From Wall Street to Silicon Valley: Using the Mathematics of Money & Risk for Fast Statistical IC Design

Rob A. Rutenbar (and Amith Singhee)
Professor, Electrical & Computer Engineering
rutenbar@ece.cmu.edu
The Problem: Statistical Variation

- At nanoscale, nothing is deterministic anymore
- How to evaluate designs?

Random Dopant Fluctuations

Line Edge Roughness
K. Shepard, U. Columbia

Gate Oxide Variation
To Evaluate Circuit Impact: Monte Carlo

- Sample each statistical variable
- Parameterize one circuit, simulate it
- Repeat--n samples
Monte Carlo Math: Just A Big Integral

\[ \text{Est} = \int \cdots \int f_{\text{ckt}}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \]

\[ \approx \frac{1}{n} \sum F(\mathbf{u}) \]

Can transform to sample uniformly from \( s \)-dim unit cube

\( s \) statistical vars
\( s \)-dimensional prob
Evaluate Circuit Impact: Monte Carlo

- **PRO:** Accurate, flexible, general
- **CON:** Slow, slow, slow...

\[ \approx \left( \frac{1}{n} \right) \sum \hat{F}(u) \]

Uniform random s-dim sample

\[ V_{T1} = \]
\[ V_{T2} = \]
\[ V_{T3} = \]
\[ V_{T4} = \]
\[ V_{TN} = \]
Why is Monte Carlo Painful?

- High-dim problems: \( s \) is big (100-1000)
- Profoundly nonlinear: *Nanoscale physics*
- Accuracy matters: \(~1-5\%\) error
- Speed matters: *Many samples*
- Samples expensive: *Simulate each circuit*
Question: Who Else Has This Problem?

Computational finance(!)

- Valuing complex financial instruments, derivatives
- High-dimensional, nonlinear, statistical integrals
- Speed+accuracy matters here, e.g., ~real-time decision-making
Big Idea: Quasi Monte Carlo (QMC)

- Classical Monte Carlo
  - Uniform pseudo-random pts
  - Surprise: *not* very uniform

- Error for n samples
  \[ O(1 / \sqrt{n}) \]

- Quasi Monte Carlo
  - *Deterministic* samples
  - “Low-discrepancy” pts

- Error for n samples
  \[ O(1 / n) \]
Computational Finance Example

- Eval 5-year discount price for a bond
  - From [Ninomiya, Tezuka, App Math Finance 1996]

![Error (log scale)]

- Ideal ~ $1/\sqrt{n}$
- Monte Carlo
- Ideal ~ $1/n$
- QMC

1439 dimensions

150x faster
Does QMC Work for Circuits? (Yes!)

- See speedups from 2X to 50X
- ...but requires some subtlety to map to QMC
- See: [Singhee, Rutenbar, ISQED 2007]

403 dimensions
Full 64b SRAM col

QMC

Pr(write < t_w) = 0.9

Pseudo Random
Very Promising Speedups

- Same 403-dimensional, 64b SRAM column

\[ n^{-0.3912} \]

\[ n^{-0.6622} \]

~9x faster for 1% error

[Singhee, Rutenbar, ISQED 2007]
SRAM reliability is all about **far tails** of stats

- **Why?** High replication (~$10^8$ bits) of core circuits
- $3\sigma$ doesn’t cut it for 100M cells; need $6\sigma$, $7\sigma$, $8\sigma$...

**Problem:** *Intractable* Monte Carlo runs

- 1M Monte Carlo sims predicts (unreliably) to $\sim 4.5\sigma$
What Do We Need To Solve This...?

- **Ultra fast sampling** of rare events
  - Put Monte Carlo *samples* out in far tails -- directly

- **Accurate analytical pdf models** of rare tails
  - Using these samples, model lets us predict *farther*
Efficiently Sampling \textit{Just} the Tail

- \textbf{Note:} \textit{Generating} MC samples is cheap, \textit{Simulating} these samples is costly

- \textbf{Idea:}
  1. \textit{Generate} regular MC samples…
  2. …but \textit{block} points that are “very probably” \textit{not} in tail
  3. \textit{Simulate} the rest – i.e., the points we do not block

\begin{itemize}
  \item \textbf{Can build this classifier filter very efficiently}
\end{itemize}
We Call the Idea: **Statistical Blockade**

- **Circuit pdf**
- **Tail**
- **Sample points**

1. **Simulate starting set** (few points, *fast*)
2. **Build classifier** (*fast*) (uses ideas from data-mining)
3. **Generate MC samples** (*fast*)
4. **Classify** sample points (*fast*)
5. **Block** nontail points (*fast*)
6. **Simulate** the rest (*slow*)

[Singhee, Rutenbar DATE 2007]
Modeling Statistics of Rare Events...

- **Extreme Value Theory (EVT)**
  - Behavior of extreme (rare) values of distributions
  - (If hurricanes are i.i.d random variables, we’d like to know the statistics of the *largest* waves...)}
EVT: Modeling the PDF in the Tail:

- Recall Central Limit Theorem: \( \sum \) (i.i.d. samples) \( \rightarrow \) Gaussian
  - Question: Is there a similar result for these tails of “extreme” results ...?  
  - Answer: YES – Extreme Value Theory (EVT)

On each of \( N \) wafers, identify cells slower than threshold \( t \). What is their distrib? EVT tells us!

- EVT gives simple analytical form for conditional tail distrib

\[
G_{a,k}(x) = \begin{cases} 
1 - \left( \frac{1 + \frac{kx}{a}}{\alpha} \right)^{\frac{1}{k}}, & k \neq 0 \\
1 - e^{-x/a}, & k = 0
\end{cases}
\]
Result: Complete 64b SRAM Column

- 90nm 64b SRAM column with write driver and column mux
- ~ 400 devices; model Write-time CDF
- Speedup: ~16X

 STD Monte Carlo: 100,000 sims
 Statistical Blockade: 6,314 sims

1000 sims to build classifier
100,000 points → 5314 sims
Result: Analytical EVT Model

- Recently validated novel analytical DRV model
  - Model of Data Retention Voltage, [Calhoun et al. UVa, ESSCIRC’07]
  - Validated to $6\sigma$, via billion element Monte Carlo run…
  - …but only did 41,721 SPICE sims; speedup $\sim 23,000X$
Summarizing

- Brute-force Monte Carlo *hurts* a lot …
  - For large, nonlinear circuit yield calculations
  - For rare event simulation
  - For… just about everything, actually

- We can do *better* with smart Monte Carlo
  - From computational finance
  - From insurance risk
  - From other apps which involve $$$ + probability

- Early CMU results: QMC, Statistical Blockade
  - On real circuits, speedups of 10x – 10,000x
Thank You!

Acknowledgements

- My CMU student, Amith Singhee, whose PhD is the basis of all the results shown in this talk

- Prof. Benton Calhoun and Jiajing Wang of U Virginia, for sharing their statistical DRV model

- Funding from Semiconductor Research Corporation

- Funding from the Focus Center for Circuit & System Solutions (C2S2), one of five such focus centers managed by the Focus Center Research Program, an SRC program.