9. Modeling and Analyzing Yield, Burn-In and Reliability for Semiconductor Manufacturing: Overview

The demand for proactive techniques to model yield and reliability and to deal with various infant mortality issues are growing with increased integrated circuit (IC) complexity and new technologies toward the nanoscale. This chapter provides an overview of modeling and analysis of yield and reliability with an additional burn-in step as a fundamental means for yield and reliability enhancement.

After the introduction, the second section reviews yield modeling. The notions of various yield components are introduced. The existing models, such as the Poisson model, compound Poisson models and other approaches for yield modeling, are introduced. In addition to the critical area and defect size distributions on the wafers, key factors for accurate yield modeling are also examined. This section addresses the issues in improving semiconductor yield including how clustering may affect yield.

The third section reviews reliability aspects of semiconductors such as the properties of failure mechanisms and the typical bathtub failure rate curve with an emphasis on the high rate of early failures. The issues for reliability improvement are addressed.

The fourth section discusses several issues related to burn-in. The necessity for and effects of burn-in are examined. Strategies for the level and type of burn-in are examined. The literature on optimal burn-in policy is reviewed. Often percentile residual life can be a good measure of performance in addition to the failure rate or reliability commonly used.

The fifth section introduces proactive methods of estimating semiconductor reliability from yield information using yield–reliability relation models. Time-dependent and -independent models are discussed.

The last section concludes this chapter and addresses topics for future research and development.

9.1 **Semiconductor Yield** 154
9.1.1 Components of Semiconductor Yield 155
9.1.2 Components of Wafer Probe Yield 155
9.1.3 Modeling Random Defect Yield ... 155
9.1.4 Issues for Yield Improvement..... 158

9.2 **Semiconductor Reliability** 159
9.2.1 Bathtub Failure Rate 159
9.2.2 Occurrence of Failure Mechanisms in the Bathtub Failure Rate 159
9.2.3 Issues for Reliability Improvement 160

9.3 **Burn-In** 160
9.3.1 The Need for Burn-In 160
9.3.2 Levels of Burn-In 161
9.3.3 Types of Burn-In 161
9.3.4 Review of Optimal Burn-In Literature 162

9.4 **Relationships Between Yield, Burn-In and Reliability** 163
9.4.1 Background 163
9.4.2 Time-Independent Reliability without Yield Information 164
9.4.3 Time-Independent Reliability with Yield Information 164
9.4.4 Time-Dependent Reliability 165

9.5 **Conclusions and Future Research** 166

References 166

Since Jack Kilby of Texas Instruments invented the first integrated circuit (IC) in 1958, the semiconductor industry has consistently developed more complex chips at ever decreasing cost. Feature size has shrunk by 30% and die area has grown by 12% every three years [9.1]. The number of transistors per chip has grown exponentially while semiconductor cost per function has been reduced at the historical rate of 25% per year. As shown in Table 9.1, the semiconductor market will reach $213 billion in 2004, which represents 28.5% growth over

Table 9.1 Industry sales expectations for IC devices [9.2]

Device type	Billion dollars				Percent growth			
	2003	2004	2005	2006	03/02	04/03	05/04	06/05
Discretes	13.3	16.0	17.0	16.7	8.1	20.2	6.2	−2.0
Optoelectronics	9.5	13.1	14.9	15.3	40.6	37.3	13.4	2.9
Actuators	3.5	4.8	5.7	6.3	[a]	35.3	18.9	9.1
Bipolar digital	0.2	0.2	0.2	0.2	−4.2	10.6	−16.3	−25.0
Analog	26.8	33.7	37.0	37.0	12.0	25.6	9.9	−0.1
MOS micro	43.5	52.4	57.2	57.6	14.3	20.4	9.2	0.6
MOS logic	36.9	46.4	50.6	49.6	18.1	25.7	9.1	−2.1
MOS memory	32.5	46.9	49.1	47.6	20.2	44.4	4.6	−3.1
Total	166.4	213.6	231.7	230.0	18.3	28.5	8.5	−0.7

[a] A growth rate is not meaningful to show since WSTS included actuators from 2003

2003. Growth of 8.5% is forecasted for 2005, followed by virtually zero growth in 2006. In 2007, however, another recovery cycle is expected to begin with market growth in the 10% range.

Clearly, yield and reliability are two of the cornerstones of successful IC manufacturing as they measure semiconductor facility profitability and post-manufacturing device failures. Yield and reliability have played a key role in many aspects of semiconductor operations such as determining the cost of new chips under development, forecasting time-to-market, defining the maximum level of integration possible and estimating the number of wafers to start with. Traditionally, reactive techniques have been used to analyze yield and reliability, and an investigation was launched to determine the cause of yield loss once a low yield was observed during production. Stress testing and failure analysis were commonly performed at the end of the manufacturing line [9.3, 4]. However, as the rapid increase in IC complexity has resulted in multi-billion-dollar semiconductor fabrication facilities, IC manufacturers struggle to obtain a better return on their investment by introducing new process technologies and materials at an accelerated rate to satisfy narrowing market windows. Given this trend, the demand for proactive techniques has strengthened in order to achieve the desired yield and reliability goals early in the process or even before production begins. The demand for these proactive techniques will be even bigger in emerging nanotechnology, which is known to have low yield and reliability [9.5, 6].

Yield and reliability modeling and analysis is a means of achieving proactive yield and reliability management. The purpose of this paper is to review the modeling and analysis of yield and reliability with an additional burn-in step. The importance of yield modeling is emphasized for obtaining better yields quickly after new technologies are introduced. In particular, the relationship between yield, burn-in and reliability will be thoroughly addressed. The relation model between yield and reliability can aid in design for manufacturability (DFM) by improving device layouts for better manufacturing yield and reliability during their early development prior to manufacturing.

9.1 Semiconductor Yield

Yield in semiconductor technology is the most important index for measuring success in the IC business. In general, yield is defined as the fraction of manufactured devices that meet all performance and functionality specifications. Higher yield tends to produce more chips at the same cost, thus allowing prices to decrease.

In this section, we first decompose overall yield into several components. Then, the literature on yield models is reviewed, focusing mainly on the random defect yield model. Traditional Poisson and compound Poisson yield models are thoroughly reviewed as well as some more recent yield models. Finally, issues related to proactive

yield improvement are discussed from the viewpoint of yield modeling.

9.1.1 Components of Semiconductor Yield

The overall yield Y_{overall} of a semiconductor facility can be broken down into several components: wafer process yield Y_{process}, wafer probe yield Y_{probe}, assembly yield Y_{assembly} and final test yield $Y_{\text{final test}}$ [9.7]. Wafer process yield, which is synonymous with line or wafer yield, is the fraction of wafers that complete wafer fabrication. Wafer probe yield is the fraction of chips on yielding wafers that pass the wafer probe test. The terms die yield, chip yield or wafer sort yield are used interchangeably with wafer probe yield. Overall yield is the product of these components, written as

$$Y_{\text{overall}} = Y_{\text{process}} Y_{\text{probe}} Y_{\text{assembly}} Y_{\text{final test}} \,.$$

9.1.2 Components of Wafer Probe Yield

Most semiconductor industries focus on improving the wafer probe yield, which is the bottleneck of overall yield. The importance of wafer probe yield to financial success is discussed in [9.8, 9]. Wafer probe yield is decomposed into functional yield $Y_{\text{functional}}$ and parametric yield $Y_{\text{parametric}}$ such that

$$Y_{\text{probe}} = Y_{\text{functional}} Y_{\text{parametric}} \,.$$

Parametric yield refers to the quantification of IC performance that is caused by process parameter variations. The designer attempts to increase parametric yield using several tools to check the design for process and parameter variations. Commonly used methods include corner analysis, Monte Carlo analysis, and the response surface methodology [9.10]. Corner analysis is the most widely used method due to its simplicity. The designer determines the worst-case corner under which the design can be expected to function. Then, each corner is simulated and the output is examined to ascertain whether or not the design performs as required. The disadvantages of corner analysis include the possibility that a design may function well at the corners but fail in between or that the designer may not know what the corners are. In Monte Carlo analysis, samples are generated to estimate yield based on the distributions of the process parameters. A disadvantage of Monte Carlo analysis is that the designer may not know if an increased yield is due to a change in the design parameters or is due to Monte Carlo sampling error. Another disadvantage is that a complete rerun of the analysis is required if the design variables are changed. With the response surface methodology, a set of polynomial models are created from the design of experiments that approximate the original design. These models are run so many times that the sampling error is reduced to nearly zero. A disadvantage of the response surface methodology arises from errors existing as a result of differences between the polynomial models and the original design.

Functional yield is related to manufacturing problems such as particulate matter, mechanical damage, and crystalline defects which cause dice not to function. Therefore, functional yield is a reflection of the quality of the manufacturing process and is often called the manufacturing yield [9.7, 11] or the catastrophic yield [9.12, 13]. In general, functional yield can be further partitioned into three categories: repeating yield $Y_{\text{repeating}}$, systematic yield $Y_{\text{systematic}}$ and random-defect-limited yield Y_{random} [9.14]:

$$Y_{\text{functional}} = Y_{\text{repeating}} Y_{\text{systematic}} Y_{\text{random}} \,.$$

Repeating yield is limited to reticle defects that occur when there are multiple dies on a reticle. Once reticle defects are identified using a pattern-recognition algorithm, repeating yield is calculated by the ratio of the number of dies without repeating defects to the total number of dies per wafer [9.9]. Then, repeating yield is extracted from functional yield, and tile yield is defined by

$$Y_{\text{tile}} = \frac{Y_{\text{functional}}}{Y_{\text{repeating}}} = Y_{\text{systematic}} Y_{\text{random}} \,.$$

Systematic yield is limited to nonrandom defects affecting every die in some region of the wafer. To decompose Y_{tile} into $Y_{\text{systematic}}$ and Y_{random}, $Y_{\text{systematic}}$ is assumed to be constant regardless of die size since, in a mature process, $Y_{\text{systematic}}$ is known and controllable and is often equal to one. Then, a model is selected to relate Y_{random} to the die area and the density of the random defects, and curve fitting is used with

$$\ln Y_{\text{tile}} = \ln Y_{\text{systematic}} + \ln Y_{\text{random}}$$

to estimate $Y_{\text{systematic}}$ and the parameters of a model for Y_{random} [9.9].

9.1.3 Modeling Random Defect Yield

Since the early 1960s, researchers have devoted extensive work to developing yield models that relate the mean number of random defects in a device to the device yield.

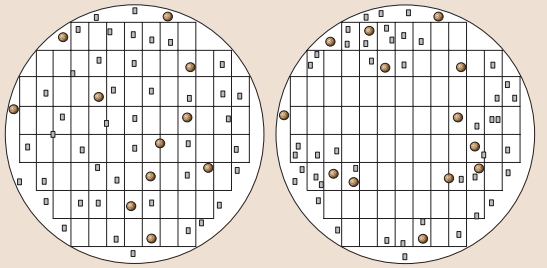

Fig. 9.1 Comparison of defect clustering for the same defect density [9.15]

The basic assumption is that yield is a function of the device area and the average number of yield defects per unit area.

During the manufacturing process, random defects can be introduced at any one of hundreds of process steps. Not all defects necessarily cause device failures. A defect that is of sufficient size and/or occurs in a place that results in an immediate device failure is called a fatal defect [9.16, 17], killer defect [9.18–22] or yield defect [9.23–27]. On the other hand, a defect that is either too small or located in a position that does not cause an immediate failure is called a latent defect, nonfatal defect or reliability defect. In this chapter, we will use the terms yield defect and reliability defect.

Poisson Model

For the purpose of yield modeling, the only yield defects that are of interest are those that can be detected by a manufacturing yield test. Let N_y be the number of yield defects introduced during fabrication on a device of area A. Assuming that the defects are randomly distributed and the occurrence of a defect at any location is independent of the occurrence of any other defect, the probability of a device having k yield defects is calculated by the Poisson probability distribution:

$$P_{\text{Poisson}}(k) = P(N_y = k) = \frac{\mathrm{e}^{-\lambda_y}\lambda_y^k}{k!}\,, \quad k = 0, 1, 2\,, \tag{9.1}$$

where λ_y is the average number of yield defects with $\lambda_y = E(N_y)$. Then, the corresponding Poisson yield is obtained by

$$Y_{\text{Poisson}} = P_{\text{Poisson}}(0) = \mathrm{e}^{-\lambda_y}\,, \tag{9.2}$$

where $\lambda_y = AD_y$, and D_y is the average number of yield defects per unit area.

Compound Poisson Model

Often defects on ICs are not uniformly distributed but tend to cluster. When defects are clustered in certain areas, the Poisson distribution is too pessimistic and the compound Poisson process is used, given by

$$P_{\text{compound}}(k) = P(N_y = k) = \int \frac{\mathrm{e}^{-AD}(AD)^k}{k!} f(D)\mathrm{d}D\,, \quad k = 0, 1, 2\,,$$

where $f(D)$ is the distribution of the defect density. The corresponding yield expression is

$$Y_{\text{compound}} = P_{\text{compound}}(0) = \int \mathrm{e}^{-AD} f(D)\mathrm{d}D\,.$$

Figure 9.1 compares two different degrees of defect clustering for the same average defect density. The left one, with low clustering, belongs more to the Poisson model and the right one, with high clustering, belongs more to the compound Poisson model.

Several distributions such as the symmetric triangle, exponential, uniform, gamma, Weibull and inverse Gaussian have been suggested for $f(D)$ [9.7, 16, 28, 29]. If D follows a uniform distribution in $[0, 2D_y]$, then the corresponding yield can be obtained by

$$Y_{\text{uniform}} = \int_0^{2D_y} \mathrm{e}^{-AD}\frac{1}{2D_y}\mathrm{d}D = \frac{1-\mathrm{e}^{-2\lambda_y}}{2\lambda_y}\,.$$

In the case where D follows a triangle distribution that approximates a normal distribution, the resulting model is called the Murphy's yield and is derived by

$$\begin{aligned} Y_{\text{Murphy's}} &= \int_0^{D_y} \mathrm{e}^{-AD}\frac{D}{D_y^2}\mathrm{d}D \\ &\quad + \int_{D_y}^{2D_y} \mathrm{e}^{-AD}\left(2-\frac{D}{D_y}\right)\frac{1}{D_y}\mathrm{d}D \\ &= \left(\frac{1-\mathrm{e}^{-\lambda_y}}{\lambda_y}\right)^2\,. \end{aligned}$$

For an exponential distribution of D, the model is called the Seed's yield and is given by

$$Y_{\text{Seed's}} = \int_0^{\infty} \mathrm{e}^{-AD}\frac{\mathrm{e}^{-D/D_y}}{D_y}\mathrm{d}D = \frac{1}{1+\lambda_y}\,.$$

For the case of the Weibull distribution, the corresponding yield is [9.29]

$$Y_{\text{Weibull}} = \int_0^\infty \mathrm{e}^{-AD}\frac{\alpha}{\beta^\alpha}D^{\alpha-1}\mathrm{e}^{-(D/\beta)^\alpha}\,\mathrm{d}D = \sum_{k=0}^{\infty}(-1)^k\frac{(AD_y)^k}{k!}\frac{\Gamma(1+k/\alpha)}{\Gamma^k(1+1/\alpha)}\,.$$

Also, if $f(D)$ is the inverse Gaussian distribution, then the yield is [9.29]

$$Y_{\text{inverse-Gaussian}} = \int_0^\infty \mathrm{e}^{-AD}\sqrt{\frac{\phi}{2\pi}}x^{-3/2}\exp\left(-\frac{\phi(D-D_y)^2}{2D_y^2D}\right)\mathrm{d}D = \exp\left\{\phi\left[1-\left(1+\frac{2AD_y}{\phi}\right)^{1/2}\right]\right\}\,.$$

When D follows a gamma distribution, the resulting model is called the negative binomial yield, which is derived as

$$Y_{\text{nb}} = \int_0^\infty \mathrm{e}^{-AD}\frac{1}{\Gamma(\alpha)\beta^\alpha}D^{\alpha-1}\mathrm{e}^{-D/\beta}\,\mathrm{d}D = \left(1+\frac{\lambda_y}{\alpha}\right)^{-\alpha}\,, \tag{9.3}$$

where α is referred to as the clustering factor. A smaller value of α means a higher degree of clustering and greater variation in defect density across the wafer. If $\alpha = 1$, then the negative binomial yield is equivalent to Seed's yield. In the case where $\alpha \to \infty$, the negative binomial yield approaches the Poisson yield. By varying the value of α, the negative binomial yield covers the whole range of yield estimations. *Cunningham* [9.28] reported methods to determine the clustering factor. *Langford* and *Liou* [9.30] presented a new technique to calculate an exact solution of α from wafer probe bin map data.

Critical Area and Defect Size Distribution in Yield Model

The random-defect-limited yield can be more accurately evaluated if the concepts of critical area and defect size distribution are incorporated.

Let $s(x)$ be the probability density function of the defect size. Although the form of $s(x)$ depends on process lines, process time, learning experience gained and other variables, it generally peaks at a critical size and then decreases on either side of the peak [9.31]. Let x_0 be the critical size of the defect that is most likely to occur. The defect size distribution is given by [9.7, 15, 23, 28]

$$s(x) = \begin{cases} cx_0^{-q-1}x^q, & 0 \le x \le x_0 \\ cx_0^{p-1}x^{-p}, & x_0 < x < \infty\,, \end{cases} \tag{9.4}$$

where $p \neq 1, q > 0$ and $c = (q+1)(p-1)/(p+q)$. While p, q and x_0 are process-dependent constants, $q = 1$ and $p = 3$ agree well with the experimental data, and x_0 must be smaller than the minimum width or spacing of the defect monitor [9.7, 23]. A gamma distribution is also used for $s(x)$ in some applications [9.32, 33].

The critical area defines the region of the layout where a defect must fall to cause device failure. Therefore, if a defect occurs in the critical area, then it becomes a yield defect. Given $s(x)$, the yield critical area is expressed by

$$A_y = \int_0^\infty A_y(x)s(x)\,\mathrm{d}x\,,$$

where $A_y(x)$ is a critical area of defect size x. Then $\lambda_y = A_yD_0$ is used in yield models where D_0 is the average defect density of all sizes. The geometric method [9.34], the Monte Carlo method [9.35] and the pattern-oriented method [9.36] have been used for critical area extraction. Critical area analysis can be used to quantify the sensitivity of a design to defects based on the layout [9.37] and can aid DFM by improving layouts for better yield [9.38].

Other Models

Sato et al. [9.39] used a discrete exponential distribution for the number of yield defects for each type of defect:

$$P_{\text{discrete expo}}(k) = (1-\mathrm{e}^{-h})\mathrm{e}^{-hk}\,, \tag{9.5}$$

where h is the parameter. Once the probability density function for m types of defects is derived by m convolution of (9.5), the yield is derived by

$$Y_{\text{discrete expo}} = (1-\mathrm{e}^{-h})^m = (1+A_yD_0/m)^{-m}\,. \tag{9.6}$$

Park and *Jun* [9.40] presented another yield model based on a generalized Poisson distribution. Assuming that the number of defect clusters in a chip and the number of defects in each cluster follows a Poisson distribution, the total number of defects in a chip follows a generalized Poisson distribution. Then yield is calculated using the

fact that the total number of yield defects in a chip is the sum of the yield defects in each cluster if the probability of each defect in the cluster becoming a yield defect is the same.

Jun et al. [9.41] developed a yield model through regression analysis in which the mean number of defects per chip and a new cluster index obtained from the defect location are used as independent variables. Simulation results showed that yield is higher for the higher index, but the rate of yield growth decreases as the cluster index increases.

Carrasco and *Suñé* [9.42] presented a methodology for estimating yield for fault-tolerant systems-on-chip, assuming that the system fails with probability $1 - C_i$ if component i fails. The system failure probability is independent of the subsets of components which failed before. For each component, an upper bound on the yield loss is obtained, which can be formalized as the probability that a Boolean function of certain independent integer-valued random variables is equal to one. The reduced-order multiple-valued decision diagram is used to compute the probability.

Noting that interconnect substrates face low yield and high cost, *Scheffler* et al. [9.43] presented a yield estimation approach to assess the impact on overall substrate cost of changing design rules. Given a defect size distribution, if a design rule is relaxed, for instance, if the line width and line spacing are widened, the total number of yield defects decreases and the critical area increases. From the limitation of applications to interconnect substrates, the critical area can be obtained by the union of the critical area for line shorts and the critical area for line opens. Then, the Poisson yield is expressed as a function of line width, and trade-offs of the design rule change can be studied. If an increase in the design rule has minimal impact on the overall substrate area, then yield improvement by increasing the design rules can lead to a more cost-effective substrate.

Cunningham et al. [9.44] presented a common-yield model to analyze and compare the yield of products from different facilities using a linear regression model. *Berglund* [9.45] developed a variable defect size yield model. *Milchalka* [9.17] presented a yield model that considers the repair capability in a part of the die area. *Stapper* and *Rosner* [9.37] presented a yield model using the number of circuits and average number of yield defects per circuit. *Dance* and *Jarvis* [9.46] explained the application of yield models to accelerate yield learning and to develop a performance–price improvement strategy.

Choosing a yield model is basically an experiential process. IC manufacturers compare data from a specific process for yield versus die size using various models and select the best fit. Depending on the distribution of die sizes of a given product and the distribution pattern of the defects, different yield models will best fit the data [9.47].

9.1.4 Issues for Yield Improvement

Achieving high-yield devices is a very challenging task due to reduced process margins and increased IC design complexity. Recent research has emphasized the role of parametric yield loss as well as that of functional yield loss in proactive yield management. Although random yield loss typically dominates in high-volume production, systematic and parametric yield losses become more important when a fabrication process is newly defined and is being tuned to achieve the necessary processes and device parameters [9.48]. Considerable attention has been paid thus far to improving random yield, but relatively little attention has been paid to systematic and parametric yield problems. With new technologies, a process may never be stabilized and statistical device-parameter variations will be a big headache. Traditionally, parametric yield problems were addressed after a design was manufactured. Low-yielding wafers were investigated to identify what process variations caused the yield loss. Then, simulations were used to see where the design should be changed to improve the yield. The traditional redesign approach is very costly compared to handling design at the front-end of the design process using design for yield (DFY) techniques. The use of DFY techniques accelerates the design flow, reduces cycle times and provides higher yield.

Before a high-volume chip comes to market, it must be manufacturable at an acceptable yield. Although traditionally yield issues have been in the domain of manufacturing teams, a new approach to bridge the gap between design and manufacture is necessary as chip geometry shrinks. *Peters* [9.49] emphasized the increasing role of DFY approaches in leading-edge device manufacturability to allow for tuning of all test programs and models so that design, manufacturing and testing provide high-yielding devices.

Li et al. [9.48] presented a holistic yield-improvement methodology that integrates process recipe and design information with in-line manufacturing data to solve the process and design architecture issues that

affect yield and performance. The approach suggests improving yield not just by eliminating defects but also by resolving parametric problems.

Nardi and *Sangiovanni-Vincentelli* [9.12] observed that for complex nanodesigns functional yield might depend more on the design attributes than on the total chip area. Given that the current yield-aware flow optimizes yield at the layout level after optimizing speed and area, a synthesis-for-manufacturability approach is suggested in which manufacturability replaces area in the cost function.

Segal [9.38] claimed that each new technology generation will see lower and lower yields if the defect level of well-running processes is not reduced. A strategy to reduce the defect level is to encompass techniques for responding quickly to defect excursion using in-line wafer scanners and wafer position tracking. The defect excursion strategy eliminates wafers and lots with very high defect densities.

9.2 Semiconductor Reliability

Once an IC device is released to the user, an important and standard measure of device performance is reliability, which is defined as the probability of a device conforming to its specifications over a specified period of time under specified conditions. A failure rate function is usually used to describe device reliability, which is defined for a population of nonrepairable devices as the instantaneous rate of failure for the surviving devices during the next instant of time. If $h(x)$ denotes a failure rate function, the corresponding reliability function is expressed by

$$R(t) = \mathrm{e}^{-\int_0^t h(x)\,\mathrm{d}x} .$$

In this section, the failure rate in semiconductor device reliability is explained. Then, we discuss where each semiconductor failure mechanism occurs in the bathtub failure rate. Finally, techniques used for reliability improvement are reviewed.

9.2.1 Bathtub Failure Rate

When engineers have calculated the failure rate of a semiconductor population over many years, they have commonly observed that the failure rate is described by a bathtub shape.

Initially, semiconductor devices show a high failure rate, resulting in an infant mortality period. The infant mortality period results from weak devices that have shorter lifetimes than the normal stronger devices, implying that infant mortality period applies to a whole population rather than a single device. The operating period that follows the infant mortality period has a lower, and almost constant, failure rate and is called the useful life period. Infant mortality and useful life failures are due to defects introduced during the manufacturing process, such as particle defects, etch defects, scratches and package assembly defects.

A device that has reached the end of its useful life enters the final phase called the aging period. Failures during the aging period are typically due to aging or cumulative damage, and these can be avoided by careful technology development and product design. These failures are inherent process limitations and are generally well-characterized.

The semiconductor manufacturing process requires hundreds of sequential steps and thus hundreds, or even thousands, of process variables must be strictly controlled to maintain the device reliability. Despite the exponential scaling of semiconductor size and chip complexity, IC reliability has increased at an even faster rate as reliability engineers reduce infant mortality and useful life failure rate and push the aging period beyond the typical usage period through a variety of reliability improvement techniques.

9.2.2 Occurrence of Failure Mechanisms in the Bathtub Failure Rate

Failure mechanisms of semiconductor devices can be classified into three groups: electrical stress failures, intrinsic failures and extrinsic failures [9.7, 50].

Electrical stress failures are user-related, and the major causes are electrical-over-stress (EOS) and electrostatic discharge (ESD) due to improper handling. ESD and EOS problems are thoroughly discussed in *Vinson* and *Liou* [9.51]. Because this failure mechanism is event-related, it can occur anywhere in the infant mortality period, the useful life period or the aging period.

The intrinsic failure mechanism results from all crystal-related defects, and thus it occurs predominantly in the infant mortality period but rarely in the aging period.

On the other hand, extrinsic failures are the result of device packaging, metallization and radiation and they can occur any time over the device's lifetime. Extrinsic failures that are due to process deficiencies, such as migration and microcracks, occur during the infant mortality period. The extrinsic failure mechanisms related to packaging deficiency, such as bond looping and metal degradation, occur in the aging period. The radiation-related extrinsic failure mechanisms, such as bit flips due to external radiation, occur continuously over the device lifetime [9.50].

The terms extrinsic and intrinsic failure have also been used in different contexts. First, intrinsic failure is used to describe those failures that are due to internal causes of a device, while failures due to forces external to the product, such as mishandling or accidents, are called extrinsic failures [9.52]. In this case, intrinsic failures occur in the infant mortality period or in the aging period, while extrinsic failures occur in the useful life period. Secondly, the terms intrinsic and extrinsic failure are used to classify oxide failures [9.53]. In this case, intrinsic failures are due to the breakdown of oxide which is free of manufacturing defects, and thus is usually caused by an inherent imperfection in the dielectric material. These failures occur in the aging period at an increasing failure rate. On the other hand, extrinsic failures that result from process defects in the oxide or problems in the oxide fabrication occur in the infant mortality period.

9.2.3 Issues for Reliability Improvement

As reliability engineers have recognized that it is no longer affordable to handle reliability assurance as a back-end process in IC product development, the reliability emphasis has been shifted from end-of-line statistical-based stress testing to new proactive techniques such as design for reliability (DFR), built-in reliability (BIR), wafer-level reliability (WLR), qualified manufacturing line (QML), and physics-of-failure (POF) approaches [9.54, 55].

DFR means building reliability into the design rather than incorporating it after development [9.56]. The importance of DFR increases as stress testing becomes increasingly difficult as the allowable stress is decreased.

The effectiveness of BIR has been outlined in [9.57, 58] for manufacturing highly reliable ICs through the elimination of all possible defects in the design stage.

WLR represents a transition from the end-of-line concept toward the concept of BIR, because the testing is performed at the wafer level reducing the time and expense of packaging. Examples of WLR implementation into a production line or a testing method are given in [9.59–62].

QML is another evolutionary step devised for the purpose of developing new technologies where the manufacturing line is characterized by running test circuits and standard circuit types [9.63]. Understanding failure mechanisms and performing failure analysis are critical elements in implementing the BIR and QML concept.

In cases where the fundamental mechanical, electrical, chemical, and thermal mechanisms related to failures are known, it is possible to prevent failures in new products before they occur. This is the basic idea of POF, which is the process of focusing on the root causes of failure during product design and development in order to provide timely feedback.

9.3 Burn-In

Burn-in is a production process that operates devices, often under accelerated environments, so as to detect and remove weak devices containing manufacturing defects before they are sold or incorporated into assemblies. Because the design rules change so quickly, burn-in today is an essential part of the assembly and testing of virtually all semiconductor devices. To burn-in or not to burn-in and how long the burn-in should be continued are perennial questions.

In this section, we discuss several issues related to burn-in, such as key questions for burn-in effectiveness, burn-in level and burn-in types. Then, the previous burn-in literature is reviewed based on the level of burn-in application.

9.3.1 The Need for Burn-In

Since most semiconductor devices ordinarily have an infant mortality period, the reliability problem during this period becomes extremely important. Manufacturers use burn-in tests to remove infant mortality failures for most circuits, especially where high reliability is a must. Burn-in ensures that a circuit at assembly has moved to the

useful life period of the bathtub curve. During burn-in, elevated voltage and temperature are often combined to activate the voltage- and temperature-dependent failure mechanisms for a particular device in a short time. Careful attention to design of stress burn-in is necessary to ensure that the defect mechanism responsible for infant mortality failures is accelerated while normal strong devices remain unaffected.

Although burn-in is beneficial for screening in the infant mortality period, the burn-in cost ranges from 5–40% of the total device cost depending on the burn-in time, quantities of ICs and device complexity [9.64], and it might introduce additional failures due to EOS, ESD or handling problems. Solutions to the key questions posed by *Kuo* and *Kuo* [9.65] will continue to be found with new technologies for exercising burn-in effectively:

1. How much should infant mortality be reduced by burn-in?
2. Under what environmental conditions should burn-in be performed?
3. Should burn-in be accomplished at the system, subsystem, or component level?
4. Who should be in charge of burn-in, the vender, the buyer, or a third party?
5. Are there any side-effects of burn-in?
6. How will the industry benefit from burn-in data?
7. What physics laws should be followed to conduct burn-in?

9.3.2 Levels of Burn-In

There are three burn-in types based on levels of a device: package-level burn-in (PLBI), die-level burn-in (DLBI), and wafer-level burn-in (WLBI) [9.66–68].

PLBI is the conventional burn-in technology where dies are packed into the final packages and then subjected to burn-in. Although PLBI has the advantage of assuring the reliability of the final product, repairing or discarding a product after PLBI is far too costly.

The strong demand for known good dies (KGD) has motivated the development of more efficient burn-in technology. Generally, KGD is defined as a bare unpacked die that has been tested and verified as fully functional to meet the full range of device specifications at a certain level of reliability [9.68, 69]. KGD enables manufacturers to guarantee a given quality and reliability level per die before integration and assembly. Optimizing burn-in is a key aspect of KGD [9.69].

In DLBI, dies are placed in temporary carriers before being packed into their final form to reduce the cost of added packaging. DLBI and testing of the individual die before packaging ensures that only KGD are packaged and thus produces a quality product at a reduced cost.

Considerations of how to reduce burn-in cost and solve KGD issues have led to the concept of WLBI. WLBI achieves burn-in on the wafer as soon as it leaves the fab. Though WLBI can result in less-reliable final products than PLBI, the trend in industry is to do more testing at the wafer level due to the cost and KGD issues [9.70].

Recently, the line between burn-in and testing has begun to blur as far as reducing testing costs and cycle times. For example, some test functions have moved to the burn-in stage and multi-temperature environments have moved to final testing. DLBI and WLBI that have evolved from burn-in to include testing are called die-level burn-in and testing (DLBT) and wafer-level burn-in and testing (WLBT), respectively. It is reported that DLBT is an expensive step in memory production and the transfer to WLBT can reduce the overall back-end cost by 50% [9.71].

9.3.3 Types of Burn-In

A basic burn-in system includes burn-in sockets to provide a temporary electrical connection between the burn-in board (BIB) and the device under test (DUT) package. Each BIB might accommodate 50 or more sockets, and a burn-in system might hold 32 BIBs. To develop a successful burn-in strategy, detailed knowledge is necessary about temperature distributions across a DUT package, across a BIB, and throughout the burn-in oven [9.72].

Three burn-in types are known to be effective for semiconductor devices: steady-state or static burn-in (SBI), dynamic burn-in (DBI) and test during burn-in (TDBI) [9.7, 73].

In SBI, DUTs are loaded into the burn-in boards (BIB) sockets, the BIBs are put in the burn-in ovens and the burn-in system applies power and an elevated temperature condition (125–150 °C) to the devices for a period ranging from 12 to 24 h. Once the devices cool down, the BIBs are extracted from the boards. These devices are placed in handling tubes and mounted on a single-device tester. Functional tests are then applied on the devices to sort them according to failure types. Because the DUT is powered but not exercised electrically, SBI may not be useful for complex devices because external biases and loads may not stress internal nodes.

In DBI, the DUT is stimulated at a maximum rate determined by the burn-in oven electronics, which can

propagate to internal nodes. Neither SBI not DBI monitors the DUT response during the stress, and thus dies that fail burn-in cannot be detected until a subsequent functional test.

Beyond static and dynamic burn-in is so-called intelligent burn-in [9.72]. Intelligent burn-in systems not only apply power and signals to DUTs, they also monitor DUT outputs. Therefore, they can guarantee that devices undergoing burn-in are indeed powered up and that input test vectors are being applied. In addition, they can perform some test functions. TDBI is a technique for applying test vectors to devices while they are being subjected to stresses as part of the burn-in process. Though function testing is not possible due to the burn-in stress, idle time can be used advantageously to verify circuit integrity, permitting abbreviated functional testing after burn-in.

9.3.4 Review of Optimal Burn-In Literature

While a considerable number of papers have dealt with burn-in at one level, recent research has been directed to the study of burn-in at multiple levels. In this section, we will review the burn-in literature based on the burn-in level being analyzed.

One-Level Burn-In

To fix burn-in at one level, previous work has taken two different approaches: the black-box approach and the white-box approach. In the black-box approach, each device is treated as a black box and a specific failure rate distribution is assumed for the device. In the white-box approach, the device is decomposed into smaller components and a failure rate distribution is assumed for each component. Then the whole-device failure rate is obtained from the structure function and component failure rate.

Many papers have taken the black-box approach and determined the optimal burn-in time to minimize a cost function. *Mi* [9.74, 75] showed that optimal burn-in times that minimize various cost functions occur in the infant mortality period. *Sheu* and *Chien* [9.76] showed the same result for two different types of failures. Assuming that the device lifetime follows a Weibull distribution, *Drapella* and *Kosznik* [9.77] obtained optimal burn-in and preventive replacement periods using Mathcad code. *Cha* [9.78–80] considered a minimally repaired device and derived the properties of optimal burn-in time and block replacement policy. *Tseng* and *Tang* [9.81] developed a decision rule for classifying a component as strong or weak and an economical model to determine burn-in parameters based on a Wiener process. Assuming a mixed Weibull distribution, *Kim* [9.82] determined optimal burn-in time with multiple objectives of minimizing cost and maximizing reliability. A nonparametric approach [9.83] and a nonparametric Bayesian approach [9.84] have been used to estimate the optimal system burn-in time that minimizes a cost function.

The first report that takes a white-box approach appears in *Kuo* [9.85]. The optimal component burn-in time was determined to minimize a cost function subject to a reliability constraint, assuming that the failure of each component follows a Weibull distribution. *Chi* and *Kuo* [9.86] extended it to include a burn-in capacity constraint. *Kar* and *Nachlas* [9.87] consider a series structure, assuming that each component has a Weibull distribution. Given that each component that fails system burn-in is replaced, the optimal system burn-in time was determined to maximize a net-profit function that balances revenue and cost. For the case where percentile residual life is the performance measure of burn-in, *Kim* and *Kuo* [9.88] studied the relationship between burn-in and percentile residual life.

Multi-level Burn-in

For studying burn-in at various levels, the white-box approach must be asked to characterize the failure time distribution of the whole device. Because system burn-in is never necessary after component burn-in if assembly is perfect [9.89, 90], modeling of burn-in at multiple levels must focus on the quantification of assembly quality. *Whitbeck* and *Leemis* [9.91] added a pseudo-component in series to model the degradation of a parallel system during assembly. Their simulation result showed that system burn-in is necessary after component burn-in to maximize the mean residual life. *Reddy* and *Dietrich* [9.92] added several connections to explain an assembly process and assumed that each of components and connections followed a mixed exponential distribution. The optimal burn-in time at the component and system levels were determined numerically to minimize the cost functions, given that the components were replaced and the connections minimally repaired upon failure. *Pohl* and *Dietrich* [9.93] considered the same problem for mixed Weibull distributions. *Kuo* [9.94] used the term *incompatibility* for reliability reduction realized during assembly process. The incompatibility factor exists not only at the component level but also at the subsystem and the system levels due to poor manufacturability, workmanship, and design strategy. *Chien* and *Kuo* [9.95] proposed a nonlinear model

to estimate the optimal burn-in times for all levels as well as to determine the number of redundancies in each subsystem when incompatibility exists. To quantify the incompatibility factor, *Chien* and *Kuo* [9.96] added a uniform random variable to the reliability function. Optimal burn-in times at different levels were determined to maximize the system reliability, subject to a cost constraint via simulation, assuming that the component followed a Weibull distribution. A conceptual model has been developed [9.97] that considers PLBI and WLBI for minimizing a cost function subject to the reliability requirement. *Kim* and *Kuo* [9.98, 99] analytically derived the conditions for system burn-in to be performed after component burn-in using a general system distribution to which the component burn-in information and assembly problems were transferred. *Kim* and *Kuo* [9.100] presented another model for quantifying the incompatibility factor when the assembly adversely affected the components that were replaced at failure. Optimal component and system burn-in times were determined using nonlinear programming for various criteria.

9.4 Relationships Between Yield, Burn-In and Reliability

As semiconductor technology advances, burn-in is becoming more expensive, time-consuming and less capable of identifying the failure causes. Previous research has focused on the determination of burn-in time based on a reliability function estimated from the time-to-first-failure distribution. However, newer proactive methods to determine the burn-in period in the early production stage are of great interest to the semiconductor industry.

One such approach is based on the relation model of yield, burn-in and reliability, which we will review in this section.

9.4.1 Background

Observing that high yield tends to go with high reliability, it was conjectured that defects created on IC devices during manufacturing processes determine yield as well as reliability [9.31]. Subsequent experiments confirmed that each defect in a device affects either yield or reliability, depending on its size and location. This is illustrated in Fig. 9.2 for oxide defects. Therefore, reliability can be estimated based on yield if the relationship between yield and reliability is identified. A model that relates yield and reliability has many applications, such as in yield and reliability predictions for future devices, device architecture design, process control and specification of allowable defect density in new processes for achieving future yield and reliability goals. As a result, the start-up time of new fabrication facilities and cycle times can be shortened by reducing the amount of traditional stress testing required to qualify new processes and products.

Developing a relation model of yield and reliability has been an active research area in the past decade. Three different definitions have been used for reliability in previous research. First, reliability is defined by the probability of a device having no reliability defects, where a reliability defect is defined not as a function of the operating time but as a fixed defect size. Secondly,

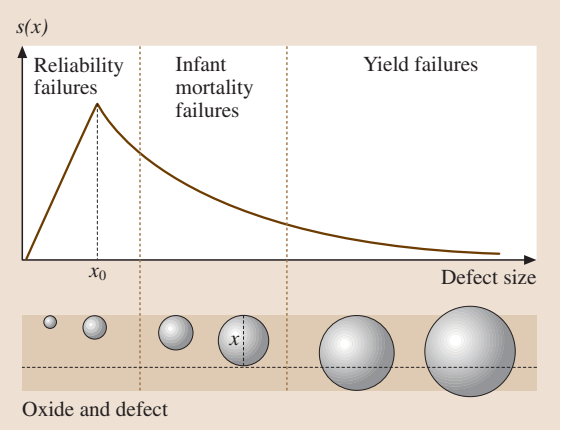

Fig. 9.2 Defect size distribution and oxide problems [9.15]

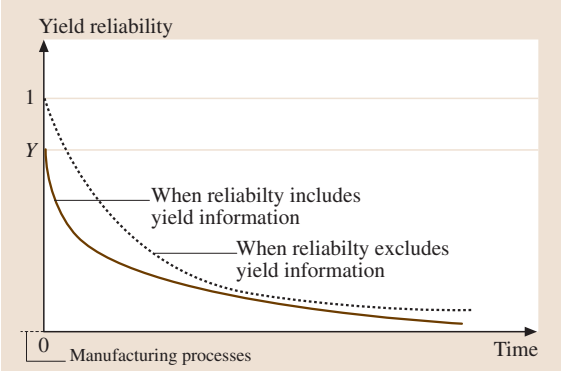

Fig. 9.3 Yield-reliability relationship depending on the definition of reliability [9.24]

reliability denotes the probability of a device having no reliability defects given that there are no yield defects. This reliability is equivalent to yield at time zero, as depicted in Fig. 9.3. This definition of reliability is useful from the designer's point-of-view as it incorporates yield information, but it is not consistent with the traditional definition. Thirdly, reliability is defined as the probability of a device having no reliability defects by time t. In this case, reliability is defined as a function of the operating time without incorporating the yield defects, after assuming that any device that is released to field operation has passed a manufacturing yield test, implying that no yield defects exist. Such a reliability is always 1 at time zero.

9.4.2 Time-Independent Reliability without Yield Information

The first model that related yield and reliability was reported by *Huston* and *Clarke* [9.23]. Let N_y be the number of yield defects and N_r be the number of reliability defects in a device. Reliability defects were defined as a specific defect size rather than as a function of time. Assume that N_y and N_r are independent and each follows a Poisson distribution. Thus, the distribution of N_y is given in (9.1) and the distribution of N_r is given by

$$P^*_{\text{Poisson}}(k) = P(N_r = k) = \frac{e^{-\lambda_r}\lambda_r^k}{k!}, \quad k = 0, 1, 2,$$

where λ_r is the average number of reliability defects per chip. Then, for a device without reliability defects, the Poisson reliability model is obtained by

$$R_{\text{Poisson}} = P^*_{\text{Poisson}}(0) = e^{-\lambda_r}. \tag{9.7}$$

The Poisson yield–reliability relation is obtained from (9.2) and (9.7) by

$$R_{\text{Poisson}} = Y^{\gamma}_{\text{Poisson}} \tag{9.8}$$

where

$$\gamma = \frac{\lambda_r}{\lambda_y}. \tag{9.9}$$

Next, they expressed $\lambda_y = A_y D_0$ and $\lambda_r = A_r D_0$ where D_0 is the common defect density for the yield and reliability defects, and A_y and A_r are the yield and reliability critical areas, respectively.

Subsequently, *Kuper* et al. [9.101] used a similar model given by

$$R_{\text{Poisson}} = (Y_{\text{Poisson}}/M)^{\gamma} \tag{9.10}$$

where M is the maximum possible yield fraction considering clustering effects and edge exclusions. The value of γ depends on the technology and process and on the conditions under which the product is used. They assumed that $\lambda_y = AD_y$ and $\lambda_r = AD_r$ where A is the device area and D_y and D_r are the yield and reliability defect density, respectively. The model was verified with high-volume ICs manufactured by several processes.

Riordan et al. [9.27] verified that (9.10) agrees well for yields based on the lot, the wafer, the region of the wafer and the die in a one-million-unit sample of microprocessors. *Van der Pol* et al. [9.102] used (9.10) to study the IC yield and reliability relationship further for 50 million high-volume products in bipolar CMOS and BICMOS technologies from different wafer fabrication facilities. Experiments showed that a clear correlation exists among functional yield, burn-in failures and field failures.

Zhao et al. [9.103] used a discrete exponential yield model given in (9.6) for yield and (9.7) for reliability. Then, the relation model is obtained by

$$R = \exp\left(-\frac{m\left(1 - Y^{1/m}_{\text{discrete expo}}\right)}{Y^{1/m}_{\text{discrete expo}}}\gamma\right).$$

9.4.3 Time-Independent Reliability with Yield Information

Barnett et al. [9.19] developed a relation model for the negative binomial model, rather than for the Poisson model, assuming that the number of reliability defects is proportional to the number of yield defects in a device. Let N be the total number of defects, where $N = N_y + N_r$. Then

$$P(N_y = m, N_r = n | N = q) = \binom{q}{m} p_y^m p_r^n, \tag{9.11}$$

where p_y is the probability of a defect being a yield defect, and $p_r = 1 - p_y$) is the probability of a defect being a reliability defect. Let $\lambda = E(N)$. If N is assumed to follow a negative binomial distribution

$$P(N = q) = \frac{\Gamma(\alpha + q)}{q!\Gamma(\alpha)} \frac{\left(\frac{\lambda}{\alpha}\right)^q}{(1 + \frac{\lambda}{\alpha})^{\alpha+q}},$$

then the wafer probe yield can be obtained by

$$Y_{\text{nb}} = P(N_y = 0) = \left(1 + \frac{\lambda_y}{\alpha}\right)^{-\alpha}, \tag{9.12}$$

where $\lambda_y = \lambda p_y$ is the average number of yield defects. Let R be the conditional probability that there are no

reliability defects given that there are no yield defects. Then,

$$R = P(N_r = 0 | N_y = 0) = \left(1 + \frac{\lambda_r(0)}{\alpha}\right)^{-\alpha} , \quad (9.13)$$

where

$$\lambda_r(0) = \frac{\lambda p_r}{1 + \lambda p_y / \alpha}$$

is the average number of reliability defects given that there are no yield defects. Using (9.12) and (9.13), the relation model is derived as

$$R = \left[1 + \gamma \left(1 - Y^{1/\alpha}\right)\right]^{-\alpha} ,$$

where $\gamma = \lambda_r / \lambda_y = p_r / p_y$. Numerical examples were used to show that the number of reliability failures predicted by the negative binomial model can differ from the prediction by the Poisson model because of clustering effects. *Barnett* et al. [9.18] modified the model in order to consider the possibility of repair in a certain area of a chip and experimentally verified that the reliability of an IC with a given number of repairs can be accurately quantified with the model. *Barnett* et al. [9.21] and [9.20] validated the yield–reliability relation model using yield and stress test data from a 36-Mbit static random-access memory (SRAM) memory chip and an 8-Mbit embedded dynamic random-access memory (DRAM) chip and from 77 000 microprocessor units manufactured by IBM microelectronics, respectively.

9.4.4 Time-Dependent Reliability

Van der Pol et al. [9.104] added the time aspect of reliability to their previous model [9.102] to suggest detailed burn-in. From an experiment, a combination of two Weibull distributions was employed for the time-to-failure distribution by which $1 - R_{\text{Poisson}}$ in (9.8) is replaced. Similarly, *Barnett* and *Singh* [9.22] introduced the time aspect of reliability in (9.13) using a Weibull distribution. *Forbes* and *Arguello* [9.105] expressed the reliability by time t by

$$\begin{aligned} R(t) &= 1 - e^{-\lambda_r(t)} \simeq 1 - \lambda_r(t) \\ &= 1 - A D_r(t) = 1 - A D_y \gamma(t) , \end{aligned} \quad (9.14)$$

where $\gamma(t) = \frac{D_r(t)}{D_y}$. Then, the Weibull distribution reliability replaces the left-hand side of (9.14) and the corresponding relationship of yield and reliability is used to optimize the burn-in period. All of these models are based on the assumption that the device time-dependent reliability is available in advance from experiments or field failure data.

Kim and *Kuo* [9.26] suggested using $\lambda_r(t) = A_r(t) D_0$ in (9.8), where $\lambda_r(t)$ denotes the mean number of reliability defects realized by time t, and $A_r(t)$ is the reliability critical area by time t. Assuming that the defect growth for operation time t is a known increasing function of time, they calculated $\lambda_r(t)$ and derived a relation model of oxide yield and time-dependent reliability. This is the first model in which time-dependent reliability is estimated from yield and critical area analysis, rather than from field failure data. Because of the properties of the assumed defect growth function, the resulting reliability function has an increasing failure rate. The effect of burn-in on yield, using yield and reliability critical area, was studied by *Kim* et al. [9.106]. *Kim* et al. [9.24] presented another model to tie oxide yield to time-dependent reliability by combining the oxide time to a breakdown model with the defect size distribution given in (9.4). This reliability model predicted from the yield has an infant mortality period such that the optimal burn-in policy for burn-in temperature, burn-in voltage and burn-in time can be determined based on the model.

To handle the dependence between the numbers of yield and reliability defects, *Kim* and *Kuo* [9.25, 107] used a multinomial distribution for the number of yield defects, the number of reliability defects that fail during burn-in and the number of reliability defects that are eventually released to field operation. The distribution of the number of defects is arbitrary. From a feature of multinomial distribution, the number of yield defects

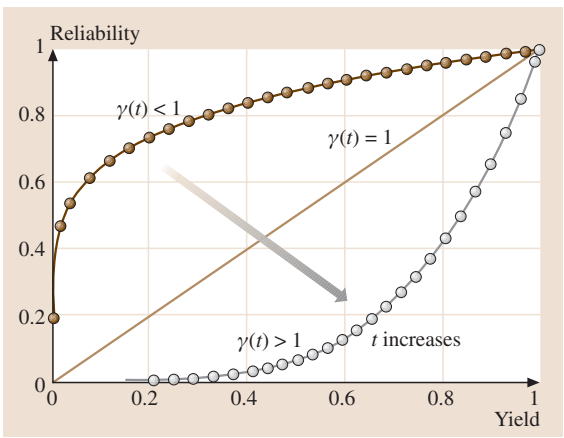

Fig. 9.4 Relation between yield and time-dependent reliability [9.25]

and the number of reliability defects are negatively correlated if the total number of defects in a device is fixed. An analytical result showed that two events, the number of yield defects being zero and the number of reliability defects that fail during burn-in being zero, are positively correlated. This explains the correlated improvement between yield and burn-in fallout. It was also shown that burn-in may be useful if device-to-device variability in the number of defects passing yield tests is greater than a threshold, where the threshold depends on the failure rate of a defect occurrence distribution and the number of defects remaining after the test. Let $\gamma(t)$ be the scaling factor from yield to reliability such that

$$\gamma(t) = \frac{\lambda_r(t)}{\lambda_y} ,$$

where $\lambda_r(t)$ is the number of reliability defects failed by time t. Figure 9.4 shows that a larger value of the scaling factor gives a smaller value of reliability for a given yield value. Clearly, burn-in, reliability and warranty cost can be controlled in an actual process by considering yield and the scaling factor. One can conjecture that burn-in should be performed if the scaling factor is large.

9.5 Conclusions and Future Research

In this chapter, we reviewed semiconductor yield, burn-in and reliability modeling and analysis as a fundamental means of proactive yield and reliability management. It was emphasized that with new technologies the consideration of parametric and systematic yield loss is increasingly important in addition to the consideration of yield defects. Therefore, developing a robust design methodology that can be used to improve parametric and systematic yield becomes a promising research area. Statistical softwares for easily implementing the response surface methodology and Monte Carlo simulation are necessary to overcome the limitations of the current corner analysis method that is widely used in parametric yield analysis.

As design rules tend to change quickly, whether or not to perform burn-in is a perennial question. Previously, a considerable number of papers have studied ways to determine optimal burn-in times based on time-to-first-failure distributions, such as the Weibull distribution or the mixed Weibull distribution. Since burn-in is expensive and time-consuming, more proactive approaches are necessary for determining optimal burn-in time, for example POF analysis.

As correlated improvements in yield, burn-in failures and reliability have occurred, the development of a model relating them has been an active research area in the last decade. Such a model is a prerequisite to predict and control burn-in and reliability based on the device layout in the design stage. Through the model, cycle times and testing costs can be reduced significantly. Currently, experiments are validating the relationship between yield and time-independent reliability. Experiments are necessary to confirm the time-dependent relationship as well. Validation of the time-dependent behavior of reliability defects using IC devices is necessary to determine optimal burn-in periods through the relation model. To do this, physical models must be available to characterize defect growth during operation for various device types, which will enable the estimation of reliability defects as a function of operation time. Also, some future research should be conducted to generalize the yield–reliability relation model to other defect density distributions besides the Poisson and negative binomial models.

References

9.1 Semiconductor Industry Association: *The National Technology Roadmap for Semiconductors* (Semiconductor Industry, San Jose, CA 2003)

9.2 World Semiconductor Trade Statistics: *Semiconductor Market Forecast* (Miyazaki, Japan May 2004)

9.3 W. T. K. Chien, W. Kuo: Use of the Dirichlet process for reliability analysis, Comput. Ind. Eng. **27**, 339–343 (1994)

9.4 W. T. K. Chien, W. Kuo: Extensions of Kaplan–Meier estimator, Commun. Stat. Simul. Comp. **24**, 953–964 (1995)

9.5 National Research Council Committee: *Implications of Emerging Micro- and Nano-technologies*, National Research Council Committee, Division of Engineering and Physical Sciences (National Academy, Washington 2003) p. 251

9.6 W. Luo, Y. Kuo, W. Kuo: Dielectric relaxation and breakdown detection of doped tantalum oxide high-k thin films, IEEE Trans. Dev. Mater. Reliab. **4**, 488–494 (2004)

9.7 W. Kuo, W. T. K. Chien, T. Kim: *Reliability, Yield, and Stress Burn-in* (Kluwer Academic, Norwell 1998)

9.8 C. J. McDonald: New tools for yield improvement in integrated circuit manufacturing: can they be applied to reliability, Microelectron. Reliab. **39**, 731–739 (1999)

9.9 L. Peters: Introduction to Integrated Yield Analysis, Semicond. Int. **22**, 56 (January 1999)

9.10 http://www.cadence.com/whitepapers /5047ParametricYieldWPfnl.pdf

9.11 R. K. Ravikumar: Viewpoint addressing design yield prior to silicon, Chip Des. Mag. **40** (April/May 2004)

9.12 A. M. Nardi, A. Sangiovanni-Vincentelli: Logic synthesis for manufacturability, IEEE Des. Test Comput. **21**(3), 192–199 (2004)

9.13 A. J. Strojwas: *Yield of VLSI Circuits: myths vs. reality*, 23th Design Automation Conference (IEEE Computer Society, Las Vegas, NV 1986) pp. 234–235

9.14 C. H. Stapper: Large area fault clusters and fault tolerance in VLSI circuits: a review, IBM J. Res. Develop. **33**, 174–177 (1989)

9.15 W. Kuo, T. H. Kim: An overview of manufacturing yield and reliability modeling for semiconductor products, Proc. IEEE **87**, 1329–1344 (1999)

9.16 A. V. Ferris-Prabhu: *Introduction to Semiconductor Device Yield Modeling* (Artech House, Boston 1992)

9.17 T. L. Michalka, R. C. Varshney, J. D. Meindl: A discussion of yield modeling with defect clustering, circuit repair and circuit redundancy, IEEE Trans. Semicond. Manuf. **3**, 116–127 (1990)

9.18 T. S. Barnett, A. D. Singh, V. P. Nelson: *Estimating burn-in fall-out for redundacy memory*, Proc. 2001 VLSI Test Symp. (IEEE Computer Society, Los Angeles, CA May 2001) pp. 340–347

9.19 T. S. Barnett, A. D. Singh, V. P. Nelson: Extending integrated-circuit yield models to estimate early life reliability, IEEE Trans. Reliab. **52**, 296–301 (2003)

9.20 T. S. Barnett, A. D. Singh, M. Grady, V. P. Nelson: *Yield-reliability modeling: experimental verification and application to burn-in reduction*, Proc. 20th IEEE VLSI Test Symp. (IEEE Computer Society, Monterey, CA 2002) pp. 1–6

9.21 T. S. Barnett, M. Grady, K. Purdy, A. D. Singh: Redundancy implications for early-life reliability: experimental verification of an integrated yield–reliability model. In: *ITC International Test Conference* (IEEE Computer Society, Baltimore, MD 2002) pp. 693–699

9.22 T. S. Barnett, A. D. Singh: *Relating yield models to burn-in fall-out in time*, ITC Int. Test Conf. (IEEE Computer Society, Charlotte, NC 2003) pp. 77–84

9.23 H. H. Huston, C. P. Clarke: *Reliability defect detection and screening during processing: theory and implementation*, Proc. Int. Reliability Physics Symposium (IEEE Reliability Society, San Diego, CA 1992) pp. 268–275

9.24 K. O. Kim, W. Kuo, W. Luo: A relation model of gate oxide yield and reliability, Microelectron. Reliab. **44**, 425–434 (2004)

9.25 K. O. Kim, W. Kuo: A unified model incorporating yield, burn-in and reliability, Naval Res. Log. **51**, 704–719 (2004)

9.26 T. H. Kim, W. Kuo: Modeling manufacturing yield and reliability, IEEE Trans. Semicond. Manuf. **12**, 485–492 (1999)

9.27 W. C. Riordan, R. Miller, J. Hicks: Reliability versus yield and die location in advanced VLSI, Microelectron. Reliab. **39**, 741–749 (1999)

9.28 J. A. Cunningham: The use and evaluation of yield models in integrated circuit manufacturing, IEEE Trans. Semicond. Manuf. **3**, 60–71 (1990)

9.29 M. Raghavachari, A. Srinivasan, P. Sullo: Poisson mixture yield models for integrated circuits: a critical review, Microelectron. Reliab. **37**, 565–580 (1997)

9.30 R. E. Langford, J. J. Liou: *Negative binomial yield model parameter extraction using wafer probe bin map data* (IEEE Electron Devices Meeting, Hong Kong 1998) pp. 130–133

9.31 F. Jensen: *Yield, quality and reliability – a natural correlation?*, Reliability 1991, ed. by R. H. Matthews (Elsevier Applied Science, London 1993) pp. 739–750

9.32 Z. Stamenkovic, N. Stojadinovic: New defect size distribution function for estimation of chip critical area in integrated circuit yield models, Electron. Lett. **28**, 528–530 (1992)

9.33 Z. Stamenkovic, N. Stojadinovic: Chip yield modeling related to photolithographic defects, Microelectron. Reliab. **32**, 663–668 (1992)

9.34 G. Allen, A. Walton, R. Holwill: A yield improvement technique for IC layout using local design rules, IEEE Trans. CAD ICs Syst. **11**, 1355–1362 (1992)

9.35 S. Williamson, A. Singh, V. Nelson: *Fault and yield modeling of MCMs for automotive applications*, Proc. Int. MCM Conf. (IEEE Society, Denver 1996) pp. 147–154

9.36 J. Mattick, R. Kelsall, R. Miles: Improved critical area prediction by application of pattern recognition techniques, Microelectron. Reliab. **36**, 1815–1818 (1996)

9.37 C. H. Stapper, R. J. Rosner: Integrated circuit yield management and yield analysis: development and implementation, IEEE Trans. **8**, 95–102 (1995)

9.38 J. Segal, L. Milor, Y. K. Peng: Reducing baseline defect density through modeling random defect-limited yield, Micro Mag. **18**, 61–71 (2000)

9.39 H. Sato, M. Ikota, A. Sugimoto, H. Masuda: A new defect distribution metrology with a consistent discrete exponential formula and its application, IEEE Trans. Semicond. Manuf. **12**, 409–418 (1999)

9.40 K. S. Park, C. H. Jun: Semiconductor yield models using contagious distributions and their limiting forms, Comput. Eng. **42**, 115–125 (2002)

9.41 C. H. Jun, Y. Hong, Y. K. Kim, K. S. Park, H. Park: A simulation-based semiconductor chip yield model incorporating a new defect cluster index, Microelectron. Reliab. **39**, 451–456 (1999)

9.42 J. A. Carrasco, V. Suñé: Combinatorial methods for the evaluation of yield and operational reliability of fault-tolerant systems-on-chip, Microelectron. Reliab. **44**, 339–350 (2004)

9.43 M. Scheffler, D. Cottet, G. Tröster: A simplified yield modeling method for design rule trade-off in interconnection substrates, Microelectron. Reliab. **41**, 861–869 (2001)

9.44 S. P. Cunningham, C. J. Spanos, K. Voros: Semiconductor yield improvement: results and best practices, IEEE Trans. Semicond. Manuf. **8**, 103–109 (1995)

9.45 C. N. Berglund: A unified yield model incorporating both defect and parametric effects, IEEE Trans. Semicond. Manuf. **9**, 447–454 (1996)

9.46 D. Dance, R. Jarvis: Using yield models to accelerate learning curve progress, IEEE Trans. Semi. Manuf. **5**, 41–45 (1992)

9.47 L. Peters: Choosing the Best Yield Model for Your Product, Semicond. Int. **23**, 52 (May 2000)

9.48 X. Li, J. Strojwas, M. F. Antonelli: Holistic Yield Improvement Methodology, Semiconductor Fabtech **27**(7), 257–265 (1998)

9.49 L. Peters: Demystifying Design-for-Yield, Semicond. Int. **27**, 42 (July 2004)

9.50 A. Amerasekera, D. S. Campbell: *Failure Mechanisms in Semiconductor Devices* (Wiley, New York 1987)

9.51 J. E. Vinson, J. J. Liou: Electrostatic discharge in semiconductor devices: overview, Proc. IEEE **86**, 399–419 (1998)

9.52 N. Wan, K. Manning: *Exceeding 60-year life expectancy from an electronic energy meter*, Metering Asia Pac. Conf. (Kuala Lumpur, Malaysia 20-22 February 2001)

9.53 W. W. Abadeer, A. Bagramian, D. W. Conkle, C. W. Griffin, E. Langois, B. F. Lloyd, R. P. Mallette, J. E. Massucco, J. M. McKenna, S. W. Mittl, P. H. Noel: Key measurements of ultrathin gate dielectric reliability and in-line monitoring, IBM J. Res. Devel. **43**, 407–417 (1999)

9.54 C. Hu: Future CMOS scaling and reliability, Proc. IEEE **81**, 682–689 (May 1993)

9.55 J. Stevenson, J. A. Nachlas: *Microelectronics reliability prediction derived from component defect densities* (Pro Annual Reliab. Maintainab. Sympos., Los Angeles, CA 1990) pp. 366–370

9.56 R. Doyle, J. Barrett: Design for Reliability: a Focus on Electronics Assembly, Solder. Surf. Mount Technol. **9**(3), 22–29 (1997)

9.57 M. Baz, M. Udrea-Spenea, E. Tsoi, F. Turtudau, V. Ilian, G. Papaioannou, L. Galateanu, A. Stan, A. Tserepi, M. Bucur: Building in reliability technology for diodes manufacturing, IEEE CAS 2000 Proceedings **1**, 333–337 (2000)

9.58 W. T. K. Chien, C. H. J. Huang: Pracitical building in reliability approaches for semiconductor manufacturing, IEEE Trans. Reliab. **52**, 469–482 (2002)

9.59 S. Garrard: *Production implementation of a practical WLR program* (1994 IRW Final Report IEEE International, Lake Tahoe, CA 1995) pp. 20–29

9.60 L. N. Lie, A. K. Kapoor: *Wafer level reliability procedures to monitor gate oxide quality using V ramp and J ramp test methodology*, 1995 IRW Final Report (IEEE International, Gold Canyon, AZ 1996) pp. 113–121

9.61 A. P. Bieringer, D. Koch, H. Kammer, a. Kohlhase, A. Lill, A. Preusser, A. Schlemm, M. Schneegans: *Implementation of a WLR–program into a production line* (1995 IRW Final Report, Gold Canyon, AZ 1996) pp. 49–54

9.62 T. A. Dellin, W. M. Miller, D. G. Pierce, E. S. Snyder: Wafer level reliability, SPIE Microelectron. Manuf. Reliab. **18**(2), 144–154 (1992)

9.63 J. M. Soden, R. E. Anderson: IC failure analysis: techniques and tools for quality and reliability improvement, Proc. IEEE **81**, 703–715 (1993)

9.64 A. W. Righter, C. F. Hawkins, J. M. Soden, P. Maxwell: *CMOS IC reliability indicators and burn-in economics* (International Test Conference, IEEE Computer Society, Washington D.C. 1998) pp. 194–204

9.65 W. Kuo, Y. Kuo: Facing the headaches of early failures: a state-of-the-art review of burn-in decisions, Proc. IEEE **71**, 1257–1266 (1983)

9.66 A. Vassighi, O. Semenov, M. Sachdev: CMOS IC technology scaling and its impact on burn-in, IEEE Trans. Dev. Mater. Reliab. **4**, 208–222 (2004)

9.67 D. Gralian: Next generation burn-in development, IEEE Trans. Components Packaging Manuf. Technol. **17**, 190–196 (May 1994)

9.68 B. Vasquez, S. Lindsey: *The promise of known-good-die technologies, MCM '94 Proc.* (IEEE Society, Santa Clara, CA 1994) pp. 1–6

9.69 D. Inbar, M. Murin: KGD for flash memory: burn-in is key, Semicond. Int. **8**, 1 (2004)

9.70 A. Wager: *Wafer probe tests and stresses used for quality and reliability improvement*, 8th Annual KGD packaging and Test Workshop Proceedings (Die Products Consortium, Napa, CA September 2001)

9.71 P. Garrou: The wafer level packaging evolution, Semicond. Int. **27**, 119 (10/2/2004)

9.72 R. Nelson: Hot tests whip chips into shape, Test Meas. World **20**(12), 30 (October 2000)

9.73 D. Romanchik: Why burn-in ICs ?, Test Measurement World **12**, 85–88 (Oct. 1992)

9.74 J. Mi: Warranty policies and burn-in, Naval Res. Log. **44**, 199–209 (1997)

9.75 J. Mi: minimizing some cost functions related to both burn-in and field use, Oper. Res. **44**, 497–500 (1996)

9.76 S. H. Sheu, Y. H. Chien: |Minimizing cost functions related to both burn-in and field operation under a generalized model, IEEE Trans. Reliab. **53**, 435–440 (2004)

9.77 Kosznik, Drapella: Combining preventive replacements and burn-in procedures, Qual. Reliab. Eng. Int. **18**, 423–427 (2002)

9.78 J. H. Cha: On a better burn-in procedure, J. Appl. Probab. **37**, 1099–1103 (2000)

9.79 J. H. Cha: Burn-in procedure for a generalized model, J. Appl. Probab. **38**, 542–553 (2001)

9.80 J. H. Cha, S. Lee, J. Mi: Bonding the optimal burn-in time for a system with two types of failure, Naval Res. Logist. **51**, 1090–1101 (2004)

9.81 S. T. Tseng, C. Y. Peng: Optimal burn-in policy by using an integrated Wiener process, IIE Trans. **36**, 1161–1170 (2004)

9.82 K. N. Kim: Optimal burn-in for minimizing cost and multi-objectives, Microelectron. Reliab. **38**, 1577–1583 (1998)

9.83 W. T. K. Chien, W. Kuo: A nonparametric approach to estimate system burn-in time, IEEE Trans. Semicond. Manuf. **9**, 461–466 (1996)

9.84 W. T. K. Chien, W. Kuo: A nonparametric Bayes approach to decide system burn-in time, Naval Res. Logist. **44**, 655–671 (1997)

9.85 W. Kuo: Reliability enhancement through optimal burn-in, IEEE Trans. Reliab. **R-33**, 145–156 (1984)

9.86 D. H. Chi, W. Kuo: Burn-in optimization under reliability and capacity restrictions, IEEE Trans. Reliab. **38**, 193–199 (1989)

9.87 T. R. Kar, J. A. Nachlas: Coordinated warranty & burn-in strategies, IEEE Trans. Reliab. **46**, 512–519 (1997)

9.88 K. O. Kim, W. Kuo: Percentile residual life and system reliability as performance measures in the optimal system design, IIE Trans. **35**, 1133–1142 (2003)

9.89 H. W. Block, J. Mi, T. H. Savits: Some results on burn-in, Stat. Sin. **4**, 525–534 (1994)

9.90 H. W. Block, J. Mi, T. H. Savits: Burn-in at the component and system level. In: *Lifetime Data: Models in Reliability and Survival Analysis*, ed. by N. P. Jewell, A. C. Kimber, M. L. T. Lee, G. A. Whitmore (Kluwer, Dordrecht 1995)

9.91 C. W. Whitbeck, L. M. Leemis: Component vs. system burn-in techniques for electronic equipment, IEEE Trans. Reliab. **38**, 206–209 (1989)

9.92 R. K. Reddy, D. L. Dietrich: A 2-level environmental stress screening model: a mixed distribution approach, IEEE Trans. Reliab. **43**, 85–90 (1994)

9.93 E. A. Pohl, D. L. Dietrich: *Environmental stress screening strategies for multi-component systems with weibull failure times and imperfect failure detection* (Proceedings Annual Reliability and Maintainability Symposium, Washington D.C. 1995) pp. 223–232

9.94 W. Kuo: Incompatibility in evaluating large-scale systems reliability, IEEE Trans. Reliab. **43**, 659–660 (1994)

9.95 W. T. K. Chien, W. Kuo: Modeling and maximizing burn-in effectiveness, IEEE Trans. Reliab. **44**, 19–25 (1995)

9.96 W. T. K. Chien, W. Kuo: Optimal burn-in simulation on highly integrated circuit systems, IIE Trans. **24**, 33–43 (1992)

9.97 T. H. Kim, W. Kuo: Optimal burn-in decision making, Qual. Reliab. Eng. Int. **14**, 417–423 (1998)

9.98 K. O. Kim, W. Kuo: A general model of heterogeneous system lifetimes and conditions for system burn-in, Naval Res. Log. **50**, 364–380 (2003)

9.99 K. O. Kim, W. Kuo: Some considerations on system burn-in, IEEE Trans. Reliab. **54**(2), 207–214 (2005)

9.100 K. O. Kim, W. Kuo: Two-level burn-in for reliability and economy in repairable series systems having incompatibility, J. Reliab. Qual. Saf. Eng. **11**(3), 197–211 (2004)

9.101 F. Kuper, J. Vander Pol, E. Ooms, T. Johnson, R. Wijburg: *Relation between yield and reliability of integrated circuits: experimental results and application to continuous early failure rate reduction programs* (Proc. Int. Reliab. Phys. Symp., Dallas, Texas 1996) pp. 17–21

9.102 J. Van der Pol, F. Kuper, E. Ooms: Relation between yield and reliability of ICs and application to failure rate assessment and reduction in the one digit fit and ppm reliability era, Microelectron. Reliab. **36**, 1603–1610 (1996)

9.103 T. Zhao, Y. Hao, P. Ma, T. Chen: *Relation between reliability and yield of ICs based on discrete defect distribution model*, Proc. 2001 IEEE Int. Symp. Defect Fault Tolerance in VLSI systems (IEEE Society, San Francisco, CA 2001)

9.104 J. Van der Pol, E. Ooms, T. Hof, F. Kuper: *Impact of screening of latent defects at electrical test on yield-reliability relation and application to burn-in elimination* (Int. Relia. Phys. Sympos., Reno, NV 1998) pp. 370–377

9.105 K. R. Forbes, N. Arguello: Using time-dependent reliability fallout as a function of yield to optimize burn-in time for a 130nm SRAM device. In: *Integ. Reliab. Workshop Final Rep., IEEE Int., Lake Tahoe, CA* (IEEE Society 2003) pp. 61–66

9.106 T. H. Kim, W. Kuo, K. Chien: Burn-in effect on yield, IEEE Trans. Electron. Pack. Manuf. **23**, 293–299 (2000)

9.107 K. O. Kim, M. Zuo, W. Kuo: On the relationship of semiconductor yield and reliability, IEEE Trans. Semicond. Manuf. **18**(3), 422–429 (2005)