Semiconductor Corner Lot Generation Robust to Process Variation: Modeling and Analysis

Abstract

Product characterization is an important phase in developing new semiconductors. The goal is to determine if the new product will function when produced under the extreme edge of fabrication variation; if not, the product might be considered to have insufficient design margin, necessitating circuit redesign. Achieving this goal requires producing a so-called *corner lot* that consists of skew chips, i.e., chips whose key performance parameters that are expected to be around certain targeted extreme values. These skew chips will be extensively tested to determine if their functions still meet specifications. However, due to extensive variation in the fabrication process, few skewed chips can be guaranteed in a produced corner lot, and this is a long-standing frustration in the semiconductor industry. One approach to produce a satisfactory corner lot is through variation reduction of the fabrication process. Despite being a popular research area, variation reduction is a long-term effort that involves both technical and managerial considerations. We approach this problem from a different avenue by treating process variation as given and instead identifying a design strategy that guarantees production of a good corner lot robust to the variation. Specifically, we propose a first-of-its-kind rigorous mathematical formulation about this problem, investigate the theoretical properties and practical implications of this formulation, and further propose several optimal criteria and a corresponding design search algorithm. Applications for a broad range of semiconductor products are presented to demonstrate the universal improvement of the proposed optimal design compared with the traditional design used in current industrial practice.

Keywords: semiconductor manufacturing, product characterization, variation reduction

1. Introduction

A key milestone in the development of a new integrated circuit product is completion of the design process and transfer of tooling to the fabrication plant to produce initial lots of the new product. Chips in these initial lots are extensively tested and analyzed to determine if the design meets specifications. Product characterization is an important step in determining if the new product will be robust to long-term fabrication variation. Since the product development time course is much shorter (months) than the lifetime of the product (10 years or more), variation in the initial lots of the new product will be much smaller than the long-term fabrication variation. This poses a risk that the product may not perform well in the long term. When performing product characterization, a longstanding industry practice is to produce a *corner lot* (Weste and Harris, 2011; Automotive Electronics Council, 2013), which is a lot whose recipe is manipulated to try to achieve extreme values of the long-term fabrication variation for key performance parameters of the product, such as leakage current, circuit frequency, and operating voltage (May and Spanos, 2006).

Chips in the corner lot with performance parameters within a small tolerance around the targeted extreme values are called *skew chips*. These chips will be extensively tested and analyzed to see if they can still meet product specifications. If they do not, the new product may be considered to lack sufficient design margin. This will result in certain circuits being redesigned. Both the product division, who developed the product, and the fabrication factory, which will manufacture it, have an interest in the corner lot because a new product that is robust to long-term fabrication variation will be cheaper and easier to produce, benefitting both the product division and fabrication factory.

To produce a corner lot, the industrial common practice is to make sure the *mean* of the key performance parameter of all the chips in the corner lot is equal to the targeted extreme value (Nakagawa et al., 1999; Wang et al., 2004; Gough, 2014).

Precise control on the mean of a lot is not hard in semiconductor manufacturing because this industry is mature with well-established process recipe to followed by engineers. However, there is no control over the *variance* of the key performance parameter. The variance is typically large because a semiconductor fabrication line involves of hundreds of process steps. Each step has its own inherent variation. The step-wise variation accumulates and propagates, leading to large variance in the key performance parameter of the chips at the end of the fabrication line. As a consequence, when a corner lot is being produced according to the process recipe, the variation will act as random shock that moves the key performance parameter of the chips away from the targeted value. Therefore, it is a common frustration in the current industrial practice that few skew chips can be guaranteed in a produced corner lot. This creates tremendous difficulty for subsequent design evaluation and product characterization. When this happens, additional corner lots will have to be produced, which will increase the lead time and cost of new product development.

One approach to produce a good corner lot is through variation reduction of the fabrication process. There is abundant existing work in variation reduction in semiconductor manufacturing, which can be categorized into four major sub-areas. Next, we briefly review each sub-area:

Wafer defect detection and classification: Recognizing the defect patterns is the first step toward process variation reduction. Research in this sub-area generally develops pattern recognition and statistical models to detect a specific defect or classify different defect types. For example, Wang et al. (2006) proposed an approach comprising a spatial filter, a classification module, and an estimation module to distinguish three types of defect patterns. Yuan and Kuo (2007) proposed a model-based clustering algorithm for spatial defect recognition on semiconductor wafers. Bao et al. (2014) proposed to decompose the thickness variation of wafers into macro and micro-scale variations modelled as a cubic curve and first-order intrinsic Gaussian Markov random field, respectively.

Process monitoring: In semiconductor manufacturing, the process and data have some unique characteristics that drive the development of new control charts. For example, Yeh et al. (2005) proposed a multivariate EWMA control chart for monitoring the critical dimensions of dies sampled from different sites on individual wafers. Zou et al. (2007) proposed a multivariate EWMA control chart to monitor the profiles of the DRIE process and detect changes. Zou et al. (2008) relaxed the linear assumption of the previous approach and proposed a non-linear, non-parametric method for profile monitoring.

Fault root cause diagnosis: To enable root cause diagnosis, one needs to build a model between the quality variables and process parameters such that quality problems can be traced back to the process. Fenner et al. (2009) proposed a Bayesian parallel site model to link wafer deposition uniformity measures to key process parameters. Jin and Liu (2013) proposed the use of piecewise linear regression trees to identify multiple variation propagation modes in multistage semiconductor manufacturing, and then link quality variables at each stage with upstream quality, process, and material variables. Yu and Qin (2009) tackled the fault diagnosis of multi-layer overlay lithography processes using a multistage state space model.

Automatic control: In semiconductor manufacturing, R2R process control is an important tool for stabilizing process output, reducing variation, and improving quality. Recent developments include a batch EWMA controller that considers both batch information and feedback quality information (Wang and Han, 2013), a Smith-EWMA controller that ensures stability at the presence of serious material delay (Jin and Tsung, 2009), and variable EWMA controllers for drifted processes (Tseng et al., 2007; Tseng et al., 2008). In addition to R2R control, Jin and Shi (2012) developed multistage feedforward control methods in semiconductor manufacturing.

Although being an effective approach, variation reduction is a long-term effort that involves thorough process investigation, data analytics, and interventions complicated by technical and non-technical considerations. The present study has a different

focus: we treat process variation as given and aim to produce corner lots robust to the variation. Our study is relevant but different from robust parameter design (Myers et al., 1992; Myers et al., 2016). Here, "robustness"means producing corner lots that contain a *guaranteed* number of skew chips. The basic idea of our proposed approach is that since process variation is treated as fixed, the robustness can be achieved by spreading out the mean. That is, instead of setting the mean of a key performance parameter of all the chips in the corner lot to be the same value (i.e., the targeted extreme value), we can set the means of some chips to be below and others to be above the extreme value. Compared with variation reduction that is relatively long-term effort, this study serves an immediate need in new product development. However, despite the practical value of this study goal, there is neither a rigorous mathematical formulation about the problem nor any theoretical investigations or practical algorithms to guide the search for the best mean-spread-out strategy to achieve robustness. This study aims to bridge the existing gap.

The rest of the paper is organized as follows: Section 2 presents the detailed mathematical formulation of the problem; Section 3 studies theoretical properties of the formulation and their practical implications; Section 4 proposes two optimal criteria to guide the search for the best mean-spread-out strategy, called the "optimal design", and a design search algorithm; Section 5 presents the application; Section 6 is the conclusion. Note that the "optimal design" in the context of this paper has a different meaning from that in experimental designs (Montgomery, 2012). Finally, we would like to point out that this paper focuses on development of the optimal design, not on how to generate corner lots according to the optimal design after it is developed. The latter, nevertheless, is briefly discussed in Appendix B for readers who are interested in the implementation.

2. Problem formulation

2.1 Nomenclature

X_{ijk}	Performance parameter (e.g., circuit frequency) for the kth chip on the jth wafer in the ith lot
μ_j	Mean performance parameter of the jth wafer
L_i, W_{ij}, M_{ijk}	Variance components of X_{ijk} from the lot, wafer and chips levels, $L_i \sim N(0, \sigma_L^2)$, $W_{ij} \sim N(0, \sigma_W^2)$, $M_{ijk} \sim N(0, \sigma_M^2)$
σ^2	Total variance of X_{ijk}, $\sigma^2 = \sigma_L^2 + \sigma_W^2 + \sigma_M^2$
ϵ	A pre-defined tolerance
m	Number of wafers in a corner lot
n	Number of chips on a wafer
δ	Corner lot design parameter
$\tilde{\epsilon}$, $\tilde{\delta}$, $\tilde{\sigma}_L^2$, $\tilde{L}_1^2$	Standardized parameter, $\tilde{\epsilon} = \frac{\epsilon}{\sqrt{\sigma_W^2+\sigma_M^2+\sigma_L^2}}$ $\tilde{\delta} = \frac{\delta}{\sqrt{\sigma_W^2+\sigma_M^2}}$ $\tilde{\sigma}_L^2 = \frac{\sigma_L^2}{\sigma_W^2+\sigma_M^2}$ $\tilde{L}_1 \sim N(0, \tilde{\sigma}_L^2)$

2.2 Formulation

A semiconductor fabrication line consists of hundreds of process steps that turn a bare silicon wafer into one containing hundreds of chips (i.e., integrated circuits). In the final step, key performance parameters of the chips such as leakage current, circuit frequency, and operating voltage, will be tested to see how well the chips work. There are three variation sources for performance parameters of the chips:

- Lot-to-lot variation: a lot is the smallest batch of wafers that will be operated on at a process step. Once a lot is started at a process step, all of the wafers in that lot will be processed. This processing mechanism induces lot-to-lot variation.

- Wafer-to-wafer variation: at each process step, the equipment can either process the wafers in a lot all together (i.e., the so-called batch mode) or sequentially. Both strategies induce wafer-to-wafer variation. In the batch mode, the wafer-to-wafer variation is due to each wafer being in a different position in the equipment. In the sequential mode, it is due to tool variation over time.

- Within-wafer site-to-site variation: each chip has a spatial position on the wafer. Different equipment will induce variation across the surface of wafer in different patterns. This creates within-wafer variation.

Considering the structure of the three variation sources, we can model a performance parameter of a chip as follows:

$$X_{ijk} = \mu + L_i + W_{ij} + M_{ijk}, \tag{1}$$

(1) is the well-known variance components model that has been commonly used to model semiconductor processes in the literature (Diebold, 2001; Drain, 1997; Reda and Nassif, 2010; Yashchin, 1994).

In our case, (1) needs two modifications:

- We focus on corner lots. As discussed in the Introduction, just one corner lot is typically produced due to cost and scheduling considerations. Therefore, we change the lot index i to 1 in (1).

- μ is allowed to vary for different wafers, so μ is replaced by μ_j. In fact, μ_j is a decision variable in our case, i.e., it can be set by the fabrication engineers to achieve a desired value, and how to set μ_j to achieve robustness in corner lot generation is the research question we want to address in this paper.

Considering these modifications, (1) is changed to (2), which is used in the rest of this paper.

$$X_{1jk} = \mu_j + L_1 + W_{1j} + M_{1jk}. \tag{2}$$

Furthermore, recall that the purpose of a corner lot is to have the performance parameter of its chips achieve a targeted extreme value μ_0, which is "extreme value" in the sense that it is usually far from the nominal value of the performance parameter μ_{nominal}, e.g., $\mu_0 = \mu_{\text{nominal}} + 3\sigma$. The chips in the corner lot that have performance parameters within a small tolerance around the targeted extreme value

μ_0 are successful chips, also known as *skew chips*. Mathematically, we can define a skew chip in the following way: $I_{[X_{1jk}]} = 1$ if X_{1jk} is a skew chip in the corner lot and $I_{[X_{1jk}]} = 0$ otherwise, i.e.,

$$I_{[X_{1jk}]} = \begin{cases} 1 & \mu_0 - \epsilon \leq X_{1jk} \leq \mu_0 + \epsilon \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

In this paper, we focus on performance parameters with two-sided tolerance limits and leave one-sided limits (i.e., performance parameters that are the-larger-the-better or the-smaller-the-better) for future investigation. Then, the expected proportion of skew chips in the corner lot is:

$$\begin{aligned} f(L_1; \mu_1, \ldots, \mu_m) &\triangleq E[\frac{\sum_{j=1}^{m}\sum_{k=1}^{n} I_{[X_{1jk}]}}{mn}|L_1] = \frac{\sum_{j=1}^{m}\sum_{k=1}^{n} E[I_{[X_{1jk}]}|L_1]}{mn} \\ &= \frac{\sum_{j=1}^{m}\sum_{k=1}^{n} P(\mu_0 - \epsilon \leq X_{1jk} \leq \mu_0 + \epsilon | L_1)}{mn}. \end{aligned} \tag{4}$$

If the corner lot had been produced, then L_1 would have been observed, i.e., $L_1 = l_1$. l_1 denotes a realization for L_1. Then, the conditional distribution of $X_{1jk}|L_1$ is $N(\mu_j + l_1, \sigma_W^2 + \sigma_M^2)$. Then, (4) becomes:

$$\frac{1}{m}\sum_{j=1}^{m}\left(\phi(\frac{\mu_0 + \epsilon - (\mu_j + l_1)}{\sqrt{\sigma_W^2 + \sigma_M^2}}) - \phi(\frac{\mu_0 - \epsilon - (\mu_j + l_1)}{\sqrt{\sigma_W^2 + \sigma_M^2}})\right), \tag{5}$$

where $\phi(\cdot)$ is the cumulative probability function of the standard normal distribution. However, we are designing the corner lot generation in this paper, i.e., L_1 is not yet observed at this stage. Therefore, the l_1 in (5) should be replaced by the random variable L_1, i.e.,

$$f(L_1; \mu_1, \ldots, \mu_m) = \frac{1}{m}\sum_{j=1}^{m}\left(\phi(\frac{\mu_0 + \epsilon - (\mu_j + L_1)}{\sqrt{\sigma_W^2 + \sigma_M^2}}) - \phi(\frac{\mu_0 - \epsilon - (\mu_j + L_1)}{\sqrt{\sigma_W^2 + \sigma_M^2}})\right). \tag{6}$$

For a fixed design in corner lot generation, i.e., $\mu_1, \cdots, \mu_m$ are given, $f(L_1; \mu_1, \cdots, \mu_m)$ is a random variable because it is a function of the random variable L_1. For example, a *target design* is a design used in the current industrial practice that sets all the μ_j's to the targeted extreme value μ_0, i.e., $(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$. Under the target design, the histogram of $f(L_1; \mu_0, \cdots, \mu_0)$ for a semiconductor process with $m = 24, \mu_0 = \mu_{\text{nominal}} + 3\sigma, \epsilon = 0.5\sigma, \sigma_L^2 = 6, \sigma_W^2 = 2$, and $\sigma_M^2 = 3$, is shown in Figure 1. Note that we are only able to show the histogram using a Monte Carlo simulation but not the probability density function because the distribution of $f(L_1; \mu_0, \cdots, \mu_0)$ does not follow any known parametric distribution.

Figure 1 sheds some light on why the target design can be unsatisfactory. For one corner lot, we get one realization for $f(L_1; \mu_0, \cdots, \mu_0)$, which has a certain (non-zero) probability of being a very small number. For example, $f(L_1; \mu_0, \cdots, \mu_0)$ has 3.1%, 7%, and 11.4% probabilities of being less than 0.05, 0.10, and 0.15, respectively. This means that very few skew chips may be generated in the corner lot, making it difficult for the subsequent design evaluation and product characterization.

To compare with the target design, we show the histogram of $f(L_1; \mu_1^*, \cdots, \mu_m^*)$ in Figure 2 for an optimal design, i.e., the design under the maximum-single-lot-probability criterion proposed in Section 4. Compared with the target design in Figure 1, $f(L_1; \mu_1^*, \cdots, \mu_m^*)$ has only 0.05%, 0.30%, and 0.92% probabilities of being less than 0.05, 0.10, and 0.15, respectively. Especially, there is zero probability for $f(L_1; \mu_1^*, \cdots, \mu_m^*)$ to be zero while this probability is non-zero for $f(L_1; \mu_0, \cdots, \mu_0)$. Also, it is quite obvious that $f(L_1; \mu_1^*, \cdots, \mu_m^*)$ has a much smaller variance than $f(L_1; \mu_0, \cdots, \mu_0)$. All these imply that the optimal design is more likely to produce a guaranteed number of skew chips, a highly favorable property for corner lot generation in practice. Therefore, the research question we will need to tackle is the following: *what design in terms of the setting for* $(\mu_1, \cdots, \mu_m)^T$ *will lead to a favorable shape for the distribution of* $f(L_1; \mu_1, \cdots, \mu_m)$? This is an extremely challenging question because the distribution of $f(L_1; \mu_1, \cdots, \mu_m)$ is non-parametric; thus theoretically,

to fully characterize this distribution requires an infinite number of parameters. To provide a tractable solution, we focus on studying the mean and variance of the distribution of $f(L_1; \mu_1, \cdots, \mu_m)$ in this paper (Section 3), which further leads to two proposed optimal criteria and an algorithm for searching the optimal designs that satisfy the industrial need (Section 4).

3. Theoretical Properties and Practical Implications

Let $E(f(L_1; \mu_1, \cdots, \mu_m))$ and $Var(f(L_1; \mu_1, \cdots, \mu_m))$ denote the mean and variance of $f(L_1; \mu_1, \cdots, \mu_m)$, respectively. The mean $E(f(L_1; \mu_1, \cdots, \mu_m))$ and the variance $Var(f(L_1; \mu_1, \cdots, \mu_m))$ are functions of the design parameters $(\mu_1, \cdots, \mu_m)^T$. In this section, we will first show that the target design, i.e.,$(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$, maximizes $E(f(L_1; \mu_1, \cdots, \mu_m))$ (Theorem 1). This implies that among all possible designs, the target design achieves the highest mean proportion of skew chips in a corner lot. This seems to suggest that the target design is favorable. However, the mean only reflects the proportion of skew chips in a corner lot *in the long term*. For a *single* corner lot that is to be produced, the proportion of skew chips in the lot is also heavily influenced by the variance, i.e., $Var(f(L_1; \mu_1, \cdots, \mu_m))$. Even though the target design has the highest mean proportion of skew chips, if the variance is large, a single corner lot produced under the target design may still have a chance of including zero or few skew chips. This is well demonstrated by the example in Figure 1. Therefore, we further study the property of $Var(f(L_1; \mu_1, \cdots, \mu_m))$ in this section. Specifically, Theorem 2 proves that the target design does not minimize $Var(f(L_1; \mu_1, \cdots, \mu_m))$, i.e., there are other designs with smaller variances.

Theorem 1 *The target design, i.e.,*$(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$ *, is the global maximum solution for* $E(f(L_1; \mu_1, \cdots, \mu_m))$*, i.e.,*

$$(\mu_0, \cdots, \mu_0)^T = \underset{\mu_1, \cdots, \mu_m}{\operatorname{argmax}} E(f(L_1; \mu_1, \cdots, \mu_m)).$$

Theorem 2 *When* $m > 1$ *and* $\tilde{\epsilon}^2 \leq \frac{-20a_2 + \sqrt{400a_2^2 - 840a_1a_3}}{14a_3(1+\tilde{\sigma}_L^2)}$*, where*

$$a_1 = -(1 + 2\tilde{\sigma}_L^2)(1 + \tilde{\sigma}_L^2)^5 + (1 + 2\tilde{\sigma}_L^2)^{\frac{5}{2}}(1 + \tilde{\sigma}_L^2)^2,$$

$$a_2 = -(1+2\tilde{\sigma}_L^2)^{\frac{5}{2}}(1+\tilde{\sigma}_L^2) + (\frac{3}{2}\tilde{\sigma}_L^4 + 2\tilde{\sigma}_L^2 + 1 + \frac{3}{8m})(1+\tilde{\sigma}_L^2)^4, \textit{ and}$$

$$a_3 = (1+2\tilde{\sigma}_L^2)^{\frac{5}{2}},$$

there exists a design $(\mu_1^*, \cdots, \mu_m^*)^T$*, such that*

$$Var(f(L_1; \mu_1^*, \cdots, \mu_m^*)) < Var(f(L_1; \mu_0, \cdots, \mu_0)).$$

Here we discuss the practical validity of two conditions in Theorem 2: m is the number of wafers in a corner lot, which is typically greater than one. The condition on ϵ is also easily satisfied, as demonstrated in the Application section that includes a broad range of semiconductor products for all of which this condition is met.

Theorems 1 and 2 imply that the target design maximizes the mean of the proportion of skew chips in a corner lot, but it does not minimize the variance of the proportion. This suggests that we can potentially search for a design that has a smaller variance with some acceptable sacrifice of the mean. In practice, a design with a smaller variance is desirable because it provides assurance for producing at least *some* skew chips in a single corner lot. To search for such a design, we need to identify the direction in the m-dimensional space of the design parameters $(\mu_1, \cdots, \mu_m)^T$, along which the variance is decreasing. Corollary 2.1 presents the variance-decreasing directions, as a result from Theorem 2. Please see the proofs of Theorems 1 and 2 and Corollary 2.1 in the Appendix.

Corollary 2.1 *Let* $(1, \cdots, 1)^T$ *be a vector consisting of* m *ones. (i) Any direction orthogonal to* $\frac{1}{\sqrt{m}}(1, \cdots, 1)^T$ *is a decreasing direction for* $Var(f(L_1; \mu_1, \cdots, \mu_m))$ *at* $(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$*.(ii) Furthermore,* $Var(f(L_1; \mu_1, \cdots, \mu_m))$ *along any of the variance-decreasing directions in (i) decreases at the same rate.*

Corollary 2.1 (ii) suggests that among the variance-decreasing directions identified in (i), there is no mathematical preference because the variance along any of these directions decreases at the same rate. However, from the engineering perspective, practical considerations must be taken into account when choosing a direction.

Specifically, for cost-saving and error-prone purposes, a variance-decreasing direction that needs minimum process adjustments is desirable. This leads to choosing the direction of $\frac{1}{\sqrt{m}}(-1, \cdots, -1, 1, \cdots, 1)^T$ when m is an even number (i.e., there is an even number of wafers in the corner lot). Here, $\frac{1}{\sqrt{m}}(-1, \cdots, -1, 1, \cdots, 1)^T$ is a vector for which the first $m/2$ elements are -1 and the remaining $m/2$ elements are 1. With this direction, there are only two different settings for the m wafers, i.e., $\mu_0 - \delta$ and $\mu_0 + \delta$. Therefore, process adjustment is minimized. Factoring in this engineering consideration, we can replace the μ_j in (6) with $\mu_0 - \delta$ for $j = 1, \cdots, m/2$ and with $\mu_0 + \delta$ for $j = m/2 + 1, \cdots, m$. Then, (6) becomes:

$$\begin{aligned} f(L_1;\delta) &= \frac{1}{2}\big(\phi(\frac{\epsilon+\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}) - \phi(\frac{-\epsilon+\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}})\big) \\ &+ \frac{1}{2}\big(\phi(\frac{\epsilon-\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}) - \phi(\frac{-\epsilon-\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}\big). \end{aligned} \tag{7}$$

Likewise, when m is an odd number, the direction of $\frac{1}{\sqrt{m}}(-1, \cdots, -1, 0, 1, \cdots, 1)^T$ can be chosen. $\frac{1}{\sqrt{m}}(-1, \cdots, -1, 0, 1, \cdots, 1)^T$ is a vector for which the first $(m-1)/2$ elements are -1, last $(m-1)/2$ elements are 1, and middle element is 0. With this direction, there are three different settings for the m wafers, i.e., $\mu_0 - \delta$, μ_0, and $\mu_0 + \delta$. Then, (6) becomes:

$$\begin{aligned} f(L_1;\delta) &= \frac{m-1}{2m}\big(\phi(\frac{\epsilon+\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}) - \phi(\frac{-\epsilon+\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}})\big) \\ &+ \frac{m-1}{2m}\big(\phi(\frac{\epsilon-\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}) - \phi(\frac{-\epsilon-\delta-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}\big) \\ &+ \frac{1}{m}\big(\phi(\frac{\epsilon-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}}) - \phi(\frac{-\epsilon-L_1}{\sqrt{\sigma_W^2+\sigma_M^2}})\big). \end{aligned} \tag{8}$$

Summarizing the results of this section, we can conclude that δ is a design parameter. When $\delta = 0$, the corresponding design is the target design (i.e., the design used in the current industrial common practice). As δ increases, the corresponding design will have a decreasing $Var(f(L_1;\delta))$ accompanied by a decreasing $E(f(L_1;\delta))$. Therefore, the research question boils down to identifying the value for δ that

corresponds to an optimal design. To answer this question, we first need to define the optimal criterion and then develop an algorithm that searches for a design satisfying the optimal criterion (the optimal design), which will be presented in the next section.

4. Optimal Criteria and Optimal Design Search

We propose two optimal criteria:

(i) Maximum-single-lot-probability criterion. In industry, it is common that only one corner lot is allowed to be produced due to resource or time constraints. Since skew chips in the corner lot will be used for design evaluation and product characterization, sufficient skew chips need to be produced. That is, there is usually a requirement on the proportion of skew chips in the corner lot. Denote the required proportion by α. Then, we would like to find a design that maximizes the probability for a single corner lot to contain at least α proportion of skew chips, i.e.,

$$\delta^{(i)*} = \underset{\delta}{\operatorname{argmax}}\, P(f(L_1; \delta) > \alpha). \tag{9}$$

Note that this criterion maximizes the *probability* of a corner lot having at least α proportion of skew chips. As long as this maximum probability is not one, it is still possible for a single corner lot to contain less than α proportion of skew chips. When this happens, the decision can be to take whatever skew chips that have been produced for design evaluation and product characterization, if the resource or/and production timeline is so tight that no more corner lots can be afforded. Alternatively, if having a required number of skew chips is more important and it outweighs the concerns of more resource spent and possible delay in the production schedule, the decision can be to produce more corner lots until the required number of skew chips is reached. This motivates the second criterion presented as follows.

(ii) Minimum-expected-cost criterion. When a single corner lot fails to produce at least α proportion of skew chips, a second corner lot will be produced. In general, corner lots have to be continuously generated until at least α proportion of skew chips is accumulated. Each corner lot generation is associated with a production cost, c.

Then, the expected production cost for the needed corner lots is: $c \times E_{L_1}(M(\delta, \alpha))$, $M(\delta, \alpha) = \min\{n : Y_1 + Y_2 + \cdots + Y_n \geq \alpha\}$, where Y_1, Y_2, $\cdots$, Y_n are independently and identically distributed with $f(L_1; \delta)$. Without loss of generality, we assume unit cost, i.e., $c = 1$, in the remainder of this paper. Then, the optimal design is one for which the δ minimizes the expected production cost, i.e.,

$$\delta^{(ii)*} = \underset{\delta}{\operatorname{argmin}}\, E_{L_1}(M(\delta, \alpha)). \tag{10}$$

The optimization problems in (9) and (10) have the properties of no analytical forms for the objective functions, no constraints, and only one decision variable δ. Considering these properties, numerical search methods are appropriate for solving the optimizations. We adopt a popular search algorithm that is a combination of golden section search and successive parabolic interpolation and is also implemented in R software. This algorithm is guaranteed to converge with a super-linear converging rate (Brent, 2013). More detailed steps for solving the optimizations in (9) and (10) are presented by a flow chart in Figure 3. Finally, we want to point out that both criteria were not designed to "guarantee" a certain pre-defined proportion of number of skewed chips. The latter would be ideal, but not possible because the proportion of skewed chips in a corner lot is a random variable.

5. Application

In this section, we present the results of applying the proposed optimal criteria and associated design search algorithm to a broad range of semiconductor products. We focus on one important performance parameter of semiconductor chips, which is the circuit frequency. Standardized notations are used for clarity of the presentation.

Note that it is customary in industry to present the tolerance in units of the total standard deviation. Then, (7) becomes

$$\begin{aligned} f(\tilde{L}_1; \tilde{\delta}) &= \frac{1}{2}\big(\phi(\tilde{\epsilon}\sqrt{1+\tilde{\sigma}_L^2} + \tilde{\delta} - \tilde{L}_1) - \phi(-\tilde{\epsilon}\sqrt{1+\tilde{\sigma}_L^2} + \tilde{\delta} - \tilde{L}_1)\big) \\ &+ \frac{1}{2}\big(\phi(\tilde{\epsilon}\sqrt{1+\tilde{\sigma}_L^2} - \tilde{\delta} - \tilde{L}_1) - \phi(-\tilde{\epsilon}\sqrt{1+\tilde{\sigma}_L^2} - \tilde{\delta} - \tilde{L}_1)\big), \end{aligned} \tag{11}$$

and the optimal criteria in (9) and (10) become (13) and (14), respectively.

$$\tilde{\delta}^{(i)*} = \underset{\tilde{\delta}}{\text{argmax}}\, P(f(\tilde{L}_1; \tilde{\delta}) > \alpha). \tag{12}$$

$$\tilde{\delta}^{(ii)*} = \underset{\tilde{\delta}}{\text{argmin}}\, E_{\tilde{L}_1}(M(\tilde{\delta}, \alpha)). \tag{13}$$

We start out by focusing on one type of product and illustrating how to search for the optimal design for corner lot generation. For this product, the product division requires $\alpha = 0.15$. A corner lot consists of 24 wafers and 800 chips. $\alpha = 0.15$ translates into $0.15 \times 24 \times 800 = 2880$ skew chips, which is considered to be an adequate number for product characterization and design evaluation. The tolerance of this product is 0.5σ, which is given by engineering design. Furthermore, we need to estimate the variance components. This requires a large amount of historical data from the same distribution in order to ensure the quality of statistical estimation. To this end, we retrieve historical data from the old generations of this (same) product from our industry collaborator. Because semiconductor manufacturing has high production volume, a large quantity of data is available, which includes measurements on circuit frequency for 13076487 chips contained on 9425 wafers in 400 lots. Variance component estimation in semiconductor processes has been a well-studied topic and the best approach that has been recommended in the literature and also widely used in industry is the ANOVA method (Jenson, 2002). Using the ANOVA method, we obtain the estimates to be $\hat{\sigma}_L^2 = 0.24$,$\hat{\sigma}_W^2 = 0.15$, $\hat{\sigma}_M^2 = 0.61$. Note that these are point estimates for the variance components. To account for the estimation uncertainty, we can employ the sampling distributions that $\frac{df_L \hat{\sigma}_L^2}{\sigma_L^2} \sim \chi^2_{df_L}$, $\frac{df_W \hat{\sigma}_W^2}{\sigma_W^2} \sim \chi^2_{df_W}$, and $\frac{df_M \hat{\sigma}_M^2}{\sigma_M^2} \sim \chi^2_{df_M}$, where $df_L = 399$, $df_W = 9424$, and $df_M = 13076487$, and get confidence interval estimation. Specifically, the 95% confidence intervals for the variance components are $\sigma_L^2 \in [0.210, 0.277]$, $\sigma_W^2 \in [0.146, 0.154]$, and $\sigma_M^2 \in [0.6095, 0.6104]$. These confidence intervals are very narrow due to the large sample size, suggesting that estimation uncertainty is not much of a concern. Therefore, we decide to use the point estimates

to conduct an initial search for the optimal design, i.e., $\hat{\tilde{\sigma}}_L^2 = \frac{\hat{\sigma}_L^2}{\hat{\sigma}_W^2 + \hat{\sigma}_M^2} = 0.316$.

Give all the above information as input, we first choose the maximum-single-lot-probability criterion. The design search algorithm finds the optimal design parameter to be $\hat{\tilde{\delta}}^{(i)*} = 1.24$. Under this optimal design, the probability for a single lot to consist of at least 0.15 proportion of skew chips is $P(f(L_1; 1.24) > 0.15) = 1.000$. Furthermore, to evaluate how this finding is impacted by the variance component estimation uncertainty, we generate 100 random samples for σ_L^2, σ_W^2, σ_M^2 by sampling from $\frac{df_L \hat{\sigma}_L^2}{\chi^2_{df_L}}$, $\frac{df_W \hat{\sigma}_W^2}{\chi^2_{df_W}}$, and $\frac{df_M \hat{\sigma}_M^2}{\chi^2_{df_M}}$, respectively. This allows us to obtain an empirical distribution for the optimal design parameter $\tilde{\delta}^{(i)*}$. Based on the empirical distribution, we can obtain a 95% confidence interval for $\tilde{\delta}^{(i)*}$, which is $\tilde{\delta}^{(i)*} \in [0.95, 1.47]$. This confidence interval is very narrow, indicating that the uncertainty in finding the optimal design parameter that is introduced by the variance component estimation uncertainty is not much of a concern.

Alternatively, we can choose the minimum-expected-cost criterion, under which the design search algorithm finds the optimal design to be $\hat{\tilde{\delta}}^{(i)**} = 1.35$. Under this optimal design, the expected production cost (i.e., the expected number of corner lots to be produced) is $E_{\tilde{L}_1}(M(1.35, 0.15)) = 1.000$. This means that one corner lot is expected to be sufficient. Furthermore, to evaluate how this finding is impacted by the variance component estimation uncertainty, we use the same sampling approach as the one previously used for the maximum-single-lot-probability criterion and obtain a 95% confidence interval for $\tilde{\delta}^{(ii)*}$, which is $\tilde{\delta}^{(ii)*} \in [1.23, 1.53]$. This confidence interval is also very narrow. Note that the confidence intervals for the optimal design parameters under the two optimal criteria overlap, which implies that the two parameters have no statistical difference. This explains why the optimal designs under the two criteria both suggest that a single corner lot is sufficient for satisfying the requirement of producing at least 0.15 proportion of skew chips in the corner lot.

Furthermore, we would like to repeat the above analysis on a variety of other semiconductor products. The products are different in terms of their variance

components and tolerances. Typical semiconductor products have a tolerance ranging from 0.1 to 0.5 of the total process standard deviation. Using the standardized notation in (11), this means $\tilde{\epsilon} \in [0.1, 0.5]$. Also, according to (11), individual variance components do not matter but only the ratio of the lot-to-lot variance to the sum of wafer-to-wafer and within-wafer variances, i.e., $\tilde{\sigma}_L^2$, do. Typical semiconductor products have $\tilde{\sigma}_L^2 \in [0.2, 2]$, which is the range we focus on in this study. Each combination of $\tilde{\epsilon} \in [0.1, 0.5]$ and $\tilde{\sigma}_L^2 \in [0.2, 2]$ corresponds to a different type of product. For each type of product, we run the design search algorithm according to the two proposed optimal criteria. Results associated with the maximum-single-lot-probability criterion are shown in Tables 1-3. Specifically, Table 1 shows the optimal design parameter $\tilde{\delta}^{(i)*}$ including a point estimate and a 95% confidence interval. Table 2 shows the probability for a single lot to consist of at least 0.15 proportion of skew chips under the optimal design, i.e., $P(f(\tilde{L}_1; \tilde{\delta}^{(i)*}) > 0.15)$ including a point estimate and a 95% confidence interval. Table 3 shows the percentage of increase in this probability by comparing the optimal design and the target design with a p value indicating the statistical significant of this comparison. Likewise, results associated with the minimum-expected-cost criterion are shown in Tables 4-6. We summarize the observations on Table 1-6 as follows:

1) Table 1 shows that when $\tilde{\epsilon} = 0.1$, the optimal design parameter under the maximum-single-lot-probability criterion is not available (NA), i.e., no design exists to generate a corner lot with at least 0.15 proportion of skew chips. This is because the tolerance, $\tilde{\epsilon}$, is so small that it is very difficult for a chip to fall within this tolerance limit and be qualified as a skew chip. When the tolerance gets larger, i.e., $\tilde{\epsilon} \in [0.2, 0.5]$, the optimal design parameter exists.

2) At $\tilde{\epsilon} = 0.2$ and $\tilde{\sigma}_L^2 \in [0.2, 1.4]$, the optimal design parameter is zero, meaning that the optimal design is found to be the target design. At other settings of $\tilde{\epsilon}$ and $\tilde{\sigma}_L^2$, the optimal design is different from the target design. A general observation on the optimal design parameters in Table 1 is that at a fixed $\tilde{\epsilon}$, $\tilde{\delta}^{(i)*}$ increases as $\tilde{\sigma}_L^2$

increases. This makes sense because a larger $\tilde{\sigma}_L^2$ means a larger lot-to-lot variation, in which case the optimal design would need to spread the means of the wafers in the corner lot to a greater extent in order to protect against the larger variation. Also, at a fixed $\tilde{\sigma}_L^2$, $\tilde{\delta}^{(i)*}$ generally increases as $\tilde{\epsilon}$ increases. A possible explanation is that a larger tolerance makes it easier for a chip to be qualified as a skew chip, and therefore the optimal design could spread the means more to allow more chips inside the tolerance limits.

3) Focusing on products with $\tilde{\epsilon} \geq 0.3$, i.e., when the optimal design is different from the target design, Table 2 shows a high probability (0.885-1.000) for a single lot to consist of at least 0.15 proportion of skew chips under the optimal design. Table 3 shows that in terms of this probability, the optimal design has a statistically significant improvement over the target design ($p < 0.001$). Greater improvement is seen for products with larger lot-to-lot variation, i.e., $\tilde{\sigma}_L^2$. For products with $\tilde{\epsilon} = 0.2$ and $\tilde{\sigma}_L^2 \in [1.6, 2.0]$, although the optimal design still has a statistically significant improvement over the target design (Table 3), the optimal design is unable to generate a single lot with at least 0.15 proportion of skew chips with a high probability. This is a joint effect of the small tolerance that makes it difficult for a chip to be qualified as a skew chip and the large lot-to-lot variation that makes corner lot generation inherently more difficult.

4) Under the minimum-expected-cost criterion, Table 4 shows that when $\tilde{\epsilon}$ is small, the optimal design parameter is zero, meaning that the optimal design is found to be the target design. Also, at a fixed $\tilde{\epsilon}$, $\tilde{\delta}^{(ii)*}$ increases as $\tilde{\sigma}_L^2$ increases; at a fixed $\tilde{\sigma}_L^2$, $\tilde{\delta}^{(ii)*}$ generally increases as $\tilde{\epsilon}$ increases. These trends are similar to those observed for Table 1 under the maximum-single-lot-probability criterion and the same explanations apply.

5) Table 5 shows that for products with $\tilde{\epsilon} \geq 0.3$, the expected number of corner lots to be produced is very close to one. Table 6 shows that the optimal design has a statistically significant improvement over the target design ($p < 0.001$) in terms

of reducing the expected number of corner lots. Greater improvement is seen for products with larger lot-to-lot variation, i.e., $\tilde{\sigma}_L^2$.

6) Comparing the results under the two optimal criteria, we find similar trends in terms of how the optimal design parameter varies with respect to $\tilde{\sigma}_L^2$ and $\tilde{\epsilon}$ (Tables 1 and 4). Comparing Tables 2 and 5, we can see that the higher the probability for a single lot to consist of at least 0.15 proportion of skew chips (Table 2), the smaller the expected number of corner lots that need to be produced (Table 5). This correlation is intuitive and makes sense. Also, we observe significant improvement of the optimal design compared with the target design over different products under both criteria (Tables 3 and 6). Despite the consistency of the results by the two criteria, they each have a unique value in practice. The maximum-single-lot-probability criterion is more appropriate if the resource or/and production timeline is so tight that only a single corner lot can be afforded. The minimum-expected-cost criterion is more appropriate if having a required number of skew chips for product characterization is more important so that corner lots must be continuously produced until the required number is met. The former criterion is usually adopted for relatively more mature products while the latter for new products with potentially high risk and high return.

6. Conclusion

Corner lot generation is an important task in the product characterization of new semiconductors. However, the current industrial practice is primarily empirically based and ineffective. In this paper, we provided a first-of-its-kind rigorous mathematical formulation for corner lot generation, investigated the theoretical properties of the formulation and its practical implications, and further developed a practical algorithm to identify the optimal design under two proposed optimal criteria. Applications on a broad range of semiconductor products were presented that demonstrated universal improvement of the optimal design compared with the target design across various products. An immediate future research direction is to extend the proposed methodology to additional dimensions, i.e., to consider multiple

key performance parameters of the chips in corner lot generation.

Appendix A

Proof of Theorem 1

This proof uses the original definition of $f(L_1; \mu_1, \ldots, \mu_m)$ in (4), i.e.,

$$\begin{aligned} E(f(L_1;\mu_1,\ldots,\mu_m)) &= E[\frac{\sum_{j=1}^{m}\sum_{k=1}^{n} E[I_{[X_{1jk}]}|L_1]}{mn}] = \frac{1}{m}\sum_{j=1}^{m} E[I_{[X_{1jk}]}] \\ &= \frac{1}{m}\sum_{j=1}^{m} P(\mu_0 - \epsilon \leq X_{1jk} \leq \mu_0 + \epsilon). \end{aligned} \tag{A-1}$$

The distribution of X_{1jk} according to (2) is $X_{1jk} \sim N(\mu_j, \sigma^2)$, where $\sigma^2 = \sigma_L^2 + \sigma_W^2 + \sigma_M^2$. Therefore, (A-1) becomes:

$$E(f(L_1;\mu_1,\ldots,\mu_m)) = \frac{1}{m}\sum_{j=1}^{m}\int_{\mu_0-\epsilon}^{\mu_0+\epsilon}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu_j)^2}{2\sigma^2}\}dx. \tag{A-2}$$

Based on the property of normal distributions, it is clear that (A-2) is maximized when $(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$. Consequently, $E(f(L_1; \mu_1, \ldots, \mu_m))$ is maximized at the target design.

Proof of Theorem 2

The basic concept of the proof is to demonstrate that the target design, i.e., $(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$, is a local maximum or a saddle point for $Var(f(L_1; \mu_1, \cdots, \mu_m)$ when ϵ is below an upper bound. Being a local maximum or saddle point naturally means that a design with a smaller variance than the target design exists (i.e., the statement in Theorem 2).

Briefly, the proof consists of the following three steps: (i) we derive the first-order partial derivatives of $Var(f(L_1; \mu_1, \cdots, \mu_m)$ with respect to $\mu_i, i = 1, \cdots, m$, and further show that they are equal to zero at the target design. (ii) (i) implies that the target design is either a local extreme or saddle point for $Var(f(L_1; \mu_1, \cdots, \mu_m)$. To confirm, we derive the second-order partial derivatives and further the Hessian matrix at the target design. The derivation shows that, under some condition, the

Hessian matrix can either have both positive and negative eigenvalues (meaning that the target design is a saddle point) or all negative eigenvalues (meaning that the target design is a local maximum). In both cases, there is a design with a smaller variance than the target design. (iii) Finally, we derive the condition in (ii) for the two cases to hold, which turns out to be an upper bound on ϵ.

In what follows, we will present the detailed derivation in each of the three steps. For notation simplicity, let

$$\check{\epsilon} = \frac{\epsilon}{\sqrt{\sigma_W^2 + \sigma_M^2}}, \tilde{\mu}_j = \frac{\mu_j - \mu_0}{\sqrt{\sigma_W^2 + \sigma_M^2}}, \tilde{L}_1 \sim N(0, \tilde{\sigma}_L^2).$$

Then, the target design is $(\tilde{\mu}_1, \cdots, \tilde{\mu}_m)^T = (0, \cdots, 0)^T$. Using the new notations, (6) becomes (A-3):

$$\begin{aligned} f(L_1; \tilde{\mu}_1, \cdots, \tilde{\mu}_m) &= \frac{1}{m}\sum_{j=1}^{m}\big(\phi(\tilde{\mu}_j + \tilde{L}_1 + \check{\epsilon}) - \phi(\tilde{\mu}_j + \tilde{L}_1 - \check{\epsilon})\big) \\ &= \frac{1}{m}\sum_{j=1}^{m}\int_{-\check{\epsilon}}^{\check{\epsilon}} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x - \tilde{\mu}_j - \tilde{L}_1)^2}{2}\}dx. \qquad \text{(A-3)} \end{aligned}$$

For notation simplicity, we will reuse μ_j, L_1, σ_L^2, ϵ as $\tilde{\mu}_j$, $\tilde{L}_1$, $\tilde{\sigma}_L^2$, $\check{\epsilon}$ in the subsequent derivations. Then $Var(f(L_1; \mu_1, \cdots, \mu_m))$ can be written as

$$\begin{aligned} Var(f(L_1; \mu_1, \cdots, \mu_m)) &= Var(\frac{1}{m}\sum_{j=1}^{m}\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx) \\ &= \frac{1}{m^2}\sum_{j=1}^{m} Var(\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx) + \\ &\quad \frac{2}{m^2}\sum_{i<j} Cov(\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_i-L_1)^2}{2}\}dx, \\ &\quad \int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx) \\ &\triangleq \frac{1}{m^2}\sum_{j=1}^{m} g(\mu_j) + \frac{2}{m^2}\sum_{i<j} h(\mu_i, \mu_j) \qquad \text{(A-4)} \end{aligned}$$

Here,

$$g(\mu_j) = Var(\int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx),$$

and is therefore a function of μ_j.

$$h(\mu_i,\mu_j) = Cov(\int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x-\mu_i-L_1)^2}{2}\}dx, \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx),$$

and is therefore a function of μ_i and μ_j.

(i) Derive the first-order partial derivatives of $Var(f(L_1;\mu_1,?\mu_m))$ at the target design

According to (A-4),

$$\frac{\partial Var(f(L_1;\mu_1,\cdots,\mu_m))}{\partial \mu_j} = \frac{1}{m^2}\frac{\partial g(\mu_j)}{\partial \mu_j} + \frac{2}{m^2}\sum_{i\neq j}\frac{\partial h(\mu_i,\mu_j)}{\partial \mu_j}. \qquad \text{(A-5)}$$

Next, we derive $\frac{\partial g(\mu_j)}{\partial \mu_j}$ and $\frac{\partial h(\mu_i,\mu_j)}{\partial \mu_j}$.

$$\begin{aligned}
\frac{\partial g(\mu_j)}{\partial \mu_j} &= \frac{Var(\int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx)}{\partial \mu_j} \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_L} \exp\{-\frac{L_1^2}{2\sigma_L^2}\} \int_{-\epsilon}^{\epsilon} \exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx \\
&\quad (\exp\{-\frac{(-\mu_j-\epsilon-L_1)}{2}\} - \exp\{-\frac{(-\mu_j+\epsilon-L_1)}{2}\})dL_1 - \\
&\quad \frac{1}{\pi}\int_{-\epsilon}^{\epsilon} \frac{1}{1+\sigma_L^2} \exp\{-\frac{(x-\mu_j)^2}{2(1+\sigma_L^2)}\}(\exp\{-\frac{(\mu_j+\epsilon)^2}{2(1+\sigma_L^2)}\} - \exp\{-\frac{(\mu_j-\epsilon)^2}{2(1+\sigma_L^2)}\})dx \\
&= (A-B)-(C-D),
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_L} \exp\{-\frac{L_1^2}{2\sigma_L^2}\} \int_{-\epsilon}^{\epsilon} \exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx \\
&\quad \exp\{-\frac{(-\mu_j-\epsilon-L_1)}{2}\}dL_1 \\
&= \frac{1}{\pi}\int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{1+2\sigma_L^2}} \exp\{-\frac{(x-\mu_j)^2+(\mu_j+\epsilon)^2}{2}\} \exp\{\frac{\sigma_L^2(x-2\mu_j-\epsilon)}{2(1+2\sigma_L^2)}\}dx,
\end{aligned}$$

$$\begin{aligned} B &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_L}\exp\{-\frac{L_1^2}{2\sigma_L^2}\}\int_{-\epsilon}^{\epsilon}\exp\{-\frac{(x-\mu_j-L_1)^2}{2}\}dx \\ &\quad \exp\{-\frac{(-\mu_j+\epsilon-L_1)}{2}\}dL_1 \\ &= \frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{1+2\sigma_L^2}}\exp\{-\frac{(x-\mu_j)^2+(\mu_j-\epsilon)^2}{2}\}\exp\{\frac{\sigma_L^2(x-2\mu_j+\epsilon)}{2(1+2\sigma_L^2)}\}dx, \end{aligned}$$

$$C = \int_{-\epsilon}^{\epsilon}\frac{1}{\pi}\frac{1}{1+\sigma_L^2}\exp\{-\frac{(x-\mu_j)^2}{2(1+\sigma_L^2)}\}\exp\{-\frac{(\mu_j+\epsilon)^2}{2(1+\sigma_L^2)}\}dx,$$

$$D = \int_{-\epsilon}^{\epsilon}\frac{1}{\pi}\frac{1}{1+\sigma_L^2}\exp\{-\frac{(x-\mu_j)^2}{2(1+\sigma_L^2)}\}\exp\{-\frac{(\mu_j-\epsilon)^2}{2(1+\sigma_L^2)}\}dx.$$

$$\begin{aligned} \frac{\partial h(\mu_i,\mu_j)}{\partial \mu_j} &= \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{(1+2\sigma_L^2)}} \\ &\quad \exp\{-\frac{(1+\sigma_L^2)(\epsilon+\mu_i)^2+(1+\sigma_L^2)(x-\mu_j)^2+2\sigma_L^2(\epsilon+\mu_i)(x-\mu_j)}{2(1+2\sigma_L^2)}\}dx - \\ &\quad \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon}\frac{1}{\sqrt{(1+2\sigma_L^2)}} \\ &\quad \exp\{-\frac{(1+\sigma_L^2)(\epsilon-\mu_i)^2+(1+\sigma_L^2)(x+\mu_j)^2+2\sigma_L^2(\epsilon-\mu_i)(x+\mu_j)}{2(1+2\sigma_L^2)}\}dx + \\ &\quad \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon}\frac{1}{(1+\sigma_L^2)}\exp\{-\frac{(x-\mu_j)^2}{2(1+\sigma_L^2)}\}\exp\{-\frac{(\epsilon-\mu_i)^2}{2(1+\sigma_L^2)}\}dx - \\ &\quad \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon}\frac{1}{(1+\sigma_L^2)}\exp\{-\frac{(x-\mu_j)^2}{2(1+\sigma_L^2)}\}\exp\{-\frac{(\epsilon+\mu_i)^2}{2(1+\sigma_L^2)}\}dx. \end{aligned}$$

Furthermore, it can be shown that $(A-B)-(C-D)=0$ when $\mu_j=0$. Therefore, $\frac{\partial g(\mu_j)}{\partial \mu_j}|_{\mu_j=0}=0$. It can also be shown that $\frac{\partial h(\mu_i,\mu_j)}{\partial \mu_j}|_{\mu_i=0,\mu_j=0}=0$. These results imply that the first-order partial derivatives of $Var(f(L_1;\mu_1,\cdots,\mu_m))$ at the target design $(\mu_1,\cdots,\mu_m)^T=(\mu_0,\cdots,\mu_0)^T$ are zero.

(ii) Derive the second-order partial derivatives and Hessian matrix of

$Var(f(L_1;\mu_1,\cdots,\mu_m))$ **at the target design**

$$\begin{aligned}\frac{\partial^2 g(\mu_j)}{\partial \mu_j^2}|_{\mu_j=0} &= \frac{1}{\pi}\int_{-\epsilon}^{\epsilon} \exp\{-\frac{x^2+\epsilon^2}{2}\}\exp\{\frac{\sigma_L^2(x-\epsilon)}{2(1+2\sigma_L^2)}\}\frac{2(x-\epsilon)}{(1+2\sigma_L^2)^{3/2}}dx + \\ &\quad \frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\frac{2\epsilon}{(1+\sigma_L^2)^2}\exp\{-\frac{x^2+\epsilon^2}{2(1+\sigma_L^2)}\}dx \\ &= \frac{1}{\pi}\int_{-\epsilon}^{\epsilon} s(x)\psi(x)dx, \end{aligned} \tag{A-6}$$

where

$$s(x) = \exp\{-\frac{\sigma_L^4(x^2+\epsilon^2)+2(\sigma_L^2+\sigma_L^4)x\epsilon}{2(1+2\sigma_L^2)(1+\sigma_L^2)}\}\frac{2(x-\epsilon)}{(1+2\sigma_L^2)^{3/2}} + \frac{2\epsilon}{(1+\sigma_L^2)^2},$$

$$\psi(x) = \exp\{-\frac{x^2+\epsilon^2}{2(1+\sigma_L^2)}\}.$$

Furthermore,

$$\begin{aligned}\frac{\partial^2 h(\mu_i,\mu_j)}{\partial \mu_j^2}|_{\mu_i=0,\mu_j=0} &= -\frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\exp\{-\frac{(1+\sigma_L^2)(x^2+\epsilon^2)+2\sigma_L^2 x\epsilon}{2(1+2\sigma_L^2)}\}\frac{(\sigma_L^2 x+(1+\sigma_L^2)\epsilon)}{(1+2\sigma_L^2)^{3/2}}dx + \\ &\quad \frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\frac{2\epsilon}{(1+\sigma_L^2)^2}\exp\{-\frac{x^2+\epsilon^2}{2(1+\sigma_L^2)}\}dx, \end{aligned} \tag{A-7}$$

and

$$\frac{\partial^2 h(\mu_i,\mu_j)}{\partial \mu_i \partial \mu_j}|_{\mu_i=0,\mu_j=0} = \frac{1}{\pi\sqrt{(1+2\sigma_L^2)}}(\exp\{-\frac{\epsilon^2}{1+2\sigma_L^2}\}-\exp\{-\epsilon^2\}). \tag{A-8}$$

Using the results in (A-6),(A-7), and (A-8), the Hessian matrix for $Var(f(L_1;\mu_1,\cdots,\mu_m))$ at $\mu_i=0$, $i=1,\cdots,m$ is:

$$\begin{pmatrix} a & c & c & \cdots & c \\ c & a & c & \cdots & c \\ c & c & a & \cdots & c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \cdots & a \end{pmatrix}, \tag{A-9}$$

where

$$a = \frac{1}{m^2}\frac{\partial g(\mu_j)}{\partial \mu_j^2}|_{\mu_j=0} + \frac{2(m-1)}{m^2}\frac{\partial h(\mu_i,\mu_j)}{\partial \mu_j^2}|_{\mu_i=0,\mu_j=0},$$

$$c = \frac{2}{\pi\sqrt{(1+2\sigma_L^2)}m^2}(\exp\{-\frac{\epsilon^2}{1+2\sigma_L^2}\} - \exp\{-\epsilon^2\}).$$

The eigenvalues of the Hessian matrix are $\lambda_1 = a+(m-1)c$, $\lambda_2 = \cdots = \lambda_m = a-c$. According to the properties of a Hessian matrix, if $\lambda_1 > 0$ and $\lambda_2 < 0$, then $(0,\cdots,0)^T$ is a saddle point for $Var(f(L_1;\mu_1,\cdots,\mu_m))$; if $\lambda_1 < 0$ and $\lambda_2 < 0$, then $(0,\cdots,0)^T$ is a local maximum point for $Var(f(L_1;\mu_1,\cdots,\mu_m))$. Next, we will derive a sufficient condition for $\lambda_2 = a-c < 0$.

(iii) Derive the condition on ϵ

Note that $e^{-x} \geq 1-x$ and $e^{-x} \leq 1-x+\frac{x^2}{2}$ for all $x \geq 0$. Substituting all exponential functions, we have

$$a - c \;\leq\; \frac{2\epsilon^2}{15m\pi}\frac{1}{(1+2\sigma_L^2)^{\frac{5}{2}}(1+\sigma_L^2)^4}(30a_1 + 20a_2\epsilon^2 + 7a_3\epsilon^4),$$

where

$$a_1 = -(1+2\sigma_L^2)(1+\sigma_L^2)^5 + (1+2\sigma_L^2)^{\frac{5}{2}}(1+\sigma_L^2)^2,$$

$$a_2 = -(1+2\sigma_L^2)^{\frac{5}{2}}(1+\sigma_L^2) + (\frac{3}{2}\sigma_L^4 + 2\sigma_L^2 + 1 + \frac{3}{8m})(1+\sigma_L^2)^4,$$

$$a_3 = (1+2\sigma_L^2)^{\frac{5}{2}}.$$

Since $\sigma_L^2 > 0$, it is obvious that $a_1 < 0, a_2 > 0, a_3 > 0$. To make $\lambda_2 = a-c < 0$, we must ensure that $30a_1 + 20a_2\epsilon^2 + 7a_3\epsilon^4 < 0$. Thus, a sufficient condition for $30a_1 + 20a_2\epsilon^2 + 7a_3\epsilon^4 < 0$ is

$$\epsilon^2 \leq \frac{-20a_2 + \sqrt{400a_2^2 - 840a_1a_3}}{14a_3}. \tag{A-10}$$

By replacing all notations back, we have

$$\check{\epsilon}^2 \le \frac{-20a_2 + \sqrt{400a_2^2 - 840a_1a_3}}{14a_3}, \tag{A-11}$$

where

$$a_1 = -(1+2\tilde{\sigma}_L^2)(1+\tilde{\sigma}_L^2)^5 + (1+2\tilde{\sigma}_L^2)^{\frac{5}{2}}(1+\tilde{\sigma}_L^2)^2,$$

$$a_2 = -(1+2\tilde{\sigma}_L^2)^{\frac{5}{2}}(1+\tilde{\sigma}_L^2) + (\frac{3}{2}\tilde{\sigma}_L^4 + 2\sigma_L^2 + 1 + \frac{3}{8m})(1+\tilde{\sigma}_L^2)^4,$$

$$a_3 = (1+2\tilde{\sigma}_L^2)^{\frac{5}{2}}.$$

Also note that $\tilde{\epsilon}^2 = \frac{1}{1+\tilde{\sigma}_L^2}\check{\epsilon}^2$, then (A-11) becomes

$$\tilde{\epsilon}^2 \le \frac{-20a_2 + \sqrt{400a_2^2 - 840a_1a_3}}{14a_3(1+\tilde{\sigma}^2)}, \tag{A-12}$$

which is the condition in Theorem 2.

Proof of Corollary 2.1:

The proof of Theorem 2 shows that the Hessian matrix in (A-9) at the target design has one eigenvalue $\lambda_1 = a+(m-1)c$ and $m-1$ other identical eigenvalues, i.e., $\lambda_2 = \cdots = \lambda_m = a-c$. Furthermore, it can be derived that the eigenvector associated with λ_1 is $\boldsymbol{e}_1 = \frac{1}{\sqrt{m}}(1,\cdots,1)^T$. Let $\boldsymbol{e}_2,\cdots,\boldsymbol{e}_m$ be the eigenvectors associated with eigenvalues $\lambda_2,\cdots,\lambda_m$, respectively. Then, any vector of unit length orthogonal to $\boldsymbol{e}_1$ can be expressed as $\boldsymbol{x} = \sum_{i=2}^{m}\tau_i\boldsymbol{e}_i$, where $\tau_i, i=2,\cdots,m$ satisfy $\sum_{i=2}^{m}\tau_i^2 = 1$. Let $\boldsymbol{\mu} = (\mu_1,\cdots,\mu_m)^T$ be a design in a small neighborhood of the target design along the direction of $\boldsymbol{x}$. Then, we can write the Taylor's expansion of $Var(f(L_1;\mu_1,\cdots,\mu_m))$ with respect to the target design as:

$$\begin{aligned} Var(f(L_1;\mu_1,\cdots,\mu_m)) \approx\ & Var(f(L_1;0,\cdots,0)) + \\ & \sum_{i=1}^{m}\mu_i \cdot \frac{\partial Var(f(L_1;\mu_1,\cdots,\mu_m))}{\partial\mu_i}|_{\boldsymbol{\mu}=\boldsymbol{0}} + \\ & \boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu} \end{aligned} \tag{A-13}$$

Here, $\boldsymbol{H}$ is the Hessian matrix in (A-9). In (A-13), the first-order term is zero because the $\frac{\partial Var(f(L_1;\mu_1,\cdots,\mu_m))}{\partial \mu_i}|_{\boldsymbol{\mu}=\mathbf{0}} = 0$ according to the derivation of Theorem 2. The second-order term can be further derived as follows:

$$\begin{aligned} \boldsymbol{\mu}^T \boldsymbol{H} \boldsymbol{\mu} &= ||\boldsymbol{\mu}||^2 (\sum_{i=2}^{m} \tau_i \boldsymbol{e}_i)^T \boldsymbol{H} \sum_{i=2}^{m} \tau_i \boldsymbol{e}_i \\ &= ||\boldsymbol{\mu}||^2 \sum_{i=2}^{m} \tau_i^2 \boldsymbol{e}_i^T \boldsymbol{H} \boldsymbol{e}_i = ||\boldsymbol{\mu}||^2 (a-c) < 0. \end{aligned} \tag{A-14}$$

where $||\boldsymbol{\mu}||^2 = \sum_{i=1}^{m} \mu_i^2$. Therefore, $Var(f(L_1;\mu_1,\cdots,\mu_m)) < Var(f(L_1;0,\cdots,0))$. This proves (i) in Corollary 2.1. Moreover, because the $m-1$ eigenvalues are the same, this naturally implies that (ii) in Corollary 2.1 holds.

Appendix B

Corner Lot Generation According to the Optimal Design

We would like to discuss practical aspects related to how corner lots will be generated according to the optimal design. After the product division receives a recommended optimal design parameter, δ^*, the engineers need to manipulate the process recipe so as to produce half of the wafers with average chip performance (e.g., circuit frequency) equal to $\mu_0+\delta^*$ and the other half equal to $\mu_0-\delta^*$ in each corner lot. Recall that μ_0 is an engineering-defined extreme value of the performance parameter for purpose of product characterization and design evaluation. Semiconductor is a mature manufacturing process in the sense that there are well-established empirical and physical models (Gray et al., 2009) for guiding the manipulation of recipes to achieve desired average chip performance. Take circuit frequency as an example. To achieve a desired average circuit frequency for the chips on a wafer, e.g., to make $\mu = \mu_0 + \delta^*$, there is a well-known empirical model that links μ with the N-type and P-type MOSFET device currents, I_{DN} and I_{DP}, i.e., $\mu = f(I_{DN}, I_{DP})$. Using this model, we can identify the specific I_{DN}^* and I_{DP}^* that help achieve the desired $\mu_0 + \delta^*$. Next, to decide what process parameter settings can lead to the I_{DN}^* and I_{DP}^*, two well-known physical models exist, i.e., $I_{DN} = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TN})^2$

and $I_{DP} = \mu_p C_{ox} \frac{W}{L}(V_{SG} - V_{TP})^2$. μ_n and μ_p are electron (N) and hole (P) carrier mobilities. L is the transistor length. V_{TN} and V_{TP} are threshold voltages of the N-type and P-type transistors. In theory, all these process parameters can be modulated to achieve the desired I^*_{DN} and I^*_{DP}. In practice, some may be easier to modulate than others and which process parameter(s) to modulate for each particular product is known from process design. There are also detailed recipes/guidance on how to adjust the process parameters. For example, adjustment on L can be achieved by adjusting exposure energy in lithography or the etch time in plasma etch. Adjustment on V_{TN} and V_{TP} can be achieved by adjusting the implant dose in ion implantation. In summary, a combination of mature empirical/physical models and process/product design knowledge exists to make sure average chip performance in each corner lot can be achieved as recommended by the optimal design. This can also be achieved with high precision, except when the amount of adjustment needed on certain equipment is so fine that it is even smaller than the smallest adjustment that is physically possible. However, such fine adjustment is rarely needed in practice. Despite the high precision of recipe manipulation, it would still be of great practical interest to study how small deviation in the recipe from the desire level would impact the level of achievement on the optimal criteria, which is a future research direction we would like to pursue.

References

[1] Automotive Electronics Council, Component Technical Committee, Guideline for Characterization of Integrated Circuits, February 18, 2013.

[2] Brent, R.P., Algorithms for minimization without derivatives, Dover publications, New York (2013).

[3] Bao, L., Wang, K., Jin, R., A hierarchical model for characterising spatial wafer variations, International Journal for Production Research, Vol. 52, No. 6, 1827-1842 (2014).

[4] Diebold, A.C., Handbook of Silicon Semiconductor Metrology, CRC Press (2001).

[5] Drain, David C., Statistical Methods for Industiral Process Control, CRC Press (1997).

[6] Gough, A.M., Semiconductor Sample Generation Experimental Designs Robust to Random Process Shocks, Technical report (2014).

[7] Gray, Paul R., Hurst, Paul J., Lewis, Stephen H., Meyer, Robert G., Analysis and Design of Analog Integrated Circuits, fifth edition, Wiley (2009).

[8] Fenner, J.S., Jeong, Y.S., Jeong, M.K., Lu, J.C. A Bayesian parallel site methodology with an application to uniformity modeling in semiconductor manufacturing, IIE Transactions, Vol. 41, Issue 9, 754-763 (2009).

[9] Jensen, C.R. , Variance Component Calculations: Common Methods and Misapplications in the Semiconductor Industry, Quality Engineering, Vol. 14, Issue 4, 647-657 (2002).

[10] Jin, R. and Shi, J., Reconfigured piecewise linear regression tree for multistagemanufacturing process control, IIE Transactions, Vol. 44, Issue 4, 249-261 (2012).

[11] Jin, R., Liu, K., Multimode Variation Modeling and Process Monitoring for Serial-Parallel Multistage Manufacturing Processes, IIE Transactions, Vol. 45, Issue 6, 617-629 (2013).

[12] Jin, M., Tsung, F., Smith-EWMA run-to-run control schemes for a process with measurement delay, IIE Transactions, Vol. 41, Issue 4, 346-358 (2009).

[13] May, Gary S., Costas J. Spanos, Fundamentals of semiconductor manufacturing and process control, John Wiley & Sons, Inc. (2006).

[14] Montgomery, D.C., Introduction to Statistical Quality Control, 7th Edition, John Wiley & Sons, Inc. (2012).

[15] Montgomery, D.C., Design and Analysis of Experiments, 8th Edition, John Wiley & Sons, Inc. (2012).

[16] Myers, R.H., Khuri, A.I., Vining G., Response surface alternatives to the Taguchi robust parameter design approach. The American Statistician, The American Statistician, Vol. 46, Issue 9, 131-139 (1992).

[17] Myers, R.H., Montgomery, D.C., Anderson-Cook, C.M., Response surface methodology: process and product optimization using designed experiments, John Wiley & Sons (2016).

[18] Nakagawa, O.S., Chang, N., Lin. S., et al., Circuit impact and skew-corner analysis of stochastic process variation in global interconnect, IEEE International Conference in Interconnect Technology, 230-232 (1999).

[19] Reda, S., Nassif, S.R., Accurate Spatial Estimation and Decomposition Techniques for Variability Characterization, IEEE Transactions on Semiconductor Manufactoring, Vol. 23, No. 3, 345-357 (2010).

[20] Tseng, S.T., Tsung, F., Liu, P,Y., Variable EWMA run-to-run controller for drifted processes, IIE Transactions, Vol. 39, Issue 3, 291-301 (2007).

[21] Tseng, S.T., Jou, B.Y., Liao, C.H., Adaptive variable EWMA controller for drifted processes, IIE Transactions, Vol. 42, Issue 4, 247-259 (2010).

[22] Wang, C.H., Kuo, W., Bensmail, H., Detection and classification of defect patterns on semiconductor wafers, IIE Transactions, Vol. 38, Issue 12, 1059-1068 (2006).

[23] Wang, D.T., McNall, W.A., Statistical Model based ASIC Skew Selection Method, IEEE Workshop on Microelectronics and Electron Devices, 64-66 (2004).

[24] Wang, K., Han, K., A batch-based run-to-run process control scheme for semiconductor manufacturing, IIE Transactions, Vol. 45, Issue 6, 658-669 (2013).

[25] Weste, Neil H.E., Harris, D.M., CMOS VLSI Design: A Circuits and Systems Perspective, 4th Edition, Addison-Wesley (2011).

[26] Yashchin, Emmanuel, Monitoring Variance Components, Technometrics, Vol. 36, No.4, 379-393 (1994).

[27] Yeh, Authur B., Huwang, L., Wu, C.W., A multivariate EWMA control chart for monitoring process variability with individual observations, IIE Transactions, Vol. 37, Issue 11, 1023-1035 (2005).

[28] Yuan, T., Kuo, W., A model-based clustering approach to the recognition of the spatial defect patterns produced during semiconductor fabrication, IIE Transactions, Vol. 40, Issue 2, 93-101 (2007).

[29] Yu, J., Qin, S.J., Variance component analysis based fault diagnosis of multi-layer overlay lithography processes, IIE Transactions, Vol. 41, Issue 9, 764-775 (2009).

[30] Zou, C., Tsung, F., Wang, Z., Monitoring General Linear Profiles Using Multivariate Exponentially Weighted Moving Average Schemes, Technometrics, Vol. 49, No.4 (2007).

[31] Zou, C., Tsung, F., Wang, Z., Monitoring Profiles Based on Nonparametric Regression Methods, Technometrics, Vol. 50, No. 4, 512-526 (2008).

Table 1: The optimal design parameter $\tilde{\delta}^{(i)*}$ (point estimate and 95% confidence interval) under the maximum-single-lot-probability criterion for different combinations of $\tilde{\epsilon}$ and $\tilde{\sigma}_L^2$ (i.e., different products).

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
0.2	*NA*	0.00 [0.00,0.00]	0.55 [0.45, 0.64]	1.19 [1.18, 1.24]	1.13 [1.01, 1.20]
0.4	*NA*	0.00 [0.00,0.00]	0.88 [0.81, 1.07]	1.37 [1.32, 1.39]	1.48 [1.39, 1.54]
0.6	*NA*	0.00 [0.00,0.00]	1.18 [1.16,1.20]	1.38 [1.35, 1.40]	1.54 [1.50, 1.62]
0.8	*NA*	0.00 [0.00,0.00]	1.24 [1.22, 1.27]	1.47 [1.45, 1.50]	1.65 [1.60, 1.68]
1.0	*NA*	0.00 [0.00,0.00]	1.29 [1.26, 1.32]	1.52 [1.49, 1.55]	1.71 [1.66, 1.73]
1.2	*NA*	0.00 [0.00,0.00]	1.33 [1.30, 1.37]	1.56 [1.54, 1.59]	1.75 [1.71, 1.78]
1.4	*NA*	0.00 [0.00,0.00]	1.37 [1.33, 1.39]	1.56 [1.53, 1.59]	1.78 [1.73, 1.82]
1.6	*NA*	0.37 [0,00 0.58]	1.40 [1.37,1.43]	1.63 [1.60, 1.67]	1.82 [1.78, 1.85]
1.8	*NA*	0.63 [0.32,0.83]	1.43 [1.39,1.47]	1.66 [1.64,1.70]	1.86 [1.82,1.90]
2.0	*NA*	0.80 [0.46,1.08]	1.46 [1.42,1.49]	1.69 [1.66,1.73]	1.89 [1.83,1.93]

Table 2: The probability for a single lot to consist of at least 0.15 proportion of skew chips under the optimal design in Table 1, i.e., $P(f(\tilde{L}_1; \tilde{\delta}^{(i)*}) > 0.15)$ (point estimate and 95% confidence interval).

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
0.2	*NA*	0.772 [0.754, 0.791]	0.982 [0.980, 0.990]	1.000 [0.999, 1.000]	1.000 [0.999, 1.000]
0.4	*NA*	0.714 [0.698, 0.725]	0.940 [0.933, 0.951]	0.998 [0.997,0.999]	1.000 [0.999, 1.000]
0.6	*NA*	0.673 [0.662, 0.690]	0.932 [0.927, 0.937]	0.995 [0.990, 0.997]	0.999 [0.998, 0.999]
0.8	*NA*	0.652 [0.637, 0.667]	0.927 [0.922, 0.933]	0.989 [0.986,0.993]	0.998 [0.996, 0.999]
1.0	*NA*	0.631 [0.616, 0.648]	0.921 [0.914, 0.923]	0.982 [0.979, 0.988]	0.995 [0.993, 0.997]
1.2	*NA*	0.615 [0.598, 0.627]	0.914 [0.904, 0.924]	0.978 [0.971, 0.985]	0.992 [0.989, 0.995]
1.4	*NA*	0.603 [0.586, 0.617]	0.907 [0.900, 0.916]	0.978 [0.977, 0.979]	0.989 [0.983, 0.993]
1.6	*NA*	0.590 [0.576, 0.604]	0.901 [0.893, 0.910]	0.965 [0.957, 0.972]	0.985 [0.981, 0.990]
1.8	*NA*	0.581 [0.573,0.596]	0.895 [0.884,0.904]	0.959 [0.950,0.966]	0.982 [0.976,0.987]
2.0	*NA*	0.579 [0.574,0.596]	0.885 [0.876,0.899]	0.951 [0.943,0.961]	0.978 [0.972,0.984]

Table 3: The percentage of improvement of the optimal design compared with the target design, i.e., $[P(f(\tilde{L}_1;\tilde{\delta}^{(i)*}) > 0.15) - P(f(\tilde{L}_1;0) > 0.15)]/P(f(\tilde{L}_1;0) > 0.15)$, with p value indicating statistical significance of the improvement.

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	0.1	0.2	0.3	0.4	0.5
0.2	NA	0 (p=0.191)	0.3% (p<0.001)	0.2% (p<0.001)	0.1% (p<0.001)
0.4	NA	0 (p=0.150)	1.6% (p<0.001)	3.2% (p<0.001)	1.6% (p<0.001)
0.6	NA	0 (p=0.060)	6.9% (p<0.001)	6.5% (p<0.001)	3.8% (p<0.001)
0.8	NA	0 (p=0.188)	11.3% (p<0.001)	9.2% (p<0.001)	6.8% (p<0.001)
1.0	NA	0 (p=0.062)	14.6% (p<0.001)	12.9% (p<0.001)	9.1% (p<0.001)
1.2	NA	0 (p=0.55)	18.2% (p<0.001)	15.2% (p<0.001)	11.0% (p<0.001)
1.4	NA	0 (p=0.97)	20.8% (p<0.001)	15.8% (p<0.001)	14.4% (p<0.001)
1.6	NA	1.4% (p<0.001)	23.6% (p<0.001)	20.1% (p<0.001)	15.1% (p<0.001)
1.8	NA	0.9% (p<0.001)	25.9% (p<0.001)	22.7% (p<0.001)	17.6% (p<0.001)
2.0	NA	2.5% (p<0.001)	27.1% (p<0.001)	23.6% (p<0.001)	20.7% (p<0.001)

Table 4: The optimal design parameter $\tilde{\delta}^{(ii)*}$ (point estimate and 95% confidence interval) under the minimum-expected-cost criterion for different combinations of $\tilde{\epsilon}$ and $\tilde{\sigma}_L^2$ (i.e., different products).

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	0.1	0.2	0.3	0.4	0.5
0.2	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.25 [0.98,1.27]	1.17 [0.94,1.30]
0.4	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.92 [0.87,1.04]	1.33 [1.21,1.38]	1.45 [1.31,1.57]
0.6	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.14 [1.08,1.20]	1.35 [1.30,1.44]	1.56 [1.42, 1.63]
0.8	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.21 [1.16,1.25]	1.46 [1.34,1.51]	1.65 [1.50,1.67]
1.0	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.23 [1.16,1.31]	1.49 [1.37,1.54]	1.69 [1.57,1.73]
1.2	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.31 [1.21,1.35]	1.53 [1.45,1.58]	1.71 [1.61, 1.76]
1.4	0.00 [0.00,0.00]	0.00 [0.00,0.00]	1.33 [1.27,1.38]	1.57 [1.50,1.63]	1.71 [1.63, 1.78]
1.6	0.00 [0.00,0.00]	0.65 [0.54,0.75]	1.39 [1.31,1.43]	1.60 [1.55,1.66]	1.75 [1.62, 1.78]
1.8	0.00 [0.00,0.00]	0.82 [0.72,0.93]	1.41 [1.33,1.45]	1.64 [1.56,1.69]	1.85 [1.74,1.88]
2.0	0.67 [0.55,0.82]	0.95 [0.83,1.06]	1.44 [1.33,1.47]	1.66 [1.58,1.71]	1.86 [1.77,1.91]

Table 5: The expected production cost of corner lots under the optimal design in Table 4, i.e., $E_{\tilde{L}_1}(M(\tilde{\delta}^{(ii)*}, 0.15))$ (point estimate and 95% CI).

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	0.1	0.2	0.3	0.4	0.5
0.2	2.208 [2.191, 2.232]	1.225 [1.209,1.242]	1.017 [1.011,1.026]	1.000 [1.000,1.000]	1.000 [1.000,1.000]
0.4	2.308 [2.285,2.323]	1.293 [1.277,1.307]	1.060 [1.051,1.068]	1.002 [1.000,1.004]	1.001 [1.000,1.001]
0.6	2.364 [2.340,2.396]	1.332 [1.318,1.356]	1.071 [1.064,1.078]	1.006 [1.003,1.009]	1.001 [1.000,1.002]
0.8	2.416 [2.389,2.451]	1.378 [1.358,1.400]	1.074 [1.067,1.084]	1.010 [1.007,1.019]	1.002 [1.001,1.005]
1.0	2.463 [2.433,2.502]	1.408 [1.389,1.437]	1.088 [1.073,1.099]	1.019 [1.012,1.024]	1.006 [1.003,1.007]
1.2	2.508 [2.473,2.549]	1.450 [1.416,1.483]	1.090 [1.080,1.100]	1.024 [1.019,1.029]	1.009 [1.005,1.012]
1.4	2.548 [2.512,2.591]	1.467 [1.448,1.495]	1.095 [1.087,1.115]	1.031 [1.023,1.039]	1.013 [1.008,1.020]
1.6	2.586 [2.552,2.625]	1.483 [1.464,1.500]	1.100 [1.096,1.119]	1.040 [1.028,1.046]	1.018 [1.012,1.027]
1.8	2.618 [2.576,2.661]	1.476 [1.468,1.497]	1.118 [1.101,1.131]	1.042 [1.036,1.054]	1.019 [1.014,1.025]
2.0	2.638 [2.613,2.673]	1.476 [1.464,1.493]	1.121 [1.109,1.135]	1.056 [1.041,1.065]	1.024 [1.018,1.031]

Table 6: The percentage of reduction in expected production cost of the optimal design compared with the target design, i.e., $[E_{\tilde{L}_1}(M(0, 0.15)) - E_{\tilde{L}_1}(M(\tilde{\delta}^{(ii)*}, 0.15))]/E_{\tilde{L}_1}(M(0, 0.15))$, with p value indicating statistical significance of the reduction.

$\tilde{\sigma}_L^2$ \ $\tilde{\epsilon}$	0.1	0.2	0.3	0.4	0.5
0.2	0 (*p=0.27*)	0 (*p=0.87*)	0 (*p=0.15*)	0.2% (*p<0.001*)	0.1% (*p<0.001*)
0.4	0 (*p=0.07*)	0 (*p=0.61*)	1.0% (*p<0.001*)	2.6% (*p<0.001*)	1.1% (*p<0.001*)
0.6	0 (*p=0.07*)	0 (*p=0.55*)	4.8% (*p<0.001*)	5.4% (*p<0.001*)	4.0% (*p<0.001*)
0.8	0 (*p=0.65*)	0 (*p=0.33*)	8.4% (*p<0.001*)	7.9% (*p<0.001*)	6.0% (*p<0.001*)
1.0	0 (*p=0.18*)	0 (*p=0.66*)	10.4% (*p<0.001*)	10.4% (*p<0.001*)	7.6% (*p<0.001*)
1.2	0 (*p=0.95*)	0 (*p=0.16*)	12.1% (*p<0.001*)	11.5% (*p<0.001*)	9.4% (*p<0.001*)
1.4	0 (*p=0.32*)	0 (*p=0.47*)	14.2% (*p<0.001*)	13.3% (*p<0.001*)	11.6% (*p<0.001*)
1.6	0 (*p=0.06*)	2.8% (*p<0.001*)	16.3% (*p<0.001*)	14.5% (*p<0.001*)	12.9% (*p<0.001*)
1.8	0 (*p=0.31*)	4.6% (*p<0.001*)	16.8% (*p<0.001*)	16.2% (*p<0.001*)	14.1% (*p<0.001*)
2.0	0.3% (*p<0.001*)	6.6% (*p<0.001*)	19.3% (*p<0.001*)	16.8% (*p<0.001*)	14.9% (*p<0.001*)

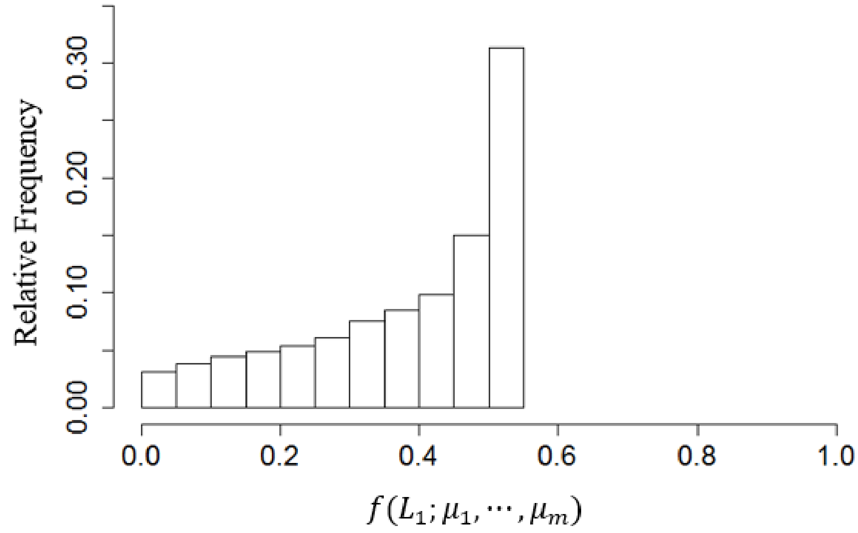

Figure 1: Histogram of $f(L_1; \mu_1, \cdots, \mu_m)$ under the target design
$(\mu_1, \cdots, \mu_m)^T = (\mu_0, \cdots, \mu_0)^T$.
$m = 24, \sigma_W^2 = 2, \sigma_M^2 = 3, \sigma_L^2 = 6, \epsilon = 0.5\sigma, \mu_0 = \mu_{\text{nominal}} + 3\sigma$.

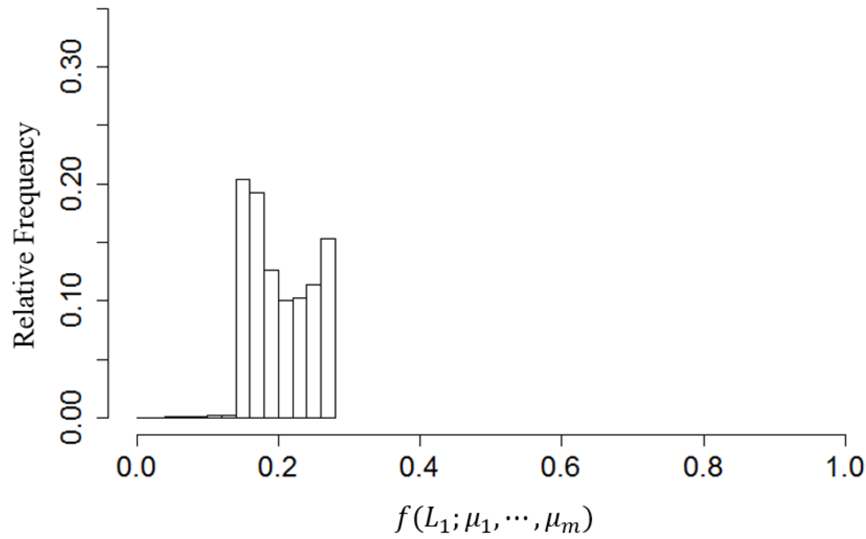

Figure 2: Histogram of $f(L_1; \mu_1, \cdots, \mu_m)$ under an optimal design, i.e., the design under the maximum-single-lot-probability criterion proposed in Section 4.
$(\mu_1, \cdots, \mu_m)^T = (\mu_1^*, \cdots, \mu_m^*)^T$,
$m = 24, \sigma_W^2 = 2, \sigma_M^2 = 3, \sigma_L^2 = 6, \epsilon = 0.5\sigma, \mu_0 = \mu_{\text{nominal}} + 3\sigma$, $\mu_i^* = \mu_0 - 3.89$ for $i = 1, \cdots, 12$, $\mu_i^* = \mu_0 + 3.89$ for $i = 13, \cdots, 24$.

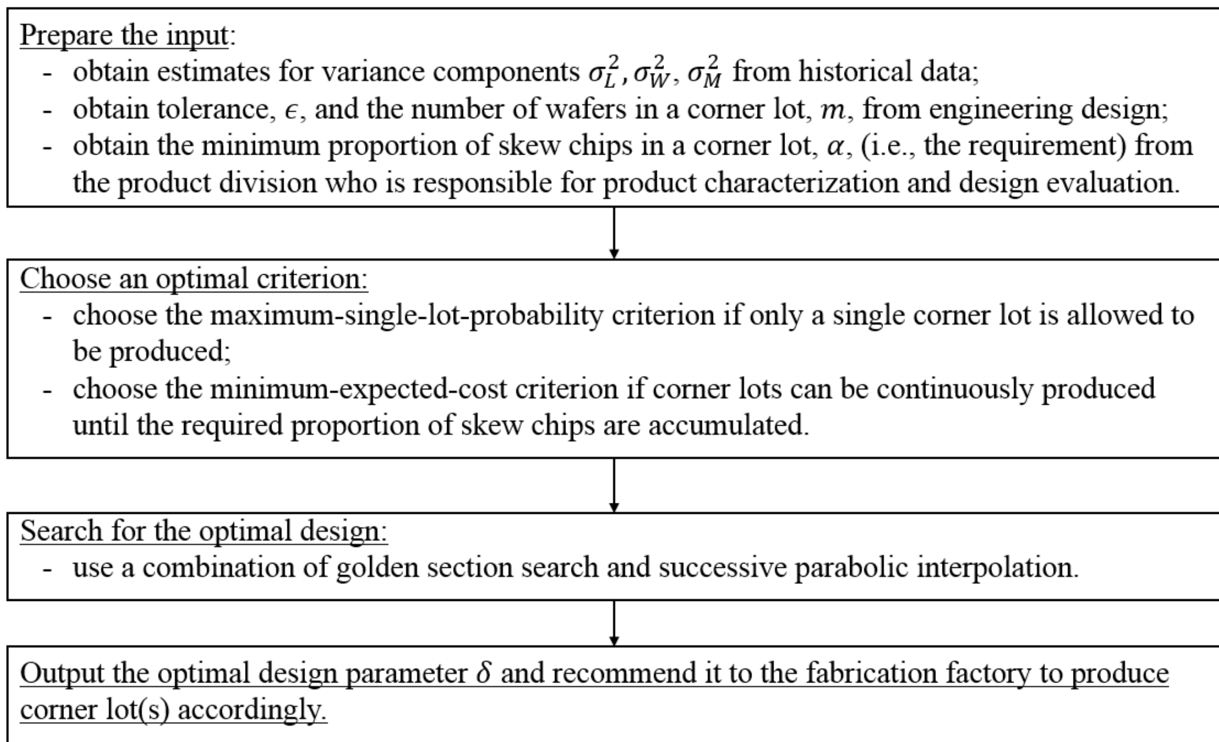

Figure 3: Steps of the optimal design search algorithm.