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

Wei Wu¹, Srinivas Bodapati², Lei He¹,³

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

Evolution of Process Variation

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

- 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)

- How to efficiently and accurately estimate $P_{\text{fail}}$ (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

[^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 unstable 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

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.

Failing Samples with relaxed boundary

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 Mathematical interpretation of Monte Carlo
  - $P_{\text{Fail}} = \int I(x) \cdot f(x)\,dx$
  - $I(x)$ is the indicator function

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

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

- Failing samples close to nominal case have high weights.
- Weight can be extremely large

To Capture More Important Samples

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

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

---

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

\[
\text{Euclidean Distance}(X^{(1)}, X^{(2)}) = \|X^{(1)} - X^{(2)}\|
\]
\[
\text{Cosine Distance}(X^{(1)}, X^{(2)}) = 1 - \frac{X^{(1)^T}X^{(2)}}{\|X^{(1)}\| \|X^{(2)}\|}
\]

[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 $X = \{X^{(1)}, X^{(2)}, ..., X^{(M)}\}$
- Sample weights: $w^{(1)}, w^{(2)}, ..., w^{(M)}$
- Number of initial clusters: $k$

Output: Cluster label for samples $Y = \{y^{(1)}, y^{(2)}, ..., y^{(M)}\}$
- Updated number of clusters: $k$

1. Randomly initialize the unit length cluster centroids $U = \{\mu^{(1)}, \mu^{(2)}, ..., \mu^{(k)}\}$;
2. repeat
3. Cluster Assignment (update $Y$):
   - For each sample $X^{(i)}$, set $y^{(i)} = \arg\max_j X^{(i)^T} \mu^{(j)}$;
4. Remove Empty Clusters (update $k$)
   - Remove $X^{(i)}$ if $X^{(i)} \notin \{y^{(1)}, y^{(2)}, ..., y^{(M)}\}$
   - Update number of cluster $k$;
5. Weighted Centroid Update (update $U$):
   - For cluster $k$, let $X^{(i)} \in X^{(i)}|y^{(i)} = j$; update centroid as
     - $\mu^{(j)} = \sum_{X^{(i)} \in X^{(i)}|y^{(i)} = j} w^{(i)} X^{(i)}$
     - $\mu^{(j)} = \mu^{(j)} / ||\mu^{(j)}||$;
6. until $\langle Y \rangle$ remains unchanged
7. 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
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, $R_{\text{min}}$

1: $R_{\text{max}} = R$
2: $R_{\text{min}} = 0$
3: repeat
4: $R = (R_{\text{max}} + R_{\text{min}})/2$
5: simulate a small set of samples at $R$
6: if any failed sample captured then
7: $R_{\text{max}} = R$
8: else
9: $R_{\text{min}} = R$
10: end if
11: until $R_{\text{max}} - R_{\text{min}} < \text{threshold}$
12: Return $R$
```

Importance Sampling:
- Covering multiple failure regions

---

[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:
  \[ P_{\text{Fail}} = \int I(x) \cdot \frac{f(x)}{g(x)} \cdot g(x) dx \]
  - Where \( g(x) = \alpha f(x) + (1 - \alpha) \sum_{\text{clusters}} \beta_i f(X - C_i) \)
  - \( g(x) \) samples around multiple min-norm points
  - \( \frac{f(x)}{g(x)} \) is always bounded by \( \frac{1}{\alpha} \)

**Scale of likelihood ratios:**

\[ f(X) \quad g(X) \]

**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)
  - $S_1 = \{X | \|X\| > 3.8 \text{ and } \phi(X) \in [\frac{2}{3} \pi, \frac{3}{4} \pi]\}$
  - $S_2 = \{X | \|X\| > 3.9 \text{ and } \phi(X) \in [\frac{4}{3} \pi, \frac{3}{2} \pi]\}$

Step 1: Spherical Presampling

Step 2: Clustering

Potential clustering failure

Can be avoided by random initialization

Step 3 & 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.

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, $V_{\text{TH}}$ 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
  - $\Delta$’s are the sample means of different IS implementations.
Accuracy and Speedup

- On high dimensional setup (70-dimensional)

About 700X speedup over MC

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

<table>
<thead>
<tr>
<th></th>
<th>Monte Carlo</th>
<th>HDIS [8]</th>
<th>Spherical IS</th>
<th>Proposed HSCS with 10 replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure probability</td>
<td>4.904e-5</td>
<td>3.9e-3</td>
<td>8.788e-7</td>
<td>3.89e-5 ~ 5.88e-5 (mean 4.82e-5)</td>
</tr>
<tr>
<td>Total #sim. runs</td>
<td>1.584e7</td>
<td>3.8e4</td>
<td>&gt;7.4e5</td>
<td>4.6e3 ~ 5.5e4 (mean 2.3e4)</td>
</tr>
<tr>
<td>#sim. for presampling</td>
<td>-</td>
<td>1.1e4</td>
<td>4e3</td>
<td>4.2e3</td>
</tr>
<tr>
<td>#sim. for IS</td>
<td>-</td>
<td>3.8e4</td>
<td>&gt;7e5</td>
<td>410 ~ 5.1e4 (mean 1.9e4)</td>
</tr>
</tbody>
</table>
Robustness

- HSCS is executed with **10 replications**, yielding very consistent results.
  - Failure rate: $3.89 \times 10^{-5} \sim 5.88 \times 10^{-5}$ (mean $4.82 \times 10^{-5}$, MC: $4.904 \times 10^{-5}$)
  - # of simulation: $4.6 \times 10^3 \sim 5.5 \times 10^4$ (mean $2.3 \times 10^4$, MC: $1.584 \times 10^7$)

![Graphs of Failure Probability (Pfail) vs. # of samples and Standard Deviation of Pfail vs. # 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
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