Electrical & Computer ENGINEERING

5th Int'l System-on-Chip (SoC) Conference

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

Rob A. Rutenbar (and Amith Singhee)

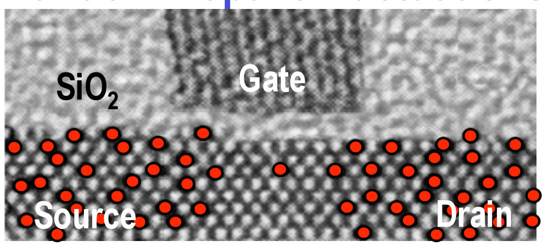
Professor, Electrical & Computer Engineering

rutenbar@ece.cmu.edu

Carnegie Mellon

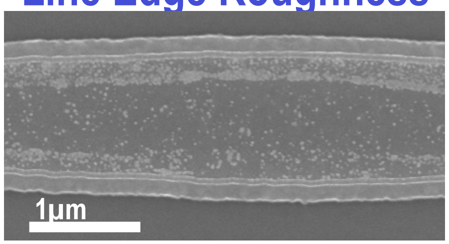
The Problem: Statistical Variation

Random Dopant Fluctuations



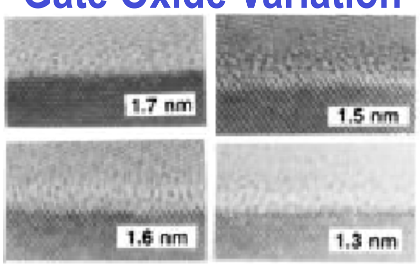
A. Brown *et al.*, *IEEE Trans. Nanotechnology*, p. 195, 2002

Line Edge Roughness

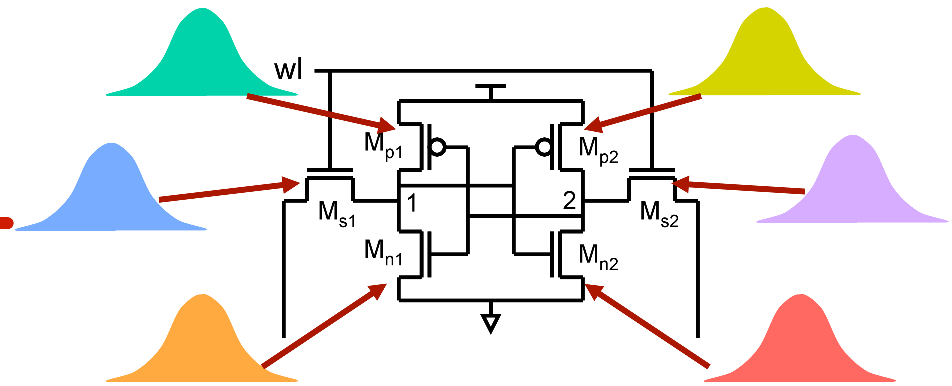


K. Shepard, U. Columbia

Gate Oxide Variation

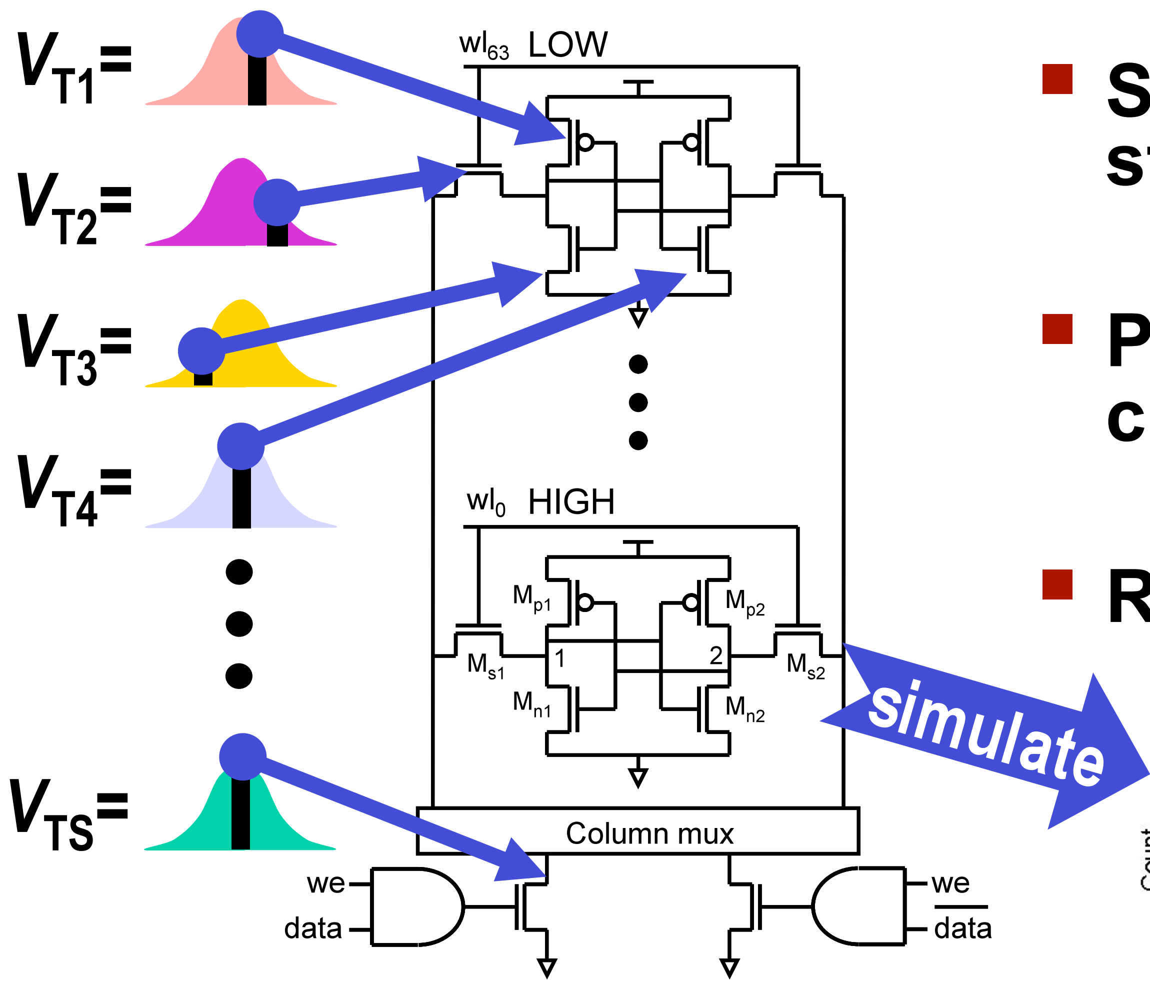


Momose et al, *IEEE Trans. Electron Devices*, 45(3), 1998

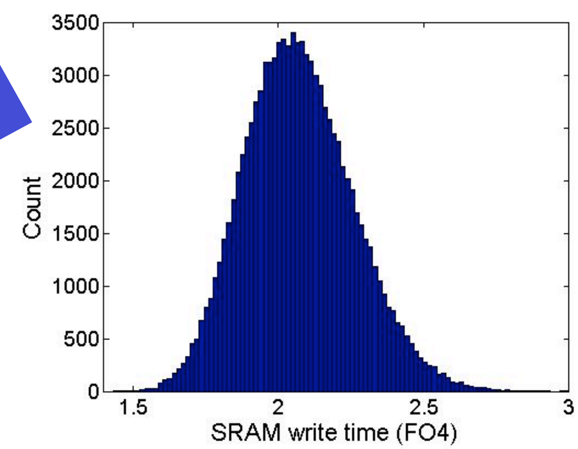


- **At nanoscale, nothing is *deterministic* anymore**
- **How to evaluate designs?**

To Evaluate Circuit Impact: *Monte Carlo*

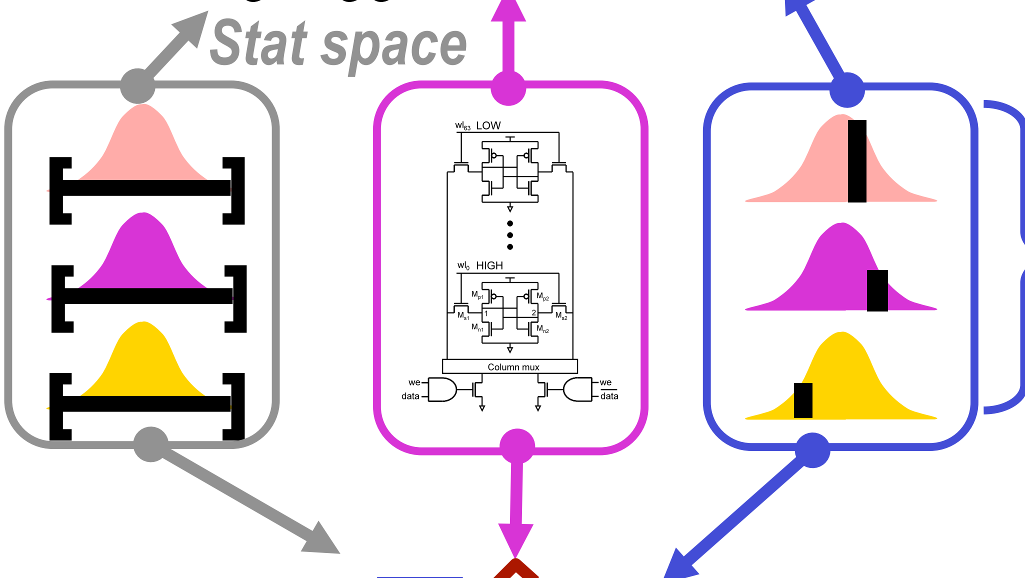


- Sample each statistical variable
- Parameterize one circuit, simulate it
- Repeat--n samples



Monte Carlo Math: Just A Big Integral

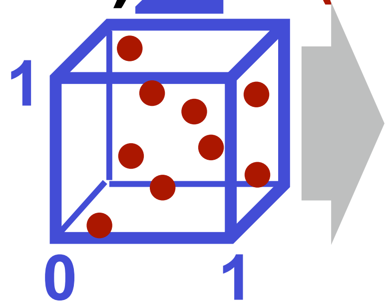
$$\text{Est} = \int \cdots \iint f_{ckt}(\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$$



s statistical vars

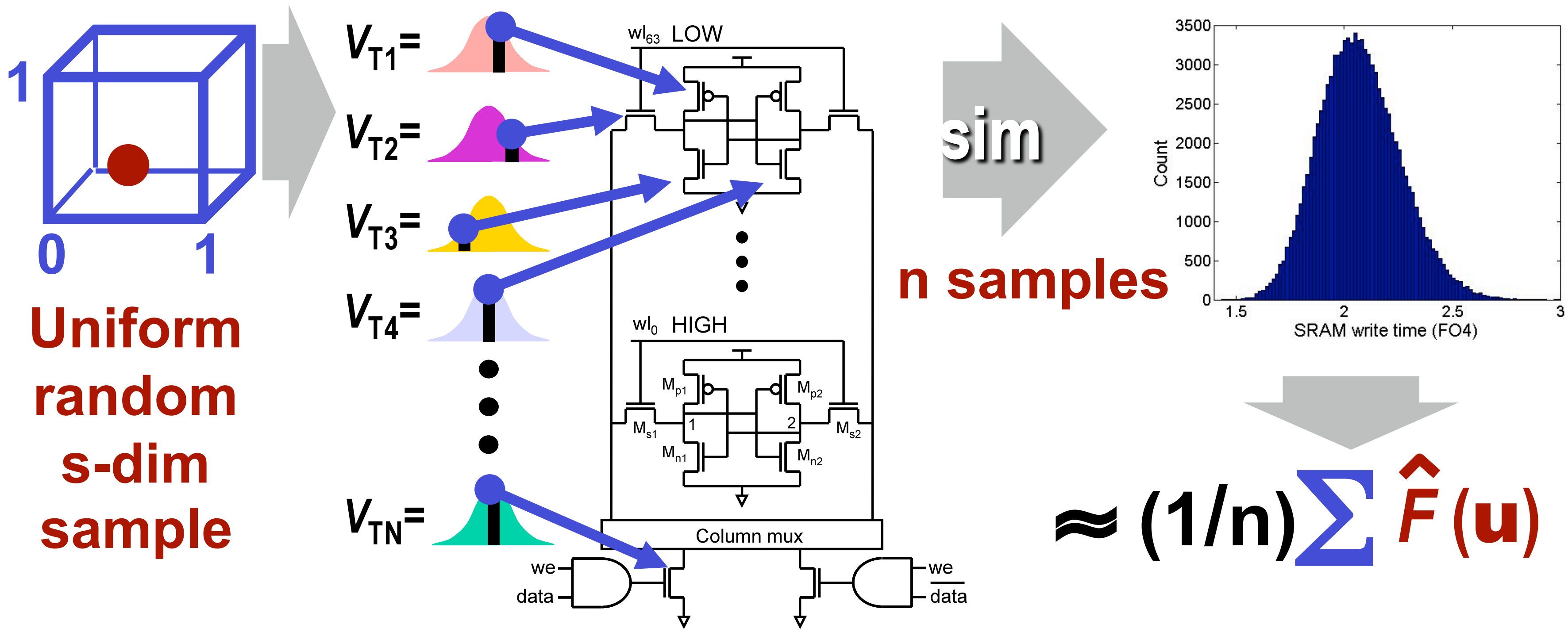
s-dimensional prob

$$\approx (1/n) \sum \hat{F}(\mathbf{u})$$



Can transform to sample uniformly from s-dim unit cube

Evaluate Circuit Impact: Monte Carlo



- **PRO:** Accurate, flexible, general
- **CON:** Slow, slow, s l o w…

Why is Monte Carlo Painful?

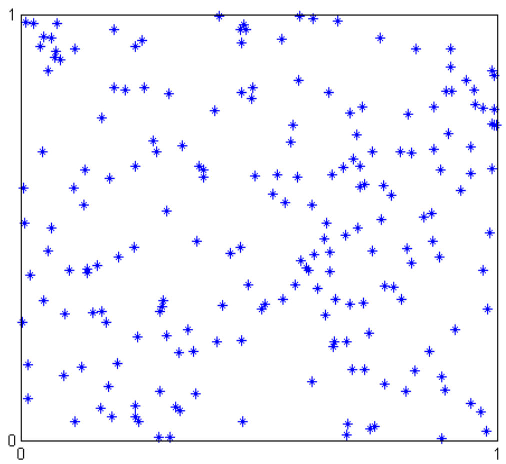
- **High-dim problems:** ***s*** ***is big (100-1000)***
- **Profoundly nonlinear:** ***Nanoscale physics***
- **Accuracy matters:** ***~1-5% error***
- **Speed matters:** ***Many samples***
- **Samples expensive:** ***Simulate each circuit***

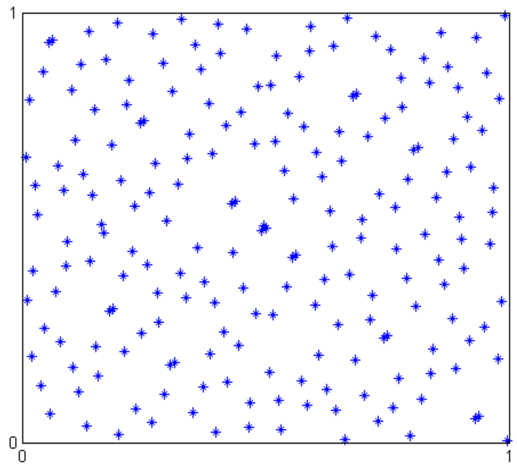
Question: Who *Else* Has This Problem?

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**
- **Quasi Monte Carlo**
  - ***Deterministic* samples**
  - **"Low-discrepancy" pts**

- **Error for n samples**

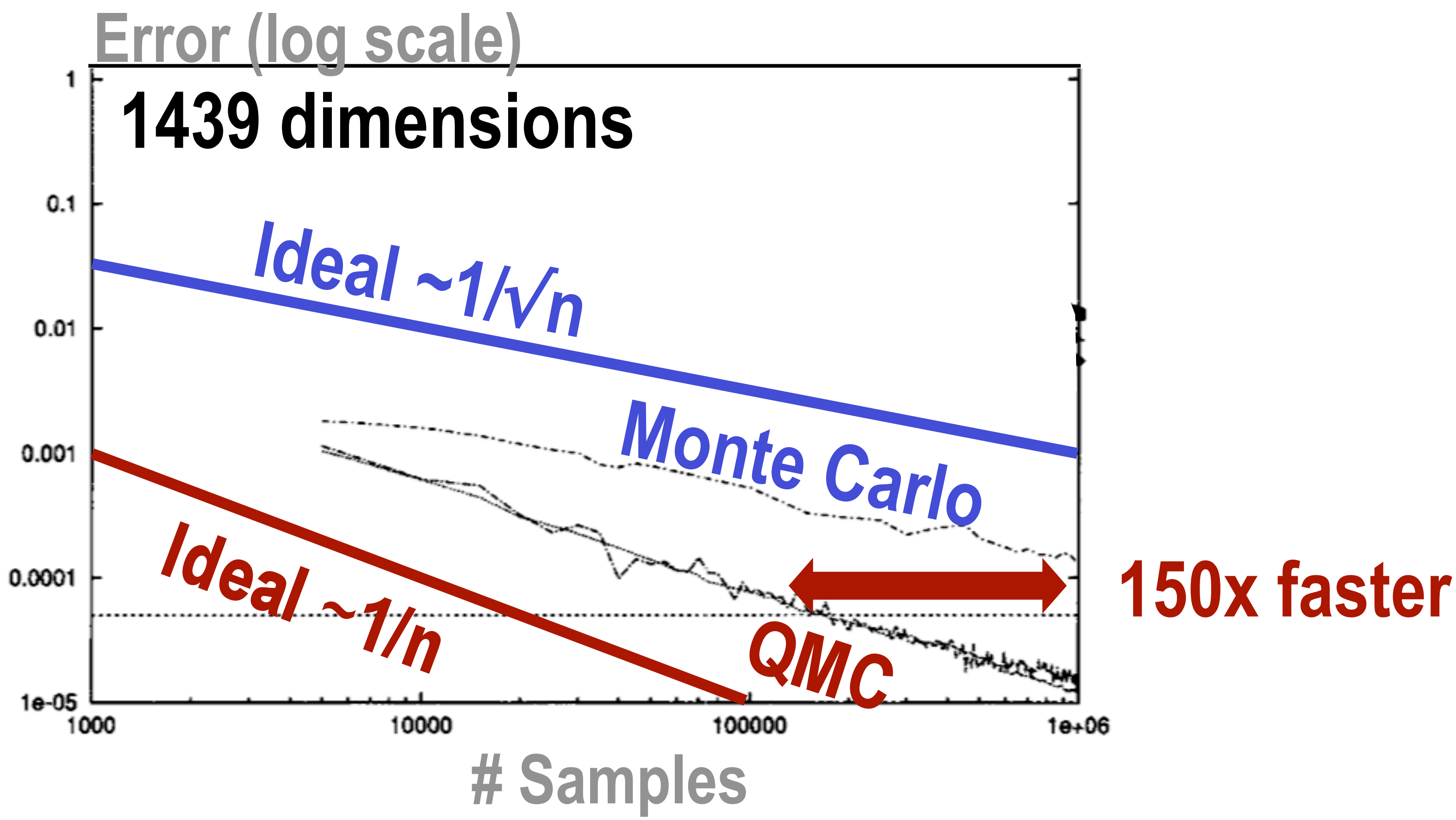
  $O(1 / \sqrt{n})$

- **Error for n samples**

  $O(1 / n)$

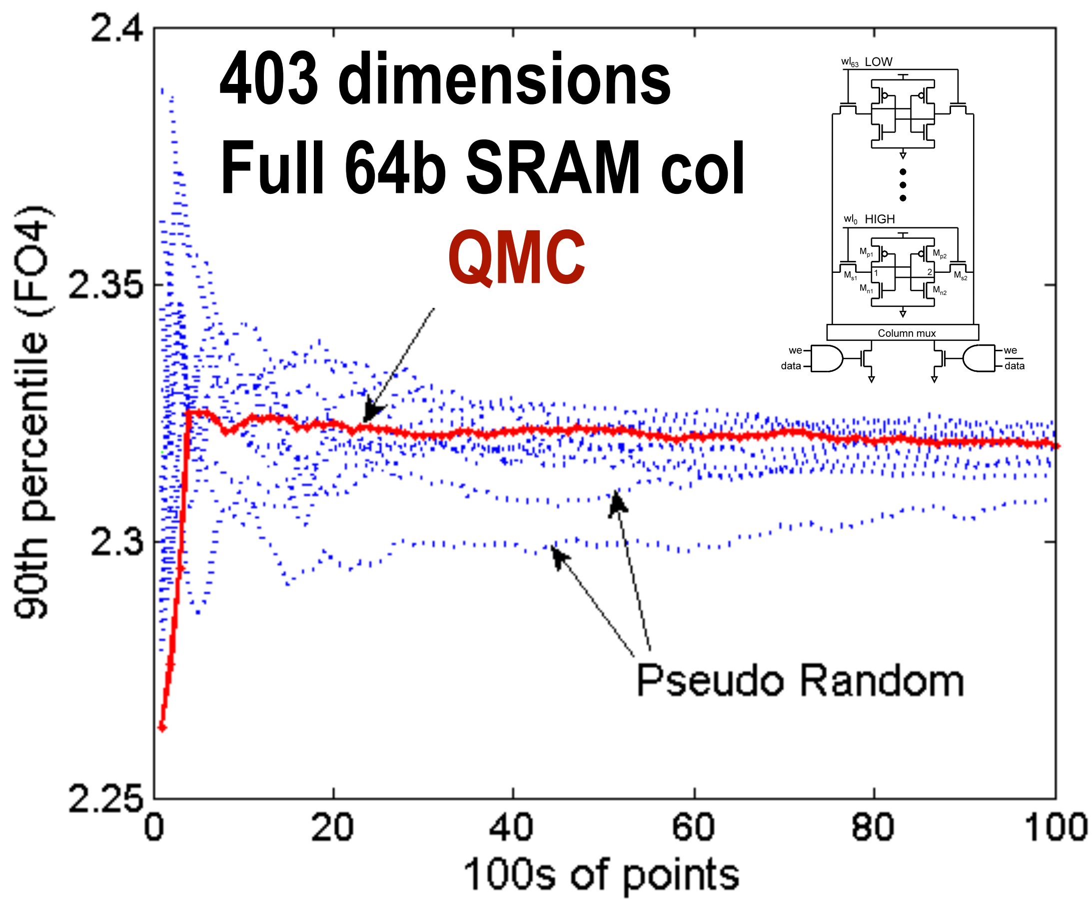
Computational Finance Example

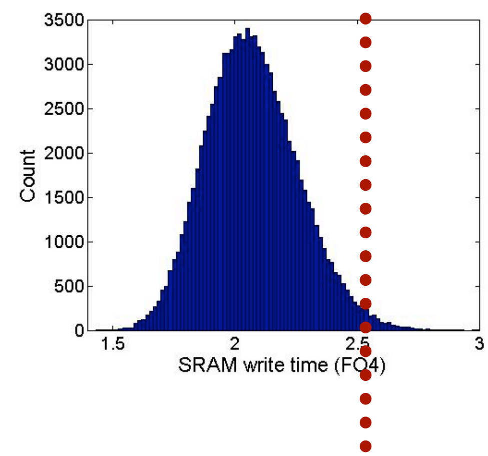
- **Eval 5-year discount price for a bond**
  - From [Ninomiya,Tezuka, App Math Finance 1996]



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]

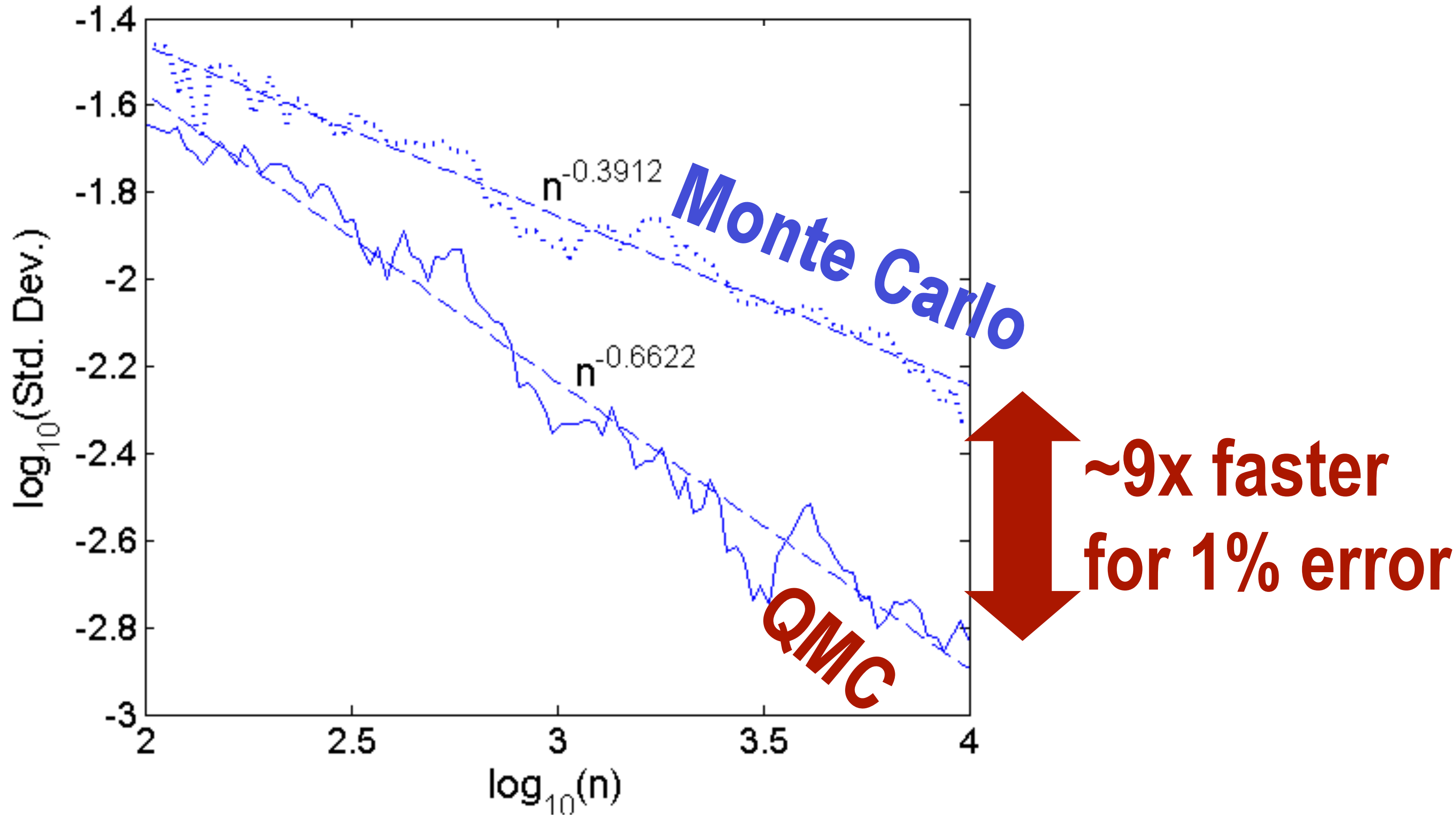




$Pr(write < t_W) = 0.9$

Very Promising Speedups

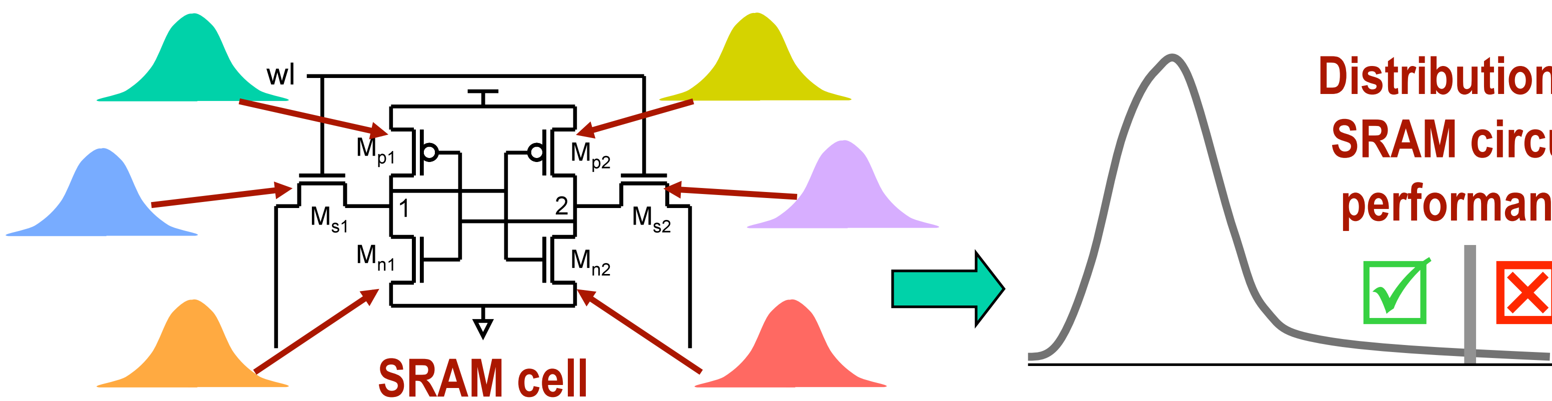
- Same 403-dimensional, 64b SRAM column



[Singhee, Rutenbar, ISQED 2007]

Next Problem: "Rare Event" Statistics

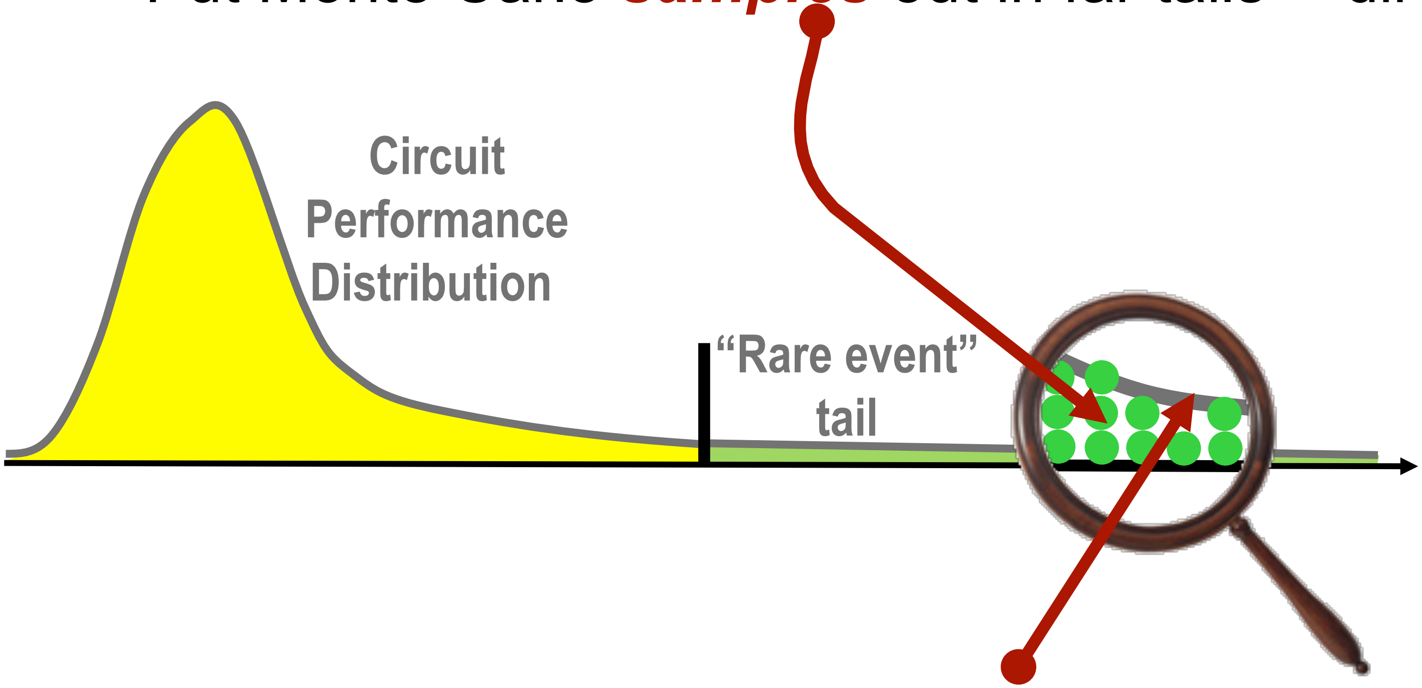
- **SRAM reliability is all about *far tails* of stats**
  - **Why?** High replication (~10^8 bits) of core circuits
  - 3σ doesn't cut it for 100M cells; need **6σ, 7σ, 8σ**…



- **Problem: *Intractable* Monte Carlo runs**
  - 1M Monte Carlo sims predicts (unreliably) to **~4.5σ**

What Do We Need To Solve This…?

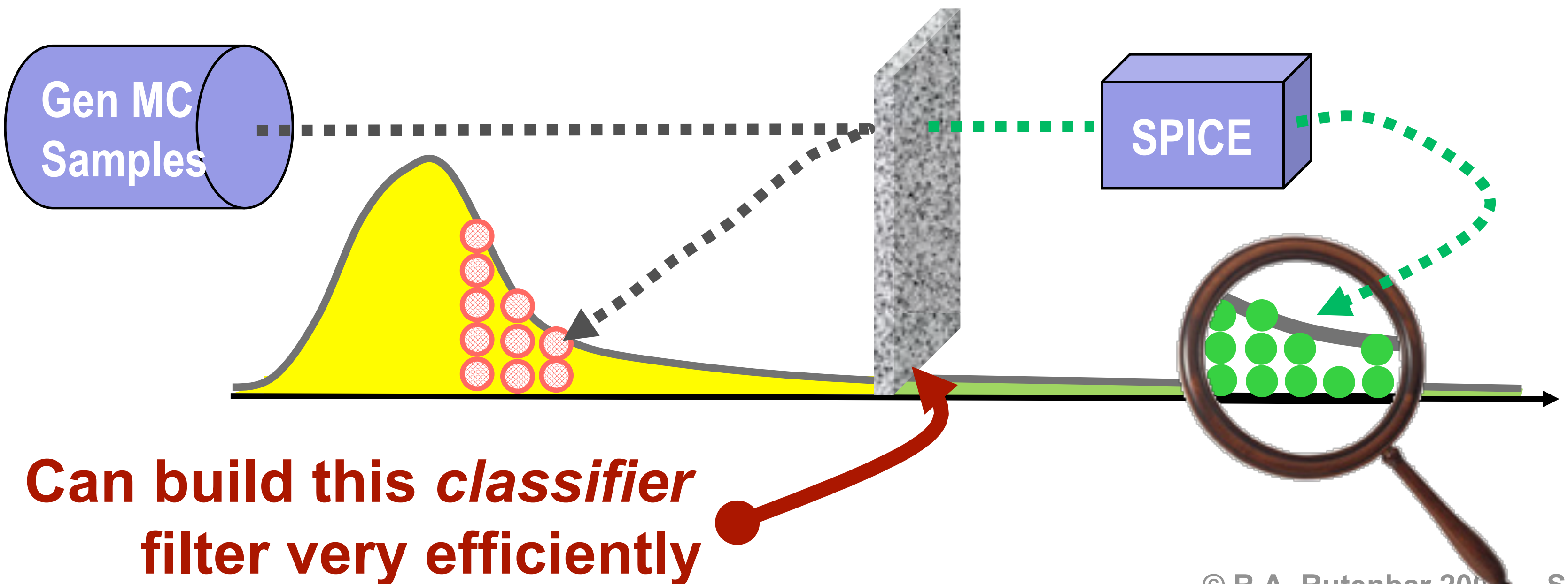
- **Ultra fast *sampling* of rare events**
  - Put Monte Carlo ***samples*** out in far tails -- directly



- **Accurate analytical *pdf models* of rare tails**
  - Using these samples, model lets us predict ***farther***

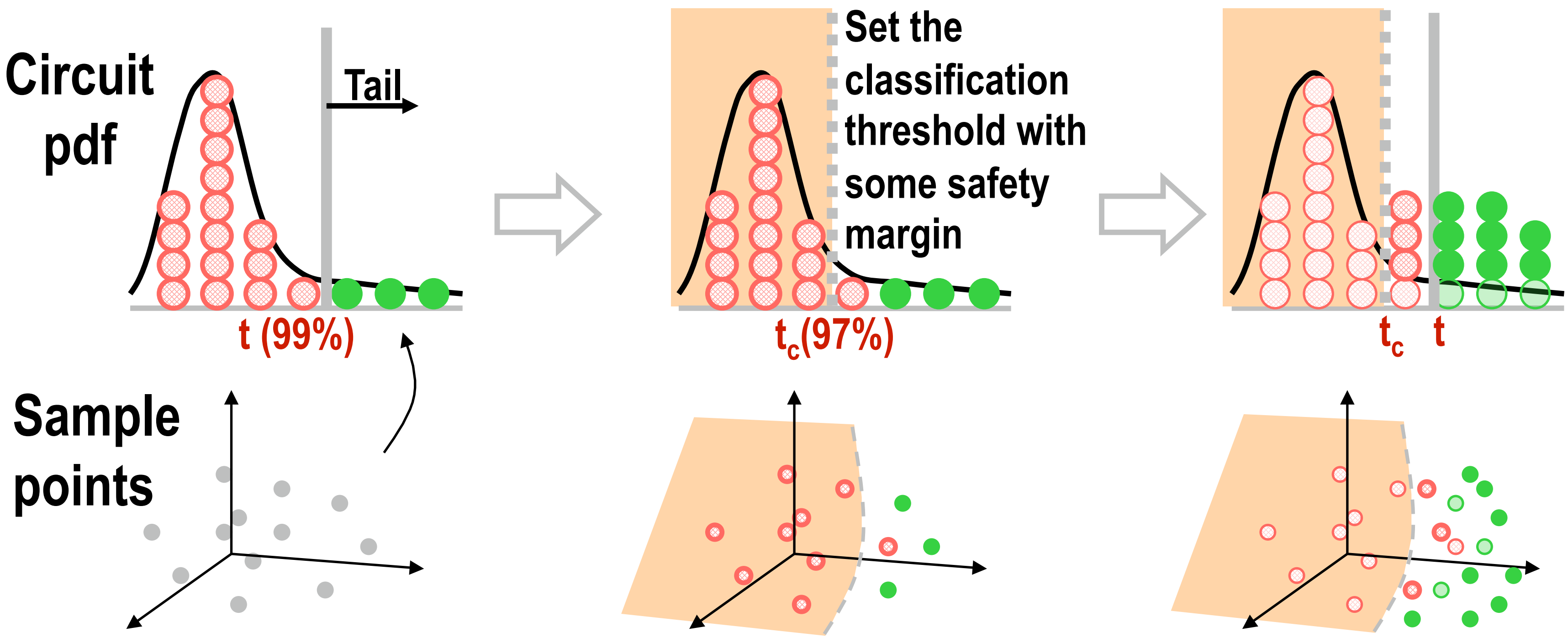
Efficiently Sampling *Just* the Tail

- **Note:** ***Generating*** **MC samples is cheap,**
  Simulating **these samples is costly**
- **Idea:**
  1. **Generate** regular MC samples…
  2. …but ***block*** points that are "very probably" ***not*** in tail
  3. **Simulate** the rest – i.e., the points we do not block



Can build this *classifier* filter very efficiently

We Call the Idea: *Statistical Blockade*



Simulate starting set (few points, *fast*)

Build *classifier* (*fast*) (uses ideas from data-mining)

Generate MC samples (*fast*)
Classify sample points (*fast*)
Block nontail points (*fast*)
Simulate the rest (*slow*)

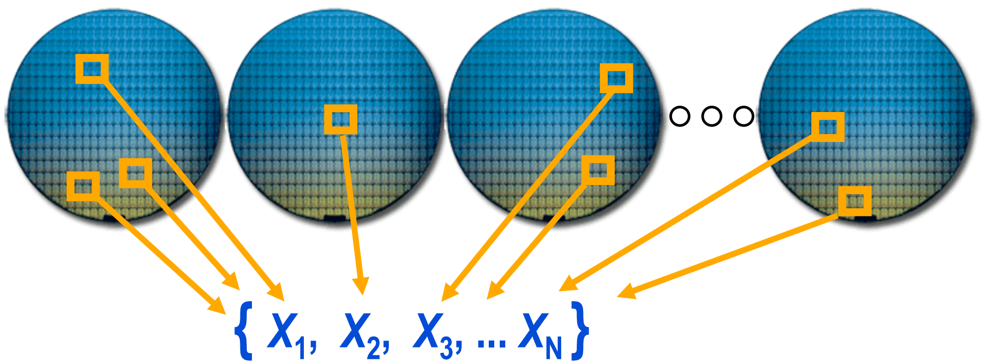
[Singhee, Rutenbar DATE 2007]

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*…)**

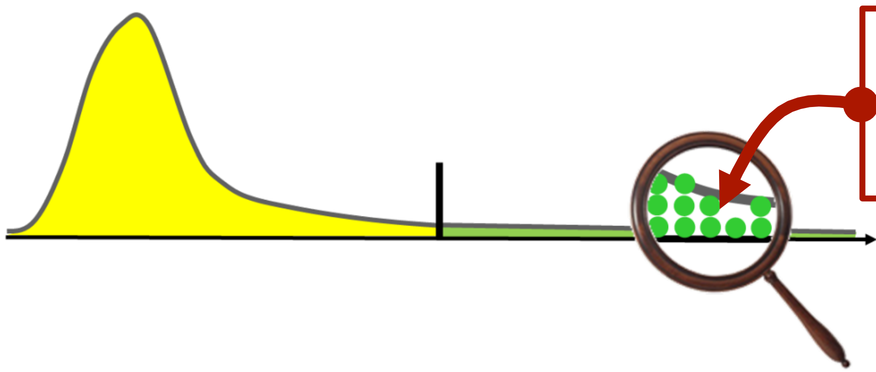
EVT: Modeling the PDF in the Tail:

- **Recall Central Limit Theorem: Σ(i.i.d. samples) → Gaussian**
  - **Question**: Is there a similar result for these tails of "extreme" results ...?
  - **Answer**: YES – **Extreme Value Theory** (EVT)



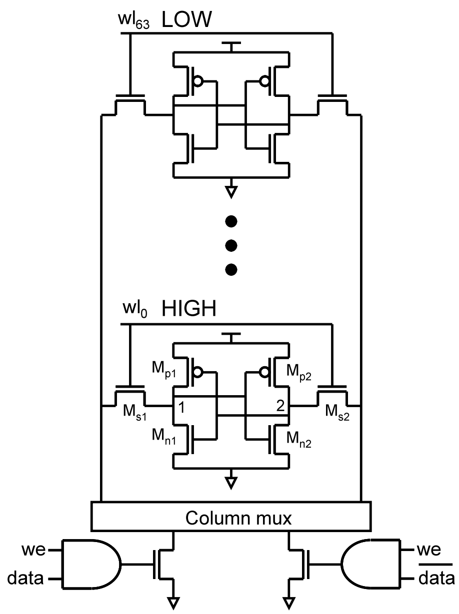
On each of *N* wafers, identify cells slower than threshold t. What is their distrib? EVT tells us!

- **EVT gives simple analytical form for *conditional tail distrib***

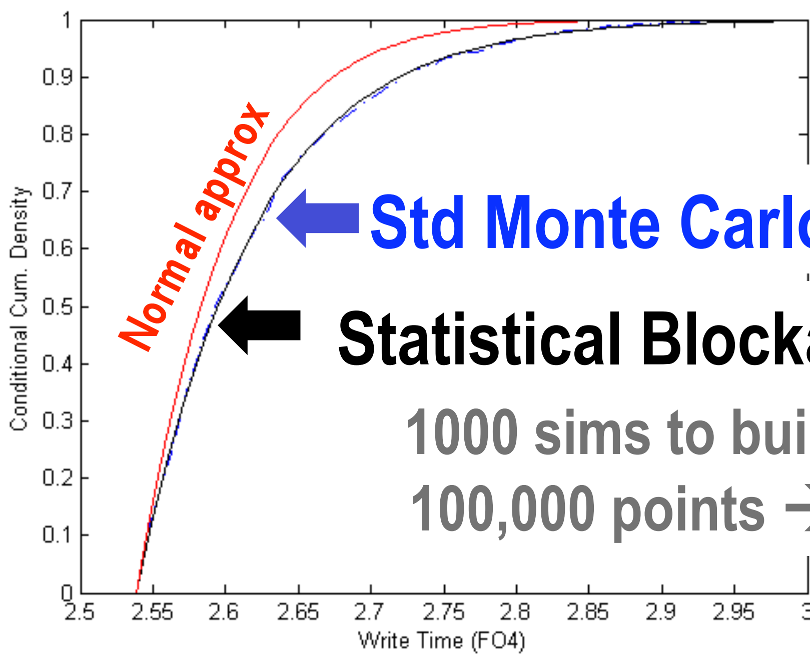


$$G_{a,k}(x) = \begin{cases} 1 - \left(1 + \frac{kx}{a}\right)^{1/k} & , k \neq 0 \\ 1 - e^{-x/a} & , k = 0 \end{cases}$$

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

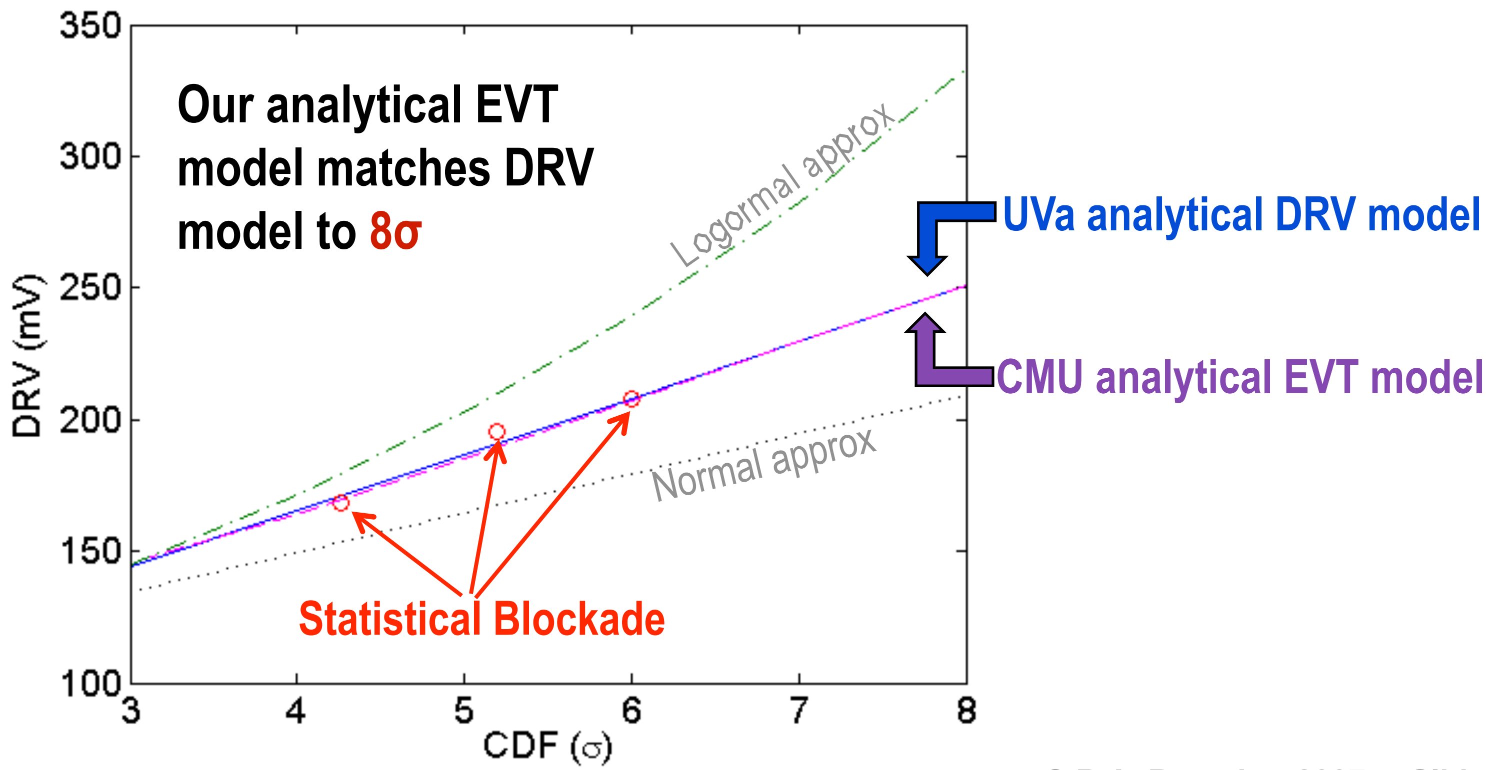
Statistical Blockade: 6,314 sims

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 **6σ**, via **billion element** Monte Carlo run…
  - …but only did 41,721 SPICE sims; speedup **~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.