Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

An efficient prediction framework for multi-parametric yield analysis under parameter variations*

Xin LI[†‡1,2], Jin SUN[3], Fu XIAO[1]

([1]*Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing 210013, China*)

([2]*Technology Innovation Center, Jiangsu Academy of Safety Science and Technology, Nanjing 210042, China*)

([3]*School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*)

[†]E-mail: lin65002@hotmail.com

Received May 16, 2016; Revision accepted Sept. 11, 2016; Crosschecked Nov. 8, 2016

Abstract: Due to continuous process scaling, process, voltage, and temperature (PVT) parameter variations have become one of the most problematic issues in circuit design. The resulting correlations among performance metrics lead to a significant parametric yield loss. Previous algorithms on parametric yield prediction are limited to predicting a single-parametric yield or performing balanced optimization for several single-parametric yields. Consequently, these methods fail to predict the multi-parametric yield that optimizes multiple performance metrics simultaneously, which may result in significant accuracy loss. In this paper we suggest an efficient multi-parametric yield prediction framework, in which multiple performance metrics are considered as simultaneous constraint conditions for parametric yield prediction, to maintain the correlations among metrics. First, the framework models the performance metrics in terms of PVT parameter variations by using the adaptive elastic net (AEN) method. Then the parametric yield for a single performance metric can be predicted through the computation of the cumulative distribution function (CDF) based on the multiplication theorem and the Markov chain Monte Carlo (MCMC) method. Finally, a copula-based parametric yield prediction procedure has been developed to solve the multi-parametric yield prediction problem, and to generate an accurate yield estimate. Experimental results demonstrate that the proposed multi-parametric yield prediction framework is able to provide the designer with either an accurate value for parametric yield under specific performance limits, or a multi-parametric yield surface under all ranges of performance limits.

Key words: Yield prediction, Parameter variations, Multi-parametric yield, Performance modeling, Sparse representation
http://dx.doi.org/10.1631/FITEE.1601225 **CLC number:** TP312

1 Introduction

As the integrated circuit (IC) process scales, process, voltage, and temperature (PVT) variability becomes increasingly critical (Li, 2010; Trejo-Guerra *et al*., 2012; Nateghi and El-Sankary, 2015; Tlelo-Cuautle and Sanabria-Borbon, 2016). This increasing variability leads to widespread fluctuations in IC performance and makes it continually more challenging to create a reliable, robust design with high yield (Radfar and Singh, 2014; Banerjee and Chatterjee, 2015; Kondamadugula and Naidu, 2016). In addition, because of the fluctuations in PVT parameters, there exist correlations among performance metrics, such as gain, bandwidth, leakage current, offset, and gate delay. Improving one of these may deteriorate the others (Srivastava *et al*., 2008), and further cause serious parametric yield loss. For analog/mixed-signal circuits, parametric yield loss has become a significant or even dominant part of the total yield loss. Hence, within today's IC design phase,

‡ Corresponding author

* Project supported by the Natural Science Foundation of Jiangsu Province (Nos. BK20161072, BK20150785, and BK20130877) and the National Natural Science Foundation of China (Nos. 61502234 and 71301081)

ORCID: Xin LI, http://orcid.org/0000-0002-4859-2477

it is critical to consider multiple performance metrics simultaneously as part of the parametric yield prediction procedure under PVT parameters.

Most of the existing work in yield prediction has been limited to considering one single performance metric, and therefore has neglected the correlations among performance metrics under PVT parameter variations. For instance, Liu *et al*. (2013) presented a yield analysis of analog circuits considering process variations. This approach first estimates the frequency response bounds for a performance metric. Then, the analog circuit yield is analyzed according to the frequency-domain variation bound of the performance metric. Thus, this situation will bring significant parametric yield loss and great difficulty in yield optimization. Similar work was reflected in Haghdad and Anis (2012) and Bayrakci (2015). To maintain the correlations among performance metrics, there is an urgent need to incorporate multiple performance metrics simultaneously in the yield prediction procedure. This paper suggests a novel multi-parametric yield prediction framework under PVT parameter variations using several efficient techniques.

Recent research has been fully aware of the importance of the correlations among performance metrics. Accordingly, these studies have focused on considering simultaneously multiple performance metrics, especially some conflicting metrics, in parametric yield analysis. For instance, Haghdad and Anis (2012) presented a statistical analysis of the timing yield by considering both process and environmental variations. This algorithm first generates a statistical thermal profile and maps its associated voltage drops across the power grid. Then, the statistics of voltage drop and temperature fluctuations are used to predict the chip-level timing yield. Hwang *et al*. (2013) used yield slack to perform a statistical optimization. This method was verified to achieve a good leakage estimate with high computational efficiency. However, the approaches mentioned above fail to consider the close correlations among performance metrics (e.g., leakage current and gate delay). On the contrary, they predict/optimize one single parametric yield under other performance metric constraint(s), without predicting/optimizing the parametric yield considering multiple performance metrics simultaneously.

To address the above-mentioned issue, Tlelo-Cuautle and Sanabria-Borbon (2016) applied an evolutionary algorithm called a non-dominated sorting genetic algorithm (NSGA-II) in the optimization of operational transconductance amplifiers, which can guarantee the robustness of the feasible solutions to PVT variations among multiple performance metrics. However, this is a performance modeling method to describe the relationship between performance metrics and PVT variations. It does not describe the relationship between parameter yield and PVT variations by performance statistic information. Furthermore, Tang and Jha (2016) and Li *et al*. (2016) proposed two kinds of new methodologies that target multi-objective yield optimization. By applying adaptive multi-objective optimization, these methodologies can simultaneously optimize the timing and leakage power under PVT parameter variations. However, the proposed methodologies provide only trade-off information among multiple parametric yields. That is to say, they still predict the parametric yield for one single performance metric. In contrast to these multi-objective optimization methods, we propose an efficient multi-parametric yield prediction framework which can provide the designer with either an accurate value for parametric yield under specific performance limits, or a multi-parametric yield surface under all ranges of performance limits, instead of a set of feasible solutions among multiple parameter yields. Thus, our framework will have higher accuracy and practicability.

This paper aims to predict parametric yield considering multiple performance metrics in a simultaneous manner. The prediction framework regards multiple performance metrics as the constraint conditions and provides a multi-parametric yield estimate. In this work, we propose a multi-parametric yield prediction framework based upon the adaptive elastic net (AEN) method, multiplication theorem, Markov chain Monte Carlo (MCMC) method, and copula theory. The proposed framework maintains the correlations among performance metrics by modeling them in terms of PVT parameter variations sparsely with the AEN method. Then, based on the multiplication theorem and the MCMC method, a single-parametric yield prediction methodology has been developed to estimate the parametric yield for one single performance metric. Finally, a copula-based multi-parametric yield prediction procedure is suggested to predict the parametric yield considering multiple performance metrics simultaneously. As a result, designers can conveniently and efficiently obtain an accurate value for multi-parametric yield considering arbitrary performance metrics. The effectiveness of the proposed approach has been verified on various ISCAS benchmark circuits. To summarize, the multi-parametric yield prediction framework can provide the designer with an accurate value of parametric yields under specific performance limits, or a multi-parametric yield surface under all ranges of performance limits.

2 Sparse modeling for an arbitrary performance metric

The response surface modeling (RSM) (Xu *et al.*, 2015) method has been widely applied to estimate the performance metric of IC. However, the main disadvantage of the RSM method is the over-fitting problem that it may cause. Furthermore, most RSM methods apply least-squares fitting to obtain the parameter sensitivities of performance metrics from an over-determined linear equation, which leads to a very high computational cost. In this section, we introduce the AEN method (Zou and Zhang, 2009) to model the stochastic sparse models for the performance metrics of IC. The AEN method is an effective variable selection strategy that requires just a small set of sample points without over-fitting and possesses sparsity.

2.1 Stochastic model for performance metrics

Without loss of generality, in this study we use a common vector $\boldsymbol{p}=[p_1, p_2, \ldots, p_m]$ to represent the PVT parameters. The parameter variations $\Delta\boldsymbol{p}=\boldsymbol{p}-\boldsymbol{p}_0$ are usually considered as random variables, where $\boldsymbol{p}_0$ contains the mean values of $\boldsymbol{p}$. To analyze the variability of an arbitrary performance metric f, such as gain, bandwidth, leakage current, offset, and gate delay, the following model is applied to approximate f by a linear combination of n basis functions:

$$f(\Delta\boldsymbol{p}) = f_{\mathrm{L}}(\Delta\boldsymbol{p}) + \varepsilon = \sum_{i=1}^{n} \alpha_i B_i(\Delta\boldsymbol{p}) + \varepsilon, \tag{1}$$

where α_i (i=1, 2, …, n) denote the sensitivity coefficients of the performance metric with regard to the corresponding PVT parameters taken into account. $B_i(\Delta\boldsymbol{p})$ (i=1, 2, …, n) are regarded as the basis functions of interest. The notation ε describes the approximation error term.

In general, the unknown sensitivity coefficients in Eq. (1) can be obtained by solving the linear equations at k sample points:

$$\begin{cases} f^{(1)}(\Delta\boldsymbol{p}) \approx f_{\mathrm{L}}^{(1)}(\Delta\boldsymbol{p}) = \sum_{i=1}^{n} \alpha_i B_i(\Delta\boldsymbol{p}^{(1)}), \\ f^{(2)}(\Delta\boldsymbol{p}) \approx f_{\mathrm{L}}^{(2)}(\Delta\boldsymbol{p}) = \sum_{i=1}^{n} \alpha_i B_i(\Delta\boldsymbol{p}^{(2)}), \\ \quad\vdots \\ f^{(k)}(\Delta\boldsymbol{p}) \approx f_{\mathrm{L}}^{(k)}(\Delta\boldsymbol{p}) = \sum_{i=1}^{n} \alpha_i B_i(\Delta\boldsymbol{p}^{(k)}), \end{cases} \tag{2}$$

where $\Delta\boldsymbol{p}^{(j)}$ and $f^{(j)}(\Delta\boldsymbol{p})$ are the values of $\Delta\boldsymbol{p}$ and $f(\Delta\boldsymbol{p})$ evaluated at the jth sample point, respectively.

If we attempt to obtain the least squares (LS) solutions to Eq. (2), the size of samples k must be equal to or greater than the number of basis functions. When the number of basis functions is large, the general RSM methods become intractable due to the high computational cost. Furthermore, given the approximate Eq. (1), not all basis functions play an important role for a specific performance metric. In other words, a considerable number of these sensitivity coefficients are generally zero. Only a small set of critical basis functions that impact $f(\Delta\boldsymbol{p})$ significantly are necessary for approximating the performance metric. Thus, it is critical to identify these critical basis functions and further represent the performance metric sparsely to reduce the computational complexity.

2.2 AEN-based sparse representation modeling

AEN is a regularization method for dealing with the common linear problem (Zou and Zhang, 2009). This section details the AEN method for sparse representation of performance metrics, which will be incorporated into the multi-parametric yield prediction framework in subsequent sections.

Without loss of generality, substituting the sampling data into Eq. (1), it can be equivalently represented as a matrix form:

$$\boldsymbol{F} = \boldsymbol{B}\cdot\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \sum_{i=1}^{n} \alpha_i \cdot \boldsymbol{B}_i + \boldsymbol{\varepsilon} \approx \sum_{i=1}^{n} \alpha_i \cdot \boldsymbol{B}_i, \tag{3}$$

where

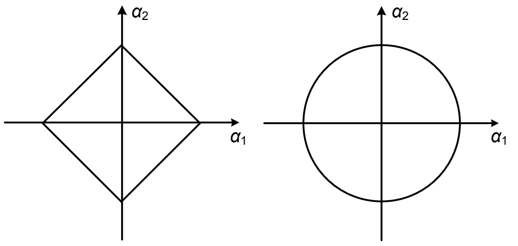
$$\begin{cases} \boldsymbol{F} = (f^{(1)}, f^{(2)}, \cdots, f^{(k)})^{\mathrm{T}}, \\ \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)^{\mathrm{T}}, \\ \boldsymbol{B} = (\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_n), \\ \boldsymbol{B}_i = (B_i(\Delta \boldsymbol{p}^{(1)}), B_i(\Delta \boldsymbol{p}^{(2)}), \cdots, B_i(\Delta \boldsymbol{p}^{(k)}))^{\mathrm{T}}, \\ \boldsymbol{\varepsilon} = (\varepsilon^{(1)}, \varepsilon^{(2)}, \cdots, \varepsilon^{(k)})^{\mathrm{T}}. \end{cases} \tag{4}$$

In Eq. (4), vector $\boldsymbol{F}$ denotes the k sample values of the performance metric. Vector $\boldsymbol{B}$ represents the sample matrix for the basis functions—the size of basis functions may be large, e.g., 10^4–10^6 (Li, 2010). Vector $\boldsymbol{B}_i \in \mathbb{R}^k$ contains the sample points for the ith basis function, which can be regarded as the basis vector associated with $B_i(\Delta \boldsymbol{p})$. The n-dimensional column vector $\boldsymbol{\alpha}$ describes the sensitivity coefficients of the corresponding basis functions. In addition, $\boldsymbol{\varepsilon}$ is the approximate error vector.

Now we introduce the L_1- and L_2-penalty functions, which will be considered as the sparsity constraints in yield estimation. The AEN criterion objective function of a performance metric can be formulated as follows:

$$\mathrm{AEN}(\lambda_1, \lambda_2, \boldsymbol{\alpha}) = \|\boldsymbol{F} - \boldsymbol{B} \cdot \boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_2^2 + \lambda_2 \|\boldsymbol{\omega}.*\boldsymbol{\alpha}\|_1, \tag{5}$$

where λ_1 and λ_2 denote the regularization coefficients of the AEN method. $\|*\|_2^2$ and $\|*\|_1$ stand for the L_2- and L_1-penalty functions, respectively. The L_2-penalty $\|\boldsymbol{\alpha}\|_2^2$ equals the summation of squares of the elements in vector $\boldsymbol{\alpha}$. The L_1-penalty $\|\boldsymbol{\omega} \cdot \boldsymbol{\alpha}\|_1$ equals the summation of the absolute values of the elements in vector $\boldsymbol{\omega}.*\boldsymbol{\alpha}=(\omega_1\alpha_1, \omega_2\alpha_2, \ldots, \omega_n\alpha_n)^{\mathrm{T}}$, where $\omega_i=(|\alpha_i(\mathrm{EN})|)^{-\gamma}$, $\alpha_i(\mathrm{EN})$ is the ith parameter estimation of the elastic net (EN) method, and γ is a positive constant. The penalty function contours of L_1 and L_2 are illustrated in Fig. 1.



(a) (b)

Fig. 1 The penalty function contours of L_1 (a) and L_2 (b)

To interpret the idea of L_1- and L_2-penalty functions, we formulate the following optimization problem to obtain an AEN estimate of $\boldsymbol{\alpha}$:

$$\begin{aligned} \boldsymbol{\alpha}' &= \arg\min_{\boldsymbol{\alpha}} \{\mathrm{AEN}(\lambda_1, \lambda_2, \boldsymbol{\alpha})\} \\ &= \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{F} - \boldsymbol{B} \cdot \boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_2^2 + \lambda_2 \|\boldsymbol{\omega}.*\boldsymbol{\alpha}\|_1 \right\} \\ &= \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{F} - \boldsymbol{B} \cdot \boldsymbol{\alpha}\|_2^2 + \lambda_1 \sum_{i=1}^{n} \alpha_i^2 + \lambda_2 \sum_{i=1}^{n} \omega_i |\alpha_i| \right\}, \end{aligned} \tag{6}$$

where $\boldsymbol{\alpha}'$ is the optimal estimate of the sensitivity coefficients. Some of them will be compressed to zeros by the optimization of the AEN method, therefore making the model in Eq. (1) sparse. The parameters λ_1 and λ_2 in Eq. (6) manipulate the amount of regularization applied to the estimation. If $\lambda_1=0$, it leads the AEN estimation back to the adaptive Lasso (least absolute shrinkage and selection operator) estimate.

Now, to deal with the optimization problem (6), we construct a new sampling data set using the original sampling data set ($\boldsymbol{B}$, $\boldsymbol{F}$) and the regularization coefficients λ_1, λ_2, which are represented as ($\boldsymbol{B}^*$, $\boldsymbol{F}^*$):

$$(\boldsymbol{B}^*_{(k+n)\times n}, \boldsymbol{F}^*_{k+n}) = \left((1+\lambda_1)^{-1/2} \begin{bmatrix} \boldsymbol{B} \\ \sqrt{\lambda_1} \boldsymbol{I} \end{bmatrix}, \begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{0} \end{bmatrix} \right). \tag{7}$$

Let $\eta=\lambda_2(1+\lambda_1)^{-1/2}$ and $\boldsymbol{\alpha}^*=(1+\lambda_1)^{1/2}\boldsymbol{\alpha}$. The optimal estimation of the sensitivity coefficients obtained by AEN is equivalent to the estimation of an adaptive Lasso method (Zou, 2006), which is defined as

$$\begin{cases} \boldsymbol{\alpha}'^{*} = \arg\min_{\boldsymbol{\alpha}} \{\mathrm{AL}(\eta, \boldsymbol{\alpha}^*)\} \\ \quad = \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{F}^* - \boldsymbol{B}^* \cdot \boldsymbol{\alpha}^*\|_2^2 + \eta \|\boldsymbol{\omega}.*\boldsymbol{\alpha}^*\|_1 \right\}, \\ \boldsymbol{\alpha}' = \boldsymbol{\alpha}'^{*} / \sqrt{1+\lambda_1}. \end{cases} \tag{8}$$

In practice, we can solve the adaptive Lasso problem by converting it into a traditional Lasso issue. Let $\boldsymbol{B}^{**}=\boldsymbol{B}^*./\boldsymbol{\omega}=(\boldsymbol{B}_1^*/\omega_1, \boldsymbol{B}_2^*/\omega_2, \ldots, \boldsymbol{B}_n^*/\omega_n)$ and $\boldsymbol{\alpha}^{**}=\boldsymbol{\omega}.*\boldsymbol{\alpha}^*$. Then Eq. (8) can be converted as follows:

$$\begin{cases} \boldsymbol{\alpha}'^{**} = \arg\min_{\boldsymbol{\alpha}} \{\mathrm{Lasso}(\eta, \boldsymbol{\alpha}^*)\} \\ \quad = \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{F}^* - \boldsymbol{B}^{**} \cdot \boldsymbol{\alpha}^{**}\|_2^2 + \eta \|\boldsymbol{\alpha}^{**}\|_1 \right\}, \\ \boldsymbol{\alpha}' = (1+\lambda_1)^{-1/2} (\boldsymbol{\alpha}'^{**} ./ \boldsymbol{\omega}). \end{cases} \tag{9}$$

It is clear that Eq. (9) is a traditional Lasso expression. The optimal estimation of Eq. (9) can be solved effectively by the least angle regression (Lars) (Zhang and Zamar, 2014). Once the optimal estimate is obtained, we can easily obtain the sparse estimate of the sensitivity coefficients $\boldsymbol{\alpha}'$. Note that to avoid the over- or under-fitting situation, a parameter selection criterion, e.g., the Akaike information criterion (AIC) (Panchal *et al*., 2010), Bayesian information criterion (BIC) (Lan *et al*., 2012), or cross-validation (Kaneda *et al*., 2015), is necessary for selecting the optimal path of Eq. (9).

According to the AEN method, the crucial basis functions are automatically selected by compressing the sensitivity coefficients of unimportant basis functions to zeros. Thus, the model of a performance metric can be represented sparsely, established by

$$f(\Delta\boldsymbol{p}) = f_{\text{sparse}}(\Delta\boldsymbol{p}) = \sum_{i=1}^{n} \alpha_i' B_i(\Delta\boldsymbol{p}), \tag{10}$$

where $f(\Delta\boldsymbol{p})$ is sparse, indicating that $\exists i\in[1, n]$, $\alpha_i'=0$.

By performing the algorithmic flow mentioned above, the model of the performance metric can be constructed with sparse coefficients and without over-fitting. The resulting performance metric model will later be employed to predict the parametric yield.

3 Parametric yield analysis for one single performance metric

In this section we introduce the parametric yield prediction method for one single performance metric under PVT parameter variations. The parametric yield for an arbitrary performance metric is assumed to be subject to a given metric limit, denoted by y_{nom}. As a result, the parametric yield information can be provided as

$$Y_{\text{single}} = F(y_{\text{nom}}) = \text{Prob}(f(\Delta\boldsymbol{p}) < y_{\text{nom}}), \tag{11}$$

where vector $\Delta\boldsymbol{p}$ denotes the PVT parameter variations, and $f(\Delta\boldsymbol{p})$ is the approximation of performance metric f obtained by the AEN method described previously.

It is obvious that the parametric yield can be determined by the cumulative probability function (CDF) of performance fluctuations propagated from parameter variations, which return the cumulative probability at a certain limit. Accordingly, the key issue here is to predict the CDF of a performance metric under parameter variations. The CDF provides the mapping relationship between the performance metric and parameter variations. In this work, the parametric yield model is further given by

$$Y_{\text{single}} = F(y_{\text{nom}}) = \text{Prob}(\chi_{\text{f}}), \tag{12}$$

where χ_{f} represents the parameter variation set, denoted by $\chi_{\text{f}}=\{\Delta\boldsymbol{p}: f(\Delta\boldsymbol{p})<y_{\text{nom}}\}$. According to Eq. (12), once $\text{Prob}(\chi_{\text{f}})$ is estimated, the parametric yield can be predicted. The prediction procedure takes advantage of the multiplication theorem and the MCMC method to estimate the CDF of the performance metric.

The MCMC method is a powerful technique for simulating arbitrary probability distributions (Yuan, 2009). In this study, we apply the MCMC method to estimate the CDF of performance distribution, denoted by $\text{Prob}(\chi_{\text{f}})$. Here we denote the probability distribution $\text{Prob}(\chi_{\text{f}})=\text{Prob}(f(\Delta\boldsymbol{p})<y_{\text{nom}})$ as the stationary distribution of a Markov chain. The state points can be regarded as the sample points in χ_f obtained by Markov chain simulation. Moreover, the approximate maximum likelihood point, denoted by $\Delta\boldsymbol{p}^*$, can be obtained from these state points easily (Yuan, 2009). To approximate the performance metric $f(\Delta\boldsymbol{p})$, a first-order Taylor series expansion is appropriate for expanding the sparse model in Eq. (10) at $\Delta\boldsymbol{p}^*$. The linear approximation of the performance metric under parameter variations can be established as follows:

$$\begin{aligned} L(\Delta\boldsymbol{p}) &= f(\Delta\boldsymbol{p}^*) + \sum_{x\in\Delta\boldsymbol{p}} \left.\frac{\partial f(\Delta\boldsymbol{p})}{\partial x}\right|_{\Delta\boldsymbol{p}^*} (x - x^*) \\ &= \left[f(\Delta\boldsymbol{p}^*) - \sum_{x\in\Delta\boldsymbol{p}} \left.\frac{\partial f(\Delta\boldsymbol{p})}{\partial x}\right|_{\Delta\boldsymbol{p}^*} x^* \right] + \sum_{x\in\Delta\boldsymbol{p}} \left.\frac{\partial f(\Delta\boldsymbol{p})}{\partial x}\right|_{\Delta\boldsymbol{p}^*} x. \end{aligned} \tag{13}$$

With the sparse model of the performance metric approximated linearly, the next step is to estimate the CDF $\text{Prob}(\chi_{\text{f}})$. To do that, a parameter variation set of the linear approximation χ_{l} is introduced, defined as $\chi_{\text{l}}=\{\Delta\boldsymbol{p}: L(\Delta\boldsymbol{p})<y_{\text{nom}}\}$. As indicated in the multiplication theorem, the probability relationship can be described as

$$\begin{aligned}\mathrm{Prob}(\chi_\mathrm{f}\chi_\mathrm{l}) &= \mathrm{Prob}(\chi_\mathrm{l}\mid\chi_\mathrm{f})\cdot\mathrm{Prob}(\chi_\mathrm{f})\\ &= \mathrm{Prob}(\chi_\mathrm{f}\mid\chi_\mathrm{l})\cdot\mathrm{Prob}(\chi_\mathrm{l}).\end{aligned} \tag{14}$$

It is obvious that the CDFs, $\mathrm{Prob}(\chi_\mathrm{f})$ and $\mathrm{Prob}(\chi_\mathrm{l})$, are proportional to each other when parameter variation sets χ_f and χ_l have intersection. As a result, if $\mathrm{Prob}(\chi_\mathrm{l})$ and the two conditional probabilities, $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$ and $\mathrm{Prob}(\chi_\mathrm{l}|\chi_\mathrm{f})$, are obtained, $\mathrm{Prob}(\chi_\mathrm{f})$ can be calculated. Following this flow, the saddle point estimation method (Jin, 2013) is further introduced to calculate $\mathrm{Prob}(\chi_\mathrm{l})$. Without loss of generality, for an arbitrary parameter variation Δp, the cumulant generating function (CGF) can be described as

$$K_{\Delta p}(t) = \log\{M_{\Delta p}(t)\} = \log\left\{\int_{-\infty}^{+\infty} \mathrm{e}^{t\Delta p}\mathrm{pdf}(\Delta p)d\Delta p\right\}, \tag{15}$$

where $M_{\Delta p}(t)$ denotes the moment generating function and $\mathrm{pdf}(\Delta p)$ is the probability density function (PDF) of Δp.

We now focus on estimating the CGF of a linear expression, denoted by $J(\Delta\boldsymbol{p})=L(\Delta\boldsymbol{p})-y_\mathrm{nom}$. The linear expression is in terms of parameter variations $\Delta\boldsymbol{p}$ and performance metric limit y_nom. According to Huang and Du (2008) and Yuan (2009), the CGF of linear expression $J(\Delta\boldsymbol{p})$ can be defined as

$$\begin{aligned}K_{J(\Delta\boldsymbol{p})}(t) &= K_{L(\Delta\boldsymbol{p})}(t) - K_{y_\mathrm{nom}}(t)\\ &= \left[f(\Delta\boldsymbol{p}^*) - \sum_{x\in\Delta\boldsymbol{p}} \left.\frac{\partial f(\Delta\boldsymbol{p})}{\partial x}\right|_{\Delta\boldsymbol{p}^*} x^*\right]t\\ &\quad + \sum_{x\in\Delta\boldsymbol{p}} K_x\left(\left.\frac{\partial f(\Delta\boldsymbol{p})}{\partial x}\right|_{\Delta\boldsymbol{p}^*} t\right) - y_\mathrm{nom}t.\end{aligned} \tag{16}$$

We then use the saddle point estimation method to estimate $\mathrm{Prob}(\chi_\mathrm{l})$ in Eq. (14). To be specific, given parameter variations $\Delta\boldsymbol{p}$, $\mathrm{Prob}(\chi_\mathrm{l})$ can be described as

$$\begin{aligned}\mathrm{Prob}(\chi_l) &= \mathrm{Prob}\{L(\Delta\boldsymbol{p}) \le y_\mathrm{nom}\}\\ &= \mathrm{Prob}\{J(\Delta\boldsymbol{p}) \le 0\}\\ &= \Phi(u) + \phi(u)\left(\frac{1}{u} - \frac{1}{v}\right),\end{aligned} \tag{17}$$

where $\Phi(u)$ and $\phi(u)$ denote the CDF and PDF of a standard normal distribution, respectively. The random variables u and v are obtained by the GGF of $J(\Delta\boldsymbol{p})$ described in Eq. (16). Specifically, u and v are given by $u=\mathrm{sign}(t_\mathrm{s})[-2K_{J(\boldsymbol{p})}(t_\mathrm{s})]^{1/2}$ and $v=t_\mathrm{s}[K''_{J(\boldsymbol{p})}(t_\mathrm{s})]^{1/2}$, where t_s denotes the saddle point, obtained by solving equation $K'_{J(\boldsymbol{p})}(t_0)=0$, and '′' and '″' denote the first- and second-order derivatives, respectively.

By now, we have estimated $\mathrm{Prob}(\chi_\mathrm{l})$, which is able to provide the cumulative probability of the linear approximation of performance under parameter variations $\Delta\boldsymbol{p}$. As long as we can calculate the values of the conditional probabilities, $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$ and $\mathrm{Prob}(\chi_\mathrm{l}|\chi_\mathrm{f})$ demonstrated in Eq. (14), $\mathrm{Prob}(\chi_\mathrm{f})$ can be estimated. In what follows, the conditional probability Markov chain method (Yuan *et al.*, 2010) is applied to estimate $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$ and $\mathrm{Prob}(\chi_\mathrm{l}|\chi_\mathrm{f})$.

The conditional probability $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$ is calculated by a digital simulation method. Specifically, regarding the probability distributions in χ_l as the stationary distribution of a Markov chain, the sample points in χ_l can be obtained by Markov chain simulation. The sample size is represented by n_l. By counting the number of the sample points falling into region χ_f, the value of the conditional probability $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$ can be calculated as

$$\mathrm{Prob}(\chi_\mathrm{f}\mid\chi_\mathrm{l}) = \frac{1}{n_\mathrm{l}}\sum_{i=1}^{n_l} I_\mathrm{f}(\Delta\boldsymbol{p}^{(i)}) = \frac{n_\mathrm{f|l}}{n_\mathrm{l}}, \tag{18}$$

where $I_\mathrm{f}(\Delta\boldsymbol{p}^{(i)})$ is the indicator function in region χ_f. When $\Delta\boldsymbol{p}^{(i)}\in\chi_\mathrm{f}$, $I_\mathrm{f}(\Delta\boldsymbol{p}^{(i)})=1$; otherwise, $I_\mathrm{f}(\Delta\boldsymbol{p}^{(i)})=0$. $n_\mathrm{l|f}$ denotes the size of the samples falling into region χ_f.

On the other hand, the conditional probability $\mathrm{Prob}(\chi_\mathrm{l}|\chi_\mathrm{f})$ can be calculated in a similar manner:

$$\mathrm{Prob}(\chi_\mathrm{l}\mid\chi_\mathrm{f}) = \frac{1}{n_\mathrm{f}}\sum_{i=1}^{n_\mathrm{f}} I_\mathrm{l}(\Delta\boldsymbol{p}^{(i)}) = \frac{n_\mathrm{l|f}}{n_\mathrm{f}}, \tag{19}$$

where $I_\mathrm{l}(\Delta\boldsymbol{p}^{(i)})$ denotes the indicator function in region χ_l, n_f is the sample size in region χ_f, and $n_\mathrm{l|f}$ is the size of the n_f samples falling into region χ_l.

Having calculated $\mathrm{Prob}(\chi_\mathrm{f}|\chi_\mathrm{l})$, $\mathrm{Prob}(\chi_\mathrm{l}|\chi_\mathrm{f})$, and $\mathrm{Prob}(\chi_\mathrm{l})$, we can estimate $\mathrm{Prob}(\chi_\mathrm{f})$ as follows:

$$\begin{aligned}\mathrm{Prob}(\chi_\mathrm{f}) &= \frac{\mathrm{Prob}(\chi_\mathrm{f}\mid\chi_\mathrm{l})}{\mathrm{Prob}(\chi_\mathrm{l}\mid\chi_\mathrm{f})}\cdot\mathrm{Prob}(\chi_\mathrm{l})\\ &= \frac{n_\mathrm{f|l}}{n_\mathrm{l}}\frac{n_\mathrm{f}}{n_\mathrm{l|f}}\mathrm{Prob}(\chi_\mathrm{l}).\end{aligned} \tag{20}$$

The algorithmic flow for estimating the cumulative distribution probability is summarized in Algorithm 1.

Algorithm 1 Cumulative distribution probability estimation

```
Input: fun, y_nom.
/* Sparse model for a performance metric under PVT parameter variations, and performance metric limit */
Output: prob.
  Δp*←Likelihood_Point(fun)
  linear_Fun←Linear(fun, Δp*)
  J_Stack←Decompose(linear_Fun, y_nom)
  CGF_Stack=∅
  while J_Stack≠∅ do
    F←POP(J_Stack)
    CGF_F=CGF_Computation(F)
    PUSH(CGF_F, CGF_Stack)
  end while
  J_CGF=Combine(CGF_Stack)
  CDF_Linear=CDF_Computation(J_CGF)
  (num_Fun, sample_Fun, num_Linear, sample_Linear)=
    Simu_Markov(fun, linear_Fun, y_nom)
  (num_F_L, num_L_F)=Count(fun, linear_Fun, y_nom,
    sample_Fun, sample_Linear)
  prob=Prob_Computation(CDF_Linear, num_Fun,
    num_Linear, num_F_L, num_L_F)
  return prob
```

Having the sparse model for a performance metric under parameter variations and the performance limit y_{nom}, the Likelihood_Point subroutine constructs the sampling space by Markov chain simulation and returns the approximate maximum likelihood point $\Delta\boldsymbol{p}^*$. The Linear subroutine uses a first-order Taylor series expansion to expand the sparse model for the performance metric at $\Delta\boldsymbol{p}^*$. The Decompose and Combine subroutines decompose and compose the linear expression $J(\Delta\boldsymbol{p})$, respectively. The CGF_Computation subroutine computes the CGF of the components of $J(\Delta\boldsymbol{p})$ according to Huang and Du (2008) and Yuan (2009). The CDF_Computation subroutine can compute Prob(χ_{l}) by a saddle point estimation method. The simu_Markov subroutine returns the sample points and sample size in regions χ_{f} and χ_{l} by Markov chain simulation sampling. The Count subroutine is responsible for counting the numbers of sample points falling into χ_{f} and χ_{l}. The Prob_Computation returns the exact value of Prob(χ_{f}). By performing the algorithmic flow presented in Algorithm 1, Prob(χ_{f}) can be estimated efficiently, which will subsequently be applied to predict the parametric yield for one single performance metric under parameter variations according to Eq. (12).

4 Copula-based multi-parametric yield prediction procedure

In Section 3, we concluded that the parametric yield of one single performance metric can actually be represented by the CDF of the performance metric. However, in practice, parametric yield prediction often considers multiple performance metrics simultaneously, e.g., gain, bandwidth, leakage current, offset, and gate delay. Therefore, it is necessary to predict the multi-parametric yield under parameter variations, which can be modeled as the joint CDF of the multiple performance metrics. Likewise, the multi-parametric yield model can be constructed as

$$Y_{\text{multi}} = F(\boldsymbol{y}_{1,\text{nom}}, \boldsymbol{y}_{2,\text{nom}}, \cdots, \boldsymbol{y}_{d,\text{nom}}) = \text{Prob}(f_1(\Delta\boldsymbol{p}) < \boldsymbol{y}_{1,\text{nom}}, f_2(\Delta\boldsymbol{p}) < \boldsymbol{y}_{2,\text{nom}}, \cdots, f_d(\Delta\boldsymbol{p}) < \boldsymbol{y}_{d,\text{nom}}), \tag{21}$$

where vector $\Delta\boldsymbol{p}$ denotes PVT parameter variations, $f_i(\Delta\boldsymbol{p})$ (i=1, 2, …, d) are the sparse models of d performance metrics, and $\boldsymbol{y}_{i,\text{nom}}$ (i=1, 2, …, d) are the limits of the d performance metrics.

Similar to the flow in the previous section, to predict the multi-parametric yield, we focus on estimating the joint CDF of the performance metrics. In this work, copula theory is applied to predict the joint CDF considering multiple performance metrics.

4.1 Basic facts about copula theory

Copula theory is an emerging branch in the field of statistics (Binois *et al.*, 2015). It studies how to construct an appropriate multivariate joint CDF, which has correlations among designated single variable marginal CDFs and all random variables. As of now, copula theory is well known in mathematical statistics. However, to our knowledge, it has not yet been used for a multi-parameter yield prediction procedure. In this subsection, we first present several basic facts about copula theory, which are necessary to follow the analysis.

Definition 1 (Copula function) (Nelson, 2006) Copula is a function C defined on $[0, 1]^n$, which satisfies the following conditions: (1) $\forall \boldsymbol{u}\in[0, 1]^n$, part of $\boldsymbol{u}$ are zeros, let $C(\boldsymbol{u})=0$; (2) $C(\boldsymbol{u})$ is an n-dimensional increasing function; (3) $\forall \boldsymbol{u}\in[0, 1]^n$ and $i=1, 2, \ldots, n$, $u_i=1$, $i\neq k$, $C(\boldsymbol{u})=u_k$.

Proposition 1 (Sklar's theorem) (Nelson, 2006) For any n-dimensional distribution function F: $\mathbb{R}^n\rightarrow[0, 1]^n$ with marginals $F_1, F_2, \ldots, F_n$, there exists a copula C such that $\forall z\in \mathbb{R}^n$, $F(z)=C(F_1(z_1), F_2(z_2), \ldots, F_n(z_n))$. Moreover, if F_k are continuous, then C is uniquely given by $C(\boldsymbol{u})=F(F^{-1}(u_1), F^{-1}(u_2), \ldots, F^{-1}(u_n))$; otherwise, C is uniquely determined on the range $F_1\times\ldots\times$ range F_n.

Through Sklar's theorem, we are pleasantly surprised to find an efficient general technology for handling an arbitrary correlation structure. As long as the marginal CDFs F_i ($i=1, 2, \ldots, n$) are known together with the copula function describing the correlation, we can estimate the joint CDF. In this work, the marginal CDFs of performance metrics have been unambiguously estimated by the multiplication theorem and the MCMC method in the previous section. It is obvious that as long as we determine the form of the copula function, the multi-parameter yield (also known as the joint CDF of the performance metrics) can be predicted effectively, which will be discussed in the next subsection.

4.2 Copula-based joint CDF prediction procedure

This subsection details the joint CDF prediction procedure for multiple performance metrics based on copula theory. The focus here is on optimizing the parameter(s) of the selected copula function that effectively describe(s) the correlations among performance metrics. To predict the joint CDF, our method starts with a copula function selection performed on various copula functions obtained by different construction methods. At present, copula functions include mainly elliptic copula functions (e.g., Gaussian and t-copula), Archimedean copula functions, and Plackett copula functions. However, the elliptic copula has a quite complicated form (not even analytic) and is intractable in our multi-parameter yield prediction problem. Instead, we turn to a different, simpler class of copula functions, found in so-called Plackett and Archimedean copulas. Plackett copulas are functions based on a constant cross-product ratio (Kao and Govindaraju, 2008). The function forms of Plackett copulas were detailed in Kao and Govindaraju (2008), while Archimedean copulas can be defined in what follows.

Definition 2 (Archimedean copula) (Houda and Lisser, 2015) A copula C is called Archimedean if there exists a continuous strictly decreasing function φ: $[0, 1]\rightarrow[0, +\infty]$, called the generator of C, such that $\varphi(1)=0$ and $C(\boldsymbol{u})=\varphi^{-1}[\varphi(u_1)+\varphi(u_2)+\ldots+\varphi(u_k)]$. If $\lim_{t\rightarrow 0}\varphi(t)=+\infty$, then C is called a strict Archimedean copula and φ is called a strict generator.

Archimedean and Plackett copulas have many families for different types of functions. In this work, the main family types are listed in Table 1. This list includes well-known families of Archimedean copulas such as Gumbel, Clayton, and Frank, but also the Plackett family of copulas at the bivariate level. These copula families all have different properties and, due to their flexibility, serve as a good starting point for dependence modeling.

After determining the copulas to be used, the parameter estimation procedure is applied to explore the optimal parameter θ for each copula. To do so, a pseudo maximum likelihood estimator (PMLE) (Kim *et al.*, 2007) method is adopted here to estimate the parameter θ. First, we construct the marginal empirical CDFs of the performance metrics, based on the sampling data. The marginal empirical CDF of the ith performance metric can be defined as

$$F_i(y_{i,\mathrm{nom}})=\frac{1}{s}\sum_{j=1}^{s}I(f_i(\Delta\boldsymbol{p}^{(j)})<\boldsymbol{y}_{i,\mathrm{nom}})=\frac{s_i}{s}, \quad (22)$$

where $I(f_i(\Delta\boldsymbol{p}^{(j)})<\boldsymbol{y}_{i,\mathrm{nom}})$ is the indicator function of the ith performance metric. When $f_i(\Delta\boldsymbol{p}^{(j)})<\boldsymbol{y}_{i,\mathrm{nom}}$, $I(f_i(\Delta\boldsymbol{p}^{(j)})<\boldsymbol{y}_{i,\mathrm{nom}})=1$; otherwise, $I(f_i(\Delta\boldsymbol{p}^{(j)})<\boldsymbol{y}_{i,\mathrm{nom}})=0$. The notation s denotes the size of the sampling data for the ith performance metric. s_i denotes the size of samples satisfying $f_i(\Delta\boldsymbol{p})<\boldsymbol{y}_{i,\mathrm{nom}}$.

We randomly choose n performance limits for performance constraints, denoted by $\boldsymbol{y}_{i,\mathrm{nom}}=(y_{i,\mathrm{nom}}^{(1)}, y_{i,\mathrm{nom}}^{(2)}, \ldots, y_{i,\mathrm{nom}}^{(n)})^{\mathrm{T}}$ ($i=1, 2, \ldots, d$). Based on the marginal empirical CDFs determined by Eq. (22), the marginal empirical CDF matrix can be estimated easily, where it is described as $\boldsymbol{u}_m=\{\boldsymbol{u}_i=(u_i^{(1)}, u_i^{(2)}, \ldots, u_i^{(n)})^{\mathrm{T}}, i=1, 2, \ldots, d\}$. In this work, we apply matrix $\boldsymbol{u}_m$ and the copula demonstrated in Table 1 to calculate the maximum likelihood estimation (MLE) of parameter θ. Without loss of generality, the likelihood function of the copula in Table 1 can be represented as

Table 1 Main Archimedean and Plackett copula families

Family	Copula function	Parameter θ
Archimedean copulas		
Gumbel	$C=\exp\left(-\left(\sum_{i=1}^{n}(-\ln u_i)^{\theta}\right)^{1/\theta}\right)$	[1, +∞)
Clayton	$C=\left(1+\sum_{i=1}^{n}(u_i^{\theta}-1)\right)^{-1/\theta}$	(0, +∞)
Frank	$C=-\frac{1}{\theta}\ln\left(1+\frac{\prod_{i=1}^{n}(\mathrm{e}^{-\theta u_i}-1)}{(\mathrm{e}^{-\theta}-1)^{n-1}}\right)$	(−∞, +∞)\{0}
AMH	$C=\frac{1-\theta}{\prod_{i=1}^{n}\frac{1-\theta(1-u_i)}{u_i}-\theta}$	[−1, 1)
Joe	$C=1-\left[1-\prod_{i=1}^{n}(1-(1-u_i)^{\theta})\right]^{1/\theta}$	[1, +∞)
Plackett copulas (2-dimensional)		
FGM	$C=u_1u_2+\theta u_1u_2(1-u_1)(1-u_2)$	[−1, 1]
Plackett	$C=(\theta-1)^{-1}\left(w-\sqrt{w^2-v}\right)/2$ $w=1+(\theta-1)(u_1+u_2)$ $v=4u_1u_2\theta(\theta-1)$	(0, +∞)\{1}

$$L_c(\theta)=\prod_{i=1}^{n}c((u_1^{(i)},u_2^{(i)},\cdots,u_d^{(i)});\theta), \quad (23)$$

where $\boldsymbol{u}=(u_1, u_2, \ldots, u_d)$ denotes the PDF of the copula function, and $c(\boldsymbol{u};\theta)=\frac{\partial^d C(\boldsymbol{u};\theta)}{\partial u_1\partial u_2\ldots\partial u_d}$.

Correspondingly, the log-likelihood function of the copula in Table 1 can be obtained by

$$l_{\mathrm{c}}(\theta)=\ln L_{\mathrm{c}}(\theta)=\sum_{i=1}^{n}\ln c((u_1^{(i)},u_2^{(i)},\cdots,u_d^{(i)});\theta). \quad (24)$$

As a general principle of parameter point estimation, PMLE seeks to maximize the log-likelihood function. Denoting $\theta_{\mathrm{PMLE}}:=\arg\max_{\theta\in\Theta} l_c(\theta)$, the value of θ_{PMLE} is considered the optimal estimate of parameter θ, which can be obtained either numerically or analytically.

Having estimated the optimal parameters for the copulas listed in Table 1, we focus on selecting the best fitting copula to predict the joint CDF of the performance metrics. The optimal copula can be selected by the goodness of the fit evaluation method, such as AIC (Panchal *et al.*, 2010) or ordinary least square (OLS) (Wang and Hutson, 2015). By comparing the evaluation results, the copula with the best goodness of fit (e.g., the minimum value of AIC or OLS) can be considered as the most suitable copula. Then according to copula theory, the multi-parameter yield (joint CDF) of the performance metrics can be constructed by the copula function obtained in the previous step and the marginal CDFs of the performance metrics estimated in Section 3. The expression of the multi-parameter yield can be represented as

$$\begin{aligned}Y_{\mathrm{multi}}&=F(\boldsymbol{y}_{1,\mathrm{nom}},\boldsymbol{y}_{2,\mathrm{nom}},\cdots,\boldsymbol{y}_{d,\mathrm{nom}})\\&=C(F_1(\boldsymbol{y}_{1,\mathrm{nom}}),F_2(\boldsymbol{y}_{1,\mathrm{nom}}),\cdots,F_d(\boldsymbol{y}_{1,\mathrm{nom}});\theta_{\mathrm{PMLE}}).\end{aligned} \quad (25)$$

So far, we have predicted the multi-parametric yield under parameter variations following the procedures mentioned above. Fig. 2 shows a flowchart summarizing the elementary flows of the proposed yield prediction framework, which is built step by step from circuit to parametric yield.

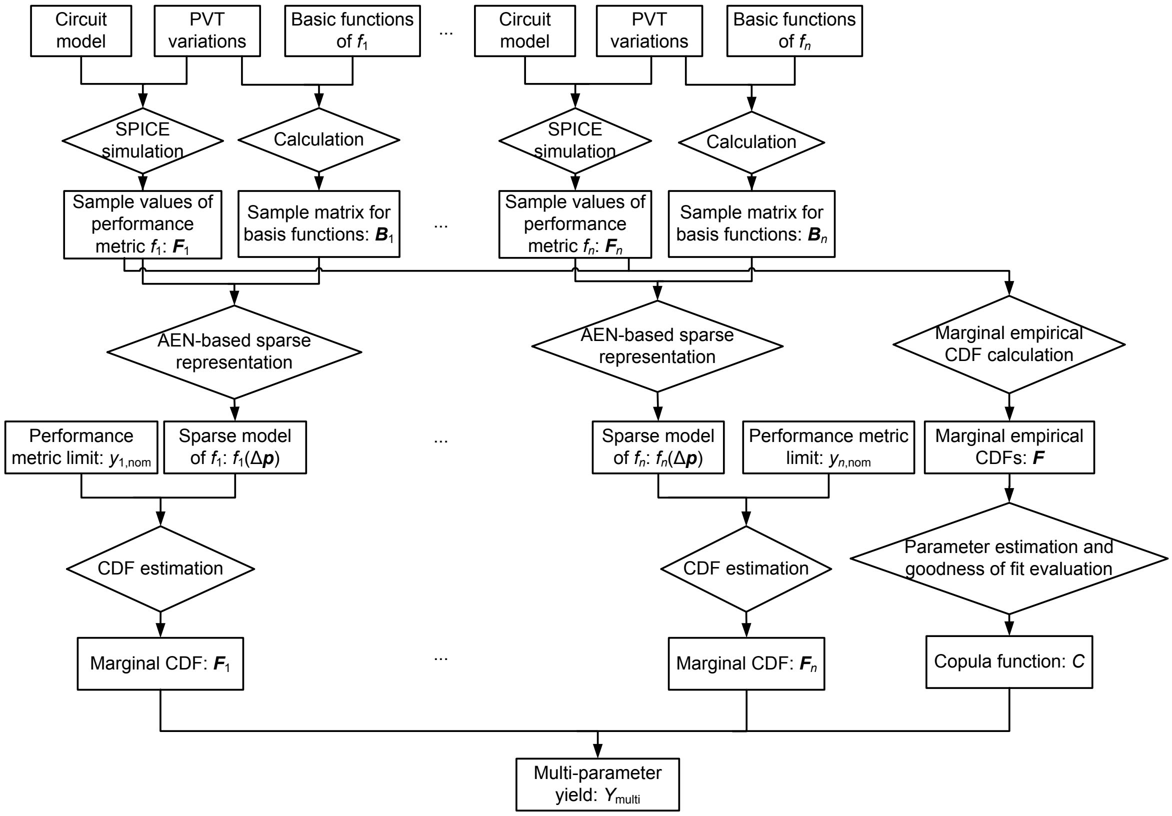


Fig. 2 Flowchart of our multi-parametric yield analysis framework

5 Experimental results

This section presents the experimental results of the proposed multi-parameter yield prediction framework implemented by MATLAB programming. The gate model is a linear model, and the technology node is 65 nm. The variations in the following key PVT parameters are taken into consideration: effective channel length variation (ΔL_{eff}), threshold voltage variation (ΔV_{th}), oxide thickness variation (ΔT_{ox}), power-supply voltage variation (ΔV_{dd}), and on-chip temperature variation (ΔT). Here, according to the empirical data in Visweswariah (2003), all parameter variations are modeled as truncated Gaussian distributions. The 3σ values of ΔL_{eff}, ΔV_{th}, and ΔT_{ox} are 20%, 10%, and 8% of their nominal values, respectively. The global and local variations of the process parameters are assumed to account for half of the total variation, respectively (Mande *et al.*, 2013). Besides, with regard to environmental parameters, the maximum deviations on voltage drop and on-chip temperature are 10% and 40% of their nominal value, respectively. In this study, the nominal values for power-supply voltage and on-chip temperature are set as 1.1 V and 25 °C, respectively.

The effectiveness of the proposed framework is demonstrated with ISCAS benchmark circuits. As mentioned above, the proposed framework is capable of representing performance metrics sparsely. Just for the convenience of comparative analysis, in this experiment, we aim to model two performance metrics, leakage current and gate delay, which both have explicit empirical expressions. It is worth emphasizing that the proposed prediction framework is not limited to modeling the performance metrics with empirical expressions. It is also suitable for modeling any other performance metrics such as gain, bandwidth, and offset.

According to Sun *et al.* (2008), the leakage current can be described as a sum of the subthreshold current and the gate leakage current. In addition, the subthreshold current and the gate leakage current can be approximated by a linear exponential function of PVT parameters. On the other hand, the gate delay can be approximated by a linear function. Without loss of generality, we randomly select the basis functions (log form) for subthreshold current as $\{\Delta L_{\text{eff}}^2, \Delta L_{\text{eff}}, \Delta V_{\text{th}}, \Delta V_{\text{th}}\Delta L_{\text{eff}}, \Delta V_{\text{dd}}, \Delta V_{\text{dd}}\Delta V_{\text{th}}, \Delta T, \Delta T\Delta L_{\text{eff}}\Delta V_{\text{th}}\}$ (the basis functions can include, as much as possible, the basis functions of interest). The corresponding coefficients are denoted by k_i (i=1, 2, …, 8). Similarly, the basis functions (log form) for gate leakage current are selected as $\{\Delta V_{\text{dd}}^3, \Delta V_{\text{dd}}^2, \Delta V_{\text{dd}}, \Delta T_{\text{ox}}^3, \Delta T_{\text{ox}}^2, \Delta T_{\text{ox}}, \Delta V_{\text{dd}}\Delta T_{\text{ox}}\}$, and the basis functions of gate delay are $\{\Delta L_{\text{eff}}^3, \Delta L_{\text{eff}}^2, \Delta L_{\text{eff}}, \Delta V_{\text{th}}^2, \Delta V_{\text{th}}, \Delta T_{\text{ox}}, \Delta T_{\text{ox}}\Delta L_{\text{eff}}, \Delta V_{\text{dd}}, \Delta V_{\text{dd}}\Delta T_{\text{ox}}, \Delta T\}$.

To verify the sparsity of the proposed performance metric modeling, we take the subthreshold current as an instance and apply circuit C432 to model it under different regularization coefficients. Gate leakage current, total leakage current, and gate delay are similar. In this step, different regularization coefficients λ_1 are used to select the sensitivity coefficients of basis functions. The AIC criterion is necessary here to optimize the model of subthreshold current. Fig. 3 demonstrates the coefficient selection path of basis functions for the subthreshold current under a few of different regularization coefficients. As illustrated in Fig. 3, the dashed part is the optimal sensitivity coefficients obtained based on the AIC criterion. Besides, it is clear that under a fixed regularization coefficient λ_1, different L_1-penalty constraints lead to a different compression of the coefficients of basis functions. When the value of AIC is very large, the selected coefficients will lead to the decrease of the accuracy of the model.

To further demonstrate the sparsity of our modeling method, the sensitivity coefficients of basis functions are optimized under a specific regularization coefficient, λ_1=1e-13. The coefficient selection results under different L_1-penalty constraints are listed in Table 2. It is easy to know that if the L_1-penalty constraint is very loose, the model is under-fitting. All the coefficients are compressed to zero. Otherwise, if the L_1-penalty constraint is very strict, all the coefficients are selected, and the model is not sparse. In addition, by comparing the values of AIC, Table 2 demonstrates the optimal sensitivity

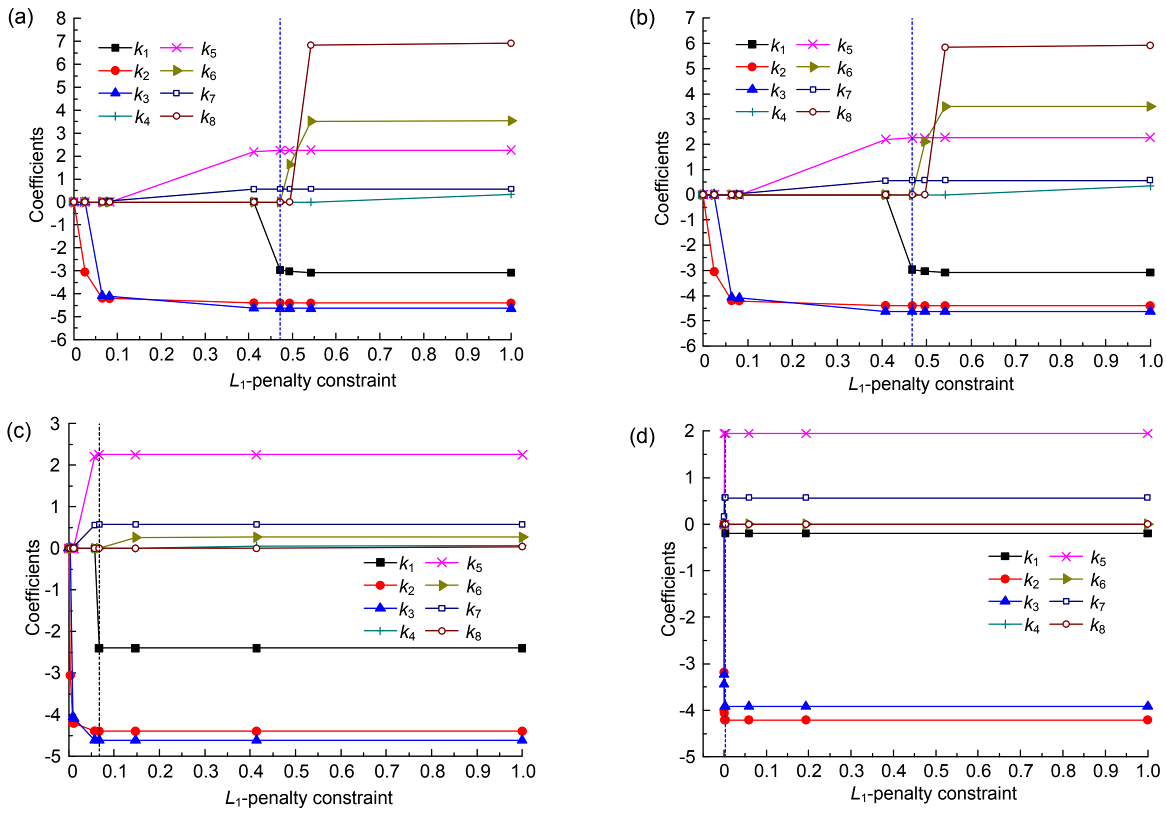


Fig. 3 The coefficient selection paths of the subthreshold current under different conditions: (a) λ_1=1e-13, min(AIC)=63.0989; (b) λ_1=1e-5, min(AIC)=63.0994; (c) λ_1=1e-2, min(AIC)=63.6353; (d) λ_1=0.5, min(AIC)=84.1606

Table 2 The coefficients obtained under different L_1-penalty constraints

L_1-penalty constraint	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	AIC
0	0	0	0	0	0	0	0	0	408.3657
0.0251	0	−3.0511	0	0	0	0	0	0	221.0552
0.0647	0	−4.1958	−4.0653	0	0	0	0	0	144.1919
0.0815	0	−4.2087	−4.1012	0	0	0	0.0334	0	136.7368
0.4119	0	−4.3990	−4.6220	0	2.1919	0	0.5534	0	63.4891
0.4718	**−2.9716**	**−4.4078**	**−4.6363**	**0**	**2.2575**	**0**	**0.5688**	**0**	**63.0989**
0.4937	−3.0216	−4.4079	−4.6368	0	2.2583	1.6367	0.5691	0	63.1038
0.5419	−3.0807	−4.4080	−4.6380	0	2.2594	3.5244	0.5695	6.8400	63.1104
1	−3.0817	−4.4080	−4.6384	0.3493	2.2594	3.5350	0.5695	6.9185	63.1226

Optimal coefficients (the minimum value of AIC) are in boldface

coefficients obtained based on the AIC criterion. From the optimal coefficients, some of them are compressed to zero. Thus, the model is sparse.

Having verified the sparsity of the performance metric modeling method, we can perform parametric yield analysis for one single performance metric. To verify the reliability of our method, we present the comparison results on a C432 circuit between the CDFs of subthreshold current obtained by our method and a Monte Carlo (MC) analysis (Fig. 4a). Similar results are obtained for the gate leakage current, total leakage current, and gate delay (Figs. 4b–4d). Fig. 4 shows that the results of our method fit well with those of MC analysis. From this we know that our method can predict the parametric yield of one single performance metric effectively.

As mentioned above, in this experiment, an important reason for choosing leakage current and gate delay as example performance metrics is that they have explicit empirical expressions. By these empirical expressions, we can better verify the effectiveness of our method. Fig. 5 demonstrates the CDFs of subthreshold current, gate leakage current, total leakage current, and gate delay obtained by our prediction method and the explicit empirical expressions (EE). As illustrated in Fig. 5, the results obtained by our method and the explicit EE analysis are basically consistent. This further verifies the effectiveness of our method.

In addition, to further demonstrate the effectiveness of our single-parametric yield prediction method, the relative errors of the aforementioned four performance metrics among our method, EE analysis, and MC analysis are listed in Table 3. We have evaluated the relative errors at the 95th, 50th, and 20th percentiles, respectively. All performance metrics have been normalized. It is obvious that the relative errors are relatively small, around 2%, verifying that the results obtained by our method fit those of EE and MC analysis very well.

After verifying the effectiveness of our single-parametric yield prediction method, we present the final results in the proposed multi-parametric yield prediction framework. In this set of experiments, we still take the C432 circuit as an instance to select the optimal copula from those in Table 1. The optimal parameter estimations and the values of goodness of fit evaluations (AIC and OLS) are listed in Table 4. In addition, by comparing the values of AIC and OLS, Table 4 lists the optimal copula as well, which will be applied to predict the multi-parametric yield in subsequent experiments.

In the next step, we perform the multi-parametric yield prediction framework using the copula selected in the previous experiment. The total leakage current and gate delay limits are set as some random multiples of the nominal total leakage current and gate delay, respectively. In such a case, the multi-parametric yields of the total leakage current and gate delay obtained by our framework and MC analysis are listed in Table 5. Table 5 also presents the relative errors between our framework and MC analysis. As illustrated, the relative errors are acceptable and the absolute deviations are all less than 0.05. The results justify that our framework could effectively predict the multi-parametric yield of performance metrics.

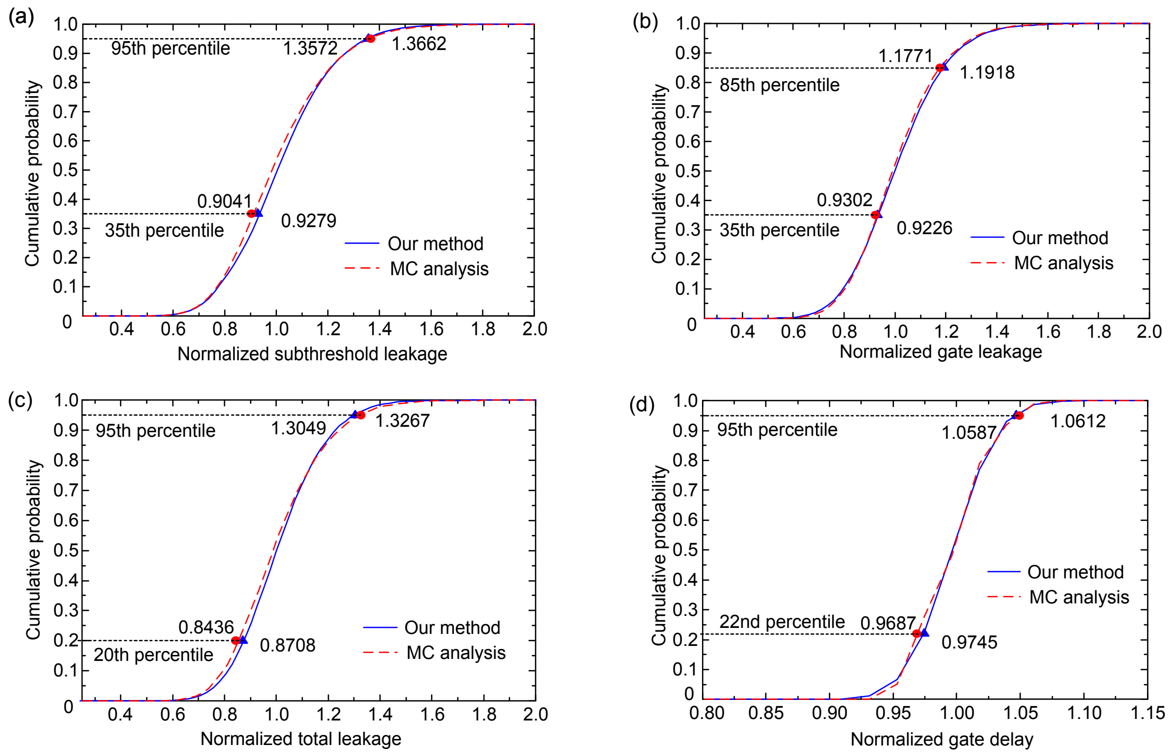


Fig. 4 The CDF obtained by our method and MC analysis: (a) subthreshold current; (b) gate leakage current; (c) total leakage current; (d) gate delay

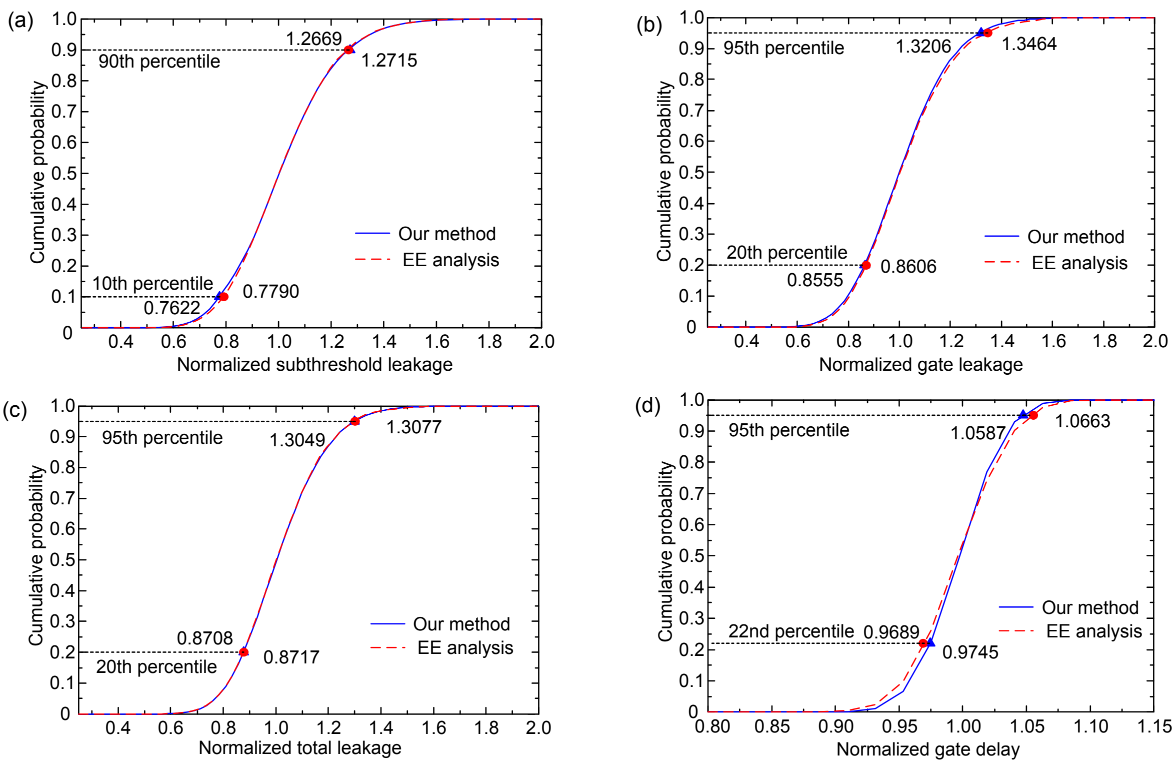


Fig. 5 The CDF obtained by our method and EE analysis: (a) subthreshold current; (b) gate leakage current; (c) total leakage current; (d) gate delay

We further chose the C432 circuit to provide the multi-parametric yield surface. Fig. 6 illustrates the multi-parametric yields generated by the proposed prediction framework for various limits of leakage current and gate delay. The multi-parametric yield surface obtained by MC analysis is provided in Fig. 7 for comparison. Each point in the surface represents a multi-parametric yield under specific leakage current and gate delay limits. All these points compose a smooth multi-parametric yield surface and provide the designers with useful and flexible yield information considering multiple performance metrics.

Table 3 Relative errors of different performance metrics

Performance metric	Method	95th percentile		50th percentile		20th percentile	
		Value	Error	Value	Error	Value	Error
Subthreshold current	Our method	1.3572		0.9993		0.8413	
	EE analysis	1.3548	0.18%	0.9985	0.08%	0.8496	0.99%
	MC analysis	1.3662	0.66%	0.9779	2.19%	0.8258	1.55%
Gate leakage current	Our method	1.3206		0.9964		0.8555	
	EE analysis	1.3464	1.92%	0.9998	0.34%	0.8606	0.59%
	MC analysis	1.3286	0.60%	0.9851	115%	0.8538	0.20%
Total leakage current	Our method	1.3049		1.0013		0.8708	
	EE analysis	1.3077	0.21%	1.0009	0.04%	0.8717	0.10%
	MC analysis	1.3267	1.64%	0.9821	1.95%	0.8436	3.21%
Gate delay	Our method	1.0587		1.0005		0.9713	
	EE analysis	1.0663	0.71%	0.9989	0.16%	0.9658	0.57%
	MC analysis	1.0612	0.25%	1.0016	0.11%	0.9664	0.51%

Table 4 The optimal copula selection from different copula families

Copula family	Parameter θ	AIC	OLS
Gumbel	1.0443	−2.7226e4	0.0109
Clayton	0.0124	−2.8359e4	0.0089
Frank	0.2052	−2.7854e4	0.0096
AMH	0.0258	−2.8381e4	0.0088
Joe	**1**	**−2.8617e4**	**0.0087**
FGM	0.0644	−2.8115e4	0.0092
Plackett	1.1603	−2.7361e4	0.0104

Optimal copula is in boldface

Table 5 The multi-parametric yields and relative errors under the selected copula

No.	Total leakage current	Gate delay	Yield		Relative error
			Our method	MC analysis	
1	0.943	1.025	0.2753	0.3020	8.84%
2	1.057	1.031	0.5081	0.5460	6.94%
3	1.143	1.055	0.7443	0.7775	4.27%
4	1.257	1.062	0.8804	0.9020	2.39%
5	1.323	1.069	0.9334	0.9430	1.02%
6	1.446	1.075	0.9773	0.9770	0.03%
7	1.625	1.092	0.9935	0.9950	0.15%
8	1.713	1.112	0.9987	0.9995	0.08%

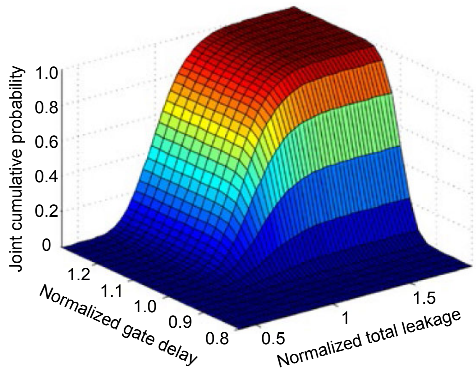


Fig. 6 The multi-parametric yield surface for the C432 circuit obtained by our framework

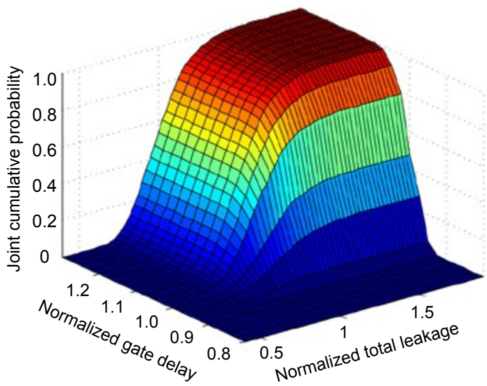


Fig. 7 The multi-parametric yield surface for the C432 circuit obtained by MC analysis

Table 6 A few multi-parametric yields obtained for various benchmark circuits

Circuit	Copula	θ	Evaluation		Example result 1			Example result 2			Running time (s)
			AIC ($\times 10^4$)	OLS	Yield		Relative error	Yield		Relative error	
					Our method	MC analysis		Our method	MC analysis		
C432	Joe	1	−2.861	0.0087	0.6026	0.6435	6.36%	0.9802	0.9815	0.13%	13.44
C499	AMH	0.0248	−2.776	0.0098	0.6173	0.6565	6.35%	0.9843	0.9820	0.23%	29.01
C880	AMH	0.0260	−2.845	0.0087	0.5911	0.6545	9.68%	0.9737	0.9700	0.37%	29.22
C1355	FGM	0.0661	−2.751	0.0102	0.6278	0.6350	1.13%	0.9857	0.9835	0.22%	20.94
C1980	AMH	0.0247	−2.779	0.0097	0.6200	0.6585	5.84%	0.9845	0.9820	0.25%	28.52
C2670	FGM	0.0667	−2.901	0.0079	0.6119	0.6555	6.65%	0.9823	0.9805	0.18%	21.34
C3540	Clayton	0.0129	−2.848	0.0086	0.6164	0.6550	5.89%	0.9812	0.9800	0.12%	14.31
C5315	FGM	0.0686	−2.857	0.0086	0.6083	0.6545	7.05%	0.9809	0.9800	0.09%	20.95
C7552	FGM	0.0683	−2.853	0.0086	0.6083	0.6541	7.00%	0.9811	0.9810	0.01%	20.75

At the last step, we perform the proposed multi-parametric yield prediction framework on various benchmark circuits. The optimal copula for each circuit, the corresponding parameter estimation, and the values for AIC and OLS are listed in Table 6. A few example yields selected from the multi-parametric yield surface randomly, and the corresponding yields obtained by MC analysis, are also listed. The example yields are described as $Y_{\text{multi},1}$=Prob($I_{\text{Leak}}(\Delta\boldsymbol{p})<1.12I_{\text{nom}}$, $D_{\text{Delay}}(\Delta\boldsymbol{p})<1.03D_{\text{nom}}$) and $Y_{\text{multi},2}$=Prob($I_{\text{Leak}}(\Delta\boldsymbol{p})<1.45I_{\text{nom}}$, $D_{\text{Delay}}(\Delta\boldsymbol{p})<1.08D_{\text{nom}}$). In addition, the relative errors between these two methods and the running time are calculated. According to the data in Table 6, different copulas will be selected for different benchmark circuits. However, our method can predict the multi-parametric yield effectively for all benchmark circuits. Moreover, the data in the table shows that the running time for our method does not increase with the increase of the scale of the benchmark circuits. It changes with the type of copula function, and the running times of the same copula function are close. Finally, it is worth emphasizing that the running time in Table 6 is the execution time of our prediction framework—it does not include the time for SPICE simulation. The time for SPICE simulation increases with the increase of the scale of the benchmark circuits.

6 Conclusions

We have proposed an effective multi-parametric yield prediction framework that integrates the AEN method, multiplication theorem, MCMC method, and copula theory. Considering multiple performance metrics simultaneously, the proposed strategy can be used to predict the multi-parameter yield by estimating the joint CDF of the performance fluctuations propagated from parameter variations. The efficiency of this new method has been verified with various ISCAS benchmark circuits.

References

Banerjee, A., Chatterjee, A., 2015. Signature driven hierarchical post-manufacture tuning of RF systems for performance and power. *IEEE Trans. VLSI Syst.*, **23**(2):342-355. http://dx.doi.org/10.1109/TVLSI.2014.2309114

Bayrakci, A.A., 2015. Stochastic logical effort as a variation aware delay model to estimate timing yield. *Integr. VLSI J.*, **48**(1):101-108. http://dx.doi.org/10.1016/j.vlsi.2014.07.003

Binois, M., Rullière, D., Roustant, O., 2015. On the estimation of Pareto fronts from the point of view of copula theory. *Inform. Sci.*, **324**:270-285. http://dx.doi.org/10.1016/j.ins.2015.06.037

Haghdad, K., Anis, M., 2012. Timing yield analysis considering process-induced temperature and supply voltage variations. *Microelectron. J.*, **43**(12):956-961. http://dx.doi.org/10.1016/j.mejo.2012.07.014

Houda, M., Lisser, A., 2015. Archimedean copulas in joint chance-constrained programming. *Commun. Comput. Inform. Sci.*, **509**:126-139. http://dx.doi.org/10.1007/978-3-319-17509-6_9

Huang, B., Du, X., 2008. Probabilistic uncertainty analysis by mean-value first order Saddlepoint approximation. *Reliab. Eng. Syst. Safety*, **93**(2):325-336. http://dx.doi.org/10.1016/j. ress.2006.10.021

Hwang, E.J., Kim, W., Kim, Y.H., 2013. Timing yield slack for timing yield-constrained optimization and its application to statistical leakage minimization. *IEEE Trans. VLSI Syst.*, **21**(10):1783-1796. http://dx.doi.org/10.1109/TVLSI.2012.2220792

Jin, Y.J., 2013. Reliability-based sensitivity analysis for machining precision by saddle-point approximation. *Appl. Mech. Mater.*, **241-244**:280-283.

http://dx.doi.org/10.4028/www.scientific.net/AMM.241-244.280

Kaneda, S., Mizumoto, T., Maeno, T., *et al*., 2015. A cross validation of network system models for delay tolerant networks. Int. Conf. on Mobile Computing and Ubiquitous Networking, p.185-190. http://dx.doi.org/10.1109/ICMU.2015.7061064

Kao, S.C., Govindaraju, R.S., 2008. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resources Res.*, **44**(2):333-341. http://dx.doi.org/10.1029/2007WR006261

Kim, G., Silvapulle, M.J., Silvapulle, P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, **51**(6):2836-2850. http://dx.doi.org/10.1016/j.csda.2006.10.009

Kondamadugula, S., Naidu, S.R., 2016. Parameter-importance based Monte-Carlo technique for variation-aware analog yield optimization. Proc. 26th edition on Great Lakes Symp. on VLSI, p.51-56. http://dx.doi.org/10.1145/2902961.2903018

Lan, W., Wang, H., Tsai, C., 2012. A Bayesian information criterion for portfolio selection. *Comput. Statist. Data Anal.*, **56**(1):88-99. http://dx.doi.org/10.1016/j.csda.2011.06.012

Li, X., 2010. Finding deterministic solution from underdetermined equation: large-scale performance variability modeling of analog/RF circuits. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.*, **29**(11):1661-1668. http://dx.doi.org/10.1109/TCAD.2010.2061292

Li, X., Sun, J., Xiao, F., *et al*., 2016. An efficient bi-objective optimization framework for statistical chip-level yield analysis under parameter variations. *Front. Inform. Technol. Electron. Eng.*, **17**(2):160-172. http://dx.doi.org/10.1631/FITEE.1500168

Liu, X., Tan, S.X.D., Palma-Rodriguez, A.A., *et al*., 2013. Performance bound analysis of analog circuits in frequency- and time-domain considering process variations. *ACM Trans. Des. Autom. Electron. Syst.*, **19**(1):6. http://dx.doi.org/10.1145/2534395

Mande, S.S., Chandorkar, A.N., Iwai, H., 2013. Computationally efficient methodology for statistical characterization and yield estimation due to inter- and intra-die process variations. Proc. 5th Asia Symp. on Quality Electronic Design, p.287-294. http://dx.doi.org/10.1109/ASQED.2013.6643602

Nateghi, H., El-Sankary, K., 2015. A self-healing technique using ZTC biasing for PVT variations compensation in 65nm CMOS technology. Canadian Conf. on Electrical and Computer Engineering, p.128-131. http://dx.doi.org/10.1109/CCECE.2015.7129173

Nelson, R.B., 2006. An Introduction to Copulas. Springer, New York. http://dx.doi.org/10.1007/0-387-28678-0

Panchal, G., Ganatra, A., Kosta, Y.P., *et al*., 2010. Searching most efficient neural network architecture using Akaike's information criterion (AIC). *Int. J. Comput. Appl.*, **1**(5): 41-44. http://dx.doi.org/10.5120/126-242

Radfar, M., Singh, J., 2014. A yield improvement technique in severe process, voltage, and temperature variations and extreme voltage scaling. *Microelectron. Reliab.*, **54**(12): 2813-2823. http://dx.doi.org/10.1016/j.microrel.2014.07.138

Srivastava, A., Chopra, K., Shah, S., *et al*., 2008. A novel approach to perform gate-level yield analysis and optimization considering correlated variations in power and performance. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.*, **27**(2):272-285. http://dx.doi.org/10.1109/TCAD.2007.907227

Sun, J., Li, J., Ma, D., *et al*., 2008. Chebyshev affine-arithmetic-based parametric yield prediction under limited descriptions of uncertainty. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.*, **27**(10):1852-1865. http://dx.doi.org/10.1109/TCAD.2008.2003300

Tang, A., Jha, N.K., 2016. GenFin: genetic algorithm-based multiobjective statistical logic circuit optimization using incremental statistical analysis. *IEEE Trans. VLSI Syst.*, **24**(3):1126-1139. http://dx.doi.org/10.1109/TVLSI.2015.2442260

Tlelo-Cuautle, E., Sanabria-Borbon, A.C., 2016. Optimising operational amplifiers by evolutionary algorithms and g_m/I_d method. *Int. J. Electron.*, **103**(10):1665-1684. http://dx.doi.org/10.1080/00207217.2016.1138522

Trejo-Guerra, R., Tlelo-Cuautle, E., Jiménez-Fuentes, J.M., *et al*., 2012. Integrated circuit generating 3- and 5-scroll attractors. *Commun. Nonl. Sci. Numer. Simul.*, **17**(11): 4328-4335. http://dx.doi.org/10.1016/j.cnsns.2012.01.029

Visweswariah, C., 2003. Death, taxes and failing chips. Proc. 40th Annual Design Automation Conf., p.343-347. http://dx.doi.org/10.1145/775919.775921

Wang, D., Hutson, A.D., 2015. Inversion theorem based kernel density estimation for the ordinary least squares estimator of a regression coefficient. *Commun. Statist. Theory Methods*, **44**(8):1571-1579. http://dx.doi.org/10.1080/03610926.2013.781633

Xu, F., Li, C., Jiang, T., 2015. Printed circuit board model updating based on response surface method. *J. Beijing Univ. Aeronaut. Astronaut.*, **41**(3):449-455 (in Chinese).

Yuan, X., 2009. Application Research of Markov Chain Simulation in Reliability Analysis. PhD Thesis, Northwestern Polytechnical University, Xi'an, China (in Chinese).

Yuan, X.K., Lu, Z.Z., Qiao, H.W., 2010. Conditional probability Markov chain simulation based reliability analysis method for nonnormal variables. *Sci. China Technol. Sci.*, **53**(5):1434-1441. http://dx.doi.org/10.1007/ s11431-010-0138-9

Zhang, H., Zamar, R.H., 2014. Least angle regression for model selection. *Wiley Interdiscipl. Rev. Comput. Statist.*, **6**(2):116-123. http://dx.doi.org/10.1002/wics.1288

Zou, H., 2006. The adaptive Lasso and its Oracle properties. *J. Am. Statist. Assoc.*, **101**(476):1418-1429. http://dx.doi.org/10.1198/016214506000000735

Zou, H., Zhang, H.H., 2009. On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.*, **37**(4): 1733-1751. http://dx.doi.org/10.1214/08-AOS625