A New Sampling Technique for Monte Carlo-based Statistical Circuit Analysis

Hiwa Mahmoudi and Horst Zimmermann

Institute of Electrodynamics, Microwave and Circuit Engineering, Technische Universität Wien,
Gußhausstraße 25/E354, A-1040 Wien, Austria
Telephone: +43-1-58801-354621, Fax: +43-1-58801-9354600

***Abstract*—Variability is a fundamental issue which gets exponentially worse as CMOS technology shrinks. Therefore, characterization of statistical variations has become an important part of the design phase. Monte Carlo-based simulation method is a standard technique for statistical analysis and modeling of integrated circuits. However, crude Monte Carlo sampling based on pseudo-random selection of parameter variations suffers from low convergence rates and thus, providing high accuracy is computationally expensive. In this work, we present an extensive study on the performance of two widely used techniques, Latin Hypercube and Low Discrepancy sampling methods, and compare their speed-up and accuracy performance properties. It is shown that these methods can exhibit a better efficiency as compared to the pseudo-random sampling but only in limited applications. Therefore, we propose a new sampling scheme that exploits the benefits of both methods by combining them. Through a representative example, it is shown that the proposed sampling technique provides significant improvement in terms of computational efficiency and offers better properties as compared to each solely.**

I. Introduction

As integrated circuits scale to finer feature sizes the process variability gets increasingly important and difficult to capture as has been highlighted in several ITRS reports [1]. Therefore, it is crucial that the variability is addressed at the design stage and thus, advanced statistical methods are required to understand variation effects, ensure the manufacturability and improve the parametric yield [2], [3], [4].

A strong method to model and analyze variations of circuit properties is Monte Carlo-based modeling and simulation. In this context, Monte Carlo is a circuit simulation method which tries to act like nature by assuming random variations on process and device parameters and then performs SPICE (transistor-level) simulations to precisely track the propagation of the parameters' variation effects and to obtain the circuit behavior/performance of interest in presence of variations. Furthermore, this method is applicable to any circuit with any number of varying parameters exhibiting arbitrary statistical distributions. Therefore, it has become the standard technique for statistical analysis and modeling of integrated circuits.

Despite the advantages mentioned above, the convergence of Monte Carlo-based estimation error is rather slow and is proportional to $1/\sqrt{n}$, where n is the number of simulation runs. A strategy to speed up Monte Carlo simulation and still retain its advantages is to place control on how the samples are drawn from the parameter space by using effective sampling methods as compared to random selection of the parameter variations. Latin Hypercube and Low-Discrepancy (also termed as quasi-Monte Carlo) methods are two widely used sampling techniques to provide effective spreading properties [5], [6], [7], [8] and both techniques are now available in the Cadence Spectre simulation tool. The next two sections describe these techniques and present their superiorities and limitations. In Sec. IV, we propose a new sampling scheme that combines the two techniques optimally and exploits the benefits of both methods. Sec. V presents the simulation results and discussions and finally, we conclude in Sec. VI.

II. Latin Hypercube Sampling: Performance and Convergence Properties

A. One-Dimensional LH Sampling

The Latin Hypercube (LH) sampling method shows an excellent performance and projection properties in one dimension by providing an optimum spreading of the sampling points in the variation range. In order to generate a set of n LH sample points with uniform distribution ($\hat{x_{\mathrm{i}}}, i \in \{0,\ 1,\ \ldots,\ n-1\}$), a set of n uniformly distributed random sample points (x_{i}) is used as $\hat{x_{\mathrm{i}}} = i/n + x_{\mathrm{i}}/n$. Here, the term i/n is used to construct n equal non-overlapping bins and the term x/n is used to make a random selection of a point in each bin. This guarantees a good spreading of the points over the whole variation range on one hand, and still shows a random nature on the other hand. Therefore, it can be thought as a high quality random sampling method to generate well-spread samples in one dimension.

As the parameter variations in circuit simulation are usually normally distributed, we use an efficient method called inverse cumulative distribution function (CDF) transformation [9] to transform samples drawn from a uniform LH distribution into another distribution with any arbitrary probability density function. Therefore, in order to generate normally distributed LH samples, first, we generate uniformly distributed LH samples and then, exploit inverse-CDF method as $\hat{x_{\mathrm{i}}} = \sqrt{2} \times \mathrm{erf}^{-1}(2u_{\mathrm{i}} - 1)$, where erf^{-1} is the inverse error function and u_{i} denotes a set of uniformly distributed samples.

978-3-9815370-8-6/17/$31.00 ©2017 IEEE

B. Multi-Dimensional Sampling

Above we described the method to generate one-dimensional LH samples. However, in statistical circuit simulation and analysis we have multi-dimensional problems where each dimension corresponds to a varying parameter. The number of variable parameters (shown in this report by m) is usually in the range of $(10^3 - 10^5)$ as industrial circuits involve hundreds to tens of thousands devices (transistors) and each of them may have several varying parameters.

Fig. 1 shows the worst case and a random case of two-dimensional coverage. It is important to note that both $x1$ and $x2$ are one-dimensional LH and thus high quality sample sets. However, in the worst case, when simply the first, second, etc. points from both sets $x1$ and $x2$ are picked up to form the sample points in the multi-dimensional space, the coverage is very poor (Fig. 1a). Furthermore, the variables are heavily correlated and behave like one statistical variable in the analyses. That means, if in reality only one has strong effect on the output, the statistical result will be misleading as both variables may be recognized as important input parameters with high impact on the output performance. On the other hand, when the combination of the one-dimensional points is performed in a random manner (the $i-$th point from $x1$ corresponds to the $j-$th point form $x2$ where i and j are random numbers $\in \{1,\ 2,\ \ldots,\ n\}$), a two-dimensional coverage like Fig. 1b will be achieved. A multi-dimensional LH set constructed based on such random permutation method is the traditional design for LH samples. It should be noted that the traditional LH may show a better performance than a pure random in low-dimension problems. However, its superiority degrades with the dimension but is never worse than random sampling.

A possible improvement is to optimize the LH permutations to provide better multi-dimensional spreading properties. For example, in [10] and [11] search heuristics are used to find optimal design for multi-dimensional LH samples. However, these methods are computationally very expensive in high-dimensional problems as the amount of calculations increases exponentially with the number of dimensions (varying parameters). An alternative approach is to use a different class of sampling methods called low-discrepancy (LD) sequences which can provide better multi-dimensional spreading properties. As these methods generate deterministic sequences, simulations that are based on LD sequences are termed as quasi-Monte Carlo.

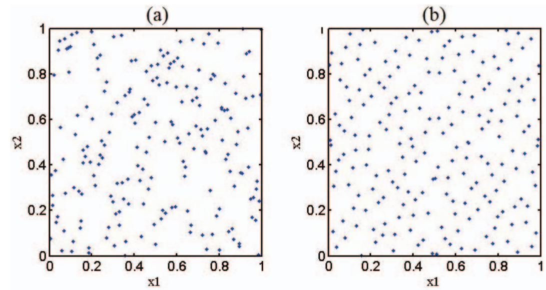

Fig. 2. A set of 200 two-dimensional sample points drawn from a uniform distribution using (a) traditional LH sampling method and (b) Sobol LD sequences generator.

III. Low-Discrepancy Sequences: Performance and Convergence Properties

Two well-known LD sample generators (provided also by Matlab) are Halton and Sobol generators. Due to higher efficiency, usually the Sobol generator is employed as the representative LD sample generator [7]. Fig. 2 shows the two-dimensional coverage of traditional LH sampling method and Sobol LD sequences generators. It is clear that the Sobol points provide a better multi-dimensional spreading property as compared to random permutation of one-dimensional LH sample points. As a result, LD sampling can often surpass both random and LH sampling in statistical circuit simulation and offers attractive runtime speedups [7]. However, despite the superiority of Sobol points in multi-dimensional spreading properties, one cannot guarantee that the Sobol points exhibit better spread of the samples in each single dimension.

Therefore, in order to compare the sampling qualities for different values of n and the corresponding convergence rates, Fig. 3 compares the mean and the standard deviation errors (obtained by Eq. 1) for one-dimensional random, LH, Sobol, and Halton samples as a function of the number of samples n. The result is quite interesting and shows that the Sobol and Halton generators offer a convergence rate of better than $1/\sqrt{n}$ but still not as good as LH offering the convergence rate of $1/n$. Here, we emphasize that the Latin Hypercube has the best performance for one-dimensional problems, or in other

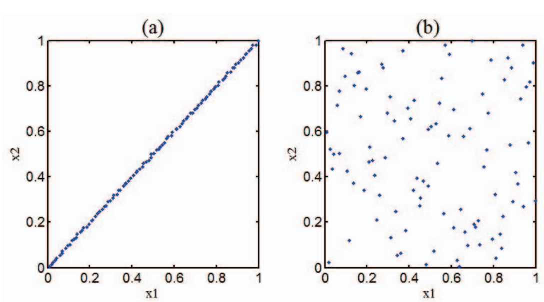

Fig. 1. A set of two-dimensional sample points drawn from a uniform distribution using LH sampling method where the sample points of $\boldsymbol{x1}$ and $\boldsymbol{x2}$ are arranged (a) one-by-one and (b) randomly.

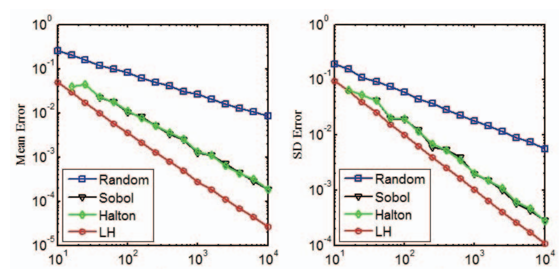

Fig. 3. Convergence of mean and standard deviation of samples drawn from a normal distribution using different sampling methods. The results are averaged over 500 sample sets.

words, over each single dimension in a multi-dimensional problem. Therefore, the best scenario would be to gain from both methods. Next section describes how a new sampling scheme can employ the benefits form both methods.

$$
\begin{aligned}
&Err(\text{mean}) = |\mu - \bar{x}|, \\
&Err(\text{SD}) = |\sigma - [1/n \sum_{i=1}^{n} (x_{\text{i}} - \bar{x})^2]^{1/2}|,
\end{aligned} \tag{1}
$$

where $\mu = 0$ and $\sigma = 1$ for a standard normal distribution and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_{\text{i}}$.

IV. A New LH-LD Sampling Scheme

Above we made an important conclusion that to provide a set of m dimensional samples (i.e. m variables each with given number of points n) every dimension should be sampled using LH method as this will guarantee the best one-dimensional spreading and projection properties. Then, as explained before, we need to find an efficient design to combine all one-dimensional LH sample sets to have an optimum $n \times m$ set. Therefore, our goal is to find a combinatorial design method which works better than the traditional LH sampling method based on random combinations of one-dimensional LH points. To employ the superiority of LD method in multi-dimensional coverage, instead of using an LD method directly to assign the sample points, we propose to assign the points by LH and then arrange them by LD methods as is explained in the following. As a result, we gain both one- and multi-dimensional advantages of LH and LD, respectively.

For example, X_{LH} represents a two-dimensional LH sample set with $n = 5$ points in each dimension drawn from a uniform distribution. This corresponds to 5 Monte Carlo runs where each column forms an input sample for the simulations.

$$
X_{\text{LH}} = \begin{bmatrix} 0.11 & 0.27 & 0.42 & 0.63 & 0.91 \\ 0.09 & 0.31 & 0.55 & 0.69 & 0.94 \end{bmatrix}
$$

We use a two-dimensional LD (Sobol) sample set of the same size as

$$
X_{\text{LD}} = \begin{bmatrix} 0.15 & 0.65 & 0.40 & 0.90 & 0.09 \\ 0.77 & 0.27 & 0.02 & 0.52 & 0.59 \end{bmatrix},
$$

and obtain the matrix $\mathrm{I(X_{LD})}$ which returns the (increasing) rank order of the elements of X_{LD} in each row as

$$
\mathrm{I(X_{LD})} = \begin{bmatrix} 2 & 4 & 3 & 5 & 1 \\ 5 & 2 & 1 & 3 & 4 \end{bmatrix}
$$

Now, an optimal LH-LD sample set can be obtained as

$$
X_{\text{LH}-\text{LD}} = \begin{bmatrix} 0.27 & 0.63 & 0.42 & 0.91 & 0.11 \\ 0.94 & 0.31 & 0.09 & 0.55 & 0.69 \end{bmatrix}
$$

where the rows of X_{LH} are sorted according to the permutations returned by $\mathrm{I(X_{LD})}$. As a result, the obtained sample set $X_{\text{LH}-\text{LD}}$ exhibits both one-dimensional and multi-dimensional spreading properties of the LH and LD sampling methods, respectively. In fact, this guarantees that the $X_{\text{LH}-\text{LD}}$ exploits the benefits of both methods simultaneously. Next section presents the corresponding simulation results.

V. Results and Discussions

Characterizing the circuit performance uncertainties (e. g. mean, standard deviation, and parametric yield estimation) in the presence of parameter variation and identifing the important variation sources that contribute to the performance variations (e. g. mismatch/variation contributions computation) are two main goals in statistical circuit analysis. In this section we evaluate the effect of the proposed LH-LD sampling method on the performance of the estimations providing these goals based on simulations in Matlab.

A. Uncertainty Estimation

We use a representative mathematical example to evaluate the uncertainty estimations as explained in the following. An output (performance) parameter is defined as $y = \sum a_{\text{j}} x_{\text{j}}$, where x_{j} ($j \in \{1, \ \ldots, \ m\}$) is a normally distributed variable ($\mu = 0$, $\sigma = 1$) and $a_{\text{j}} \in \mathbb{R}$ denotes its contribution coefficient to the output y. Here, y can be supposed as the output variation when m resistors of $a_{\text{j}}\Omega$ are connected in series and the output is defined as the total resistance and x_{j} express the random variation of the $j-$th resistor. It is clear that when the variations are all mutually independent, the true values of mean and the standard deviation of y are given as $\mu_{\text{y}} = 0$, $\sigma_{\text{y}} = \sqrt{\sum a_{\text{j}}^2}$, respectively. Now, we sample x_{j} at n points by using different sampling methods and estimate the statistical characteristics of y. Thus, the estimation errors are given by

$$
\begin{aligned}
&Err(\text{mean}) = |\bar{y} - \mu_{\text{y}}| = |\bar{y}| = |\frac{1}{n}\sum_{i=1}^{n} y_{\text{i}}|, \\
&Err(\text{SD}) = |(\sum_{j=1}^{m} a_{\text{j}}^2)^{1/2} - (\sum_{i=1}^{n} (y_{\text{i}} - \bar{y})^2/n)^{1/2}|.
\end{aligned} \tag{2}
$$

Here, $y_{\text{i}} = \sum_{j=1}^{m} a_{\text{j}} x_{\text{i,j}}$ and $x_{\text{i,j}}$ denotes the $i-$th sample point of the $j-$th variable x_{j}. Fig. 4 compares these errors for different sampling techniques. The results depicted by LH-LD is based on our proposed sampling technique explained before and demonstrate that the proposed method has both one-dimensional and multi-dimensional advantages of LH and LD, respectively and therefore, acts as excellent as LH for mean and at the same time keeps the good performance of LD in multi-dimensions ($m = 10$).

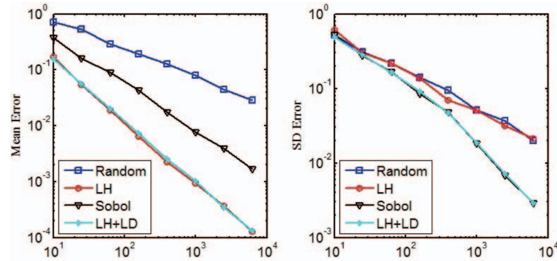

Fig. 4. Convergence of (a) mean and (b) standard deviation estimation errors based on different sampling methods.

Another important property of performance uncertainties of integrated circuits is yield. In our example the variations are normally distributed and thus we use the C_{PK} method based on the process capability index as a statistical measure to estimate yield $\mathrm{Y}_{\mathrm{CPK}} = 0.5 + 0.5\mathrm{erf}[(3/\sqrt{2})C_{\mathrm{PK}}]$, where $C_{\mathrm{PK}} = \min[(\mathrm{USL} - \mu)/3\sigma, (\mu - \mathrm{LSL})/3\sigma]$ [12]. Here, μ and σ are the estimated mean and standard deviation of the output, respectively. Now we define the yield error estimation as $Err(\mathrm{Yield}) = |\mathrm{Y}_{\mathrm{CPK}} - \mathrm{Y}|$, where Y denotes the true yield. Fig. 5a shows the convergence of yield estimation errors based on different sampling methods where the true yield is obtained by $C_{\mathrm{PK}} = 1$ for designs with a "3σ yield" target yield. It demonstrates that the proposed LH-LD sampling method outperforms other sampling methods for all n and provides more accurate yield estimations. Intuitively, as this method exploits the improvement from both LH and LD to provide better estimations for μ and σ (Fig. 4), it exhibits more accurate estimations of C_{PK} and thus, better estimations of $\mathrm{Y}_{\mathrm{CPK}}$.

B. Uncertainty Apportionment

In order to investigate the performance of different sampling techniques when Monte Carlo simulations are used to determine variation contributions, we use the same example explained above and we evaluate the accuracy in estimations. Such estimations are of special interest as in deep sub-micron CMOS technologies, mismatch variations play an important role in design effort and thus, it is critical that the relative contributions are successfully calculated. Therefore, we use a linear model to approximate the performance variations as a function of the device parameter variations given by $y = \sum_{j=1}^{m} b_{\mathrm{j}}\ x_{\mathrm{j}} + \varepsilon$, where b_{j} denotes the unknown model coefficients and ε represents the part of the output variation that cannot be explained by the weighted sum of the parameter variations. The coefficients b_{j} are estimated as $b_{\mathrm{j}} = \frac{1}{n}\sum_{i=1}^{n} x_{\mathrm{j}}^{(\mathrm{i})} y^{(\mathrm{i})}$ [3] where $\boldsymbol{x}_{\mathbf{j}}$ is a vector representing the sample points (variations) of the $j-$th dimension. Now the relative contribution coefficients (c_{k}) are calculated as

$$c_{\mathrm{k}} = \frac{b_{\mathrm{k}}^2}{\sum_{j=1}^{m} b_{\mathrm{j}}^2}. \tag{3}$$

To evaluate the convergence rate of the estimations we define the estimation error of c_{k} as

$$Err(\mathrm{CC}) = \frac{1}{m}\sum_{k=1}^{m} |c_{\mathrm{k}} - c_{\mathrm{k}}^{*}|\,, \tag{4}$$

where c_{k} and c_{k}^{*} denote the estimated and the true relative contribution coefficients, respectively. Fig. 5b shows the convergence of $Err(\mathrm{CC})$ for different sampling methods. As is expected, the LH-LD sampling method shows a better performance for all n and giving $2\times$ speedup for errors below 1%.

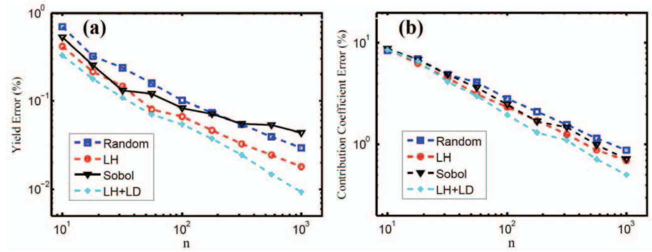

Fig. 5. (a) Yield and (b) contribution coefficient estimation errors as a function of the number of Monte Carlo runs based on different sampling methods.

VI. Conclusion

A new sampling scheme is proposed to simultaneously exploit the superiorities of both Latin Hypercube and Low Discrepancy sampling methods. It outperforms both methods and exhibits better convergence properties in Monte Carlo-based statistical analysis. Therefore, we supplement the obtained result in [7] and restate that quasi-Monte Carlo has better multi-dimensional coverage than random permutation, however Latin Hypercube is still the method of the choice to assign the sample points enabling the best spreading characteristics in each single dimension.

Acknowledgment

The authors would like to thank the Austrian BMVIT via FFG and the ENIAC Joint Undertaking for financial funding in the project eRAMP.

References

[1] International Technology Roadmap for Semiconductors, 2013, [Online] Available: http://www.itrs2.net/itrs-reports.html.
[2] S. R. Nassif, "Modeling and Analysis of Manufacturing Variations," *Proceedings of the IEEE Conference on Custom Integrated Circuits*, pp. 223–228, 2001.
[3] X. Li, and H. Liu, "Statistical Regression for Efficient High-Dimensional Modeling of Analog and Mixed-Signal Performance Variations," *Proceedings of the 45$^{\mathrm{th}}$ annual Design Automation Conference (ACM)*, pp. 38–43, 2008.
[4] X. Li, W. Zhang, and F. Wang, "Large-Scale Statistical Performance Modeling of Analog and Mixed-Signal Circuits," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 1–8, 2012.
[5] J. C. Helton, and F. J. Davis, "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering & System Safety*, vol. 81, pp. 23–69, 2003.
[6] P. Glasserman, *Monte Carlo Methods in Financial Engineering*. Springer Science & Business Media, 2003, vol. 53.
[7] A. Singhee, and R. A. Rutenbar, "Why Quasi-Monte Carlo is Better than Monte Carlo or Latin Hypercube Sampling for Statistical Circuit Analysis," *IEEE Trans. Computer-Aided Design*, vol. 29, pp. 1763–1776, 2010.
[8] F. Gong, S. Basir-Kazeruni, L. He, and H. Yu, "Stochastic Behavioral Modeling and Analysis for Analog/Mixed-Signal Circuits," *IEEE Trans. Computer-Aided Design*, vol. 32, pp. 24–33, 2013.
[9] W. T. Shaw, "Sampling Student's T Distribution-Use of the Inverse Cumulative Distribution Function," *Journal of Computational Finance*, vol. 9, pp. 37–73, 2006.
[10] J. Jaffari, and M. Anis, "On Efficient LHS-based Yield Analysis of Analog Circuits," *IEEE Trans. Computer-Aided Design*, vol. 30, pp. 159–163, 2011.
[11] H. Mahmoudi, and H. Zimmermann, "On Optimal Latin Hypercube Design for Yield Analysis of Analog Circuits," *Proceedings of the Austrian Workshop on Microelectronics (Austrochip)*, pp. 46–49, 2015.
[12] T. R. S. Weber and C. Duarte, "Yield Prediction With a New Generalized Process Capability Index Applicable to Non-Normal Data," *IEEE Trans. Computer-Aided Design*, vol. 35, pp. 931–942, 2016.