Title
Analysis and Modeling of Large-Scale Variation on DACs and SRAMs

Permalink
https://escholarship.org/uc/item/21s900g6

Author
Park, Henry Arnold

Publication Date
2013

UNIVERSITY OF CALIFORNIA

Los Angeles

Analysis and Modeling of Large-Scale
Variation on DACs and SRAMs

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Electrical Engineering

by

Henry Arnold Park

2013

ABSTRACT OF THE DISSERTATION

Analysis and Modeling of Large-Scale Variation on DACs and SRAMs

by

Henry Arnold Park

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2013

Professor Chih-Kong Ken Yang, Chair

Evolution of CMOS circuits has been leveraged by continuous scaling of the feature size. Scaling has enabled integration of several billions of transistors in a single die with lower power consumption and throughput increase for the last two decades [1]. Behind of such technological advance, however, the increasing process variability over die-to-die and within-die is a growing issue for the system reliability [2-3]. The device model for the most advanced technology is becoming too complicated [59] due to diverse physical effects arising from short channel length and high field [60-62], which makes it hard to rely on the model accuracy to precise estimate the productivity and reliability of a large-scale system. Therefore, the need of accurate yield estimation model (or tool) based on the measurement data is required.

This work presents simple and fast reliability estimation techniques for two of the most widely used systems: digital-to-analog converter (DAC) and SRAM. The DAC is exclusively adopted by most mixed-signal systems such as high performance transceivers,

digital phase-locked loop, clock-data recovery, and successive approximation. Depending on their purpose, the DAC design may demand different design targets. While many of the systematic performance degradations (especially dynamic linearity) can be handled by careful layout, circuit architecture, segmentation, and switching algorithm, the nonlinearity caused by unit element mismatch can only be handled by sizing up the device or by calibration. In any case, the achievable minimum nonlinearity should be carefully considered from a yield estimation model. This model must be based on measurable mismatch information of the unit element, such as unit current in a current-steering DAC, and it should be applicable to arbitrarily segmented structure. From the survey of existing models and their limitations, this work proposes two general models for the differential nonlinearity (DNL) and integral nonlinearity (INL) yield. The validity of the model is verified by measurement data from an 8-bit current-steering DAC fabricated in 90nm CMOS.

The second case study is for SRAM. Most microprocessors have various cache memories that are usually built by SRAM for its robust data retention and high access speed for both read and write. In a recent trend of multi-core processors in a single die in association with the decreasing feature size, the number of SRAM blocks and density of the SRAM cell increase such that reliability becomes a serious issue. As an old tradition of SRAM design, the size of the cell is generally determined by yield from static stability margin or from dynamic perspectives that are relying on the accuracy of the device model. Rapid yield estimation techniques such as importance sampling or response surface model are the extreme case, as their predictability of failure depends on the assumed variability of the few major parameters such as threshold level.

For better estimation of SRAM yield, built-in self-test (BIST) circuits are suggested in numerous literatures that can improve the predictability of failure conditions. This failure condition is particularly useful for gauging time-dependent stability variation of the memory cell due to diverse effects such as NBTI [112] and aging effects. The estimated failure condition found in BIST circuits can be used to counteract the failure mechanism

to decrease the fail bit count such as controlling the cell supply. Such varying failure condition cannot be detected by traditional pass/fail based test. In addition, knowing the analog level of the stability distribution greatly helps in reducing the power of the entire memory array by exploring the optimum lower supply level without failing read/write operation. Another particular utilization of the stability information is to correlation to the device model. SRAM designers face countless combinations of the device parameters to satisfy a certain stability margin. With large stability measurement data, their design strategy can be more reliably verified.

However, none of the proposed techniques can be applied to a large memory array because of the speed issue or relatively incorrect estimation result. This work proposes a rapid yield estimation technique for concerning static stability. By using small size on-chip ADC and direct bit-line access technique, the static read stability and write-ability of 6T SRAM cells are characterized. From the definitions of the new dynamic stability, the close correlation between the static estimation and the dynamic characteristics are demonstrated. The estimation results match very well to the measured stability from a test chip in 65nm CMOS.

The dissertation of Henry Arnold Park is approved.

_____

Mario Gerla

_____

Dejan Marković

_____

Jason Woo

_____

Chih-Kong Ken Yang, Committee Chair

University of California, Los Angeles

2013

To my family and my fiancé

Table of Contents

List of Tables

List of Figures

xiv

xv

List of Equations

Acknowledgements

VITA

1999-2003	B.S. Electrical Engineering (summa cum laude) Seoul National University Seoul, Korea
2003-2006	System Engineer Hunter Technology Seoul, Korea
2006-2009	M.S. Electrical Engineering University of California, Los Angeles Los Angeles, California
2009	Analog IC Design Intern Broadcom Irvine, California
2006-2012	Graduate Student Researcher University of California, Los Angeles Los Angeles, California
2012-present	Scientist Staff - IC Design Broadcom Irvine, California

PUBLICATIONS

- Journals:
  - Henry Park and Chih-Kong Ken Yang, "In-situ SRAM static and dynamic stability estimation in 65nm CMOS," IEEE JSSC, will be submitted.
  - Fengbo Ren, Henry Park, Chih-Kong Ken Yang and Dejan Marković, "Sensing margin improvement by reference calibration for body-voltage sensing in STT-RAMs," IEEE TCAS-I, accepted for publication.
  - Henry Park and Chih-Kong Ken Yang, "Stability estimation of a 6-T SRAM cell using a nonlinear regression," IEEE TVLSI, accepted for publication.
  - Henry Park and Chih-Kong Ken Yang, "An INL yield model of the digital-to-analog converter," IEEE TCAS-I, Dec. 2012.
  - Henry Park and Chih-Kong Ken Yang, "Nearly exact analytical formulation of the DNL yield of the digital-to-analog converter," IEEE TCAS-II, vol. 59, no. 9, Sep. 2012.
- Conference
  - Ramy Yousry, Henry Park, E-Hung Chen, and Chih-Kong Ken Yang, "A digitally-calibrated 10 GS/s reconfigurable flash ADC in 65nm CMOS," ISCAS 2013.
  - Henry Park, Richard Dorrance, Amr Amin, Fengbo Ren, Dejan Marković, and Chih-Kong Ken Yang, "Analysis of STT-RAM Cell Design with Multiple MTJs per Access," NANOARCH '11, pp. 32-36, 2011.
  - Fengbo Ren, Henry Park, Richard Dorrance, Yuta Toriyama, Chih-Kong Ken Yang and Dejan Marković, "A Body-Voltage-Sensing-Based Short Pulse Reading Circuit for Spin-Torque Transfer RAMs," ISQED'12, pp. 275-282, 2012.

Chapter 1

Introduction

Of the many important blocks consisting of a large scale mixed-signal system on a chip, DAC and SRAM are two common and important building blocks. These two systems are similar in implementation such that an identical unit element such as the current source in the DAC or the data storage cell in the SRAM is repeatedly used to comprise the whole system. As the number of the unit elements can be from several hundreds to millions in a typical mixed-signal system, the sizing of the element is the key factor in the area of the system. A smaller cell size can be used to increase the cell density (or reduce the whole system area) as well as to enhance the performance such as speed but the random fluctuation of device parameters is known to cause reliability issues. The goal of this work is to provide mechanisms to predict the reliability and yield. For DACs, we provide a model that has reduced computational order and accurate for design. For SRAMs, we introduce a measurement and estimation approach for post-silicon characterization and optimization.

One of a DAC's performance specifications is the linearity of its input-to-output (I/O) transfer characteristics. Statistically, the non-linear portion of the I/O curve caused by the random mismatch may exceed the required resolution or target output harmonic leading

to a reduced yield. The yield of a design can be estimated by running several tens of thousands MC simulations but this method can take hours for a high resolution DAC for each design iteration. Therefore, a computationally efficient yield model that relates by the unit element variation with the expected yield benefits a designer allowing her to quickly evaluate the impact of changing the size or bias of unit elements. Previously published models are discussed in Chapter 2. This work proposes yield models for both INL and DNL using multivariate Gaussian random variables that can be extended to arbitrarily segmented structures and takes less than a second to produce an accurate estimate using a mathematical tool such as MATLAB.

For an SRAM, a similar issue exists for estimating the reliability of the entire memory array. The yield is defined by the number of cells that fail a read or write operation. As the SRAM cell is made with approximately minimum size devices, device variability is the primary cause of cell instability. Many publications have quantified the read/write stability and are reviewed in Chapter 4. A function that predicts yield is not easily derived from unit cell variation due to complex transient effects. Designs that meet yield requirements across all process corners in addition to device variation are typically substantially overdesigned in terms of supply voltage or device sizing. Instead of developing a yield-prediction model as we do for DACs, by obtaining the stability of the cells directly from a post-silicon IC, supply voltage can be better optimized to reduce power or be more robust to degradation of stability over time due to NBTI. As an example, the read stability of a SRAM cell can be enhanced by increasing the cell supply [89], [91-93], [97], [106] or reducing the word-line voltage [87], [94-104], [106]. Current on-die measurement approaches are reviewed in Chapter 4. Accurate measurements typically require measurement times of more than days for even small memory sizes such as 32kb. More rapidly estimation of SRAM stability have not been accurate with fitting accuracy as low as $R^2 \sim 0.6$ [84], [117]. This dissertation proposes a rapid and accurate stability estimation technique from cell currents under varying supply that can provide

2

estimation accuracy of $R^2 > 0.8$ with measurement times under a second for 32kb memory arrays.

This chapter begins with a brief description of DAC and SRAM applications and how the basic structure depends on the matching of devices. The impact of technology scaling on device matching and as a result yield is discussed in Section 1.2. Section 1.3 presents an overview of the dissertation.

Section 1.1   DAC and SRAM Application

A DAC is an important component in mixed-signal processing and communication systems. It can be used as a full-scale signal generator like a transmitter [16-17] or can be associated with other feedback systems such as digital phase-locked loop (DPLL) [24], clock and data recovery (CDR) system [23], or successive approximation data converters [14]. A simple and commonly used approach to produce precise analog levels is to rely upon the matching of nominally identical unit cells in the fabrication process. Design of the unit cell involves diverse trade-off such as nonlinearity caused by device mismatch and by frequency dependent impedance modulation. Most of the systematic effects such as output impedance are addressed by circuit architecture. Variation of unit cells due to the inherently stochastic nature of dopant implantation and physical geometry [4-5] are among the root causes of static and dynamic nonlinearity.

For digital processing applications, an SRAM is an essential memory block. Due to its reliable data retention and rapid read and write access, SRAMs are mostly adopted as low-level caches in processors for immediate high-locality data storage. Typically, the more processing cores in a system, the larger the memory blocks that are integrated. The denser and larger memories are possible by the continued scaling the size of SRAM bit cell. As shown in Figure 1 [1][65], the size of the SRAM bit cell has maintained a scaling rate of 50% per every technology node. As a result of the scaling, the SRAM design is

becoming increasingly difficult. Stability of the stored element depends upon limiting the parametric mismatch [4-5], [65], [71-75] and process variability [2-3], [64].

Section 1.2   DAC and SRAM Reliability

Figure 2 shows the threshold voltage variability per technology node from 65nm down to a predictive 15nm [87]. In this plot, two scenarios are assumed for the scaling of the oxide thickness: 1) the effective oxide thickness (EOT) does not scale further after 65nm and 2) EOT scales. As EOT is proportional to Pelgrom's mismatch coefficient (A_{Vth} in [5]) according to [66-67], the first scenario assumes the worst threshold voltage variation with the scaling of the bit cell size. In reality after 32nm, high dielectric-K or multi-gate device such as FinFETs [63] improves the gate controllability and the threshold mismatch does not scale rapidly as the first scenario in Figure 2. However, even with the improved matching property, the scaling of the feature size increases the absolute magnitude of the threshold voltage variation as shown in the second scenario with EOT scaling in Figure 2.

The reliability issue has led to the relatively invariant unit device size for the DAC and nearly constant supply scaling for the SRAM (Figure 1). For a DAC, proper sizing of the unit cell is necessary to reduce the statistical variation. Even designs with explicit calibration such as dynamic-element matching (DEM) [6-8], foreground calibration [9, 45], and background calibration [10-14] still require a minimum yield since the calibration typically has a limited range. Hence, a proper yield model is required to explore the correction limits by calibration.

For an SRAM, while many publications suggest different techniques to reduce the minimum operating supply (VCC_{min}) such as read and write assist [87-106], the existing publications do not provide a systematical approach to set the optimum calibration targets for the assistant circuitry that is applicable to each die. Only process dependent variation can be tracked using built-in resistor divider [97-99], [103] or feedback circuitry [96]; however, neither method provides stability margin for the cells close to the failure conditions.

4
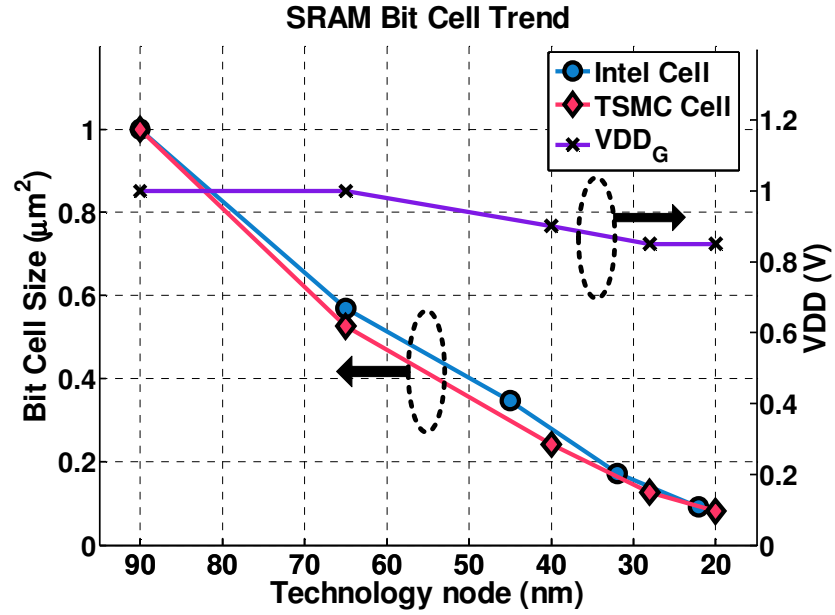
SRAM Bit Cell Trend

Figure 1.   SRAM bit cell trend (size and supply level). Courtesy of [1].

σ_{Vth} **Scaling Trends of SRAM Bit Cell**
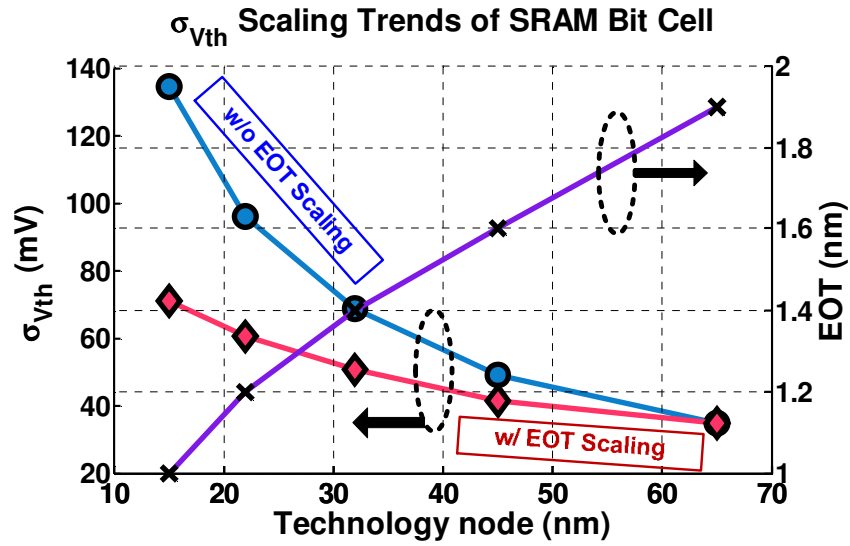
Figure 2.   Threshold voltage variation trends of the SRAM bit cell (size and

supply level). Courtesy of [87].

Section 1.3   Organization

Chapter 2 presents an overview of DAC design issues where the I/O nonlinearity caused by device mismatch is introduced. This chapter also describes the complexity of existing yield models for integral and differential nonlinearity (INL and DNL) since the yield calculation requires an integration over multi-dimensional Gaussian variables with a correlation matrix.

This work proposes general yield models based on arbitrary unit cell variation and arbitrary structure and is described in Chapter 3. The yield of the DAC can be rapidly estimated using device variation models from either device simulations or process characterization measurements. A simplification of multivariate Gaussian random variables is suggested and applied to calculating a DAC's INL and DNL by reducing by the function to a few terms. The yield model can be applied to any arbitrarily segmented DAC in contrast to current models that only apply to either binary or thermometer DACs. The model results are verified by both MC simulation and from measurement data. The data is extracted from an 8-bit current-steering DAC in 90nm CMOS. The measurement data shows excellent matching even though the model simplifies the random variation to Gaussian distributions.

Chapter 4 reviews the design issues on SRAM with particularly focus on causes of cell instability. The stored data in a cell should remain undisrupted during read while the cell should be successfully programmable during write (read-stability and write-ability [68-69], [75-80]). This chapter revisits published ideas that gauge the static and dynamic stability of the SRAM cell. The chapter ends by focusing on stability metrics that can be easily measured with built-in self-test (BIST) such as supply read-retention voltage (SRRV) and bit-line write trip voltage (BWTV).

Chapter 5 utilizes the measured stability metrics to determine an estimation function that relates the stability to a minimal set of measured current. By separately measuring the predictor (cell current) and the estimator (stability) variables, a close relation between the two variables is found from a subset of memory array by using a non-linear

regression method. This technique can characterize the device variability and its impact on the read/write stability of the cell. Various trade-offs between the measurement requirements and the estimation accuracy are described in details. This work also shows that if the target control factor is identical, then there is a close relation between the static and dynamic stability. For instance, if the cell supply is lowered until the initial state is flipped during read access, the supply voltages measured during the static access (i.e. WL is always high) and during the pulsed access are strongly correlated especially for the cells close to failure.

The proposed technique is verified by a chip implementation in Chapter 6. The built-in self-testing (BIST) circuitry can rapidly measure the stability of the SRAM cell without significant change of the memory array structure. The static stability estimation is demonstrated with a test chip in 65nm CMOS. This chapter also shows that near the edge of the stability distribution, the measured static and dynamic stability are strongly correlated. Using this fact, the estimation of the dynamic stability can be easily done from the estimated stability and a known factor from the correlation. This idea of matching static and dynamic behavior is verified across multiple dice.

Chapter 2
Background on DAC Yield

This chapter revisits the fundamentals of the device random mismatch and its impact on DAC's input/output characteristics. For simplicity, current-steering DAC is analyzed in detail but the fundamentals can be applied to other types of DACs. Based on the arguments observed in this chapter, the next chapter suggests two yield models that can effectively gauge the maximum performance boundary caused by the unit element mismatch.

Section 2.1   Current-Steering DAC Design

The current-steering DAC is comprised of identical unit current source (Figure 3). The input digital signal controls how many unit sources to be steered to OUT while the remaining units are steered to the other output terminal. This differential output current is converted to voltage signal by the termination resistance (50 ohms). Although the output level is determined by the integer number of cells that are chosen to drive the OUT

terminal, the I/O curve may not be linear due to random mismatch between the unit elements. Section 2.2 describes the statistical modeling of the random mismatch of the unit elements and their impacts on the I/O linearity.

Another key concern is the structure. The DAC structure may vary depending on how to group the current source array. If each current source is independently controlled by a dedicated bit such as thermometer-coded DAC, the area of the switch and control logics is too large for a high resolution DAC > 10-bit. For example, if a 14-bit DAC is implemented in a thermometer-coded structure, the total number of switches should be 16k and each switch should be driven by the same number of latches. The binary structure is more compact but it shows more glitch energy and more serious linearity issues than the thermometer-coded DAC. The optimum DAC structure combines the binary for LSBs and the thermometer for MSBs. Table 2-1 compares the specs between the thermometer and binary DAC [18]. While the thermometer-coded structure has better glitch and DNL (Section 2.2), the binary structure prevails in aspects of the power, area, and layout complexity. Our DAC yield models are for any arbitrarily segmented DAC that visualizes the yield variation between different structures.

Table 2-1 Comparison between thermometer and binary structure [18].

Specs	Thermometer	Binary
INL	Normal	Normal
DNL	Good	Poor
Glitch	Less harmonics	More Harmonics
Complexity (wiring, logics)	Complex	Simple
Power	High	Low
Area	Large	Small

Section 2.2   Impact of Random Mismatch

Statistical variation of the unit element can be described as the sum of a nominal value and a random variable. For an N-bit current-steering DAC, each unit current source (I_i) is

modeled with a nominal current ($\bar{I}$) and a variation (ΔI_i) as shown in Figure 3. For a large N-bit DAC, $\bar{I}$ approximately equals to the 1 LSB size. The variation defined as ΔI_i can be modeled as a Gaussian random variable with zero mean and standard deviation σ_I.
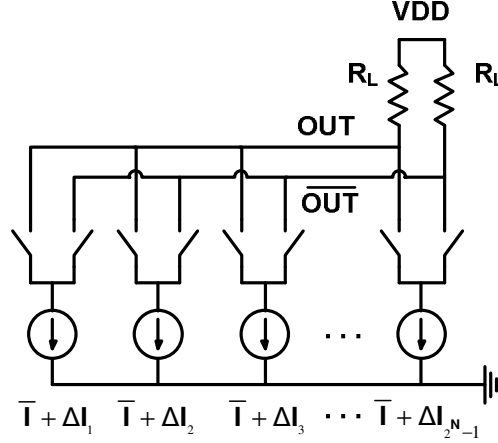
Figure 3.   An N-bit current-steering DAC with unary structure. Variation of each current source is modeled as an additive Gaussian random variable (ΔI) with a variance $\sigma_I{}^2$.

The identical variation model can be used to other types of elements such as a resistor (R + ΔR_i) or a capacitor (C + ΔC_i).

$$I_i = \bar{I} + \Delta I_i, \ 1 \le i \le 2^N - 1$$

(Eq 2.1)

The variation ΔI_i results in random fluctuation of the analog output during input transition. This variation for each code transition when normalized by the mean current ($\bar{I}$) is defined as the differential nonlinearity (DNL). As the number of unit elements switched at every input code is strongly dependent on the structure of the DAC, a binary-coded DAC shows the worst output fluctuation especially when the MSB group switches. The integral nonlinearity (INL) is defined to be the distance between each output level

and the ideal linear line from interpolating the ends of the transfer curve normalized by the LSB.

The relative importance of DNL versus INL depends on the application and architecture. For a system that uses the full-scale output of a DAC for data transmission [16-17], the output spectrum requirements (e.g. SFDR) are strongly affected by the maximum tolerable INL [18-22]. However, for a DAC that is used inside a feedback loop such as CDR [23], DPLL [24], or successive approximation [14], INL is not a critical design target as long as the I/O characteristic of the DAC is monotonic. More importantly in this case, the maximum tolerable DNL should be carefully explored to avoid substantial performance degradation by incomplete settling of the loop (or even oscillation) due to a missing state or a non-monotonic I/O. Architecture of the DAC can be important. Compared to DNL, INL is relatively less sensitive to the segmentation structure. In general, the design specifies bounds that may be different for INL and DNL to meet a target yield..

In the output structure of an N-bit segmented DAC with the last NB-LSBs assigned to the binary group, the increment of the output between the (i-1)-th and the i-th input code, $\Delta I_{out}(i)$, is expressed as

$$
\begin{aligned}
I_{out}\left(i\right) &= i \cdot \overline{I} + \sum_{j=1}^{NB} D_j\left(i\right) \cdot \Delta I_{Bj} + \sum_{k=1}^{2^{(N-NB)}-1} U_k\left(i\right) \cdot \Delta I_{Tk} \\
\Delta I_{out}\left(i\right) &= I_{out}\left(i\right) - I_{out}\left(i-1\right) = \\
\overline{I} &+ \sum_{j=1}^{NB} \left(D_j\left(i\right) - D_j\left(i-1\right)\right) \cdot \Delta I_{Bj} + \sum_{k=1}^{2^{(N-NB)}-1} T_k\left(i\right) \cdot \Delta I_{Tk}
\end{aligned}
\tag{Eq 2.2}
$$

A degenerate form of the equation can be used to express the current for either a binary or thermometer by removing the thermometer or binary current terms. In this equation, $D_j(i)$ is a binary representation (0 or 1) of the input code (i) at the j-th LSB. ΔI_{Bj} is the sum of the current variations of the j-th binary group. For the thermometer group, $U_k(i)$ is

1 if the input code (i) is equal to or greater than k×2NB otherwise $U_k(i)$ is 0. Similarly,

ΔI_{Tk} is the sum of the current variations of the k-th thermometer group. $T_k(i)$ is 1 if and only if the input code (i) is equal to $k \times 2^{NB}$. From (Eq. 2.2), the DNL and INL of the i-th input code position can be expressed as the following equations. For estimating a high DNL yield level with a small unit cell variation, the averaged variation terms in (Eq. 2.3) can be ignored compared to $\bar{I}$ leading to the approximation. Note that this (unity gain) approximation cannot be applied when expressing the INL.

$$DNL_{LSB}(i) = \frac{\Delta I_{out}(i) - \bar{I} - \dfrac{1}{2^N - 1}\displaystyle\sum_{j=1}^{2^N - 1}\Delta I_j}{\bar{I} + \dfrac{1}{2^N - 1}\displaystyle\sum_{j=1}^{2^N - 1}\Delta I_j} \approx \frac{\Delta I_{out}(i) - \bar{I}}{\bar{I}} \qquad \text{(Eq 2.3)}$$

$$INL_{LSB}(i) = \frac{I_{out}(i) - i \cdot \bar{I} - \dfrac{i}{2^N - 1}\displaystyle\sum_{j=1}^{2^N - 1}\Delta I_j}{\bar{I} + \dfrac{1}{2^N - 1}\displaystyle\sum_{j=1}^{2^N - 1}\Delta I_j}$$

$$\approx \frac{\displaystyle\sum_{j=1}^{NB} D_j(i) \cdot \Delta I_{Bj} + \displaystyle\sum_{k=1}^{2^{(N-NB)}-1} U_k(i) \cdot \Delta I_{Tk} - \dfrac{i}{2^N - 1}\displaystyle\sum_{j=1}^{2^N - 1}\Delta I_j}{\bar{I}} \qquad \text{(Eq 2.4)}$$

Equation (Eq. 2.4) is a common expression for INL for each input code, i. In (Eq. 2.4), when the resolution of the DAC is sufficiently high (≥ 8 bits), the denominator of (Eq. 2.4) can be simplified to the mean current of the unit cell, $\bar{I}$. The first term in the numerator of (Eq. 2.4) represents the ideal linear output obtained from the curve interpolation. Due to averaging by (2^N-1), the amount of this additive term between adjacent input codes is negligible.

The maximum INL and DNL are typically specified to be bounded to ±A LSBs (defined as the decision boundary) to avoid significant nonlinear distortion of the N-bit output signal. Due to the relatively small impact of a single variable on the INL (or DNL),

the INL values at different input codes are highly correlated. Due to this correlation, an accurate estimation of the INL yield would require integration of a multivariate Gaussian probability density function (PDF) with order 2^N-1. The complexity grows exponentially with the number of bits and can be numerically intractable. A simplified approach to the yield based on multivariate Gaussian random variables is proposed in chapter 3.
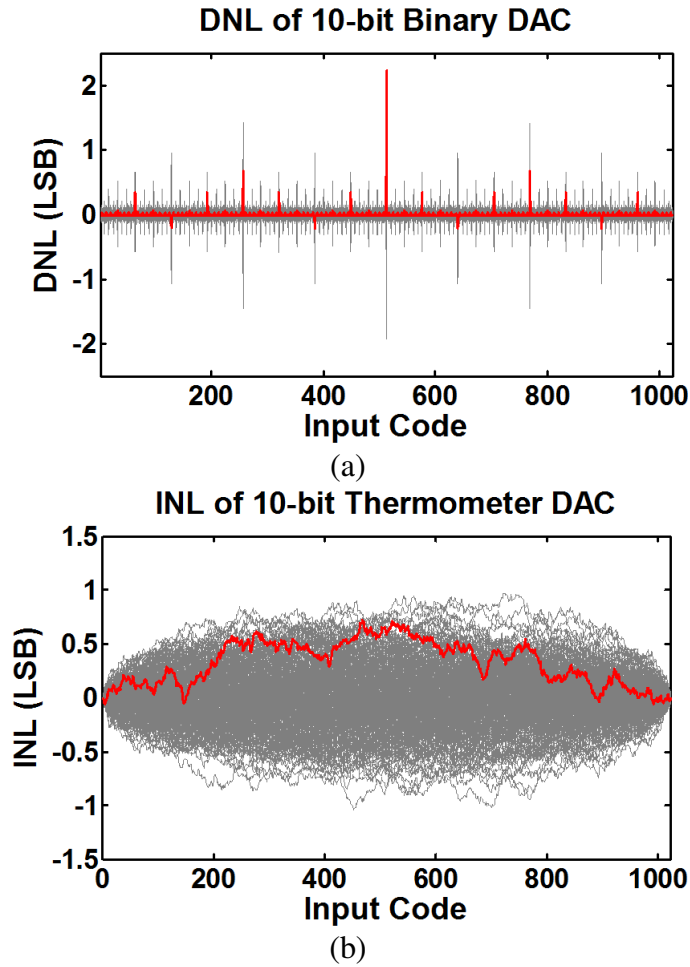
Figure 4.   (a) DNL and (b) INL of 10-bit binary DAC samples.

The SFDR caused by the element mismatch can be formulated for the binary-coded DAC [21] by using the normalized sigma of the current variation.

$$SFDR \approx 20\log\frac{3\pi}{4} + 3N - 20\log\left(\frac{\sigma_I}{I}\right)$$

(Eq 2.5)

(Eq. 2.5) clearly shows that halving the variability of the unit element can increase the dynamic linearity by 20dB. For this reason, static calibration techniques proposed in [6-10] also enhance the dynamic linearity for all input frequency range.

Section 2.3   Yield Model for DAC Linearity

   This dissertation focuses on a estimating the percentage yield for a design to meet a target INL or DNL with a given DAC architecture (number of bits and segmentation) as a function of unit device variability. A common way to estimate the INL and DNL yield of the DAC is to run behavioral Monte-Carlo (MC) simulations by assuming independent Gaussian distribution of the unit cell element. From multiple MC simulations combined with a sizing algorithm, proper requirements on the unit cell can be found that determines geometry and bias conditions [5]. However, MC simulations require a large number of runs to accurately estimate the yield. According to [25], for example, at least 40,000 runs are necessary when targeting 99% yield level with 5% estimation error sigma (i.e. the yield results is bound by [98.95%, 99.05%] for in ±1σ certainty). Since the yield should be verified by running multiple statistical conditions on the unit cell, the total amount of MC simulations can be large and require long run-times even with using high-performance computation servers. The design cycle time can be especially long when calibration is concerned or when a higher yield level is targeted with small uncertainty. For greater accuracy than behavior simulations to account for realistic device characteristics, more precise yield estimation may incorporate HSPICE MC simulation which requires substantially longer computation time. Therefore, an accurate and simple modeling of the yield can be useful to accelerate design cycle, especially one that

accounts for the effect of non-Gaussian distribution of the unit cell variation (i.e. verifiable with measured current data).

The required yield should precisely estimate the maximum bound of the INL and DNL for a given unit cell variation and a structure of the DAC. The INL yield models are well studied but no literature established a unified model for arbitrary structure. The DNL yield is overshadowed by the importance of the INL such that not a single complete study has been carried over its yield model. The next two sections briefly review the existing INL and DNL models.

2.3.1. Existing Models on INL Yield

In general, with most of the existing models, the ratio between the standard deviation and the mean of the least significant bit (LSB) current is primary piece of information used to estimate the INL yield of the current-steering DAC. The yield is defined by a decision boundary (A, maximum tolerable INL) for each input level, which is typically set at ±0.5 LSBs.

An early, commonly cited, INL yield model for a DAC [46] is based on statistical independence between different digital codes. The predicted yield from such a model has been shown to be pessimistic and hence has been modified by considering only most-probable error cases [47]. However, the modified method estimated errors on being overly optimistic. These two simple approaches can be considered as bounds and reveal the complexity of estimating INL yield. For example, using (Eq. 2.4), the required mathematics for calculating INL yield of an N-bit thermometer-coded DAC can be expressed as the following.

$$
\begin{aligned}
Yield_{INL} &= P\left(\forall i : \left|INL_{LSB}(i)\right| < A\right), 0 \le i \le 2^N - 1 \\
&= \int_{-A}^{A} \cdots \iiint P\left(INL_{LSB}(0), ..., INL_{LSB}(2^N - 1)\right) d\overline{INL}
\end{aligned}
\qquad \text{(Eq 2.6)}
$$

Note that the INL_{LSB} terms in (Eq. 2.6) are correlated as seen from (Eq. 2.4). With increasing number of codes, the order of joint Gaussian probability density function (PDF) of the INL increases exponentially; hence fully describing the random behavior of the INL at every digital code is not a practical solution.

An intuitive model has been suggested in [48] by introducing a fictive 2^N-th code. However, this model fails to account for gain error. As a result, the model generally does not match to the INL yield curve obtained from MC simulations. A Z-table-based empirical method given by [49-50] gives the most accurate estimation of INL yield. While the Z-table is invariant to various DAC resolution, it is not an efficient way to express the yield since a table entry is needed for every combination of number of bits, LSB statistics, segmentation structure, and decision boundary. Finally, a Brownian-Bridge-based analysis [51] is recently introduced and demonstrates excellent accuracy for a high resolution thermometer DAC. However, the model loses accuracy as the number of bits decreases below 8 where the INL depends on the discrete random variables of each unit cell rather than a continuous Wiener process. For the same reason, the accuracy degrades for a binary DAC or a segmented DAC.

The next chapter extends the yield estimation concept upon the previous models by introducing an intuitive INL expression with the number of bits, the unit cells' statistical properties, and the decision boundary as parameters. An empirically-fitted model based on the expression can accurately fit yield data for thermometer, binary, and arbitrarily segmented DACs.

2.3.2.   Existing Models on DNL Yield

Compared to the previously published papers on INL characteristics caused by device mismatch, an analytical model for statistically estimating the DNL yield has not been formulated even though the DNL yield may demand a more precise matching of the unit cell (e.g. in a binary DAC). A few empirical models exist based on Z-table [49] or

regression model [50]. Even published analytical models do not completely relate a target yield to the required device variation [52].

Section 2.4   Summary

This chapter discussed the random error caused by device mismatch in current-steering DAC. DNL and INL are the two key expressions of the impact of the random error. The DAC yield model can be formulated by the maximum bound of DNL/INL. However, there is a notable lack of a similar model for DNL. The limitations of the existing INL yield models lead to a low accuracy (> 7% device area estimation error, Chapter 3). The next chapter proposes a simple yet logical approach to fully address both INL and DNL of any arbitrarily segmented DAC. The function results in substantial improvement in accuracy especially in low yield situations. This low yield situation can be particularly important because of lot-to-lot yield variation (Chapter 3).

Chapter 3

DAC's Linearity Yield Model Caused by Random Mismatch

(Eq. 2.3) shows that the DNL terms are multivariate Gaussian random variables and each term is correlated. The same argument can be made for the INL terms as seen in (Eq. 2.4). From (Eq. 2.6), the PDF of the correlated random variables should be integrated. A simplified way to carry out such complicated integration is proposed in this section. An intuitive yield model for the multivariate Gaussian random variables is first proposed and analyzed. This model is then extended to precisely approximate the general DNL/INL yield models.

Section 3.1   Multivariate Gaussian Random Variables

From textbook statistics [54, 57], the joint PDF of two normal random variables can be expressed as the term inside the integral of (Eq. 3.1). A probability of success (PC) can be found by integrating the PDF within an interval [-A, A].

$$P_C = \int_{-A}^{A} \int_{-A}^{A} \frac{\exp\left(\frac{-1}{2\left(1-\rho^2\right)} \cdot \left(\frac{x^2}{\sigma_x^2} - 2\rho\frac{xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2}\right)\right)}{2\pi \cdot \sigma_x \sigma_y \sqrt{1-\rho^2}} dxdy \qquad \text{(Eq 3.1)}$$

The shape of the PDF depends on the correlation coefficient, rho, a value between 0 (uncorrelated) and 1 (fully correlated to form a single variable). Figure 5 illustrates the PDF for an intermediate value.

(a)   (b)
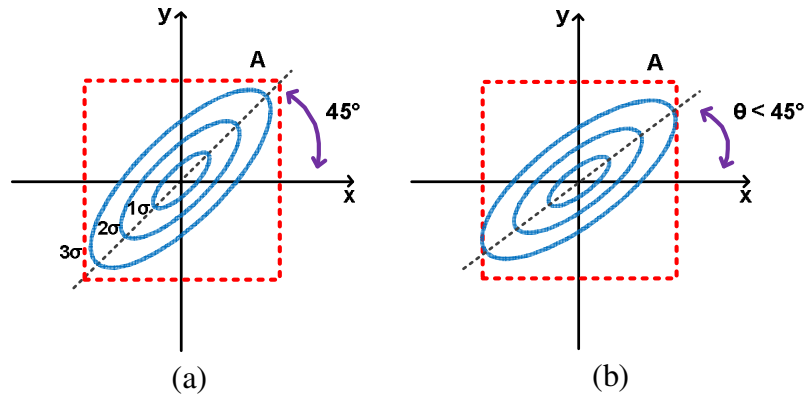
Figure 5.   Probability density function (pdf) of 2D Gaussian random variables (solid lines) and decision boundary (dotted square).

In Figure 5, the decision boundary is placed near the tails of the distribution: (a) when $\sigma_x = \sigma_y$ and the correlation coefficient is between 0 and 1, and (b) when $\sigma_x > \sigma_y$. Instead of solving (Eq. 3.1), for any correlation coefficient, we consider two observations/scenarios that simplify the model. Both can be illustrated with a two-variable PDF. In the first scenario, if σ_x and σ_y are close in value, the equation can be approximated by an error function with a fractional power between 1 and 2 as shown in (Eq. 3.2). This approximation is valid with infrequent error event. The validity of the approximation in (Eq. 3.2) as the basis for our model is discussed in the Appendix.

$$P_C \approx \left[erf\left(\frac{A}{\sqrt{2}\cdot\sigma_{max}} \right) \right]^k \qquad 1 \le k \le 2$$

$$\sigma_{max} = max\left(\sigma_x,\sigma_y\right)$$

(Eq 3.2)

Numerical examples of (Eq. 3.2) and 10,000 MC simulation results are plotted in Figure 6 to show that the approximation is accurate with the proper choice of k especially when $\sigma_x = \sigma_y$. As the correlation coefficients increase, the fitting parameter moves from 2 to 1.
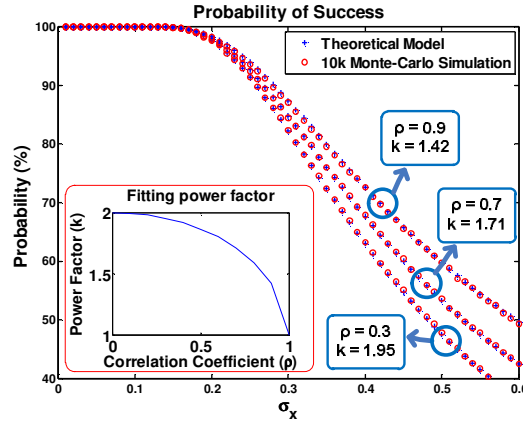
Figure 6.   Probability of success for two random variables with the decision boundary placed at ±0.5.

In the second scenario, if σ_x is several times larger than σ_y (or vice versa), the error events (for a given decision boundary where P_C is high) are usually determined by the random variable with the larger variance, regardless of the correlation coefficient. Figure 7 shows two examples where the decision boundary is placed at ±0.5 and the correlation coefficient is fixed by 0.5.

In the first case, two variables have an identical variance and the fitting power factor is 1.87. In the second case, x has a four times larger standard deviation than y. As x has a greater contribution to the error probability especially when the error rate is low, this case

is approximated by a single random variable (i.e. k = 1). The two observations can be used to simplify higher order multivariate Gaussian random variables. As discussed in the second scenario, only dominant sources of errors need to be included in the model. Using this simplified yield approximation, the DNL and INL yield can be easily formulated.
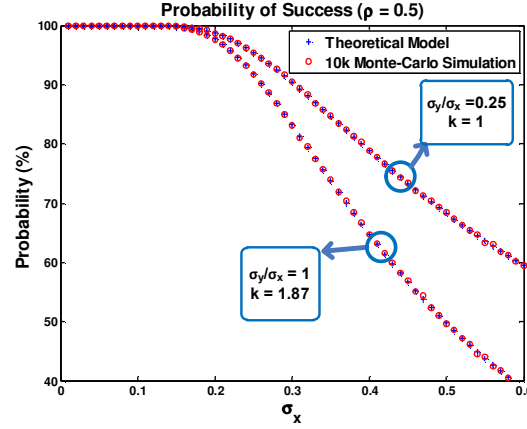
Figure 7.   Probability of success with change of variances.

Section 3.2   DNL Yield Formulation

(Eq. 2.3) shows that the DNL of each thermometer group appears once at a single code, as the input code sweeps from 0 to full-scale, while the binary group repeatedly shows up as a combination with other groups. Figure 8 shows the essential test input codes for the DNL yield estimation for an N-bit segmented DAC with the number of binary bits (NB) = N-2.

Other codes are redundant for the DNL test as identical current sources repetitively appear due to binary switching [53]. In Figure 9, 10^5 behavioral MC simulation runs are performed for characterizing the standard deviation of the DNLs of a 10-bit DAC with the last 8 bits assigned to the binary group.

Except for the three maximum peaks by the thermometer group switching, the DNLs at each lower peak level have essentially identical expressions. Due to periodic switching of the binary groups at each code (e.g. all the binary groups appear at 2^{N-2}, $2\times2^{N-2}$, and $3\times2^{N-2}$), the DNL yield calculation involves analysis on the correlated error sources. From an analysis of the DNL correlation, the next section describes simple DNL yield models that effectively reduce the integration to a single dimension for thermometer, binary, and arbitrarily segmented DAC.

Code	I_1	I_2, I_3		$I_{2^{N-3}} \sim I_{2^{N-2}-1}$	$I_{2^{N-2}} \sim I_{2^{N-1}-1}$	$I_{2^{N-1}} \sim I_{3\times2^{N-2}-1}$	$I_{3\times2^{N-2}} \sim I_{2^N-1}$	
1	⊕							
2	⊖	⊕						Weak ρ
...			...					
2^{N-3}	⊖	⊖		⊕				
2^{N-2}	⊖	⊖		⊖	⊕			
$2\times2^{N-2}$	⊖	⊖		⊖		⊕		Strong ρ
$3\times2^{N-2}$	⊖	⊖		⊖			⊕	

Binary Group   Thermometer Group

⊖ = switching to -   ⊕ = switching to +
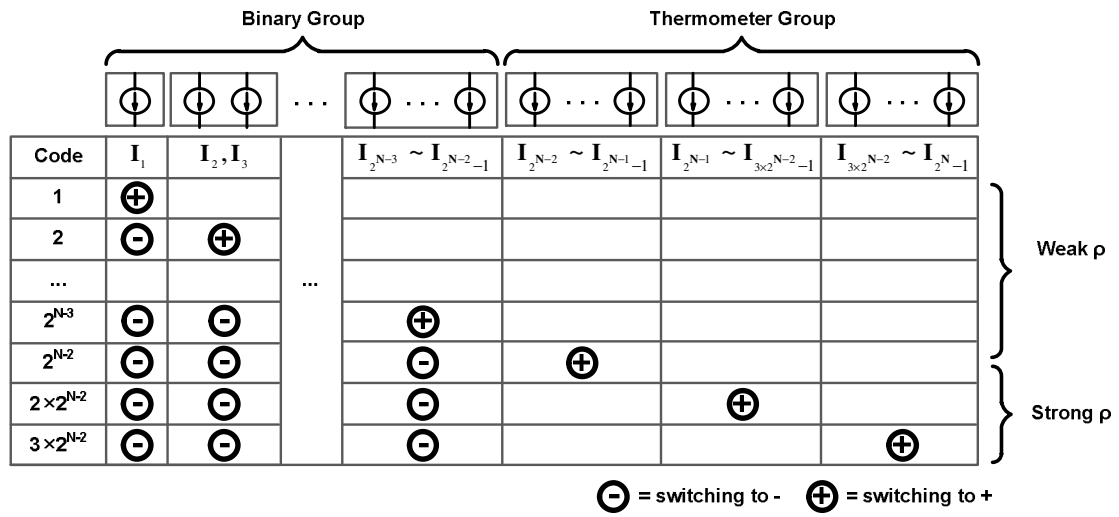
Figure 8.   Switching diagram of current sources for an N-bit segmented DAC with the last N-2 bits assigned to the binary group (NB = N-2).
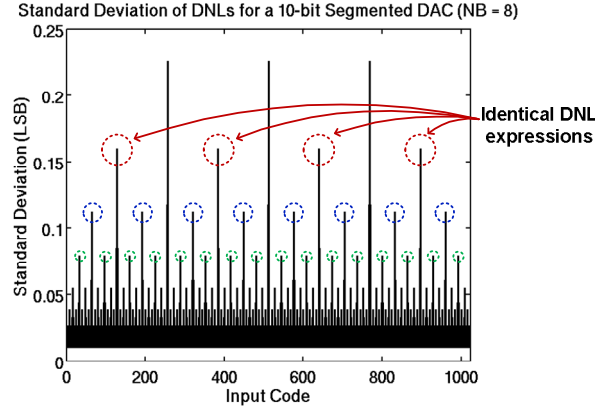
Figure 9.   Standard deviation of DNLs of a 10-bit segmented DAC with NB =
8. The sigma of ΔI is 1% of the averaged current (Ī).

3.2.1.   **Thermometer-Coded DAC**

The DNL of a thermometer-coded DAC at each input code is a normalized current variation of the unit cell. As is well known, the DNL yield can be simply calculated by using a product of multiple Gaussian PDFs due to the statistical independence of the unit cell variation. The following expression assumes that the decision boundary equals ±A.

$$DNL\ Yield = \left[erf\left(\frac{A}{\sqrt{2}} \cdot \frac{\bar{I}}{\cdot\sigma_I} \right) \right]^{2^N - 1} \qquad \text{(Eq 3.3)}$$

The erf term in (Eq. 3.3) is very close to 1 unless the normalized standard variation of the unit cell ($\sigma_I/\bar{I}$) is comparable to A. For this reason, a thermometer DAC shows a higher DNL yield than any other structure with a given $\sigma_I/\bar{I}$.

3.2.2.   **Binary-Coded DAC**

The DNL yield of a binary DAC can be approached in a similar manner. There are N binary current groups. Testing of the DNL is conducted at N input codes where each

binary group switches at least once. The essential DNL test codes are as follows (without losing generality, $\bar{I} = 1$)

$$DNL(2^{N-1}) = \sum_{i=2^{N-1}}^{2^N-1} \Delta I_i - \sum_{j=1}^{2^{N-1}-1} \Delta I_j$$

$$DNL(2^{N-2}) = \sum_{i=2^{N-2}}^{2^{N-1}-1} \Delta I_i - \sum_{j=1}^{2^{N-2}-1} \Delta I_j$$

$$\dots$$

$$DNL(1) = \Delta I_1$$

(Eq 3.4)

Any two DNL terms in (Eq. 3.4) share identical current groups implying a correlation (e.g. $\Delta I_1 \sim \Delta I_{15}$ in DNL(8) and DNL(16)). However, these current sources are subtractive in one DNL (e.g. $(\Delta I_8 + \dots + \Delta I_{15}) - (\Delta I_1 + \dots + \Delta I_7)$ for DNL(8)), and are additive for the other DNL terms (e.g. $(\Delta I_8 + \dots + \Delta I_{15}) + (\Delta I_1 + \dots + \Delta I_7)$ for DNL(16)). Therefore, the correlation coefficient between any two DNL terms can be expressed as

$$\rho_{DNL}\left(2^{n-1}, 2^{k-2}\right) = \frac{-1}{\sqrt{2^n - 1} \cdot \sqrt{2^{k-1} - 1}}$$

$$n, k = \{2, \dots, N\}, \ n \geq k$$

(Eq 3.5)

For a large n or a large k, ρ_{DNL} is sufficiently small such that the statistical correlation between two DNL terms vanishes. For a small $\rho_{DNL} < 0.2$, the two random variables can be assumed as nearly statistically independent when a yield calculation is concerned as seen from (Eq. 3.2) and Figure 6. Although there still exists a strong correlation between LSB groups such as $\rho_{DNL}(1, 2)$, their contributions on the DNL yield can be ignored especially for a high yield level (> 90%). By approximating statistical independence of the primary sources (i.e. MSB terms in (Eq. 3.4)) of the DNL errors, the DNL yield of a binary DAC is expressed as products of error functions.

$$DNL\ Yield \approx \prod_{i=1}^{N} erf\left(\frac{A}{\sqrt{2}} \cdot \frac{\overline{I}}{\sqrt{2^i - 1} \cdot \sigma_I} \right)$$

(Eq 3.6)
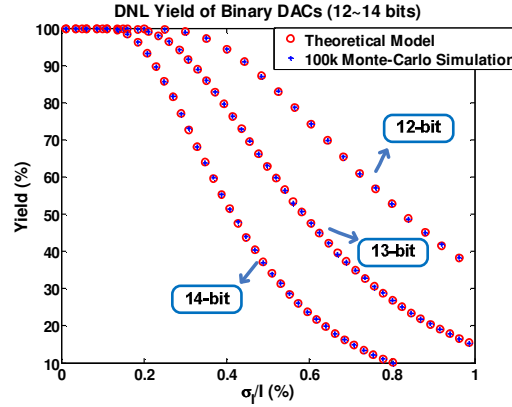
Figure 10.   DNL yield plots of the theoretical model and 100k behavioral MC

simulation results for 12~14-bit binary DACs. (A = ±0.5LSB).

Figure 10 shows the DNL yield plots of 12~14-bit resolution binary DAC with the analytical model (Eq. 3.6). The precise matching of the model and the 100k behavioral MC simulation results proves the validity of the assumption on the statistical independence between the major DNL terms in (Eq. 3.4). Impact of the correlated error sources between LSB terms in (Eq. 3.4) can be observed in Figure 11.
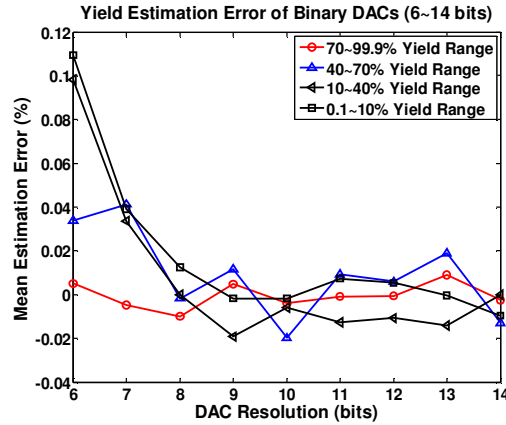
Figure 11.   Yield estimation error for 6~14-bit binary DACs. Errors are
averaged over 70~99.9%, 40~70%, 10~40%, and 0.1~10% yield ranges.

This figure shows the average yield estimation error between the simulated yield and the model yield for 6~14 bit binary DACs. For any resolution, the MSB DNL term (2^{N-1}) contributes to most of the error events when the unit element has a small variation (i.e. a high yield level > 95%). For a slightly higher device variation, DNL(2^{N-2}) can exclusively contribute to the error event although this case is infrequent. The same argument can be applied sequentially to less significant DNL terms. For this reason, our model shows excellent accuracy for estimating high yield levels > 70% where only MSB DNL terms (ρ < 0.2) contribute to the majority of the error events. For other yield ranges, the DNL yield is generally underestimated especially for a low resolution DAC as correlated LSB DNL terms may have more chances to add to the number of errors.

3.2.3.   Segmented DAC

The approximation used for the binary DAC and the resulting model in (Eq. 3.6) can be extended to explain the yield of a segmented DAC. For an N-bit segmented DAC with NB (number of binary bits), the most probable DNL error occurs when all the binary groups and one of the thermometer groups switch. For an entire input range, there are ($2^{(N-NB)}$-1) worst-case transitions. From the same viewpoint of (Eq. 3.6), a statistical independence is assumed between the binary transitions and any one of these worst-case transitions (Figure 8). Therefore, the DNL yield can be approximated as a multiplication of the yield of the binary group (P_{NB}) and the yield of the worst-case thermometer transitions (P_W).

$$DNL\ Yield \approx P_{NB} \times P_W$$

(Eq 3.7)

P_{NB} can be expressed as (Eq. 3.6) with replacing N by NB. P_W requires a 1-D integration of a conditional Gaussian PDF. To illustrate the calculation using the example of Figure 8, there are three input codes where each thermometer group switches. (Eq. 3.8) expresses the probability of the worst-case DNL being bounded by ±A.

$$
\begin{aligned}
P_W &= P\left(\forall k : \left| \Delta I_{Tk} - \sum_{i=1}^{2^{N-2}-1} \Delta I_i \right| < A \right) \\
&= \int_{-A}^{A} P_{Binary} \cdot P\left(\left| \Delta I_{T1} - x \right| < A \right) \cdot \\
&\quad P\left(\left| \Delta I_{T2} - x \right| < A \right) \cdot P\left(\left| \Delta I_{T3} - x \right| < A \right) dx \\
P_{Binary} &= P\left(\sum_{i=1}^{2^{N-2}-1} \Delta I_i = x \right)
\end{aligned}
$$

(Eq 3.8)

The contribution of the binary group's variation to the worst-case DNL is expressed in equation (Eq. 3.8) as x. The contribution of each of the three largest code transitions corresponds to the latter three terms of the equation. The difference, $\Delta I_{Tk} - x$ where ΔI_{Tk} is defined in (Eq. 2.2), is the DNL contribution of an entire group of binary bits switch from ON to OFF while a thermometer bit switches ON. The integration bound of x in (Eq. 3.8) is taken from the equivalent bounding condition on $DNL(2^{N-3})$ in (Eq. 3.4). Note that x is not necessarily bounded by ±A, but this simplification for high yield targets leads to a negligible estimation error. By substituting (Eq. 3.6) and (Eq. 3.8) to (Eq. 3.7), the DNL yield of a segmented DAC (only for NB = 1 ~ N-2) is expressed as

$$DNL\ Yield = \left\{ \prod_{i=1}^{NB} erf\left(\frac{A}{\sqrt{2}} \cdot \frac{\overline{I}}{\sqrt{2^i - 1} \cdot \sigma_I} \right) \right\} \times$$

$$\left\{ \int_{-A}^{A} \frac{1}{\sqrt{2\pi} \cdot \sigma_B} \cdot \exp\left(-\frac{x^2}{2 \cdot \sigma_B^2} \right) \times \right.$$

$$\left. \left[\left| \frac{1}{2} \cdot erf\left(\frac{x + A}{\sqrt{2} \cdot \sigma_T} \right) \right| + \left| \frac{1}{2} \cdot erf\left(\frac{x - A}{\sqrt{2} \cdot \sigma_T} \right) \right| \right]^{2^{N-NB}-1} dx \right\}$$

(Eq 3.9)

In this equation, σ_B represents the normalized standard deviation of the current sum of all binary groups while σ_T stands for the normalized current variation of a thermometer group. Using $\sigma_I/\overline{I}$, σ_B and σ_T are expressed as (Eq. 3.10).

$$\sigma_B = \sqrt{\left(2^{NB} - 1\right)} \cdot \frac{\sigma_I}{\overline{I}}, \quad \sigma_T = \sqrt{2^{NB}} \cdot \frac{\sigma_I}{\overline{I}}$$

(Eq 3.10)

(Eq. 3.9) contains a 1-D integration that can be easily implemented in any mathematical analysis tool (e.g. MATLAB). In Figure 12, the model accuracy is demonstrated for 8~14-bit segmented DACs. As two extreme cases, two segmentation structures are given for NB = 1 and NB = N-2. The model (Eq. 3.9) matches to the simulation results very well.

(a)   (b)
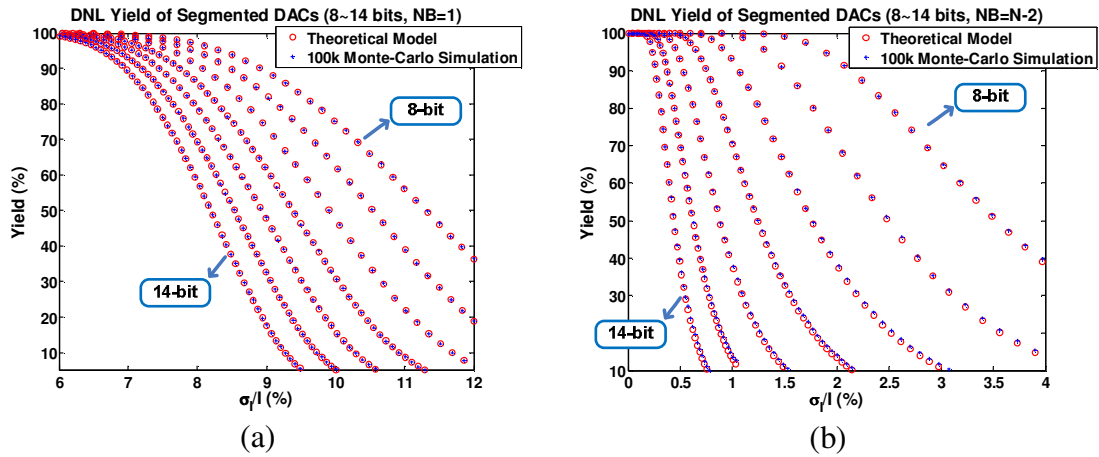
Figure 12.   DNL yield plots of the model and 100k behavioral MC simulation results for 8~14 bit segmented DACs with (a) NB = 1, and (b) NB = N-2.

Section 3.3   INL Yield Expression

The INL yield model can be analyzed in a similar way to the DNL yield formulation. However, the expression is much more complicated than (Eq. 3.9) as the INL terms do not show such nice correlation distribution between thermometer groups. As a reference case, the INL correlation characteristics of the thermometer, binary, and segmented structures are analyzed. Based on the discussions made in section 4.2, two simple models are suggested per model complexity.

3.3.1.   INL Yield Estimation of a Thermometer-Coded DAC

(Eq. 3.2) and the two scenarios approximated the yield calculation based on a single random variable having the maximum variance. The same argument can be applied to the INL. From (Eq. 2.4), the uncertainty of the INL is maximized at the middle of the input range, where squared sum of the coefficients in the numerator is maximized, and then it gradually diminishes to zero as the input code moves away from the middle. Note that this argument is not dependent on the structure of the DAC.

$$MAX\left(\sigma_{INL}^2\right) = 2^{N-1} \cdot \left(1 - \frac{2^{N-1}}{2^N - 1}\right) \cdot \frac{\sigma_I^2}{\overline{I}^2}$$

(Eq 3.11)

As for initial approach, the maximum INL is used to approximate the yield of the INL. For this purpose, the correlation between the maximum and all other INL terms should be investigated. As an example of 6-bit DACs, Figure 13(a) shows the INL's correlation coefficients between each input code and the middle code (i.e. maximum variance) for the thermometer, binary, and segmented structure (NB = 3).

(a)

(b)
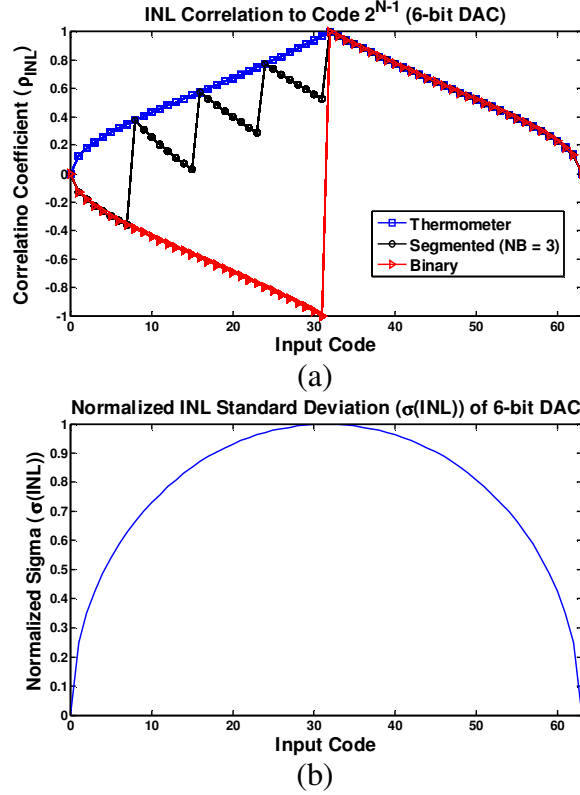
Figure 13.   INL characteristics of 6-bit segmented DACs. (a) INL correlation

coefficient between the entire input code (x-axis) and code 2^{N-1} (middle). (b) INL

variance at each input code.

   The coefficient (ρ_{INL}) is high around the middle, and it gradually diminishes to zero as the input code moves away from the middle. Similarly, Figure 13(b) shows that the variance of the INL has a flat region around the middle and it rapidly diminishes at the ends. Focused on the high yield level estimation, most of the error events are contributed from the codes near the middle. Therefore, as shown in the first scenario in Figure 6, the overall yield level can be found by adjusting the power factor k in (Eq. 3.2) where the standard deviation is found at the middle of the input range (Eq. 3.11). If the decision boundary is set to be ±A (in LSB), the INL yield of a thermometer DAC can be expressed as (Eq. 3.12).

31

$$INL\ Yield \approx \left[\frac{1}{\sqrt{2\pi} \cdot \sigma_{MAX}} \cdot \int_{-A}^{A} e^{-\frac{x^2}{2\sigma_{MAX}^2}} dx \right]^k \approx \left[erf\left(\frac{A \cdot \overline{I}/\sigma_I}{\sqrt{2^{N-1}}} \right) \right]^k \qquad \text{(Eq 3.12)}$$

The power factor k is an empirically-fitted parameter to minimize estimation error (e.g. minimum sum of squared error) between the model (Eq. 3.12) and the MC simulation data. Since the number of elements near the middle and their correlation depends on the number of current sources, the fitting parameter is a function of the number of bits. From Figure 13(a), adding 1 more LSB to the N-bit DAC (equivalently N+1 bits) introduces new error sources but the overall shape of ρ_{INL} does not change. These new error sources must be highly correlated to the existing N-bit codes as ρ_{INL} varies smoothly. Hence applying the scenario 1 in Figure 6 indicates that the fitting parameter in (Eq. 3.12) is eventually saturate as the number of bits increases. Figure 14 shows an example of a 6-bit thermometer DAC. The fitting factor is 4.9. The plot shows that the theoretical model fits well for the yield level >50%. As (Eq. 3.12) is valid for rare failure events, the theoretical model also loses accuracy for estimating low yield level.
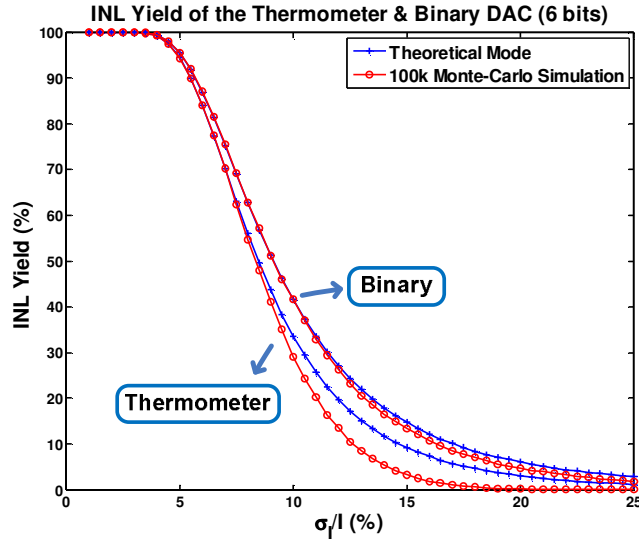
Figure 14.   INL yield plots of the theoretical model and 100k behavioral MC

simulation results for a 6 bit thermometer and binary DAC.

3.3.2.   INL Yield Estimation of a Binary-Coded DAC

While each unit current source in the thermometer-coded DAC has same contribution to the INL, a binary-coded DAC would have each bit weight contributing differently to the INL. In Figure 13(a), compared to the thermometer structure, the correlation coefficient shows anti-symmetry at the center of the input code, where the MSB group switches ($\rho_{INL} = \pm 1$). The identical behavior can be observed from (Eq. 2.4) that the INL itself is anti-symmetry near the middle of the code for the binary-coded DAC. This (anti-)symmetry implies that for the binary-coded DAC, the equivalent INL yield test can be carried over for either left or right half of the input code. Behavioral MC simulation results also support this observation: the INL test range is varied from [0, 2^{N-1}] to [2^{N-1}, 2^N-1] but the yield level remains roughly constant. Due to this strong correlation of the INL distribution over the input code, the power factor of a binary DAC is usually smaller than the power factor of a thermometer DAC for a given resolution bit.

INL yield of a 6-bit binary DAC is displayed in Figure 14 with k factor of 3.8. The theoretical model shows a higher accuracy for the entire range of yield. Since the MSB group has the largest variance and most of the errors are caused by this group, the simplification to a fewer equivalent random variable fits better for the binary structure.

3.3.3.   INL Yield Estimation of a Segmented DAC

The error in a thermometer (or binary) DAC can be attributed to approximating high correlation to a dominant error source. This approximation allowed us to use a single term in (Eq. 3.12). A segmented DAC with a few binary-coded LSBs can be expected to exhibit the same behavior. For a fixed total number of bits, the segmented DAC is known to have the worst yield compared to a fully thermometer or a fully-binary structure [50]. Figure 15 shows an example of a 14-bit DAC with the last 12 LSBs assigned to the

binary group. The yield curve of MC simulations starts to drop sharply, and the analytical model of (Eq. 3.12) does not track such transition even when the curve fitting is focused on the high yield levels > 50%.

This discrepancy can be clearly visualized in Figure 13(a). From the center of the input code, only right half region $[2^{N-1}, 2^{N}-1]$ shows similar ρ_{INL} distribution as the thermometer's case. From MC simulation results, the model of (Eq. 3.12) fits precisely for high yield level > 50% in any arbitrary segmented DAC if the INL test is confined to the right half of the input code. For the left half of the input code in Figure 13(a), the segmented DAC shows periodical jumps in ρ_{INL} distribution. Such jumps indicate that more error sources are less-correlated and the INL yield degrades more than the other two structures (unary and binary), which makes the simplification of a single term model in (Eq. 3.12) less effective.

To extend the model to include a segmented DAC, the left and right side of ρ_{INL} are separately modeled as multiplication terms. Since section 3.2 shows that the error function can effectively model a subset of the random variables (i.e. the dominant ones), multiple error functions can be employed to model the different sections in Figure 13(a).

INL Yield & Estimation Error of a 14-bit Segmented DAC
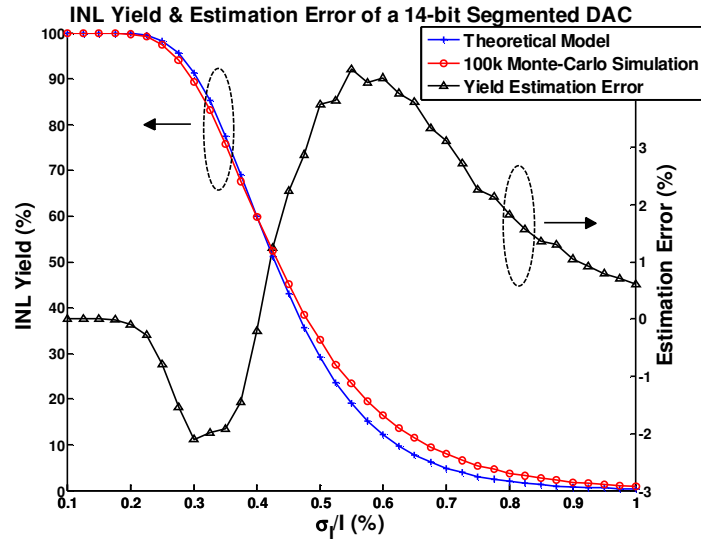
Figure 15.   INL yield curve and estimation error of a 14-bit segmented DAC
with the last 12 LSBs assigned to the binary group.

(Eq. 3.13) shows the extended model that uses a multiplication of multiple (n) error functions to fit the yield curve. The decision boundary for the yield is set to be ±A (in LSB).

$$INL\ Yield \approx \prod_{i=1}^{n} \left[erf\left(\frac{A}{\sqrt{2}} \cdot \frac{\overline{I}}{\sigma_I} \cdot \frac{1}{C(x_i)} \right) \right]^{k(i)} \qquad \text{(Eq 3.13)}$$

$$C_1(x) = \sqrt{2^{N-2} \cdot 1.2}$$

$$C_2(x) = \sqrt{2^{N-1} \cdot \left(1 - \frac{2^{N-1}}{2^N}\right)}$$

$$C_3(x) = \sqrt{2^{N-3} \cdot \left(1 - \frac{2^{N-3}}{2^N}\right)}$$

$$C_4(x) = \sqrt{2^{N-4} \cdot \left(1 - \frac{2^{N-4}}{2^N}\right)}$$

(Eq 3.14)

Note that error functions in (Eq. 3.13) can be replaced by exponential expressions [55-56] for further simplification. The number of terms can be arbitrarily large though four terms in (Eq. 3.14) show extremely precise fitting results for any segmented DACs. From (Eq. 3.14), C_2 is from the right half of the input code (Eq. 3.12) while C_1 has a slightly larger value than C_2. As explained in Appendix, when the two random variables have different variances, (Eq. 3.2) can be modified to fit better by adjusting σ_{max} with an additional fitting parameter $c \times \sigma_{max}$. Combining C_1 and C_2 enhances flexibility of the model to fit the initial curvature of the yield at a high level > 90%. The last two terms are from input code 2^{N-3} and 2^{N-4}, respectively. As the value of ρ_{INL} in Figure 13(a) drops below 0.35 at those input codes, their error contributions can be modeled as separate error function multiplied to C_1 and C_2 terms (Figure 6). All the four terms are used to fit the yield curve for the entire region with minimum sum of squared errors.

Examples of this new model are plotted in Figure 16. Each plot shows three lines: a complete model using all four terms in (Eq. 3.13), a model-order reduction by ignoring the last two terms in (Eq. 3.13), and 100k behavioral MC simulation results. The two-term model accurately tracks the simulation results for the yield   level >70%. As mentioned previously, the order of the model depends on the DAC architecture. A single-term model in (Eq. 3.12) can reliably estimate the yield > 70% for binary-coded and most thermometer-coded DACs.
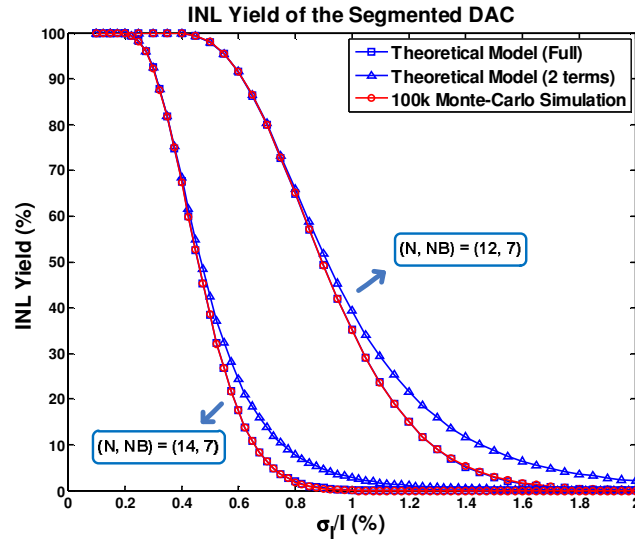
Figure 16.   INL yield plots of the segmented DACs for 12 and 14 bits (NB = 7).

3.3.4.   **Single-Term Model Simulation Results**

The simulation results of the single-term yield model in (Eq. 3.12) are plotted in Figure 17 for thermometer DACs and binary DACs, respectively. The resolution of the DACs is swept from 3-bit to 14-bit with the decision boundary placed at ±0.5 LSBs.

For each DAC resolution, the single-term yield model is listed the corresponding fitting parameter k (in (Eq. 3.12)). The fitting parameter k is found to minimize the sum of errors for the yield level above 70%, and the results are summarized in Table 3-1.

Table 3-1 Resolution bits and the required power factor in (Eq. 3.12)

Bits (N)	Power (k)		Bits (N)	Power (k)	
	Unary	Binary		Unary	Binary
3	2.0	1.3	9	6.0	5.23
4	2.97	2.05	10	6.26	5.51
5	3.86	2.87	11	6.4	5.75
6	4.63	3.65	12	6.5	5.9
7	5.2	4.3	13	6.6	6.06
8	5.7	4.83	14	6.6	6.08

(a)   (b)
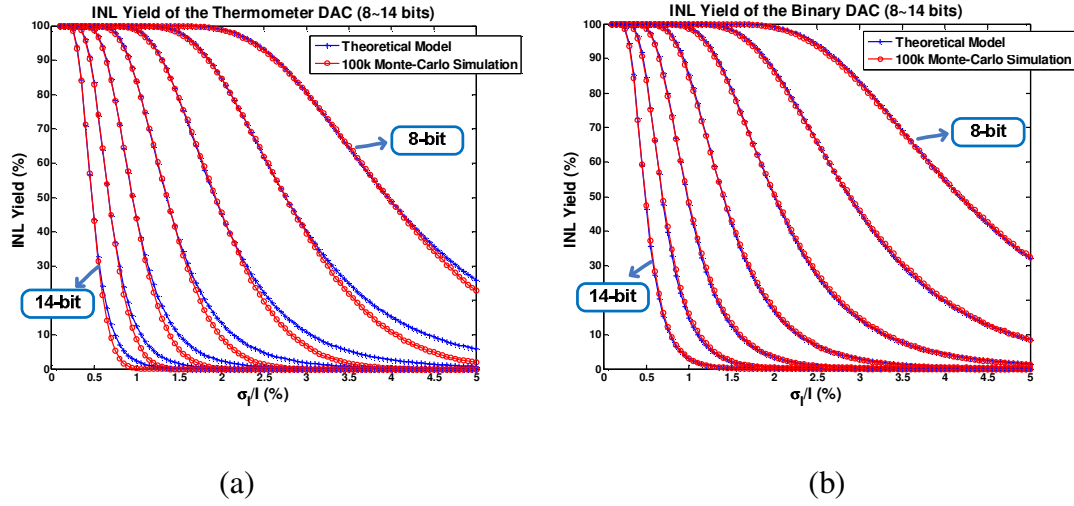
Figure 17.   INL yield plots of the theoretical model and 100k behavioral MC

simulation results for 8~14-bit (a) thermometer DAC and (b) binary DAC.
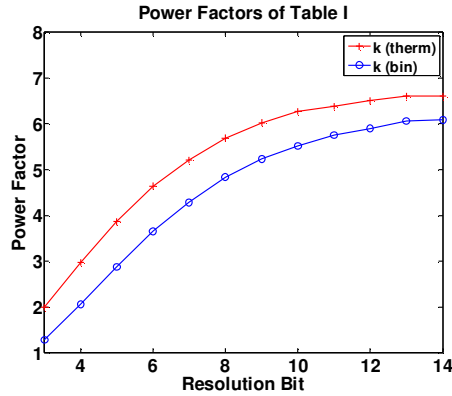
Figure 18.   Power factor k for thermometer DAC and binary DAC as a function

of the number of bits of resolution.

As expected in section 3.4.1, the power factor converges as the number of bits increases
(Figure 18). For a low resolution thermometer DAC (3~7 bits), the k factor varies rapidly
from 2.0 to 5.2. However as the number of bits exceed 8, the parameter gradually
converges to 6.6. This convergence is because at higher resolution, the added elements
are closely correlated in terms of the INL which does not increase the power factor. This

result validates similar claims [49] that the INL yield of a thermometer DAC or a binary DAC does not vary if the normalized current variation (σ_I/I) is suppressed by half for every two bit increase in resolution.

3.3.5.   Multi-Term Model Simulation Results

This section provides the complete table of the fitting parameters used in the proposed INL yield models. As the parameters are structure dependent, behavioral MC simulations are carried over for every available combination of thermometer and binary structure. The simulation is limited up to a 14-bit DAC due to convergence of the parameters. The results of the fitting parameters for the multiple-term model are presented in this section. As indicated in Section 3.4.3, the last two terms in (Eq. 3.13) do not need to be included in using the model when only estimating for high yield. The parameters (k1~k4) are fit using least square fitting across the entire yield curve. Table 3-2 shows all the parameters including fully thermometer-coded (NB = 0) and fully binary-coded (NB = N-1) structures.

Table 3-2 Resolution bits and the required power factor in (Eq. 3.13)

Bits		Power Factor				Bits		Power Factor			
N	NB	k1	k2	k3	k4	N	NB	k1	k2	k3	k4
6	0	0.53	3.7	1.7	8.3	11	5	3.03	2.63	5.2	12.4
	1	0.75	3.7	2.3	7.4		6	2.89	3.45	4.1	16.3
	2	0.77	4.13	2.5	5.1		7	3.44	3.38	4.6	8.2
	3	0.72	4.9	0.9	3.5		8	3.55	4.29	0.3	12.2
	4	1.47	3.73	0.0	0.0		9	4.7	2.19	0.0	0.0
	5	0.03	3.6	0.2	0.9		10	1.67	3.39	0.0	2.7
7	0	1.1	3.4	2.4	9.8	12	0	2.11	3.12	2.9	16.0
	1	0.9	4.05	1.8	10.9		1	1.89	3.58	2.4	17.0
	2	1.33	3.76	2.8	8.6		2	1.97	3.57	2.3	17.6
	3	1.54	4.0	3.0	5.8		3	2.26	3.23	3.2	15.7
	4	1.69	4.5	1.3	4.1		4	2.41	3.18	3.3	17.7
	5	2.5	3.2	0.0	0.0		5	2.85	2.78	4.3	15.7
	6	0.38	3.77	0.0	1.5		6	2.95	3.03	4.0	19.4

Group					
8	0	1.55	3.13	3.0	10.6
	1	1.39	3.66	2.3	12.0
	2	2.04	2.97	3.8	9.4
	3	2.29	3.0	4.4	8.4
	4	2.03	4.15	2.8	8.6
	5	2.51	4.17	1.9	2.3
	6	3.21	2.95	0.0	0.0
	7	0.88	3.55	0.4	0.9
9	0	1.54	3.48	2.7	11.9
	1	1.75	3.34	2.8	12.9
	2	1.59	3.95	1.4	19.2
	3	2.22	3.21	3.5	13.5
	4	2.56	3.2	4.0	11.8
	5	2.59	3.9	3.5	9.1
	6	2.76	4.54	0.9	5.8
	7	4.12	2.31	0.0	0.0
	8	1.19	3.54	0.0	2.0
10	0	1.96	3.02	3.6	11.2
	1	2.07	3.02	3.4	13.2
	2	1.85	3.63	2.4	15.9
	3	2.33	3.08	3.6	14.1
	4	2.71	2.8	4.9	11.9
	5	2.98	3.04	4.5	11.7
	6	3.24	3.33	4.4	8.1
	7	3.28	4.36	0.0	11.7
	8	4.48	2.21	0.0	0.0
	9	1.45	3.49	0.0	2.1
11	0	1.99	3.19	3.1	13.6
	1	2.05	3.15	3.5	12.9
	2	2.18	3.1	3.7	12.9
	3	2.3	3.11	3.6	15.1
	4	2.24	3.59	2.5	19.7

Group					
	7	3.51	2.75	5.1	15.1
	8	3.73	3.26	4.2	12.1
	9	3.69	4.34	0.4	11.1
	10	5.13	1.84	0.0	0.0
	11	1.97	3.11	0.3	2.3
13	0	2.35	2.77	3.8	13.5
	1	2.2	3.17	2.8	16.5
	2	2.28	3.17	2.3	19.6
	3	2.47	2.8	4.5	12.1
	4	2.52	2.92	3.7	16.2
	5	2.72	2.94	2.7	22.2
	6	2.66	3.35	2.6	22.1
	7	3.25	2.71	4.9	15.5
	8	3	3.74	3.9	17.7
	9	3.94	3.03	5.6	6.3
	10	4.37	3.43	1.8	8.6
	11	5.36	1.68	0.0	0.0
	12	1.97	3.2	0.0	3.1
14	0	2.4	2.8	3.8	14.0
	1	1.86	3.79	1.6	20.4
	2	2.26	3.06	3.9	13.6
	3	2.36	3	3.8	14.6
	4	2.23	3.37	2.8	18.3
	5	2.94	2.29	5.5	10.9
	6	2.87	2.63	4.9	14.2
	7	2.81	3.07	4.3	16.5
	8	3.28	2.85	4.2	19.0
	9	3.32	3.41	3.8	20.8
	10	3.91	3.16	5.8	6.6
	11	4.42	3.48	2.1	6.0
	12	5.46	1.64	0.0	0.0
	13	1.99	3.27	0.3	1.8

Similar to the single-term model, the first two power factors (k1 and k2) converge as the number of bits increases. The power factor k3 and k4 has a dependency on the number of bits assigned to the binary group (NB) and diminishes as NB is larger. This

effect is explained by the single-term model that accurately addresses the yield curve of a binary DAC for the entire range of the current variation. Figure 19 compares different yield models for a 14-bit segmented DAC (NB = 10). As seen from the figure, the multi-term model precisely matches to the MC simulation results. For other models, only [49, 51] show reasonable accuracy for high yield level > 99%. The required Z-table for [49] is generated from behavioral MC simulation results of a 10-bit thermometer DAC.

Accuracy of the models is summarized in Table 3-3 where the estimation errors of normalized current variation ($\sigma_I/\bar{I}$) are listed at different target yields. Each value in the table is   selected from the worst-case estimation error when each model is applied to 6~14 bit arbitrarily segmented DAC. The values in percentage are insensitive to the decision boundary (A).
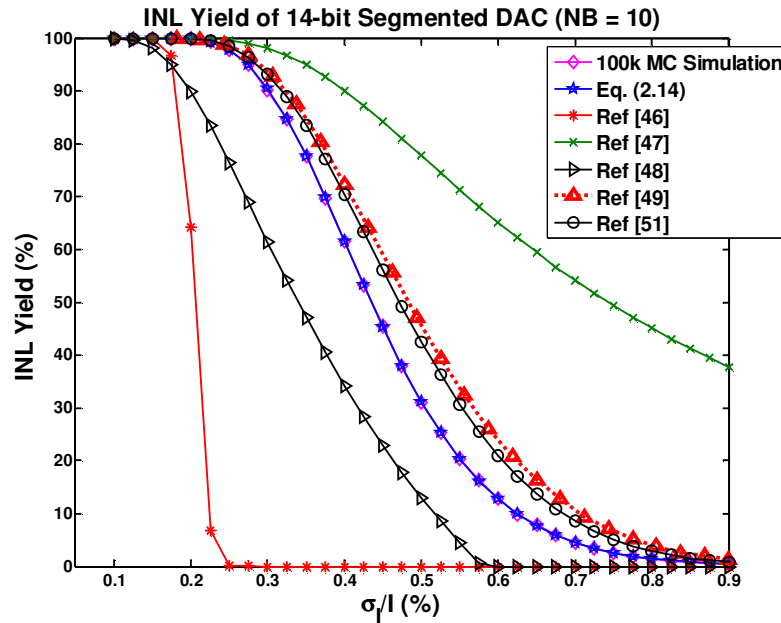
Figure 19.   Comparison of INL yield models for a 14-bit segmented DAC
(NB=10).

41

Table 3-3 The worst estimation error of the normalized current variation ($\varepsilon_{MAX}(\sigma_I/\bar{I})$) for 6~14-bit arbitrarily segmented DAC

Target Yield	[46]	[47]	[48]	[49]	[51]	(Eq. 3.13)
99%	31.3%	22.1%	46.7%	7.1%	7.3%	2.2%
95%	38.7%	29.6%	45%	8.2%	8.7%	<1%
90%	42.6%	35.2%	44%	9.4%	9.6%	<1%

In terms of geometry estimation, models in [46~48] may incorrectly estimate the area of the unit cell by more than 50%, models in [49, 51] by 15%, and (Eq. 3.13) by less than 5% when the target yield is 99%. If only high resolution (>10 bit) thermometer or binary DACs are considered, [49] and [51] produce less than 1% error. The accuracy of other models degrades as the target yield decreases to 90% while the proposed multi-term model stays in a small error bound.

Since the yield estimation must include lot-to-lot variation, the overall yield must be averaged from the expected range of the current variation. Overall yield can be greatly affected by the low-yield lots. Therefore, the yield model should maintain reasonable accuracy at moderate yield level such as > 90%.

$$INL \ Yield = \int Y\left[N, NB, \frac{\bar{I}}{\sigma_I}(x) \right] \cdot P(x) \, dx$$

Y = INL Yield at a given $\dfrac{\bar{I}}{\sigma_I}$

x = Global process corner variable

P = PDF of x

(Eq 3.15)

Section 3.4   Discussions on Current Distribution

This section provides measurement data from an 8-bit current-steering DAC fabricated in a 90-nm CMOS technology. The measured currents are used to estimate the DNL and INL yield with respect to the decision boundary. The measurement data can include

diverse effects on the current distribution such as WPE, STI-stress, geometric gradient [39-40], and device non-linearity (I-V curve). As expected, the unit cell statistical distribution is nearly but not entirely Gaussian. The impacts of device non-linearity on the INL yield are analyzed in more details using HSPICE MC simulations.
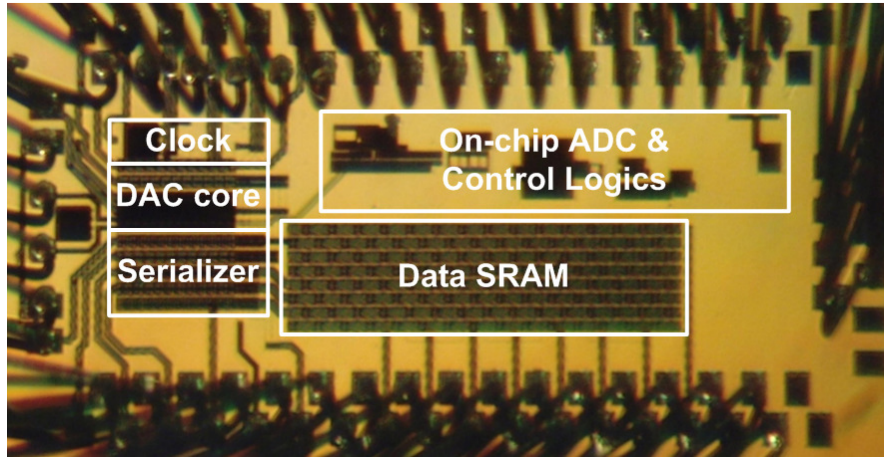
Figure 20.   Die microphotograph of an 8-bit current-steering DAC.
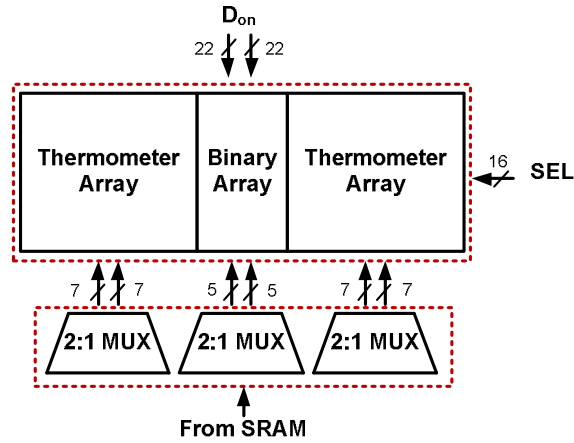
Figure 21.   Diagram of current cell distribution

3.4.1.   Yield Estimation from Measurement Data

Figure 20 shows the micrograph of the testchip. The die area is 2×1 mm^2. In Figure 21, the DAC core is comprised of unit current source arrays with seven thermometer groups and five binary groups. The twelve bit (7+5) input signals are coming from a data SRAM. Auxiliary control signals D_{on} and SEL can toggle on/off state of each current unit cell by programming its dedicated memory for biasing. Figure 22 shows the detailed schematics of the unit current source.
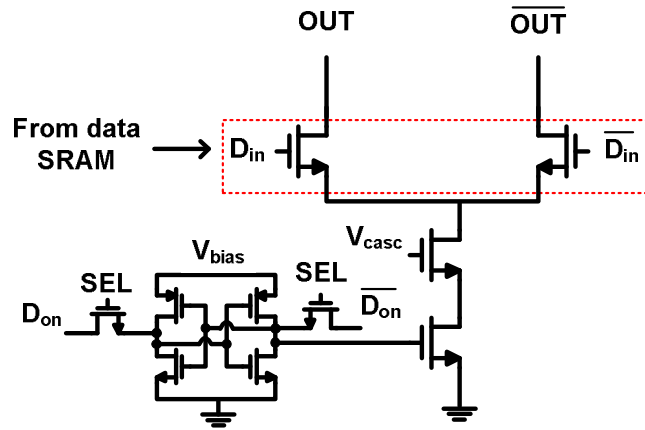
Figure 22.   Unit current cell with a dedicated memory for biasing.

There are more than 340 unit current cells in a single chip including redundancy. The testchip is designed such that the current of each unit cell can be measured independently and the DAC can be organized in arbitrary segmentation structure if the DAC output is measured statically. More than 1,000 cell currents are measured from multiple chips (three) and stored as a set of random number generator. Chip-to-chip variations are corrected by matching mean of each data group. The distribution of the measured currents is shown in Figure 23(a).

The amount of non-ideal deflection can   be visualized using a Quantile-Quantile (Q-Q) plot [58]. In the Q-Q plot of Figure 23(b), a slight deviation from the reference line indicates that the tails of the current distribution is weakly distorted by the device nonlinearity and geometric gradient. The measured normalized current variation (σ_I/I) is 9.55%. Since the current variation is a fixed quantity, the decision boundary (A) is swept to produce a yield curve varying from 0 to 100%.

As the number of current sources is limited, random data are generated from the 1,000 measurement set. Figure 23(c) shows the flow diagram of the validation methodology. A total 100k 8-bit arbitrarily segmented DACs are generated based on the measurement data. In each case, the maximum DNL and INL is compared to a variable decision boundary ranging from ±0.1 LSBs to ±4 LSBs. The DNL and INL yield curves for the analytical model and 100k Pseudo-MC simulation results are shown in Figure 22 for 8-bit segmented DACs.

Figure 24 compares the yield levels of three different structures: thermometer, binary, and segmented. The structure sensitivity on each linearity yield curve is quite different. As shown in the Figure 24(a), the DNL yield improves a lot as the structure changes from fully binary to fully thermometer. Although Figure 24(b) shows that the INL yield is relatively insensitive to the structure, the segmented DAC has the worst-case yield level, followed by the thermometer, and the binary DAC shows the best yield.

The sensitivity of the decision boundary on the yield is observed from Figure 24. As an example, the yield of the 8-bit segmented DAC with the given measurement data is 95% when the decision boundary is at ±2 LSBs. If the decision boundary is at ±2.5 LSBs, the yield level is over 99%. Since other types of nonlinearity such as finite output impedance of the unit current source may also contribute to the INL [22], a proper decision margin should be considered to maintain the target yield level.

(a)

(b)   (c)
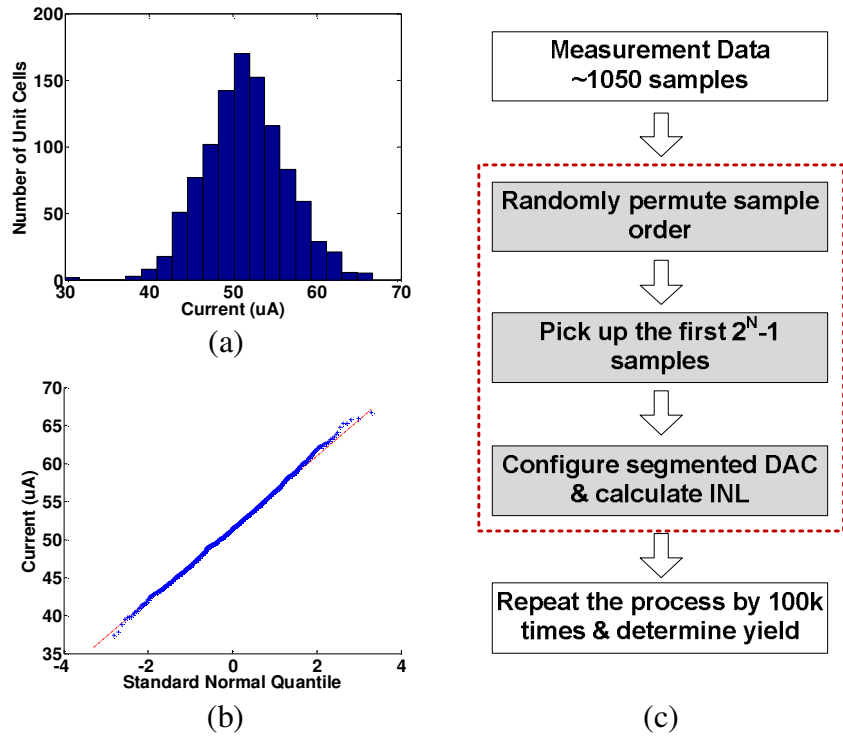
Figure 23.   (a) Histogram of the measured current data. (b) Q-Q plot of the

measured current data. (c) Procedure of pseudo MC simulations.

(a)   (b)
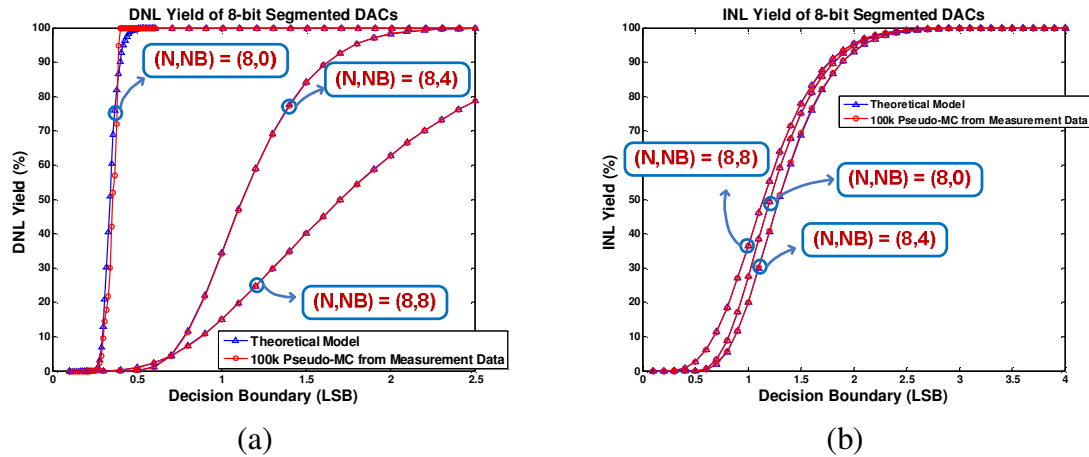
Figure 24.   (a) DNL yield and (b) INL yield curves of 8-bit segmented DACs.

Figure 24 shows that the analytical models can be reliably used for estimating the yield of the two linearity metrics. However, a slight deviation is observed for the DNL yield of the thermometer-coded DAC. This deviation caused by the slight non-Gaussian current distribution is more discussed in the next section.

(a)

(b)   (c)
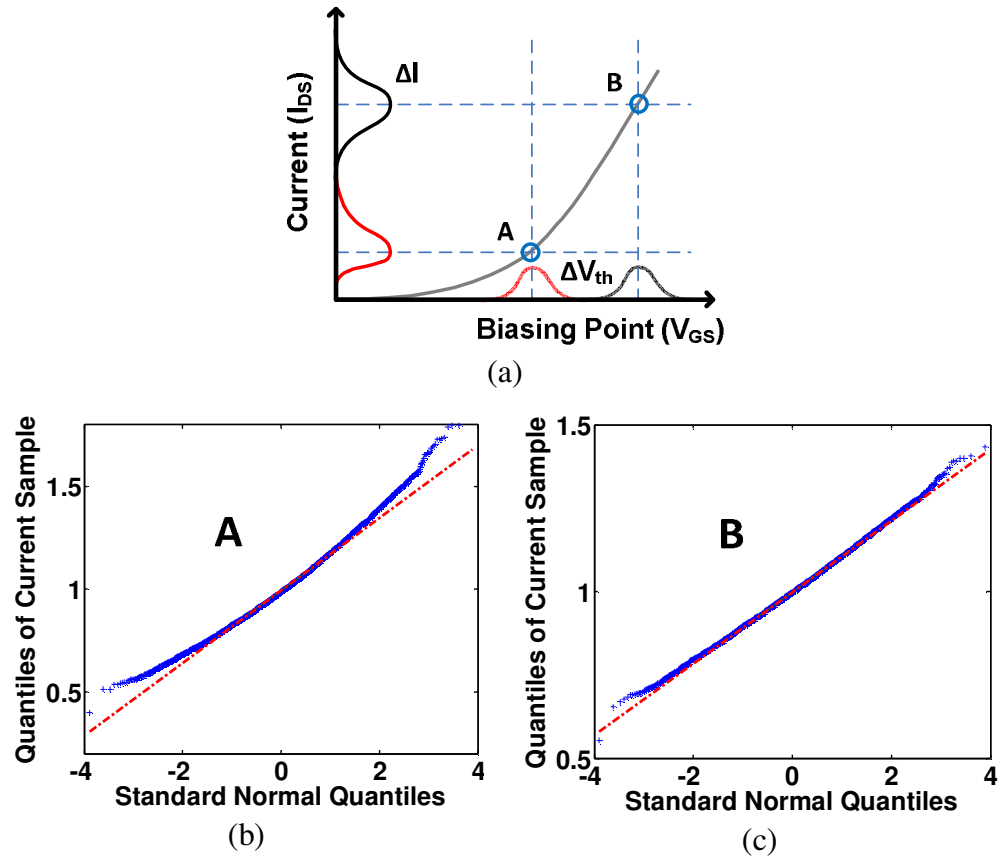
Figure 25.   (a) Bias point of a current source and its corresponding current distribution. Q-Q plots of current data (normalized by its mean) versus standard normal distribution: (b) with the current source biased at a low overdrive voltage and (c) with the current source biased at a high overdrive voltage.

3.4.2.   **Impact of Device Nonlinearity**

Device nonlinearity becomes substantial when the devices in a unit cell are biased at a low overdrive voltage and have small area ($\Delta v_{th}\uparrow$) or large aspect ratio (W/L). Figure 25(a) shows such nonlinear deflection of the current distribution, generated by HSPICE MC simulation. Near the tails of the distribution, the CDF of the current cell's variation deviates from the CDF of the standard Gaussian distribution.
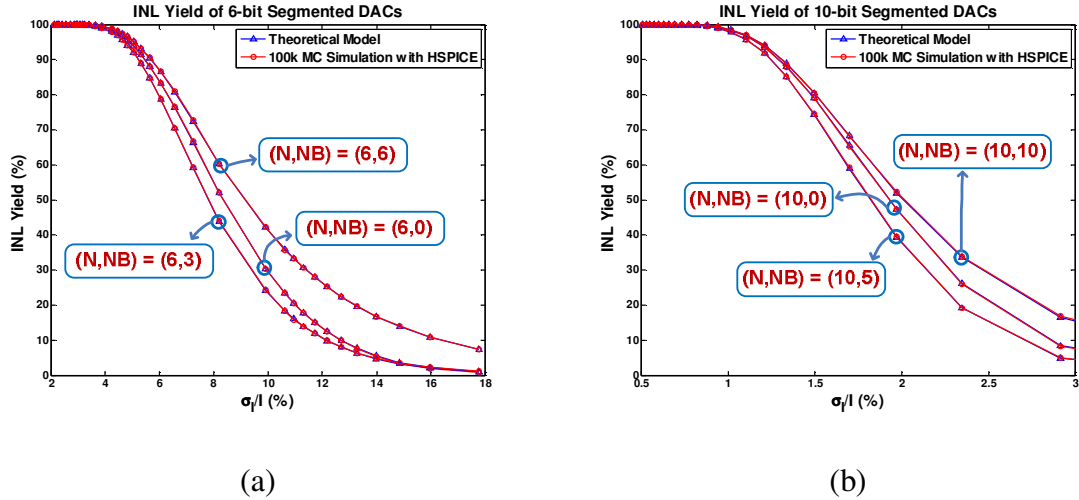
Figure 26.   (a) INL yield curve of several 6-bit segmented DACs. Current data are generated from HSPICE MC simulations. (b) Similar INL yield curve of several 10-bit segmented DACs.

Figure 26 shows the INL yield simulated using the nonlinear distribution of Figure 25(b) for several DAC designs. The yield is derived from 100k MC simulations at each case. Figure 26(a) and (b) show that the yield curves of three different 6-bit and 10-bit segmented DACs are not sensitive to nonlinearity and still matches accurately to our model. Nonlinearity does not have significant impact on the yield curve because of two reasons. First, since a noticeable nonlinear deflection is accompanied by large current

variation ($\sigma I/I$), the portion of the yield curve that is affected is where the yield is very low and not a concern for a practical high resolution DACs. Second, since INL is a cumulative sum of individual current variations, the Central Limit Theorem leads to relatively Gaussian behavior near the middle of the input code.
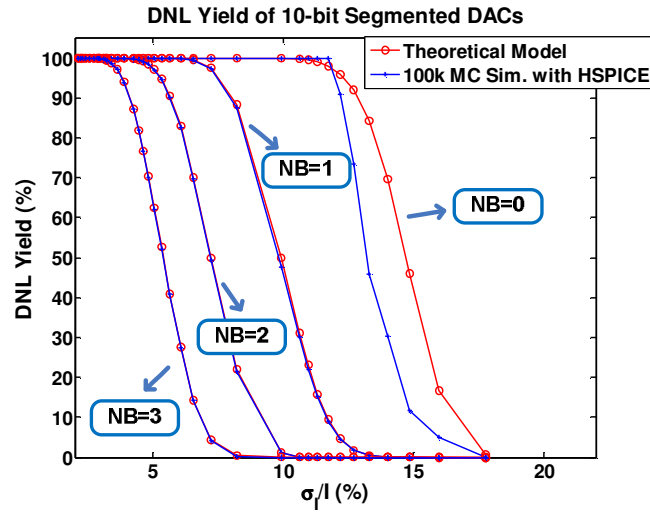
Figure 27.   DNL yield plots of the model and 100k HSPICE MC simulation results for 10-bit segmented DACs with NB = 0~3.

The same HSPICE simulation is performed for the DNL test. In Figure 27, results from 100k HSPICE MC simulation for 10-bit segmented DACs with NB = 0, 1, 2, and 3 are plotted with the analytical model. Except for a fully thermometer-coded DAC (NB=0), the model prediction matches the simulation results well. From the same viewpoint of the INL characteristics, the Central Limit Theorem leads to asymptotic Gaussian behavior for each thermometer group and MSB groups of the binary section. Therefore, the DNL yield models can be applied to most of segmented DACs that are built with active unit cells while thermometer-coded DAC may result an optimistic estimation of the DNL yield.

3.4.3.   DNL and INL Yield Comparison

Design targets typically need to satisfy both of the INL and the DNL requirements. For the same target yield level and the decision boundary, either the INL or the DNL may determine the yield. As shown in Figure 28 (only 14-bit results are shown but other resolutions have similar plots), for 6~14-bit arbitrarily segmented DACs, the DNL yield is more error-tolerant compared to the INL yield as long as NB < N-2. Using the two models, more diverse yield levels can be compared between the two metrics.
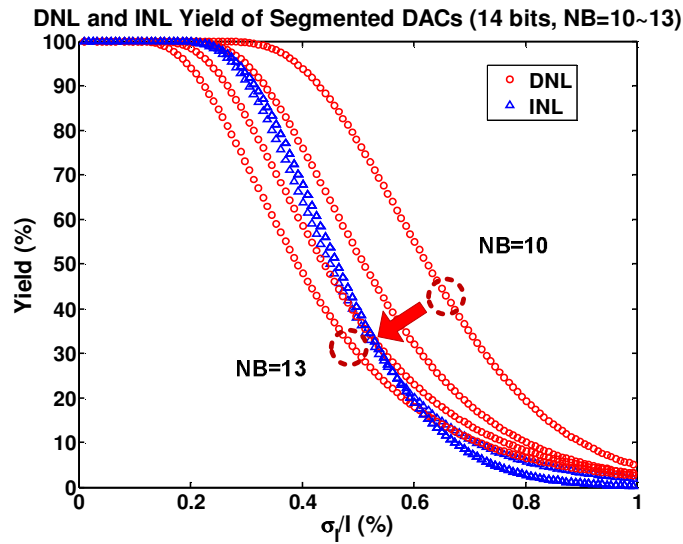
Figure 28.   DNL and INL yield plot of 14-bit segmented DACs. A = ±0.5 LSB.

Section 3.5   Summary of DAC Yield Models

Simple analytical formulations of the DNL and INL yield for any arbitrarily segmented DAC are presented. Combination of the two models can rapidly estimate yield levels for a given unit cell variation. The inverse of the yield models can be used to explore the maximum resolution of the system for a pre-defined area and a decision boundary (A) or

to pre-estimate the necessary device geometry and the bias condition for a target yield and A. The choice of the model depends on the segmentation structure of the DAC.

Each DNL yield model has a unified formulation across different resolution bits so the model can be easily implemented in a mathematical tool. The INL model is based on fitting parameters and the complete table for 6~14bit arbitrary segmented DAC is given as a table (3-2). This work clearly showed the condition when the DNL yield requirement is lower than the INL yield requirement such as NB ≥ N-2.

Chapter 4
Background on SRAM Stability

In this chapter, the existing definitions on the SRAM cell stability is reviewed. Section 4.1 and Section 4.2 discuss static and dynamic stability respectively including the definition and various metrics. Section 4.3 describes circuit techniques for assisting read and write stability. These techniques are closely related to measuring the stability of a cell without incurring hardware overhead. Section 4.4 reviews previously published BIST circuits. Two of the stability metrics reviewed in this chapter is used in Chapter 5 as the target metric for the proposed stability estimation

Section 4.1   Static Stability

SRAM failure can be generally categorized by 1) read failure and 2) write failure [107]. The read failure is either due to unstable read stability of the cell or due to too small read current to create sufficient voltage swing that can be detected by the sense amplifier. Each case is characterized as read stability, I_{READ}, and write-ability. These

error types are defined as "soft fails." According to [87], the occurrence of hard fails[1] due to defective cells is diminishing by process improvement, while the soft failure rate is increasing by the device mismatch. In this section, the static stability margins for the read and write are reviewed and their benefits and limits are discussed.
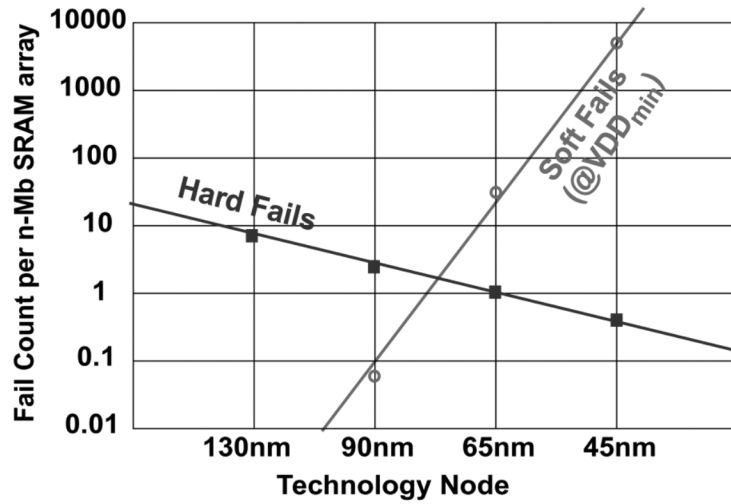
Figure 29.   Trend of SRAM hard and soft fails. Courtesy of [87].

4.1.1.   **Static Noise Margin**

Static noise margin (SNM) is a widely used term to quantify static logic's stability in terms of the tolerable amount of DC noise voltage injected at the data storage nodes without changing the stored states. Read SNM (RSNM) and write noise margin (WNM) are extended definitions of the static noise margin [68-69].

Figure 30 shows the schematics of the 6T SRAM cell for the test of the static read disturbance. The word-line (WL) is statically driven to the nominal V_{DD} level and the bit-

---

[1] Time dependent aging effects such as negative bias temperature instability (NBTI) shift the threshold voltage of the PMOS device, which affects the failure conditions [109-112].

lines (BL) are tied to V_{DD} that represents the stress during the initial phase of the read access. In a normal operation, the word-line is accessed by a fixed pulse and the bit-lines are slowly discharged from the pre-charged level. These two effects create dynamic read stability issues that are discussed in section 4.3. The voltage transfer characteristics (VTC) of the two inverter pairs can be independently drawn by controlling the internal nodes V_R and V_L. Figure 30(b) shows the two VTCs from each half of the SRAM cell. The RSNM is defined as the maximum square that can be fit into the enclosed area. Depending on the direction of the area, RSNM can be defined as $RSNM_L$ or $RSNM_R$.

The WNM is defined in a similar way. Instead of driving the two BLs to V_{DD}, one of the bit-lines is discharged to ground to change the state of the cell. Figure 31(a) shows the schematics during the static write access. The WNM is defined as the minimum square fit into the enclosed area by the two VTCs (Figure 31(b)). The RSNM and WNM are to represent the maximum tolerable noise that can be injected into the internal nodes of the SRAM cells without disturbing the intended operation.

The static margins are determined by relative strength between cells. For the read stability, the strength ratio between the pull-down (PD) and access (AC) device is important to reliably discharge the bit-line. The device size ratio between PD and AC is defined as pull-down ratio (β). During write access, the access device pulls down the node storing '1' to ground. The initial fighting between the pull-up (PU) and access (AC) device determines the write-ability. The device ratio between PU and AC is defined as pull-up ratio (γ). As the SNM and WNM are determined by device strength ratio, the variations of the margins can be expressed as threshold voltage variation of each device [71-74]. These analytical models are widely used to explore tolerable threshold voltage variations to guarantee 6σ yield of the static margins.

Despite the simple interpretation, the actual measurement of the RSNM and the WNM requires an internal node access of every single cell and graphical analysis on the measured data. This internal access demands additional wiring and switches to the 6T cell

layout structure, which cannot be feasible in a dense SRAM array. Also they cannot model the dynamic behavior of the pulsed read/write access such that the static margins generally underestimate the dynamic read stability or overestimate the dynamic write-ability.

(a)   (b)
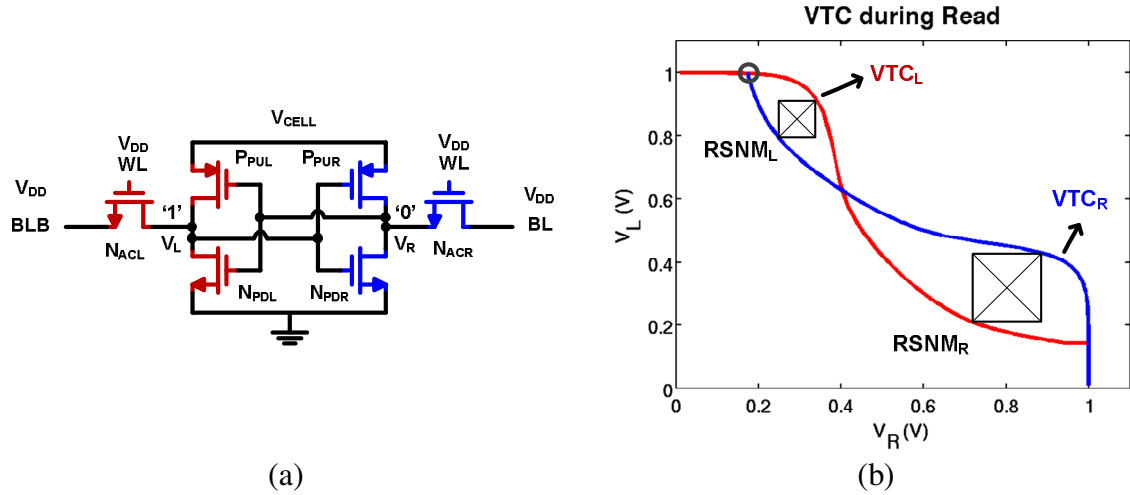
Figure 30.   (a) Schematics of the 6T SRAM cells and (b) voltage transfer characteristics (VTC) during the static read operation.

(a)   (b)
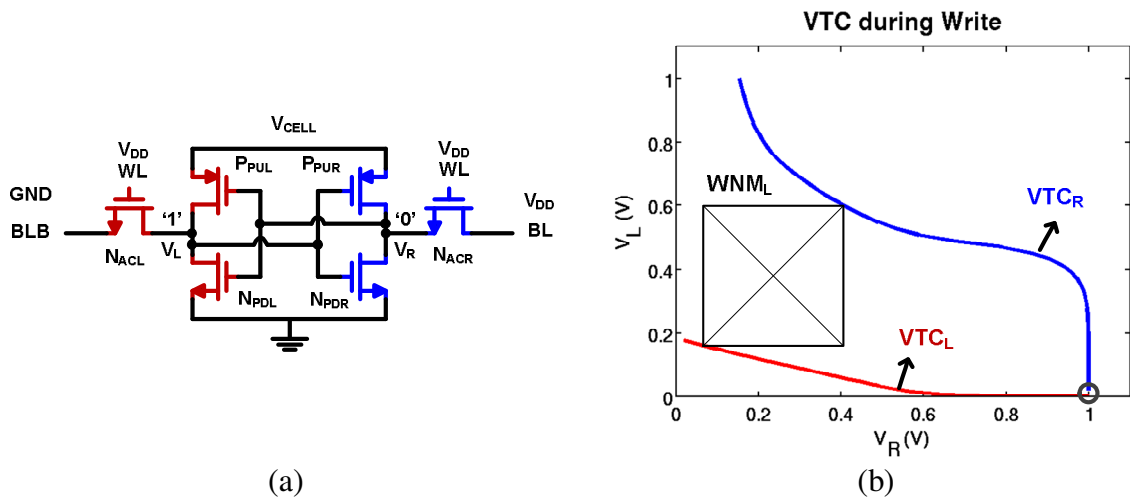
Figure 31.   (a) Schematics of the 6T SRAM cells and (b) voltage transfer

characteristics (VTC) during the static write operation.

4.1.2.   SINM and WTI

Static current noise margin (SINM) and write trip current (WTI) are alternative metrics to address the read stability and write-ability of the SRAM cell [75]. In this measurement, bit-lines are tied to the supply level, and the access transistors are turned on, and a tester voltage source is applied to the node at the "low" state (V_R in Figure 32(a)). Sweeping the test voltage gives two stable and one meta-stable points where the pull-down and pull-up currents created by each device balance. At these points, the input tester does not conduct input current (I_{in} = 0A). Figure 32(b) shows the VTC_L by sweeping the node voltage V_R using the input tester. The amount of the input current (I_{in}) is plotted in the same figure, which is called N-curve.

The voltage distance from the initial condition 'A' to the meta-stable point 'B' is the required noise magnitude to flip the initial state. This distance between A and B is called static voltage noise margin (SVNM) for the read access. The maximum amplitude of the input current between A and B is called static current noise margin (SINM), and this current quantifies the amount of charges necessary to change the cell state. [75] noted that although these two metrics represents the same read static margin, their values can be greatly different depending on their dynamic stability. For example, the voltage margin such as RSNM or SVNM is determined by device ratio. However, the dynamic characteristics are also strongly dependent on the maximum tolerable charge injection without flipping the state; hence SINM has an additional dependency on the device size (plus ratio).

In opposite way, during write operation, the initial condition at 'C' should be reversed back to 'A' for successful programming. Even during this write test, the bit-lines are still

tied to VDD that implies the worst-case scenario for pessimistic write-ability estimation. The voltage distance between C and B is defined as write trip voltage (WTV) while the peak current is defined as write trip current (WTI). While WTV mostly depends on the pull-up ratio (effective width of pull-up device/effective width of the access device), WTI also depends on the device size (e.g. cap).

From the new definitions of SINM and WTI, [75] revealed that the dynamic stability cannot be expressed by RSNM or WNM. However, the tester voltage source required for SINM or WTI has to access the internal nodes of every single SRAM cell. In addition, as the word-line and bit-line are always driven high, this testing setup does not properly addresses the dynamic issues such as pulse width of the WL access or bit-line cap. Later in [118], the authors claimed that WTI and WTV defined in N-curve do not properly characterize the write-ability under a pulsed word-line access condition.

(a)

(b)
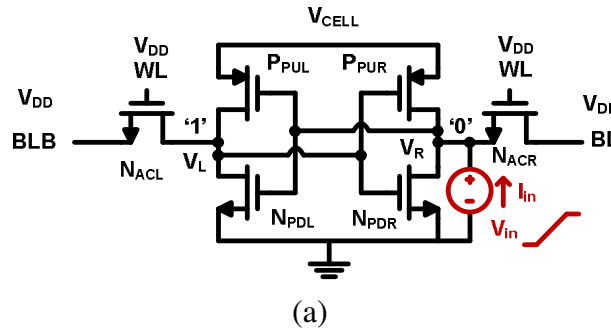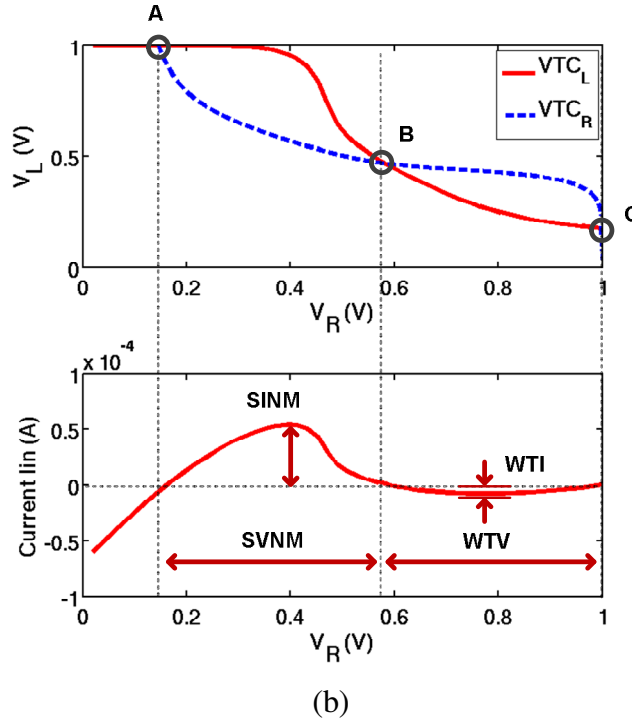
Figure 32.   (a) Test setups for SVNM and SINM (plus WTV and WTI). (b) VTC
and N-curve that defines SVNM, SINM, WTV, and WTI.

4.1.3.   Read Retention Voltage

Read retention voltage (RRV) is a measurable quantity for an in-situ cell by changing
the cell supply (SRRV) or word-line driver supply (WRRV) [80]. As these two metrics
are highly correlated, this paper focuses on SRRV. This technique can be used to extract
the cell's read stability without changing the cell layout. The approach measures the
lowest cell supply voltage before disturbing the stored bit. From Figure 33(a), all bit lines
are tied to the supply line with the word-line on, and the cell supply voltage (V_{CELL}) is
swept from high to low. At the beginning of the test, '0' is written into the target cell
hence V_R stores a low state. Next, V_{CELL} scales down by a predefined step ΔV, and then
the bit-line current (I_{BL}) is measured. If the data is not changed, non-zero amount of

current can be measured. This procedure is repeated until the stored bit is flipped. At this cell supply level, I_{BL} suddenly drops to zero. The amount of the supply voltage scaling is recorded as the SRRV.

This process is conceptually illustrated in Figure 33(b). This plot is drawn based on the read access outlined in Figure 33(a). During the SRRV test, the internal node V_L is pulled up high to V_{CELL} and this level is scaled down by a multiple of ΔV. The initial flat part (in Figure 33(b)) of the VTC_L shifts linearly downward by ΔV. One observation is that the "tail" of the VTC_R does not shift unless V_{CELL} reaches a certain voltage level such that N_{ACL} turns on. For every shift in supply of ΔV, the enclosed area is also linearly decreased. Due to the asymmetry between the left and right side inverters, one of the cross-sections of the two VTCs may have a smaller enclosed area. When V_{CELL} drops to a sufficiently low level, the VTC curves have only one stable point and the stored data is lost. In the case of Figure 33(b), where the enclosed area of the right side is larger than the left side, a 1 to 0 (i.e. B to A) transition is not likely to happen while 0 to 1 (i.e. A_1 to B_f) can be easily detected with a small V_{CELL} variation. Depending on the direction of the flipping, the SRRV can be divided into $SRRV_L$ and $SRRV_R$.

As reported in [117], SRAM cell's read stability is critically affected by the cell supply level. Other stability enhancing techniques such as reverse body bias or adjustment of the word-line driver supply level have relatively small impact on the read stability. Therefore, a proper cell supply level should be chosen based on the SRRV distribution to minimize the read disturbance.

(a)   (b)
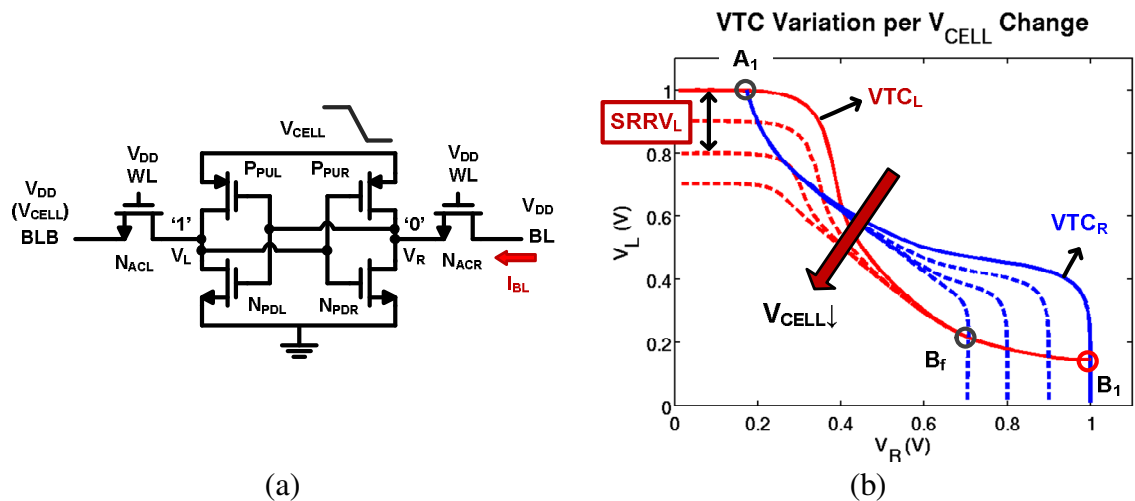
Figure 33.   (a) Schematics of the 6T SRAM cells and (b) VTC variations during

the static read access by sweeping the cell supply.

Figure 34 shows how the SRRV is correlated to the RSNM. This scatter plot is from MC simulation results in a 45nm CMOS technology. For relatively unstable cells, the SRRV test effectively extracts the enclosed area by the VTCs as seen in Figure 33(b), leading to a high correlation to the RSNM. However, the scatter plot becomes gradually dispersive as the RSNM is higher. When the cell supply scales down, the flat part of each VTC should be linearly scaled for ideal extraction of the enclosed area. When the cell is highly stable such that the cell supply needs to be scaled to a significantly low level to flip the state, the access transistor (N_{ACL}) turns on. As a result, the internal node storing '1' (V_R in Figure 33(a)) is partially pulled up by N_{ACL} preventing the cell from flipping. The measured SRRV is greater than if the measurement is without the influence of N_{ACL}. This error deflects the scatter plot in Figure 34(a) when the SRRV is larger than 35% of the nominal supply level (V_{DD}). This effect indicates that the SRRV is more useful for the cells close to failure.

The correlation can be improved with a modification of the SRRV technique by suppressing the unwanted pull-up of the access device. In Figure 33(a) during the $SRRV_L$

measurement, if BLB is tied to V_{CELL} (not to the nominal V_{DD}), then N_{ACL} does not raise V_L above V_{CELL}. The elimination of this extra pull-up strength improves the linear correlation between the SRRV and the RSNM as shown in Figure 34(b). The coefficient of determination (R^2) is improved from 0.75 to 0.84. Note that for $SRRV_R$ measurement, the voltage setups of BL and BLB should be switched.

Even though the RRV is focused on the characterization of the static stability of the SRAM cell, this dissertation proposes dynamic SRRV that captures the dynamic disturbance arising during the read access. The dynamic SRRV is introduced in section 4.3.
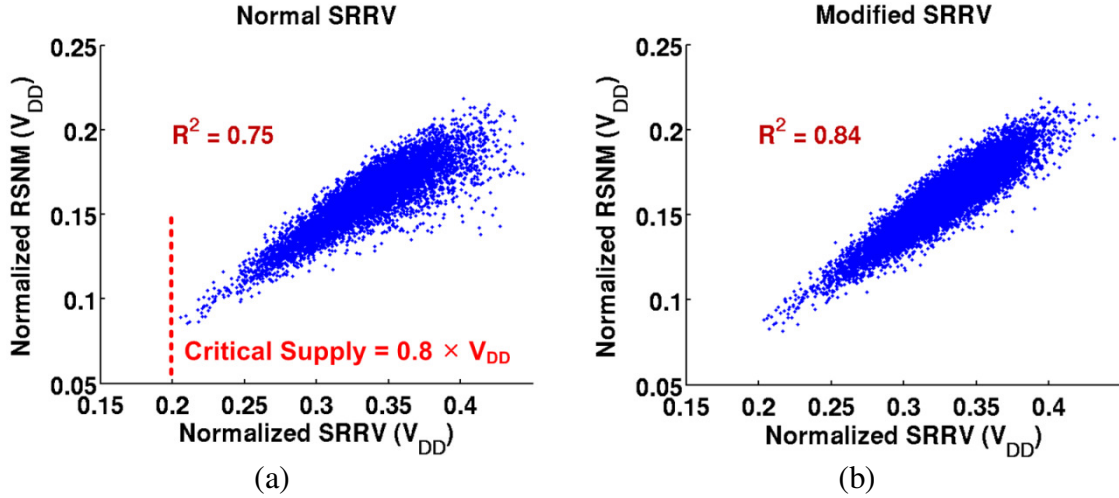
Figure 34.   Correlation scatter plot between the RSNM and the SRRV (10,000 MC simulation results in a 45nm CMOS technology). The results are normalized by the nominal supply level (V_{DD}). (b) Modified SRRV simulation.

4.1.4.   Write Trip Voltage

Similar to the concept of RRV, the write-ability of a cell can be expressed as the marginal voltage of the bit-line (BWTV) or the word-line (WWTV) during the write operation [76-77]. The BWTV is defined as the maximum tolerable voltage on the low

bit-line side for the successful writing of the wanted data bit without wasting too much energy for discharging the bit-line cap [75], [78]. In Figure 35(a), if the BLB voltage is not fully low, the internal state of the 6T cell resists against the forced state reversal by the write access. The maximum tolerable non-zero bit-line voltage effectively represents how easily the cell can change its state during the write access and how the cell is robust against noise injected at the low-side bit-line. The WWTV is defined as the minimum word-line voltage for the successful programming of the cell. As is already verified from the measurement data [80], these two metrics are highly correlated so this paper focuses on the BWTV.

(a)   (b)
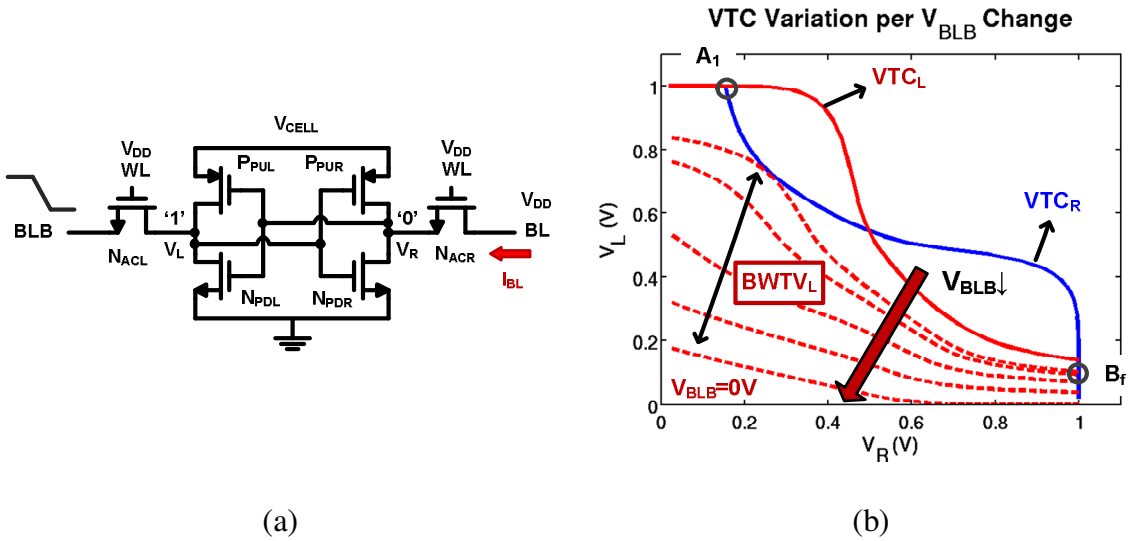
Figure 35.   (a) Schematics of the 6T SRAM cells and (b) VTC variations during the static write access by sweeping the low side bit-line voltage.

Similar to the SRRV measurement, the BWTV can be measured without change of the cell layout. From Figure 35(a), one of the bit lines storing '0' is tied to V_{DD}, and other bit-line storing '1' is swept from high to low. The VTCs under this condition is named as A_1

in Figure 35(b). If the given cell has a positive SRRV (i.e. nonnegative read stability), a non-zero I_{BL} can be measured. As V_{BLB} scales down linearly, the operating point moves from A_1 to B_f where I_{BL} becomes zero. This bit-line voltage is recorded as the $BWTV_L$. The $BWTV_R$ can be tested on the opposite bit line.

Although the BWTV is representing the static write-ability of the SRAM cell, it also characterizes well the dynamic behavior during write access. The dynamic BWTV is introduced in section 4.3.

This section reviewed the published static read stability and write-ability metrics. Among these definitions, only SRRV (or WRRV) and BWTV (or WWTV) can be applied to the given memory array without changing the major structure. In the following section, the dynamic characteristics of the read and write access is reviewed. Based on the discussions, the dynamic SRRV and BWTV are proposed that are used as a primary way to gauge the dynamic stability of 6T SRAM cell. The validity of the metrics is verified using the test chip built in a 65nm CMOS in Chapter 6.

Section 4.2   Dynamic Stability

4.2.1.   Dynamic Stability

Although the conventional design methodology of SRAM bit cell highly depends on the static margin, the soft error caused by dynamic behavior is depicted in recent literatures [113-119]. During the read access, the cell is stressed by a short pulse width for high performance SRAM. This pulse width may not be sufficient to inject a large amount of charge to the internal node to flip the cell state. Hence, the dynamic read stability is greater than the conventional RSNM that is extracted from the worst-case scenario with the infinite access time. The opposite situation can be expected for dynamic write access, where a certain amount of charge should be extracted from internal nodes

during the access time. Literatures in [117-118] used adjustable pulse generator to measure the minimum pulse width for read disturbance or successful write. In [117], the measured pulse width shows a large number of outliers for the read disturbance while the write access shows better correlations. The major problem of the author's claims is that they used timing metric to the voltage (or current) metrics defined in Section 4.2. Unfortunately, none of the literatures clearly addressed the correlation between the identical control factors such as voltage-to-voltage with fixed access timing. In the next section, this dissertation proposes two dynamic stability metrics that are more realistic in SRAM operation. The main idea is to fix the access time, and the cell stability is measured statically or dynamically by sweeping a supply source (e.g. cell supply or word-line supply). The simulation results show an excellent correlation between the two metrics.

4.2.2.   **Dynamic SRRV**

As the SRRV is measured by tying the bit-lines and the word-lines to static voltage sources, the measurement does not consider many of the dynamic behavior from a short access pulse width when a cell is normally accessed. The measurement of the static stability assumes infinite capacitance on the bit-line. This assumption leads to the worst disturbance to the internal node, which results in underestimation of the actual read stability. The actual read failure is strongly affected by the time constant of the internal node.

Figure 36(a) shows an example of pulsed read access with different cell supply levels. If the given cell is sufficiently stable at high V_{CELL}, the internal nodes does not flip during the access time (T_W) as in Figure 36(b). Figure 36(c) shows a read failure when V_{CELL} is lowered and the cell becomes unstable. As is done for the static SRRV, the dynamic SRRV with a fixed T_W can be defined by the changed in V_{CELL} required to flip the cell's

state. From the definition, the dynamic SRRV with infinitely long T_W is equivalent to the static SRRV. In Figure 36(c), the transient of the internal node voltages indicate that the dynamic stability not only depends on the static noise margin of the cell but also depends on the time constant of the internal nodes [117-118]. However, if the stability estimation is focused on the cells close to failure, a strong correlation can be observed between the static and the dynamic behaviors.

(a)

(b)

(c)
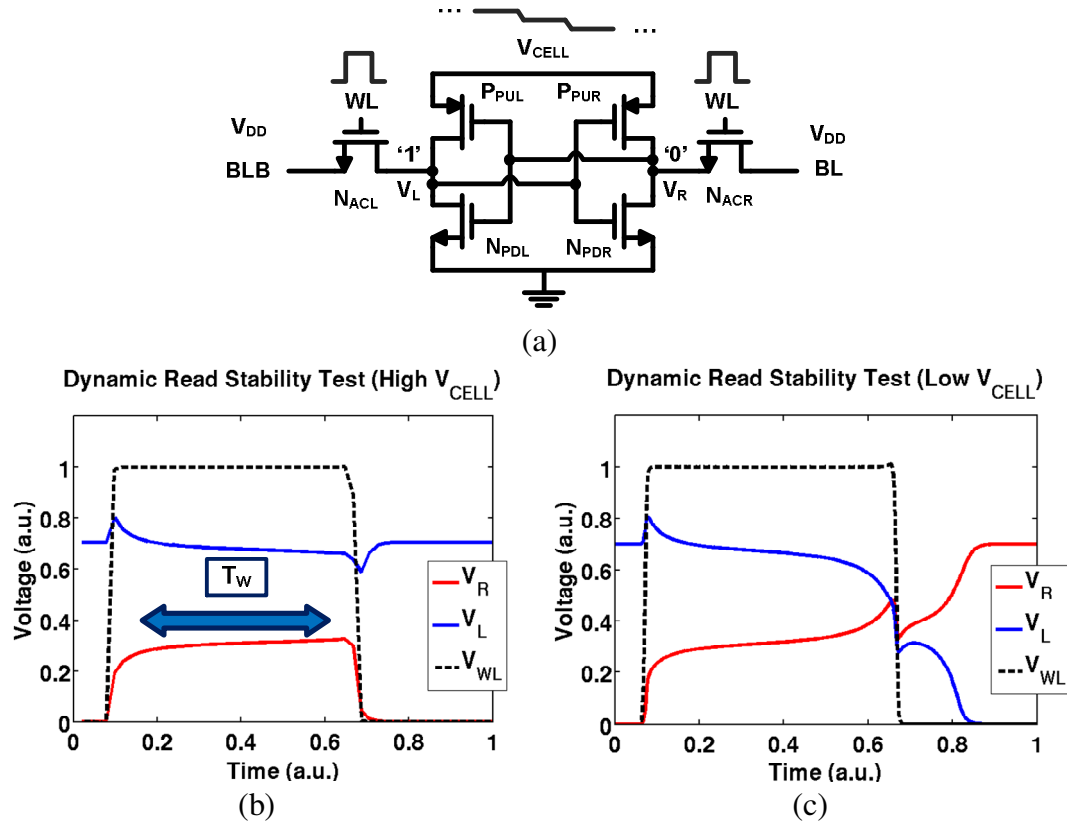
Figure 36.   (a) Simulation setups for the dynamic SRRV measurement. Internal node voltages (b) at high V_{CELL} and (c) at reduced V_{CELL}.

Figure 37 depicts such correlation when T_W is (a) 250ps and (b) 100ns. The dynamic SRRV is generally larger than the static SRRV but the amount of such improvement

depends on the access pulse width. The excellent correlation near failure is because their weak internal nodes are flipped due to the rising edge of the word-line signal and are not influenced by the time constants of internal nodes. By extending the access time, the cell behavior is closer to quasi-static such that the strong correlation extends for more stable cells as in Figure 37(b) (T_W =100ns).

Therefore, the static SRRV (SRRV hereinafter) is a valid metric to characterize both of the static and the dynamic stability of the SRAM cells near the tails of the stability distribution. Since the environmental issues such as the bit-line capacitance and the slope of the word-line transition cause estimation errors of the dynamic stability, the SRRV sets the pessimistic boundary of the cell supply voltage with the worst disturbance scenario during the read access. In Chapter 6, the relation between the static and dynamic SRRV is verified from measurement data in 65nm CMOS.
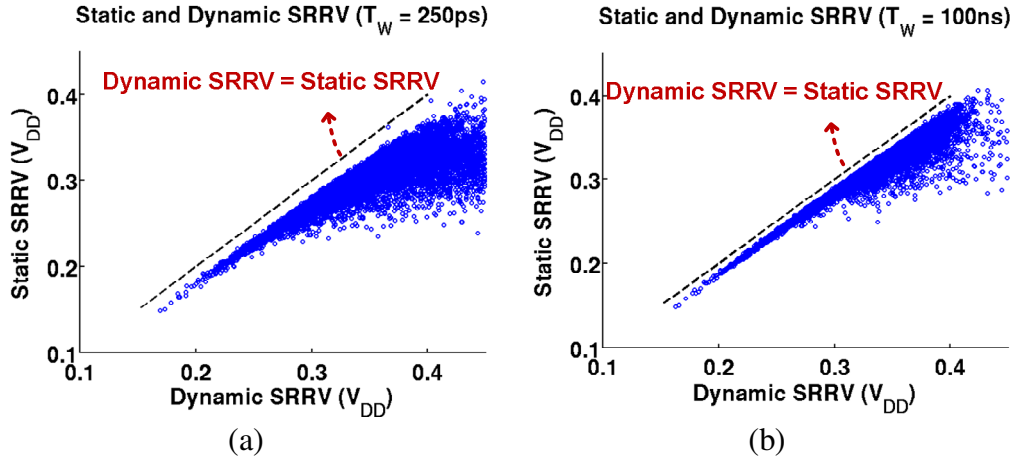
Figure 37.   Correlation plot of the static SRRV and the dynamic SRRV measured with (a) 250ps and (b) 100ns access time.

4.2.3.   Dynamic BWTV

Similar to the concepts of the dynamic SRRV, the dynamic write-ability (BWTV) can be defined with pulsed access (Figure 38(a)). The word-line is accessed for a fixed time (T_W) while the low-side bit-line voltage is slowly changing to low. If the bit-line voltage is not sufficiently low (Figure 38(b)), the internal state does not fully change during the access time and the write fails. For longer T_W, the probability of the write success increases such that the dynamic BWTV converges to the static BWTV.

(a)

(b)   (c)
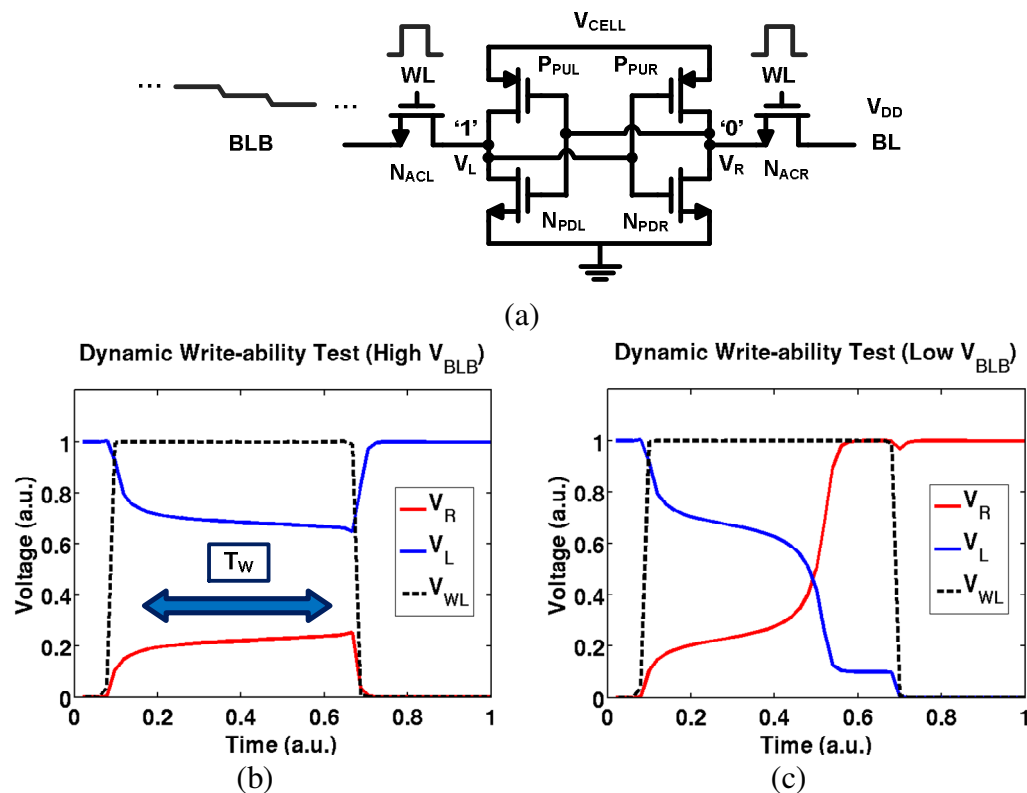
Figure 38.   (a) Simulation setups for the dynamic BWTV measurement. Internal node voltages (b) at high V_{BLB} and (c) at reduced V_{BLB}.

Figure 39 shows the correlation plots between the dynamic and the static BWTV with 250ps and 100ns access time. As the state change is strongly dependent on the discharging speed of P_{PUL} and N_{ACL} in Figure 35(a), the required bit-line voltage scales down for a narrower pulse width. Interestingly, the dynamic and the static BWTV show very linear correlation over the entire range[2]. Such high correlation enables the static BWTV to again be a valid metric to extract the distribution of the dynamic write-ability. The measured static and dynamic BWTV are provided in Chapter 6.

(a)   (b)
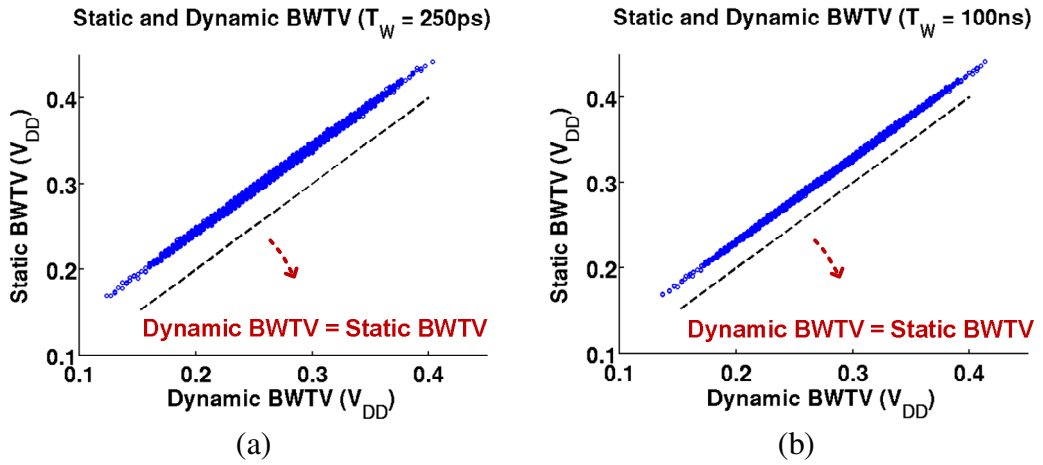
Figure 39.   Correlation plot of the static BWTV and the dynamic BWTV measured with (a) 250ps access time and with (b) 100ns access time. 10,000 MC simulation results are used in 45nm CMOS.

Section 4.3   Read and Write Assist Circuit Technique

To reliably reduce the SRAM's power supply level (for leakage and active power reduction), the read stability and write-ability should be guaranteed. Most of the existing

---

[2] Similar observation is claimed in [117-118] that the minimum write access time to flip the state under fixed bias has a close correlation to the static BWTV.

techniques are based on controlling the cell supply (V_{CELL}), the word-line signal level (V_{WL}), and bit-line level (V_{BL}).

4.3.1. Read Assist

The read stability is enhanced by increasing V_{CELL} [89], [91-93], [97], [106] or reducing V_{WL} [87], [94-104], [106]. While V_{CELL} greatly impacts the cell stability, it also increases the leakage and switching power of the entire system. Most literatures on the cell supply control are directed to reduce V_{CELL} as much as possible from the read test that does not include other source of errors such as power supply noise [119] or internal noise. In Chapter 6, a systematical approach to assign safety margin from measured stability is discussed. Control on V_{WL} has relatively a lower impact on the read stability, but lowering the gate voltage of the access device can significantly reduce the read current that leads to increase of the access time [85]. Note that V_{WL} has a contradictory tradeoff between the read stability and write-ability since lowering V_{WL} reduces the pull-down strength of the access device during write access. More detailed descriptions on the word-line control is presented in the next section.

Another read stability issue is from half-cell select for column interleaving structure [86-87]. Figure 40 shows the typical memory array structure that multiple columns share a single periphery such as sense amplifier. During read or write access, the entire row is selected while only one of them is connected to the sense amplifier. Other columns left with the pre-charged bit-line. The unselected cells should discharge the large bit-line caps, and the discharging speed is usually slow such that these cells experience stress from nearly static bit-line voltage. The half-cell select issue can be mitigated by using cross-point array such as 10T cell [86], placing sense amplifier per each column to write back unselected cells [87], or column-based VDD control [88]. All of these solutions demands increased area and complex wiring issues that is not suitable for high performance SRAM (but preferred from low power systems).

69

The amount of the required V_{CELL} or V_{WL} is expressed as SRRV or WRRV as defined in Section 4.3. Chapter 5 provides a rapid SRRV estimation technique that is applicable to the actual memory array (measurement data in Chapter 6).
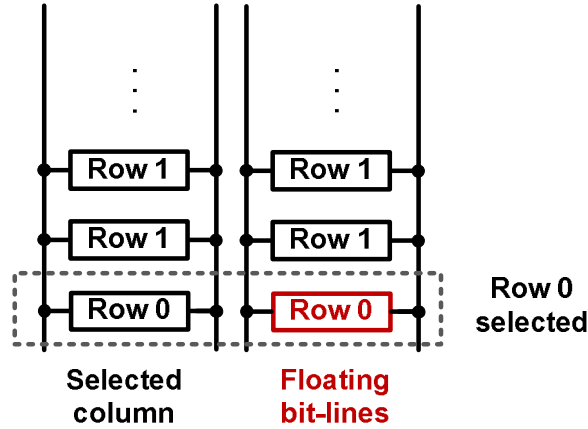
Figure 40.   Half-cell select issue

4.3.2.   **Write Assist**

The write-ability can be enhanced by boosting V_{WL} [87], [94-104], [106] or by driving the bit-line lower than ground [103-105]. The V_{WL} control affects both of the read stability and write-ability, literatures typically incorporates V_{CELL} control as well. The optimum V_{WL} can be found from equal read and write margin that is provided from modified cell structure with OP-Amp based feedback circuitry [76], [96] or by manual calibration with resistive divider to track process variation [97-99], [101-103].   The negative bit-line is especially useful for a low-power SRAM < 0.6V. This technique, however, requires a negative source generator by capacitive coupling or regulator. If the amount of the negative voltage is too large, it may cause device reliability issue [86]. For normal high performance SRAM, the bit-line does not necessarily be discharged below

ground. In this case, the bit-line can be driven to a non-zero potential to save active power to fully discharge the bit-line cap.

The amount of the required V_{WL} or V_{BL} is expressed as WWTV or BWTV as defined in Section 4.3. Chapter 5 provides a rapid BWTV estimation technique that is applicable to the actual memory array (measurement data in Chapter 6).

Section 4.4   Built-In Self-Test Circuits

From the increasing use of the read/write assist circuits, the need for built-in self test (BIST) circuits is also growing up. Primary purpose of BIST is to obtain the distribution of safety margin that can be used to properly set the optimum supply level. As the device model is becoming complicated and the 6σ boundary can be estimated too optimistically or too pessimistically from the simulation results, the read/write margin obtained from the BIST can precisely estimate the failure condition of the cells.

From the conventional definition of the static noise margin, [81-82] integrated the 6T cell array with additional switches to access the internal nodes. As a result, the size of the entire array increased and the wiring is very complicated. From measurements of I/O characteristics of each half cell, the authors successfully obtained the distribution of RSNM and WNM of the entire array that shows Gaussian distribution. However, the change of the original cell structure and the excessive increase of the area are the major limits in applying this technique to a large size memory array.

In [83], the read current and BWTV are measured for a 1Mb SRAM array by accessing each bit-line. This technique does not require access to the internal node of each cell, and the area required for the bit-line access can be small (~10% of the original array size). From the measurements, the distribution of the read current and BWTV are displayed to the 5σ tails that are useful in building database of the cell variability. In [80], the same bit-line access technique is expanded for other stability metrics such as SRRV or

WWTV, and the authors claim that the measurement technique can be applied to large scale memory arrays. However, the supply control cannot be rapidly done for large size memory array due to long time constant at the supply node, which limits the measurement time.

In [84], the authors used the read current and write current produced by the selected cell to build a relaxation oscillator. The amount of time to charge/discharge the large bit-line cap is a good indicator of the cell stability. For the first time, this technique also addresses how to measure the dynamic stability of the cell by periodically switching on/off the gate voltage of the access device. However, the measurement results in [84] do not show a good agreement to the existing stability metrics. Even simulation results were not nicely matched.

Despite the early works in SRAM BIST circuitry, none of the methods can be applied to the large scale memory array due to 1) excessive measurement time or 2) poor indicator of the cell stability. This dissertation proposes a rapid stability estimation technique that reliably predicts the failure conditions for both of static and dynamic concerns.

Section 4.5   Summary

This chapter reviewed the design trend of SRAM bit cell, and the associated reliability issues. The continuous 50% cell area reduction per technology node has induced the growing concerns on the reliability issues. Even though various read/write assist circuit techniques are suggested, none of the literatures clearly addressed a proper way to set the optimum margin per die. From the efforts to gauge the cell stability in a large scale, two major concerns in BIST circuits are discussed. The measurement time and prediction accuracy for both static and dynamic stability are considered by the proposed stability estimation technique in Chapter 5 and 6.

Chapter 5
Static Stability Estimation Technique for SRAM

Static noise margin is one of the key metrics to estimate the likelihood of failure of a 6T-SRAM cell. This chapter proposes a technique to accurately estimate the stability of a conventional SRAM cell without modifying the cell structure. The main idea is to measure specific cell's currents with variant supply levels via the bit-lines. The measured currents are used to estimate the read stability and the write-ability by nonlinear regression. Compared to the published stability testing circuits, the cell stability can be rapidly estimated with the estimation accuracy R^2 (coefficient of determination) as high as 0.95 once applied to an arbitrary data set. Simulation results show that the estimation error sigma is as small as 2.44% for the read stability and 3% for the write-ability estimation. Validity of the idea is verified by Monte-Carlo simulations using SRAM models in a 45nm CMOS technology. In chapter 6, the proposed technique and its estimation for the dynamic stability are discussed with measurement data from a standard SRAM block fabricated in 65nm CMOS.

Section 5.1   On-Chip Measurement for Estimating Cell Stability

As discussed in Chapter 4, the required precise supply control was the main issue in measurement time of the SRAM cell stability such as SRRV or BWTV. These types of measurements cannot be applied to entire cells in a large array as a method of pre-screening unstable cells (or determining the optimum supply voltage). To save the measurement time and efforts, this section proposes a static margin estimation technique with the cell currents under controllable operating conditions.

The estimation strategy is shown in Figure 41. For a given memory array, a small number of cells are randomly selected as a reference data set. This group is called the regression group. The way to select cells is more systematically described in Chapter 6 from an actual test chip. For the selected cells, the read stability (or write-ability) is measured not precisely but coarsely such as by sweeping the supply by tens of mVs. This coarse measurement of the stability induces quantization error. The maximum tolerable resolution of the stability measurement is discussed in section 5.4.

As a next step, the cell currents are measured under specific supply levels. Combining the measured stability and the cell currents under different supply levels, the relation between the two measured data is established. This formulation is applied to the remaining set of cells. By using the formulation and the cell current, the read stability (or write-ability) is estimated.

The standard metrics for the read stability and write-ability are presented as SRRV and BWTV, respectively. Compared to other published stability metrics, only these two definitions can be measured from an intact memory array without changing its layout structure. Nevertheless, the suggested technique can be applied to any stability metrics such as the conventional RSNM.
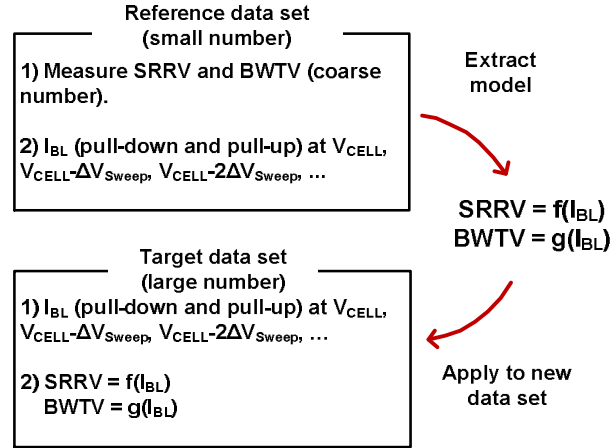
Figure 41.   Concepts of the stability estimation using cell currents.

5.1.1.   Cell Current Measurement

One way of accelerating the characterization of the variability of an intact 6T-SRAM cell structure is to measure the cell current via bit-lines while other peripheral circuits such as sense-amplifiers are turned off [70]. The cell current can be measured in two different ways: pull-down of the two NMOS devices and pull-up of the PMOS and the access NMOS devices. In Figure 42, the 6-T SRAM cell is driven either for the pull-down or pull-up (in parenthesis) current measurement. As the pull-down current is strongly dependent on the combined strength of the pull-down (triode region) and the access transistor (velocity saturation), this current contains information of the device variations of N_{ACR} and N_{PDR} (or N_{ACL} and N_{PDL}). Similarly, the pull-up current contains device variations of P_{PUR} and N_{ACR} (or P_{PUL} and N_{ACL}).

For most stable cells, when measuring the pull-down current I_{BL} in Figure 42, the node voltage V_L is approximately equal to V_{CELL} since N_{PDL} (in subthreshold region) does not conduct substantial amount of current to affect the node voltage. Consequently, the left half of the cell (N_{ACL}, N_{PDL}, and P_{PUL}) is nearly inactive during the measurement of I_{BL},

75

and their variations can be isolated. This can be observed in Figure 33 where V_R is solely determined by V_{TCR} not by V_{TCL}. In the same way, I_{BLB} is purely determined by the left half cell.
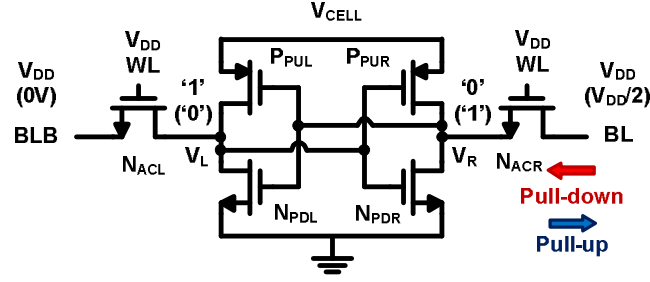
Figure 42.　　Pull-down (or pull-up) current simulation setup.

(a)　　　　　　　　　　　　　　　　　　　　(b)
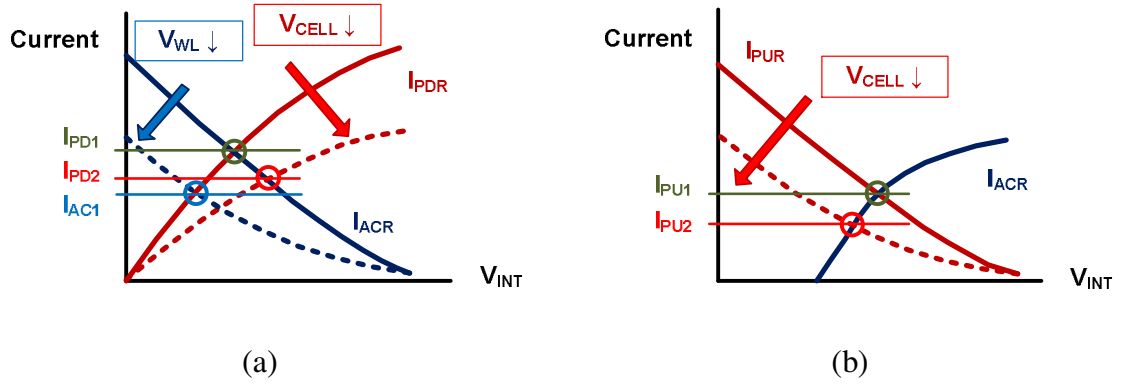
Figure 43.　　(a) Load lines and operating points of N_{PDR} and N_{ACR}. V_{INT} means internal voltage (V_R or V_L). (c) Load lines and operating points of P_{PUR} and N_{ACR}.

If only variation of the pull-down transistors is estimated, a single measurement of I_{BL} can be used as a direct indicator of the read stability. However, combined variation of multiple devices requires more measurements for precise extraction of the device

variability. The sensitivity of I_{BL} to V_{CELL} can be used as an indicator of the threshold variation of N_{PDR}. Note that this sensitivity also contains the threshold variation of N_{ACR} as I_{BL} is dependent on $V_{th,ACR}$ as well. Alternatively, the word-line voltage (V_{WL}) can be controlled to obtain I_{BL} sensitivity to V_{WL} (Figure 43(a)). Although either method can extract information on the device variations of N_{PDR} and N_{ACR}, controlling both of V_{CELL} and V_{WL} do not improve the extraction results as shown later from the simulation results in section 5.2. By considering the layout complexity, V_{CELL} can be chosen as an independent control factor while other supply voltages remain at the nominal V_{DD}.

Similar to the pull-down current measurement, the pull-up strength of P_{PUR} is tested with changing levels of V_{CELL}. In Figure 42, BL is tied to half of the main supply (V_{DD}). At this level, the operating point of N_{ACR} is in triode region. Such weak pull-down strength of N_{ACR} prevents turn-on of P_{PUL}, and V_R is only determined by P_{PUR} and N_{ACR}. From this measurement, the device variations on each side can be measured independently. In Figure 43(b), I_{PU} indicates a pull-up current with changing levels of V_{CELL}.

To determine the sensitivity, multiple discrete control levels for V_{CELL} are introduced. By scaling down V_{CELL} from a nominal level (V_{DD}) to a minimum value, the variation of I_{BL} is recorded by running HSPICE Monte-Carlo (MC) simulation. The minimum cell supply voltage to safely hold the stored data is defined as critical VDD [84]. In Figure 34(a), the simulation results show that the critical V_{DD} ($V_{Critical}$) is roughly 80% of the nominal V_{DD}. If a stored bit in a test cell flips at $V_{Critical}$, the test cell must be at the verge of unstable state; equivalently its read stability is close to 0. Note that V_{CELL} can be swept in multiple steps from V_{DD} to $V_{Critical}$ to more accurately track variation of I_{BL}.

In summary, the following steps are required for each cell to characterize the device variability.

1)   For pull-down current measurements via BL, the cell should be programmed to store '0' at V_R. Tie BLB and WL to the nominal V_{DD}. By driving BL at the nominal V_{DD} level, measure the bit-line current. Repeat this process with different levels of V_{CELL}.

2)   For pull-up current measurements via BL, the cell should be programmed to store '1' at V_R. Tie BLB to ground and tie WL to the nominal V_{DD}. By driving BL at half of the nominal V_{DD} level, measure the bit-line current. Repeat this process with different levels of V_{CELL}.

5.1.2.   **Measurement Concerns**

Since many SRAM cells share a single bit-line pair, leakage of inactive cells may introduce random error in the measurement data. To avoid leakage added to the active bit-line current, all the cells in the same bit-line should be uni-directionally programmed before measurement. For I_{PDR} measurement in Figure 42, other idle-state cells must be programmed such that their internal node voltage V_R is pulled-up high. Note that the residual leakage current can be measured first and then subtracted from the active bit-line current. For the MC simulations in the following sections, the effect of leakage current is considered by placing inactive (low word-line voltage) 63 SRAM cells in the same bit-line pair with appropriate internal states.

For the current measurement, the stored data precision is limited by the measurement circuits. Here assumes that an N-bit on-chip ADC converts the cell current into digital data. For the analysis in section 5.2 and 5.3, the resolution is infinite (ideal ADC). The required minimum resolution of the on-chip ADC and the controllability on the supply voltage is analyzed in section 5.4 with detailed simulation results.

Section 5.2   Estimation of the Read Stability

The goal of the estimation is to establish a functional relationship between the measured currents and the read stability. In this section, the metric for the read stability is the SRRV. Instead of measuring the read current while lowering the supply voltage until the cell fails, the proposed model makes a few key measurements at the periphery to extract the SRRV (i.e. failure point). This concept is illustrated in Figure 44.
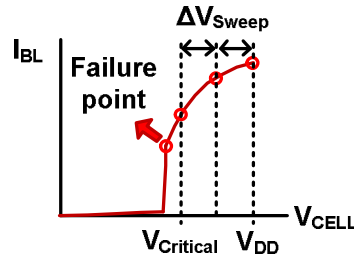
Figure 44.   Estimation of the failure condition. The failure condition can be extrapolated from the cell current variation with variant supply levels.

5.2.1.   Nonlinear Regression

From the pull-down and pull-up currents measured as in Figure 43, the cell currents are predictors and the SRRV is an estimation target. From the discussions in section 5.1, a polynomial fitting is used to resolve the inherent nonlinearity between the predictors and the estimators.

$$
\begin{aligned}
SRRV_L = \big(a_0 \quad a_{PDR1} \quad .. \quad a_{PDRk} \quad a_{PDL1} \quad .. \quad a_{PULk}\big) \\
\cdot \big(1 \quad I_{PDR} \quad .. \quad I_{PDR}^k \quad I_{PDL} \quad .. \quad I_{PUL}^k\big)^T
\end{aligned}
\tag{Eq 5.1}
$$

$$SRRV_R = \begin{pmatrix} b_0 & b_{PDL1} & .. & b_{PDLk} & b_{PDR1} & .. & b_{PURk} \end{pmatrix}$$
$$\cdot \begin{pmatrix} 1 & I_{PDL} & .. & I_{PDL}^k & I_{PDR} & .. & I_{PUR}^k \end{pmatrix}^T \qquad \text{(Eq 5.2)}$$

In (Eq. 5.1) and (Eq. 5.2), each column of the predictor matrix (e.g. I_{PDL}) is comprised of an n-by-m vector where n is the total number of samples and m denotes the number of supply levels. Subscript PD, AC, and PU are defined in Figure 43(a) and (b). Subscript L and R are used to separate each half cell and the two squares enclosed by the VTCs in Figure 33. The fitting coefficients used for the $SRRV_L$ and the $SRRV_R$ are identical if the SRAM cell is symmetric (e.g. $a_{PDR1} = b_{PDL1}$). Each coefficient is a 1-by-m matrix except for a_0. The coefficient a_0 can handle any type of constant error that is added during the measurement. In (Eq. 5.1), the coefficients $a_{PDR1}\sim a_{PDRk}$ estimate variation of the VTC_R curve from I_{PDR}. The coefficients $a_{PDL1}\sim a_{PDLk}$ and $a_{PUL1}\sim a_{PULk}$ combined with I_{PDL} and I_{PUL} estimate the initial curvature of VTC_L. Hence (Eq. 4.1) is an estimate of the $SRRV_L$.

(Eq. 5.1) and (Eq. 5.2) are solved using polynomial regression. To account for any types of errors caused by non-ideal models, this work uses the robust regression by iteratively re-assigning weight factors to each data point to calculate the weighted square sum of errors [58]. If data sets are away from the main cluster, small weights are assigned in the next iteration while data sets close to the cluster have higher weights. Within a few iterations, erroneous data sets are placed out of the main cluster and the curve fitting is done with the cluster data sets. The correlation between the predictors and the estimators is numerically calculated by the coefficient of determination (R^2) without the weight factors.

Figure 45(a) shows a regression example when m is 5, k is 2, and all six predictors are used for 4,000 samples. The estimator of the SRRV can be found by comparing $SRRV_L$ to $SRRV_R$ and taking the minimum of these two. The high coefficient of determination 0.95 validates the approach of using the measured currents to estimate the SRRV.

Once the fitting coefficients are found from a reference data group, the resulting relationship can be applied to the measured I_{BL} from any cell in the target data group to estimate its SRRV. Figure 45(b) is a scatter plot that shows the performance of the estimation by comparing the actual SRRV (from simulation) with the estimated SRRV of the target data group. The R^2 is as high as the correlation of the reference data group. Normalized errors of the SRRV estimation for the target data group are plotted in Figure 46(a). This histogram shows Gaussian distribution of the normalized error, and the standard deviation is 2.44%. With 3σ certainty, most of the estimation errors are confined within ±7.32% of the original SRRV. The predictability of the failure condition can be more clearly visualized by comparing the cumulative distribution of the simulated SRRV and the estimated SRRV. Figure 46(b) shows the difference for the cells close to the failure. The two CDFs deviate slightly for the relatively less stable cells but the amount is smaller than 5mV (for 1V nominal VDD).

(a)   (b)
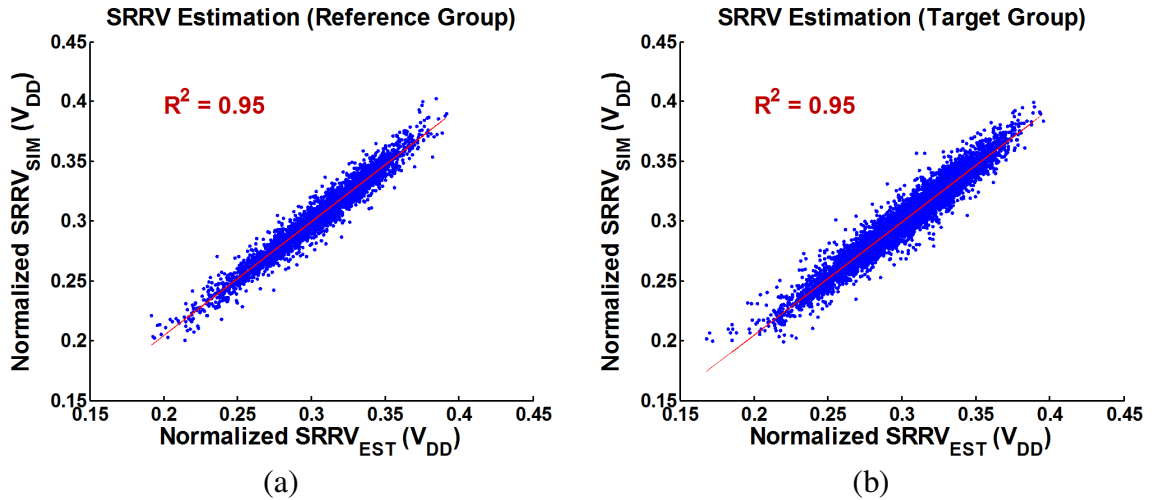
Figure 45.   SRRV estimation from the cell current measurement. Six types of

currents (I_{PD}, I_{AC}, I_{PU}, and L-R for each) with five supply steps are used. From

(Eq. 5.1) and (Eq. 5.2), n = 4,000, m = 5, and k = 4. (a) SRRV versus its

estimator for 4,000 MC runs (regression data set). (b) Another 12,000 MC runs

(target data set) with the estimation model (Eq. 5.1) and (Eq. 5.2).
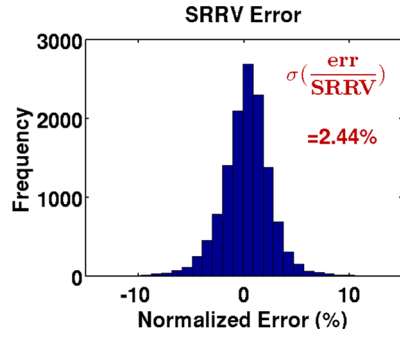
(a)   (b)

Figure 46.   (a) Normalized error (in %) histogram for the target data group. (b)

Cumulative distribution of the simulated SRRV and the estimated SRRV.
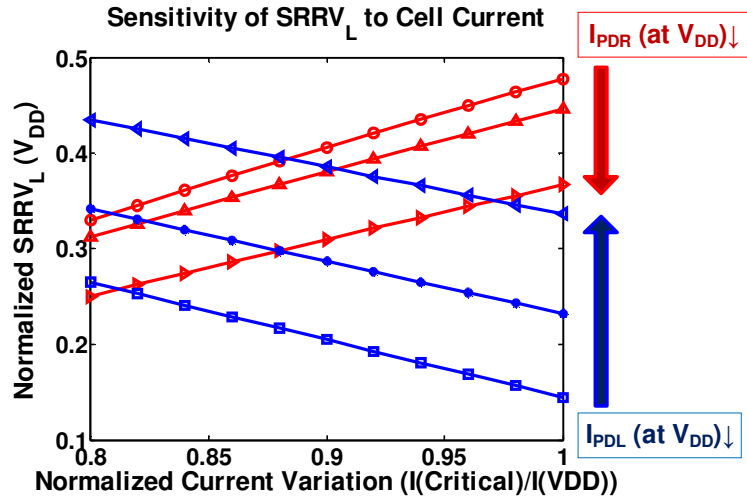
Figure 47.   Sensitivity of $SRRV_L$ to the cell current ($m = 2$, $k = 2$).

5.2.2.   Sensitivity of SRRV to Cell Currents

Each coefficient in (Eq. 5.1) and (Eq. 5.2) determines the sensitivity of the SRRV to the cell current variation at different V_{CELL}. In Figure 47, the regression coefficients are derived with two supply levels (m=2) and with quadratic equation (k=2). Using the fitting parameters extracted from the reference group, the $SRRV_L$ is estimated for various combinations of the cell currents at the nominal V_{DD} and at the critical V_{DD}. As discussed in section 5.1.1, a higher I_{PDR} at the nominal V_{DD} corresponds to a strong pull-down at the BL side, thus the $SRRV_L$ improves. Also, the threshold level of N_{PDR} and N_{ACR} should be small to reliably pull down the bit-line. Therefore, a low supply sensitivity of I_{PDR} indicates a larger $SRRV_L$ value. The opposite behavior is observed for the case of I_{PDL} since the VTC_L in Figure 33 is shifting to the left as N_{PDL} is stronger (reduced $SRRV_L$).
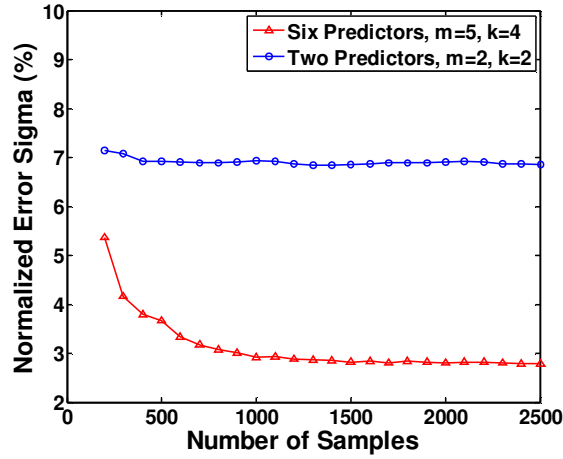
Figure 48.   Normalized error sigma (in %) as a function of the reference data
size for different number of measurements per cell.

5.2.3.   Required Data Size of the Reference Group

The size of the reference data group must be sufficiently large to accurately characterize the estimation coefficients. If the regression is performed with a small data set that does not cover the full range of the SRRV distribution, any estimation target out of the guaranteed range may produce over-fitting errors. Since a larger data size requires more number of the SRRV measurements with the increased regression complexity, there is a trade-off between the estimation accuracy and the data size. Figure 48 shows the normalized error sigma of the target data group when the coefficients are extracted from a different size of the reference data group. The first case (upper triangle) shows the normalized error sigma of the target group with the regression parameters identical to Figure 45. To avoid substantial estimation error, the reference data size should be greater than 1,000 samples. The second case (circle) used only I_{PDL} and I_{PDR} with two supply sweeps (m=2) and the second order fitting (k=2). Note that for this case only 500 samples are sufficient to suppress the error. The convergence speed of the normalized error varies since more predictors increase the regression complexity.

5.2.4.   Number of Predictor Variables

(Eq. 5.1) and (Eq. 5.2) have three variable parameters: the number of supply sweeps (m), the maximum polynomial order (k), and the number of predictor variables (I_{PD}, I_{AC}, I_{PU}, and L-R for each). By sweeping all the three fitting variables (m, k, and the number of predictor variables), the impact of each parameter on the estimation accuracy can be visualized. With each given parameter set, estimation accuracy is expressed by R^2. Figure 49 shows how R^2 of the $SRRV_L$ varies as a function of the number of predictor variables with changes of (a) the order of the polynomial functions, k, and (b) the step size of the supply sweeps, m. The included predictor variables at each number of the x-axis are tabulated in Table 5-1. At each column in Table 5-1, the most dominant predictor variables are selected to maximize the estimation accuracy, which matches well to the published sensitivity analysis of the RSNM to each device variation [71-73].

In each case, R^2 is affected mostly by the pull-down current (I_{PD}). Figure 49(a) shows how the $SRRV_L$ estimation is improved by the choice of the predictor variables. The initial jumps of R^2 by the first two predictors (I_{PDL} and I_{PDR}) imply that the NMOS devices (pull-down and access) have a dominant impact on the $SRRV_L$. I_{AC} does not contribute to the characterization of the NMOS device variability since the information it provides is redundant as indicated in Figure 43(a). In Figure 49(a), a small improvement of R^2 from the second order to the fourth order fitting indicates that k can be as small as 2 without significant degradation. Figure 49(b) shows variation of R^2 as the number of supply sweeps, m, is changed from 2 to 5. The maximum sweep range is fixed by 20% of the nominal V_{DD} (Figure 34(a)). The number of supply sweeps determines how finely the supply should be controlled. If m is 2, the cell current is measured at V_{DD} and $0.8 \times V_{DD}$. If m is 3, an additional cell current is measured at $0.9 \times V_{DD}$. Note that I_{ACL} can have a modest contribution to the $SRRV_L$ estimation. Since I_{PUL} contains device characteristics of P_{PUL} and N_{ACL}, including I_{ACL} eliminates uncertainty in characterizing P_{PUL} variability from the measured pull-up current.

Figure 49.   Coefficient of Determination (R^2) for $SRRV_L$ estimation with respect to the number of predictor variables (Table 5-1). (a) With $\Delta V_{sweep} = 0.1 \times V_{DD}$, the

fitting order (k) is varied from 1 to 4. (b) With k = 2, ΔV_{sweep} is varied from

$0.05 \times V_{DD}$ to $0.2 \times V_{DD}$. The sweep range is fixed by $0.2 \times V_{DD}$ and the fitting order

is 2 (square).

Table 5-1 Predictor Variables used in Figure 49.

	Number of Predictor Variables					
	1	**2**	**3**	**4**	**5**	**6**
Predictor Variables used in (Eq. 5.1) and Figure 49	I_{PDR}	I_{PDR} I_{PDL}	I_{PDR} I_{PDL} I_{PUL}	I_{PDR} I_{PDL} I_{PUL} I_{ACL}	I_{PDR} I_{PDL} I_{PUL} I_{ACL} I_{ACR}	I_{PDR} I_{PDL} I_{PUL} I_{ACL} I_{ACR} I_{PUR}
Predictor Variables used in (Eq. 5.2)	I_{PDL}	I_{PDL} I_{PDR}	I_{PDL} I_{PDR} I_{PUR}	I_{PDL} I_{PDR} I_{PUR} I_{ACR}	I_{PDL} I_{PDR} I_{PUR} I_{ACR} I_{ACL}	I_{PDL} I_{PDR} I_{PUR} I_{ACR} I_{ACL} I_{PUL}

Section 5.3   Estimation of the Write-Ability

5.3.1.   Nonlinear Regression

The write-ability (i.e. BWTV) of the SRAM cell can be estimated using a similar nonlinear regression. The basic equations used for the BWTV estimation are shown in (Eq. 5.3) and (Eq. 5.4).

$$BWTV_L = \begin{pmatrix} c_0 & c_{PDR1} & .. & c_{PDRk} & c_{PDL1} & .. & c_{PULk} \end{pmatrix}$$
$$\cdot \begin{pmatrix} 1 & I_{PDR} & .. & I_{PDR}^k & I_{PDL} & .. & I_{PUL}^k \end{pmatrix}^T \qquad \text{(Eq 5.3)}$$

$$BWTV_R = \begin{pmatrix} d_0 & d_{PDL1} & .. & d_{PDLk} & d_{PDR1} & .. & d_{PURk} \end{pmatrix}$$
$$\cdot \begin{pmatrix} 1 & I_{PDL} & .. & I_{PDL}^k & I_{PDR} & .. & I_{PUR}^k \end{pmatrix}^T \qquad \text{(Eq 5.4)}$$

The coefficients defined in (Eq. 5.3) and (Eq. 5.4) are counter parts of the coefficients in (Eq. 5.1) and (Eq. 5.2). Their size should be identical if the same fitting order and the same supply sweeps are used for the SRRV and the BWTV estimation. With the six predictors, five supply sweeps, and the second order fitting, the regression result of the first group is shown in Figure 50(a). As the $BWTV_L$ and the $BWTV_R$ specify the maximum tolerable noise voltage on each bit-line, the BWTV is defined as minimum of these two. Figure 50(b) is the estimation result by applying (Eq. 5.3) and (Eq. 5.4) with the coefficients found from the reference group. Both of the two groups show R^2 as high as 0.95. From Figure 50(b), the normalized estimation errors of the BWTV can be calculated, which is drawn in Figure 51(a). The estimation error is roughly Gaussian with the standard deviation of 3%, which is close to the error sigma of the SRRV estimation. In Figure 51(b), the CDF difference between the simulated and the estimated BWTV is approximately 10mV (for 1V nominal V_{DD}) which corresponds to 5% estimation error. From the small estimation error, it is concluded that the pull-down and pull-up currents contain sufficient information of the device variations.

(a)   (b)
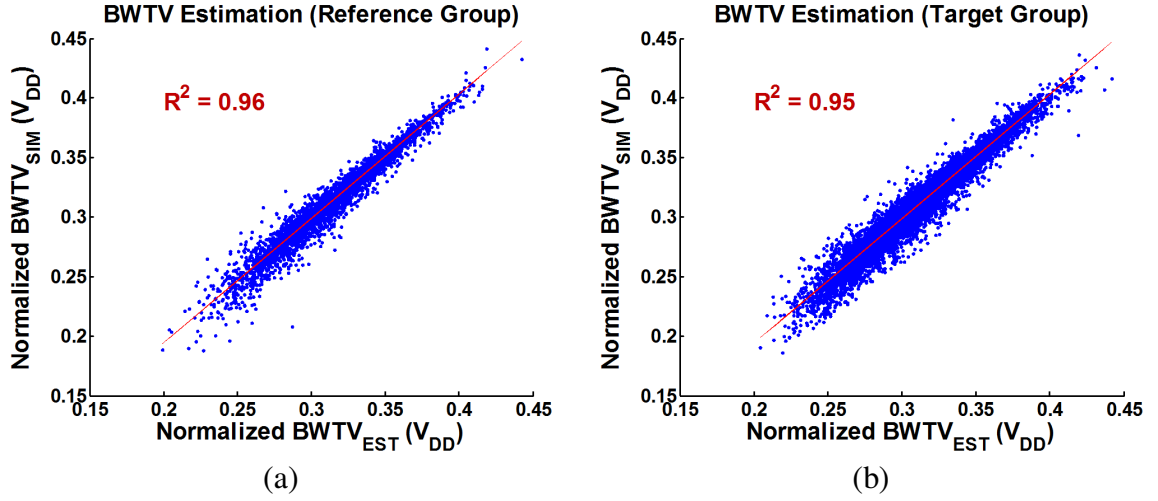
Figure 50.   BWTV estimation from the cell current measurement. Six types of currents (I_{PD}, I_{AC}, I_{PU}, and L-R for each) with five supply steps are used. From (Eq. 5.3) and (Eq. 5.4), n = 4,000, m = 5, and k = 4. (a) BWTV versus its estimator for 4,000 MC runs (regression data set). (b) Another 12,000 MC runs (target data set) with the estimation model (Eq. 5.3) and (Eq. 5.4).

(a)   (b)
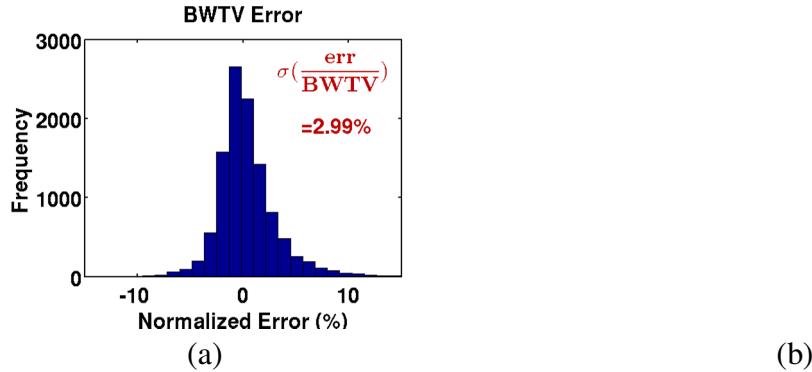
Figure 51.   (a) Normalized error (in %) histogram for the target data group. (b) Cumulative distribution of the simulated BWTV and the estimated BWTV.

5.3.2.   Sensitivity of BWTV to Cell Currents

The sensitivity of the BWTV to the cell currents are depicted in Figure 52. The total number of supply sweeps is two (m=2) and a quadratic equation is used for the regression (k=2). The write-ability is strongly affected by the pull-down strength of the access device to force the high state node to the ground. In Figure 43(b), this pull-down strength is created by N_{ACL} in association with P_{PUL}, denoted by I_{PUL} in Figure 52. As I_{PUL} at the nominal V_{DD} is smaller, the pull-down strength is weaker hence leading to a reduced write-ability or a reduced BWTV. On the other hand, if I_{PUL} varies significantly for the cell supply change from the nominal V_{DD} to the critical V_{DD}, P_{PUL} must be weak due to such high sensitivity to the cell supply. Therefore the flipping of the state can easily take place as seen in Figure 52 (higher BWTV). The amount of the voltage drop at V_L, created by the pull-down, is regenerated by N_{PDR}. If N_{PDR} is strong, the gate voltage of P_{PUL} does not change sufficiently and the regeneration may fail. The contribution of N_{PDR} to the write-ability can be clearly seen from Figure 52. If I_{PDR} at the nominal V_{DD} is small, the BWTV improves as N_{PDR} is weak. If I_{PDR} shows less supply sensitivity, N_{PDR} is strong and the BWTV   degrades.
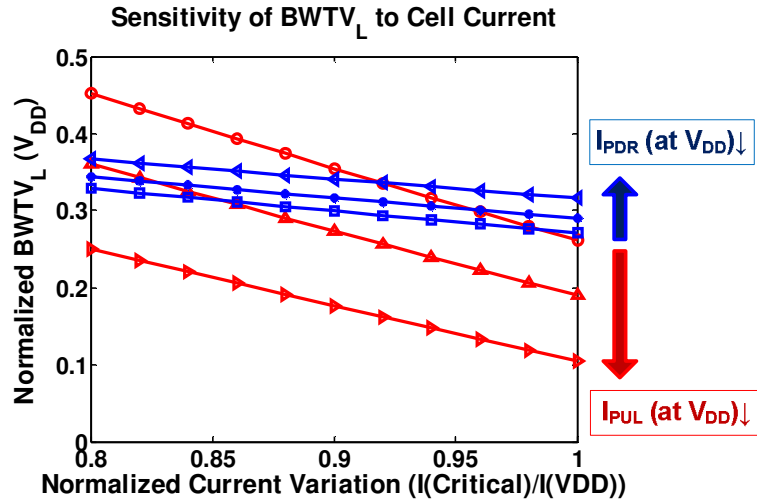
Figure 52.   Sensitivity of $BWTV_L$ to the cell current (m = 2, k = 2).

5.3.3.   Number of Predictor Variables

The sensitivity to the number of predictor variables and the resulting regression complexity are similar to that of the SRRV. Figure 53 shows how R^2 of the $BWTV_L$ varies as a function of the number of predictor variables with changes of (a) the order of the polynomial functions, k, and (b) the step size of the supply sweeps, m. The predictor variables are ordered from the maximum to the minimum impact on the $BWTV_L$.

In most cases, $BWTV_L$ is determined by I_{PUL} and I_{PDR}. The addition of I_{ACL} helps characterize the variations of N_{ACL} and P_{PUL}. The remaining three predictors have very small impact on the $BWTV_L$. Therefore, compared to the SRRV estimation, the   BWTV estimation is less complex with the reduced number of the predictor variables to achieve similar R^2.

(a)   (b)

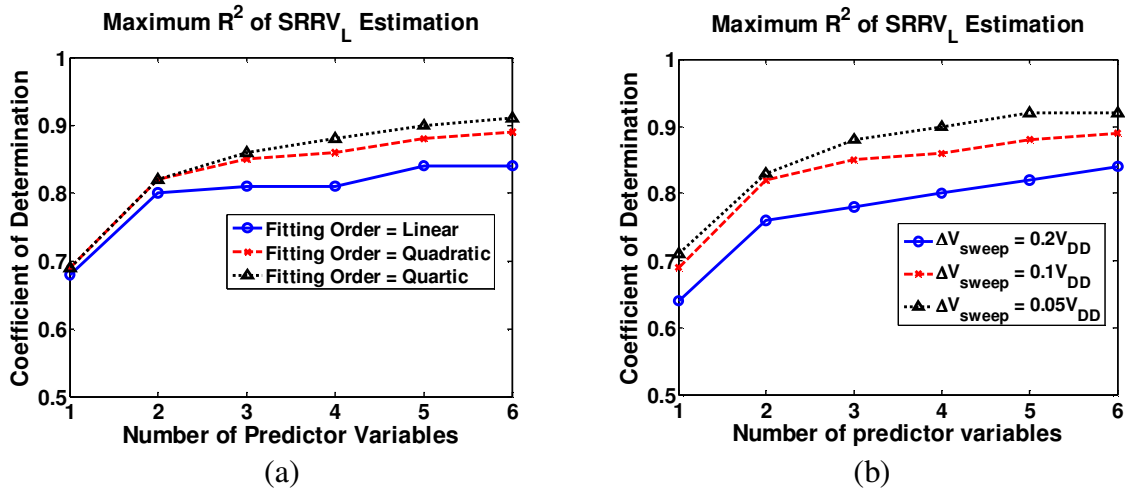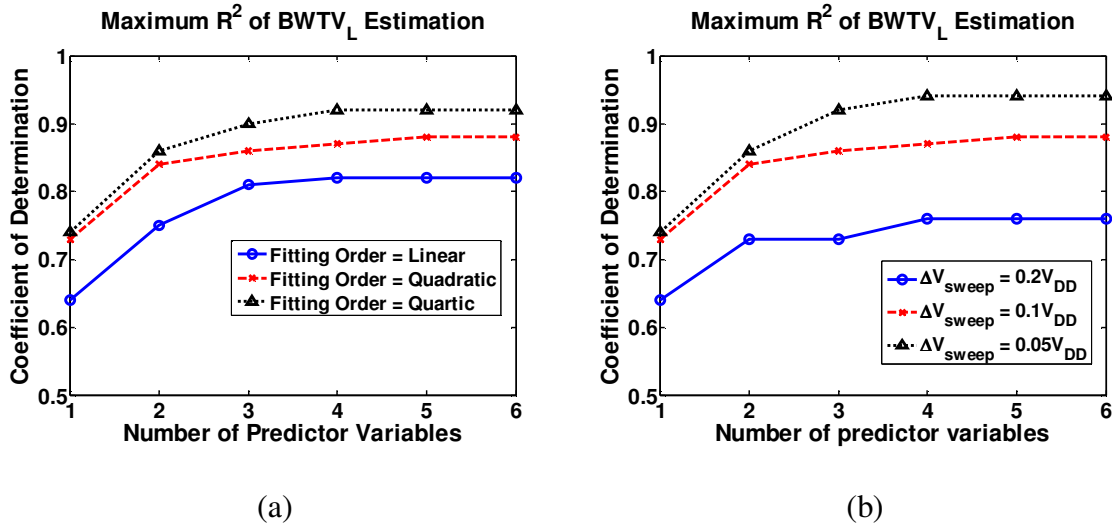Figure 53.   Coefficient of Determination (R^2) for $BWTV_L$ estimation with respect to the number of predictor variables (Table 5-2). (a) With $\Delta V_{sweep} = 0.1 \times V_{DD}$, the fitting order (k) is varied from 1 to 4. (b) With k = 2, ΔV_{sweep} is varied from $0.05 \times V_{DD}$ to $0.2 \times V_{DD}$. The sweep range is fixed by $0.2 \times V_{DD}$ and the fitting order is 2 (square).

Table 5-2 Predictor Variables used in Figure 53.

	Number of Predictor Variables					
	1	**2**	**3**	**4**	**5**	**6**
Predictor Variables used in (3) and Figure 53	I_{PUL}	I_{PUL} I_{PDR}	I_{PUL} I_{PDR} I_{ACL}	I_{PUL} I_{PDR} I_{ACL} I_{PDL}	I_{PUL} I_{PDR} I_{ACL} I_{PDL} I_{ACR}	I_{PUL} I_{PDR} I_{ACL} I_{PDL} I_{ACR} I_{PUR}
Predictor Variables used in (4)	I_{PUR}	I_{PUR} I_{PDL}	I_{PUR} I_{PDL} I_{ACR}	I_{PUR} I_{PDL} I_{ACR} I_{PDR}	I_{PUR} I_{PDL} I_{ACR} I_{PDR} I_{ACL}	I_{PUR} I_{PDL} I_{ACR} I_{PDR} I_{ACL} I_{PUL}

Section 5.4   Discussion

In this section, the estimation approach is validated for different device ratios, process corners, temperatures, and the power supply levels. Unless otherwise noted, the estimation model uses the same condition as Figure 45 and Figure 50 where all the six predictors are used with 5 supply sweeps (m=5) and the second order fitting (k=2). The measurement limitation due to the finite resolution of the ADC and the supply control is discussed in the last sub-section.

5.4.1.   Device Ratio

The device ratio used in the previous example is 1.5 for the cell ratio (PD/AC) and 0.8 for the pull-up ratio (PU/AC). The estimation accuracy with variable device ratios is depicted in Figure 54. Figure 54(a) is the SRRV estimation of the target group for different cell ratios. There is not a noticeable change in the estimation result. The BWTV

estimation result is also shown in Figure 54(b) with different pull-up ratios. These two figures indicate that the polynomial fit appropriately accounts for device ratios.
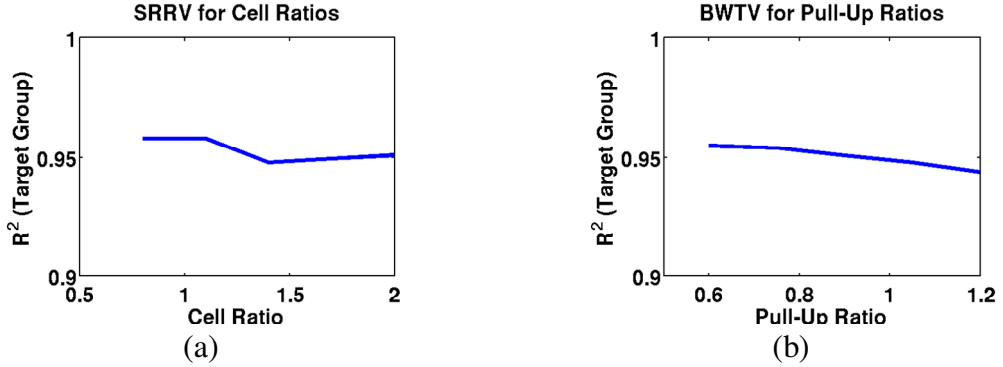
Figure 54.   Correlation of the estimation for different device ratios. (a) SRRV estimation with cell ratio sweep. (b) BWTV estimation with pull-up ratio sweep. Six predictor variables, $\Delta V_{sweep} = 0.05 \times V_{DD}$, and k = 2.

5.4.2.   **Global Process Variation**

The estimation results so far focus on random local mismatch and do not account for variation across dice or wafers [2-3]. The estimation results for different process corners are shown in Figure 55. In Figure 55(a), the first result set is based on the standard current measurement as explained in Section 5.1. At the SS, SF, and FS corner, the R^2 drops substantially. According to the simulation data at those corners, the internal node voltage (V_L or V_R in Figure 42) is not precisely determined by the cell supply during the pull-down current test. As discussed in section 5.1.2, if the node voltage is set by the feedback devices and not by the cell supply, the pull-down current would contain more device variation hence leading to lower estimation accuracy.

92

(a)   (b)
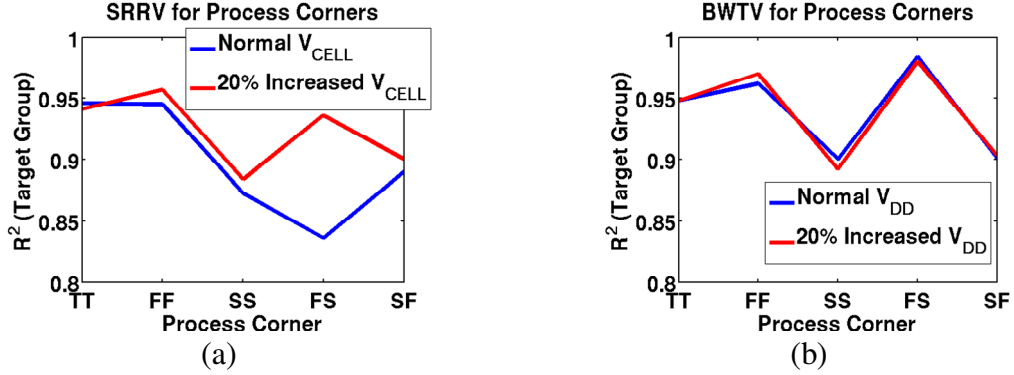
Figure 55.   Correlation of the estimation for different process corners. (a) SRRV
estimation with corner variations. The second case increased the cell supply
voltage by 20% higher than the nominal level during the pull-down current
measurement. (b) BWTV estimation with corner variation. Six predictor
variables, $\Delta V_{sweep} = 0.05 \times V_{DD}$, and k = 2.

To avoid the internal nodes being arbitrarily settled, one minor modification to the
estimation is to increase the cell supply level during the pull-down current measurement.
By using a 20% increase in V_{CELL} to measure the pull-down current while keeping other
measurements the same, the result of the FS corner improves to be comparable to the
typical corner. The estimation result can be seen as the second data set in Figure 55(a).
The other two corners at SS and SF are not improved. As mentioned previously in Figure
33, the SRRV measurement is less accurate for the stable cells and both corners have
higher read stability. It is noteworthy that while the accuracy is degraded across corners,
an R^2 of 0.9 is still sufficiently accurate such that this technique can be viable for the cell
stability estimation across wafers using regression parameters from a subset in a run of
wafers.

The BWTV estimation results at different corners are displayed in Figure 55(b).
Compared to the SRRV estimation, the BWTV has lower process sensitivity on the
estimation result. As the BTWV estimation is highly dependent on the pull-up current,
increasing the cell supply for the pull-down current does not enhance R^2.

93

5.4.3.   Temperature Variation

To reduce the sensitivity to temperature variation, estimation of the stability and the cell currents is accurate when measured at the desired operating junction temperature. The estimation results with temperature sweeps are displayed in Figure 56 for wide operating temperature ranging from -20ºC to 85ºC. Both of the SRRV and the BWTV estimations show reliable accuracy. However, there is a loss of accuracy if the fitting coefficients are extracted at one reference temperature and used for prediction at a different temperature. The regression model should be simplified to prevent over-fitting. As an example of the SRRV estimation, three supply sweeps with the second order fitting (k=2) leads to 0.85~0.88 R^2 variation for the temperature range of 20~60ºC. With the same condition, the BWTV's R^2 varies from 0.85~0.87.

(a)   (b)
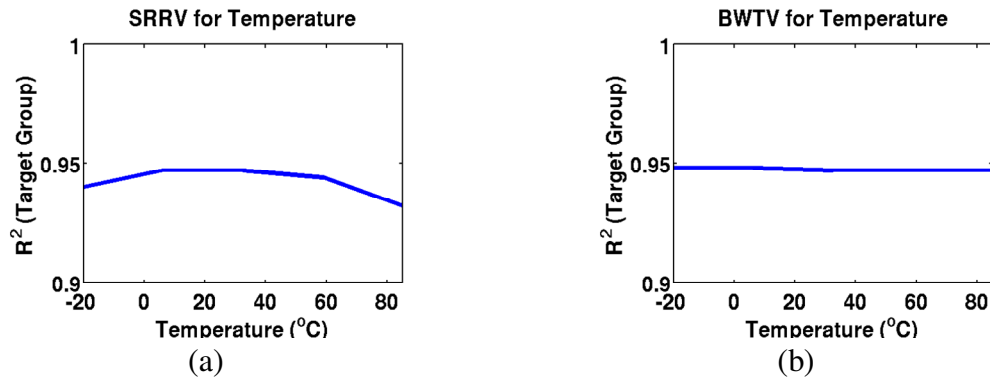
Figure 56.   Correlation of the estimation per temperature sweeps. (a) SRRV estimation and (b) BWTV estimation for -20ºC~85ºC. Six predictor variables, $\Delta V_{sweep} = 0.05 \times V_{DD}$, and k = 2.

5.4.4.   Supply Level

The estimation technique is verified for a single supply level (nominal V_{DD}). The supply levels (cell supply, word-line driver, and the bit-line driver) are scaled down by 40% to test if the identical measurement data can be used for estimation of the stability at such reduced supply level. Figure 57 shows the estimation results for the SRRV and the BWTV. The cell-current measured at the nominal VDD still fits well to the stability information at the reduced supply level.
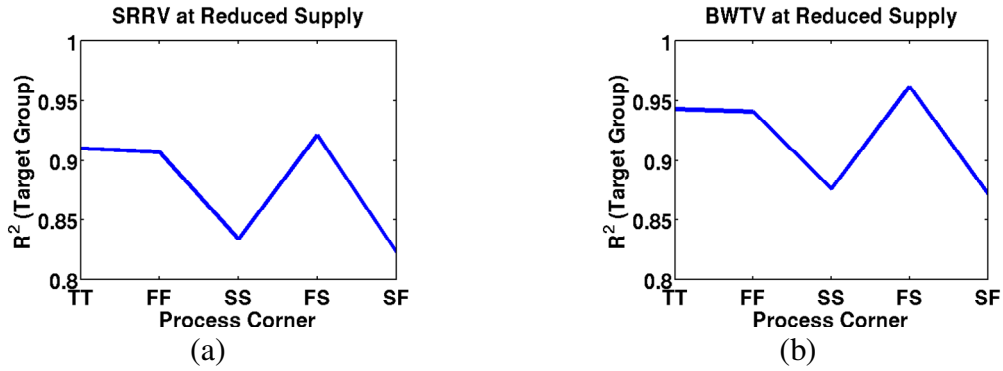
Figure 57.   Correlation of the estimation for different process corners at a reduced supply level (40% reduction from the nominal V_{DD}). (a) SRRV estimation and (b) BWTV estimation. Six predictor variables, $\Delta V_{sweep} = 0.05 \times V_{DD}$, and k = 2.

5.4.5.   Measurement Limitation

This section provides analysis on two types of quantization noise from the measurements of the stability (SRRV and BWTV) and the cell currents. To obtain the data set for the reference group, the supply voltage is lowered by a fixed step ΔV_{CELL} (or ΔV_{BL}) until read failure (or write success) is detected. The controllability of the supply voltage (ΔV_{CELL}) involves a tradeoff between the measurement time and the estimation

accuracy. Figure 58 shows an example that used 20mV as the minimum supply step in the measurement. Interestingly, the SRRV quantization noise has a small impact on the estimation results since the regression can be done based on the distribution of the SRRV. However, if the quantization step is too large such that the distribution is not clearly seen from the SRRV data, the regression fails and the errors start to increase rapidly.
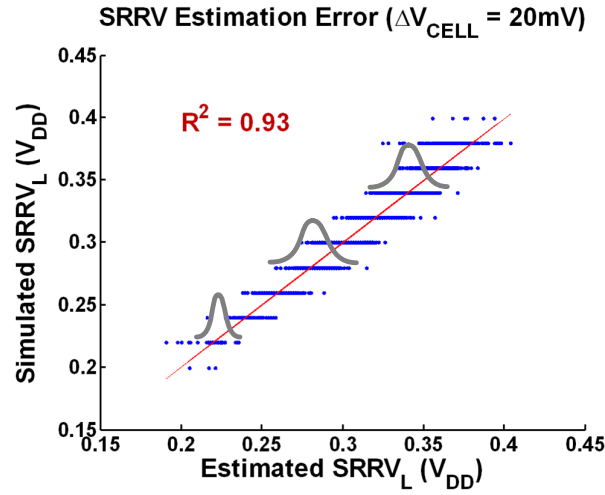
Figure 58. $SRRV_L$ estimation error with discrete supply control (ΔV_{CELL} = 20mV).

The amount of the cell current can be measured by an N-bit ADC and would have the associated quantization noise. Figure 59 shows the estimation accuracy of the SRRV and the BWTV per ΔV_{CELL} (or ΔV_{BL}) and the ADC resolution. Figure 59(a) and (b) shows that the SRRV estimation is generally insensitive to the supply control up to 50mV. To obtain R^2 higher than 0.9, the required ADC resolution is larger than 10 bits.

The BWTV shows a higher sensitivity on the ADC resolution. As the BWTV is strongly affected by the pull-up current that is usually much smaller than the pull-down

current, the quantization noise imposed on such tiny current variation can quickly degrade the estimation accuracy.

Figure 59 clearly indicates that the finite resolution of the supply control and the measurement ADC limits the improvement from increasing of the number of the predictor variables. Practical estimation target including the circuit complexity can be focused on around 0.8 as the R^2 with a reduced number of measurements such as 4 predictors with 3 supply sweeps.

(a)   (b)

(c)   (d)
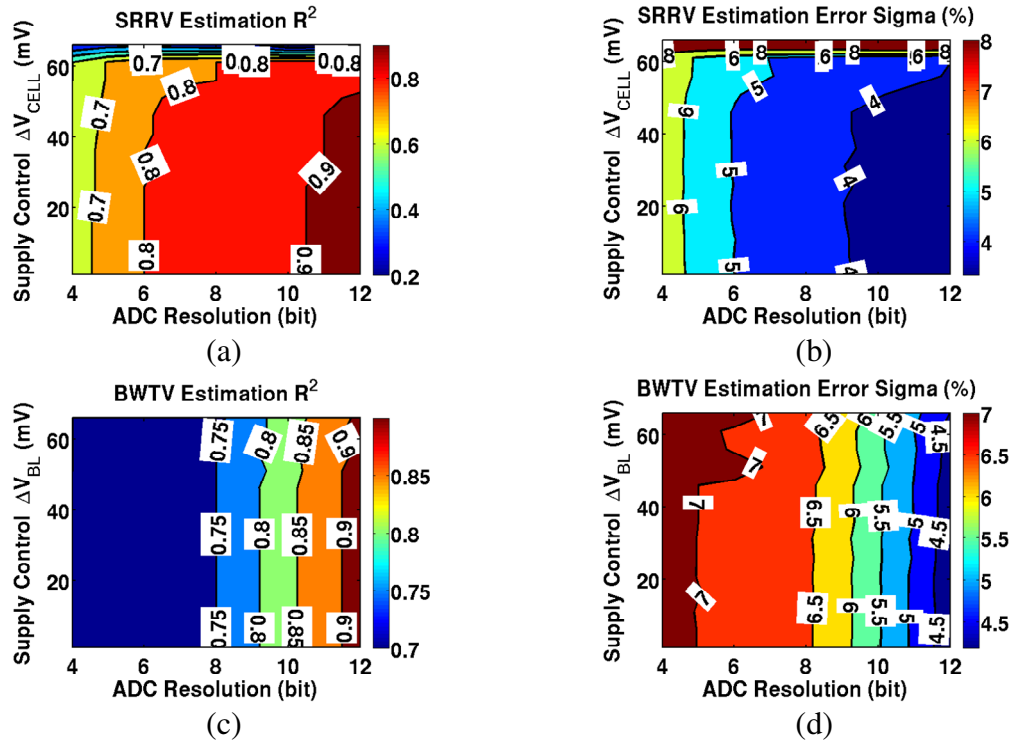
Figure 59.   Estimation results of the target group for the SRRV and the BTWV for the supply control and the ADC resolution. (a) R^2 of the SRRV estimation. (b) Error sigma of the SRRV estimation. (c) R^2 of the BWTV estimation. (b) Error sigma of the BWTV estimation. With 6 predictor variables, $\Delta V_{sweep} = 0.05 \times V_{DD}$, and the fitting order is 2 (k=2).

97

Section 5.5   Conclusion

The stability of a 6T-SRAM cell is estimated using a nonlinear regression from cell currents under variant supply levels. To apply this technique in a real IC, the modified measurable failure conditions (SRRV and BWTV) are considered. The static cell current measurements show the best read stability and the write-ability estimation results with R^2 = 0.95 when all the six predictor variables are used with the 5 supply sweeps and the 4th fitting order. The estimation also shows a small error sigma less than 3%. Since the limitations of the measurement circuits place the practical bound on R^2 (e.g. 0.8), a much reduced number of measurement data set can be used such as a reduced number of the predictor variables (without I_{AC}), a lower order fitting (2nd), and a smaller supply sweeps (3). From the estimation model, any specific SRAM cell can be tested to quantify the likelihood of failure. This technique can be implemented in a modern CMOS SRAM chip with small extra cost for the area of the measurement circuitry without change of the SRAM array structure. Based on the discussions made in the Chapter, the next Chapter presents a test chip implementation in 65nm CMOS. The measured data shows promising estimation results.

Chapter 6

Measurement Data for SRAM Stability Estimation

This Chapter extends the previous discussions with actual circuit implementation in a 65nm CMOS technology. The entire arguments made in Chapter 5 still apply to the estimation setups for the actual measurement data except for the supply sweep range for the pull-down current. In Chapter 5, the supply sweep is limited by the critical V_{DD} where none of the cells flips. As $V_{Critical}$ is typically close to the nominal V_{DD}, the supply sweep range is on the order of few hundred mVs. From the simulation data, where the cell current and stability can be measured without any internal or external error such as noise and line impedance, the estimation worked very well with such limited sweep range. However, the actual measurement is highly affected by diverse source of errors. These error sources are described in section 6.2 with circuit details to minimize their effects. Another major reason of the error is that the pull-down currents measured near the nominal V_{DD} do not properly predict the failure supply levels for the stable cells that demand huge supply sweep > 400mV.

This fact implies a major discrepancy between the device model and actual device characteristics. The conventional device model like BSIM4 still relies on the threshold

voltage based equation with a substantial set of fitting coefficients that are useful in predicting the worst-case boundary of the device variations. While the major device parameters can be effectively extracted from the simulation data to improve the stability estimation technique, the actual individual device may not be fully described by those fitting coefficients. Consequently, the estimation was not accurately done with small supply sweep. As an example, the maximum R^2 of 0.7 is obtained from 150mV supply sweep range. The R^2 dropped by more than 0.25 compared to the simulation results. To enhance the estimation accuracy, the supply level is swept down to 0.5V for the current measurement. By measuring the pull-down current at the supply level where the stable cells flip, the R^2 increased over 0.8.

The next section briefly summarizes the stability estimation concepts introduced in Chapter 5 with different supply sweep range. The following two sections address the overall architecture and design issues. Section 6.3 provides measurement data for static stability estimation. The fundamental correlation between the static and dynamic stability is discussed in section 6.4. Finally, die-to-die variation of the stability estimation results (static and dynamic) is investigated.

Section 6.1   Stability Estimation

6.1.1.   Overview on Stability Estimation

The SRAM cell's characteristics can be obtained from the measured current created by the pull-down or pull-up device in combined with the access transistor. As the current contains multiple device information, a single measurement is not sufficient as a primary indicator of either read stability or write-ability. If the currents are measured under multiple supply levels, however, the combined data show an excellent estimation result for any stability metrics such as RRV, WTV, or even the conventional RSNM. This Chapter focuses on estimation of SRRV and BWTV as they are useful in determining of

the cell supply and bit-line level[3] for successful read and write access.

The read/write stability of an SRAM cell is a function of the characteristics of multiple devices (e.g. six degree of freedom for a 6T cell). The literatures on the stability analysis have shown that the contribution to the stability variation from each device can be modeled as a nonlinear function of the device parameters such as the threshold voltage. Extraction of such individual device parameters from the dense memory arrays may demand substantial area overhead due to extra wirings and transmission gates to access the internal nodes of an SRAM cell.

Alternatively, the read/write current from a selected cell can be measured by accessing the bit-line that can used to characterize the parameter variation of each device. Since the read/write current is also a function of multi-dimensional random variables, a single current measurement is not sufficient to determine either the read-stability or the write-ability. For precise extraction of the device parameters, the cell currents need to be measured under multiple supply levels. The collected current data set are combined to accurately estimate the stability metrics such as the read-stability (RRV or RSNM) and the write-ability (WTV). This paper focuses on the estimation of the supply read retention voltage (SRRV) and the bit-line write trip voltage (BWTV) as they are useful in determining of the necessary cell supply and the bit-line level for successful read and write access.

In the previous chapter, a non-linear function was proposed to translate the current data set into the stability metric. The function has arbitrary polynomial expansions of the measured current variables with undetermined coefficients. The coefficients can be found using regression with a complete set of the measured currents and the measured stability data. In a large memory array, a subset of the array should be tested first to obtain the coefficients. The size of the subset should be as large as 1,000 cells to reduce the incorrect coefficient extraction. However, increasing the subset size would not help in improving the overall estimation accuracy as the maximum bound is more heavily limited

---

[3] The word-line level for successful write access can be gauged by WWTV. BWTV and WWTV are highly correlated [xx].

by the other sources of error such as the measurement noise from ADC.

The estimation procedure is conceptually visualized in Figure 60. At the first step, the bit-line currents are measured under different supply levels. This measurement can be rapidly done by help of on-chip ADC with intermediate resolution such as 8~10 bits. The choice of the on-chip ADC is discussed in the next section. Next, 1,000 cells are selected and their stability is measured not precisely but coarsely (e.g. 25mV step). This 1,000-cell group is called a regression group. From the regression group, the relation between the measured currents and the stability is formulated. This formula is applied to the remaining cells to estimate the stability.

The regression set should be a sample of the cells that covers the worst case variations across an array. The procedure for determining the regression set for stability is shown in **Error! Reference source not found.**Figure 61. Initially the stability (either SRRV or BWTV) of the entire memory array is measured by stepping down the control variable (cell supply or bit-line voltage) at 25mV steps. The stability does not necessarily be measured precisely as the regression formula in section II.D can handle the quantization noise such as 25mV step size. The measured SRRV or BWTV may show a slightly distorted Gaussian distribution as seen in Figure 61. If the regression is done with the entire data set, the estimation result is focused at the center of the distribution leading to reduced accuracy at the tails. For the linear estimation over the entire stability range, the regression data set is selected from the N uniformly spaced bins. In this paper, the total number of bins (N) is five. The maximum samples at each bin are limited by 1000/N.

The sampling methodology appeared in Figure 61 is focused on result precision. We have chosen this sampling method to apply the extracted formula to the other dice (hence, the errors in the coefficients must be small). In fact, the stability measurement over the entire array, although it is measured coarsely, may sound contradictory to the fast estimation. If the stability estimation is for a single die, a normal write/read test under multiple supply levels also leads to close sampling results especially for the less reliable cells. Using this faster sampling method, the selected subset's characteristics can be

quickly tested to estimate the stability over the entire array.
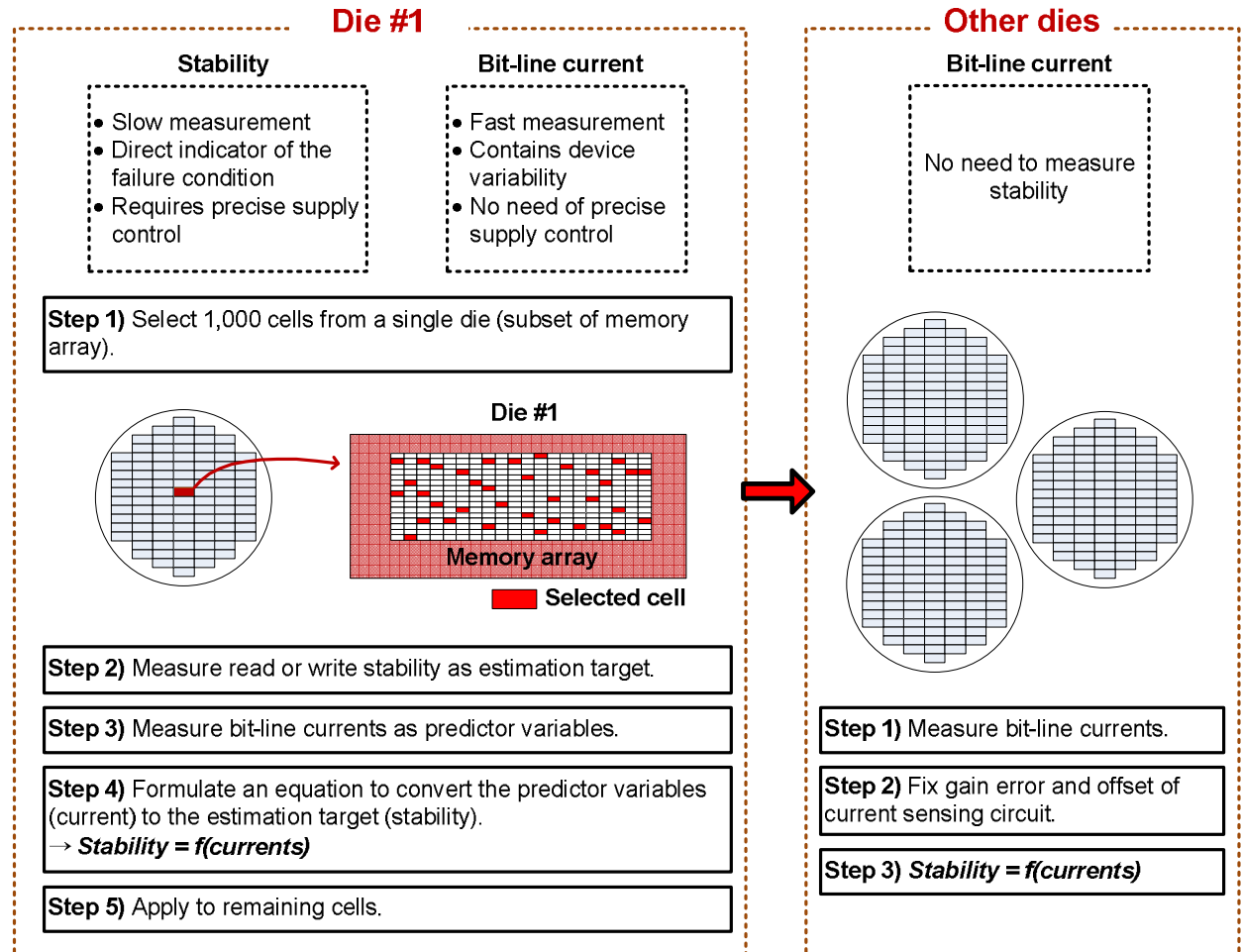
Figure 60.   Flow diagram for the stability estimation from the measured currents. A 1,000 sample group (regression group) is selected to formulate the relation between the bit-line currents (predictor variables) and stability (estimation target). This formula can be applied to the other dies to estimate their stability.
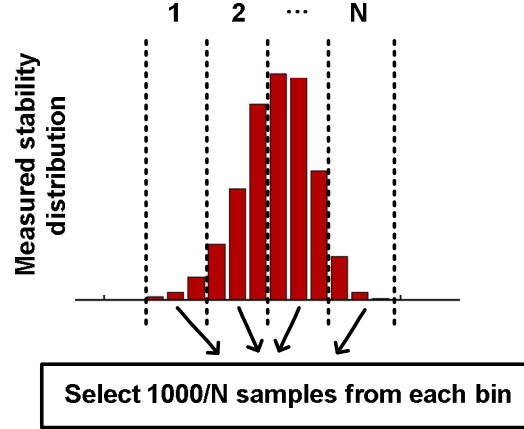
Figure 61.   Uniform sampling of the subset of the memory array during step 1 in Figure 60. In this paper, N is 5 for the read-stability (SRRV) and write-ability (BWTV).

6.1.2.   Current Measurement

The read/write currents are measured for the cells in the regression set to find the estimation function. Figure 62 shows the measurement setups for the pull-down and pull-up current for the read stability and write-ability respectively. For the pull-down current measurement (or equivalently read current), BL (storing '1') side is tied to the nominal pre-charge level (e.g. 1V) while the other bit-line is tied to the cell supply. The current is measured by directly accessing the bit-line at the end of the memory array. The cell supply (V_{CELL}) is varied from the nominal supply level (1V) to 0.5V while the two bit-lines are tied to 1V. The number of supply steps between these two supply levels determines the estimation accuracy at the cost of measurement time and it should be chosen to avoid over-fitting. Section 6.3 provides the estimation result with 4 supply steps (150mV step size) and 6 supply steps (100mV step size) for comparison. The pull-up current (or equivalently write current) in Figure 62 (b) quantifies the strength of the pull-up device that resists against the write access. The bit-line of the pull-up side is fixed by 0.4V while the other bit-line is tied to 1V. The pull-up current need less data and is

measured at 2 V_{CELL} values of 1V and 0.7V (300mV step size).

It is important to note that the bit-line current can be as small as a few tens of micro-amperes as the supply level is lowered. Leakage current and long-term device noise arising from the sensing circuits may perturb the cell current measurement. Prior to measuring each column, a calibration current is measured without activating the word-line. This calibration current contains the leakage of the measurement path. The actual cell current is obtained by calculating the difference of the pull-down (or pull-up) current with this calibration current.

(a)   (b)
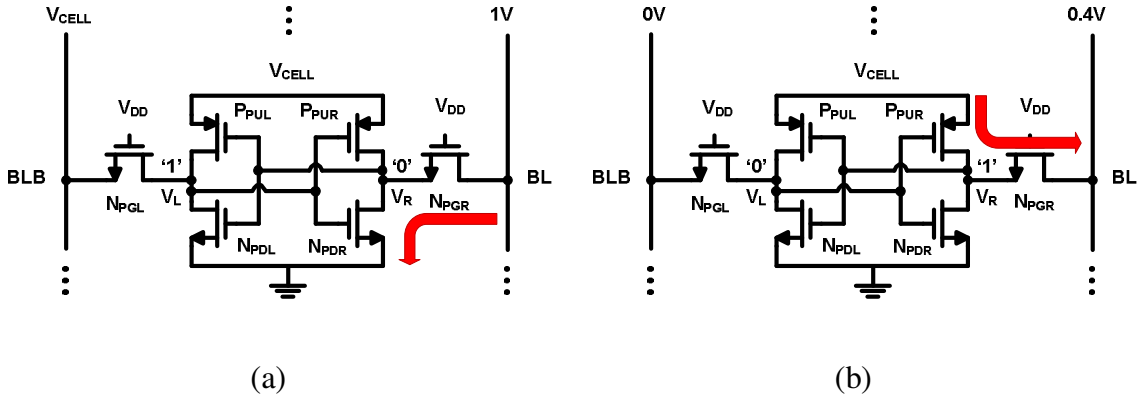
Figure 62.   Test setups for (a) pull-down current measurement and (b) pull-up current measurement. All the currents are measured using the direct bit-line access.

6.1.3.   Estimation Formula

The measured currents at each of BL and BLB side are correlated to a stability metric using a second-order polynomial equation (Eq. 5.1 ~ Eq. 5.4). There are four types of measurements per cell (I_{PDR}, I_{PDL}, I_{PUR}, and I_{PUL}) with 'm' sweeps of V_{CELL} for I_{PD} and 2 sweeps of V_{CELL} for I_{PU}. Note that the current measurement is required once and any stability (even at a different supply level) can be estimated from the same current data.

Section 6.2   System Architecture

6.2.1.   Chip Overview

A memory test chip is built in a 65-nm CMOS technology. Figure 63 shows the die micrograph and Figure 64 show the block diagram of the memory array. This chip is comprised of two banks of 16kb SRAM (DUT) with in-situ current-sensing circuits embedded in the array, on-chip ADC, memory control unit, and storage of the measured data. This section discusses each of these blocks. The specifications on the measurement circuits are based on a 0.8 R^2 for the estimation accuracy (of SRRV and BWTV).
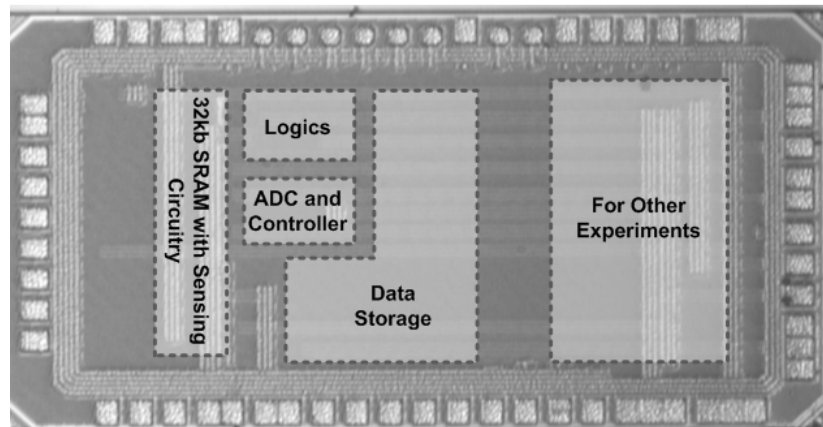
Figure 63.   Die microphotograph of the test chip. The chip area is 2×1.1 mm^2.
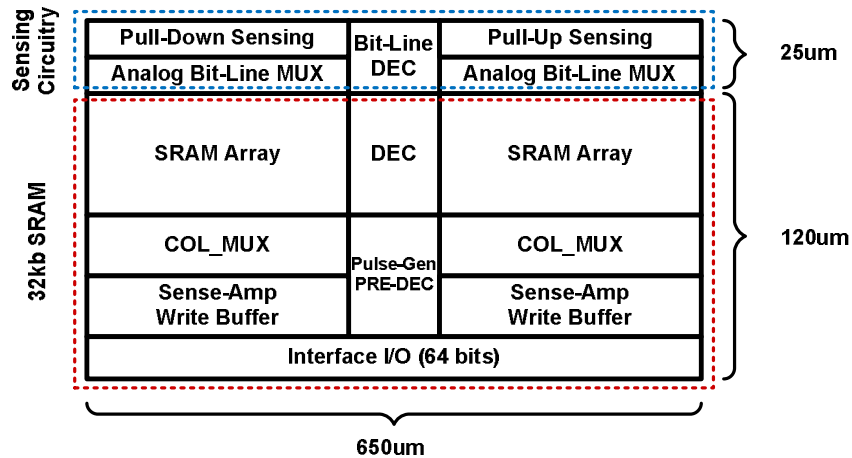
Figure 64.   Simplified block diagram of the main 32kb SRAM with the sensing circuitry.
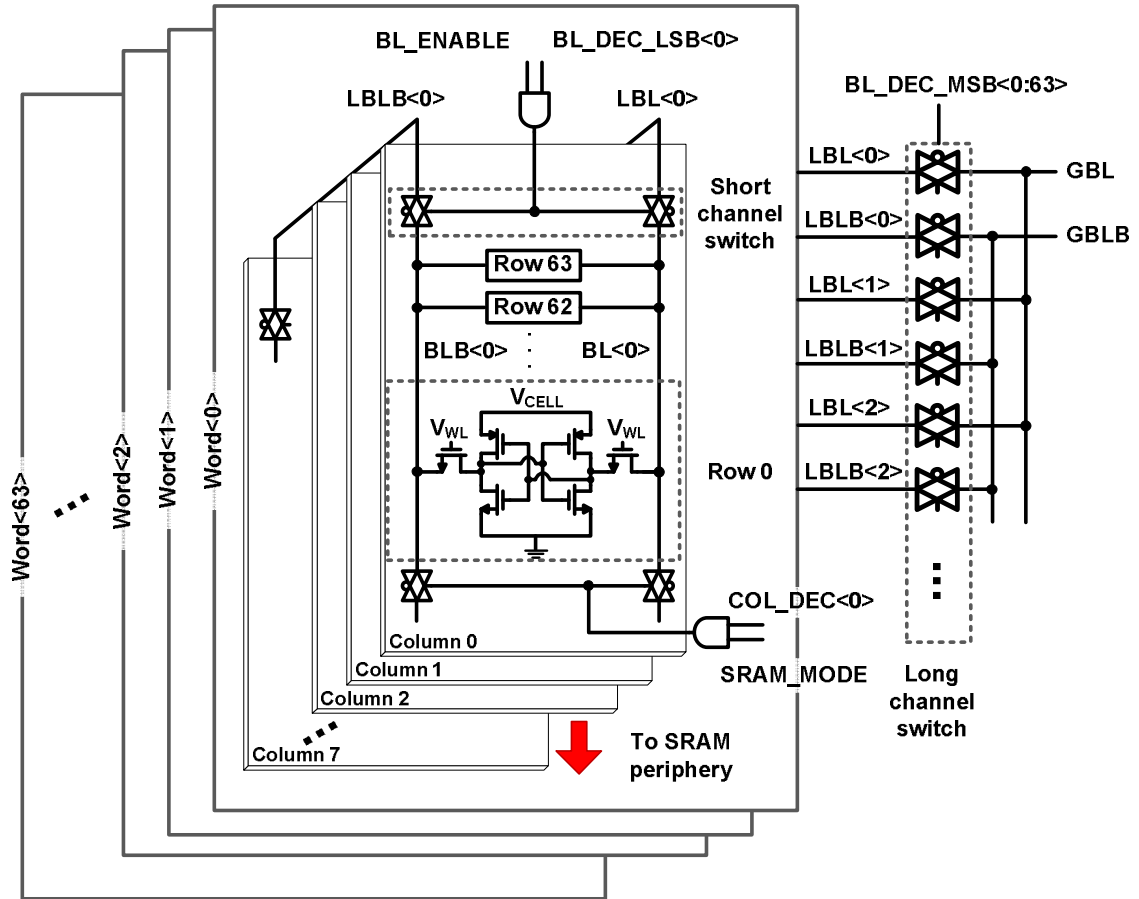
Figure 65.   The architecture of the memory array with the mode switches and the bit-line MUX arrays.

The SRAM array uses the 6T-cell available in the technology library. The word-line drivers and bit-line sensing and driving circuits are modified versions of the library elements. Figure 6 shows the detailed schematics of the memory and the interface. The SRAM operates normally as a memory when the mode selection signals (SRAM_MODE

= 1, BL_ENABLE = 0) connects the bit-line pairs to the sense amplifiers and write buffers. When SRAM_MODE = 0, the memory is configured in the current-sensing mode. The normal sense-amplifiers are disabled. At the top end of the bit-lines, an additional MUX array and sensing circuits are added to measure the bit-line current. A cell's current from one of the bit-lines of a sub-array is multiplexed onto one of the local bit-line pairs (LBL and LBLB) and then the LBL pair from each sub-array is multiplexed again onto global bit-lines (GBL and GBLB). To measure the write-ability metric, the bit-lines can also be connected to a variable voltage by tying the global bit-lines to a voltage generator and connecting the multiplexed switches appropriately. To minimize the impact on speed and performance, the switches at each end of the each bit-line pair are sized to minimize parasitic capacitance, leakage current, and voltage drop across the switch. The parasitic capacitance of the switch is less than 10% of the total bit-line capacitance with 64 cells on each column.

The word-line signal (V_{WL}) for measuring the cell current must be statically high. The wordline buffer is designed for both a dynamic pulse for a normal access and statically high for the stability estimation. This feature is used in Section 6.4 where we show the relationship between the static and dynamic stability.
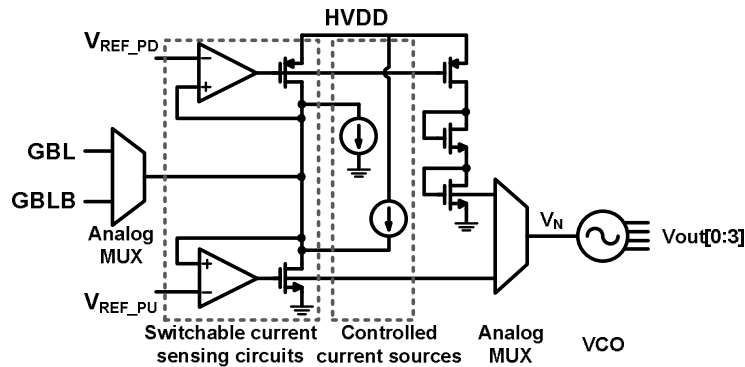
Figure 66.   Schematics of the current sensing units. Each pull-down or pull-up sensing circuit and its associated dummy current source can be switched on/off. The bit-line current is transferred to the voltage-controlled oscillator (VCO).

108

6.2.2.   Sensing Circuits

The pull-down and pull-up current sensing circuits are shown in Figure 66. The selected global bit-lines (either GBL or GBLB) are connected to a low-dropout regulator (LDO) that provides low output impedance to the subsequent circuit so that GBL and GBLB are not perturbed. Only one of the two amplifiers (and the associated current source) on the pull-up and pull-down paths is turned on to transfer the bit-line current of the proper polarity to the next stage. The current sources at the top and bottom branches of the sensing circuit are independently adjustable. Since the cell value can be flipped leading to very low bit-line currents, this current source stabilizes the feedback loop. The current source also enhances the linearity of the VCO-based ADC that follows the buffer. The area occupied by the bit-line multiplexer array is $650\times12\mu m^2$ (10% of the memory array) and the area of the sensing circuitry is $650\times13\mu m^2$. The area of the sensing circuitry can be reduced by using a simpler LDO amplifier.

6.2.3.   On-Chip ADC

The on-chip ADC is designed for 1.25MHz sampling rate, low power, small area, and 10-bit resolution. A VCO-based ADC is chosen. Even though this type of ADC has inherent non-linearity [120], the I/O characteristics are inherently monotonic since every cycle (and phase increment) accumulates as a thermometer-coded ADC. Furthermore, the estimation equations (5.1~5.4) can tolerate this second-order nonlinearity in the measurement data. Therefore, the linearity correction such as look-up table [121] is not required.

The schematic of the VCO-based ADC is shown in Figure 67(a). The measurement is done by applying a pulse of a fixed width such 400ns. As the input current is integrated during the measurement, the SNR improves due to sinc filtering of the wideband noise [121-122]. The phase difference captured at the rising and falling edge of the clock

determines the 3 LSBs while the number of cycles counted within the pulse determines the 7 MSBs. Figure 67(b) shows details of the oscillator. The cell current information buffered from the sensing circuit sets the bias voltage of the current source of the oscillator, hence determining the oscillation frequency. Note that measurement time can be reduced if using a linear ADC because only 8-bits of linear conversion levels are needed for a 0.8 R^2 during the read-stability estimation (Chapter 5). The area of the VCO and phase sampler is 32×48μm2.

(a)   (b)
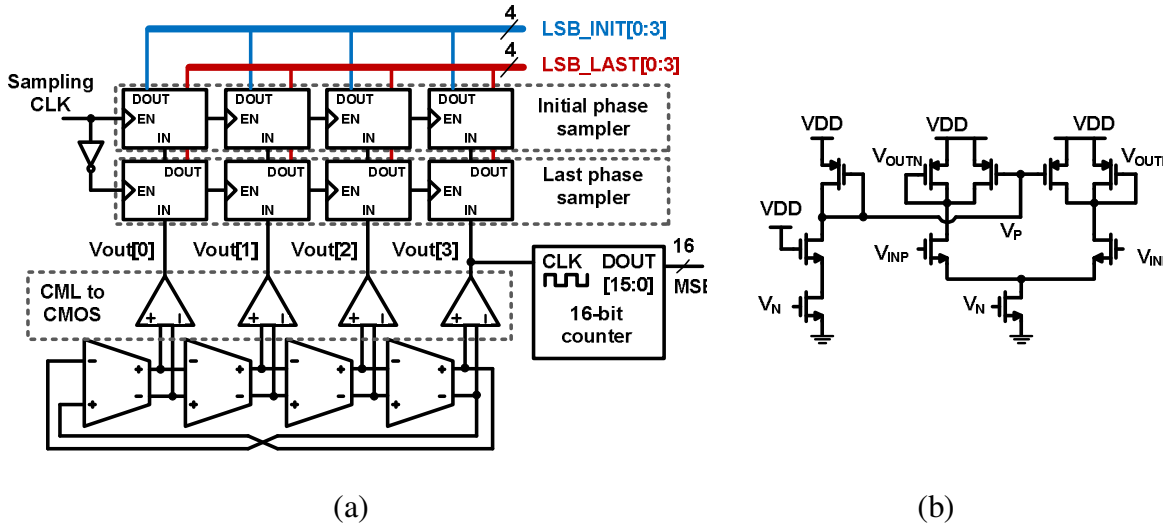
Figure 67.   (a) Schematics of the VCO-based ADC. The ring oscillator's initial and last phase determine the three LSBs while the number of cycles produces the MSBs. Depending on the target resolution bit, the counter can be made smaller.

(b) Schematics of the delay stage.

Section 6.3   Static Stability Estimation

6.3.1.   Read Stability Estimation Results

   The read-stability is estimated using the measured current data at different supply voltages (SRRV). Figure 68 shows the estimation results of the entire 32k cells using the coefficients from the regression set. The pull-down currents are measured at four supply levels (1, 0.85, 0.7, and 0.55V) while the pull-up currents are measured at two supply levels (1 and 0.7V). The coefficient of determination (R^2) is as high as 0.8 at the nominal supply voltage (1V). We also measured the SRRV at 0.8V, and 0.6V. The R^2 drops slightly but is still higher than 0.7 indicating that the current measured under the nominal supply reliably estimates the read-stability at a lower supply level. In Figure 69(a), the cumulative distributions of the estimated and measured SRRV at 1V supply match very well. From the estimated stability distribution, the tails of the distribution is accurately predicted. Figure 69(b) shows that the normalized estimation error has zero mean and 4.77% sigma. The same technique is applied with six supply steps (1, 0.9, 0.8, 0.7, 0.6 and 0.55V) and all other conditions are identical. The new estimation result increased R^2 only from 0.8 to 0.83 with the error sigma decreased from 4.77% to 4.39%. Considering the increase of the regression complexity and measurement time, increasing the number of supply sweeps more than four is not practically necessary.

   The estimation results are more heavily dependent on the noise. As the pull-down currents are measured with lowering the cell supply, the testing cell can be marginally stable at a low V_{CELL} level. With small noise injection, the cell may flip and no read current is measured. As the currents are measured on a few discrete supply levels, this random cell flipping can lead to substantial estimation error from the model (1). As shown in Figure 70, we measured the currents by multiple times (e.g. 10 times) under every supply level and the majority value is selected for the estimation. This way the effect of noise is eliminated but the entire estimation time increases accordingly. The pull-up current data do not require this type of majority selection as the testing cell is

111

more robust during the pull-up measurement.
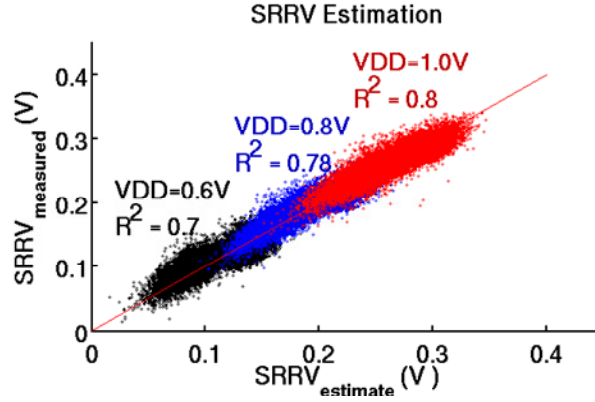
Figure 68.   SRRV estimation results of a 32kb memory array with four V_{CELL} sweeps for the pull-down and with two V_{CELL} sweeps for the pull-up. Correlation (R^2) of the measured and estimated SRRV for V_{DD} at 1.0, 0.8, and 0.6V.

<center>(a)   (b)</center>
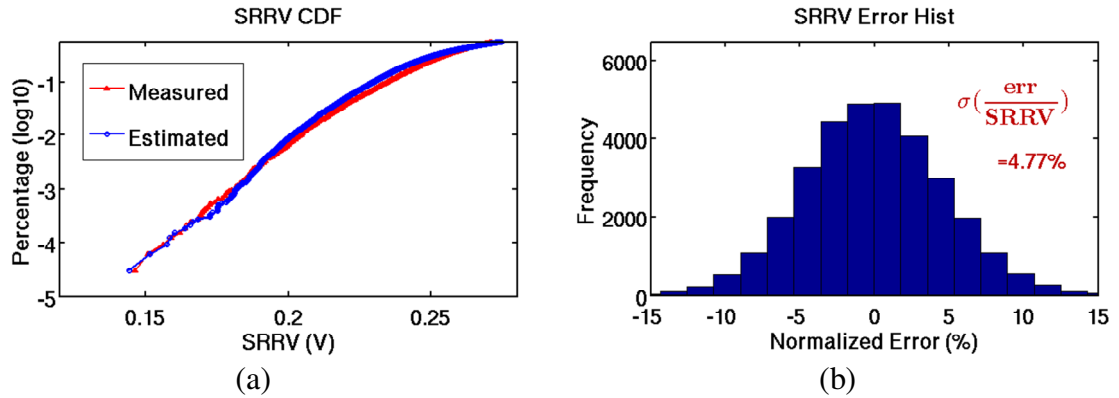
Figure 69.   (a) Cumulative distribution (CDF) of the measured and estimated SRRV at 1V supply. (b) Histogram of the estimation error at 1V supply.
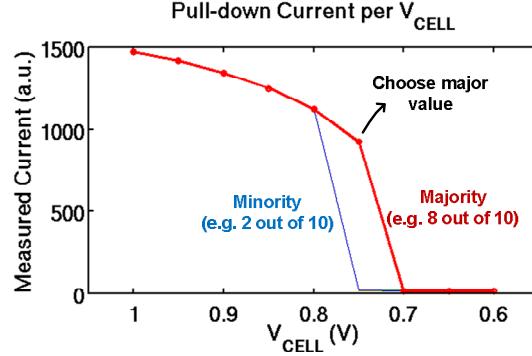
Figure 70.   Majority current selection from data. The cell near the stability edge can be flipped by the noise during the current measurement.

6.3.2.   **Write-Ability Estimation**

The write-ability estimation using BWTV is strongly dependent on the pull-up current. The estimation example in Figure 71 uses the pull-down currents at two supply levels (1 and 0.85V) and pull-up currents at 1 and 0.7V. Note that the current data used in the SRRV estimation can be shared for the BWTV estimation. Even with fewer current measurements, the BWTV estimation shows better estimation results when compared to the SRRV results due to a strong dependency of the metric on the pull-up current. In Figure 71Figure 72(a), the estimation R^2 is 0.91 at 1V supply and this number does not substantially decrease when the cell has lower supply of 0.8V. However, at 0.6V, the estimation is no longer sufficiently accurate because the transistors enter into weak inversion which is not properly estimated by using the currents measured at strong inversion.

In Figure 72(b), the CDF for the estimation and measured data shows slightly larger deviation when predicting the tails of the distribution. Nevertheless, the estimation error is bounded to several millivolts. In Figure 72(c), the error sigma of the BWTV estimation at 1V supply of only 1.29% indicating very good estimation of the write-ability.

113
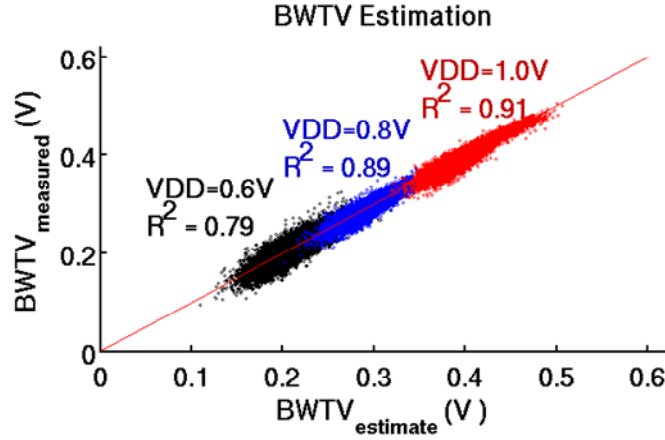
Figure 71.   BWTV estimation results of a 32kb memory array with two V_{CELL}
sweeps for the pull-down and with two V_{CELL} sweeps for the pull-up. Correlation
(R^2) of the measured and estimated BWTV for V_{DD} at 1.0, 0.8, and 0.6V.

(a)   (b)
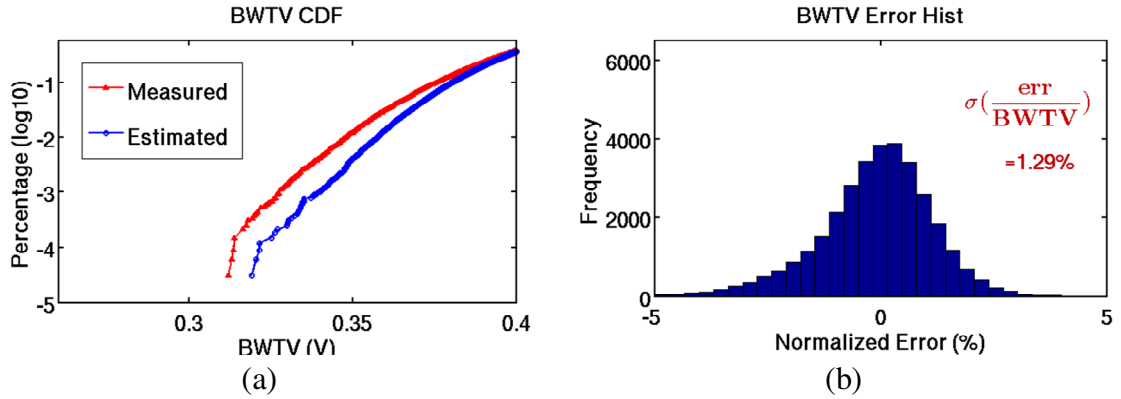
Figure 72.   (a) Cumulative distribution (CDF) of the measured and estimated
BWTV at 1V supply. (b) Histogram of the estimation error at 1V supply.

6.3.3.   Impact of the ADC Quantization Noise

Figure 73 shows the impact of the ADC resolution on the stability estimation from
another die measurement results. In this plot, the current data are initially measured by

114

10-bit resolution, and then the MSBs are taken to effectively model a lower resolution ADC. As the measurement time does not change, the noise filtering effect described in Section 6.2.3 is still valid and we can observe the pure impact of the ADC quantization noise without changing the magnitude of the device or supply noise. Figure 73(a) visualizes that the ADC resolution is not a primary factor in improving the read-stability estimation accuracy. Rather, above 8 bits the estimation error is limited by the other source of noise such as device and supply. Below 7 bits, however, the error rapidly surges up as the tiny cell current near the critical supply level (around V_{DD}-SRRV) may not be properly measured due to the limited resolution. Similarly for the write-ability estimation, the estimation accuracy does not improve for the ADC resolution above 8 bits (Figure 73(b)).

(a)   (b)
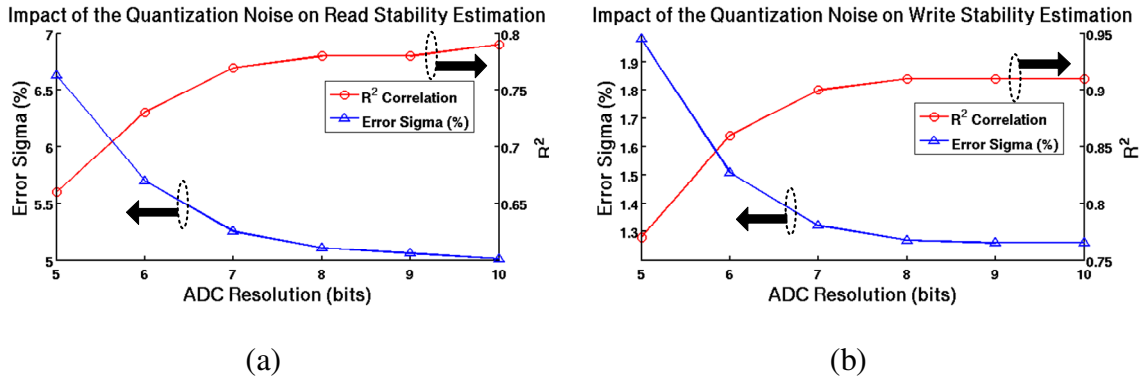
Figure 73.   Impact of the ADC quantization noise on the read stability estimation of a die. The error surges below 7 bits as the current information at a lower supply level is lost.

Section 6.4   Dynamic Stability

6.4.1.   Dynamic Read Stability

During a read, the cell is accessed with a word-line pulse of T_{WL} often with a pre-

charged bit-line [114-115], [123]. SRRV with a finite access time differs from the static measurement (static SRRV or SSRRV). Since the internal nodes of an SRAM cell has capacitance and has a transient response during a read, the cell may not be perturbed even when the supply is at the critical level (i.e. $V_{CELL}=V_{DD}$-SRRV).

To illustrate, consider a '0' node (V_R) that is charged by the access device during T_{WL} leading to a ΔV_R. The internal capacitance (C_R) is critical in determining the dynamic stability of the cell [75] since a larger C_R effectively prevents a large initial ΔV_R and subsequent ΔV_L. For a dynamically stable cell, since ΔV_R is not sufficiently large, the residual pull-down current I_{PDL}-I_{PUL} is too small to discharge V_L and change the state. For less dynamically stable cells, the residual current is larger and the internal capacitance is smaller leading to faster state changes during a read.

We follow the notation in Chapter 5 and define the value of V_{CELL} when the read is disturbed during a pulsed word-line access as dynamic SRRV (DSRRV). The definition of DSRRV is illustrated in Figure 74. This technique is an improvement from measuring the require time to flip the state [117] because it is substantially less sensitive to variations and can be related to SSRRV

The concept of DSRRV can be illustrated using state space trajectories [116]. Figure 75(a) shows simulated voltage transfer characteristics (VTC) of an SRAM cell during the hold and read access. The cell supply is slightly below the critical level ($V_{CELL} < V_{DD} -$ SRRV) such that the cell is statically unstable. The hold VTC divide the state space into two regions of convergence (ROC0 and ROC1) that regenerate any intermediate state to either S0 ("0" cell value) or S1 ("1" cell value). Figure 75(b) shows the three trajectories during the single read access with a step or unlimited pulse width (step response, solid line), $0.5 \times T_O$ (triangle), and T_O (circle and rectangle). For $T_{WL}=T_O$ (circle), the cell is marginally stable. If V_{CELL} is reduced by a small amount, the final state changes from S0 to S1 (rectangle). If the pulse width is half (triangle), V_{CELL} needs to be reduced further to flip the state as the distance between the return path of the trajectory and the ROC

boundary increases.

<center>(a)   (b)</center>

<center>Figure 74.   Dynamic SRRV test setup. (a) Schematics during the initial read</center>

<center>disturbance. (b) Testing waveform.</center>

<center>(a)   (b)   (c)</center>

<center>Figure 75.   (a) Simulated VTCs of an SRAM cell during hold and read access.</center>

(b) Trajectories of the internal voltage (V_L and V_R) during single read access with

variable pulse widths. The initial state S_0 is transitioning to S_1 as the pulse width

increases. (c) During repeated access with different duty cycle and a fixed pulse

width ($T_{WL} = T_O$). For all cases, V_{CELL} is slightly lower than the critical level (i.e.

<center>$V_{CELL} < V_{DD}$-SRRV).</center>

The difference of these two metrics can be visualized from the scatter plot of the

measured SSRRV and DSRRV in Figure 76(a). The word line pulse is generated

<center>117</center>

internally with $T_{WL}=1ns$ $(13\times\tau_{regeneration})$. The DSRRV is measured for the two different cases: loaded bit-lines and floating bit-lines. Only marginal difference is observable with floating bit-line showing slightly better stability. The reason is that since the bit-line capacitance (C_{BL}) is typically large for bit-lines with more than 64 cells, the transient voltage of the bit-line does not vary substantially during the narrow word-line pulse and hence there is little impact on the DSRRV.

As shown in Figure 76(a), for each SSRRV value, there is a distribution of DSRRV. The distribution is due to variation in the internal capacitances $(C_L$ and $C_R)$. The smaller the internal capacitance, the closer the DSRRV is to the SSRRV because the internal time constant of the cell has less impact. Along the left envelope of the DSRRV distribution, we can observe excellent correlation to the SSRRV especially for cells with low noise margin or small SSRRV. In this test chip, the measured envelope of the DSRRV shows a constant shift from the SSRRV (min(DSRRV–SSRRV)) of 15mV. This curve shift is shown in simulations in Figure 76(b) to be inversely proportional to the access time. The SSRRV becomes very close in to DSRRV when the T_{WL} is $>30\tau_{regeneration}$.

The dynamic read-stability can also be measured under repeated access [115-117]. As one expects, more frequent access does not allow cells to fully recover back to the initial state (S0) resulting in degraded dynamic read-stability. However, as observed in [117], the cells with a lower dynamic margin have smaller internal capacitances, and hence the recovery time is shorter such that repeated access does not substantially impact the DSRRV. Simulation results in Figure 77 support this fact such that the difference in DSRRV between the single and repeated access is negligible for relatively weak cells regardless of the cycle time.

Our results indicate that, for a given word-line pulse width, we can measure the DSRRV for cells with low SSRRV. The difference in the measurement can be directly used to determine the lower bound of the DSRRV by subtracting the difference from the estimated SSRRV.

(a)   (b)
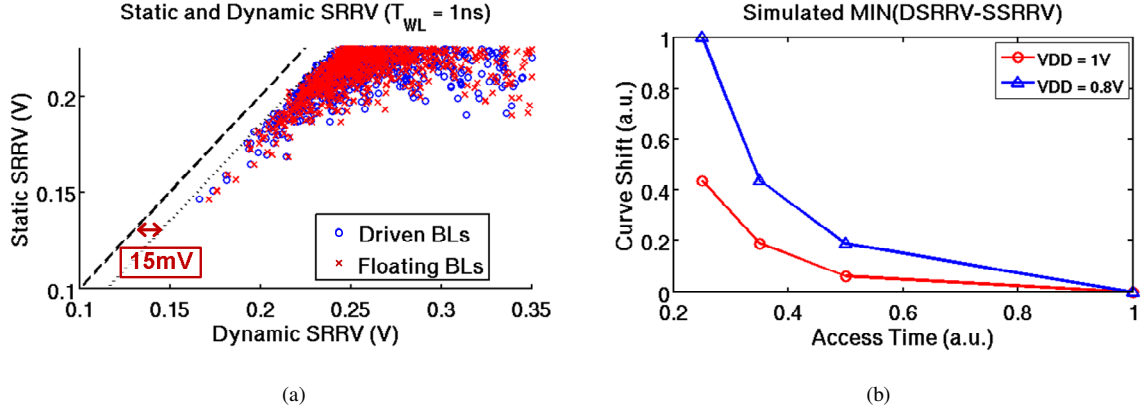
Figure 76.   (a) Scatter plot of the static and dynamic SRRV. The dynamic SRRV
is measured with the bit-lines driven by external sources (circle) and with pre-
charging the bit-lines (cross). (b) The amount of the curve shift during the read
operation with variant access time (simulation results).

(a)   (b)
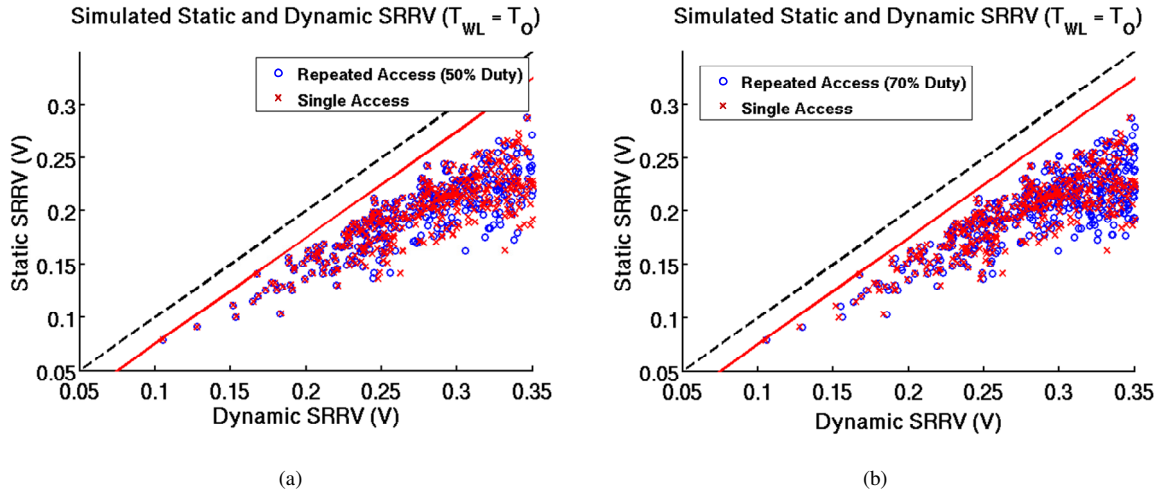
Figure 77.   Simulated static and dynamic SRRV with a fixed pulse width (T_O)
for a single time read access (cross) and for repeated read access with (a) 50% and
(b) 70% duty cycle (circle). The two dynamic stabilities match exactly for the
relatively unstable cells. Stables cells show larger stability degradation by
repeated access.

6.4.2.   Dynamic Write-Ability

We apply a similar approach as dynamic read-stability to measure dynamic write-ability. Figure 78(a) shows the schematics of a cell under test at the beginning of a write access. During a write, both sides of a cell inject current resulting in ΔV_L and ΔV_R. As illustrated in Figure 78(b), the dynamic BWTV (DBWTV) is defined to be highest level of the bit-line while still flipping the cell. With limited access time (T_{WL}), a lower V_{BL} is needed to flip a cell and hence a lower write margin than static BWTV (SBWTV). To better illustrate the difference, the state-space trajectories are plotted in Figure 79. Figure 79(a) shows the two VTCs during a hold and a write operation. Figure 79(b) shows two trajectories with the same pulse width (T_O) but different V_{BL}. At the higher V_{BL}, with a step at V_{WL}, the cell's state from S0 to S1; however, with a pulsed write access, the state stays in ROC 0 converging to S0. When V_{BL} is reduced by a small amount, the new trajectory successfully ends at S1.

Figure 80(a) shows the scatter plot of the measured SBWTV and DBWTV with 1ns write access time. A dynamic write operation has less noise margin, and the amount of the shift in the BWTV is within 25mV for the entire write-ability distribution. Unlike read-stability, DBWTV is not very different from SBWTV and the estimating the write-ability using SBWTV can be directly applied to dynamic behavior. Figure 80(b) shows the curve shift (max(SBWTV–DBWTV)) for the write-ability with various access times. The shift is considerable only for very narrow access pulses (<500ps or $7\tau_{regeneration}$). The dynamic write-ability can be predictably bounded to within a few tens of millivolts from the SBWTV.

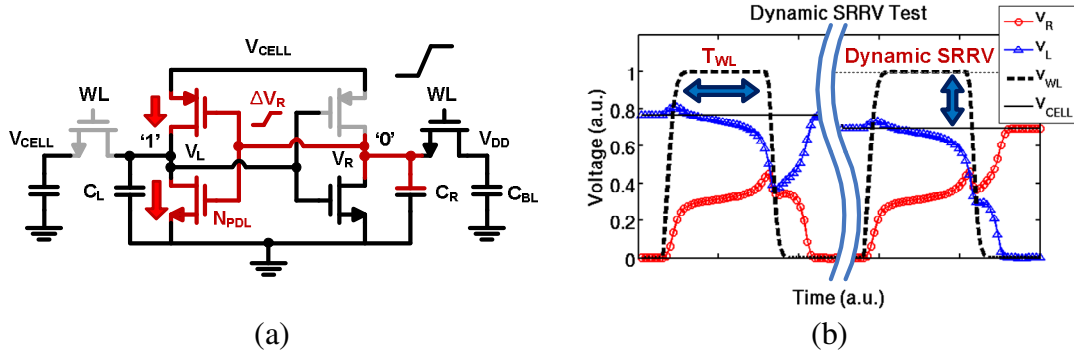(a)   (b)
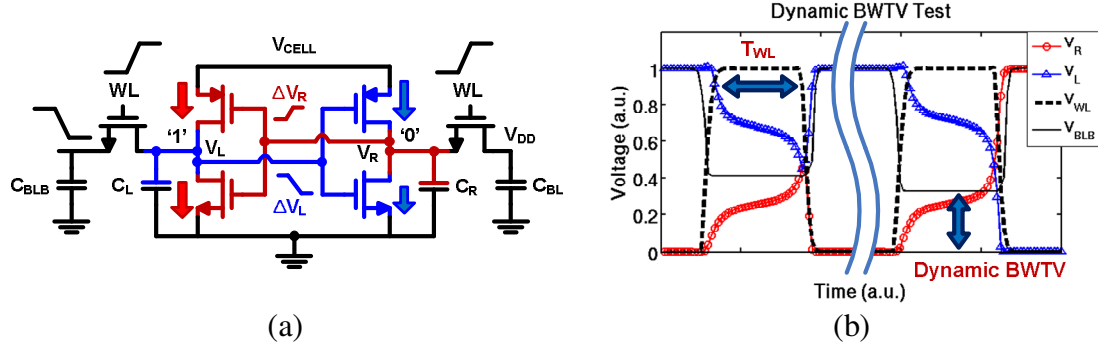
Figure 78.   Dynamic WWTV test setup. (a) Schematics during the initial write

access. (b) Testing waveform.

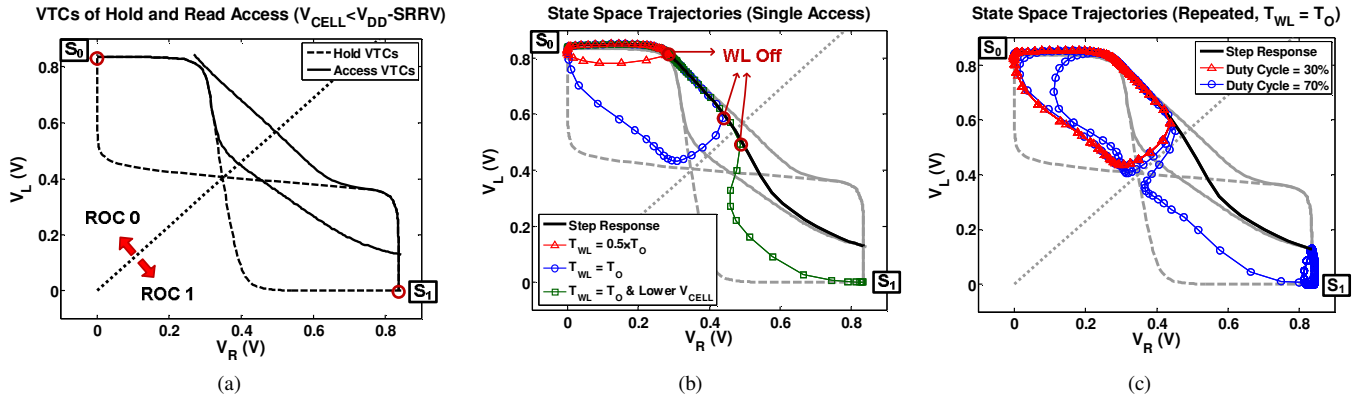(a)   (b)
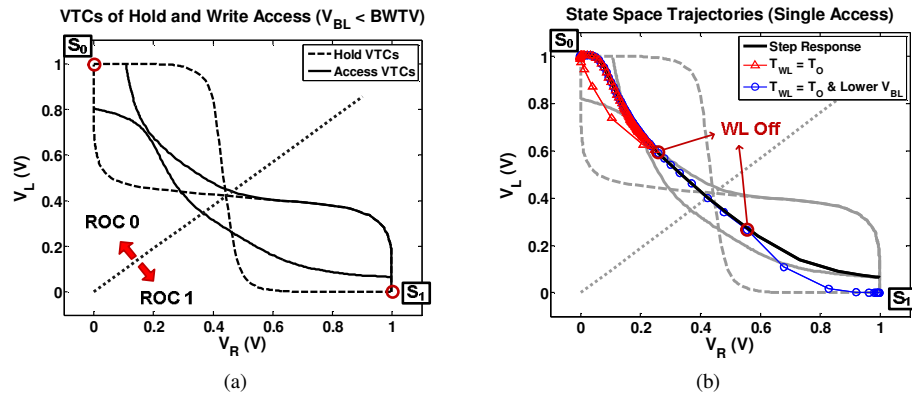
Figure 79.   (a) Simulated VTCs of an SRAM cell during hold and write access.

(b) Simulated trajectories of the internal voltage (V_L and V_R) during single write

access with variable V_{BL}. For all cases, V_{BL} is slightly lower than the write-trip

voltage level (i.e. $V_{BL} <$ BWTV).
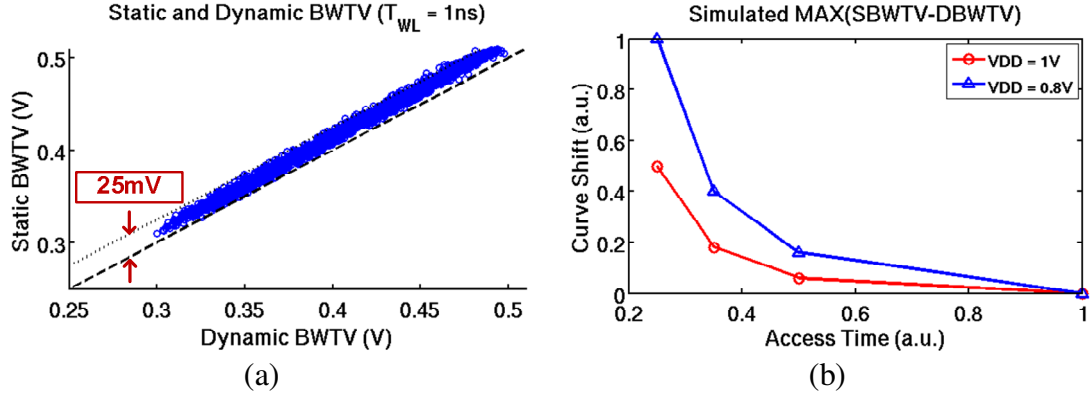
Figure 80.   (a) Scatter plot of the static and dynamic BWTV. (b) Extra dynamic
noise margin (EDNM) for the write operation with variant access time.

6.4.3.   Predictability of the Failure Condition for Multiple Dies

The estimated data in Section 6.3 are used to predict the read and write failure conditions. Figure 81(a) shows the number of failed reads per V_{CELL} sweep. The triangle is from normal read operation with T_{WL} = 1ns and the circle is from the static stability estimation data. The difference of the two curves at the tails is the estimation error of the failure condition. In this case the dynamic read-stability is underestimated when using the static estimates by about 22mV. Figure 81(b) shows the number of failed writes per V_{BL} sweep with T_{WL} = 1ns.

Once the fitting coefficients in (5.1~5.4) are found from a subset of a memory array, this coefficient set can be applied to other memory arrays in different dies. Figure 82(a) shows the absolute error distribution for 11 dies. Die-to-die variation of the DC offset and gain error of the current sensing circuits is properly corrected based on their I/O characteristics. This distribution shows that the error is confined to a range less than +/-5mV. Figure 82(b) shows the difference between the static and dynamic stability for read/write. The amount of the difference does not vary too much across multiple dies. Even if the internal pulse generator may produce variable T_{WL}, its impact on the curve

shift is relatively small unless the pulse width is much narrower than 1ns (Figure 76 and Figure 80). If the difference can be obtained from a few sample dies, the prediction error in Figure 82 can be reduced further.

<div align="center">(a)   (b)</div>
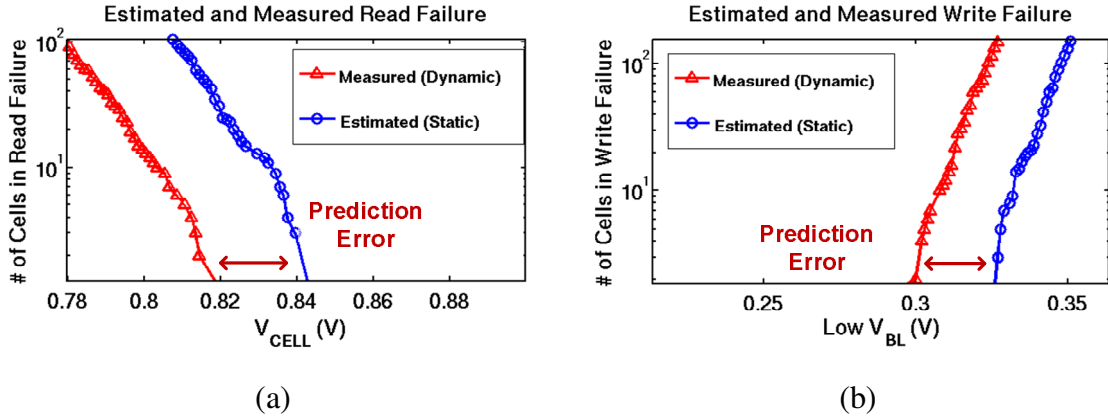
Figure 81.   (a) Measured and estimated read fail bit count per V_{CELL} sweep. (b) Measured and estimated write fail bit count per low side V_{BL} sweep.

<div align="center">(a)   (b)</div>
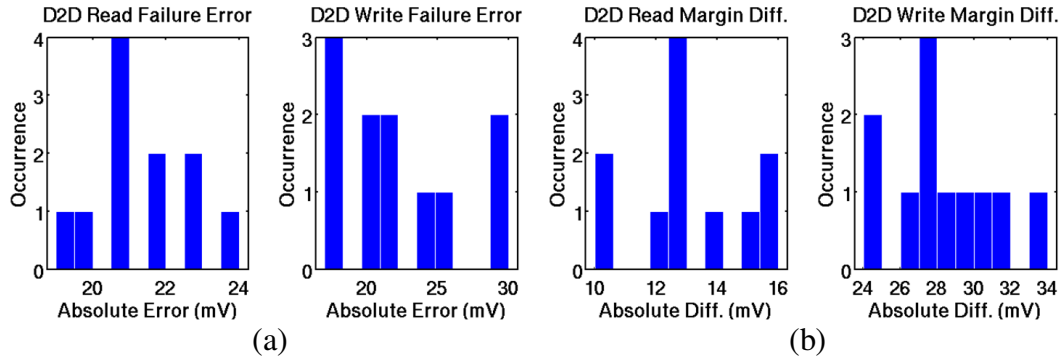
Figure 82.   Die-to-die variation of (a) failure prediction error between the dynamic read/write operation and the statically estimated results from the bit-line current measurements and (b) the difference between the measured dynamic and static margin.

Section 6.5   Summary

A rapid SRAM cell stability estimation is enabled by using the embedded current sensing circuits with controlled supply. The estimation accuracy is much improved compared to the prior works ($R^2 > 0.8$). The dynamic concerns arising from diverse effects are investigated and proper testing setups are demonstrated. A close correlation to the static estimation is verified from measurement data. Using the self-testing technique explored in this work, each supply level for a 6T-SRAM array can be more systematically controlled to guarantee read/write stability with dynamic characteristics.

Chapter 7

Conclusions

This dissertation has addressed improving the yield and reliability of two common IC building blocks, DACs and SRAMs. Instead of performing exhaustive MC simulations, this dissertation focusses on estimation functions for either pre- or post-silicon designs. The key element of the proposed approaches is maintaining the accuracy without computational or extensive measurement complexity leading to fast estimation.

The simple mathematical derivation of the integration of the multivariate Gaussian random variables enables intuitive formulation of the linearity yield model of the DAC. The proposed models can handle the yield estimation for arbitrary structure from the measured current variation. The DNL yield model is analyzed and suggested for the first time with dramatic matching to the simulation and measurement data. While the existing INL yield models typically do not accurately predict the yield for segmented DAC, this work used multiple error functions with fractional power factors that match to the measured yield very well. The yield models are verified by measured current data from 8-bit current-steering DACs fabricated in 90nm CMOS. Even with non-Gaussian

distribution the accuracy of the models does not degrade. Therefore, the proposed yield models are reliable and applicable to future technologies. By using these models, the time and efforts of traditional MC simulation-based methods can be greatly reduced. This dissertation does not extend the suggested static models to include dynamic effects; a dynamic yield model is a potential future work.

The second technique used built-in testing circuits for the embedded SRAM as a way to rapidly and reliably estimate the cell stability. With a test chip made in 65nm CMOS, this technique improved the estimation accuracy and measurement speed greatly compared to any published literature. Another contribution of this work is establishing excellent correlation between the static and dynamic stability. From the scatter plot and failure prediction, the extra dynamic noise margin is defined and is used to gauge the dynamic stability from the static noise margin. The correlation observed between the static and dynamic stability defined in this work shows small variation across the multiple dies, hence more reliable dynamic yield estimation is feasible using the technique.

The use of the technique (not only from yield estimation) can be found from many practical needs. The extracted parameters from a subset of a die are useful as they can be applied to the other dies for the stability estimation. This basic information leads to more possible applications of the technique. Primary application is to obtain the accurate analog amount of the stability information that is not only useful in adjusting the supply level of the chip to enhance the yield, but it is also helpful in placing a margin on the control variable (e.g. V_{CELL}) from a known expected device variation due to environment stress. Also the analog stability level is determined by six random variables in a 6T-cell. The stability number is purely random and is obtained from every single cell. The whole SRAM cell array can be treated as a random code generator. Both these applications have potential as future extensions of this work.

Appendices

A.1 Simplified Yield Model of 2-D Correlated Gaussian Random Variables

This section discusses the conditions for the approximation in (Eq. 3.3). This discussion starts with two nearly identical variables ($\sigma_x \sim \sigma_y$). When the variables are weakly correlated ($\rho < 0.2$), the xy-term in (Eq. 3.2) can be ignored and square root of ($1-\rho^2$) is approximately 1. Thus, (Eq. 3.2) can be simplified as multiplication of two error functions. As seen from Figure 6, the fitting coefficient of (Eq. 3.3) stays close to 2 when $\rho < 0.2$. For highly correlated variables, (Eq. 3.2) can be simplified by rotating the x-y axis by 45 degrees.

$$P_{C\_2D} \approx 4 \cdot \int_0^{\sqrt{2}\cdot A} \int_0^{\sqrt{2}\cdot A-x} \frac{\exp\left(\dfrac{-x^2}{2(1+\rho)\sigma_x^2} - \dfrac{y^2}{2(1-\rho)\sigma_x^2} \right)}{2\pi \cdot \sigma_x \sigma_x \sqrt{1-\rho^2}} dydx \qquad \text{(Eq A.1)}$$

$$P_{C\_2D} \approx 2 \cdot \int_{0}^{\sqrt{2} \cdot A} \frac{f(x)}{\sqrt{2\pi} \cdot \sigma_x \sqrt{1+\rho}} \cdot g(x)\,dx$$

$$f(x) = \exp\left(\frac{-1}{2(1+\rho)} \cdot \left(\frac{x^2}{\sigma_x^2} \right) \right)$$

$$g(x) = erf\left(\frac{\sqrt{2} \cdot A - x}{\sqrt{2} \cdot \sigma_x \sqrt{1-\rho}} \right)$$

(Eq A.2)

Note that if the two variables have different variances but in close value such as $0.9 < \sigma_x/\sigma_y < 1$ (Figure 5(b)), the rotation angle can be properly adjusted to eliminate the xy-term in (Eq. 3.2). In that case, P_{C_2D} can be simplified as (Eq. A.2) by adding proper gain factors to σ_X in f(x) and g(x) while the integration bound is approximately identical. The g(x) function can be simplified to be a step function with the step transition where the error function is at 0.5 (Figure 83). This simplification reduces the equation to (Eq. A.3). Note that this simplification is accurate when ρ is close to 1, and when the curvature of f(x) is not too sharp which implies a high probability of success or small σ_X.
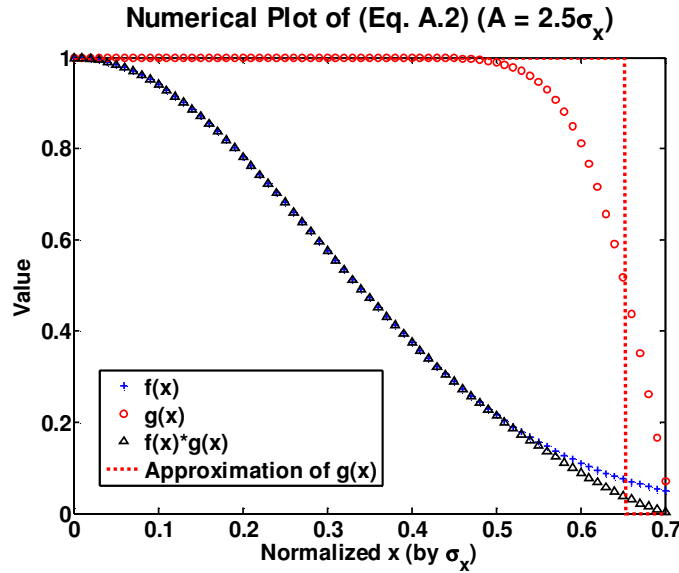
Figure 83.   Numerical approximation of (Eq. A.2). The decision boundary A is 2.5 times σ_x in this figure.

$$P_{C\_2D} = 2 \times \int_0^{\sqrt{2} \cdot A - erf^{-1}(0.5)\sqrt{2(1-\rho)}\sigma_x} \frac{f(x)}{\sqrt{2\pi} \cdot \sigma_x \sqrt{1+\rho}} dx \qquad \text{(Eq A.3)}$$

$$P_{C\_1D} = 2 \times \int_0^{\sqrt{1+\rho} \cdot A} \frac{f(x)}{\sqrt{2\pi} \cdot \sigma_x \sqrt{1+\rho}} dx \qquad \text{(Eq A.4)}$$

When comparing this reduced equation with P_C of a 1-D normal random variable (Eq. A.4), the probability of success of 2-D normal random variables (P_{C_2D}) can be recognized as a single error function with modified decision boundary that is a function of σ_x and ρ. The amount of yield loss due to shrinking of the decision boundary has a close correlation to the original yield (P_{C_1D}) and hence can be expressed as a function of P_{C_1D}. An arbitrary polynomial expansion is tried at a certain estimation target P_0.

$$P_{C\_2D} \approx b_0 + b_1 \cdot \left(P_{C\_1D} - P_0\right) + b_2 \cdot \left(P_{C\_1D} - P_0\right)^2 + \ldots$$
$$\approx \left[P_{C\_1D}\right]^k \qquad \text{(Eq A.5)}$$

By numerically analyzing the polynomial coefficients (b_0, b_1, etc.) as seen in Table A-1, the polynomial fit (third order) closely matches the expression for an exponential. The fitting factor k in this table can be different from the value in Figure 6 as ρ becomes smaller than 0.9. Due to small curvature of f(x) in (Eq. A.2), the step transition is slightly modified (from erf-1(0.5) to erf-1(0.51)) such that (Eq. A.3) is valid for high probability of success. Although the derivation is based on a high ρ value, (Eq. 3.3) shows reasonable accuracy for entire ρ value if focused on high yield level as seen in Figure 6.

Table A- 1 Numerical analysis of (Eq. A.5) with $P_0 = 0.9$

ρ	Polynomial Coefficients				Equivalent k
	b0	b1	b2	b3	
0.999	0.9	1.03	0.05	0.09	1.03
0.99	0.9	1.09	0.15	0.26	1.1
0.95	0.88	1.19	0.27	0.42	1.22
0.9	0.87	1.26	0.3	0.39	1.31
0.85	0.86	1.31	0.28	0.24	1.37

Bibliography

[1]   K. C. Smith, A. Wang, and L. C. Fujino (2012). Trend tracking for INT. SOLID-STATE CIRCUITS CONF. 2012. Available: http://Int. Solid-State Circuits Conf. .org/doc/2012/2012_Trends.pdf.

[2]   K.A. Bowman, S.G. Duvall, J.D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," IEEE Journal of Solid State Circuits, pp. 183-190, Feb. 2002.

[3]   J.W. Tschanz, et al., "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," IEEE Journal of Solid State Circuits, pp. 1396-1402, Nov. 2002.

[4]   J.-B. Shyu, G. C. Temes, and F. Krummenacher, "Random error effects in matched MOS capacitors and current sources," IEEE Journal of Solid State Circuits, pp. 948-955, Dec. 1984.

[5]   M.J.M. Pelgrom, A.C.J. Duinmaijer, A.P.G. Welbers, "Matching properties of MOS transistors," IEEE Journal of Solid State Circuits, pp. 1433-1439, Oct. 1989.

[6]   W.-T. Lin and T.-H. Kuo, "A Compact dynamic-performance-improved current-steering DAC with random rotation-based binary-weighted selection," IEEE Journal of Solid State Circuits, pp.444-453, Feb. 2012.

[7]   K. L. Chan, J. Zhiu, and I. Galton, "Dynamic element matching to prevent

nonlinear distortion from pulse-shape mismatches in high-resolution DACs," IEEE Journal of Solid State Circuits, pp. 2067-2078, Sep. 2008.

[8] K. O' Sullivan, C. Gorman, M. Hennessy, and V. Callaghan, "A 12-bit 320-MSamples/s current-steering CMOS D/A converter in 0.44mm^2," IEEE Journal of Solid State Circuits, pp. 1064-1071, July 2004.

[9] Y. Cong, R.L. Geiger, "A 1.5-V 14-bit 100-MS/s self-calibrated DAC," IEEE Journal of Solid State Circuits, pp. 2051- 2060, Dec. 2003.

[10] J. Hyde, T. Humes, C. Diorio, M. Thomas, and M. Figueroa, "A 300-MS/s 14-bit digital-to-analog converter in logic CMOS," IEEE Journal of Solid State Circuits, pp.734-740, May 2003.

[11] M. Clara, W. Klatzer, B. Seger, A. D. Giandomenico, and L. Gori, "A 1.5V 200MS/s 13b 25mW DAC with randomized nested background calibration in 0.13um CMOS," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 250-252, Feb. 2007.

[12] W.-H. Tseng, C.-W. Fan, J.-T. Wu, "A 12-Bit 1.25-GS/s DAC in 90 nm CMOS with > 70 dB SFDR up to 500 MHz", IEEE Journal of Solid State Circuits, pp. 2845-2856, Dec. 2011.

[13] Q. Huang, P.A. Francese, C. Martelli, J. Nielsen, "A 200MS/s 14b 97mW DAC in 0.18μm CMOS," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 364-532, Feb. 2004.

[14] A. R. Bugeja and B.-S. Song, "A self-trimming 14-b 100-MS/s CMOS DAC," IEEE Journal of Solid State Circuits, pp. 1841- 1852, Dec. 2000.

[15] B. Razavi, Principles of Data Conversion System Design, IEEE Press, 1995.

[16] J. Savoj, et al., "A 12-GS/s phase-calibrated CMOS digital-to-analog converter for backplane communications," IEEE Journal of Solid State Circuits, pp.1207-1216, May 2008.

[17] H. Zheng, et al., "A 3.1 GHz–8.0 GHz single-chip transceiver for MB-OFDM UWB in 0.18um CMOS process," IEEE Journal of Solid State Circuits, pp.414-426, Feb.

2009.

[18] A. V. D. Bosch, et al., "A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter," IEEE Journal of Solid State Circuits, pp.315-324, Mar. 2001.

[19] A. V. D. Bosche, M. Steyaert, and W. Sansen, "SFDR-bandwith limitations for high speed high resolution current steering CMOS D/A converters," in Proc. IEEE Int. Conf. Electronics, Circuits and Systems, pp. 1193-1196, 1999.

[20] T. Chen and G. Gielen, "Modeling of the impact of the current source output impedance on the SFDR of current-steering CMOS D/A converters," in Proc. IEEE Int. Conf. Electronics, Circuits and Systems, pp. 293-296, 2004.

[21] J. J. Wikner and N. Tan, "Modeling of CMOS digital-to-analog converters for telecommunication," IEEE Transactions on Circuits and Systems-II: Express Briefs, pp. 489-499, May 1999.

[22] S. Luschas, H.S. Lee, "Output impedance requirements for DACs," in Proc. IEEE Int. Symp. Circuits and Systems, vol.1, pp. 861-864, 2003.

[23] C. Kromer, et al., "A 25-Gb/s CDR in 90-nm CMOS for high-density interconnects," IEEE Journal of Solid State Circuits, pp.2921-2929, Dec. 2006.

[24] J.A. Tierno, et al., "A wide power supply range, wide tuning range, all static CMOS all digital PLL in 65 nm SOI," IEEE Journal of Solid State Circuits, pp.42-51, Jan. 2008.

[25] R. Kanj, R. Joshi, S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," In IEEE/ACM Proceedings of the 43rd Design Automation Conference, pp.69-72, 2006

[26] A. Jha and P. R. Kinget, "Wideband signal synthesis using interleaved partial-order hold current-mode digital-to-analog converters," IEEE Transactions on Circuits and Systems-II: Express Briefs, pp.1109-1113, Nov. 2008.

[27] C.-H. Lin, et al., "A 12 bit 2.9 GS/s DAC with IM3 < -60 dBc beyond 1GHz in 65nm CMOS," IEEE Journal of Solid State Circuits, pp. 3285-3293, Dec. 2009.

[28] Y. Tang, et al., "A 14 bit 200MS/s DAC with SFDR > 78dBc, IM3 < -83 dBc and

NSD < -163 dBm/Hz across the whole Nyquist band enabled by dynamic-mismatch mapping," IEEE Journal of Solid State Circuits, pp. 1371-1381, Jun. 2011.

[29] T. Chen and G. G. E. Gielen, "The analysis and improvement of a current-steering DACs dynamic SFDR-I: the cell-dependent delay differences," IEEE Transactions on Circuits and Systems-I: Regular Papers, pp. 3-15, Jan. 2006.

[30] C.-H. Lin and K. Bult, "A 10-b, 500-MSample/s CMOS DAC in 0.6mm^2," IEEE Journal of Solid State Circuits, pp. 1948-1958, Dec. 1998.

[31] J. Bastos, A.M. Marques, M.S.J. Steyaert, W. Sansen, "A 12-bit intrinsic accuracy high-speed CMOS DAC," IEEE Journal of Solid-State Circuits, vol.33, no.12, pp.1959-1969, Dec. 1998.

[32] D. A. Mercer, "Low-power approaches to high-speed current-steering digital-to-analog converters in 0.18-um CMOS," IEEE Journal of Solid State Circuits, pp. 1688-1698, Aug. 2007.

[33] B. Schafferer and R. Adams, "A 3V CMOS 400mV 14b 1.4GS/s DAC for multi-carrier applications," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, 2004.

[34] S. Park, G. Kim, S.-C. Park, W. Kim, "A digital-to-analog converter based on differential-quad switching," IEEE Journal of Solid State Circuits, pp. 1335-1338, Oct. 2002.

[35] T. Chen and G. G. E. Gielen, "The analysis and improvement of a current-steering DACs dynamic SFDR-II: the output-dependent delay differences," IEEE Transactions on Circuits and Systems-I: Regular Papers, pp. 268-279, Feb. 2007.

[36] D. Seo and G. H. McAllister, "A low-spurious low-power 12-bit 160-MS/s DAC in 90-nm CMOS for baseband wireless transmitter," IEEE Journal of Solid State Circuits, pp.486-495, Mar. 2007.

[37] W. Schofield, D. Mercer, and L. St. Onge, "A 16b 400MS/s DAC with < -80dBc IMD to 300MHz and <-160dBm/Hz noise power spectral density," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, 2003.

[38] C. S. G. Conroy, W. A. Lane, and M. A. Moran, "statistical design techniques for

D/A converters," IEEE Journal of Solid State Circuits, pp. 1118-1128, Aug. 1989.

[39] K.-W. Su, et al., "A scaleable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics," in Proc. IEEE Custom Integrated Circuits Conference, pp. 245-248, Sep. 2003.

[40] J.V. Faricelli, "Layout-dependent proximity effects in deep nanoscale CMOS," in Proc. IEEE Custom Integrated Circuits Conference, pp.1-8, Sep. 2010.

[41] G.A.M. Van Der Plas, J. Vandenbussche, W. Sansen, M.S.J. Steyaert, G.G.E. Gielen, "A 14-bit intrinsic accuracy Q2 random walk CMOS DAC," IEEE Journal of Solid-State Circuits, vol.34, no.12, pp.1708-1718, Dec. 1999.

[42] Y. Cong and R. L. Geiger, "Switching sequence optimization for gradient error compensation in thermometer-decoded DAC arrays," IEEE Transactions on Circuits and Systems-II: Express Briefs, pp. 585-595, July 2000.

[43] K.-C. Kuo and C.-W. Wu, "A switching sequence for linear gradient error compensation in the DAC design," IEEE Transactions on Circuits and Systems-II: Express Briefs, pp502-506, Aug. 2011.

[44] D. A. Mercer, "Low-power approaches to high-speed current-steering digital-to-analog converters in 0.18-um CMOS," IEEE Journal of Solid State Circuits, pp. 1688-1698, Aug. 2007.

[45] T. Chen and G. G. E. Gielen, "A 14-bit 200-MHz current-steering DAC with switching-sequence post-adjustment calibration," IEEE Journal of Solid State Circuits, pp. 2386-2394, Nov. 2007.

[46] K.R. Lakshmikumar, R.A. Hadaway, M.A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analog design," IEEE Journal of Solid State Circuits, pp. 1057- 1066, Dec. 1986.

[47] C.S.G. Conroy, W.A. Lane, M.A. Moran, K.R. Lakshmikumar, M.A. Copeland, R.A. Hadaway, "Comments, with reply, on `Characterization and modeling of mismatch in MOS transistors for precision analog design'," IEEE Journal of Solid State Circuits, pp.294-296, Feb. 1988.

[48] A. Van den Bosch, M. Steyaert, and W. Sansen, "An accurate statistical yield model for CMOS current-steering D/A converters," in Proc. IEEE Int. Symp. Circuits and Systems, pp.105-108, vol.4, 2000.

[49] Y. Cong and R.L. Geiger, "Formulation of INL and DNL yield estimation in current-steering D/A converters," in Proc. IEEE Int. Symp. Circuits and Systems, pp. 149–152, vol.3, 2002.

[50] M. Kosunen, J. Vankka, F. Teikari, K. Halonen, "DNL and INL yield models for a current-steering D/A converter," in Proc. IEEE Int. Symp. Circuits and Systems, pp. 969–972 vol.1, 2003

[51] G. I. Radulov, M. Heydenreich, R. W. van der Hofstad, J. A. Hegt, A. H. M. van Roermund, "Brownian-bridge-based statistical analysis of the DAC INL caused by current mismatch," IEEE Transactions on Circuits and Systems-II: Express Briefs, pp.146-150, Feb. 2007.

[52] M. Saberi, et al., "Analysis of power consumption and linearity in capacitive digital-to-analog converters used in successive approximation ADCs," IEEE Transactions on Circuits and Systems-I: Regular Papers, pp. 1736–1748, Aug. 2011.

[53] J. A. Fredenburg and M. P. Flynn, "Statistical analysis of ENOB and yield in binary weighted ADCs and DACS with random element mismatch," IEEE Transactions on Circuits and Systems-I: Regular Papers, to appear.

[54] A. Leon-Garcia, Probability and Random Process for Electrical Engineering (2nd ed), Addison Wesley, 1994.

[55] G.K. Karagiannidis, A.S. Lioumpas, "An improved approximation for the Gaussian Q-function," IEEE Communications Letters, vol.11, no.8, pp.644-646, Aug. 2007.

[56] Y. Isukapalli, B. D. Rao, "An analytically tractable approximation for the Gaussian Q-function," IEEE Communications Letters, vol.12, no.9, pp.669-671, Sep. 2008.

[57] R. N. McDonough and A. D. Whalen, Detection of Signals in Noise (2nd ed), Academic Press, 1995.

[58] G. A. F. Seber, Multivariate Observations. Wiley, 2004.

[59] F. Pregaldiny and C. Lallement, "Fourth generation MOSFET model and its VHDL-AMS implementation," Int. Journal of Numerical Modelling: Electronic Networks, Devices and Fields, vol. 18, no. 1, Jan. 2005.

[60] D. J. Frank, et al. "Device scaling limits of Si MOSFETs and their application dependencies," in Proc. IEEE, vol. 89, pp. 259-288, Mar. 2001.

[61] F. Assaderaghi, D. Sinitsky, J. Boker, P. K. Ko, H. Gaw, and C. Hu, "High-field transport of inversion-layer electrons and holes including velocity overshoot," IEEE Transaction on Electron Devices, vol. 44, no. 4, pp. 664-671, 1997.

[62] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," IEEE Electron Device Letters, no. 7, pp. 361-363, 1997.

[63] D. Lu, "Compact models for future generation CMOS," Ph.D Dissertation, Dept. Electrical Engineering, University of California, Berkeley, 2011.

[64] B. Nikolic, et al., "Technology variability from a design perspective," IEEE Transactions on Circuits and Systems I: Regular Papers, pp. 1996-2009, Sep. 2011.

[65] J. Hicks, et al., "45nm transistor reliability", Intel Technology Journal, vol. 12, Issue 2, pp. 131–144, Jun. 2008.

[66] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in Int. Electron Devices Meeting (IEDM), pp. 471-474, 2007.

[67] K. Takeuchi, et al., "Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies," in Int. Electron Devices Meeting (IEDM), pp. 467-470, 2007.

[68] J. Lohstroh, "Static and dynamic noise margins of logic circuits," IEEE Journal of Solid-State Circuits, vol.14, no.3, pp. 591–598, Jun. 1979.

[69] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells", IEEE Journal of Solid-State Circuits, vol. SC-22, no. 5, pp. 748–754, Oct. 1987.

[70] J. Wang, S. Yaldiz, X. Li, and L.T. Pileggi, "SRAM parametric failure analysis," in

Proc. ACM/IEEE 46th DAC, pp.496-501, 2009.

[71] A.J. Bhavnagarwala, X. Tang, and J.D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE Journal of Solid-State Circuits, vol.36, no.4, pp.658-665, Apr. 2001.

[72] B. H. Calhoun and A. P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS," IEEE Journal of Solid State Circuits, pp. 1673-1679, Jul. 2006.

[73] K. Agarwal and S. Nassif, ,"The impact of random device variation on SRAM cell stability in sub-90-nm CMOS technologies", IEEE Transactions On Very Large Scale Integration Systems, Vol. 16, No. 1, pp. 86–97, Jan 2008.

[74] T. Tsukamoto, et al., "Worst-case analysis to obtain stable read/write DC margin of high density 6T-SRAM-array with local Vth variability," in ICCAD Dig. Tech. Papers, pp. 398-405, Nov. 2005.

[75] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies", IEEE Journal of Solid-State Circuits, Vol. 41, Issue 11, pp. 2577–2588, Nov. 2006.

[76] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, and H. Kobatake, "Redefinition of write margin for next-generation SRAM and write-margin monitoring circuit," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 34.5, 2006.

[77] N. Gierczynski, B. Borot, N. Planes, and H. Brut, "A new combined methodology for write-margin extraction of advanced SRAM," IEEE Int. Conf. on Microelectronic Test Structures, pp.97-100, 2007.

[78] R. Herald, P. Wang, "Variability in sub-100nm SRAM designs," IEEE/ACM Int. Conf. on Computer Aided Design, pp. 347-352, Nov. 2004.

[79] Z. Guo, et al., "Large-scale read/write margin measurement in 45 nm CMOS SRAM arrays," in Symp. VLSI Circuits Dig., pp. 42–43, 2008.

[80] Z. Guo, et al., "Large-scale SRAM variability characterization in 45nm CMOS", IEEE Journal of Solid-State Circuits, Vol. 44, Issue 11, pp. 3714–3192, Nov. 2009.

[81] A. Bhavnagarwala, et al., "Fluctuation limits & scaling opportunities for CMOS SRAM cells," in Int. Electron Devices Meeting (IEDM), pp. 659-662, Dec. 2005.

[82] A.J. Bhavnagarwala, et al., "A Sub-600-mV, fluctuation tolerant 65-nm CMOS SRAM array with dynamic cell biasing," IEEE Journal of Solid-State Circuits, vol.43, no.4, pp.946–955, Apr. 2008.

[83] T. Fischer, et al., "Analysis of read current and write trip voltage variability from a 1-MB SRAM test structure," IEEE Transactions on Semiconductor Manufacturing, vol. 21, no. 4, pp. 534–541, Nov. 2008.

[84] J. Tsai, et al., "SRAM stability characterization using tunable ring oscillators in 45nm CMOS," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 354–356, Feb 2010.

[85] H. Yamauchi, "Embedded SRAM trend in nano-scale CMOS," in IEEE MTDT Dig. Tech. Papers, Dec. 2007, pp. 19-22.

[86] H. Yamauchi, "A discussion on SRAM circuit design trend in deeper nanometer-scale technologies," IEEE Transactions on Very Large Scale Integration Systems, pp. 763-774, May 2010.

[87] H. Pilo, C. Barwin, G. Braceras, C. Browning, S. Lamphier, and F. Towler, "An SRAM design in 65-nm technology node featuring read and write-assist circuits to expand operating voltage," IEEE Journal of Solid-State Circuits, pp. 813-819, Apr. 2007.

[88] M. Yamaoka, et al., "Low-power embedded SRAM modules with expanded margins for writing," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 26.4, 2005.

[89] K. Zhang, et al., "A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply," IEEE J. Solid-State Circuits, pp. 146–151, Jan. 2006.

[90] H. Mair, et al., "A 65-nm mobile multimedia applications processor with an adaptive power management scheme to compensate for variations," in Symp. VLSI

Circuits Dig. Tech. Papers, pp. 224–225, Jun. 2007.

[91] J. Pille, et al., "Implementation of the cell broadband engine in a 65 nm SOI technology featuring dual-supply SRAM arrays supporting 6 GHz at 1.3 V," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 322–323, 2007.

[92] Y. H. Chen, et al., "A 0.6 V 45 nm adaptive dual-rail SRAM compiler circuit design for lower VDDmin VLSIs," IEEE Journal of Solid State Circuits, pp. 1209–1215, Apr. 2009.

[93] M. Yamaoka, K. Osada, and K. Ishibashi, "0.4-V logic-library-friendly SRAM array using rectangular-diffusion cell and delta-boosted-array voltage scheme," IEEE J. Solid-State Circuits, pp. 934–940, Jun. 2004.

[94] A. Raychowdhury, et al., "PVT-and-aging adaptive wordline boosting for 8T SRAM power reduction," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 19.6, 2010.

[95] G. Panagopoulos, G. Karakonstantis, D. Wisland, H. Mahmoodi, J.K. Madsen, K. Roy, "Multi-level wordline driver for low power SRAMs in nano-scale CMOS technology," IEEE 29th Int. Conf. on Computer Design (ICCD), pp. 326-331, Oct. 2011.

[96] A. Carlson, Z. Guo, L.-T. Pang, T.-J. K. Liu, and B. Nikolic, "Compensation of systematic variations through optimal biasing of SRAM wordlines," in Proc. IEEE Custom Integrated Circuits Conf., pp. 411-414, Sep. 2008.

[97] M. Khellah, et al., "Process, temperature, and supply-noise tolerant 45 nm dense cache arrays with diffusion-notch-free (DNF) 6T SRAM cells and dynamic multi-Vcc circuits," IEEE Journal of Solid-State Circuits, pp. 1199-1208, Apr. 2009.

[98] S. Ohbayashi, et al., "A 65-nm SoC embedded 6T-SRAM designed for manufacturability with read and write operation stabilizing circuits," IEEE Journal of Solid-State Circuits, pp. 820-829, Apr. 2007.

[99] O. Hirabayashi, et al., "A process-variation-tolerant dual-power-supply SRAM with 0.179μm cell in 40 nm CMOS using level-programmable wordline driver," IEEE

Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 27.2, 2009.

[100] B. Mohammad, M. Saint-Laurent, P. Bassett, and J. Abraham, "Cache design for low power and high yield," in Proc. 9th Int. Symp. Quality Electronic Design, pp. 103-107, 2008.

[101] M. Yabuuchi, et al., "A 45nm low-standby-power embedded SRAM with improved immunity against process and temperature variations," in IEEE Int. Solid State Circuits Conf. Dig. Tech. Papers, pp. 18.3, 2007.

[102] K. Nii, et al., "A 45-nm bulk CMOS embedded SRAM with improved immunity against process and temperature variations," IEEE Journal Solid-State Circuits, pp. 180–191, Jan. 2008.

[103] K. Nii, et al., "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," in IEEE Symp. VLSI Circuits Dig., pp. 212–213, 2008.

[104] Y. Fujimura, et al., "A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149 m cell in 32 nm high-k metal-gate CMOS," in IEEE Int. Solid State Circuits Conf. Dig. Tech. Papers, pp. 348-349, 2010.

[105] N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M. Tan'no, and T. Douseki, "A 0.5-V 25-MHz 1-mW 256-kb MTCMOS/SOI SRAM for solar-power-operated portable personal digital equipment—Sure write operation by using step-down negatively overdriven bitline scheme," IEEE Journal of Solid State Circuits, pp. 728-742, Mar. 2006.

[106] P. Kolar, et al., "A 32nm high-k metal gate SRAM with adaptive dynamic stability enhancement for low-voltage operation," IEEE Jounal of Solid State Circuits, pp. 76-84, Jan. 2011.

[107] M. M. Khellah, A. Keshavarzi, D. Somasekhar, T. Karnik, and V. De, "Read and write circuit assist techniques for improving Vccmin of dense 6T SRAM cell," in IEEE Int. Conf. Integrated Circuit Design and Technology and Tutorial, pp. 185-

188, 2008.

[108] N. Verma and A. P. Chandrakasan, "A high-density 45nm SRAM using small-signal non-strobed regenerative sensing," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 21.3, 2008.

[109] S. Rauch, "The statistics of NBTI-induced V_T and β mismatch shifts in pMOSFETs," IEEE Transactions on Device and Materials Reliabilit, pp. 89-93, Dec. 2002.

[110] A. T. Krishnan, et al., "SRAM cell static noise margin and VMIN sensitivity to transistor degradation," in Int. Electron Devices Meeting (IEDM), pp. 1-4, Dec. 2006.

[111] S. Pae, J. Maiz, C. Prasad, and B. Woolery, "Effect of BTI degradation on transistor variability in advanced semiconductor technologies," IEEE Transactions on Device and Materials Reliability, pp. 519-525, Sep. 2008.

[112] A. Carlson, "Mechanism of increase in SRAM Vmin due to negative-bias temperature instability," IEEE Transactions on Device and Materials Reliability, pp. 473-478, Sep. 2007.

[113] M. Khellah, et al., "Wordline and bitline pulsing schemes for improving SRAM cell stability in low-vcc 65nm CMOS designs," in Symp. VLSI Circuits Dig. Tech. Papers, pp. 9-10, 2006.

[114] B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky, "Analytical modeling of SRAM dynamic stability," in IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 315-322, 2006.

[115] D. Khalil, et al., "Accurate estimation of SRAM dynamic stability," IEEE Transactions On Very Large Scale Integration Systems, Vol. 16, No. 12, pp. 1639–1647, Dec. 2008.

[116] M. Sharifkhani and M. Sachdev, "SRAM cell stability: A dynamic perspective," IEEE Journal of Solid State Circuits, pp. 609-619, Feb. 2009.

[117] S. O. Toh, Z. Guo, T.-J. K. Liu, and B. Nikolic, "Characterization of dynamic

SRAM stability in 45nm CMOS," IEEE Journal of Solid-State Circuits, Vol. 46, Issue 11, pp. 2702–2712, Nov. 2011.

[118] J. Wang, S. Nalam, and B. H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," in ACM/IEEE Int. Symp. Low Power Electronics and Design, pp.129-134, 2008.

[119] M. Khellah, D. Khalil, D. Somasekhar, Y. Ismail, T. Karnik, and V. De, "Effect of supply noise on SRAM dynamic stability," in IEEE Symp. VLSI Circuits Dig., pp. 76–77, 2007.

[120] M. Z. Straayer and M. H. Perrott, "A 12-bit, 10-MHz bandwidth, continuous-time Σ-Δ ADC with a 5-bit, 950-MS/s VCO-based quantizer," IEEE Journal of Solid State Circuits, pp. 805-814, Apr. 2008.

[121] J. Kim, T.-K. Jang, Y.-G. Yoon, and S.-H. Cho, "Analysis and Design of Voltage-Controlled Oscillator-Based Analog-to-Digital Converter," IEEE Transactions on Circuits and Systems-I: Regular Papers, pp. 18-30, Jan. 2010.

[122] A. A. Abidi, "Phase Noise and Jitter in CMOS Ring Oscillators", IEEE Journal of Solid-State Circuits, Vol. 41, Issue 8, pp. 1803–1816, Aug. 2006.

[123] J. Wang, S. Nalam, and B. H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," in ACM/IEEE Int. Symp. Low Power Electronics and Design, pp.129-134, 2008.