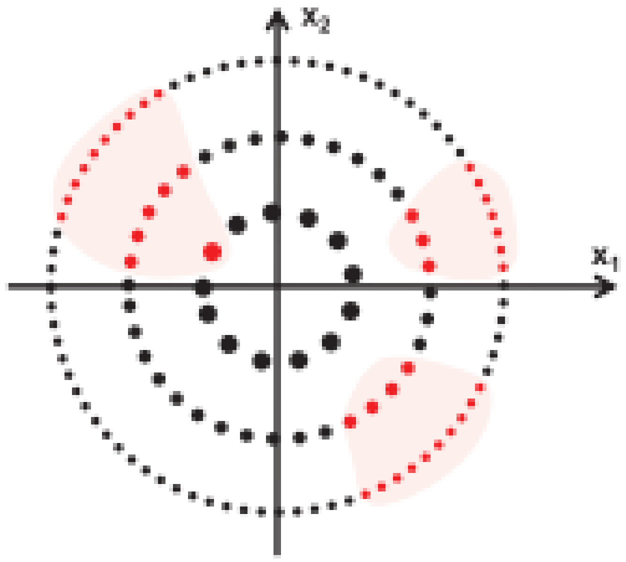
- Likelihood ratio or weight: $\frac{f(X)}{g(X)}$
- Samples with higher likelihood ratio has higher impact to the estimation of Pfail
  - Larger $f(x)$, Smaller $g(x)$
- Weight $f(x)/g(x)$ might be extremely large at high dimension

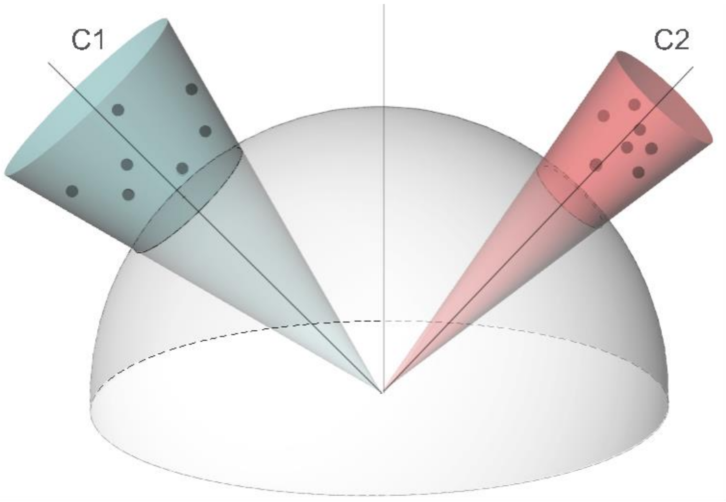


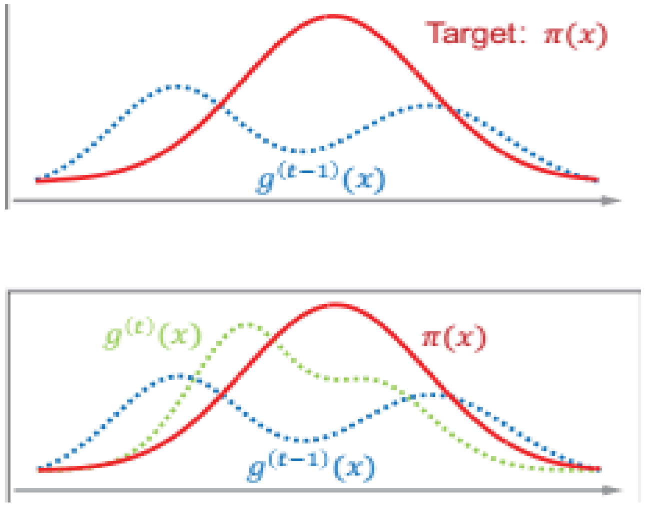
Adaptive Clustering and Sampling(ACS)

- Algorithm overview:

Hyperspherical Presampling → Multi-cone Clustering → Adaptive Sampling

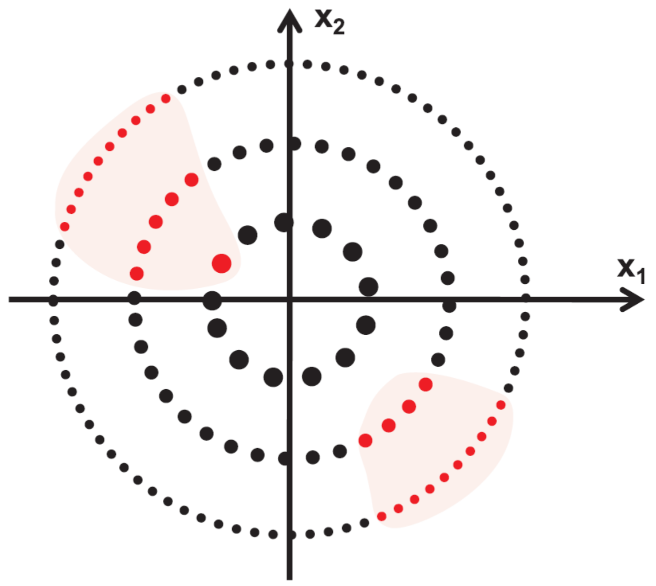






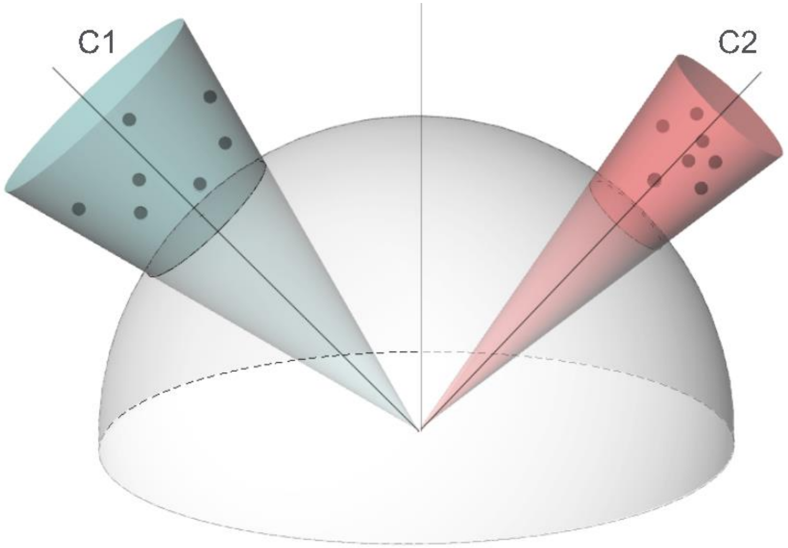
ACS Phase 1: Hyperspherical Presampling

- Purpose
  - Construct the initial sampling distribution before the first iteration
- Restrict the samples to hyper-spherical surfaces
  - Dimension reduction
- Samples with smaller Euclidean norm has higher importance



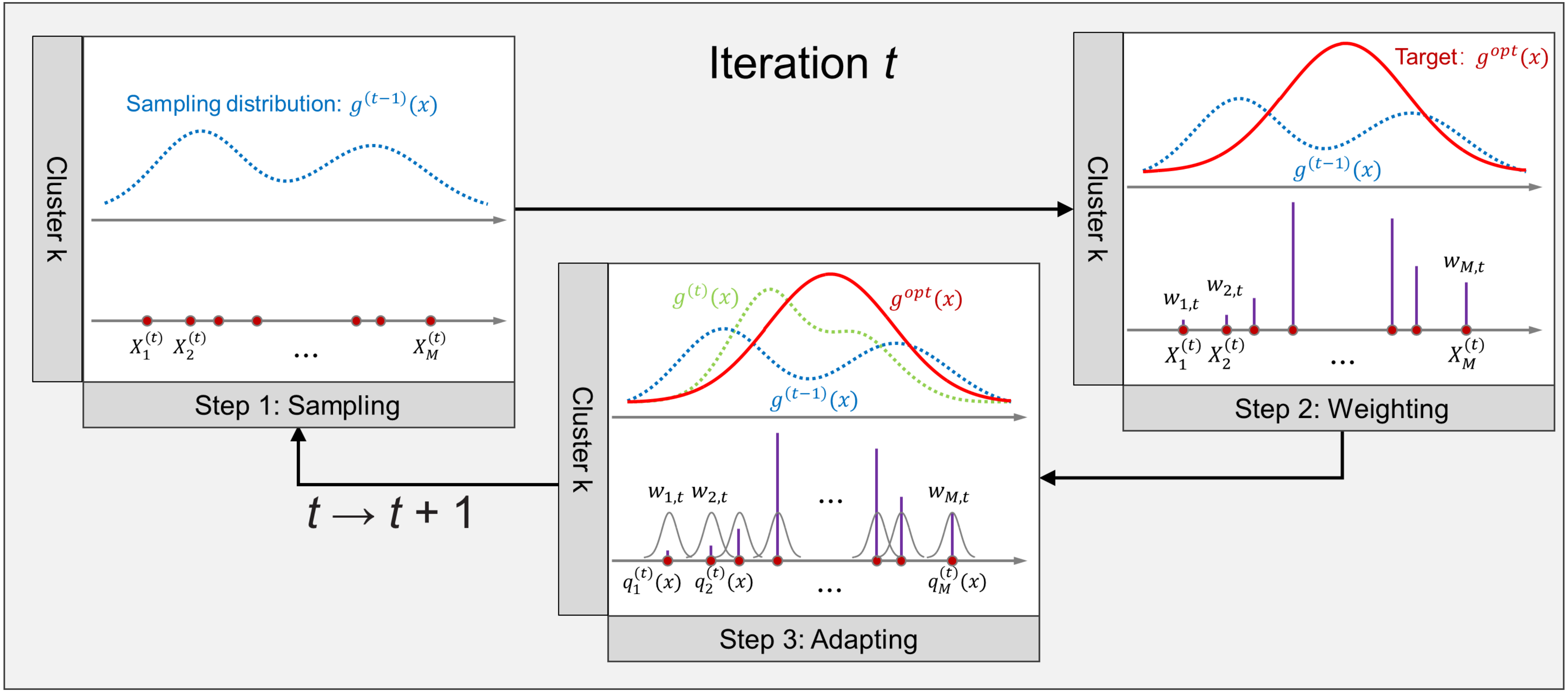
ACS Phase 2: Multi-cone Clustering

- Purpose
  - Cluster failure samples based on their direction
  - Project sample points to the unit sphere surface in the radial direction

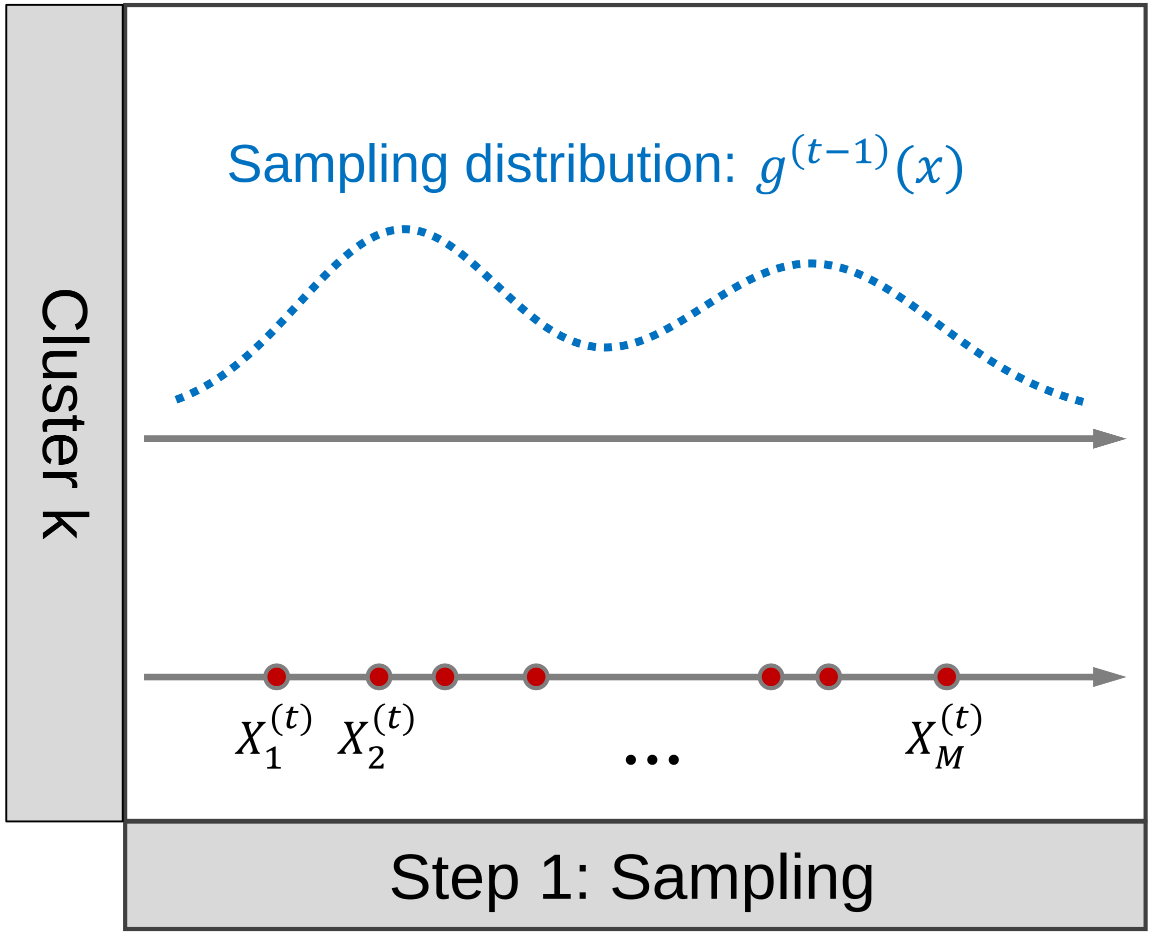


- Modified k-means algorithm
  - Distance metric: $CosineDistance\left(X^{(1)}, X^{(2)}\right) = 1 - \frac{X^{(1)} \cdot X^{(2)}}{|X^{(1)}||X^{(2)}|}$
  - Number of clusters: $k = \sqrt{N}$

ACS Phase 3: Adaptive Sampling

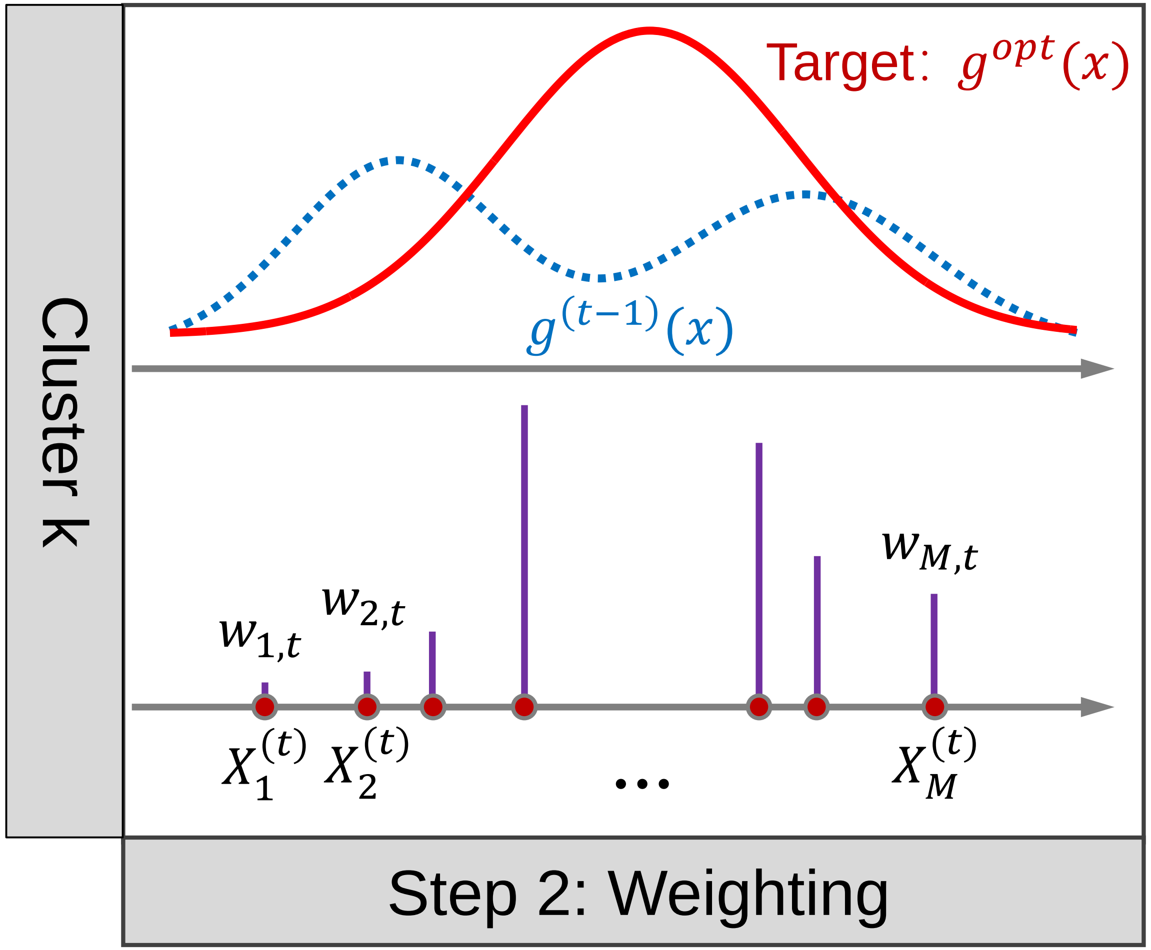


ACS Phase 3: Adaptive Sampling



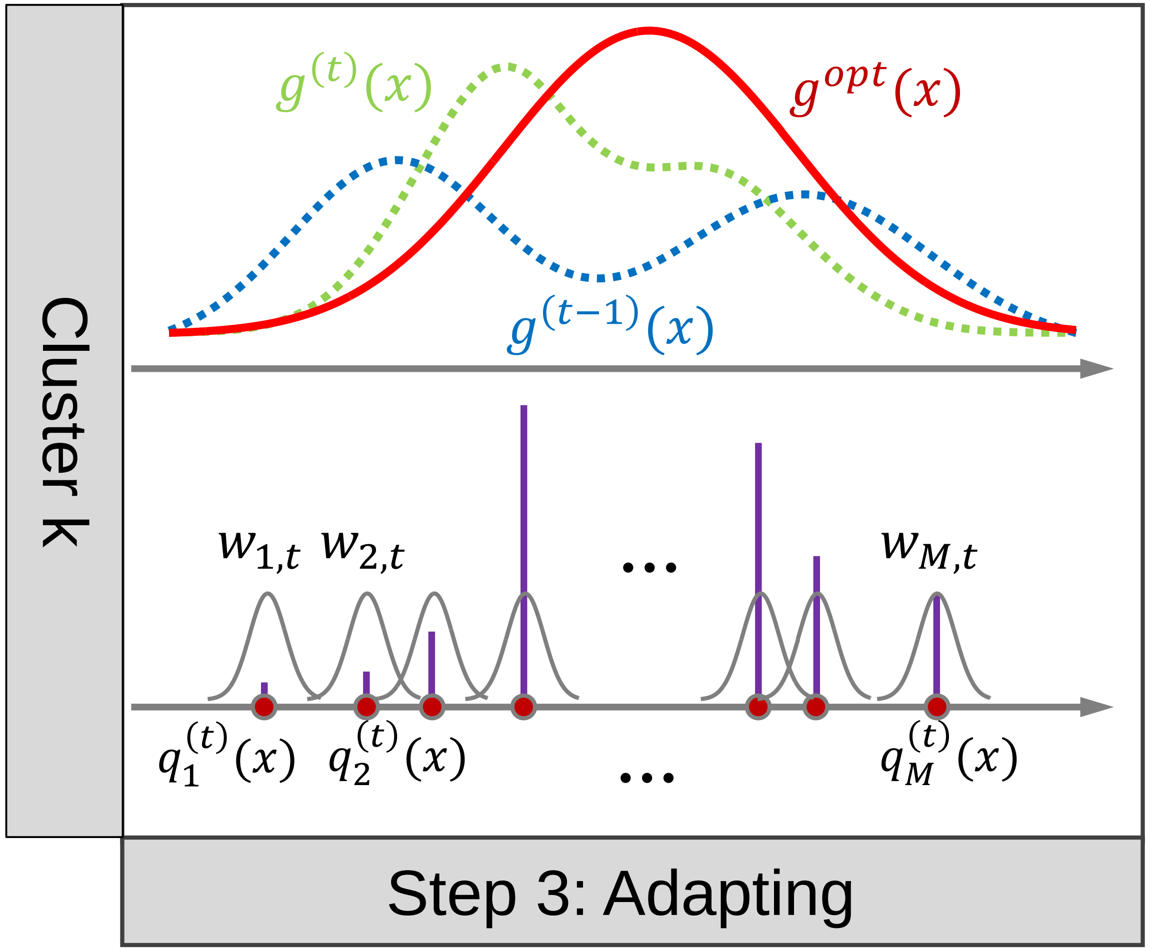
- Generate samples from previous sampling distribution $g^{(t-1)}(x)$

ACS Phase 3: Adaptive Sampling



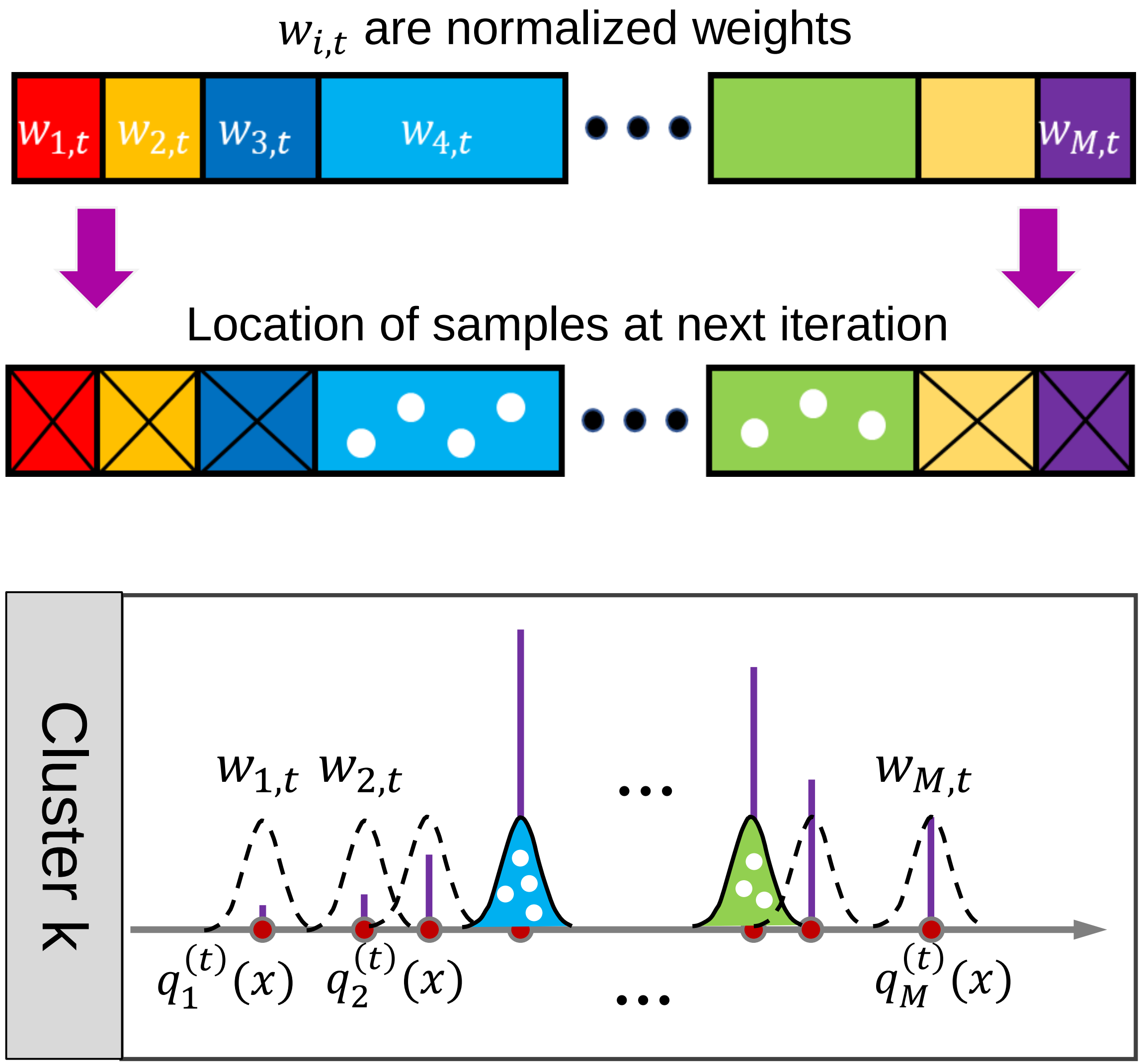
- Compute discrepancy ratio of iteration t
  - $w_{i,t} = \frac{\pi(X)}{g^{(t-1)}(X)} = \frac{f(X)I(X)}{\sum_{i=1}^{N} w_i \cdot q^{(t-1)}(X)}$

ACS Phase 3: Adaptive Sampling



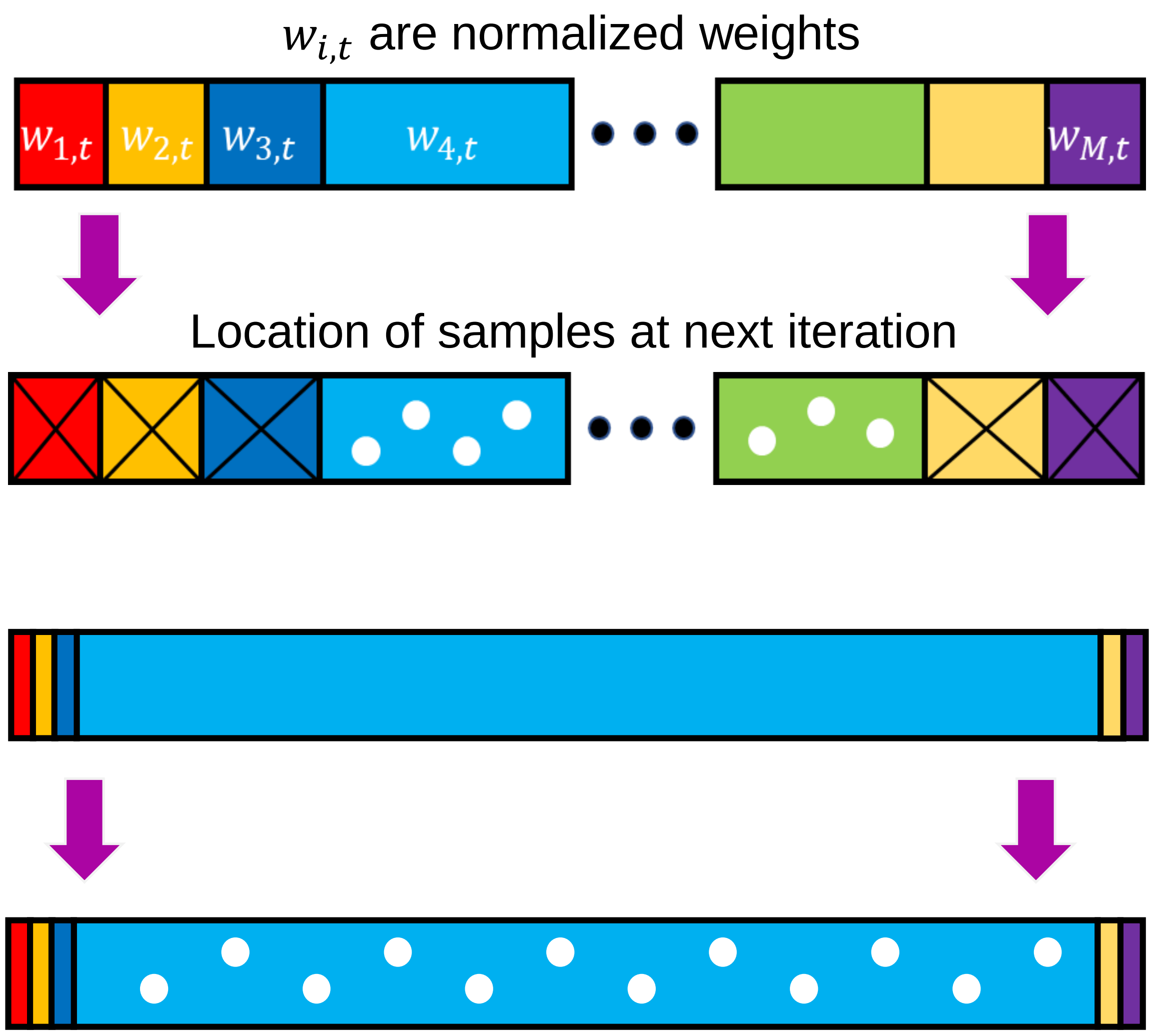
- Weighted gaussian Mixture Distribution
  - Probability mass
  - Kernel density estimation

ACS Phase 3: Distribution adaptation



- Normalize discrepancy ratio for sampling
  - $\overline{w_{i,t}} = \frac{w_{i,t}}{\sum_{i=1}^{N} w_{i,t}}$
- Effective Sample Size (ESS)
  - $ESS = \frac{1}{\sum_{i}^{N} (w_{i,t})^2}$
  - Reflects the degree of weight degeneracy

ACS Phase 3: Distribution adaptation

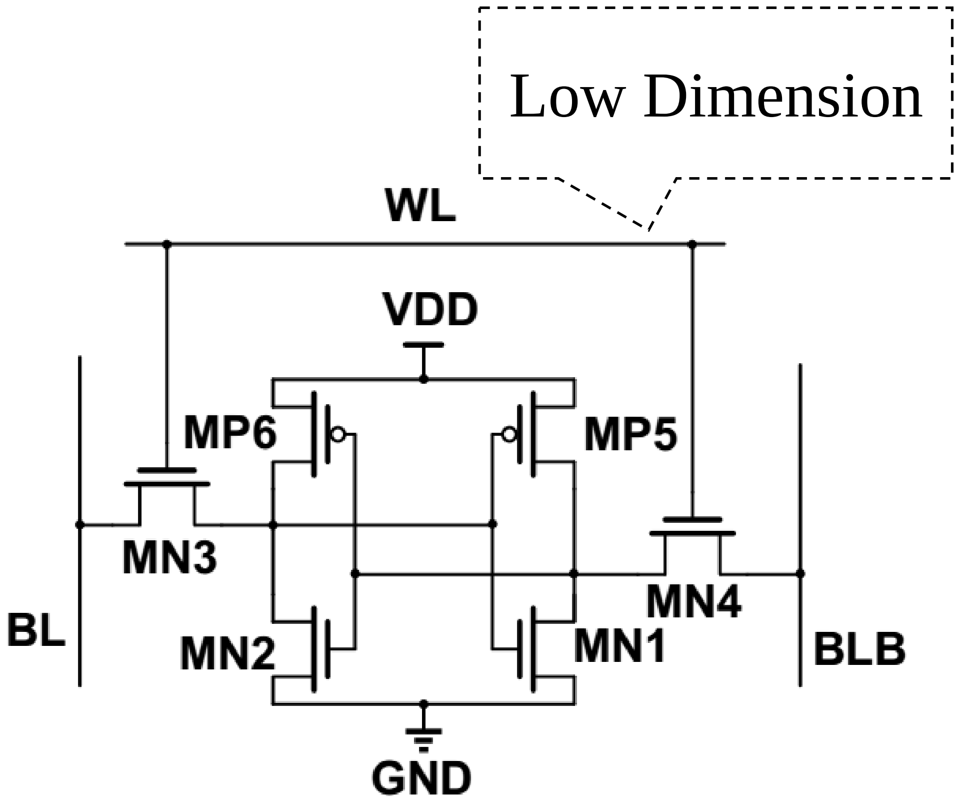


- Normalize discrepancy ratio for sampling
  - $\overline{w_{i,t}} = \frac{w_{i,t}}{\sum_{i=1}^{N} w_{i,t}}$
- Effective Sample Size (ESS)
  - $ESS = \frac{1}{\sum_{i}^{N} (w_{i,t})^2}$
  - Reflects the degree of weight degeneracy

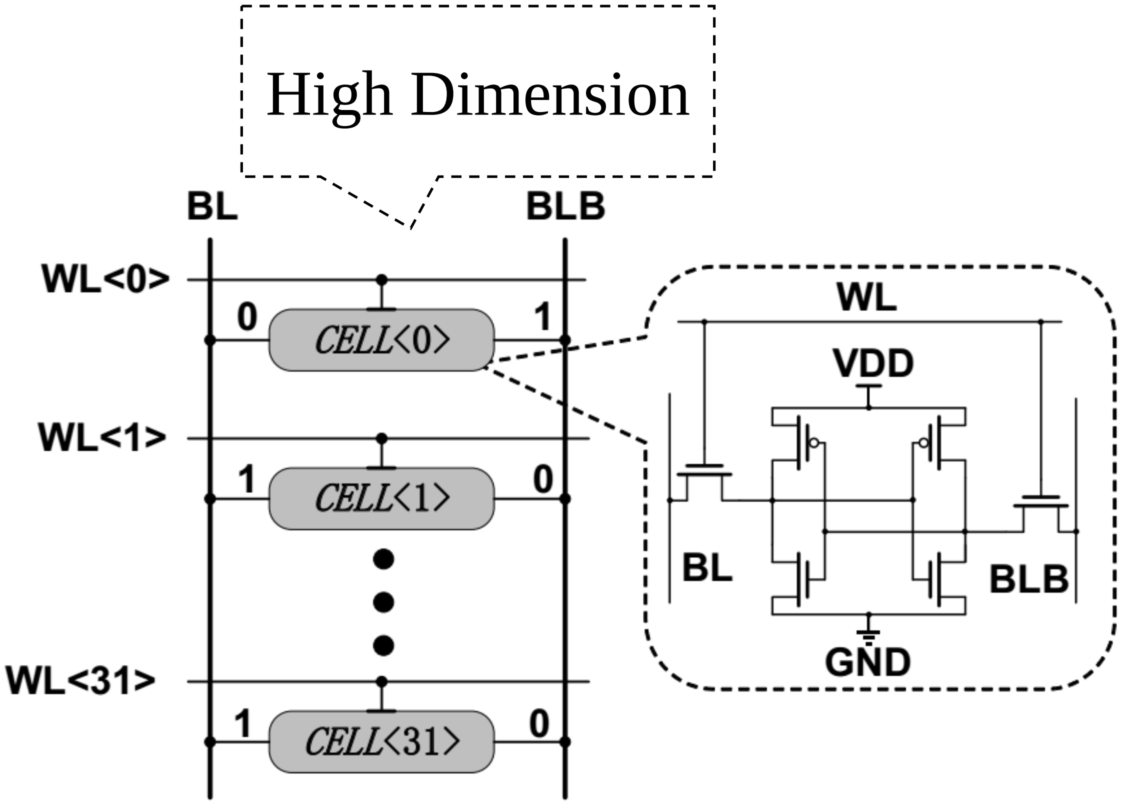
Outline

- Preliminary of High Sigma Analysis and Existing Approaches
- The Proposed Approach
- Experiment Results
- Summary

Experiments: Schematic of circuits

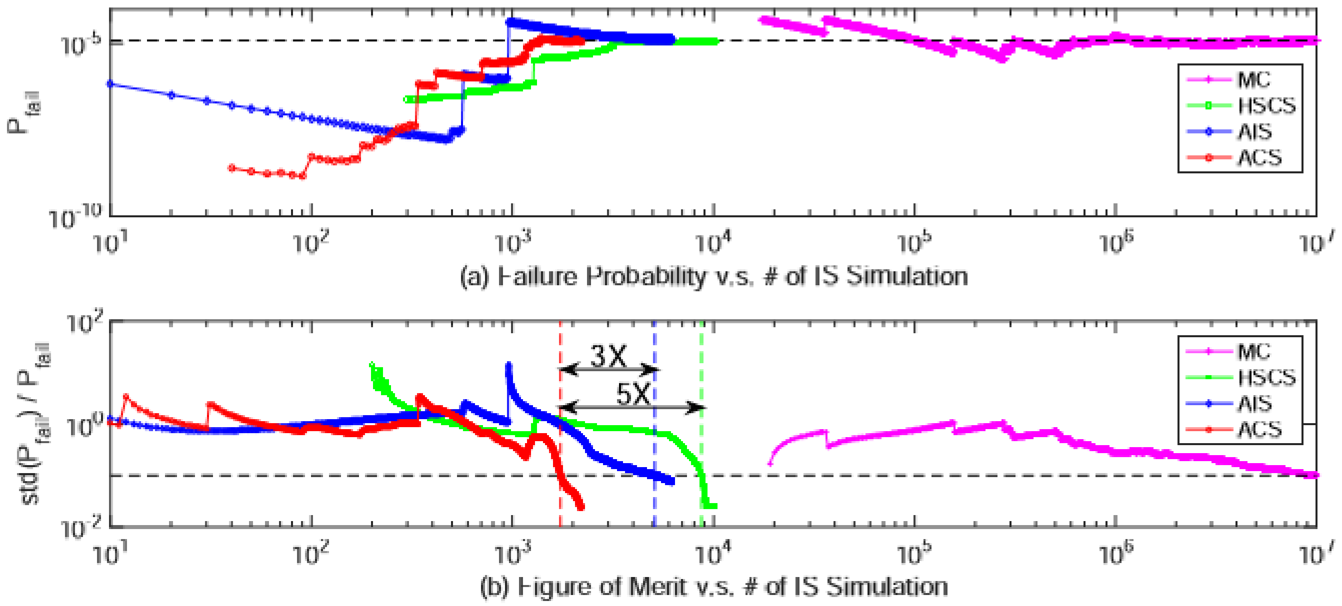


(a) SRAM Bit Cell Circuit



(b) SRAM Column Circuit

Experiments: Convergence and Runtime

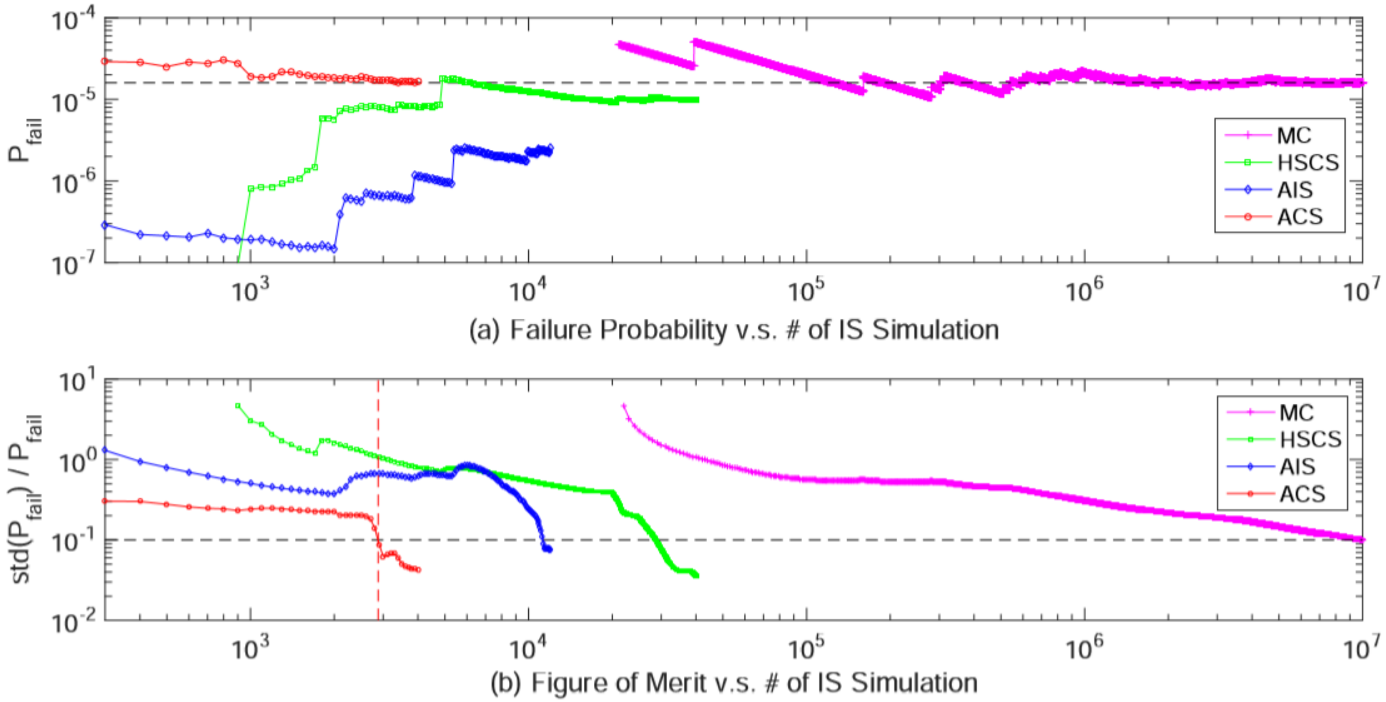


- Bit-cell experiment
  - Low dimension(18D)
  - Single failure region

	MC	HSCS	AIS	Proposed
Failure prob.(error)	1.24e-5(0%)	1.18e-5(4.6%)	1.32e-5(6.1%)	1.22e-5(1.5%)
Presampling # sim.	0	7000	3000	1100
Importance Sampling # sim.	1e7	8688	5111	1736
Total # sim. (speedup)	1e7(1X)	15688(637X)	8111(1233X)	2836(3526X)

3-5X faster than existing methods

Experiments: Convergence and Runtime

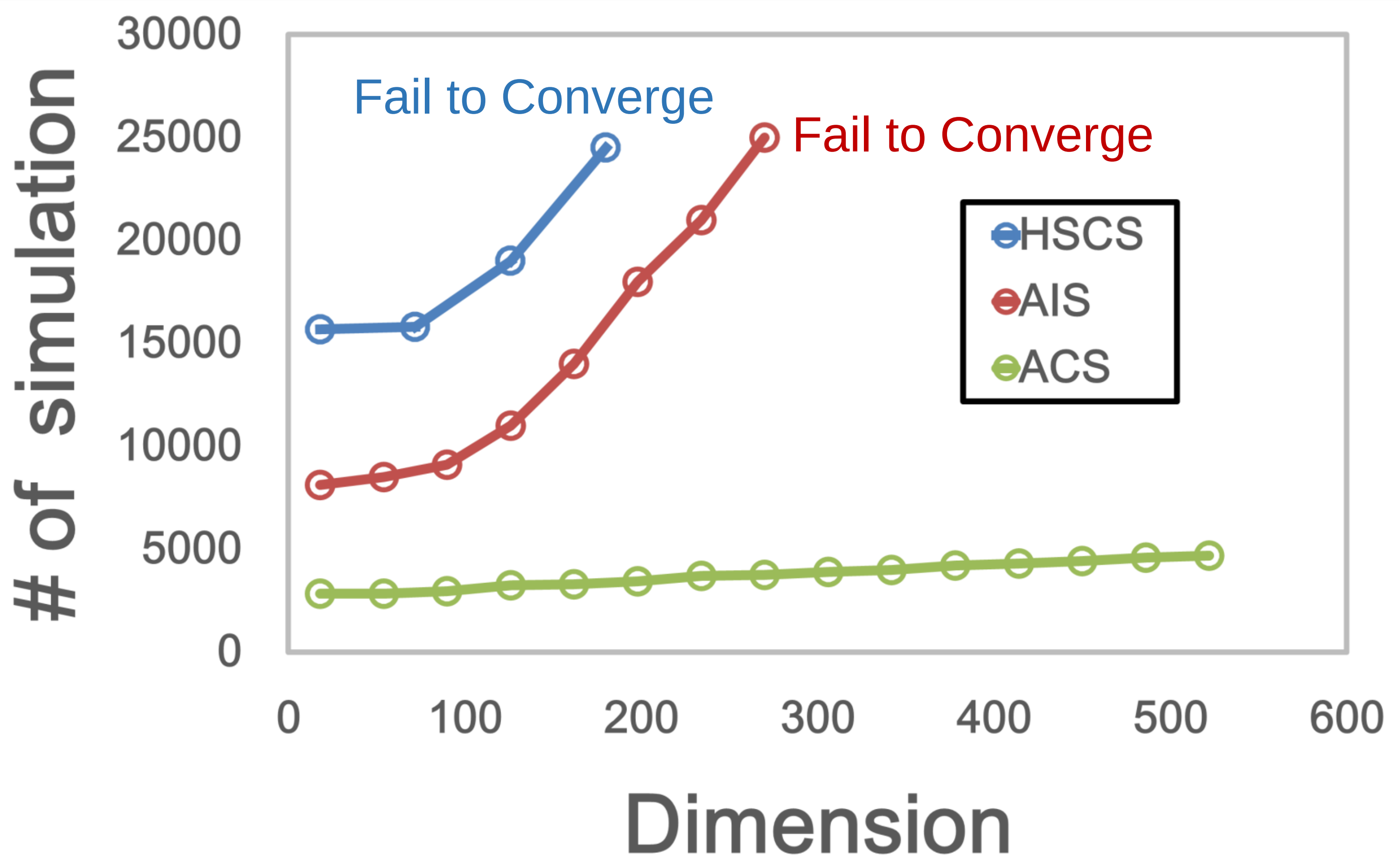


- SRAM column experiment
  - High dimension(576D)
  - Multiple failure regions

	Monte Carlo	HSCS	AIS	Proposed
Failure prob.(error)	1.60e-5(0%)	9.82e-6(error)	2.23e-6(error)	1.55e-5(3.1%)
Presampling # sim.	0	18000	5000	2000
Importance Sampling # sim.	1e7	28699	11253	2878
Total # sim. (speedup)	1e7(1X)	46699	16253	4878(2050X)

About 2050X faster than MC

Experiments: Dimension vs. # of simulations



- Vary the number of bit cells in SRAM column
- Simulation cost of ACS grows linearly with dimension

Summary

- Explore multiple failure regions
  - Adaptive sampling scheme
- Parallel computing in each failure region
  - Spherical presampling
  - Multi-cone clustering
- Better accuracy and efficiency
  - 3-5X faster than other existing methods

Q&A

Thank you for your attention!