DATE 13
Design, Automation & Test in Europe
18-22 March, 2013 - Grenoble, France
The European Event for Electronic System Design & Test

Leveraging Sensitivity Analysis for Fast, Accurate Estimation of SRAM Dynamic Write V_{MIN}

[1]Jim Boley, [2]Vikas Chandra, [2]Robert Aitken, [1]Benton H. Calhoun

[1]University of Virginia, [2]ARM R&D

University of Virginia School of Engineering and Applied Science

ARM®

Motivation

- Static stability metrics are optimistic for write and pessimistic for read
  - Assumes infinite pulse width
  - Doesn’t account for transient behavior
  - Upside: shorter simulation times
- Need dynamic metrics for more accurate prediction of V_{MIN}
  - Current metric is T_{CRIT} – the critical or minimum WL pulse width required to write the bitcell
- Focus on dynamic write-ability since write is limiting in newer technologies & static WNM is optimistic
- Problem: determining the dynamic write margin of the worst case cell in a large (i.e. >1 Mb) memory requires a prohibitively large number of Monte Carlo (MC) simulations
- Solution: create a model of the tail of the distribution based on a smaller set of MC samples to predict write margin of the worst case bitcell

Outline

- Motivation
- **Background**
- Curve Fitting
- Sensitivity Analysis Method
- Evaluation of Write Assist Techniques
- Conclusions

Background- Previous Work V_{MIN} Estimation

- **Static V_{MIN} method:**

1. **Run MC sim at multiple V_{DD} points**
2. **Fit each V_{DD} point to normal distribution (μ,σ)**
3. **Equation for μ,σ versus V_{DD}:**

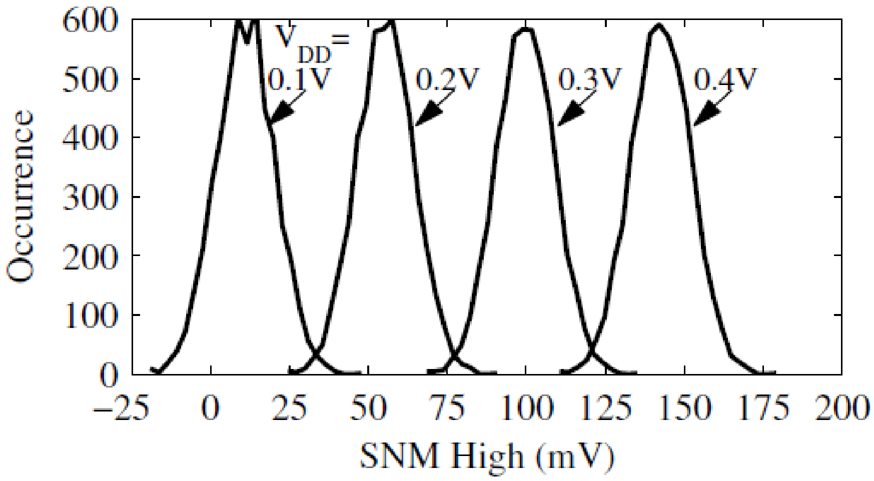
$$u = u_0 + k*(x - V_0), \sigma = \sigma_0$$

4. **Plug into CDF:**

$$F_{DRV}(x) = 1 - erfc\left(\frac{u_0 + k(x - V_0)}{\sqrt{2}\sigma}\right)$$

– **Above equation: P (DRV < V_{DD}) = x**

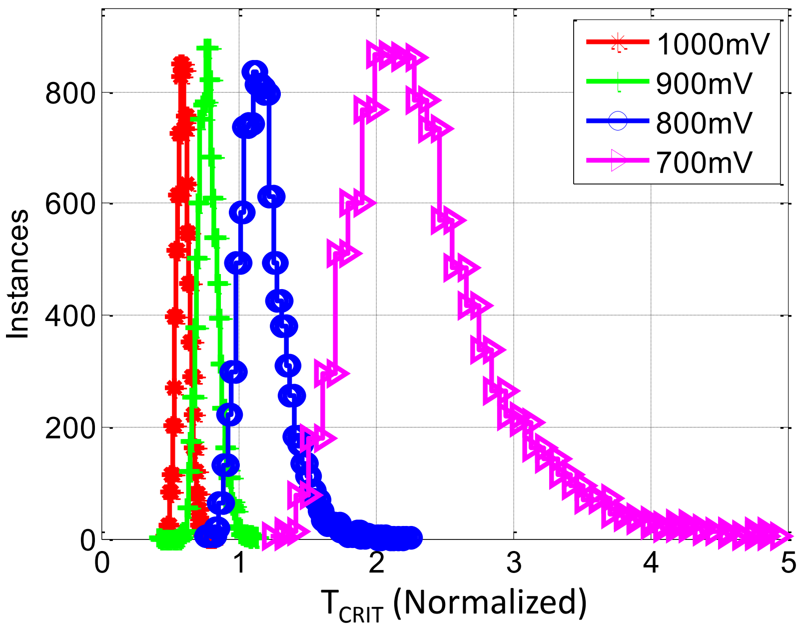
– **Same approach can be used to estimate static read/write V_{MIN}**



J. Wang et. al., "Statistical modeling for the Minimum Standby supply voltage of a Full SRAM Array," *ESSCIRC*, 2007.

Background- Dynamic Margin Distribution

- **Our approach: use a similar method as previous slide to estimate dynamic V_{MIN}**
- **Must find a long tail distribution that can accurately model**
- **Most importance characteristic of model is <u>Accuracy</u>**
  - **Underestimating V_{MIN} → reduced yields**
  - **Overestimating V_{MIN} → sacrificing potential power savings**

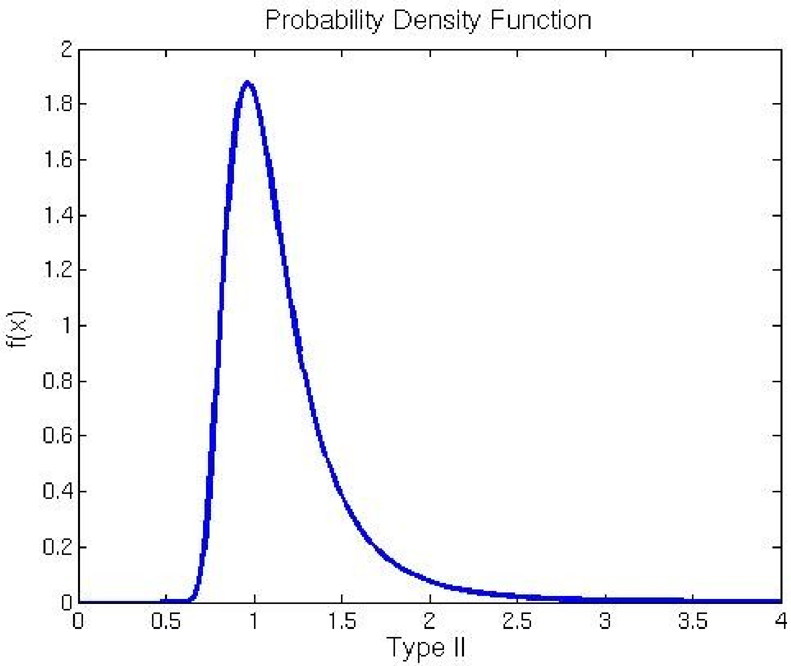


Problem: dynamic write margin distribution does not fit a normal distribution

Outline

- Motivation
- Background
- **Curve Fitting**
- Sensitivity Analysis Method
- Evaluation of Write Assist Techniques
- Conclusions

Model the Distribution

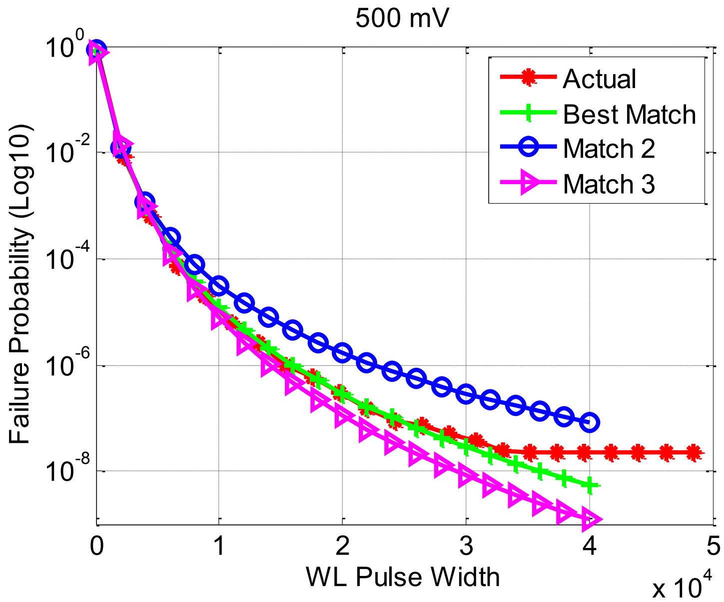


$$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x-u}\right)^{\alpha+1} \exp\left(-\left(\frac{\beta}{x-u}\right)^{\alpha}\right)$$

- **The extreme value distributions (EVD) are used to approximate the maxima of long sequences of random variables**
- **Initial approach:**
  - **Run small MC simulation with V_{DD} varying from 500 mV to 1V in increments of 100 mV**
  - **Fit each resulting distribution to the EVD**
  - **Formulate equations for the location, scale and shape parameters (μ, α, β) vs. V_{DD}**
  - **Plug these equations back into f(x) to calculate the failure probability across any V_{DD} point**

Initial Results

- Fitted distribution closely matches the MC data (~10^{-3}), but doesn't accurately model the tail region
- Curve fitting tool calculates large confidence interval for parameters (μ, α, β)
- "Actual" distribution calculated using importance sampling
- Constant failure probability → Static Failure

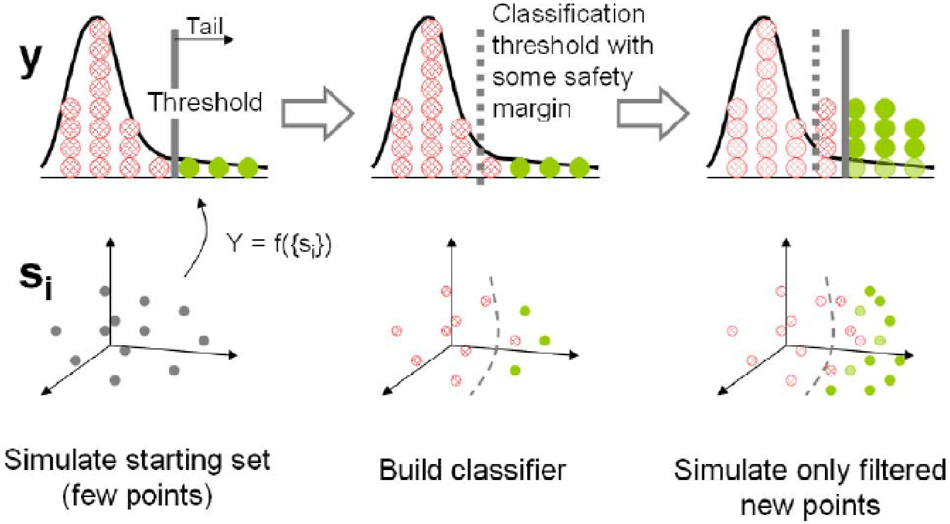


Observation: the shape of the tail can't be determined by a small MC simulation

Recursive Statistical Blockade

- **Statistical blockade- use small initial MC set to build a tail classifier**
- **Only simulate tail points**
- **Recursive statistical blockade further reduces the number of samples**

1. **Simulate initial points and build classifier**
2. **Run a 100K sample through the filter**
3. **Build a new classifier based on the 100K output to identify the 99th percentile cells**
4. **New filter now identifies 99.99th percentile**
5. **Simulate a 10M point sample set after filtering tail points**



A. Singhee and R. Rutenbar, "Statistical blockade: a novel method for very fast monte carlo simulation of rare circuit events, and its application," *DATE, 2007.*

How can we improve statistical blockade?

Downsides to statistical blockade:

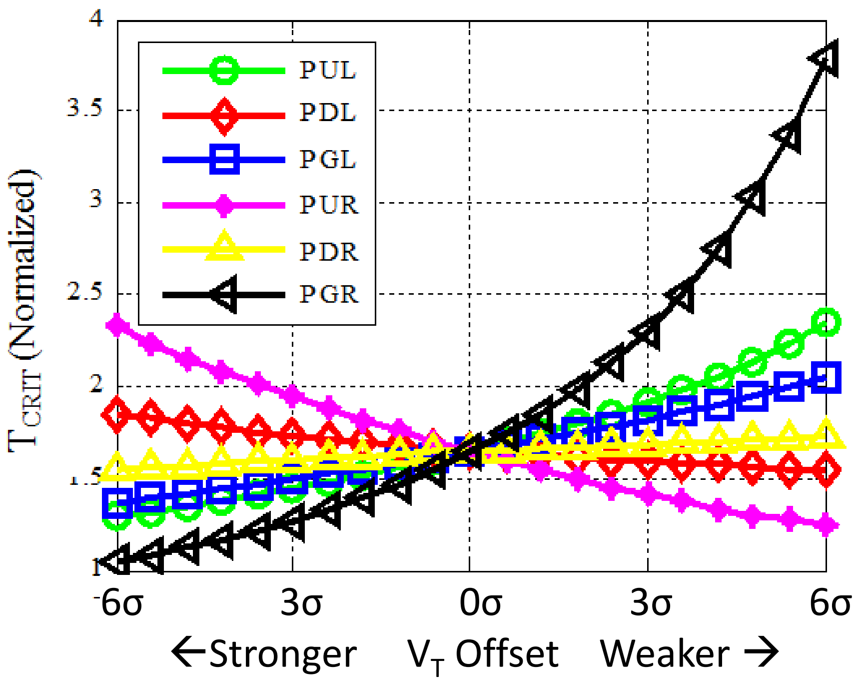
- Calculating T_{CRIT} requires running a binary search algorithm which on average takes 12 iterations
- In addition T_{CRIT0} and T_{CRIT1} require separate simulations
- Calculating the worst case bitcell in a 100 Mb array using recursive statistical blockade requires over 894,000 total simulations

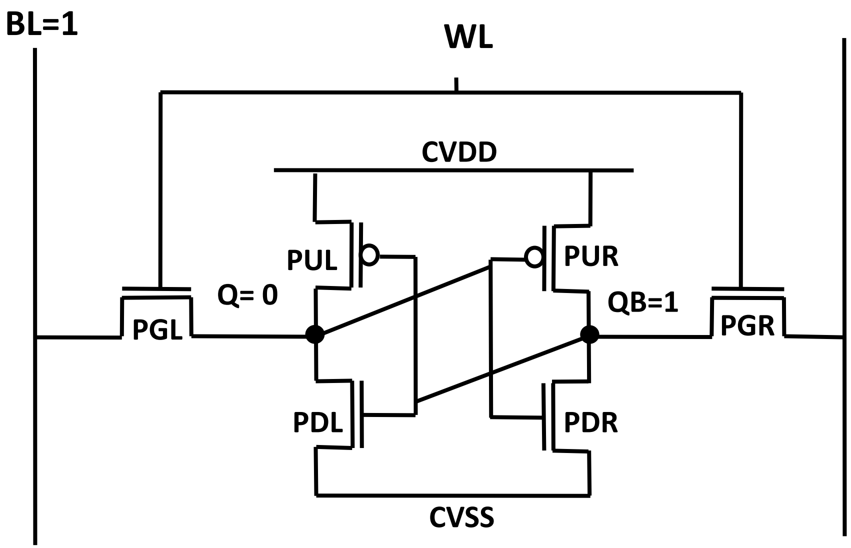
Outline

- Motivation
- Background
- Curve Fitting
- **Sensitivity Analysis Method**
- Evaluation of Write Assist Techniques
- Conclusions

New Approach: Sensitivity Analysis

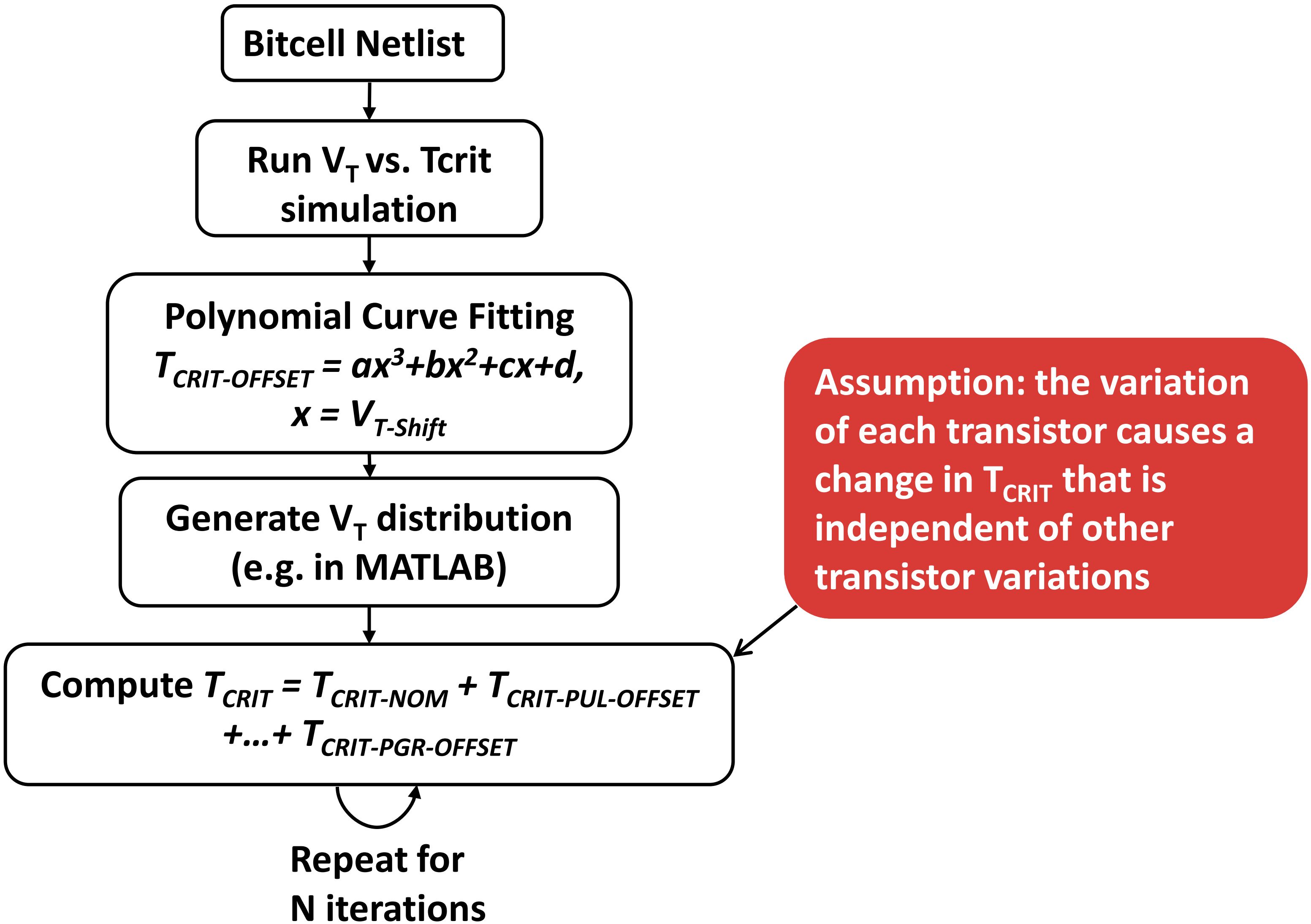
- Using sensitivity analysis we can calculate the expected T_{CRIT} value using the variation data generated by Monte Carlo
- Generating the V_T curves requires only 1080 simulations
- Once the V_T curves have been generated, the Monte Carlo data can run through the model, and the worst case bitcell can be quickly found





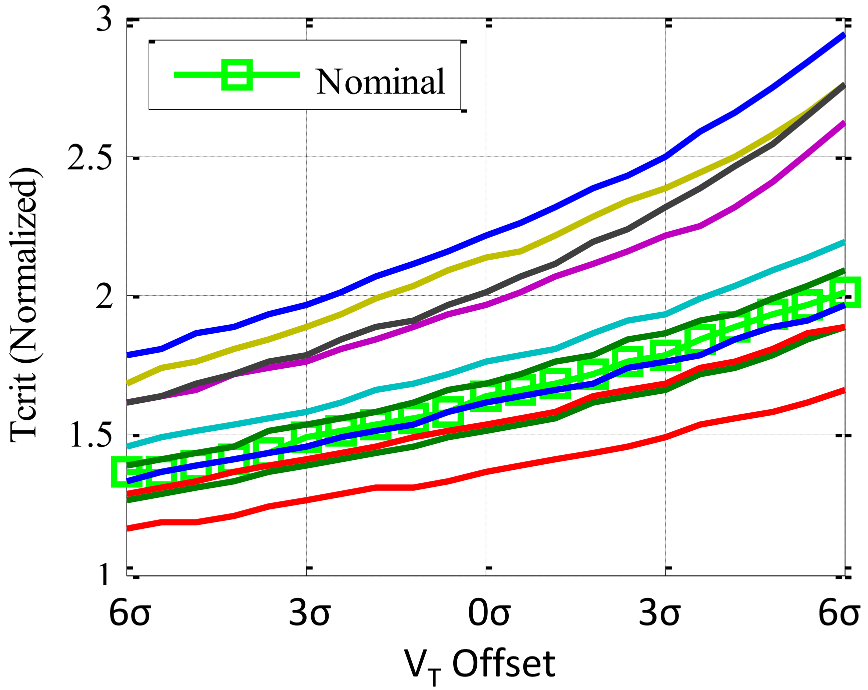
Use sensitivity analysis to calculate $\Delta T_{CRIT} / \Delta V_T$ for each transistor

Sensitivity Analysis Flowchart



Verifying Transistor Variation is Independent

- **Experimental setup:**
  - Add variation to other five transistors, sweep V_T of single transistor
  - Repeat for each transistor
- **Expected output:**
  - Shape of the sensitivity curve unchanged
  - Nominal value (0σ) shifted higher or lower



Verifying sensitivity analysis vs. statistical blockade

- **Comparison between the worst case bitcell as predicted by statistical blockade and sensitivity analysis**
- **Across VDD, the sensitivity analysis results match closely to the statistical blockade data**
- **The worst case percent error is 6.83% while the average is ~3%**

Modeled data vs. Statistical Blockade (Percentage Error)							
	500 mV	600 mV	700 mV	800 mV	900 mV	1000 mV	Average
100K	6.83	2.96	-0.18	0.83	-4.50	-2.72	3.01
10M	-4.25	-3.69	-2.64	-0.70	0.83	-2.20	2.39
100M	6.51	5.61	4.75	1.21	1.43	-2.27	3.63

What type of speed up does this method provide?

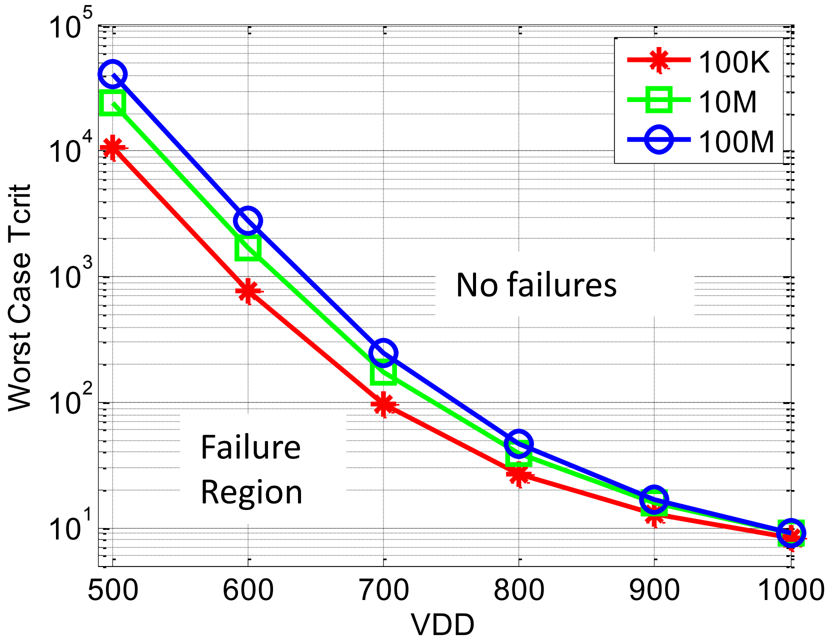
- SB analysis must be run on two cases: writing a 0 and writing a 1
- Total number of simulations: 894,288 (60 hours CPU time)
- Sensitivity analysis run time: 32 min, resulting in a speedup of 112.5x
- Note: results are for running at a single VDD point

	Statistical Blockade	Sensitivity Analysis
	Num. simulations	Run Time
Initial Simulation	24,000	18.8 min
100 Kb	107,904	0.72 s
10M	531,096	72 s
100M	231,288	12 min
Total Simulations	894,288	
Total Run Time	60 Hours	32 minutes

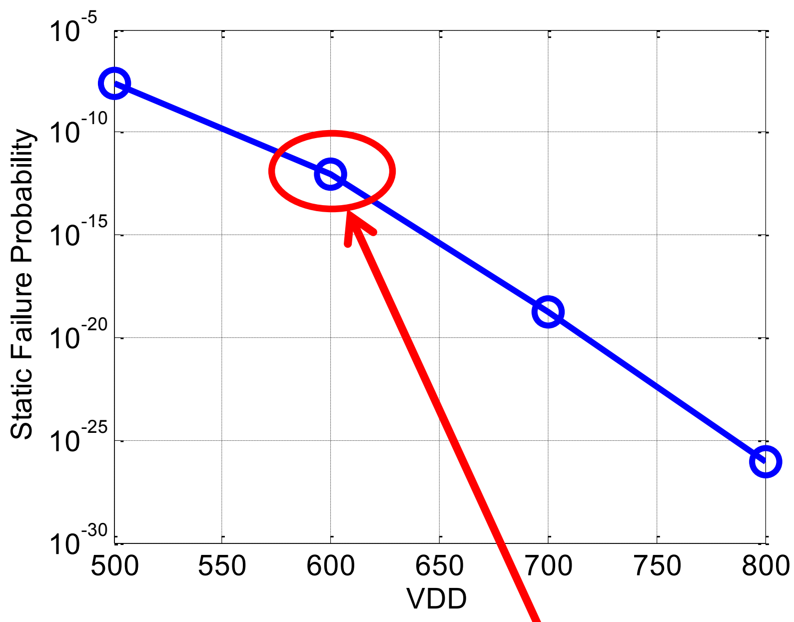
Sensitivity analysis provides a 112.5x speed up over recursive statistical blockade with an average percentage error of ~3%

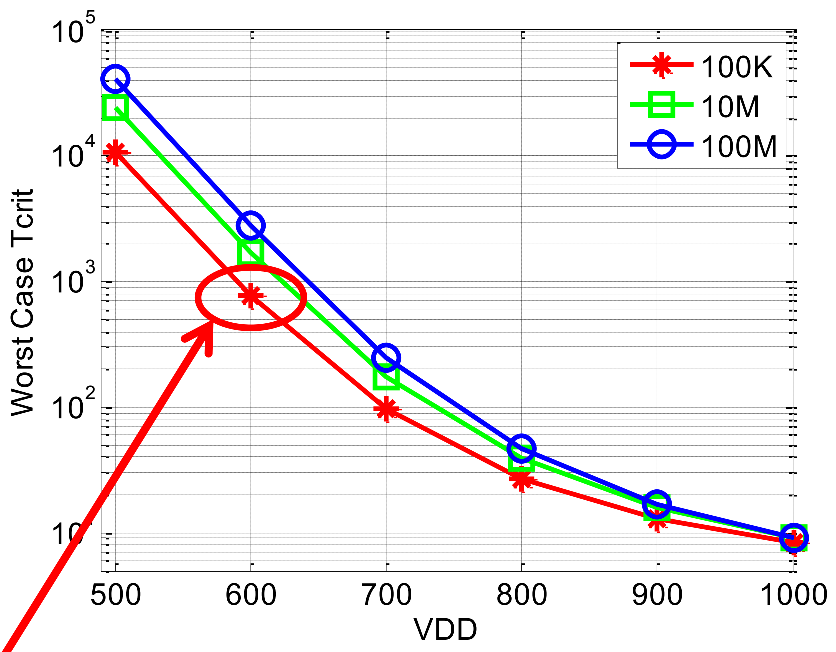
Applying Sensitivity Analysis to Dynamic V_{MIN}

- In order to calculate V_{MIN} we can repeat the procedure for varying VDD
- We chose 6 points from 0.5-1V
- The plot shows the worst case bitcell for a given VDD, varying the array size
- The curve represents the point of single bit failure, below the curve represents multiple failures, above the curve represents no failures



Comparison Between Static and Dynamic V_{MIN}



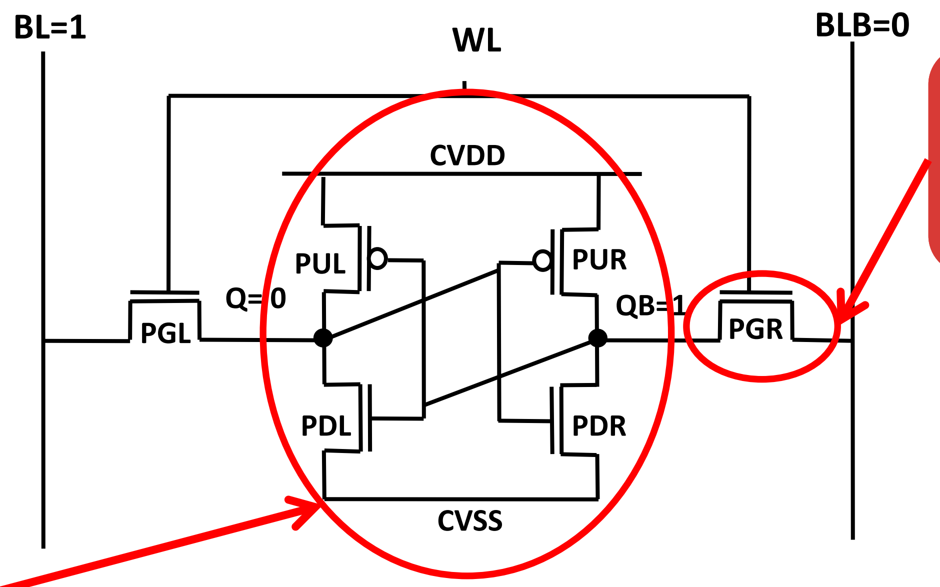


Static metrics show a failure probability <10^{-13} at 600 mV. Dynamic metrics show failure probability as high as 10^{-5} at the same V_{DD} for an aggressively scaled word line pulse width

Outline

- Motivation
- Background
- Curve Fitting
- Sensitivity Analysis Method
- **Evaluation of Write Assist Techniques**
- Conclusions

Using Sensitivity Analysis to Evaluate Write Assist Methods

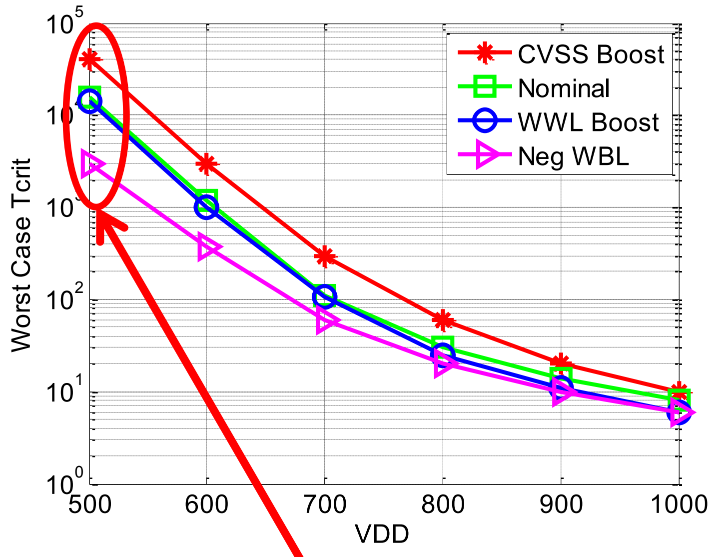


WWL Boost and Negative BL → Increase passgate strength (V_{GS}>V_{DD})

Cell VSS Boost → Reduce strength of the latch

Applying Sensitivity Analysis to Assist Methods

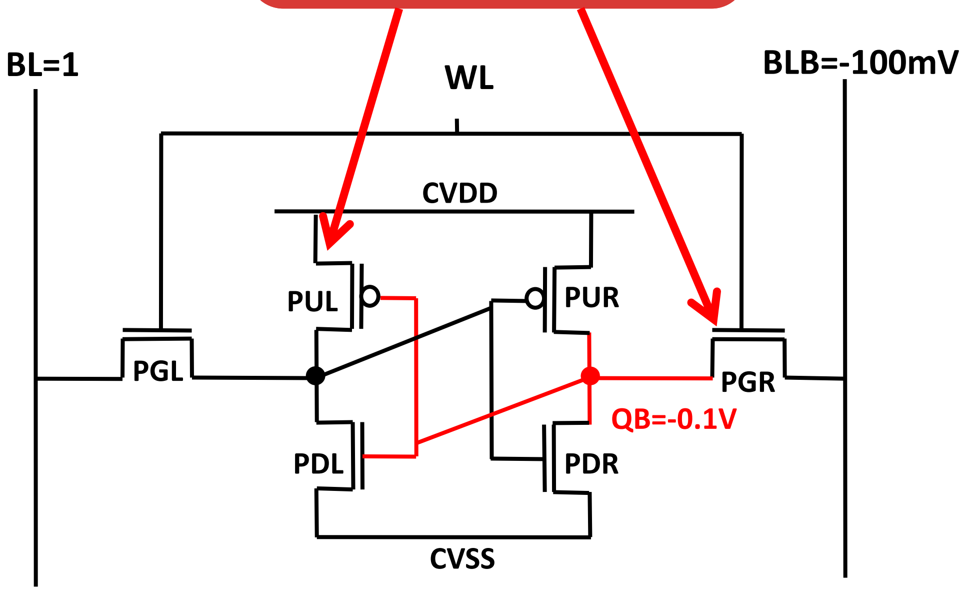
- For the example on the right the memory size is 1 Mb (i.e. P_{FAIL} = 1e-6)
- As VDD is reduced, the negative WBL technique has a significant advantage over the other assist methods
- Note: semi-log scale
- ΔV = 100 mV for each assist method



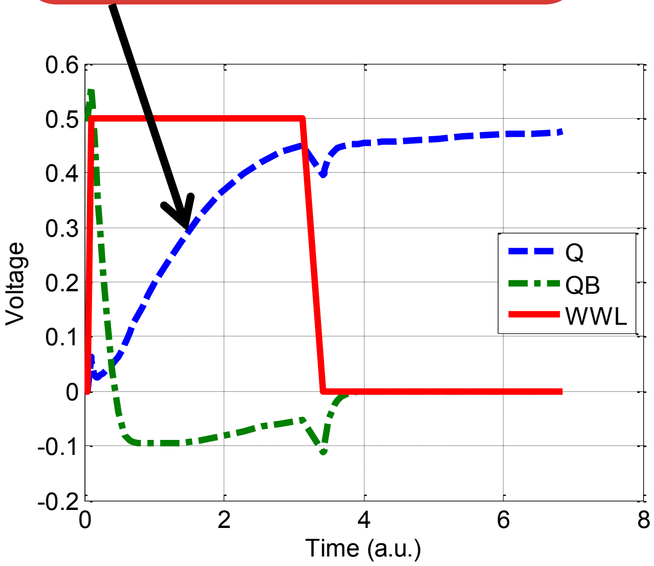
Negative BL reduction reduces the worst case T_{CRIT} close to an order of magnitude more than WWL boosting at 500 mV

Advantage of Negative BL Reduction

Negative BL strengthens pass-gate and pull-up devices



At 500 mV, the majority of write time is spent pulling Q high



Negative BL passes a stronger '0' into the cell, effectively strengthening the PUL transistor and decreasing the write time. This gives negative BL an advantage over WWL boosting at lower VDD.

Conclusions

- Modeling the tail of the dynamic write margin using a small Monte Carlo simulation is not effective
- Statistical blockade is good method for reducing simulation time, however evaluating dynamic V_{MIN} still requires a large number of simulations
- Sensitivity analysis provides a speed up over recursive statistical blockade of 112 x with an average percentage error of ~3% across VDD
- Using this analysis, we have shown that negative BL reduction is the best method for reducing dynamic V_{MIN}