Incremental Latin Hypercube Sampling

for Lifetime Stochastic Behavioral Modeling of Analog Circuits

Yen-Lung Chen*, Wei Wu*, Chien-Nan Jimmy Liu* and Lei He*
EE Dept., National Central University, Taiwan*
EE Dept., University of California, Los Angeles, CA, USA*

Speaker
Outline

- Background of Lifetime Yield Analysis
- Proposed Incremental Latin Hypercube Sampling
- Experimental Results
- Conclusions
- In deep-submicron, statistical devices induce statistical performances

- **Parametric variability induces serious yield loss issues**
  - Parametric variability will dominate yield loss
  - Yield affects the total cost of the products

Aging Effects & Parameter Degradation

- Aging effects change circuit behavior with time
  - Negative-bias temperature instability (NBTI), hot-carrier injection (HCI)...

- Performance degrades when exposed in the ambient air or under continuous bias-stress
  - $V_t$ is changed a lot over time
  - Impact the circuit performances
  - Reduce yield and reliability significantly

- New technology (ex: flexible TFT) has serious challenge from aging effects
  - $V_t$ is changed a lot in seconds

**Lifetime Yield**

- **Lifetime Yield** = Process Variations + Aging effects
  - Evaluate the reliability after a period of time

- Lifetime yield analysis often requires iterative circuit performance simulation
  - Evaluating the performance at EACH time step → high cost !!

![Diagram showing PDF and performance degradation over time](image)
Monte Carlo Simulation

- Still the golden reference for yield analysis
- Simulate a lot of random samples
  - Analyze the performance distribution under process variations
  - High analysis cost
- Infeasible to do the MC analysis at each time step …
Possible Ways to Reduce Complexity

- **Simplified simulation model**
  - Use behavioral model or equation-based model to predict the circuit performance
  - Simulation time is reduced, but estimation accuracy is also reduced

- **Compact sample generation**
  - Use special sampling techniques (ex: QMC or LHS) to generate the samples for MC simulation
  - Due to the fast convergence property, the required number of samples can be reduced

- **Performance distribution estimation**
  - Also called stochastic modeling technique
  - Use the results of a few samples to estimate the whole probability distribution
Quadratic Model for Lifetime Yield

- Use equation-based model to predict the performance distribution after a given period of time
  - Pretty fast estimation without iterations
  - Non-linear aging effects are hard to be predicted → large error exists

\[
y(t) = y(t_0) + \frac{dy(t)}{dt} \bigg|_{t_0} \cdot (t - t_0) + \frac{d^2 y(t)}{dt^2} \bigg|_{t_0} \cdot (t - t_0)^2
\]

Quasi Monte Carlo (QMC) method generates low-discrepancy sequences based on specific pseudorandom numbers.

Latin Hypercube Sampling (LHS) is a variant of QMC method:
- Each group in the sampling space contains only one single sample.
- Guarantee all the samples with low dependence.

Control the sample distribution for fast convergence:
- Less samples are required to reach the same accuracy → speedup!!

![Random Sampling](image1.png)
![Quasi-random Sampling](image2.png)
![Latin Hypercube Sampling](image3.png)
Stochastic Behavioral Modeling

- **Moment matching-based method**
  - A fast way to estimate the probability distribution with less samples
  - Calculate the probabilistic moments as
    \[ m_p^k = \frac{1}{N} \sum x_i^k \]
  - Solve the resulting nonlinear equation system to obtain residues \( a_i \) and poles \( b_i \) of \( h(t) \), which is the pdf\( (x) \)

\[ \begin{bmatrix} 1/b_1 & 1/b_2 & \cdots \ \ \\
                  1 & 1 & \cdots \ \\
                       b_1^2 & \cdots & \cdots \ \\
\end{bmatrix} \begin{bmatrix} a_1 \\
                        a_2 \\
                        \vdots \ \\
\end{bmatrix} = \begin{bmatrix} m_1 \\
                                m_2 \\
                                \vdots \ \\
\end{bmatrix} \]

- **Performance PDF**

\[ pdf(x) = \sum a_i \cdot e^{b_i \cdot y_p} \]

“Incremental” Sampling for Aging Analysis

- Circuit behavioral is not changed dramatically at each time step during aging analysis

- **Reuse most of samples** and incrementally update a small portion of samples
  - Reduce #simulations for aging analysis significantly

- How to keep the randomness property of samples?
  - follow the **LHS property** to ensure fast convergence
    - Each row and each column has only one sample !!

- **Stochastic modeling** is adopted to further reduce the samples for estimating the performance distribution
  - **Incremental moment matching** is proposed in this work
Outline

- Background of Lifetime Yield Analysis
- Proposed Incremental Latin Hypercube Sampling
- Experimental Results
- Conclusions
Flowchart of Incremental LHS

Input
- Initial samples
- Circuit database

Output
- Aging PDF

Aging Model
- Exponential Equation [5]

The purpose of sample analysis:

- Reuse the majority of samples
- Remove some redundant samples
- Add few new samples
Proposed Incremental LHS Method

- The performance of each sample may be changed after aging
  - Modify performance by estimation to consider aging
- Add/remove samples to keep the LHS property
  - Check each row and each column for required/redundant samples
- Most of the samples are reused!!

Diagram:
- Before aging:
  - Distribution of samples
- After aging:
  - Updated distribution with added and removed samples
  - Arrows indicating added and removed samples
Incremental Moment Matching

- Not all the calculations need to be redo

1) **Re**-Calculate the probabilistic moments as
   - \( m_{p, old}^k = \frac{1}{N} \cdot \sum_{old} x_i^k \)
   - \( m_{p, new}^k = \frac{1}{N} \cdot (\sum_{old} x_i^k - \sum_{reduced} x_i^k + \sum_{increased} x_i^k) \)
   - Only need to consider the “incremental” samples rather than all the samples

2) **Re**-Match to time moments \( m_t^k \)

3) **Re**-Solve the nonlinear system and obtain the new \( pdf(x) \)

**Probabilistic moments of N samples**

\[
m_p^k = \int_{-\infty}^{\infty} x^k pdf(x) dx
\]

**Time moments of LTI system \( h(t) \)**

\[
m_t^k = \frac{(-1)^k}{k!} \int_{-\infty}^{\infty} x^k h(x) dx
\]

\[
\begin{bmatrix}
\frac{1}{b_1} & \frac{1}{b_2} & \cdots \\
1 & \frac{1}{b_2} & \cdots \\
\frac{1}{b_1^2} & \cdots & \cdots
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots
\end{bmatrix}
= \begin{bmatrix}
m_1 \\
m_2 \\
\vdots
\end{bmatrix}
\]

\[pdf(x) = \sum_r a_i \cdot e^{b_i \cdot y_p}\]

Outline

- Background of Lifetime Yield Analysis
- Proposed Incremental Latin Hypercube Sampling
- Experimental Results
- Conclusions
Experimental Environment

- Perform on PC with Intel 2-core 2.50GHz CPU and 2GB memory
- Demonstrated with flexible TFT to observe clear aging effects
  - ITRI a-Si 8µm technology
  - OPA circuit with flexible TFTs [6]
  - 4-bit digital-to-analog converter (DAC) [6]

- Methods for comparison
  - MC simulation
  - Quadratic model [3]
  - Proposed incremental LHS method

Result Comparison of OPA Circuit

- iLHS achieves 243x average speedup from t=0s to t=10000s
- The accuracy of quadratic model is decreasing over time
  - The prediction error of the non-linear aging rate

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>MC (6k)</th>
<th>MC (3k)</th>
<th>Quad.</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>100%</td>
<td>98%</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>2000+2000</td>
<td>500</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>100%</td>
<td>97%</td>
<td>92%</td>
<td>99%</td>
</tr>
<tr>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
<td>13</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>100%</td>
<td>97%</td>
<td>85%</td>
<td>99%</td>
</tr>
<tr>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>100%</td>
<td>97%</td>
<td>74%</td>
<td>99%</td>
</tr>
<tr>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>10000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>100%</td>
<td>97%</td>
<td>83%</td>
<td>99%</td>
</tr>
<tr>
<td># samples</td>
<td>60000</td>
<td>30000</td>
<td>4000</td>
<td>702</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speedup</td>
<td>1x</td>
<td>2x</td>
<td>150x</td>
<td>85x</td>
</tr>
</tbody>
</table>
Proposed method achieves 99% accuracy with all time step configurations

- Because of the property of LHS can be kept at all time

![Performance Distribution of OPA Circuit Graph](image)
Result Comparison of DAC Circuit

- iLHS achieves 242x average speedup from t=0s to t=10000s
- The accuracy of quadratic model is still low
  - The prediction error of the non-linear aging rate

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>MC (6k)</th>
<th>MC (3k)</th>
<th>Quad.</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100%</td>
<td>98%</td>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>2000+2000</td>
</tr>
<tr>
<td>100</td>
<td>100%</td>
<td>97%</td>
<td>89%</td>
<td>98%</td>
</tr>
<tr>
<td></td>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
</tr>
<tr>
<td>1000</td>
<td>100%</td>
<td>96%</td>
<td>82%</td>
<td>98%</td>
</tr>
<tr>
<td></td>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
</tr>
<tr>
<td>10000</td>
<td>100%</td>
<td>96%</td>
<td>73%</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td># samples</td>
<td>6000</td>
<td>3000</td>
<td>-</td>
</tr>
<tr>
<td>Overall</td>
<td>100%</td>
<td>97%</td>
<td>82%</td>
<td>98%</td>
</tr>
<tr>
<td></td>
<td># samples</td>
<td>60000</td>
<td>30000</td>
<td>4000</td>
</tr>
<tr>
<td></td>
<td>Speedup</td>
<td>1x</td>
<td>2x</td>
<td>150x</td>
</tr>
</tbody>
</table>
Reduction on Simulation Samples

- Only hundreds of samples are required to re-simulate in proposed incremental LHS method
  - 85x ↑ speed up
Conclusions

- Incremental LHS method is proposed for aging analysis
  - Aging effects change the circuit behavior gradually

- Only a small portion of samples are incrementally updated at each time step in aging analysis
  - Reuse previous samples to greatly reduce the simulation efforts

- Stochastic modeling is adopted to further reduce #samples
  - Incremental moment matching is also proposed in this work

- Experimental results achieve 85x speedup over traditional reliability analysis method with similar accuracy
  - Demonstrated on OPA and DAC circuits
Thanks for your listening !!! ☺

Email: jimmy@ee.ncu.edu.tw