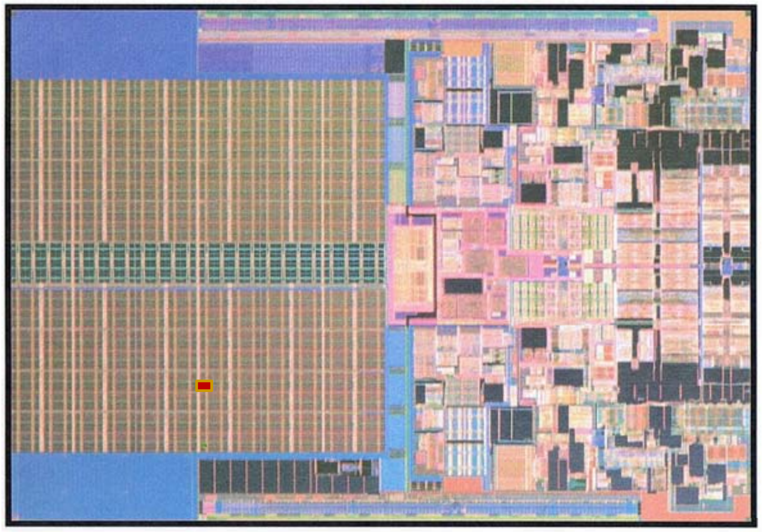
Breaking the Simulation Barrier: SRAM Evaluation Through Norm Minimization

Lara Dolecek
Masood Qazi, Devavrat Shah, Anantha Chandrakasan
Massachusetts Institute of Technology
dolecek@mit.edu

Towards high-performance SRAMs



Penryn© microprocessor by Intel
[George et al., ASSCC, 2007.]

- Significant portion of chip area is dedicated to SRAM components.
  - Need to increase the density while maintaining reliability:

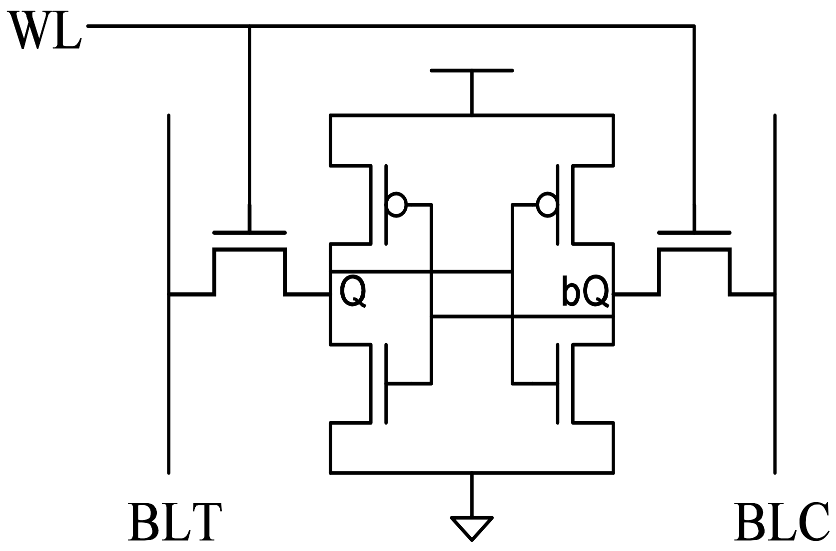
Probability of failure per cell below 10^{-6}

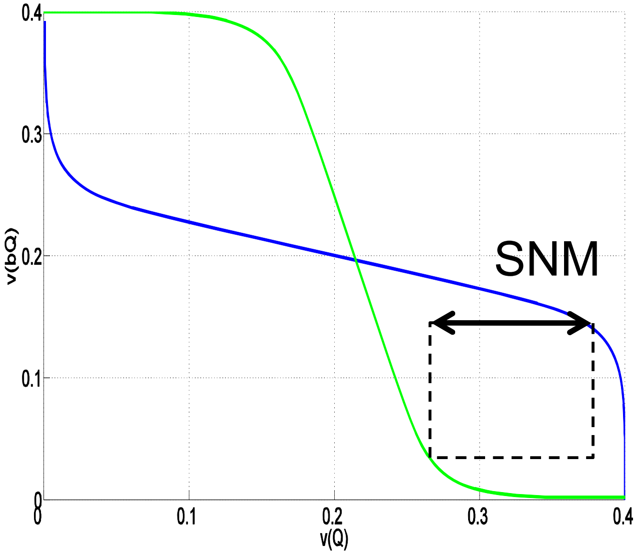
(with redundancy)

Outline

- Motivation
  - 6T SRAM cell stability
- Background
  - Monte Carlo and statistical simulation
- Norm-Minimization Importance Sampling Algorithm
- Experimental Results
  - Data retention voltage
  - Read write trade-off
- Concluding Remarks

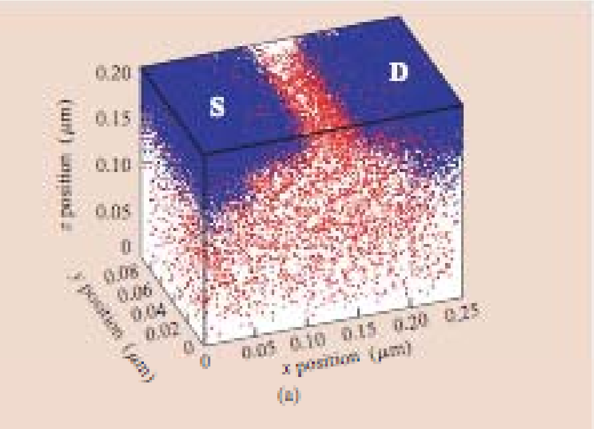
Motivation: Stability of 6T SRAM cell



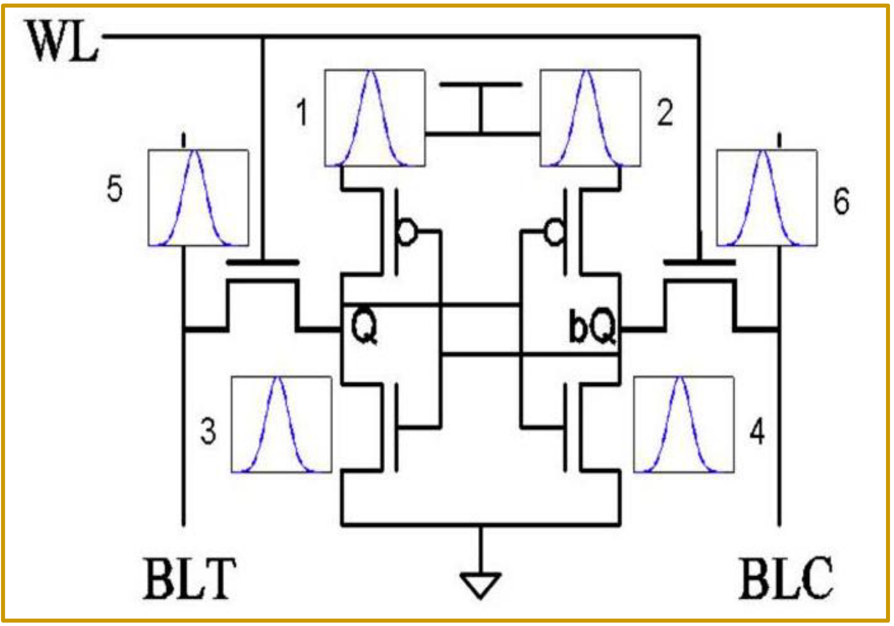


- Stability measured by the SNM gap in the butterfly curve.

Motivation: Stability of 6T SRAM cell



Reproduced from [Bernstein et al. IBM J. Res. And Dev., 2006.]



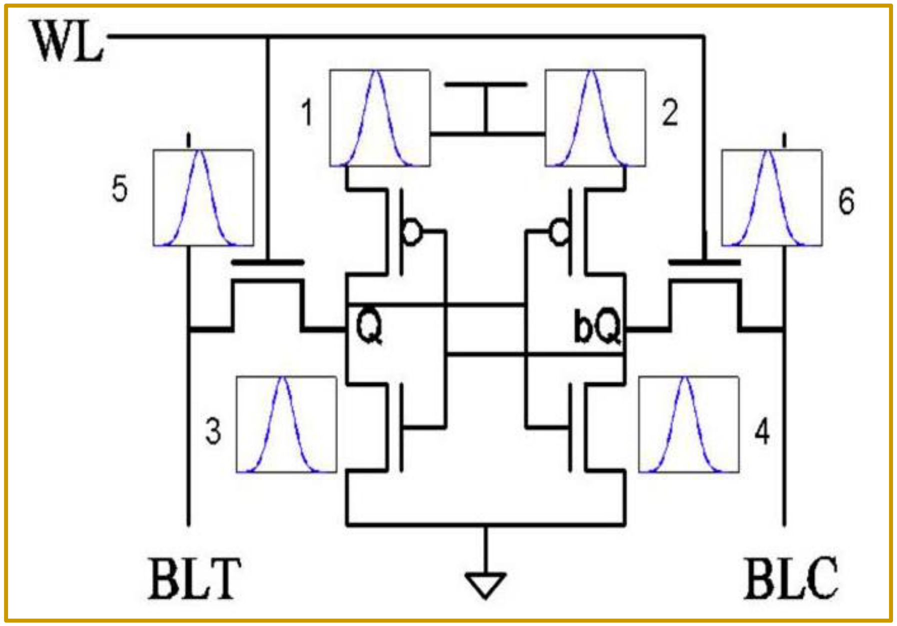
- Random dopant fluctuations

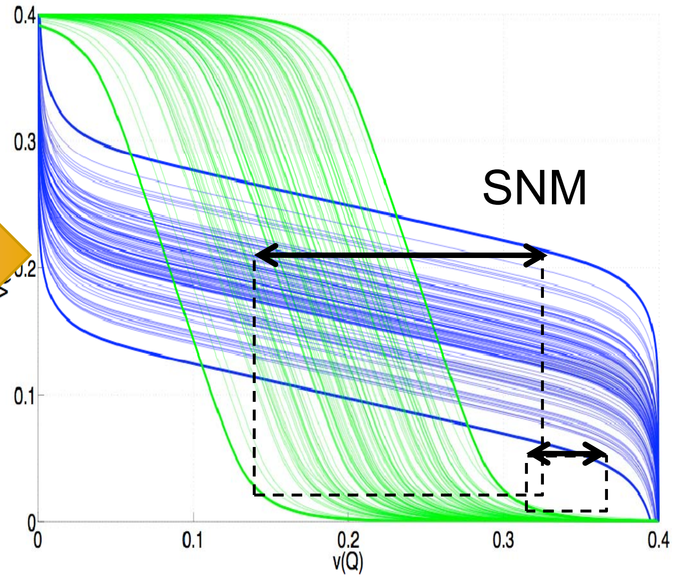
- Each threshold voltage variation is modeled by a Gaussian random variable.

  [Mukhopadhyay et al. Trans. on CAD, 2005]

  - Deviation inversely prop. to √area.

Motivation: Stability of 6T SRAM cell





- Each threshold voltage variation is modeled by a Gaussian random variable.

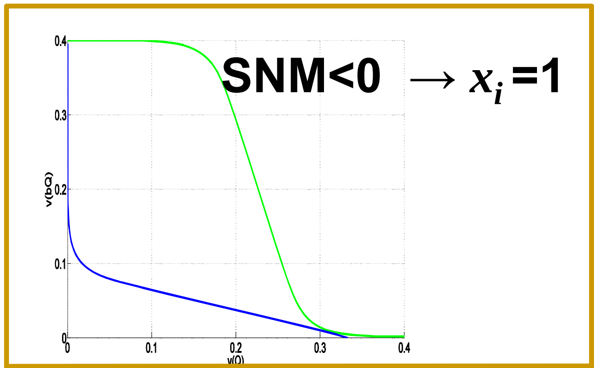
  [Mukhopadhyay et al. Trans. on CAD, 2005]

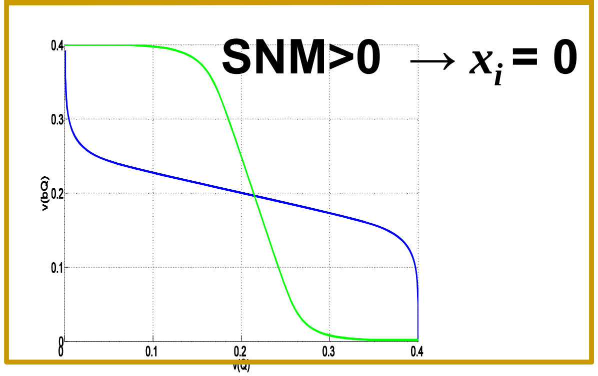
  - Deviation inversely prop. to √area.

- Cell stability decreases.

Background: A First Cut

- Monte Carlo Simulation:
  - Run the experiment for N trials.
  - On each trial *i*, $x_i = \begin{cases} 1 & \text{failure,} \\ 0 & \text{no failure} \end{cases}$
  - Estimate p_{fail} as $p_{MC} = \frac{1}{N}\sum_{i=1}^{N} x_i$


SNM<0 $\rightarrow x_i = 1$


SNM>0 $\rightarrow x_i = 0$

- For 90% accuracy and confidence:
  - $N \sim 100/p_{fail}$.
  - Example: for $p_{fail} \sim 10^{-6}$, $N \sim 10^{8}$.

Previous and Related Work

Statistical Approaches

- **Importance Sampling (IS)**

  [Kanj et al., DAC, 2006, Chen et al., ICCAD, 2007]

  - Sampling from a “tilted” distribution to reduce the variance of the estimator.
  - Characterization of a discrete set of values, e.g. failure due to SNM < 0.
- **Extreme Value Theory (EVT)**

  [Singhee and Rutenbar, DATE, 2007, VLSI, 2008]

  - Analytical description of the highest order statistic.
  - Characterizing a continuum of values, e.g. failure as a function of delay.

Previous and Related Work

Statistical Approaches

- **Importance Sampling (IS)**

  [Kanj et al., DAC, 2006, Chen et al., ICCAD, 2007]

  - Sampling from a “tilted” distribution to reduce the variance of the estimator.
  - Characterization of a discrete set of values, e.g. failure due to SNM < 0.

- **Extreme Value Theory (EVT)**

  [Singhee and Rutenbar, DATE, 2007, VLSI, 2008]

  - Analytical description of the highest order statistic.
  - Characterizing a continuum of values, e.g. failure as a function of delay.

Concept of Importance Sampling

A. Sample from the tilted distribution N times.

- On each trial *i*, $\tilde{x}_i = \begin{cases} 1 & \text{failure} \\ 0 & \text{no failure} \end{cases}$

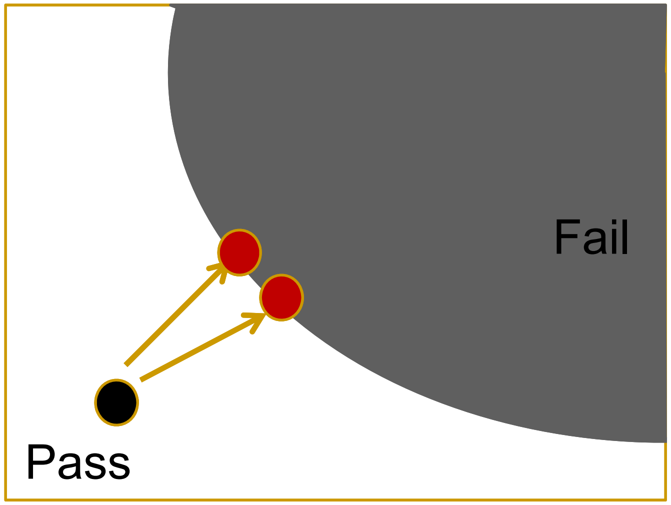
B. Estimate p_{fail} as $\boxed{p_{IS} = \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i \cdot w_i}$

- w_i is the ratio of original and tilted distributions.

◆ How to select tilting (and w_i) ?

Towards the Typical Rare Event

- Theory of Large Deviations [e.g. Bucklew, Springer, 2004]:
  - “When a rare event happens, it happens in a typical way”.



- “Tilt” the distribution towards this event.

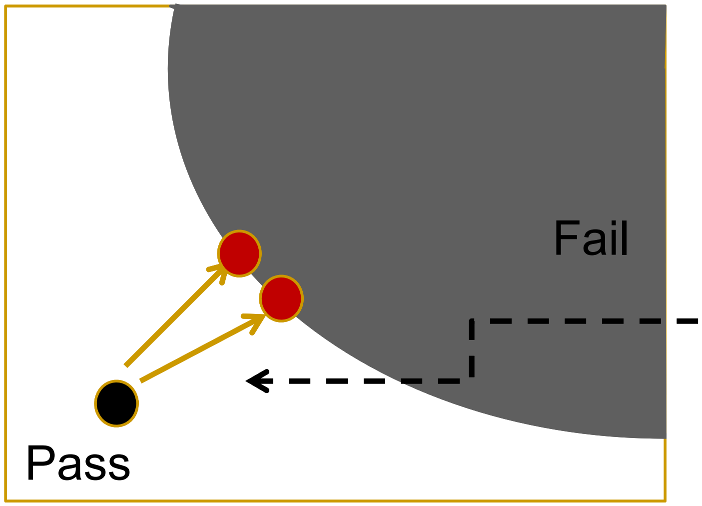
Towards the Typical Rare Event

- Underlying Gaussian r.v.s $\sim\mathcal{N}(\mu_k,\sigma_k^2)$

Tilting= shifting the mean

- Mean shift vector $\underline{s}=(s_1,\ldots,s_k,\ldots,s_M)$:

$$\mathcal{N}(\mu_k,\sigma_k^2)\rightarrow\mathcal{N}(\mu_k+s_k,\sigma_k^2)$$



Quadratic distance for Gaussian random variables

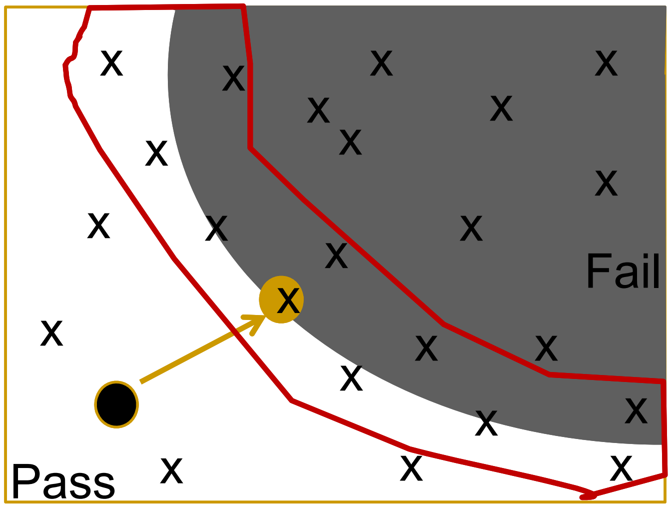
Importance Sampling Through Norm Minimization: The algorithm

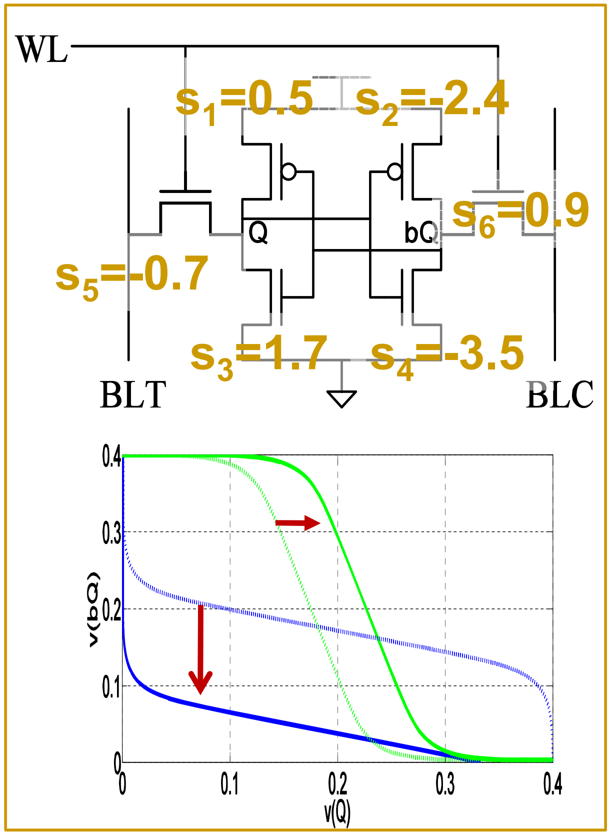
- Step 1: Exploration
  - Find the mean shift vector $\underline{s}$.

- Step 2: Exploitation
  - Run Importance Sampling based on $\underline{s}$.

Importance Sampling Through Norm Minimization: The algorithm

- <u>Step 1: Exploration</u>
  - Find the mean shift vector <u>s</u>.

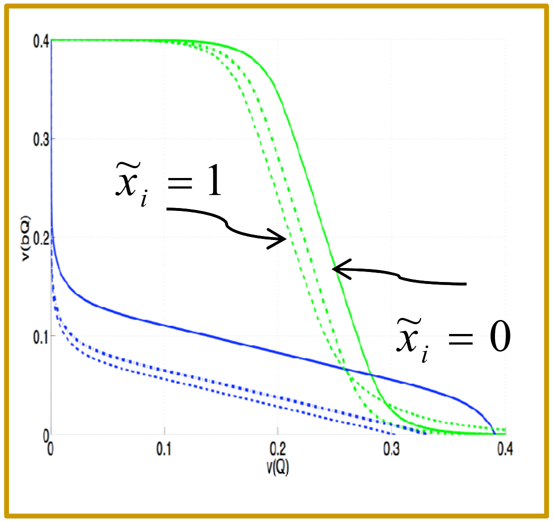




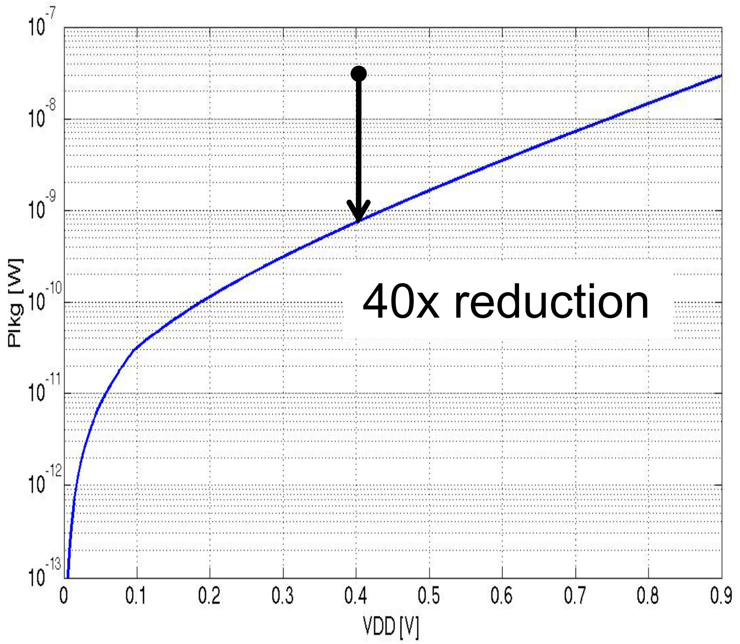
  - Uniform sampling of the space.
  - Filtering by boundary.
  - Selecting the point with lowest quadratic distance.

Importance Sampling Through Norm Minimization: The algorithm

- <u>Step 2: Exploitation</u>
  - Based on $\underline{s}=(s_1,\ldots,s_M)$ sample for N trials.
  - On each trial i:
    - Realization $\underline{y}_i=(y_{i,1},\ldots,y_{i,M})$
    - $\tilde{x}_i = \begin{cases} 1 & \text{failure} \\ 0 & \text{no failure} \end{cases}$
    - $w_i = \exp\left(-\sum_{j=1}^{M} \frac{s_j(2y_{i,j}-2\mu_j - s_j)}{2\sigma_j^2}\right)$
  - Estimate p_{fail} as $\boxed{p_{IS} = \frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i \cdot w_i}$



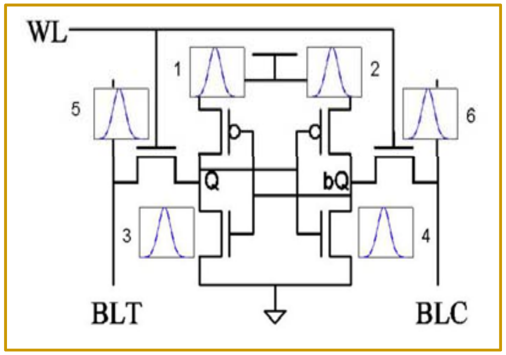
Example 1: Data Retention Voltage (DRV)



- Want to lower the supply voltage V_{dd} to reduce leakage power.
- DRV: Under **random fluctuations**, how low can V_{dd} be set to maintain cell stability ?

Example 1: Data Retention Voltage

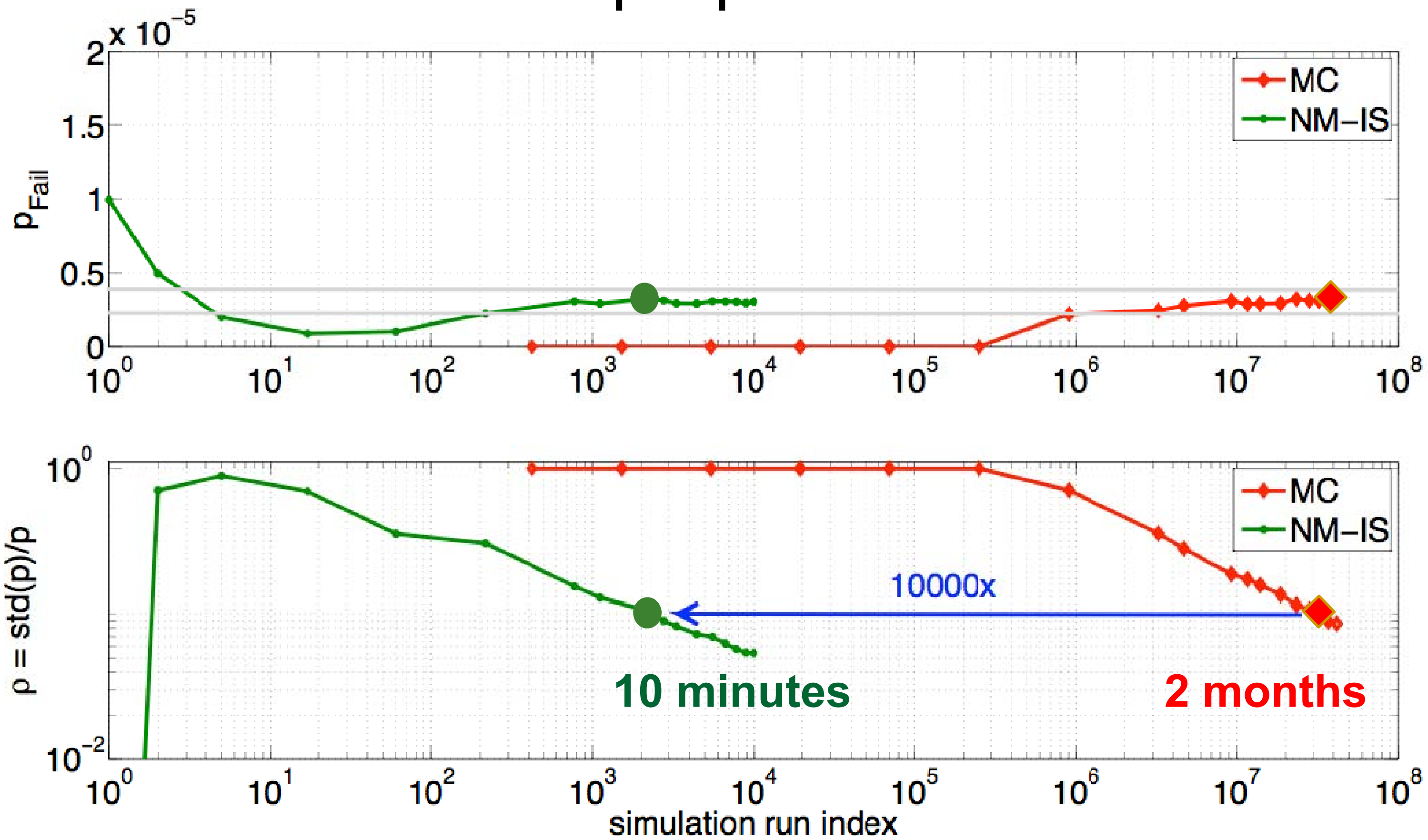
- Experimental set-up
  - Under 32nm bulk CMOS predictive technology [Zhao and Cao, ISQED 2006].
  - Variations:
    - Devices 1,2: σ = 36 mV
    - Devices 3-6: σ = 50 mV



Vdd (mV)	MC est.	MC σ-equiv.	MC no. runs	NM-IS est.	NM-IS σ-equiv.	NM-IS no. runs	speed up
275	5.4×10^{-3}	2.5491	2×10^{4}	4.9×10^{-3}	2.5828	2×10^{3}	10x
300	3.6×10^{-4}	3.3781	1.2×10^{6}	4.4×10^{-4}	3.3263	3×10^{3}	100x
400	**3.1×10^{-6}**	**4.5195**	**4×10^{7}**	**3.0×10^{-6}**	**4.5264**	**4×10^{3}**	**10000x**

Example 1: Data Retention Voltage

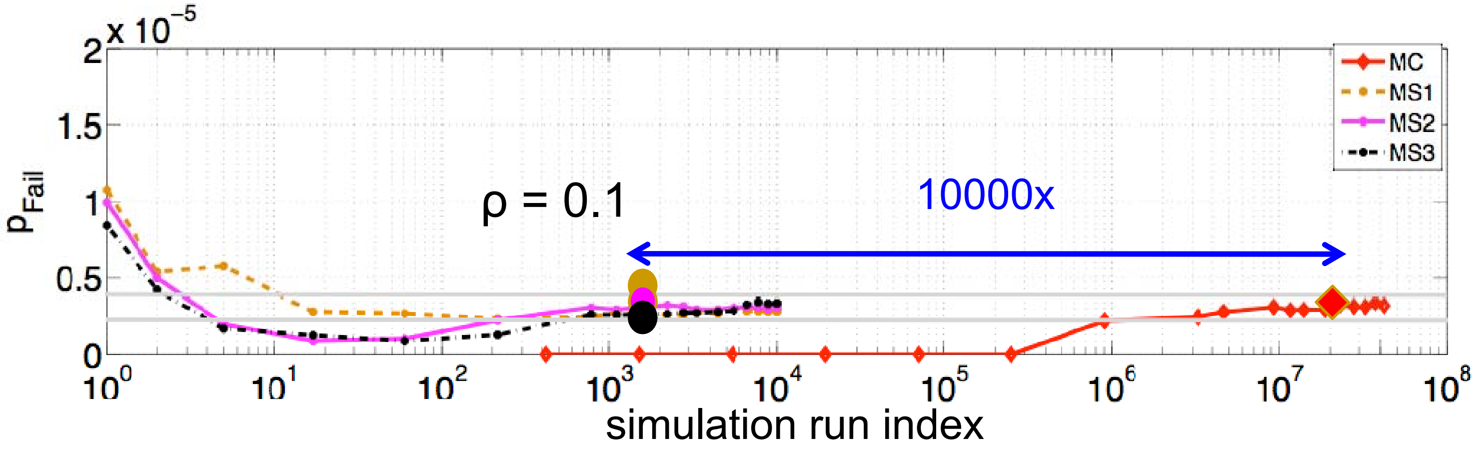
Monte Carlo vs. proposed NM-IS estimator

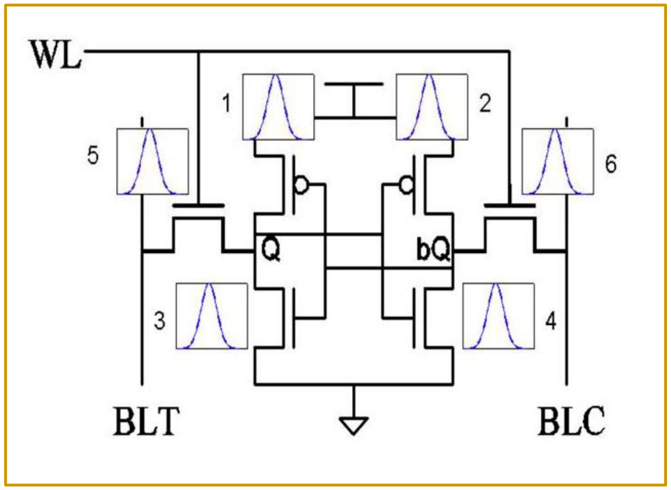


Complexity:		NM-IS trials	vs.	MC trials
Stage 1	2000	4000		40 000 000
+ Stage 2	2000			

Example 1: Data Retention Voltage

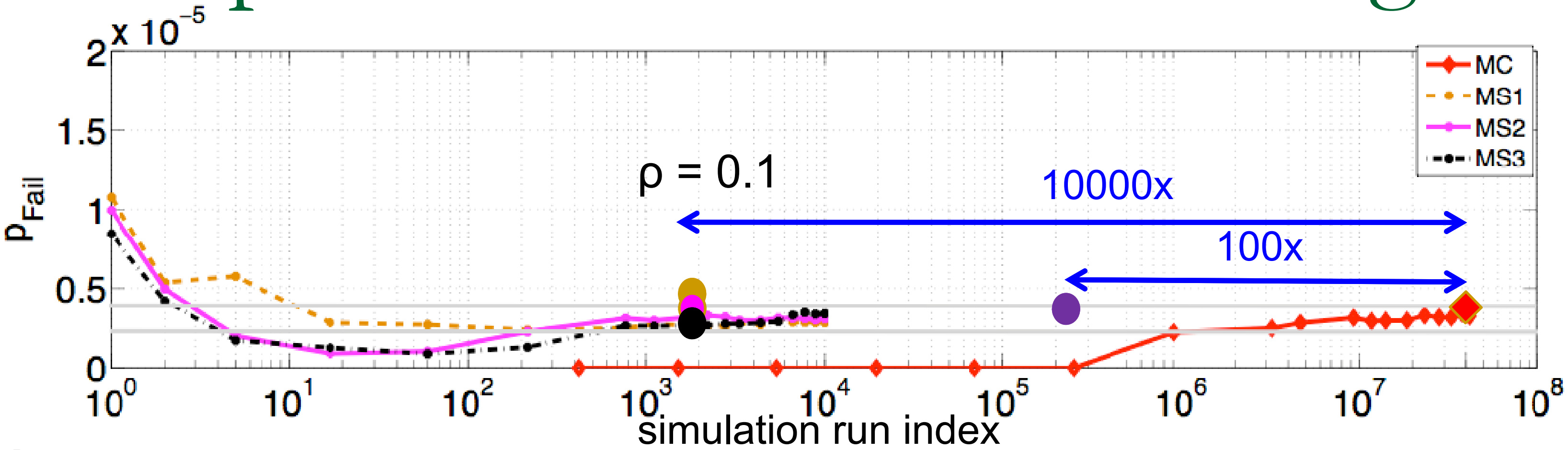
Monte Carlo vs. proposed NM-IS estimator





	VT1 shift	VT2 shift	VT3 shift	VT4 shift	VT5 shift	VT6 shift	Quad norm
MS1	0.51	-2.39	1.72	-3.54	-0.70	0.97	4.78
MS2	1.46	-2.09	0.88	-3.82	-0.78	0.73	4.79
MS3	1.17	-2.94	0.02	-3.56	-0.79	0.05	4.83

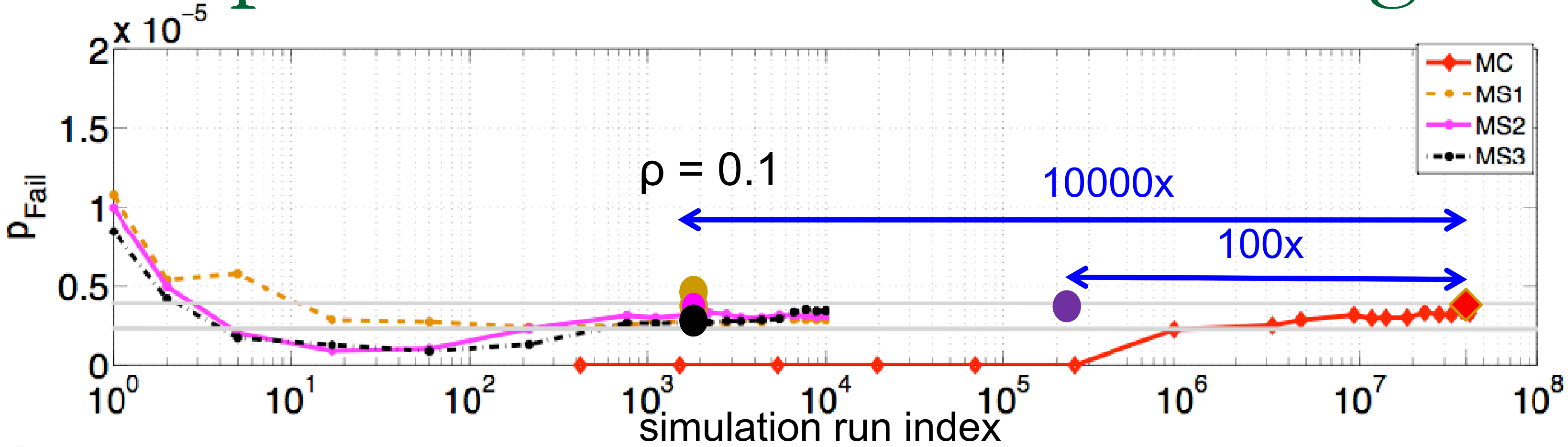
Example 1: Data Retention Voltage



Versus

1. Nominal Importance Sampling:
   - Selects mean shift based on 50% failure.

Example 1: Data Retention Voltage



Versus

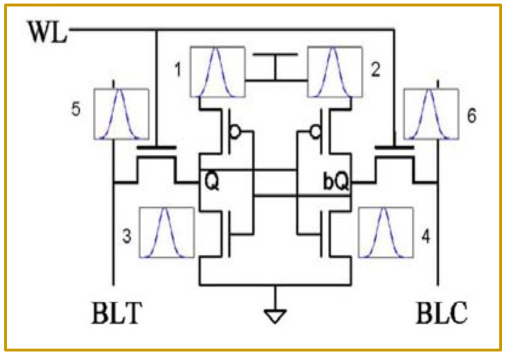
1. Nominal Importance Sampling.
2. Mixture Importance Sampling (MIS):

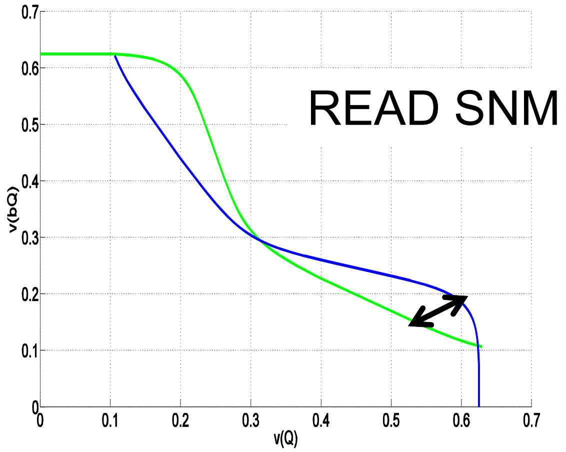
[Kanj et al., DAC 2006]

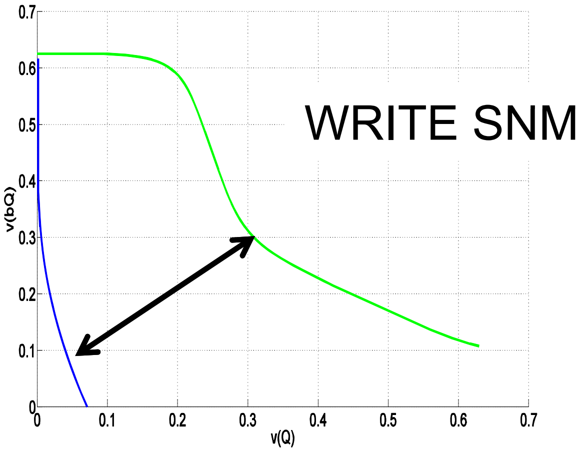
σ. equiv (≈ p_{fail})	MIS Δσ	NM-IS Δσ
2.5 (≈10^{-3})	0.05	0.03
3.4 (≈10^{-4})	0.07	0.05
4.15 (≈10^{-5})	0.07	
4.5 (≈10^{-6})		0.0069

Example 2: Read Write Trade Off

- Experimental Set-up:
  - Under 32nm bulk CMOS predictive technology [Zhao and Cao, ISQED, 2006].
  - Variations:
    - Devices 1,2: σ = 36/√2 mV
    - Devices 3,4: σ = 50/√3 mV
    - Devices 5,6: σ = 50 mV

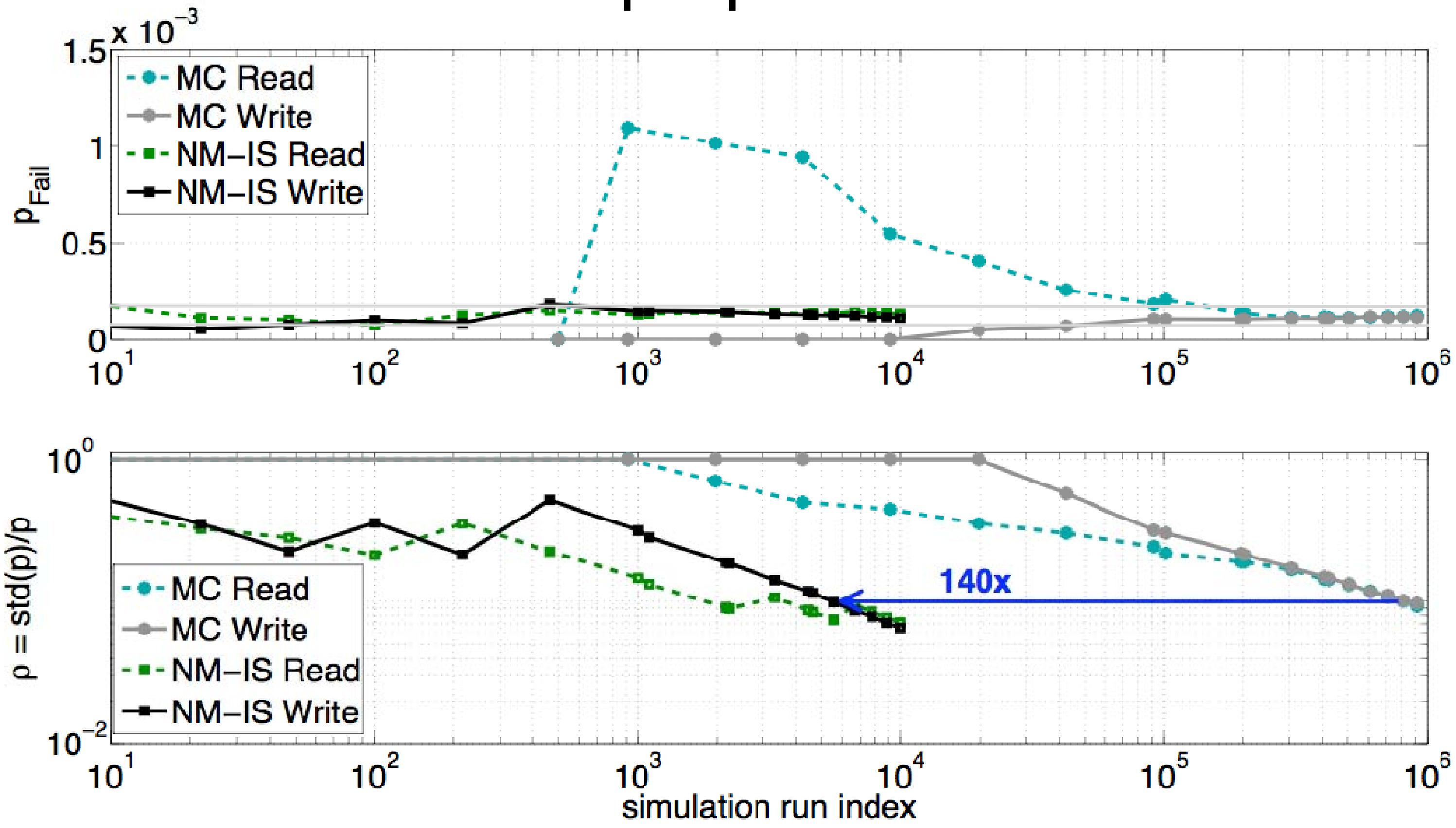






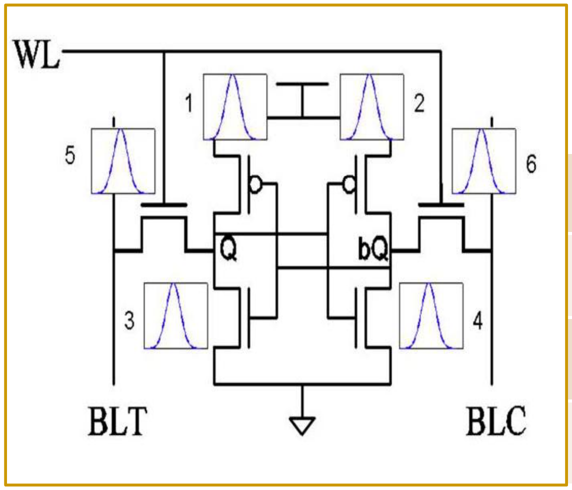
Example 2: Read Write Trade Off

Monte Carlo vs. proposed NM-IS estimator



Complexity:		NM-IS trials	vs.	MC trials
Stage 1	2000	5600		800 000
+ Stage 2	3600			

Example 2: Read Write Trade Off



	VT1 shift	VT2 shift	VT3 shift	VT4 shift	VT5 shift	VT6 shift	Quad norm
Read 1	-0.78	1.08	-2.14	2.59	-0.99	-1.99	4.24
Write 1	1.26	1.53	-1.25	0.13	1.74	3.05	4.22
Write 0	1.01	-1.09	-0.10	-1.53	2.75	2.70	4.41
Read 0	1.19	-1.98	3.03	-0.73	-1.82	-0.44	4.31

- Shift vectors generated by the algorithm agree with physical insight.
- NM-IS can serve as an auxiliary tool in establishing (in)dependence.
  - Write 1 vs. Write 0.

Concluding Remarks

- Proposed a novel simulation method based on statistical ideas.
  - Uses importance sampling approach.
  - Key contribution: systematic selection of the tilted distribution.
- Future direction
  - Employing large deviations based algorithms for statistical timing analysis.

Acknowledgement

- Work supported by DARPA ITMANET program under grant 1105741-1-TFIND and the C2S2 program, the Focus Center for Circuit & System Solutions, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.