Stochastic Behavioral Modeling of Analog/Mixed-Signal Circuits by Maximizing Entropy

Rahul Krishnan, Wei Wu, Fang Gong, Lei He
Electrical Engineering
University of California, Los Angeles
Los Angeles, CA, 90095
{rahulk, weiwu, fang08, lhe}@ee.ucla.edu

ABSTRACT

Maximum entropy (MAXENT) is a powerful and flexible method for estimating the arbitrary probabilistic distribution of a stochastic variable with moment constraints. However, modeling the stochastic behavior of analog/mixed-signal (AMS) circuits using MAXENT is still unknown. In this paper, we present a MAXENT based approach to efficiently model the arbitrary behavioral distribution of AMS circuits with high accuracy. The exact behavioral distribution can be approximated by a product of exponential functions with different Lagrangian multipliers. The closest approximation can be obtained by maximizing Shannon's information entropy subject to moment constraints, leading to a nonlinear system. Classic Newton's method is used to solve the nonlinear system for the Lagrangian multipliers, which can further recover the arbitrary behavioral distribution of AMS circuits. Extensive experiments on different circuits demonstrate that the proposed MAXENT based approach offers better stability and improves the accuracy up to 110% when compared to previous AWE-based moment matching approaches, and offers up to 592x speedup when compared to Monte Carlo method.

1. INTRODUCTION

As technology pushes the envelope for transistor sizes to get much smaller, the impact of process variations has become more prominent than ever before, causing serious issues in AMS circuit design and manufacturing [1, 2, 3]. For example, parameters such as effective channel length, width and oxide thickness can significantly deviate from their designed values due to the uncertainties from etching, lithography, and other manufacturing processes. Therefore, performance of the circuit can deviate from its nominal values by a large amount, resulting in significant yield loss [4, 5]. Consequently, efficient modeling of circuit performance in the presence of process variations is highly desired.

In the past few years, many statistical methods such as Monte Carlo (MC) simulation [6], stochastic orthogonal polynomials (SOP) [7], response surface modeling (RSM) approaches [3], and point estimation method (PEM) [1] have been proposed for behavioral modeling of AMS circuits under process variations. Among different methods, the Monte Carlo (MC) method [6] is the "golden standard" approach that reliably provides the accurate distribution of circuit performance. In general, MC method repeatedly generates random samples of all variables and evaluates their performance merits. It is obvious that MC method needs to evaluate a huge number of samples so as to cover the entire sampling space, which is extremely time-consuming. To relieve the high complexity, an SOP based method [7] has been proposed. This method can expand a probabilistic distribution with specific polynomial functions and calculate the expansion coefficients efficiently. However, SOP based methods need to know the distribution type of the circuit performance since each specific distribution type has a corresponding specific orthogonal polynomial. Unfortunately, the distribution type is usually unknown. To remedy this, [3] proposed a response-surface-model (RSM) based method in order to predict "arbitrary" behavioral distributions of AMS circuits, where the circuit performance is modeled as a polynomial function of all variable parameters. This approach shows significant improvement over MC and SOP methods, but becomes impractical for strongly nonlinear and high-dimensional cases. To deliver a practical solution, the Point Estimation Method (PEM) [1] has been proposed which estimates the high order moments using point estimation [8] and predicts the "arbitrary" behavioral distribution by the moment-matching method known as Asymptotic Waveform Estimation (AWE) [9]. It is worthwhile to point out that all previous AWE based works have two drawbacks: (1) the use of AWE may lead to instabilities, and (2) the convergence to the exact behavioral distribution is very slow as the number of moments is increased.

In this paper, a maximum entropy (MAXENT) based approach is proposed to model the "arbitrary" behavioral distributions for AMS circuits. The exact behavioral distribution can be approximated by a product of exponential functions with different Lagrangian multipliers. The closest approximation can be obtained by maximizing Shannon's information entropy subject to moment constraints, leading to a nonlinear system. Classic Newton's method is used to solve the nonlinear system for the Lagrangian multipliers, which can further recover the arbitrary behavioral distribution of AMS circuits. Extensive experiments on many circuits have validated that the proposed method is more robust and offers up to 110% higher accuracy compared to

[1], and up to 592x speedup compared to MC method [6].

The rest of the paper is organized as follows. In section 2, we give some background on the issue of process variations and previous stochastic modeling algorithms. In section 3, we describe the proposed algorithm. In section 4, we present experimental results detailing the robustness and accuracy of MAXENT. In section 5 we conclude this paper.

2. BACKGROUND

2.1 Problem Formulation

Process variations denote the uncertainties or variations in process parameters (such as channel width, oxide thickness, etc.), which differ from their designed values and cause the deviation of circuit performance. These variations are caused by the uncertainties during manufacturing processes such as etching, lithography and etc.

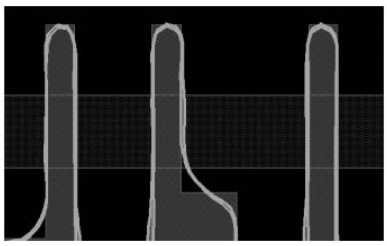

Figure 1: Transistor under Process Variation due to Lithography variations [10]

We show the deviation of geometric parameters in Figure 1 [10] that causes a significant shift in circuit performance. The circuit performance is not a deterministic value anymore but instead becomes a probabilistic value. Consequently, some fabricated circuits may lay outside the performance constraints from designers, leading to yield loss. Therefore, it is important to predict the probabilistic distribution of circuit performance.

In general, the problem can be formulated in Figure 2, which consists of two domains: the parameter domain contains probabilistic distributions of all variable parameters (such as oxide thickness), while the performance domain has the probabilistic distribution of circuit performance (such as delay). Clearly, there is a missing link/mapping between these two domains. The motivation behind this work is to explore an efficient mapping approach so as to estimate the behavioral distribution by the distributions of variable process parameters.

2.2 Sampling Methods

Sampling is a key aspect in stochastic modeling. Given a large set of data, sampling allows us to compress this large set into a smaller set while still accurately representing the statistics of the original set. There are many different types of sampling methods that we may use to achieve this such as random sampling, systematic sampling, and quasi-random sampling. Random sampling attempts to give each data point an equal probability of being selected. Systematic sampling [11] attempts to organize the target data according to a known ordering scheme. In this paper, we utilize a type of quasi-random sampling that utilizes the Sobol sequence [12]. We apply the Sobol sequence along with an inverse

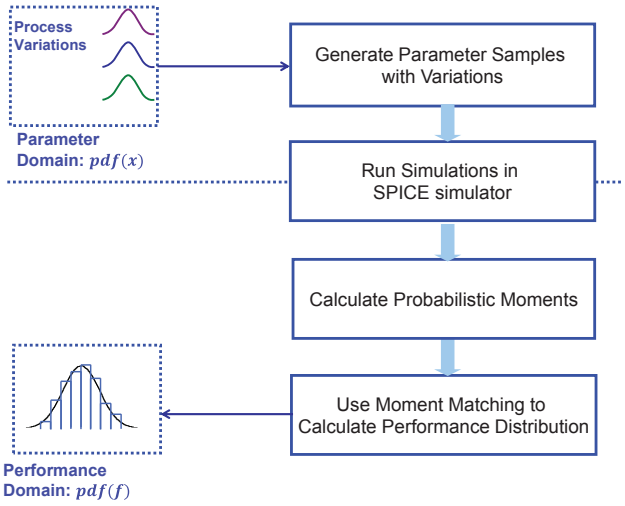

Figure 2: General Stochastic Modeling Method utilizing SPICE Simulation

CDF function to make sure the sampling space of variable parameters is uniformly covered.

2.3 Calculation of Moments

We first explain two different types of moments defined in the statistics community [13] and signal processing field [14]. These two types of moments need to be matched in the AWE moment matching method.

2.3.1 *Probabilistic Moments*

In statistics, the k-th moment of variable f is defined as:

$$m_f^k = E(f^k) = \int_{-\infty}^{+\infty} f^k \cdot pdf(f)df \tag{1}$$

where $E(\cdot)$ is the expectation operator and $pdf(f)$ is the distribution of variable f. In particular, the first four probabilistic moments are the mean, variance, skewness and kurtosis, respectively.

2.3.2 *Time Moments*

The definition of a "time moment" has been established in the signal processing field [14] for a long time and has been successfully applied to circuit analysis in the past few years [9, 15, 16]. In fact, the time moment is the coefficient of a Taylor series expansion of the homogeneous response in the Laplace domain and the k-th time moment can be expressed as:

$$m_t^k = \frac{(-1)^k}{k!} \int_{-\infty}^{+\infty} t^k \cdot h(t)dt \tag{2}$$

where t is time and $h(t)$ is the impulse response of a linear time invariant (LTI) system. Moreover, the time moments can be further expanded using the residues and poles of this LTI system as [9, 14]:

$$m_t^k = -\sum_{r=1}^{M} \frac{a_r}{b_r^{k+1}} \tag{3}$$

where a_r and b_r are residues and poles of this LTI system, respectively. Therefore, the transfer function $H(s)$ of this

LTI system can be represented as an M-order rational function (pole-residue format) as:

$$H(s) = \sum_{r=1}^{M} \frac{a_r}{s - b_r} \tag{4}$$

while its impulse response $h(t)$ can be expressed as:

$$h(t) = \begin{cases} \sum_{r=1}^{M} a_r e^{b_r \cdot t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{5}$$

2.4 AWE Based Moment Matching

An interesting observation can be found by comparing the probabilistic moments in (1) with the time moments in (2): m_f^k is different from m_t^k due to a scaling factor $(-1)^k/k!$. It is easy to represent probabilistic moments as "time moments" by multiplying by a scaling factor:

$$\hat{m}_f^k = \frac{(-1)^k}{k!} \cdot m_f^k = \frac{(-1)^k}{k!} \cdot \int_{-\infty}^{+\infty} f^k \cdot pdf(f)dx \tag{6}$$

AWE based methods such as [1, 3] treat the variable f in (1) as the time t in (2). Then, the $pdf(f)$ can be optimally approximated by impulse response $h(t)$, which can be defined in pole/residue representation (5). The time moments can also be expressed by poles b and residues a and they can be solved by matching the first $2M$ moments:

$$\begin{cases} a_1 + a_2 + \cdots a_M = -\hat{m}_x^{-1} \\ \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots \frac{a_M}{b_M} = -\hat{m}_x^0 \\ \frac{a_1}{b_1^2} + \frac{a_2}{b_2^2} + \cdots \frac{a_M}{b_M^2} = -\hat{m}_x^1 \\ \vdots \\ \frac{a_1}{b_1^{2M-1}} + \frac{a_2}{b_2^{2M-1}} + \cdots \frac{a_M}{b_M^{2M-1}} = -\hat{m}_x^{2M-2} \end{cases} \tag{7}$$

where the poles, b_i $(i = 1, 2, ..., M)$, and residues, a_i $(i = 1, 2, ..., M)$, are the $2M$ unknowns in the above nonlinear system.

This nonlinear system can be solved with many techniques which have been thoroughly discussed in [9]. Here, we briefly review the analytic solution of (7).

First, all of the poles b_i $(i = 1, 2, ..., M)$ can be solved as the eigenvalues of the matrix

$$M = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{M-1} \end{bmatrix} \tag{8}$$

where c_i $(i = 0, 1, ..., M-1)$ are solved from the linear equations as:

$$\begin{bmatrix} \hat{m}_f^{-1} & \hat{m}_f^0 & \cdots & \hat{m}_f^{M-2} \\ \hat{m}_f^0 & \hat{m}_f^1 & \cdots & \hat{m}_f^{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{m}_f^{M-2} & \hat{m}_f^{M-1} & \cdots & \hat{m}_f^{2M-3} \end{bmatrix} \begin{bmatrix} -c_0 \\ -c_1 \\ \vdots \\ -c_{M-1} \end{bmatrix} = \begin{bmatrix} \hat{m}_f^{M-1} \\ \hat{m}_f^M \\ \vdots \\ \hat{m}_f^{2M-2} \end{bmatrix} \tag{9}$$

When the poles b_i are available, the residues a_i can be solved from (7) with simple arithmetic operations. Therefore, the impulse response $h(t)$ can be calculated, with the obtained a_i and b_i in equation (5), which can be used as an approximation of the distribution $pdf(f)$.

3. PROPOSED ALGORITHM

3.1 The Maximum Entropy Distribution

Entropy is a measure of uncertainty. When choosing a distribution, one should choose a distribution that maximizes the entropy [17]. By doing this, we can ensure that the distribution is uniquely determined to be maximally unbiased with regard to missing information, while still agreeing with what is known [17].

$$W = \int -p(x) \log p(x) dx \tag{10}$$

In (10), W is the entropy and $p(x)$ is the distribution of random variable x. To find the optimal distribution, we maximize this entropy subject to moment constraints. In this paper, we consider the probabilistic moments with moment order k

$$\int x^i p(x) dx = \mu_i, \qquad i = 0, 1, ..., k. \tag{11}$$

The analytical solution to this convex optimization problem involves the use of Lagrangian Multipliers as stated in [18] and is very complex. Although a more rigorous derivation of the solution can be found in [17, 18, 19], a brief overview is given below. We start by applying Lagrangian Multipliers to maximize (10), resulting in the Lagrangian function

$$L = -\int (p(x) \log p(x)) dx + \sum_{i=0}^{k} \lambda_i (\int x^i p(x) dx - \mu_i) \tag{12}$$

Now that the problem has been reformulated, we seek to maximize this function in order to obtain a solution. To find the maximum, we must look at the partial derivative of (12) with respect to $p(x)$ and λ_i

$$\frac{\delta L}{\delta \lambda_i} = 0 \tag{13}$$

$$\frac{\delta L}{\delta p(x)} = 0 \tag{14}$$

Solving equation (13) results in our original moment constraints given in (11). However, solving equation (14) results in the following

$$\frac{\delta L}{\delta p(x)} = -\int [\ln p(x) + 1] dx + \sum_{i=0}^{k} \lambda_i \int x^i dx \tag{15}$$

(15) can be solved by noticing that the integrand over an arbitrary domain must equal zero. With this in mind, we can reduce (15) to

$$-\ln p(x) - 1 + \sum_{i=0}^{k} \lambda_i x^i = 0 \tag{16}$$

If we rearrange the terms to solve for $p(x)$

$$p(x) = \exp\left(-\sum_{i=0}^{k} \lambda_i x^i\right) \tag{17}$$

However, the solution to the above problem does not exist for values of $k \geq 2$ [20]. To remedy this issue, [21] suggests

transforming this constrained problem into an unconstrained problem by utilizing duality. By using duality, we can recast our original problem of maximizing (12) into another problem that minimizes a new "dual" function. This dual function can be obtained by plugging the results of (17) into the Lagrangian function. Using this method results in the dual objective function below

$$\Gamma = \ln Z + \sum_{i=1}^{k} \lambda_i \mu_i \tag{18}$$

$$Z = \exp(\lambda_0) = \int \exp\left(-\sum_{i=1}^{k} \lambda_i x^i\right) dx \tag{19}$$

This problem can now be solved for any value of k. Most MAXENT approaches solve this problem using an iterative method such as Newton's method [20, 22]. Here, Newton's method is used to solve for the Lagrangian multipliers $\lambda = [\lambda_1, \lambda_2, ..., \lambda_k]'$ for moments $i, j = 1, 2, ..., k$ at iteration m

$$\lambda_{(m)} = \lambda_{(m)} - H^{-1} \frac{\delta \Gamma}{\delta \lambda} \tag{20}$$

Where the gradient (21) and Hessian (22) are defined as

$$\frac{\delta \Gamma}{\delta \lambda_i} = \mu_i - \frac{\int x^i \exp\left(-\sum_{i=1}^{k} \lambda_i \mu_i\right) dx}{\int \exp\left(-\sum_{i=1}^{k} \lambda_i \mu_i\right) dx} = \mu_i - \mu_i(\lambda) \tag{21}$$

$$H_{ij} = \frac{\delta^2 \Gamma}{\delta \lambda_i \delta \lambda_j} = \mu_{i+j}(\lambda) - \mu_i(\lambda)\mu_j(\lambda) \tag{22}$$

$$\mu_{i+j}(\lambda) = \frac{\int x^{i+j} \exp\left(-\sum_{i=1}^{k} \lambda_i \mu_i\right) dx}{\int \exp\left(-\sum_{i=1}^{k} \lambda_i \mu_i\right) dx} \tag{23}$$

Since the Hessian is positive definite, there exists a unique solution to the above problem [19]. Moreover, [19] also states that for a non-negative distribution $P(x)$ integrable in [0,1] with moments $\mu_0, \mu_1, ..., \mu_k$, if $P_N(x)$ is the MAXENT density, we have the following result:

$$\lim_{N \to \infty} \int_0^1 F(x) P_N(x) dx = \int_0^1 F(x) P(x) dx \tag{24}$$

This is known as the Maximum Entropy Principle (MEP) [19]. As [22] explains, MEP indicates that the MAXENT density can be used to approximate the distribution arbitrarily well if the sample size is large enough to allow calculation of enough moments.

3.2 Stability of MAXENT based Method

To show that the distribution in (17) is stable for all cases we are concerned with, it is sufficient to show that it is non-negative and absolutely continuous in the interval [0,1] of the random variable x. Although our random variable x may fall outside of this range, it is easy to normalize it such that it falls in this interval. Moreover, we are only concerned with the interval [0,1], and not any subintervals inside of it or outside of it. Showing that (17) is non-negative and absolutely continuous is rigorously explained in [19] and an overview of the fundamentals is as follows.

The dual problem shown in (18) is everywhere convex and has an absolute minimum [19]. First, the function is everywhere convex because the Hessian, the second derivative of (18), is positive definite. Second, the dual problem has an absolute minimum if the moments in (11) are monotonic, which holds true in the case of probabilistic moments. The proofs for both of these conditions can be found in [19].

It suffices to say that if the dual problem in (18) is everywhere convex and has an absolute minimum, then the distribution that minimizes it, in this case (17), is non-negative and absolutely continuous in the interval [19]. Since the distribution $p(x)$ is non-negative and absolutely continuous, it will not have a negative probability and it will not vanish over the interval. Therefore, $p(x)$ can be considered stable.

4. EXPERIMENT RESULTS

We have implemented the proposed algorithm in MATLAB. The first circuit is a 6-T SRAM bit-cell with 54 variables, while the second circuit is a Operational Amplifier with 70 variables. HSPICE is used to simulate these 2 circuits for circuit performance. Also, MC [6] and PEM [1] are used for comparison.

4.1 Experimental Setup

Below, we will first give a brief overview of the algorithms used in the experiments to demonstrate their distinguishing characteristics, then we will give an overview of the process variations of transistors and the circuits used. The algorithms we evaluated in the experiments are listed as follows:

- **MC (Monte Carlo)[6]**: Use quasi-random sampling to gather a huge data set of "MC samples". Calculate the performance distribution from these MC samples. Use a Figure of Merit to decide when we have enough MC samples for our ground truth. Figure of Merit: If the standard deviation of error between distribution n and distribution $n-1$ is *less* than 0.01, we determine the n^{th} distribution to be the ground truth.

- **PEM (Point Estimation Method)[1]**: Use quasi-random sampling to gather a small data set in order to calculate the probabilistic moments of the random variables. Convert these probabilistic moments to time moments of the corresponding LTI system. Use AWE to perform moment-matching in order to calculate the performance distribution.

- **MAXENT (Maximum Entropy)**: Use quasi-random sampling to gather a small data set in order to calculate the probabilistic moments of the random variables. Use the Maximum Entropy [17] formulation to perform moment-matching in order to calculate the performance distribution.

Additionally, for PEM and MAXENT, we utilize a scaling technique mentioned in [23, 24]. This scaling technique is performed on the moment matrix, such as (22) and the right hand side of (7). This scaling technique is utilized to minimize numerical errors that arise due to the limitations

of software. For all experiments using PEM and MAXENT, moment matrix scaling was performed. Since the MC method does not utilize a moment matrix, but instead does direct calculation of performance values, we do not use scaling for that algorithm.

Process Variations and Circuits The statistical data for the process variations are shown in Table 1. There are a total of 9 process variations in each transistor, meaning that there are 54 variables for the 6 transistor SRAM circuit, and 70 variables for the 10 transistor OpAmp. Similar to other methods [1, 3], we model the process variations as Gaussian distributions with various mean and sigma values.

Table 1: Parameters of MOSFETs

Variable Name	σ/μ	unit
Flat-band Voltage (V_{fb})	0.1	V
Gate Oxide Thickness (t_{ox})	0.05	m
Mobility (μ_0)	0.1	m^2/Vs
Doping concentration at depletion (N_{dep})	0.1	cm^{-3}
Channel-length offset (ΔL)	0.05	m
Channel-width offset (ΔW)	0.05	m
Source/drain sheet resistance (R_{sh})	0.1	Ohm/mm^2
Source-gate overlap unit capacitance (C_{gso})	0.1	F/m
Drain-gate overlap unit capacitance (C_{gdo})	0.1	F/m

6T SRAM bit-cell: Figure 3 depicts the 6T SRAM bit cell circuit overview. The reading operation of this cell is viewed as the circuit performance. The reading operation of the cell is determined by the voltage ΔV between BL and $\overline{BL}$. If this voltage is large enough to be sensed, it is deemed to be a successful read. The discharge behavior at $\overline{BL}$ plays a crucial role in the value of ΔV. Due to process variations in all transistors, the discharge behavior of $\overline{BL}$ may not be as predicted and therefore the voltage ΔV may not be large enough.

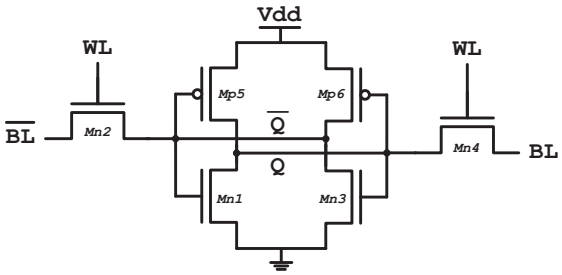

Figure 3: 6T SRAM Circuit Layout

Operational Amplifier: Figure 4 depicts the Operational Amplifier circuit overview. The bandwidth of this circuit is viewed as the circuit performance.

4.2 Stability

Figure 5 shows the performance distributions generated by MAXENT, PEM, and MC for the first 16 moments and first 18 moments using the 6T SRAM circuit using 200 samples. As we can see, MAXENT is stable under both conditions. The curves representing MAXENT for the first 16 moments and first 18 moments show very good overlap with the ground truth (MC) distribution. On the other hand, only the PEM curve corresponding to 16 moments is stable and overlaps with the ground truth distribution. The PEM

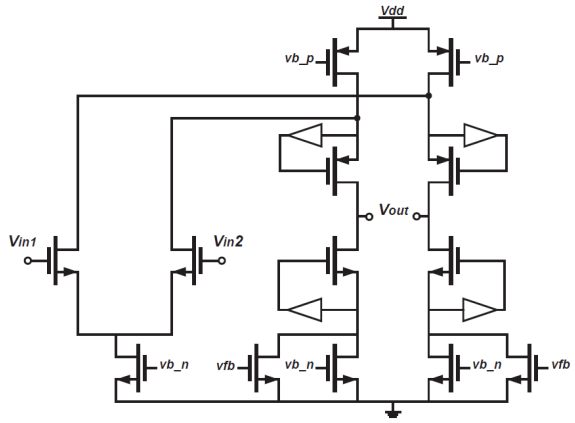

Figure 4: Operational Amplifier Circuit Layout

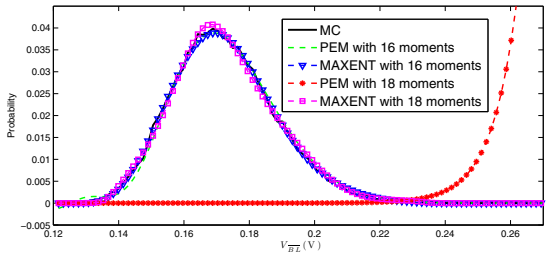

Figure 5: PEM lack of stability on SRAM circuit (200 samples)

curve corresponding to 18 moments is unstable, with a value of 0 through most of the distribution until it blows up to infinity. The only value that changed between these curves are the order of moments that were used. The sample number, circuit topology, process variations, and all other inputs were held constant. These results imply that PEM is very sensitive to the moments that are used.

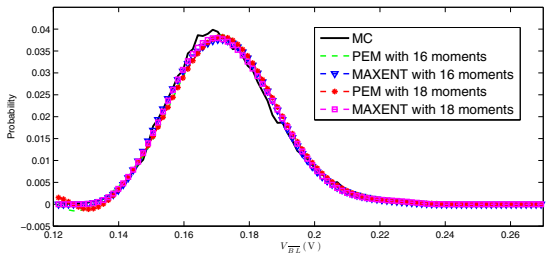

Figure 6: PEM stability on SRAM circuit (250 samples)

Figure 6 shows the performance distributions generated by MAXENT, PEM, and MC for the first 16 moments and first 18 moments using the 6T SRAM circuit using 250 samples. As we can see, MAXENT is stable under both 16 moments and 18 moments and overlap well with the ground truth distribution. Moreover, we see that PEM is now stable under both 16 moments and 18 moments and also overlap well with the ground truth distribution. Previously, PEM was

unstable for the SRAM circuit using 200 samples and 18 moments, whereas now it is stable for the SRAM circuit using 250 samples and 18 moments.

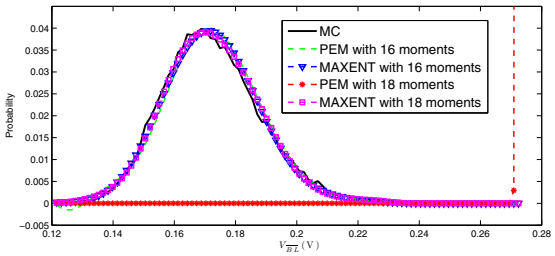

Figure 7: PEM lack of stability on SRAM circuit (300 samples)

Figure 7 shows the performance distributions generated by MAXENT, PEM, and MC for the first 16 moments and first 18 moments using the 6T SRAM circuit using 300 samples. In this case, we have returned to the instability of PEM. We see that MAXENT is still stable as always, but PEM is now unstable with 18 moments.

The above results are due to two reasons: the inaccuracies of the Pade approximation and the generation of moments from samples. PEM uses the Pade approximation to approximate the performance distribution as an impulse response function of an LTI system. In short, this impulse response function is a ratio of polynomial functions as shown in (5). The poles of this impulse response are estimated using the eigenvalues of a system matrix that is solved by AWE. When solving for the eigenvalues, AWE leverages the Pade approximation. However, the Pade approximation does not give all of the true eigenvalues of the system. Instead, it will generate poles (eigenvalues) that correspond to the dominant poles of the original system, and a few poles that do not correspond to the poles in the original system but account for the effects of the remaining poles [24]. Consequently, the Pade approximation may generate some positive eigenvalues. In the experiments above, the Pade approximation always generated 1 positive pole for the unstable cases, and 0 positive poles for the stable cases. Since the eigenvalues correspond to the poles of the impulse response function, they will take the form of a sum of weighted exponential functions. It is clear that since the Pade approximation may generate positive poles, they will correspond to unbounded exponential terms that continue to grow and lead to instability.

The number of samples also plays a key role in the stability of PEM. Since PEM uses AWE and the Pade approximation to solve the set of nonlinear functions in (7), changing the values on the RHS of (7) will change the values of b_r which are the poles of the transfer function. Using 200 samples will generate a set of moments M_1 while using 250 samples will generate a set of moments M_2. These moments will have completely different values and will lead to a new set of solutions to (7). This new set of solutions (the poles of the transfer function) may be stable or unstable.

We note that PEM, regardless of stability or instability, can run into the issue of negative probability as seen in both Figure 5 and Figure 6 where the probability dips below 0. The transfer function form of PEM allows for both positive and negative values, implying that the distribution may take on positive or negative values and this is nonsensical. On the other hand, MAXENT will never take a negative value as it is a product of exponentials, which can never take a negative value.

The key drawback of PEM is that it is unpredictable. The experimental results reinforce the idea that the instabilities in PEM are unpredictable and can occur with any number of samples that we use depending on the calculation performed by the Pade approximation. On the other hand, MAXENT is predictable. MAXENT does not use an LTI system model and does not use the Pade approximation, so it will not be subjected to this type of instability. More specifically, as was mentioned in the previous section, the distribution generated by MAXENT will always be non-negative and will always be absolutely continuous on the interval [0,1]. Clearly, the MAXENT distribution is more robust than the PEM distribution.

4.3 Accuracy

We also evaluated the accuracy of the MAXENT algorithm compared to PEM. Throughout our experiments, MAXENT consistently offers lower error relative to the ground truth than PEM does for any order of moments. We determine the error using the following equation:

$$error = \int (f_1(x) - f_2(x))\, dx \tag{25}$$

where $f_1(x)$ is our distribution from MAXENT or PEM and $f_2(x)$ is the ground truth distribution from MC. Figures (5, 6, 7) already illustrate the accuracy of MAXENT on the SRAM circuit for various samples and moment orders. Figure (8) illustrates the accuracy of MAXENT on the Operational Amplifier circuit. We see that at a moment order of 10, MAXENT already does a good job of mimicing the overall shape of the distribution, but it lacks some key details. Increasing the moment order to 12 gives an almost exact replica of the ground truth distribution. On the other hand, for a moment order of 10, PEM fails to give an accurate representation of the shape of the distribution. Moreover, increasing the moment order to 12 still yields a disappointing result. The overall shape and accuracy of the distribution from PEM is still very different from the ground truth distribution.

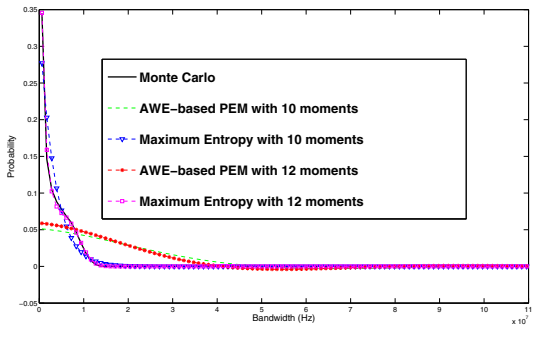

Figure 8: Operational Amplifier Accuracy (800 samples)

To quantify the results, Table 2 displays the relative error

for both MAXENT and PEM in the SRAM and Operational Amplifier circuits. We note that although the values of the variance and kurtosis (moment orders 2 and 4) themselves are accurate, the distributions generated using such few orders of moments is very inaccurate. This seems to be an issue of all moment matching algorithms. Consequently, results using such low order of moments is excluded.

Table 2: Accuracy Comparision

Circuit	# Samples	Moment Order	PEM Error(%)	MAXENT Error(%)
SRAM	200	6	46.349	11.85
		8	30.656	3.988
		10	15.577	3.281
		12	9.4457	3.394
		14	6.6038	3.181
		18	198.97	5.470
Op. Amp.	200	10	125.54	30.943
		12	116.39	30.881
		14	108.43	5.374
		16	102.05	5.506
		18	93.793	5.567
		20	111.49	5.584

The most noticeable trend is that MAXENT offers a lower relative error than PEM across all orders of moments and circuits. Both MAXENT and PEM utilized the same sampled values, and thus used the same moment values. The only difference is the way they performed their moment matching. In fact, we can see that MAXENT seems to perform its moment matching very effectively and efficiently. In the SRAM case, MAXENT seems to converge to a steady-state value of error by the time it hits 8 moments, whereas PEM continues to decrease in error (still always having a higher error than MAXENT) until it becomes unstable. In the OpAmp case, MAXENT seems to reach a steady-state value of error by the time it hits 14 moments, whereas PEM never seems to reach a steady-state error. Moreover, we see in Table 3 that changing the number of samples does not affect the result of MAXENT having a smaller relative error. The only significant difference is in the case of the OpAmp where MAXENT now achieves a steady-state value for error at only 12 moments. As we can see from both Table 2 and Table 3, once we reach a steady-state value, MAXENT offers up to 110% lower error for the OpAmp, and up to 27% lower error for the SRAM circuit.

Table 3: Accuracy Comparison

Circuit	# Samples	Moment Order	PEM Error(%)	MAXENT Error(%)
SRAM	300	6	46.117	11.043
		8	30.251	5.331
		10	15.097	6.046
		12	11.341	5.818
		14	10.74	6.516
		18	200	6.222
Op. Amp.	800	10	126.51	28.271
		12	117.26	3.851
		14	108.40	4.232
		16	101.110	3.679
		18	94.682	3.465
		20	89.264	3.568

4.4 Speedup

To observe the efficiency of MAXENT, we compare the speedup with respect to Monte Carlo while regarding the loss of accuracy. Table 4 shows the speedup in comparison to Monte Carlo with the corresponding loss in accuracy. As explained above, we use a Figure of Merit to decide when we have enough MC samples for our ground truth. Figure of Merit: If the standard deviation of error between distribution n and distribution $n-1$ is *less* than 0.01, we determine the n^{th} distribution to be the ground truth. For the 6T-SRAM circuit, we have a speedup of 195x while still maintaining a very small error of about 3%. For the OpAmp, we have a speedup of 592x while still maintaining an error of about 3.5%.

Table 4: Speedup

Circuit	Method	Samples	Speedup	Error %
SRAM	Monte Carlo	(39 x 10^3)	1x	0%
	MAXENT	200	195x	3.09%
OpAmp	Monte Carlo	(474 x 10^3)	1x	0%
	MAXENT	800	592x	3.46%

5. CONCLUSIONS

In this paper, we present a Maximum Entropy (MAXENT) based method to estimate the "arbitrary" behavioral distributions of AMS circuits. This approach approximates the exact behavioral distribution with a product of exponential distributions and finds the closest approximation by choosing the Lagrange Multipliers for these exponential distributions. To do so, the Shannon's information entropy between them has been maximized so as to reduce the distance between these two distributions. With the extensive experiments, the proposed approach has shown significant improvement in stability and accuracy (up to 110% lower error) when compared to AWE based methods and up to 592x speedup when compared with MC method.

6. ACKNOWLEDGMENTS

This work was partially supported by a UC Discovery grant sponsored by Cisco and ICscape and by a grant from Intel.

7. REFERENCES

[1] F. Gong, H. Yu, and L. He, "Stochastic analog circuit behavior modeling by point estimation method," *Proceedings of the 2011 international symposium on Physical design*, pp. 175–182, 2011.

[2] S. Nassif, "Modeling and analysis of manufacturing variations," *IEEE article on Custom Integrated Circuits*, pp. 223–228, 2001.

[3] X. Li, J. Le, P. Gopalakrishnan, and L. Pileggi, "Asymptotic probability extraction for non-normal distributions of circuit performance," *Proceedings of the 2004 IEEE/ACM International article on Computer-aided design*, pp. 2–9, 2004.

[4] F. Gong, S. Basir-Kazeruni, L. Dolecek, and L. He, "A fast estimation of sram failure rate using probability collectives," *ACM International Symposium on Physical Design*, pp. 41–47, 2012.

[5] F. Gong, S. Basir-Kazeruni, L. He, and Y. Hao, "Stochastic behavioral modeling and analysis for analog/mixed-signal circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 1, pp. 24–33, 2013.

[6] C. Jacoboni and P. Lugli, *The Monte Carlo method for semiconductor device simulation*. Springer, 2002, vol. 3.

[7] S. Vrudhula, J. Wang, and P. Ghanta, "Hermite polynomial based interconnect analysis in the presence of process variations," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 25, no. 10, pp. 2001–2011, 2006.

[8] E. Lehmann and G. Casella, *Theory of point estimation*. Springer, 1998, vol. 31.

[9] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 9, no. 4, pp. 352–366, 1990.

[10] K. Cao, S. Dobre, and J. Hu, "Standard cell characterization considering lithography induced variations," in *Proceedings of the 43rd annual Design Automation article*. ACM, 2006, pp. 801–804.

[11] H. Gundersen and E. Jensen, "The efficiency of systematic sampling in stereology and its prediction*," *Journal of Microscopy*, vol. 147, no. 3, pp. 229–263, 2011.

[12] I. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 7, no. 4, pp. 784–802, 1967.

[13] G. Casella and R. Berger, *Statistical inference*. Duxbury Press, 2001.

[14] A. V. Oppenheim, A. S. Willsky, and S. Hamid, *Signals and Systems*. Prentice Hall, 1996.

[15] W. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers," *Journal of applied physics*, vol. 19, no. 1, pp. 55–63, 1948.

[16] J. Vlach and K. Singhal, *Computer methods for circuit analysis and design*. Springer, 1983.

[17] E. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[18] J. Kapur and H. Kesavan, *Entropy optimization principles with applications*. Academic Pr, 1992.

[19] L. Mead and N. Papanicolaou, "Maximum entropy in the problem of moments," *Journal of Mathematical Physics*, vol. 25, p. 2404, 1984.

[20] X. Wu, "Calculation of maximum entropy densities with application to income distribution," *Journal of Econometrics*, vol. 115, no. 2, pp. 347–354, 2003.

[21] G. Judge and D. Miller, *Maximum entropy econometrics: Robust estimation with limited data*. John Wiley & Sons, 1997.

[22] B. Chen, J. Hu, and Y. Zhu, "Computing maximum entropy densities: A hybrid approach," *Signal Processing: An International Journal (SPIJ)*, vol. 4, no. 2, p. 114, 2010.

[23] E. Chiprout and M. Nakhla, *Asymptotic waveform evaluation and moment matching for interconnect analysis*. Kluwer Academic Publishers, 1994.

[24] P. Feldmann and R. Freund, "Efficient linear circuit analysis by padé approximation via the lanczos process," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 14, no. 5, pp. 639–649, 1995.