

Review Article

The Impact of Statistical Leakage Models on Design Yield Estimation

Rouwaida Kanj,¹ Rajiv Joshi,² and Sani Nassif¹

¹ IBM Austin Research Labs, Austin, TX 78758, USA

² IBM TJ Watson Labs, Yorktown Heights, NY 10598, USA

Correspondence should be addressed to Rouwaida Kanj, rouwaida@us.ibm.com

Received 19 October 2010; Accepted 11 December 2010

Academic Editor: Shiyan Hu

Copyright © 2011 Rouwaida Kanj et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Device mismatch and process variation models play a key role in determining the functionality and yield of sub-100 nm design. Average characteristics are often of interest, such as the average leakage current or the average read delay. However, detecting rare functional fails is critical for memory design and designers often seek techniques that enable accurately modeling such events. Extremely leaky devices can inflict functionality fails. The plurality of leaky devices on a bitline increase the dimensionality of the yield estimation problem. Simplified models are possible by adopting approximations to the underlying sum of lognormals. The implications of such approximations on tail probabilities may in turn bias the yield estimate. We review different closed form approximations and compare against the CDF matching method, which is shown to be most effective method for accurate statistical leakage modeling.

1. Introduction

With technology scaling, memory designs are the first to suffer from process variation. Density requirements in memory cells, latches, and register files aggravate variability effects and often undergo performance and yield degradation. More recently, the trend is starting to become more visible in peripheral logic [1], and random variations can lead to reduced noise margins. From a leakage perspective, general interest has been to model the statistical leakage distribution of the full-chip [2]. However, under extreme device conditions, leakages may inflict delays or faulty behavior: example false reads. In memory designs, bitline leakages can further aggravate the variability effects and impact readability of the design. Often many devices are stacked on the bitlines, and it has become necessary to statistically account for and model those many independent and, in most cases, identical, leaky sources. A typical array column can have around 16 to 64 cells per bitline. With the aim of capturing rare fail events accurately and modeling the variability space [3], it becomes important to emulate those independent variability sources by a single *statistically* equivalent device, the main goal here being to reduce the dimensionality space. This is especially

important for response surface modeling-based integration methods or variance reduction techniques.

The threshold voltage variation due to random dopant fluctuation [4] for the single device, however, is different from that of the equivalent device. In a typical linear model, the equivalent statistical model would be derived from linear relationships with respect to the individual devices, and what the designers often refer to as the square root law; as the device area (A) increases, the threshold voltage variation scales $\sim 1/\sqrt{A}$. This would let us model “ n ” small devices with one large device whose variation is $\sim 1/\sqrt{n}$ times the original variation. However, as was indicated in [5, 6] this model lacks accuracy when modeling the highly nonlinear leakage currents where the dependency is exponential on the threshold voltage (the source of variation). In fact, as we will explain in the following sections, the leakage current is distributed lognormally, and the problem of the equivalent device is known mathematically as the sum of lognormals. Mathematically, a true and exact closed form does not exist for the method, and historically there have been many approximations. There is a lot of literature on the sum of lognormals, and applications span a wide range of domains including economics, finance, actuarial sciences,

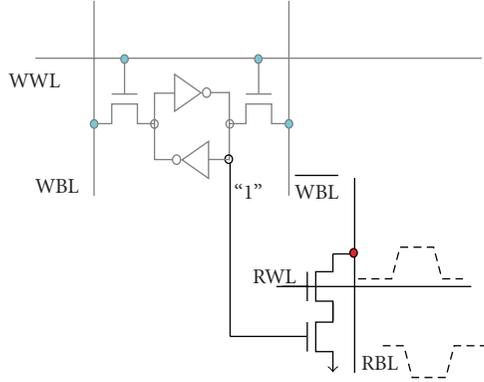


FIGURE 1: A schematic of an 8T cell/register file. During Read, the read wordline “RWL” turns ON. If the cell storage node is held high, “1”, the read bitline (RBL) is discharged.

and others, with engineering being one of the oldest. The most famous approximations to the sum of lognormals are Fenton-Wilkinson’s [7] and Schwartz-Yeh (SY) [8]. For circuit design, those models have been studied in [5, 6] to enable modeling the threshold voltage distribution of the equivalent device. The authors in [5] relied on the Fenton-Wilkinson’s (FW) method to maintain a reasonable 3-sigma estimate (99th percentile). It is often indicated that when $\sigma_{dB} > 4$, FW method may not result in proper approximation of the lognormal distribution. On the other hand, the authors in [6] proposed to use the Schwartz-Yeh (SY) methodology to model the equivalent threshold voltage distribution of a 3-fin device. Again, the approximation is qualified, in this case for a small number of summands, based on a thousand samples thereby ignoring the accuracy of the model in the tail regions which is critical to estimation of low fail probabilities. In the following sections, we review the modeling trends in the critical regions, focusing on the cumulative density function (CDF) matching method [9] as a viable solution for reliable statistical modeling when the tail region modeling is key to the accuracy of the yield estimate.

The paper is organized as follows. Section 2 introduces a basic circuit example to describe the need for the single equivalent device modeling. Section 3 reviews the mathematical background and assumptions of the most common sum of lognormals approximation methods. It also provides a summary chart of the different approaches under study. Section 4 evaluates the impact on fail probability estimations. Finally conclusions are presented in Section 5.

2. Equivalent Device Modeling

Figure 1 illustrates an 8T cell/register file. The circuit operates like a 6T SRAM cell during cell Write. During Read, the read wordline turns ON. If the cell is storing a “1” on the right-side-node, in this example, the Read stack will turn ON, thereby discharging the read bitline (RBL).

Driven by density requirements, many cells share the same bitline and sense amplifier (or read logic). Only one cell per column is accessed at a given time; other cells

remain unaccessed. For the example of Figure 1, the RWL of the unaccessed cells is turned OFF. However, based on the data storage, it is possible that the data stack device turns on, thereby rendering the stack leaky. Figure 2 illustrates a scenario where the accessed cell is OFF. RBL is expected to remain high. However, leakages through the accessed stack and unaccessed stacks may falsely discharge the RBL node. This together with noise, mismatch in the read logic, and other environmental effects can degrade the yield of the system and increase the probability of a false read.

Note that this phenomenon is also manifested in other memory and domino logic designs. For example in the case of 6T SRAM, the devices sharing the WBL (see Figure 1) can be storing 1’s thereby leaking onto the WBL and degrading the Read “0” performance in a 6T fashion. Hence there is a need to account for the statistical effects of the leaky devices in any design yield optimization and study. This can dramatically increase the dimensionality of the problem thereby reducing the efficiency of statistical integration, response surface methods, or any methods that try to model the failure region and that suffer from sparsity and the curse of dimensionality. Consider for example a study that involves variability in 64 devices on the bitline along with variability sense amp devices. Thus, a model for a statistically equivalent device to emulate the multiple leaky devices can significantly simplify the complexity. Figure 3 summarizes the problem.

Given: n devices T_1, \dots, T_n .

The threshold voltage variation δV_T is normally distributed with zero mean and standard deviation $\sigma_{V_{Ti}} = \sigma_0$.

That is, $\delta V_{Ti} \sim N(0, \sigma_0)$

Find: δV_{Teqv} distribution for the device T_{eqv} such that “ $n * I_{Teqv}$ ” is distributed according to “ $\sum I_{Ti}$ ”.

Note that we assume here independent and identical devices; the problem can be generalized to nonidentical/correlated devices.

Leakage Current Model. Leakage current as a function of the threshold voltage variation can be modeled according to (1). Hence it portrays an exponential dependency on the random variable δV_T

$$I_{leak} = e^{a\delta V_T + b}, \quad (1)$$

$$a \propto \frac{KT}{q},$$

q is the electron charge, K is boltzmann constant, and T is the temperature. For simplicity, T_{eqv} has the same device width as the other T_i , but it is followed by a current multiplier to model the equivalent leaky device and solve for the new δV_{Teqv} distribution: the set of (2) represents how the problem can be reduced. Parameters Z_i are normal variables, e^{Z_i} are lognormal variables (see next section for definitions), and $e^{Z_{eqv}}$ (and Y) are distributed according to a sum of lognormal. A sum of lognormals is not distributed lognormally, and hence Z_{eqv} is not normal. In the following

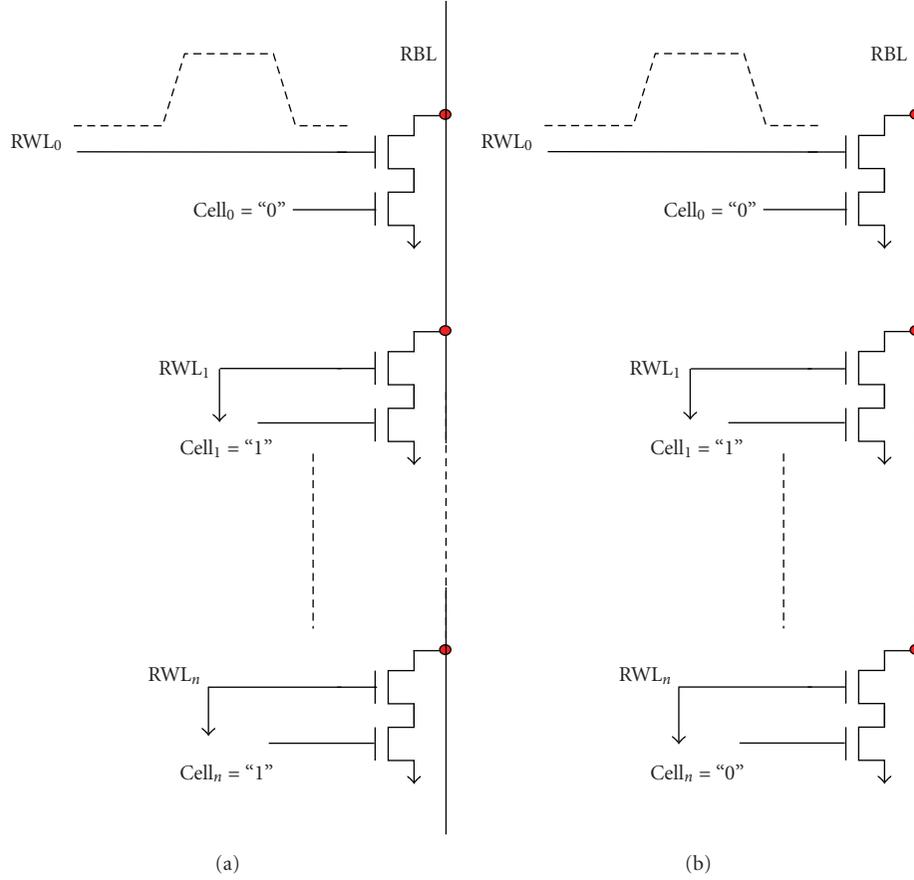


FIGURE 2: The access cell is storing a “0”. The read bitline (RBL) should stay high. Leakage from unaccessed cells can falsely discharge the bitline. Leakage can happen due to (a) all the cells on the bitline leaking, or (b) an intermediate number of cells leaking. For our work, we ignore the stack leakages if both read transistors (whose gates are the cell and RWL) are OFF.

section we review the sum of lognormals characteristics and possible methods of approximation that can be used to generate the statistical characteristics of Z_{eqv} and hence δV_{Teqv}

$$\begin{aligned}
 I_{\text{sum}} &= I_{\text{leak1}} + \dots + I_{\text{leakn}} = n * I_{\text{leak-eqv}}, \\
 I_{\text{sum}} &= \sum_{i=1}^n e^{a\delta V_{\text{Ti}}+b} = n * e^{a\delta V_{\text{Teqv}}+b}, \\
 Z_{\text{eqv}} &= a\delta V_{\text{Teqv}}; \quad Z_i = a\delta V_{\text{Ti}}, \\
 e^{Z_{\text{eqv}}} &= \frac{1}{n} \sum_{i=1}^n e^{Z_i} = \frac{1}{n} Y.
 \end{aligned} \tag{2}$$

3. Sum of Lognormals

In this section we will focus on methods to approximate models for Y (2), and compare their accuracy to the CDF matching approach. First we present some background information.

Lognormal Distribution. Let $Z = \ln(X)$. If Z is normally distributed $\sim N(\mu_z, \sigma_z)$, then X is lognormally distributed and its probability density function is [10]

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_z} e^{-(z-\mu_z)^2/2\sigma_z^2} & x > 0 \\ 0 & x \leq 0. \end{cases} \tag{3}$$

Often the Gaussian Variable $V = 10 \log_{10}(X)$ is defined. V is in decibels (dB), and $V \cong 4.34 Z$; if $Z \sim N(0, 1)$, $V \sim N(0, 4.34 \text{ dB})$.

Sum of Lognormals. A common way to compute the probability density function (pdf) of a sum of independent random variables, would be from the product of the characteristic functions, CF, of the summands [11]; a characteristic function of random variable Z is the expected value ($E[e^{jwZ}]$). However, for the case of the sum of lognormals, $Y = \sum e^{Z_i}$, a closed form of the CF does not exist, and approximations are used instead to represent the density function of the sum of lognormals. Some approximations

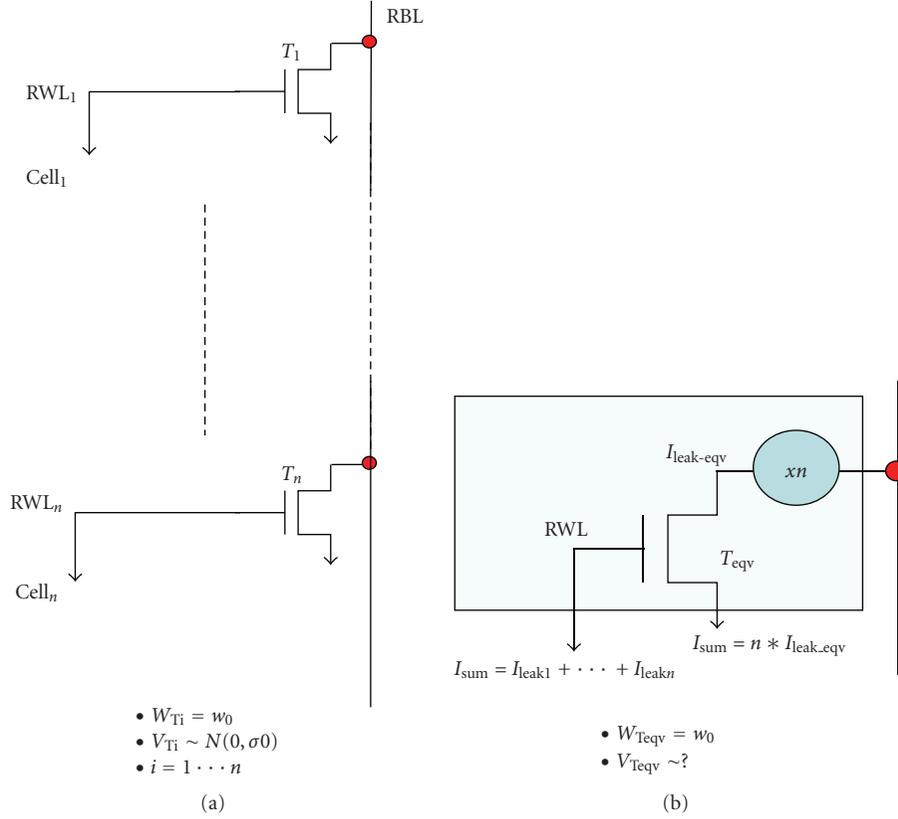


FIGURE 3: The problem of multiple devices leakages can be statistically equivalent to that of a single device. The distribution of the threshold voltage of the equivalent device is to be solved for.

Approximate $Y = \sum e^{z_i}$			
Fenton and Wilkinson $Y_{fw} = e^{Z_{fw}}$	Schwartz and Yeh $Y_{sy} = e^{Z_{sy}}$	log moment and matching $Y_{lmm} = e^{Z_{lmm}}$	CDF matching Y_{cdf}
Compute $\mu_{Z_{fw}}, \sigma_{Z_{fw}}$	Compute $\mu_{Z_{sy}}, \sigma_{Z_{sy}}$	Sample-based $\mu_{Z_{lmm}}, \sigma_{Z_{lmm}}$ estimates	Sample-based build PWL-PDF $g_{pwl}(Y_{cdf})$
$z_{fw} \sim N(\mu_{Z_{fw}}, \sigma_{Z_{fw}})$ • Y_{fw} matches 1st two moments of Y	$Z_{sy} \sim N(\mu_{Z_{sy}}, \sigma_{Z_{sy}})$ • Recursive solve for μ and σ of $\ln(Y)$	$Z_{lmm} \sim N(\mu_{Z_{lmm}}, \sigma_{Z_{lmm}})$ • Match moments of $\ln(Y)$	$\ln(Y_{cdf}) \sim g_{pwl}(e^{\ln(Y_{cdf})})$ • Non gaussian • Derived from the piecewise linear distribution

FIGURE 4: Different methods to be compared against the CDF matching approach involve moment matching to $\ln(Y)$, Schwartz-Yeh, and Fenton-Wilkinson methods.

include modeling the sum of lognormals as an “approximately” lognormal variable (see (4)); where Z_m is normally distributed (m refers to the method of approximation)

$$Y = \sum_{i=1}^n e^{z_i} \approx e^{Z_m} \tag{4}$$

Common techniques are SY and FW methods mentioned earlier. In the following section, we will revisit those methods and compare them to CDF matching. Figure 4 summarizes the four techniques that will be visited:

- (1) Fenton-Wilkinson method,
- (2) Schwartz-Yeh method,

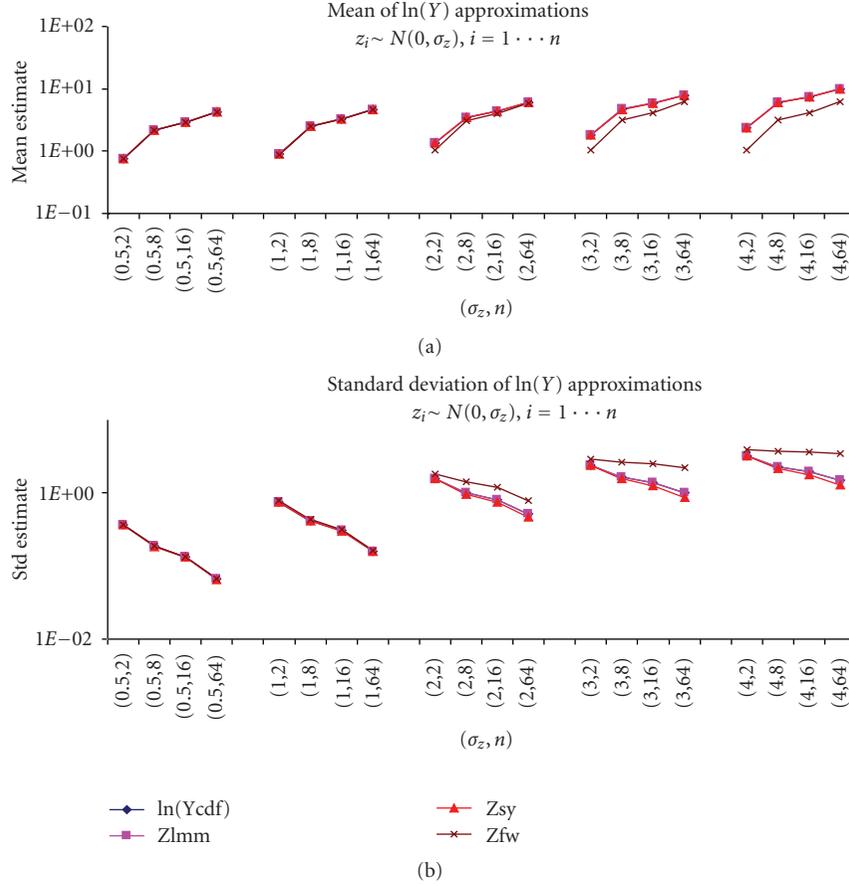


FIGURE 5: Plot of the mean and standard deviation of $\ln(Y)$ obtained from the different approximations. FW method targeted to match moments of Y falls behind for larger σ_z (>1 or 4 dB).

- (3) Log Moment Matching approximation,
- (4) CDF matching.

All these methods except the CDF matching one assume that Y is a lognormal variable; that is, that Z_m (and hence δVt) is a normally distributed Gaussian variable whose mean and standard deviation are to be found.

3.1. Fenton-Wilkinson. Fenton and Wilkinson method estimates the standard deviation and mean of the Gaussian variable Z_{fw} by matching the first two moments of Y , $E[Y]$ and $E[Y^2]$, to those of the sum of lognormals (the two sides of (4)) according to (5)

$$Y = e^{Z_{fw}},$$

$$E[e^{Z_{fw}}] = E[e^{Z_1} + \dots + e^{Z_n}], \quad (5)$$

$$E[e^{2Z_{fw}}] = E\left[\left(e^{Z_1} + \dots + e^{Z_n}\right)^2\right].$$

It is a well-known fact [10] that given a Gaussian variable, x , with mean and standard deviation μ_x and σ_x then (6) holds

$$E[e^{nx}] = e^{n\mu_x + (1/2)n^2\sigma_x^2} \quad (6)$$

by relying on (5) and (6) we can obtain a closed form for $\mu_{Z_{fw}}$ and $\sigma_{Z_{fw}}$ according to (7) for the case of independent variables $z_i \sim N(\mu_i, \sigma_z)$

$$\sigma_{Z_{fw}} = \sqrt{\ln \left[\frac{(e^{\sigma_z^2} - 1) \sum e^{2\mu_i}}{(\sum e^{\mu_i})^2} + 1 \right]}, \quad (7)$$

$$\mu_{Z_{fw}} = \ln \left(\sum e^{\mu_i} \right) + \frac{\sigma_z^2}{2} - \frac{\sigma_{Z_{fw}}^2}{2}.$$

3.2. Schwartz and Yeh. Again the methodology starts with the assumption that the sum of lognormals is approximately lognormally distributed

$$Y = e^{Z_{sy}}, \quad (8)$$

where Z_{sy} is a Gaussian variable $\sim N(\mu_{Z_{sy}}, \sigma_{Z_{sy}})$. It approximates the mean and standard deviation of the $Z_{SY} = \ln(Y)$, with numerical recursion; whereas FW estimates the first two moments of Y . Schwartz and Yeh method relies on the ability to exactly compute the mean, $\mu_{Z_{sy}}$, and standard deviation, $\sigma_{Z_{sy}}$, of $Z_{SY} = \ln(Y)$ for the case when the number of lognormal variables in the sum $n = 2$. For $n > 2$, the method

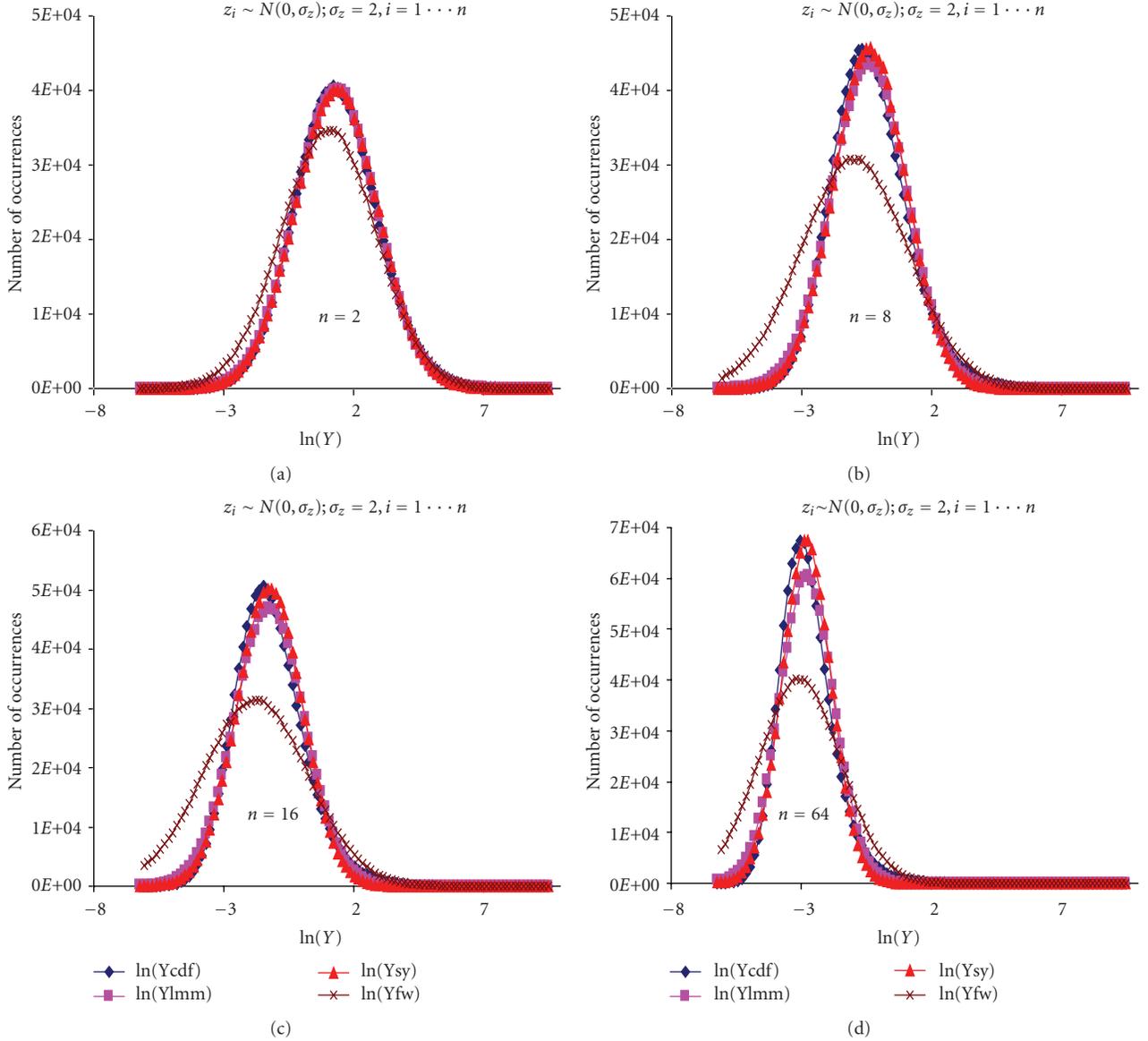


FIGURE 6: Histogram (pdf plot) of $\ln(Y)$ shows that the FW method does not model the body of the (z or δV_t) distribution well compared to other methods.

then relies on a recursive approach, adding one factor at a time. Hence, the mean and standard deviation of

$$Z_{sy} = \ln(e^{Z_1} + \dots + e^{Z_n}) \quad (9)$$

can be derived from the following set of generalized equations (10); $Z_{sy(k-1)}$ is assumed to be normally distributed; $k = 2, \dots, n$, and the Z_i are assumed to be uncorrelated in (10)

$$\begin{aligned} Z_{sy(k)} &= \ln(e^{Z_{sy(k-1)}} + e^{Z_k}), \\ \mu_{Z_{sy(k)}} &= \mu_{Z_{sy(k-1)}} + G_1, \\ \sigma_{Z_{sy(k)}}^2 &= \sigma_{Z_{sy(k-1)}}^2 + G_2 - 2 * \sigma_{Z_{sy(k-1)}} \frac{G_3}{\sigma_{w(k)}^2}, \end{aligned} \quad (10)$$

where w_k and G_i ($i = 1, 2, 3$) are defined in (11). Finally set of (12)-(13) illustrates how the G_i 's can be computed according to [12]; this slightly modified implementation was intended to circumvent the round off error of integration of the original Schwartz and Yeh implementation [13]

$$\begin{aligned} w_k &= Z_k - Z_{sy(k-1)}, \\ G_i &= G_i(\mu_{w_k}, \sigma_{w_k}) \quad i = 1, 2, 3, \\ \mu_{w_k} &= \mu_{Z_k} - \mu_{Z_{sy(k-1)}}, \\ \sigma_{w_k}^2 &= \sigma_{Z_k}^2 + \sigma_{Z_{sy(k-1)}}^2, \end{aligned} \quad (11)$$

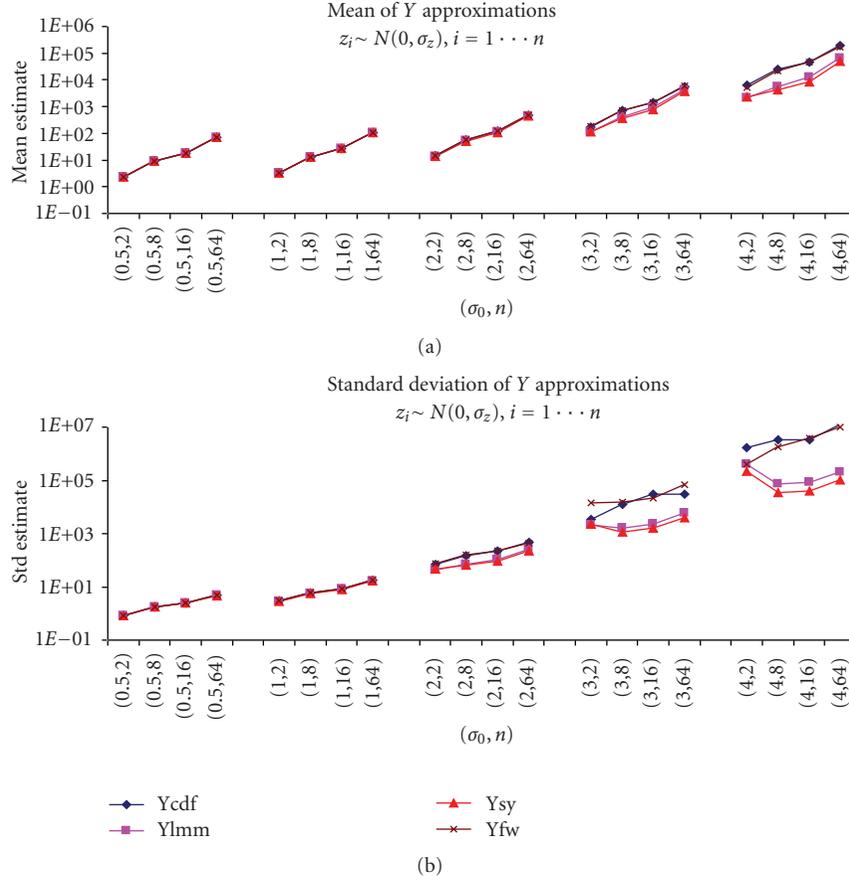


FIGURE 7: Plot of the mean and standard deviation of Y obtained from the different approximations. SY and LMM methods that match moments of $\ln(Y)$ falls behind for larger σ_z (>1 or 4 dB).

$$G_1 = A_0 + I_1,$$

$$G_2 = I_3 + 2I_4 + \sigma_{wk}^2 I_0 + \mu_{wk} A_0,$$

$$G_3 = \sigma_{wk}^2 (I_0 + I_2),$$

$$A_0 = \frac{\sigma_{wk}}{\sqrt{2\pi}} e^{-\eta^2/2} + \mu_{wk} I_0,$$

$$\eta = -\frac{\mu_{wk}}{\sigma_{wk}},$$

$$I_i = \int_0^1 h_i(v) v^{-1} dv = \int_{-\infty}^{\infty} 2h(v) v u \cosh x dx = \int_{-\infty}^{\infty} H(x) dx,$$

$$u = \exp(-2 \sinh x); \quad v = (1 + u)^{-1}. \quad (12)$$

Thus, at each step of the recursion, we compute the mean and standard deviation of w_k . The integrals I_i are then numerically computed using the functions h defined in (13). Their values are used to solve for the G_i 's in order to evaluate

(10). The final estimate for the Z_{SY} mean and standard deviation is reached at $k = n$

$$h_i(v) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-(\ln v - \eta)^2/2}, & i = 0, \\ [f_w(\ln v) - f_w(-\ln v)] \ln(1 + v), & i = 1, \\ [f_w(\ln v) - f_w(-\ln v)] (1 + v^{-1})^{-1}, & i = 2, \\ [f_w(\ln v) - f_w(-\ln v)] \ln^2(1 + v), & i = 3, \\ -f_w(-\ln v) \ln(v) \ln(1 + v), & i = 4, \end{cases} \quad (13)$$

$$f_w = \frac{1}{\sqrt{2\pi\sigma_{wk}^2}} e^{-(w_k - \mu_{wk})^2/2\sigma_{wk}^2}.$$

3.3. Log Moment Matching Approximation. While the previous two approaches find the moment matching using analytical or recursive analytical solutions, this approach relies on sampling instead to compute the moments according to (10) [9]. Similar to the SY method, it estimates the 1st two moments of $\ln(Y)$. It does not pertain to a closed-form solution, but it maintains the lognormal assumption. Thus, the recursive analytical approach in SY and this

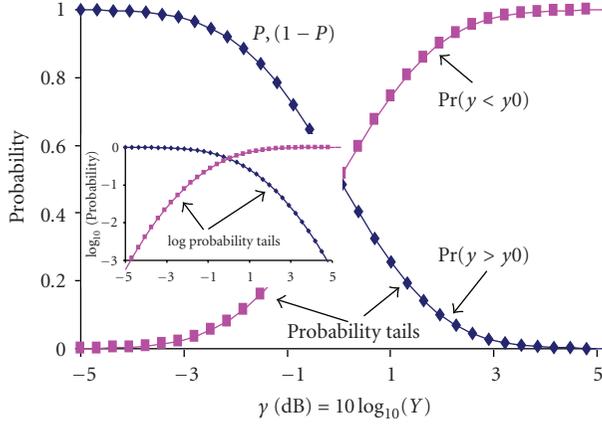


FIGURE 8: To highlight the tail region modeling, we rely on a tail log probability plot (small graph). It is derived from the density function and its complement.

approximation do not always converge as we will see in Section 3.5, especially that the SY assumptions is held exact when $n = 2$

$$\begin{aligned}
 Y &\approx e^{Z_{\text{linm}}}, \\
 \mu_{Z_{\text{linm}}} &= E[\ln(Y)], \\
 \sigma_{Z_{\text{linm}}} &= \sqrt{\left(E[(\ln Y)^2] - E[\ln(Y)]^2 \right)}.
 \end{aligned} \tag{14}$$

3.4. CDF Matching. This is the closest possible to the true sum. Unlike the previous approximations, it does not rely on the lognormal approximation. We build a piece-wise linear probability density function from the Monte Carlo samples. Our goal is to demonstrate the difference between this approach and other lognormal approximations, when it comes to tail regions modeling and the resultant probability of fail estimations. Key features of the problem are

- (i) the piece-wise linear fit for the density function of $\ln(Y)$ is non-Gaussian.
 - (a) $g_{\text{pwl}}(\cdot)$ is defined by pairs (P_j, Y_{0j}) , such that $\text{Prob}(Y > Y_{0j}) = P_j$, hence $\text{Prob}(\ln(Y) > \ln(Y_{0j})) = P_j$.
 - (b) The pwl function can be sparse in the center of the distribution and more dense in the tails to adequately model the low fail probability region. In an extreme fashion, the tail probabilities can be recomputed from tail samples to avoid interpolation errors.
- (ii) generating the Y samples is cheap, so is the I_{sum} sample once the function or even tables of $(I, \delta V_T)$ are available. Interpolation, bootstrapping, and other techniques can reduce the number of real simulations needed and still enable good confidence in the density function. After all, the previous approaches do rely on the availability of a closed form function for $I(\delta V_T)$.

The number of samples is inversely proportional to the tail probability of interest; for example, if we are looking for accurate probabilities in the range of $1e - 4$, then we need to have replications of samples that are larger than $1e4$. Replications add to the confidence in the expected tail probability. After all the interest is in the CDF tails mainly,

- (iii) finally, this model can accommodate any complex nonlinear function ($I = f(\delta V_T)$); even if it is different from the exponential approximation above,
- (iv) most importantly, once the distribution of I_{sum} is available, V_T distribution is derived by reverse fitting ($f^{-1}(I_{\text{sum}})$) samples.

Note that for purposes of comparison, in this study all methods share the same exponential current model.

3.5. Theoretical Experiments. In this section, we compare the different methods ability to model Y (Figure 4). We study both the moment matching abilities as well as the upper and lower CDF tails. The study is performed over different combinations of n and σ_z (the standard deviation of Z_i):

$$(\sigma_z = 0.5, 1, 2, 3, 4) \times (n = 2, 8, 16, 64) \tag{15}$$

σ_z range corresponds to the range 2 dB–16 dB. Recall that often 4 dB is considered as a critical threshold for how accurate FW methods can model the distribution. Also as demonstrated in [6] standard deviation of the threshold voltage of scaled devices can exceed 4–8 dB (this is based on the leakage coefficient in (1)).

For each (n, σ) combination, multiple replications of 1 million samples are studied; this enables good estimates for the low fail probabilities. Figure 5 plots the mean and standard deviation of $\ln(Y)$ obtained from the different approximations. FW method falls behind for larger σ_z (>1 or 4 dB). Recall that this method was intended to match the moments of Y and not $\ln(Y)$. *We note that the FW estimates do underestimate the mean and overestimate the variance of $\ln(Y)$ for larger dB values.* Finally, the SW and LMM methods do match the CDF well, yet they do differ from each other a bit for larger sigma values, given that the former is a recursive approximation.

Figure 6 illustrates histogram (pdf plot) of $\ln(Y)$; the FW method does not model the body of the $(Z$ or $\delta V_T)$ distribution well compared to other methods; this was also indicated in [6] for small n . The trend is even more obvious as the number of summands (n) increases. However, to obtain a complete picture, there is a need to study the lognorm matching and the tail regions which we will cover next.

Figure 7 plots the mean and standard deviation of Y obtained from the different approximations. SY and LMM methods that match moments of $\ln(Y)$ falls behind for larger σ_z (>1 or 4 dB). This is true for large n (>2), and we note that SY tends to particularly underestimate the variance of the sum of lognorms; this effect is most visible when the variables are identical and uncorrelated as is the case in this study. To study the tail regions, we rely on a “tail log plot”

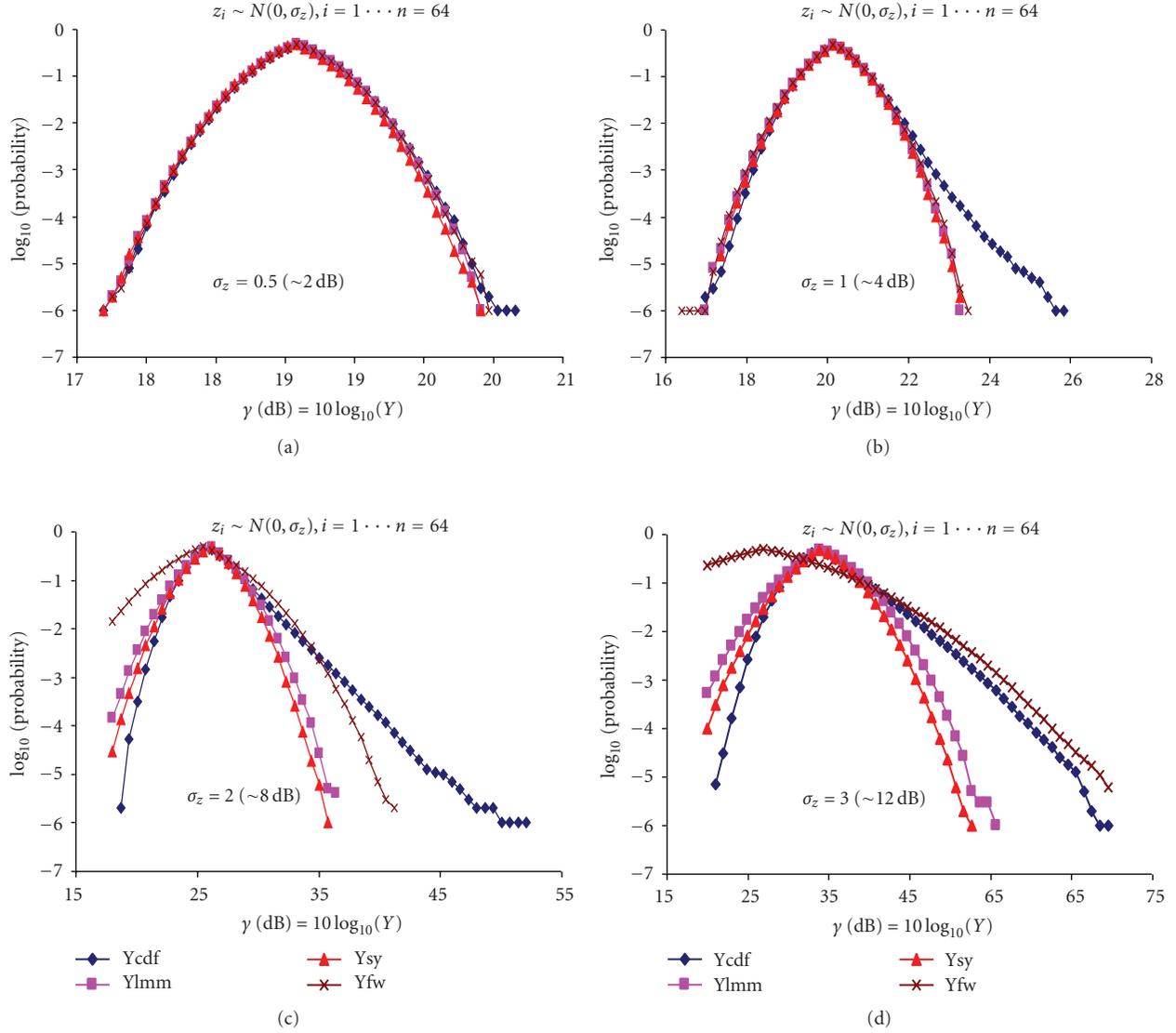


FIGURE 9: Log probability tail plot as function of σ_z ; $n = 64$. For critical σ range (4–8 dB) SY, FW and LMM methods miss the right tail.

as illustrated in Figure 8. Without loss of generality we set the x -axis in these plots to be in dB; a small shift can mean large change in Y values. The plot is derived from the density function ($P(y > y_0)$) and its complement ($1 - P$). It is such tail probabilities that are linked to the fail probabilities of a design. *Note that for the case of leakages, the right-side tail is critical for the fails (larger Y values correlate with larger leakage values in real applications).*

Figure 9 illustrates the tail probability plot as function of σ_z for $n = 64$. We get good match for all the methods for small σ . For critical σ values (4–8 dB), we note that SY and FW do miss the right tail modeling. As σ increases, FW tries to catch towards modeling the right tail by missing on the left tail model. Figure 10 illustrates the tail probability plots as function of n for $\sigma_z = 2$ (8 dB). SY and LMM methods have larger errors in modeling the right tail. FW error increases with increasing n .

4. Case Study: Leak-Down Time Comparisons

In this section, we extend the analysis to study the impact of the different modeling schemes, compared to the CDF matching method, on the probability of fail estimations. Figure 11 illustrates a summary of δV_{TeqV} distribution approximations based on the relation between δV_{TeqV} and Y in (2); except for the CDF matching method, δV_{TeqV} is modeled as Gaussian random variable $\sim N(\mu_{\text{eqV}}, \sigma_{\text{eqV}})$. The distributions are then used to analyze the leak-down of RBL in the circuit of Figure 12. The time for the bitline to discharge to 50% of the rail under extreme noisy corner voltage is then estimated.

Figures 13, 14, and 15 illustrate the normalized time-to-leak distributions for the case of 16 cells/bitline. s_0 represents a lower limit on the threshold voltage standard deviation of a 45 nm cell device; its value is set to the equivalent of $\sigma_z = 0.6$

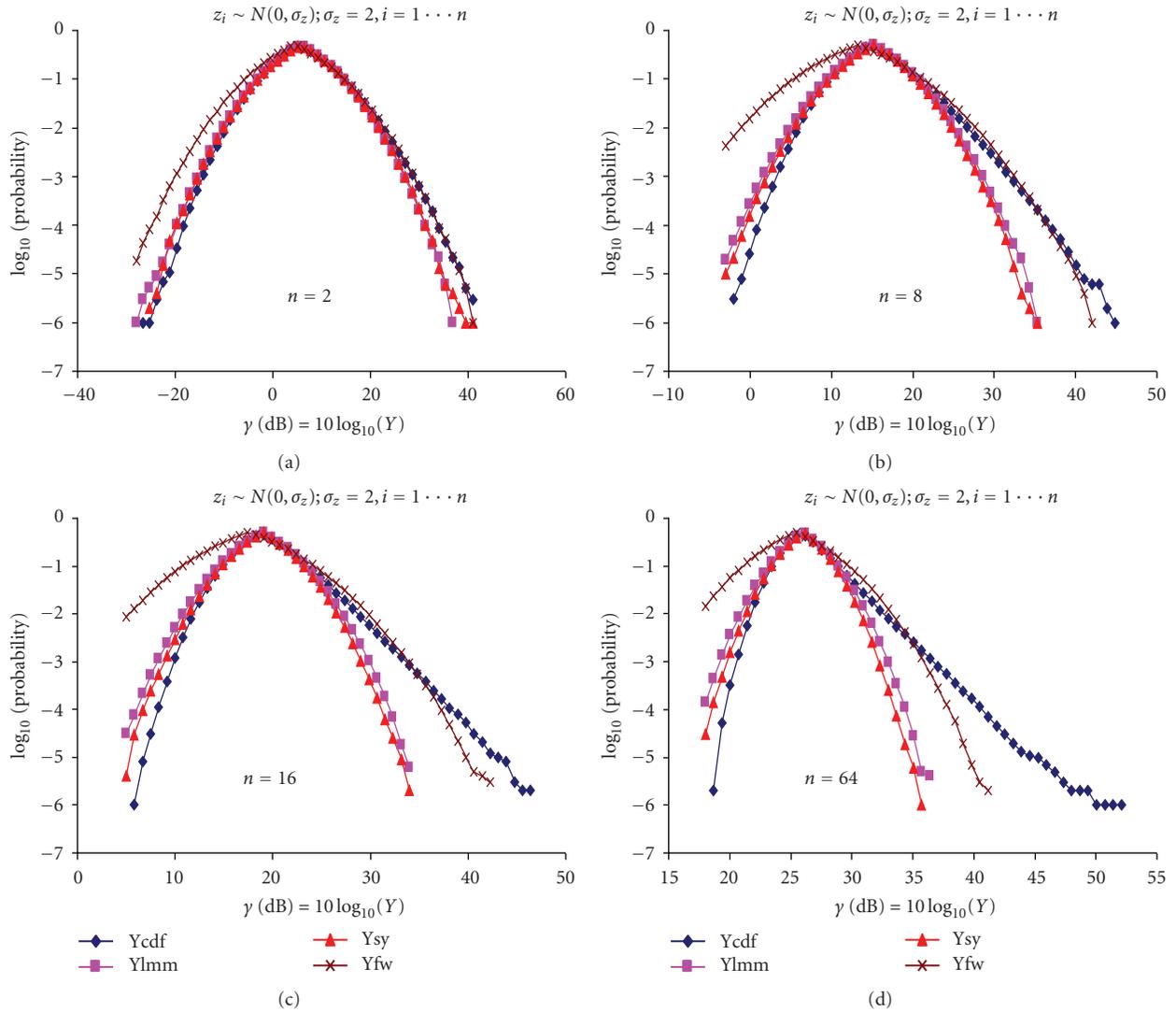


FIGURE 10: Log probability tail plot as function of n , $\sigma_z = 2$ (8 dB). SY and LMM methods show large error in the right tail model. FW error increases as the number of summands increases.

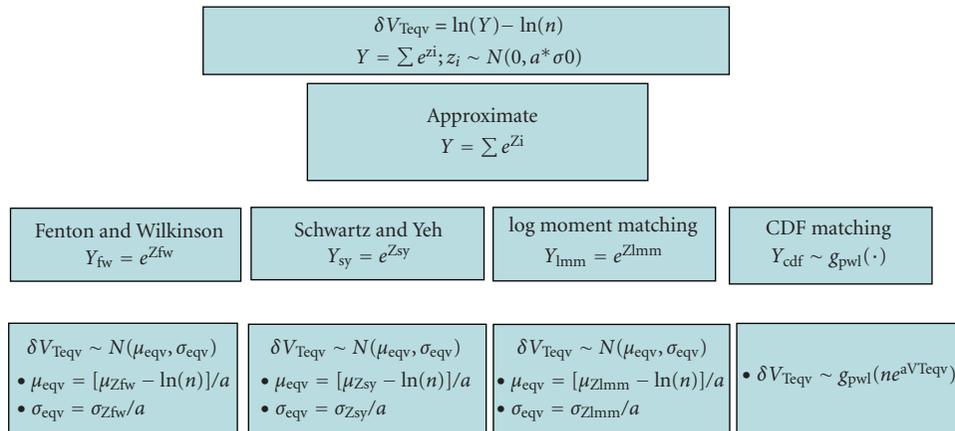


FIGURE 11: Based on Figure 4, distributions for the δV_{Teqv} can be derived from the Y approximations.

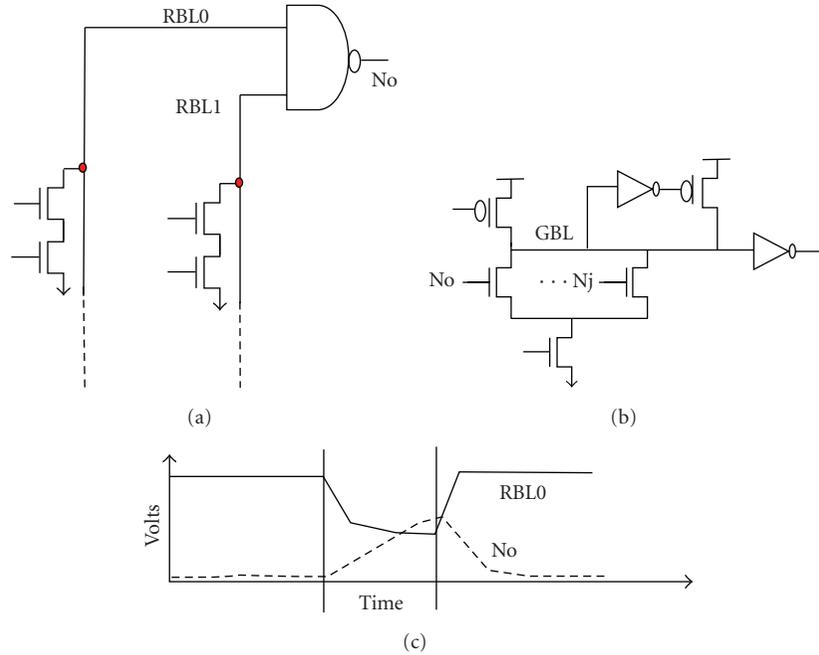


FIGURE 12: Due to leakages, the RBL may falsely discharge causing false evaluate of node N_0 .

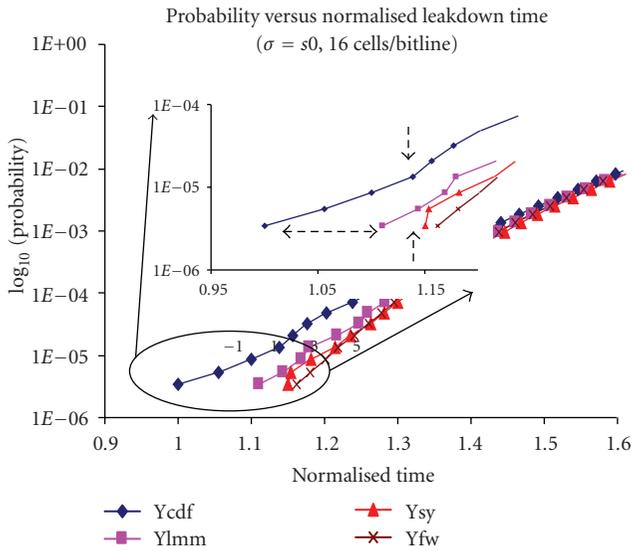


FIGURE 13: Normalized time-to-leak probability for $\sigma_{VT} \sim s_0 \sim 2.5$ dB.

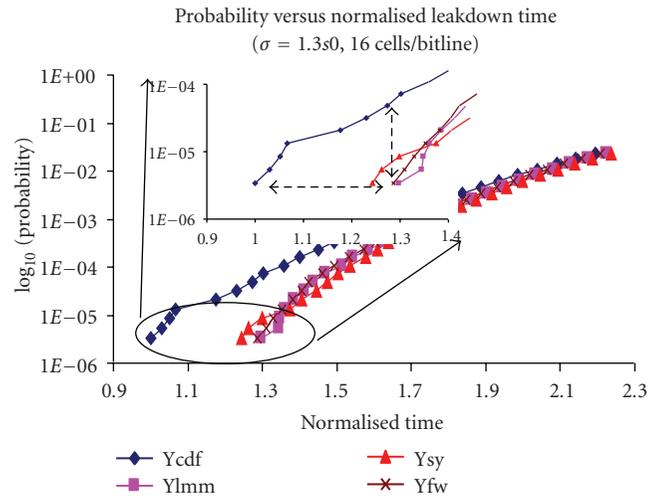


FIGURE 14: Normalized time-to-leak probability plot for $\sigma_{VT} \sim 1.3s_0$.

(or 2.5 dB). This is conservative especially that as technology scales we expect more variability and additional sources of variability like the random telegraphic noise can add to the V_T variation; cases for standard deviation of $1.3s_0$ and $1.6s_0$ are also plotted. SY, FW, and LMM methods overestimate the time-to-leak from 10% for s_0 to close to 100% for $1.6s_0$ (see horizontal arrow in the figures; this corresponds to system yield around 4.5 sigma). More importantly, this leads to underestimating the probability of the number of elements failing at a given leak-time. Note that time-to-leak values can be critical relative to operating frequencies

and accurate prediction is needed for robust designs. Thus, we are interested in computing the ratio of the probability of fails (vertical arrows in figures) for predicted time-to-leak values. Figure 16 summarizes the ratio of the true (cdf) probability of fail compared to that of the other methods (SY, FW, and LMM) at their 4.5 sigma yield leak-time. Each experiment is based on the average of 25×1 million replications. This is done for the case of 16 and 64 cells/bitline and at increments of $0.1 * s_0$. We note that the SY, FW, and LMM methods underestimate the probability of fail $10 \times - 147 \times$.

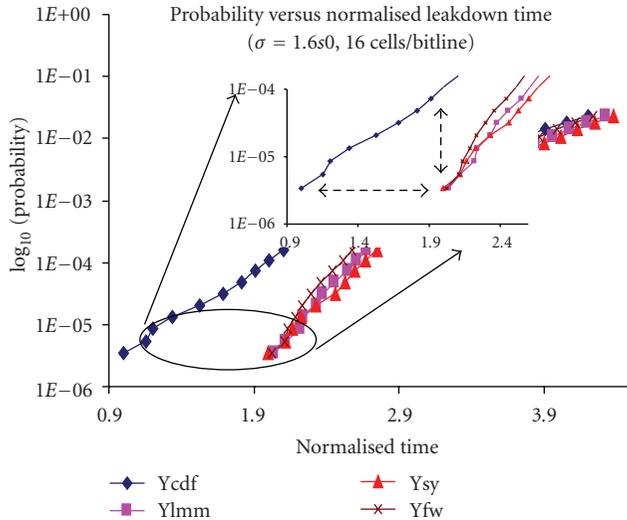


FIGURE 15: Normalized time-to-leak probability plot for $\sigma_{VT} \sim 1.3s_0$.

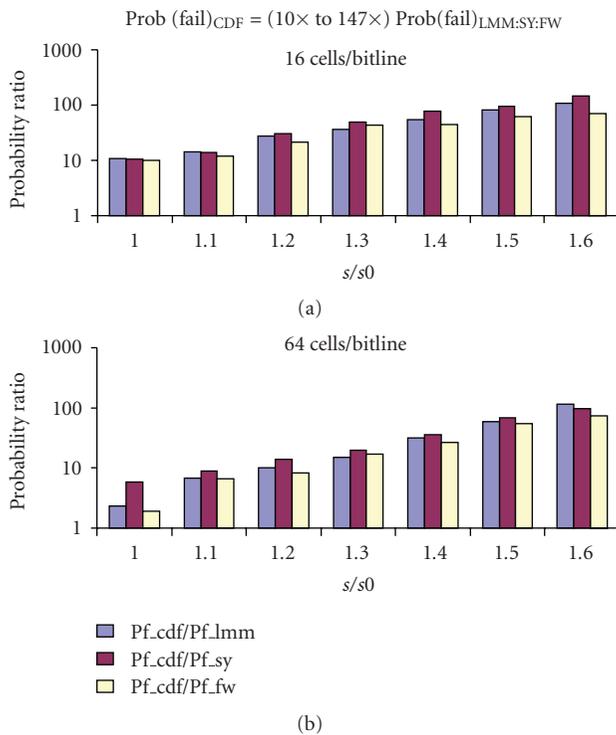


FIGURE 16: Probability of fail ratio is underestimated in SY, FW, and LMM methods.

5. Conclusions

We study the ability of different sums of lognormal approximation to emulate the leakage of multiple leaky devices by a single equivalent device. With the goal of rare event estimation, tail distributions are examined closely. Modeling the tail probability by CDF matching is found to be critical compared to the Fenton-Wilkinson and Schwartz-Yeh methods that are found to underestimate the sum of

lognorms and hence overestimate the fail probability by $10\times-147\times$; this trend is expected to increase as the variability increases with scaling technology.

References

- [1] R. Joshi, A. Pelella, A. Tuminaro, Y. Chan, and R. Kanj, "The dawn of predictive chip yield design: along and beyond the memory lane," *IEEE Design and Test of Computers*, vol. 27, no. 6, pp. 36–45, 2010.
- [2] E. Acar, K. Agarwal, and S. R. Nassif, "Characterization of total chip leakage using inverse (reciprocal) gamma distribution," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '06)*, pp. 3029–3032, May 2006.
- [3] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *Proceedings of the 43rd Annual Design Automation Conference (DAC '06)*, pp. 69–72, 2006.
- [4] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 fjm MOSFET's: a 3-D "atomistic" simulation study," *IEEE Transactions on Electron Devices*, vol. 45, no. 12, pp. 2505–2513, 1998.
- [5] J. Gu, S. S. Sapatnekar, and C. Kim, "Width-dependent statistical leakage modeling for random dopant induced threshold voltage shift," in *Proceedings of the 44th ACM/IEEE Design Automation Conference (DAC '07)*, pp. 87–92, June 2007.
- [6] S. H. Rasouli, K. Endo, and K. Banerjee, "Variability analysis of FinFET-based devices and circuits considering electrical confinement and width quantization," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD '09)*, pp. 505–512, November 2009.
- [7] L. F. Fenton, "The sum of log-normal probability distributions in scattered transmission systems," *IRE Transactions on Communications Systems*, vol. 8, pp. 57–67, 1960.
- [8] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with log-normal components," *The Bell System Technical Journal*, vol. 61, no. 7, pp. 1441–1462, 1982.
- [9] R. Kanj, R. Joshi, and S. Nassif, "Statistical leakage modeling for accurate yield analysis: the CDF matching method and its alternatives," in *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED '10)*, pp. 337–342, Austin, Tex, USA, 2010.
- [10] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Estimating the distribution of a sum of independent lognormal random variables," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2869–2873, 1995.
- [11] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw Hill, New York, NY, USA, 1991.
- [12] P. Cardieri and T. S. Rappaport, "Statistics of the sum of log-normal variables in wireless communications," in *Proceedings of the 51st Vehicular Technology Conference (VTC '00)*, pp. 1823–1827, May 2000.
- [13] C. L. Ho, "Calculating the mean and variance of power sums with two log-normal components," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 4, pp. 756–762, 1995.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

