Hyperspherical Clustering and Sampling for Rare Event Analysis with Multiple Failure Region Coverage

Wei Wu[1], Srinivas Bodapati[2], Lei He[1,3]

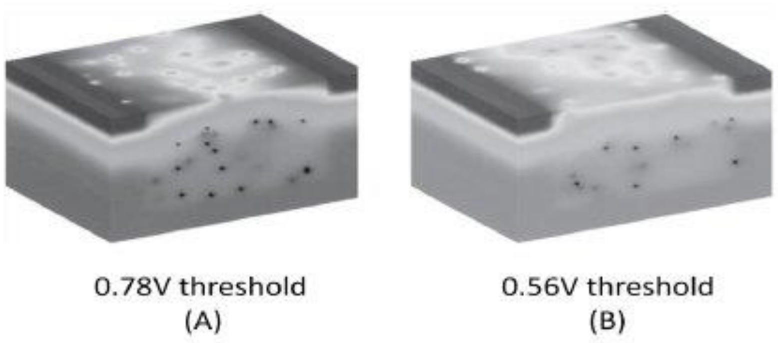
1 Electrical Engineering Department, UCLA

2 Intel Corporation

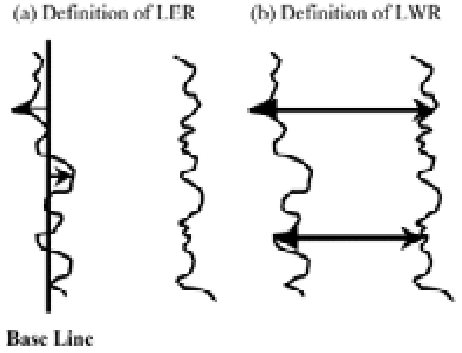
3 State Key Laboratory of ASIC and Systems, Fudan University, China

Why statistical circuit analysis? - Process Variation

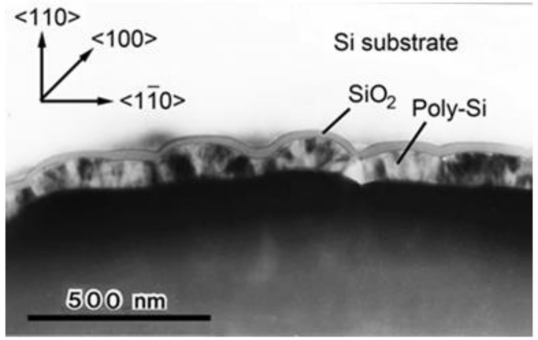
- Process Variation
  - First mentioned by William Shockley in his analysis of P-N junction breakdown[S61] in 1961
  - Revisited in 2000s for long channel devices [JSSC03, JSSC05]
  - Getting more attention at sub-100nm [IBM07, INTEL08]
- Sources of Process Variation



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



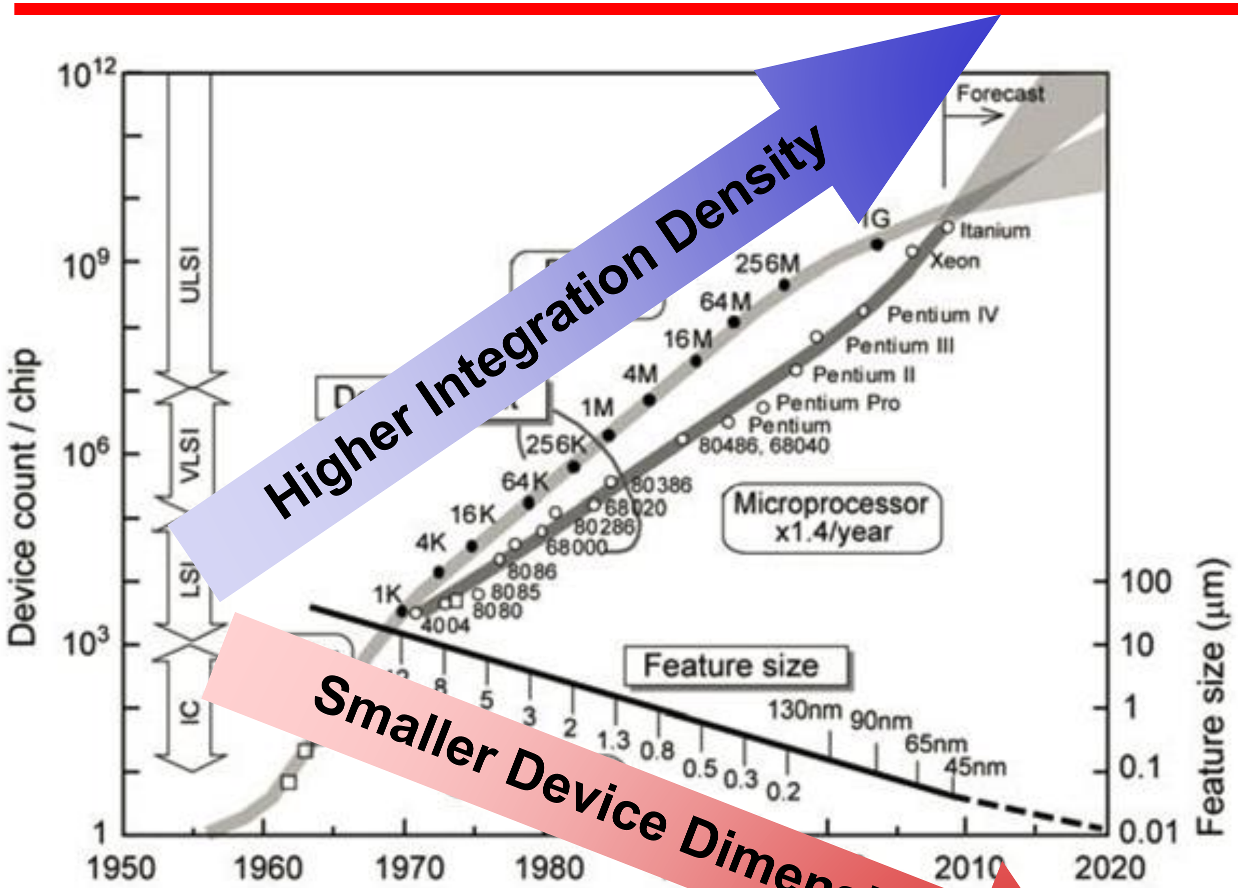
Line-edge and Line-width Roughness (LER and LWR)
-Courtesy of Technology and Manufacturing Group, Intel Corporation



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

[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).

Evolution of Process Variation



-Courtesy of Professor Hideo Sunami at Hiroshima University

Higher Density→ Rare failure event matters

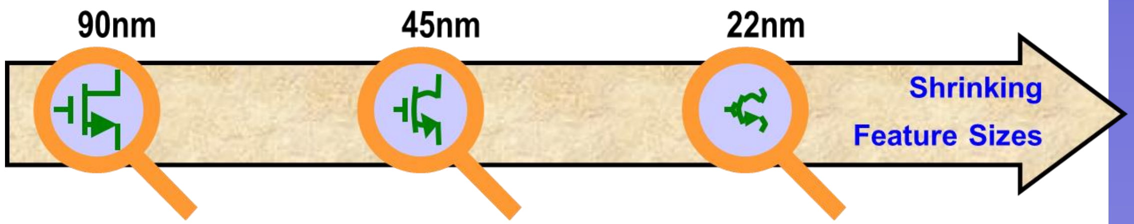
1) 10^6 independent identical standard cells

2) 10^{-6} failure probability

Probability of Single Bit Failure:

$1-(0.999999)^{1,000,000} = 0.6321$

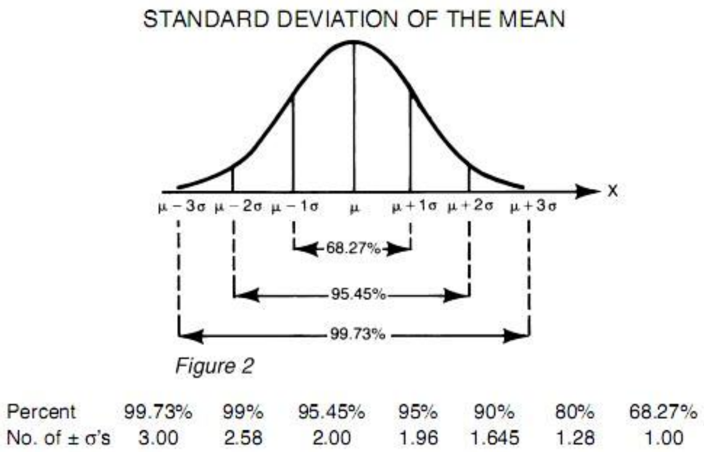
Smaller dimension → Higher impact of process variation



- Rare Event Analysis helps to debug circuits in the **pre-silicon phase** to improve yield rate

Estimating the Rare Failure Event

- **Rare event (a.k.a. high sigma) tail is difficult to achieve with Monte Carlo**
  - **# of simulations required to capture 100 failing samples**



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 required for highly-duplicated circuits and critical 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?**

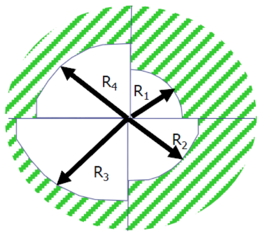
1 Cite from Solido Design Automation whitepaper (Other industrial companies: ProPlus, MunEDA, etc.)

Executive Summary

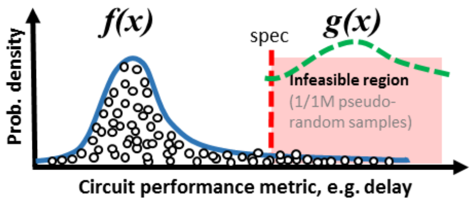
- Background
  - Why statistical circuit analysis, high sigma analysis?
  - **Existing approaches and limitations.**
- Hyperspherical Clustering and Sampling (HSCS)
  - Importance Sampling
  - Applying and optimizing clustering algorithm for high sigma analysis (Why spherical?)
  - Deterministically locating all the failure regions
  - Optimally sample all failure regions
- Experimental Results: very accurate and robust performance
  - Experimental on both mathematical and circuit-based examples

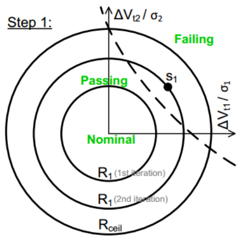
High Sigma Analysis – more details about the tail

- **Draw more samples in the tail**
- **Analytical Approach**
  - Multi-Cone[DAC12]

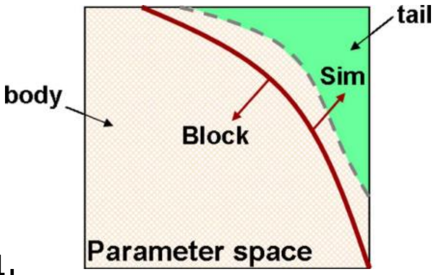


- **Importance Sampling**[DAC06]
  - Shift the sample distribution to more "important" region

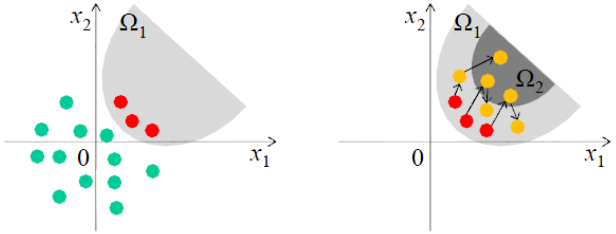




- **Classification based methods**[TCAD09]
  - Filter out unlikely-to-fail samples using classifier



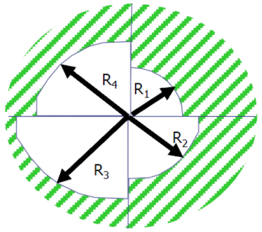
- **Markov Chain Monte Carlo (MCMC)**[ICCAD14]
  - It is difficult to cover the failure regions using a few chain of samples



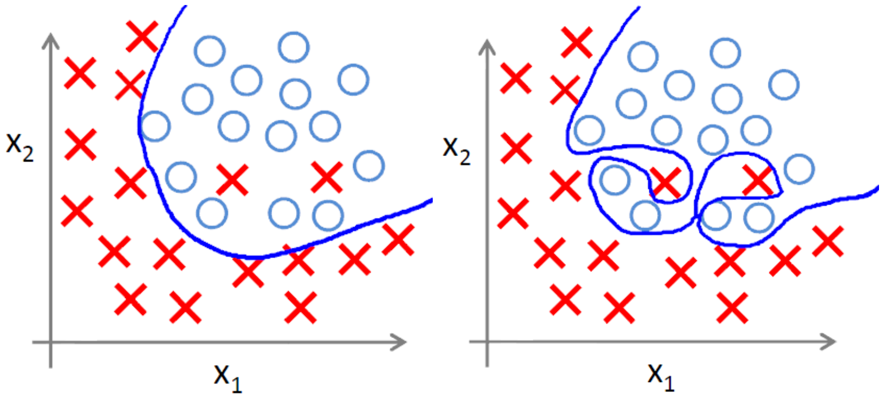
[DAC12] Kanj, Rouwaida, Rajiv Joshi, Zhuo Li, Jerry Hayes, and Sani Nassif. "Yield Estimation via Multi-Cones." DAC 2012
[DAC06] R. Kanj, R. Joshi, and S. Nassif. "Mixture Importance Sampling and Its Application to the Analysis of SRAM Designs in the Presence of Rare Failure Events." DAC, 2006
[TCAD09] 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

Challenges – High Dimensionality

- High Dimensionality
  - Analytical approaches: complexity scales exponentially to the dimension.
    - # of cones in multi-cone



  - IS: can be numerical instable at high dimensional
    - Curse of dimensionality[Berkeley08, Stanford09]
  - Classification based approaches: classifiers perform poorly at high dimensional with limited number of training samples.

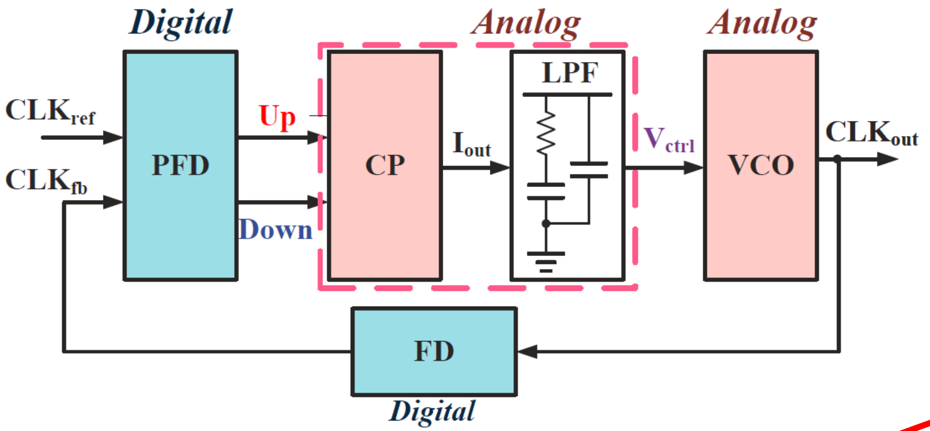


  - MCMC: It is difficult to cover the failure regions using a few chain of samples

[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.
[DAC14] Mukherjee, Parijat, and Peng Li. “Leveraging Pre-Silicon Data to Diagnose out-of-Specification Failures in Mixed-Signal Circuits.” In DAC 2014

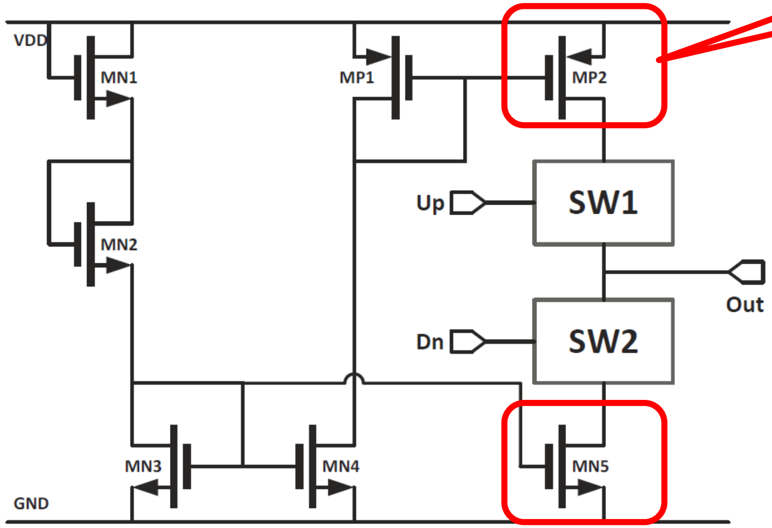
Challenge – Multiple Failure regions

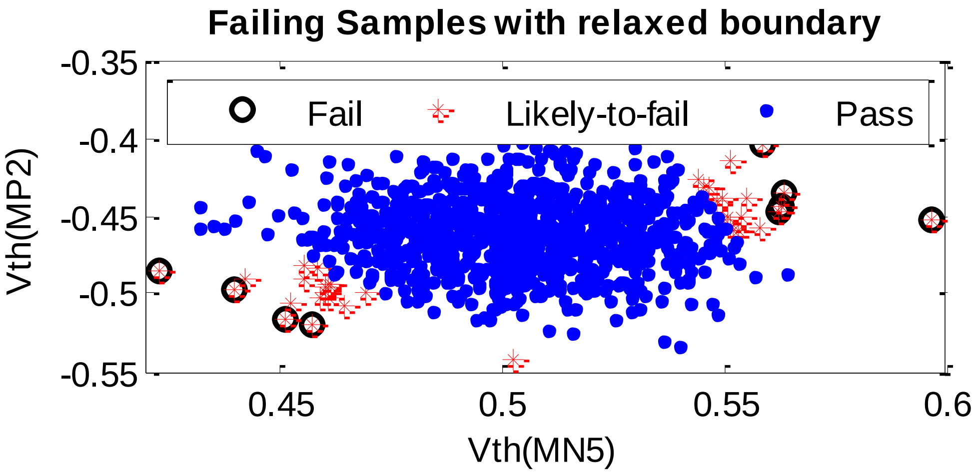
- Failing samples might distribute in multiple disjoint regions
  - A real-life example with multiple failure regions: **Charge Pump (CP) in a PLL**



PFD: phase frequency detector;
CP: Charge pump
FD: frequency divider;
VCO: voltage controlled oscillator

Mismatch between MP2 and MN5 may result in fluctuation of control voltage, which will lead to "jitter" in the clock.





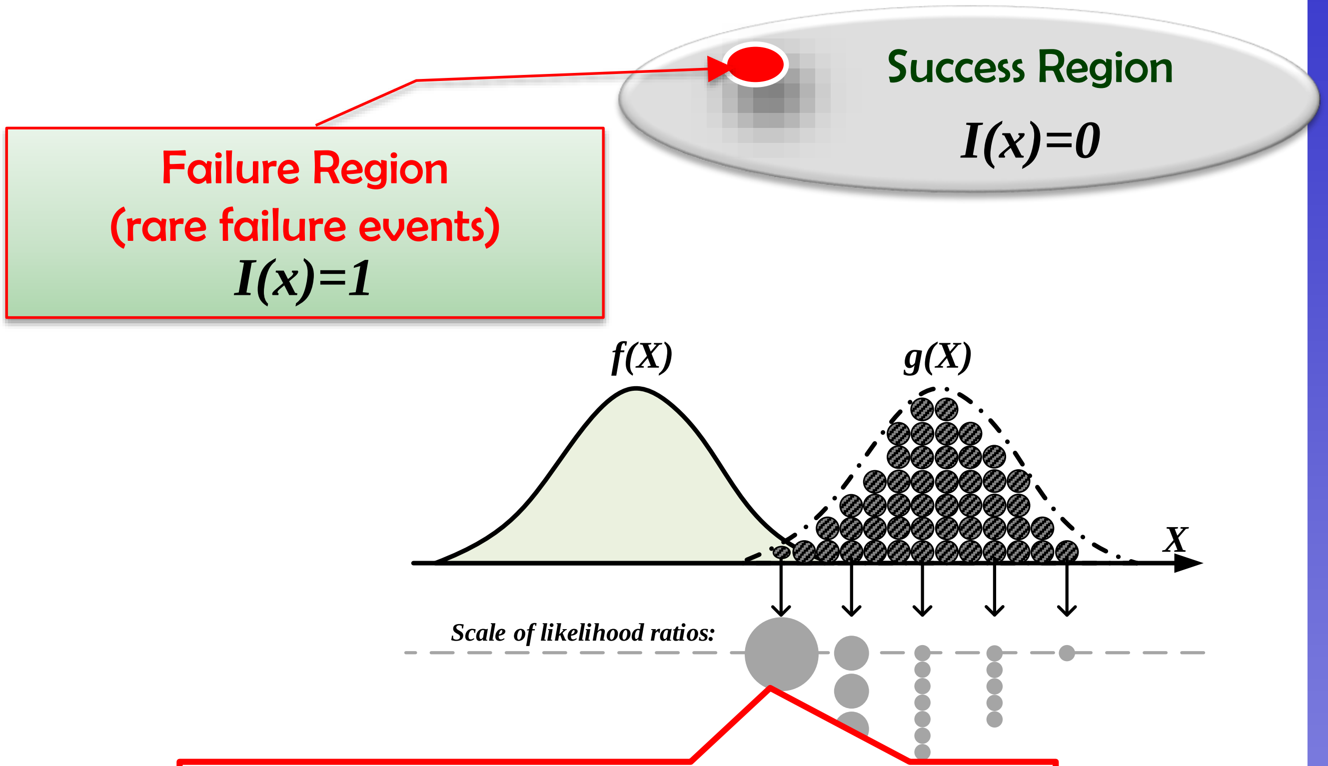
[DAC14] **Wu, Wei**, W. Xu, R. Krishnan, Y. Chen, L. He. "REscope: High-dimensional Statistical Circuit Simulation towards Full Failure Region Coverage", DAC 2014

Outline

- Background
  - Why statistical circuit analysis, high sigma analysis?
  - Limitation of existing approaches.
- **Hyperspherical Clustering and Sampling (HSCS)**
  - Importance Sampling
  - Applying and optimizing clustering algorithm for high sigma analysis (Why spherical?)
  - Deterministically locating all the failure regions
  - Optimally sample all failure regions
- Experimental Results: very accurate and robust performance
  - Experimental on both mathematical and circuit-based examples

Importance Sampling

- A Mathematic interpret of Monte Carlo
  - $P_{Fail} = \int I(x) \cdot f(x) dx$
  - $I(x)$ is the indicator function



- Importance Sampling
  - $P_{Fail} = \int I(x) \cdot f(x) dx$

$$= \int I(x) \cdot \boxed{\frac{f(x)}{g(x)}} \cdot g(x) dx$$

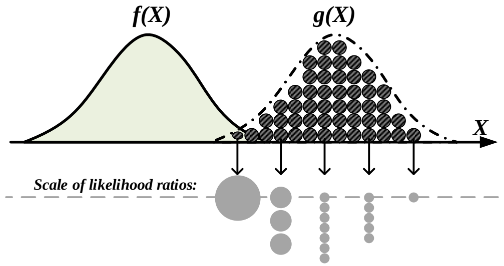
- Likelihood ratio or weight: $\frac{f(x)}{g(x)}$
  - Samples with higher likelihood ratio has high impact to the estimation of P_{fail}
    - Larger f(x), Smaller g(x)
  - Weight $f(x)/g(x)$ might be extremely large at high dimensionality

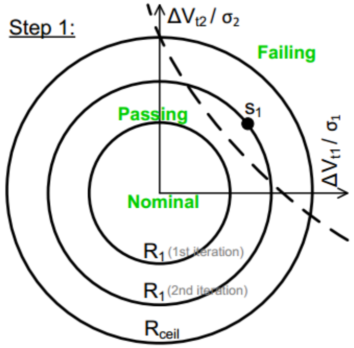
- **Failing samples closed to nominal case has high weights.**
- **Weight can be extremely largle**

[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

To Capture More Important Samples

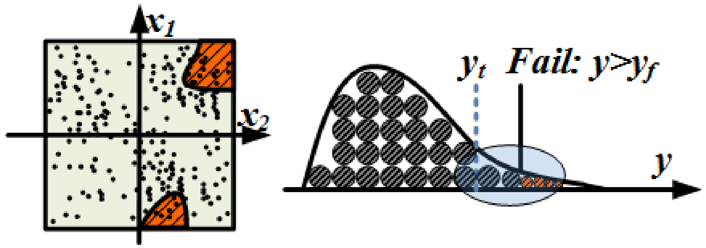
- Spherical Sampling
  - Shift the mean to the failing sample with minimal norm
    - **Min-norm point**
  - Importance Sampling Recap
    - $P_{Fail} = \int I(x) \cdot \frac{f(x)}{g(x)} \cdot g(x) dx$
    - Samples with smaller norm has higher importance
      - Smaller norm ➔ closer to mean ➔ larger ***f(x)***

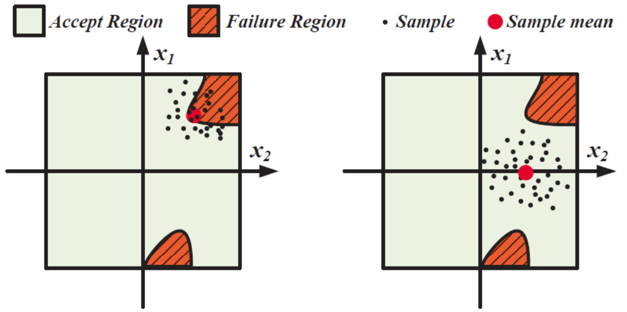




Spherical sampling[DATE10]

- Existing Importance Sampling approaches shift the sample mean to a given point
  - Do **NOT** cover multiple failure regions





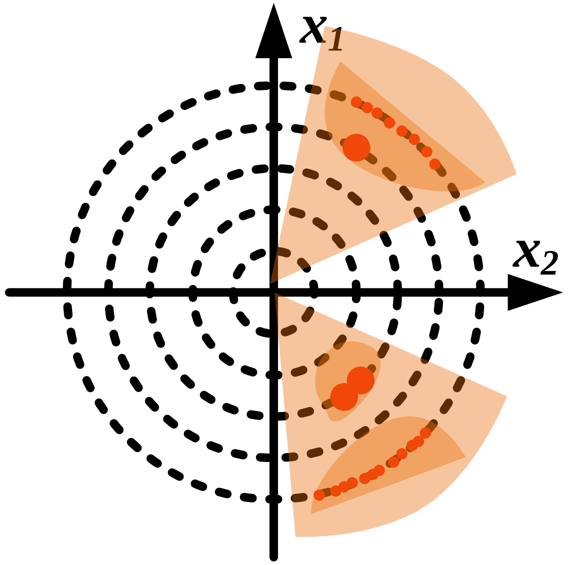
(a) Importance sampling with shifted mean on the boundary of failure region(s)

(b) Importance sampling with shifted mean on the centroid the failure region(s)

[DATE10] M. Qazi, M. Tikekar, L. Dolecek, D. Shah, and A. Chandrakasan, “Loop flattening and spherical sampling: Highly efficient model reduction techniques for SRAM yield analysis,” in DATE’2010

Hyperspherical clustering and sampling (HSCS)

- Hyperspherical Clustering and Sampling (HSCS)[ISPD16]
- Why Clustering?
  - Explicitly locating multiple failure regions
- Why Hyperspherical?
  - Direction (angle) of the failure region is more important
    - Failure regions at the same direction can be covered with samples centered at one min-norm point
    - Failure regions at different directions needs to be covered with samples centered at multiple points
- Hyperspherical Sampling?
  - Explicitly drawing samples around those failure regions

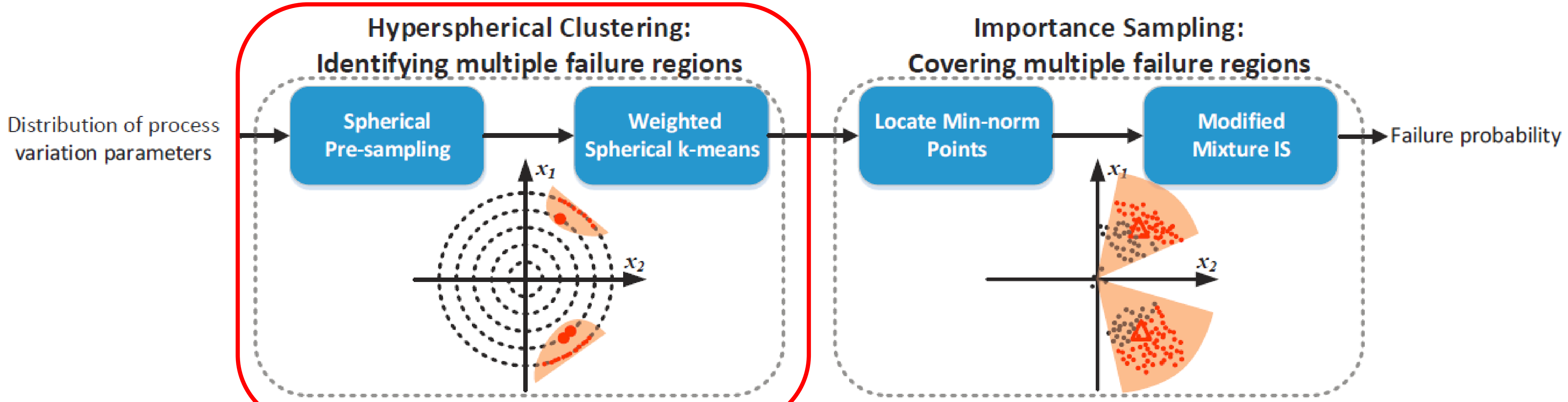


Hyperspherical clustering and sampling (HSCS)

- **Phase 1: Hyperspherical clustering**: identify multiple failure regions
  - Cosine distance v.s. Euclidean distance
    - Pay more attention to the angle over the absolute location

$$EuclideanDistance(X^{(1)}, X^{(2)}) = \left\| X^{(1)} - X^{(2)} \right\|$$

$$CosineDistance(X^{(1)}, X^{(2)}) = 1 - \frac{{X^{(1)}}^T X^{(2)}}{\|X^{(1)}\| \, \|X^{(2)}\|}$$



[ISPD16] **Wei Wu**, Srinivas Bodapati, and Lei He, “Hyperspherical Clustering and Sampling for Rare Event Analysis with Multiple Failure Region Coverage”. ISPD 2016

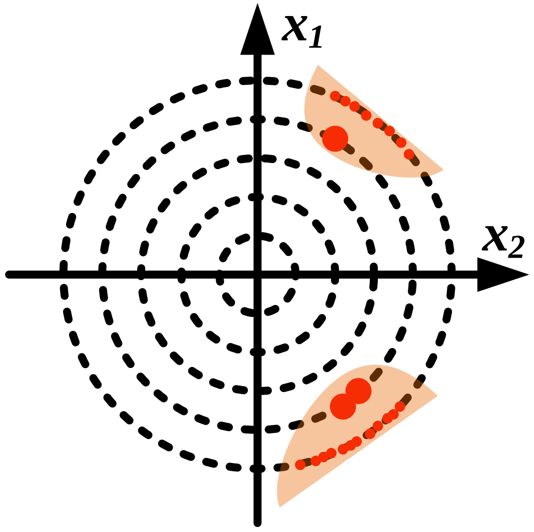
Hyperspherical clustering and sampling (HSCS)

- **Phase 1: Hyperspherical clustering**: identify multiple failure regions
  - Iteratively update cluster centroid
  - Samples are associated with different weight during clustering
    - Cluster centroid are biased to more important samples (with higher weights)

Algorithm 1 Weighted Spherical K-Means Algorithm

Input: A set of M failed samples: $\mathcal{X} = \{X^{(1)}, X^{(2)}, ..., X^{(M)}\}$
Sample weights: $w^{(1)}, w^{(2)}, ..., w^{(M)}$
Number of initial clusters: k
Output: Cluster label for samples: $\mathcal{Y} = \{y^{(1)}, y^{(2)}, ..., y^{(M)}\}$
Updated number of clusters: k

```
Randomly initialize the unit length cluster centroids U = {μ^(1), μ^(2), ..., μ^(k)};
repeat
    Cluster Assignment (update Y):
    For each sample X^(i), set y^(i) = argmax_j X^(i)T μ^(j);
    Remove Empty Clusters (update k)
    Remove X_j if X_j = {X^(i) | y^(i) = j} = ∅;
    Update number of cluster k;
    Weighted Centroid Update (update U):
    For cluster k, let X_j = {X^(i) | y^(i) = j}, update centroid as
    μ^(j) = Σ_{X^(i) ∈ X_j} w^(i) X^(i);
    μ^(j) = μ^(j) / ||μ^(j)||;
until <Y remains unchanged>
Return Y and k;
```



[ISPD16] **Wei Wu**, Srinivas Bodapati, and Lei He, “Hyperspherical Clustering and Sampling for Rare Event Analysis with Multiple Failure Region Coverage”. ISPD 2016

Hyperspherical clustering and sampling (HSCS)

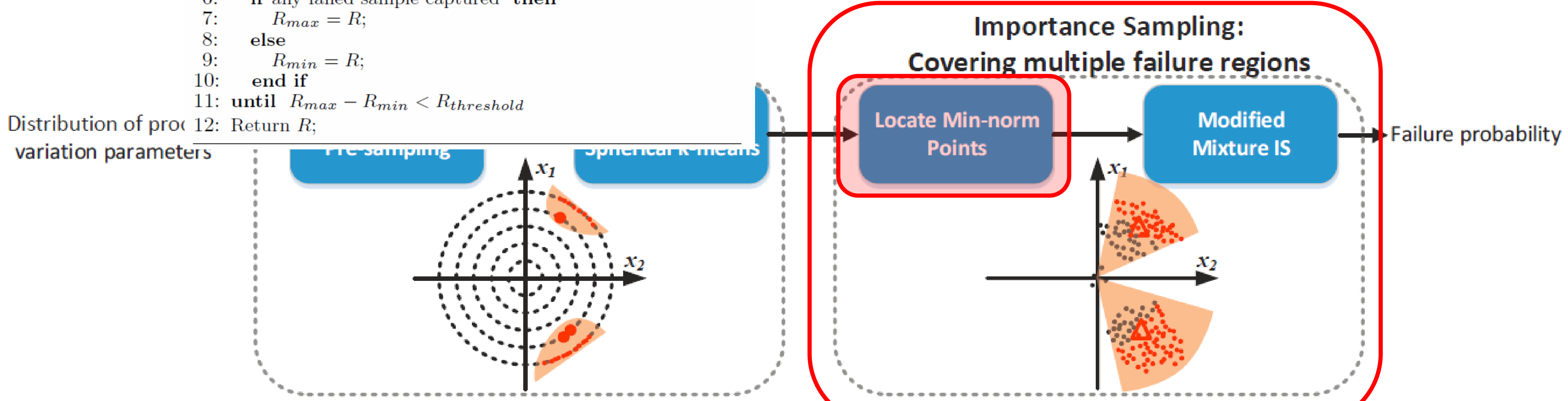
- **Phase 2: Spherical sampling**: draw samples around **multiple min-norm points**
  - Locate Min-norm Points via bisection

Algorithm 2 Locate min-norm points for each cluster with bisection

Input: Minimal radius of existing failure samples, R

Output: Radius of min-norm point: R_{min}

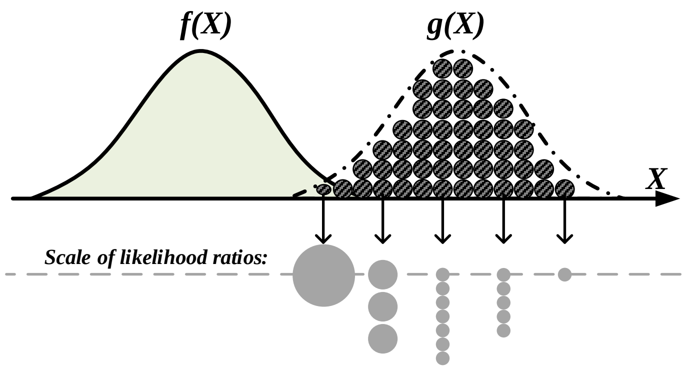
```
R_max = R;
R_min = 0;
repeat
    R = (R_max + R_min)/2;
    simulate a small set of samples at Radius = R in current cluster;
    if any failed sample captured then
        R_max = R;
    else
        R_min = R;
    end if
until R_max − R_min < R_threshold
Return R;
```

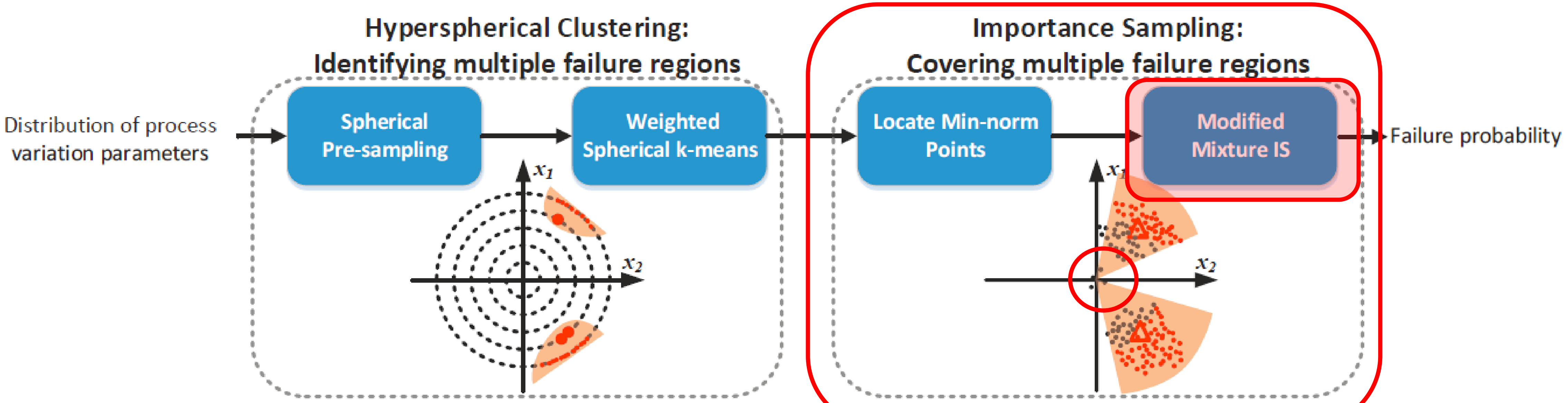


[ISPD16] **Wei Wu**, Srinivas Bodapati, and Lei He, “Hyperspherical Clustering and Sampling for Rare Event Analysis with Multiple Failure Region Coverage”. ISPD 2016

Hyperspherical clustering and sampling (HSCS)

- **Phase 2: Spherical sampling**: draw samples around **multiple min-norm points**
  - Avoid instable weights:
  - $P_{Fail} = \int I(x) \cdot \frac{f(x)}{g(x)} \cdot g(x)dx$
    - Where $g(x) = \boxed{\boldsymbol{\alpha f(x)}} + (1-\alpha)\sum_{\forall\, clusters} \beta_i f(X - C_i)$
    - $g(x)$ samples around multiple min-norm points
    - $\frac{f(x)}{g(x)}$ **is always bounded by** $\frac{1}{\alpha}$





[ISPD16] **Wei Wu**, Srinivas Bodapati, and Lei He, "Hyperspherical Clustering and Sampling for Rare Event Analysis with Multiple Failure Region Coverage". ISPD 2016

Outline

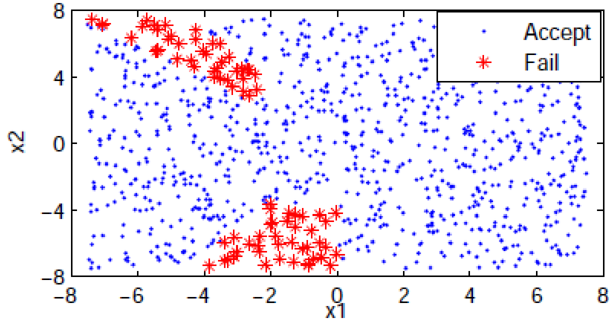
- Background
  - Why statistical circuit analysis, high sigma analysis?
  - Limitation of existing approaches.

- Hyperspherical Clustering and Sampling (HSCS)
  - Importance Sampling
  - Applying and optimizing clustering algorithm for high sigma analysis (Why spherical?)
  - Deterministically locating all the failure regions
  - Optimally sample all failure regions

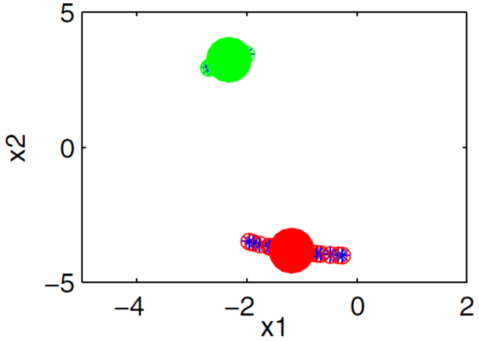
- **Experimental Results: very accurate and robust performance**
  - Experimental on both mathematical and circuit-based examples

Demo on mathematically known distribution

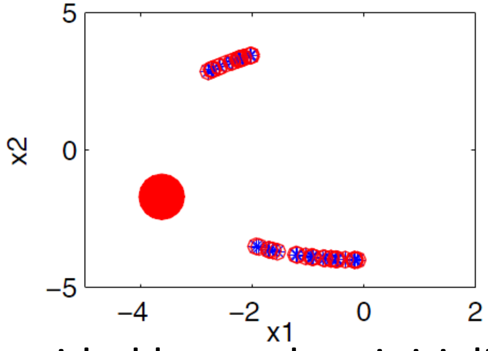
- 2-D distribution with 2 known failure regions (7.199e-5)
  - $\mathcal{S}_1 = \{X \mid \|X\| > 3.8 \text{ and } \phi(X) \in [\frac{2}{3}\pi, \frac{3}{4}\pi]\}$
  - $\mathcal{S}_2 = \{X \mid \|X\| > 3.9 \text{ and } \phi(X) \in [\frac{4}{3}\pi, \frac{3}{2}\pi]\}$



Step2: Clustering

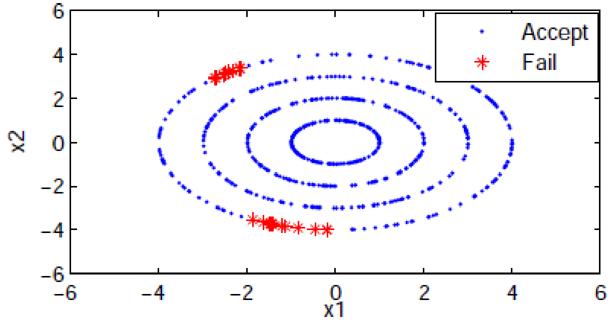


Potential clustering failure

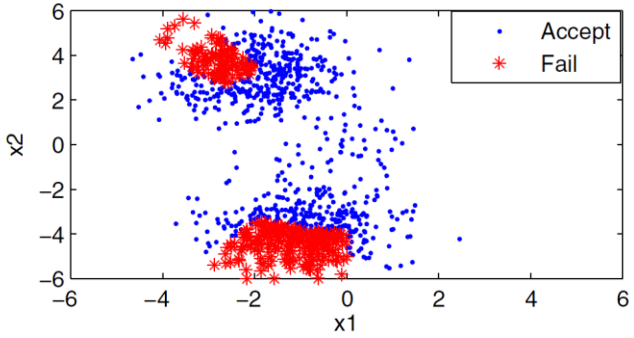


Can be avoided by random initialization

Step1: Spherical Presampling



Step3&4: locate min-norm points and IS



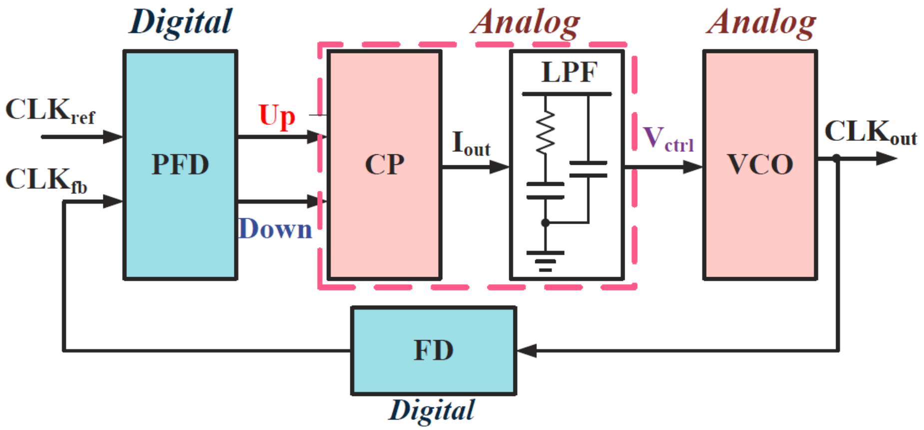
Results:

Theoretical: 7.199e-5

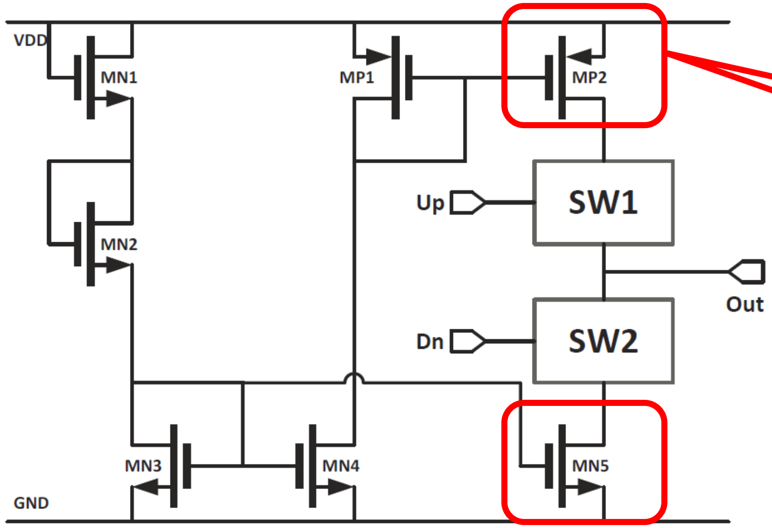
HSCS: 7.109e-5

A real-life example with multiple failure regions

- Charge Pump (CP) in a PLL



PFD: phase frequency detector;
CP: Charge pump
FD: frequency divider;
VCO: voltage controlled oscillator

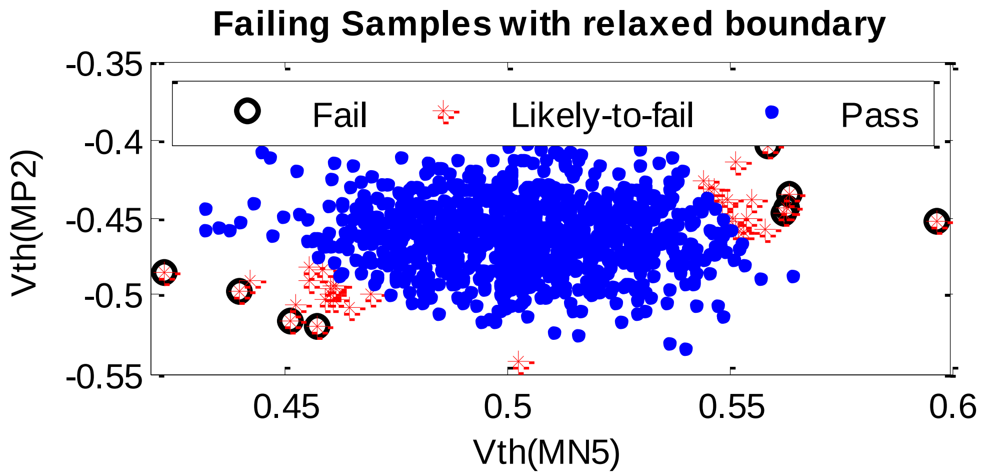


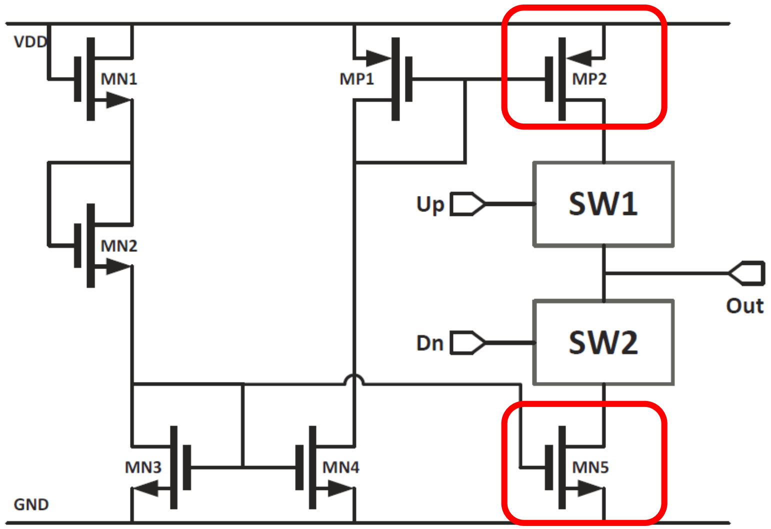
Mismatch between MP2 and MN5 may result in fluctuation of control voltage, which will lead to "jitter" in the clock.

[DAC14] **Wu, Wei**, W. Xu, R. Krishnan, Y. Chen, L. He. "REscope: High-dimensional Statistical Circuit Simulation towards Full Failure Region Coverage", DAC 2014

A real-life example with multiple failure regions

- **Two setups of this circuit**
- Low dimensional setup
  - For demonstration of multiple failure regions
  - 2 random variables, V_{TH} of MP2 and MN5

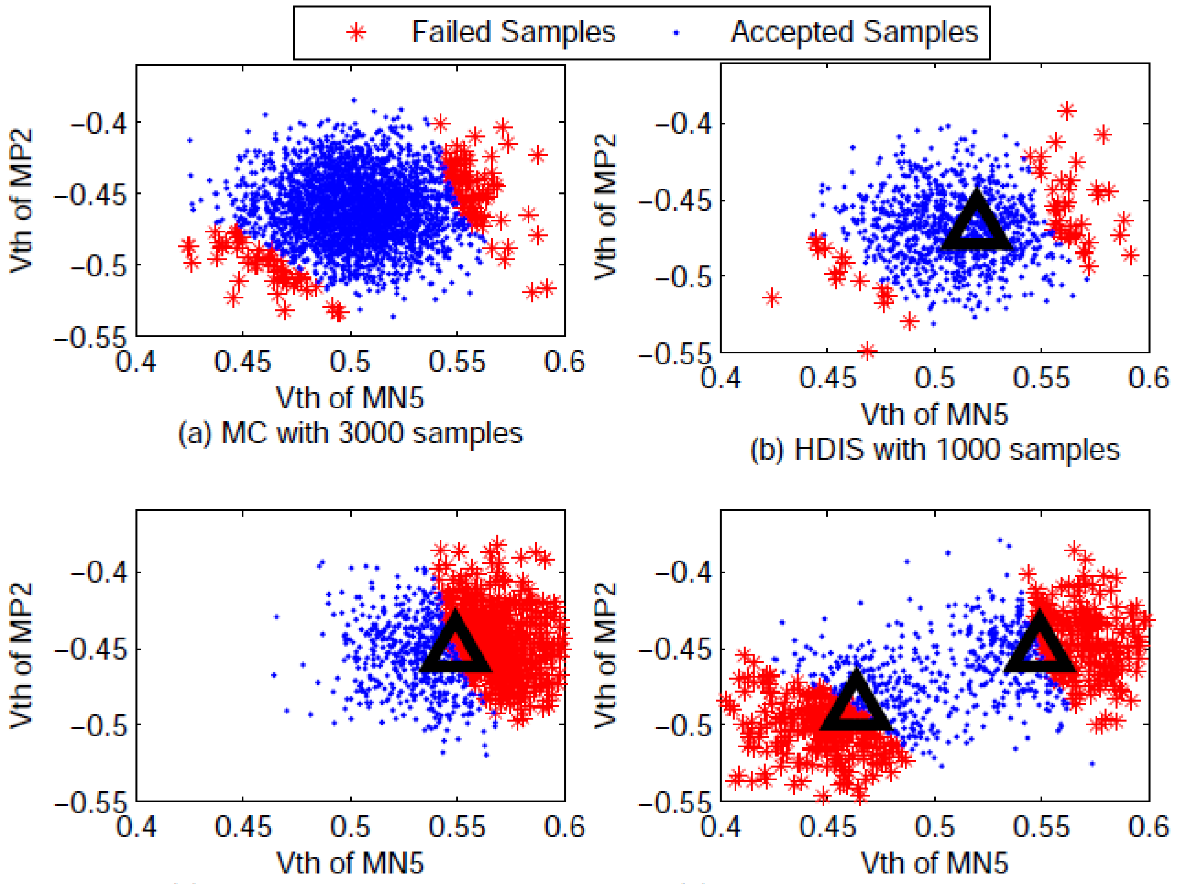




- High dimensional setup
  - A more realistic setup
  - 70 random variables on all 7 transistors (ignore the variation in SW's)

Compared with other importance sampling methods

- Importance samples drew by HDIS, Spherical Sampling, and HSCS
  - **Δ**’s are the sample means of different IS implementations.



(a) MC with 3000 samples

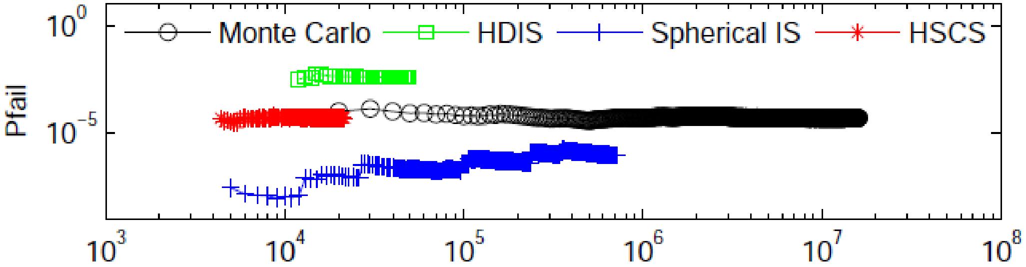
(b) HDIS with 1000 samples

(c) SPIS with 1000 samples

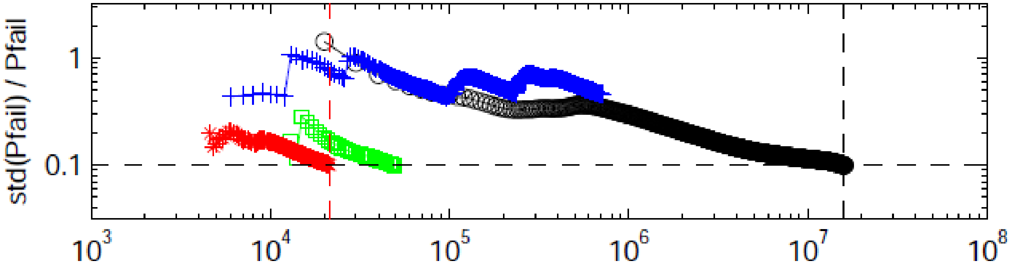
(d) Proposed HSCS with 1000 samples

Accuracy and Speedup

- On high dimensional setup (70-dimensional)



(a) Failure Probability (Pfail) v.s. # of samples



(b) Standard Deviation of Pfail v.s. # of samples

About 700X speedup over MC

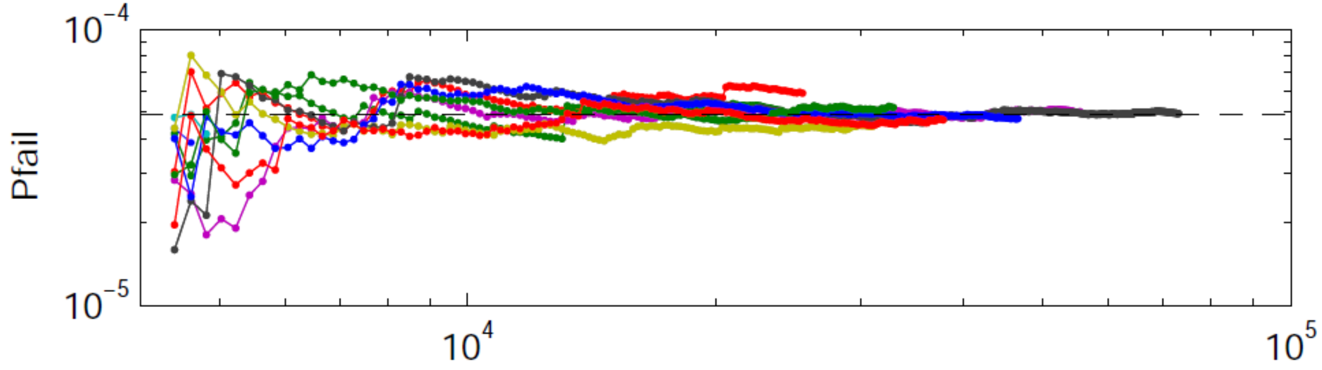
Table 1: Accuracy and efficiency evalution on 70-dimensional charge pump circuit

	Monte Carlo	HDIS [8]	SpIS [6]	Proposed HSCS with 10 replications
failure probability	**4.904e-5**	**3.9e-3**	**8.788e-7**	3.89e-5 ∼ 5.88e-5 (mean **4.82e-5**)
Total #sim. runs	**1.584e7**	**3.8e4**	**>7.4e5**	4.6e3 ∼ 5.5e4 (mean **2.3e4**)
#sim. for presampling	-	1.1e4	4e3	4.2e3
#sim. for IS		3.8e4	>7e5	410 ∼ 5.1e4 (mean 1.9e4)

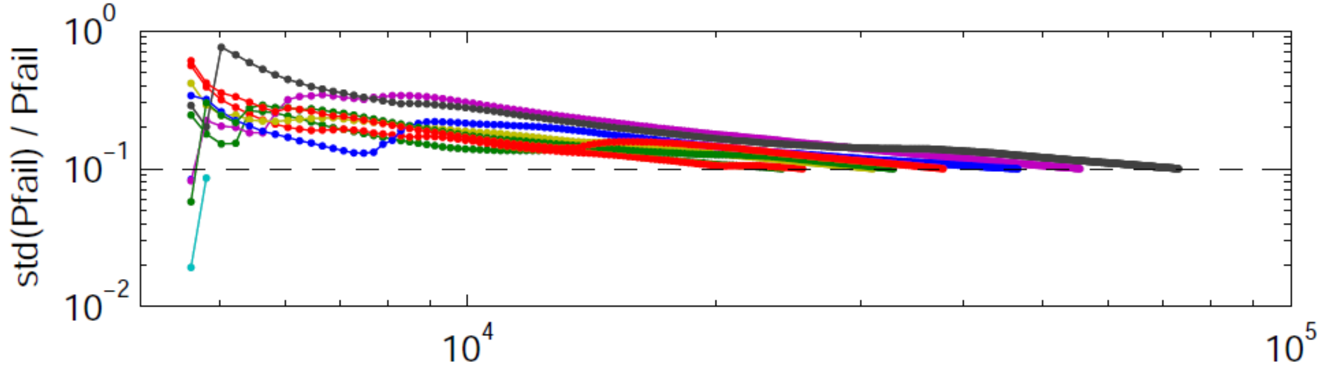
- Determine the # of clusters in HSCS

Robustness

- HSCS is executed with **10 replications**, yielding very consistent results.
  - Failure rate: : 3.89e-5 ~ 5.88e-5 (mean **4.82e-5**, **MC: 4.904e-5**)
  - # of simulation: 4.6e3 ~ 5.5e4 (mean **2.3e4**, **MC: 1.584e7**)



(a) Failure Probability (Pfail) v.s. # of samples



(b) Standard Deviation of Pfail v.s. # of samples

Summary

- Deterministically locating all the failure regions
  - Cluster samples based on **Cosine distance** instead of Euclidean distance
  - Center of failure regions are biased to important samples (higher weights)

- Optimally sampling all the failure regions
  - Locate the min-norm points of each failure region
  - Shift the sampling means to the min-norm points

- Very accurate and robust performance
  - On mathematical and circuit-based examples with multiple replications

Q&A

Thank you for attention!

Please address comments to weiw@seas.ucla.edu

Determine the # of Clusters

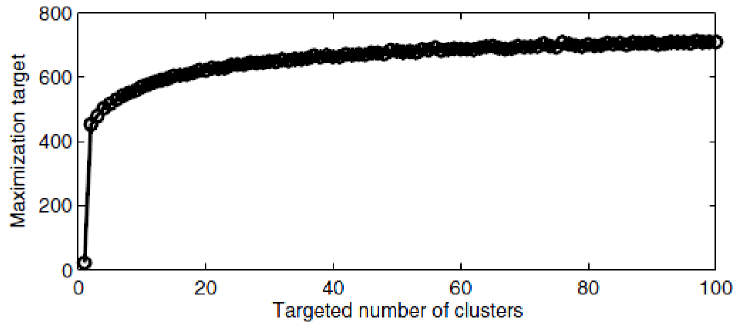


Figure 11: Clustering maximization objective while changing the targeted number of clusters

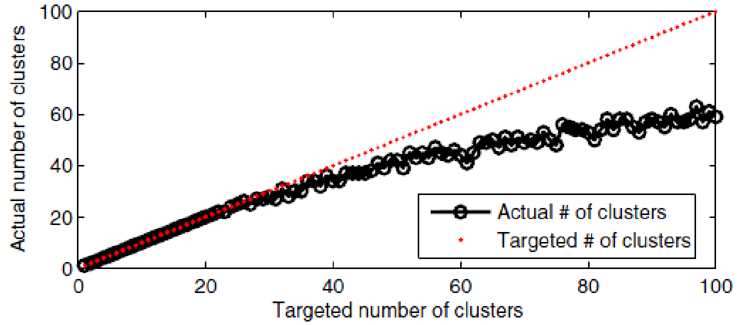


Figure 12: Number of actually clusters may be small than the targeted number of clusters

- [Go back]