

Chapter 1

Introduction

Higher yields, lower costs, and improved reliability are some of the demands that drive the production-line manufacture of integrated circuits. Shorter cycle times for chip design and shorter ramp times for new fabrication processes are among the demands that drive processing development and research. Although these forces drive the ship, the ballast is provided by a constant attention to quality.

In the case of integrated circuit (IC) fabrication, the first phase of quality assurance occurs when process tests are conducted both during and after wafer manufacture. Process tests are absolutely critical to the production of high-quality IC's. Process tests qualify wafers, not devices, however, and many serious device problems cannot be eliminated by process tests. Instead, most device problems are caught during the second phase of quality assurance, called 'wafersort,' wherein each device on the wafer is subjected to a suite of electrical tests.

1.1 Wafersort

Figure 1.1 shows the overall flow of wafers and information during the manufacturing of IC's. On the left, IC fabrication begins with polished, bare wafers of crystalline silicon. Several wafers are processed together in a 'lot,' typically of 25 wafers, and IC's are created on every wafer in the lot. After all the process steps have been completed, the wafers are

sent for final cleaning and for process tests. Most wafers pass process tests and become finished wafers that leave the fabrication facility, or ‘fab.’ These wafers enter wafersort, as illustrated in Fig. 1.1.

The purpose of wafersort is to test the many individual IC’s on the finished wafers. Every IC that fails any test is identified with an ink dot. Figure 1.2 is a photograph of a finished wafer fabricated with over 200 individual IC’s, some of which failed wafersort and were ‘inked.’ Were the wafer in the photo to be sawn, the inked IC’s would be discarded and the passing IC’s packaged.

In Fig. 1.1, the outputs of wafersort are inked wafers and IC test results. The wafersort step is often called ‘probe test’ and the IC test results referred to as the ‘probe-test data.’ Probe tests are electrical measurements, and the results of these measurements are stored and retained for some time—often a few years—until the packaged devices reach customers and potential problems have time to surface. A large test-house with fifty testers operating around-the-clock can easily produce upwards of a giga-byte¹ of data every day. Typically, these data are examined only when a problem occurs, perhaps when the yield drops or a customer reports a problem.

When the need for diagnostic work arises, the data files are retrieved so that various graphical and statistical tools can be employed to identify the cause of the problem. The failing devices often cluster together in distinctive spatial patterns on the wafer. Anyone who looks at a tested wafer can see the pattern because of the ink dots; the photograph in Fig. 1.2 provides an example. Notice that most of the inked devices are above the mid-line of the wafer, near its crown.

Two-dimensional arrays called wafer maps are constructed and examined at this point. A ‘wafer map’ reports the spatial distribution of a particular measurement. For example, Fig. 1.3 displays the pass/fail wafer map associated with the photograph. The grey squares represent the inked IC’s.² Wafer map information can be reported in various ways; for example, as pass/fail, as fail-low/pass/fail-hi, and as graduated values in a contour map.

¹A giga-byte is 10^9 bytes

²The incomplete IC’s at the rim of the wafer have been omitted from the wafer map, since they always fail.

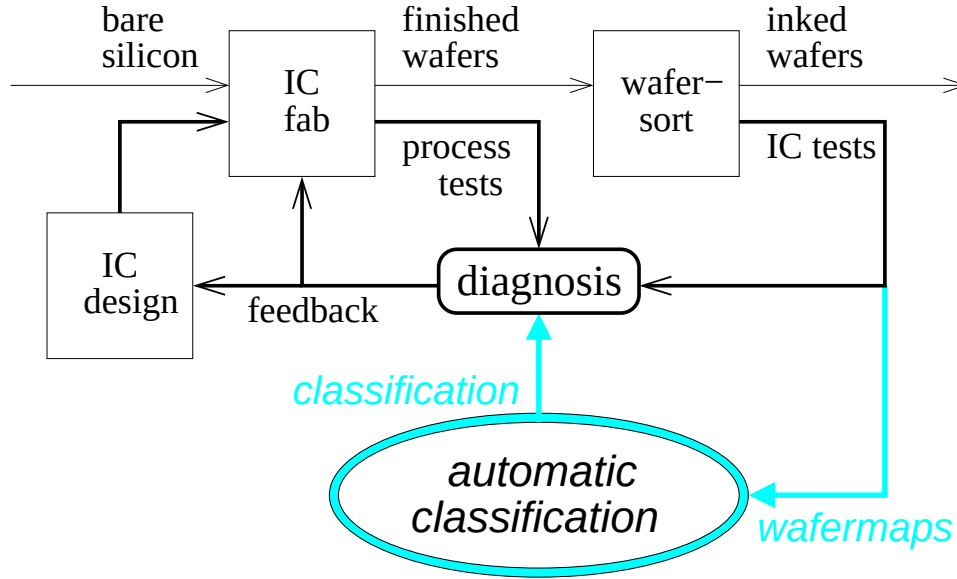


Figure 1.1: Wafer and Information Flow during IC Manufacture. Wafer flow is marked by thin lines and information flow by thick lines. This research augments the information flow by introducing an automatic diagnostic, as shown.

Together with the test ID, the spatial distribution of failures—the ‘failure pattern’—is often characteristic of a specific fabrication or design problem. Unfortunately, wafer maps require human analysis and are time-consuming to study. The resources to examine and analyze the wafer maps of every wafer simply do not exist. Only automatic classification and analysis could provide the resources to examine every wafer map.

The goal of this research is to work toward the automatic classification of wafer map failure patterns. Classification requires us to choose features that can be identified and categories that describe those features. As an example, the areal extent of a failure pattern is directly related to the yield of the wafer and is an appropriate choice as a feature. Knowing that, we might choose to separate this *Area* feature into categories such as *Small*, *Medium*, and *Large*.

Two purposes drive our choice of features. First is the desire to provide a *useful* characterization of the failure pattern. To do so, as a minimum it is necessary to classify the *Shape*, *Area*, and *Location* of a failure pattern. Second is the desire to provide a *complete* characterization of a failure pattern. Shape, area, and location are functions of

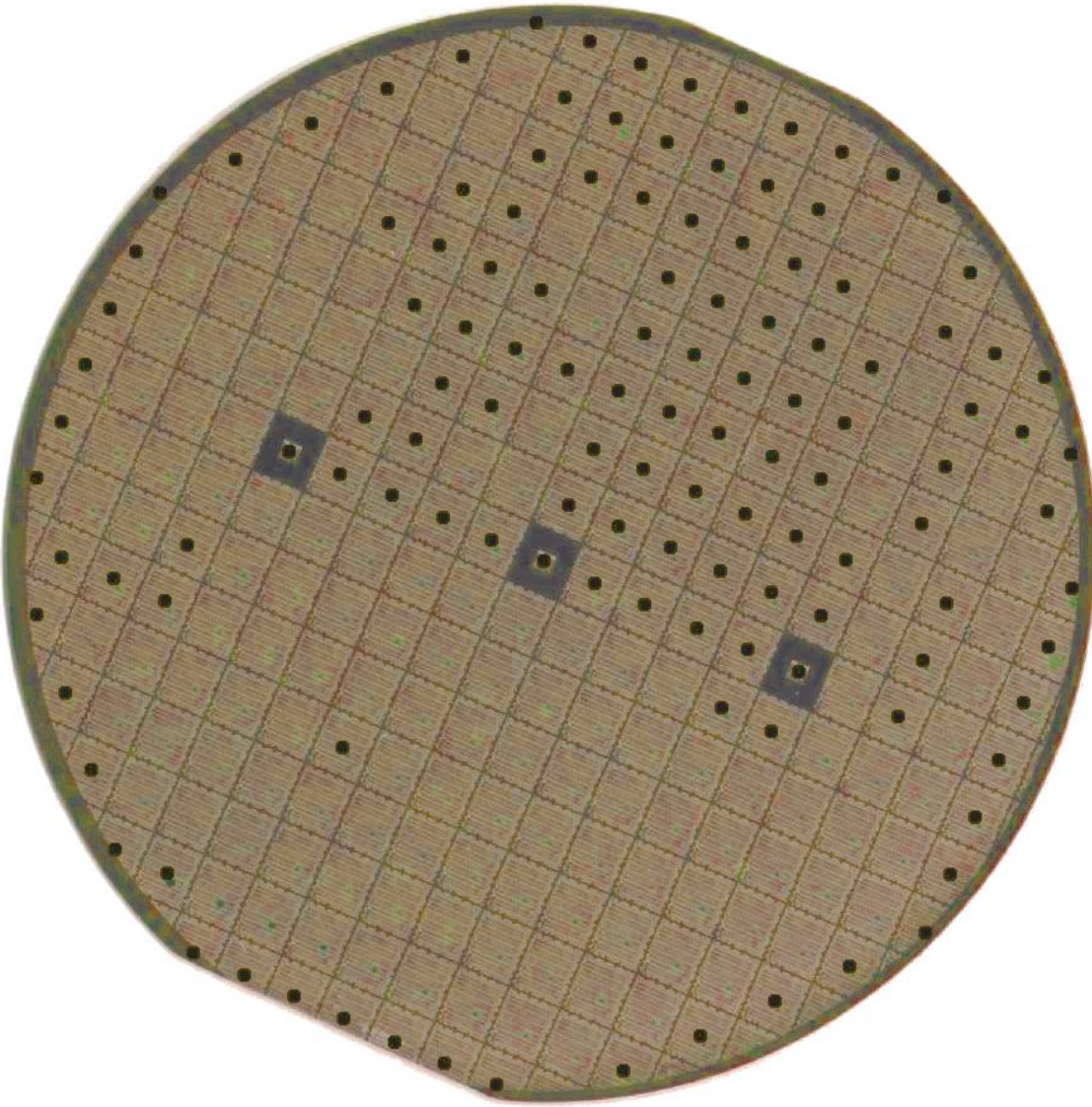


Figure 1.2: A Photograph of a Tested Wafer. Ink dots mark IC's that failed electrical testing. Notice that there are many more ink dots at the crown (upper right) than anywhere else on the wafer. Photo by Damon Hart-Davis/DHD Multimedia Gallery at <http://gallery.hd.org/>. The ink dots were added by the author for illustration.

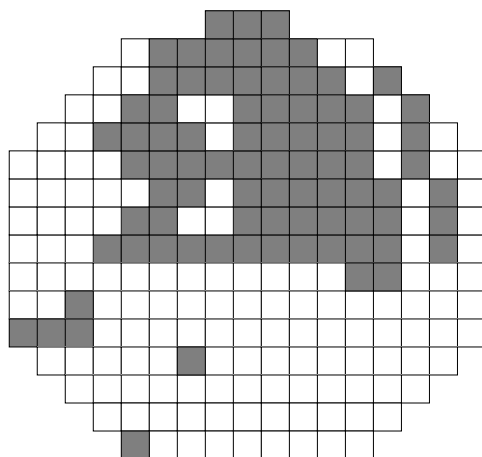


Figure 1.3: The Photograph as a Wafer Map. IC's with ink dots are represented as grey boxes; those without ink are white. An alternate format for binary maps such as these uses a '1' (fail) for a grey box and a '0' (pass) for a white box.

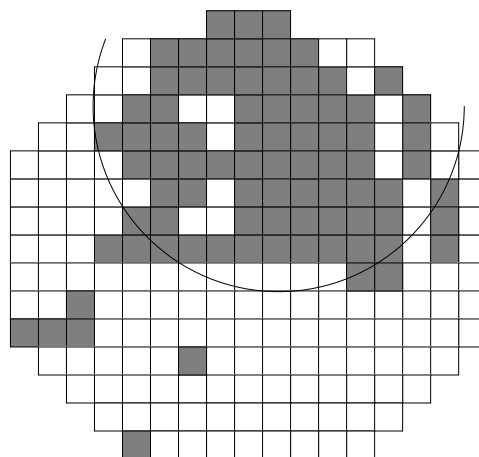


Figure 1.4: The Wafer Map Divided into Two Regions. A circular boundary curve has been superimposed upon Fig. 1.3, thereby allowing it to be described in just a few words by naming the shape and giving the size and location of the low yield region.

the pattern geometry and do not provide a complete description of complicated patterns. Additional features are selected as required.

To create a geometrical model of failure patterns, one must begin by selecting appropriate shapes. Two geometries dominate in the fabrication of wafers: straight lines (especially in lithography) and circles (*e.g.*, from the spinning of photoresist). Straight lines and circles are simple shapes, and with that simplicity comes mathematical power. Therefore, our shape classifications are based upon straight lines and circles.

Figure 1.4 provides an illustration of our approach to the classification of shape. A circular boundary has been applied to the wafer map of Fig. 1.3, dividing it into two regions. Within the circle, the yield is low; without, the yield is high. The boundary does not fit the failure pattern perfectly, but it does capture certain essential features. First, it surrounds or intersects all but a few of the devices in the largest failure grouping. Second, it faithfully recognizes that there are good devices at the left and right edges of the wafer rim. And third, it provides a reasonably accurate estimate of the area and location of the

primary low-yield region.

There are many reasons to make use of probe-test data. One is that the use of probe-test data is inexpensive, since probe tests have to be done anyway. Another reason is that probe tests can reveal both large-scale problems—those that involve much of a wafer’s surface—and small-scale problems that involve only a small part of a single device. Finally, probe-test results remain available for data mining for as long as the data are retained, which is typically a year or two. Therefore, by making use of these results, we gain access to a veritable ocean of information about design and processing problems.

1.2 Other Work

Some of the most powerful yield analysis tools exist for the diagnosis of memory IC’s such as RAM, ROM, and on-chip caches. These tools are effective because the circuitry in a memory chip is so regular and observable.

Typically, a memory diagnostic system has a library of facts that identify associations between the circuit layout, the possible defects, and the electrical signature of each defect [23], [26]. Electrical test results may be used to identify critical spots on the wafer and to direct computer-controlled tools to accomplish detailed examinations through SEM³ or FIB⁴ [30]. Graphical tools to aid engineering analysis are often another important component [19], and pattern recognition techniques are always employed to one degree or another [7], [28], [46]. There can be a remarkable amount of integration between testing, failure modeling, and process correction in the manufacture of memories. IC’s other than memories are not handled so readily, however.

Yield enhancement techniques for non-memory IC’s tend to focus on the use of probe-test data. One area of research that continues to receive a lot of attention is in the use of statistical methods to analyze wafer maps. Techniques such as principal component analysis (PCA) [31], binomial tests [20], cluster analysis [17], [25], [35], and many others [10], [11], [45], [52] have been studied.

³Scanning Electron Microscopy

⁴Focused Ion Beam etching

Zonal analysis of wafer maps is widely used as well [1], [34], [35], [36]. The wafer is divided into regions—the zones—and each region is analyzed in isolation. Afterwards, regions are compared, both within and among wafers. There are many ways to divide a wafer, and the literature contains almost as many division schemes as research groups. Some generalizations can be made, nevertheless. Wafers are often divided into concentric zones; typically two, three, or four zones are chosen. Also, wafers are often divided radially into an even number of pie wedges, typically two to eight. Both concentric and radial zones may co-exist. For example, there may be two radial zones with the outer zone subdivided into pie wedges. Some researchers have chosen to subdivide further, but most zones are radial, angular or both. We use both radial and angular zone structures when we create **Location** categories in Chapter 7.

After statistical methods and zonal analysis, quadrat analysis [51], [58] is a third common technique used to extract spatial information. A quadrat is a square region in a grid. The idea is to superimpose grids of several sizes onto a wafer map, then, as a function of quadrat size, analyze changes to various calculated parameters such as defects per quadrat.

Quadrats and zones are used also in the creation of yield models, especially by Stapper [49], [50]. Yield models are employed usually when it is necessary to predict the yield of a new IC. A more interesting application is to create a yield model to explain a set of observations [56]. Classically, yield modeling involves the derivation of an analytic expression for the yield from a set of assumptions [2], [48]. For instance, the oldest and simplest yield model is the Poisson model, which assumes that defects are uniformly and randomly distributed. Typically it also assumes that every defect causes a functional failure, so only $k = 0$ produces a working IC. Following this model, if we let λ_0 be the average number of defects per IC, let $P(k)$ be the probability that an IC will actually have k defects, and let Y be the yield (*i.e.*, $P(0)$), then

$$P(k) = \frac{e^{-\lambda_0} \lambda_0^k}{k!} \quad \text{for } k=0,1,2,\dots \quad \text{and} \quad Y = e^{-\lambda_0}. \quad (1.1)$$

Usually, one assumes that $\lambda_0 = D_0 A$, where A is the area of one IC and D_0 is the number

of defects per unit area. Cunningham [9] provides a good historical review of yield models.

One other topic should be mentioned, namely the application of various kinds of knowledge systems to failure analysis. Two early examples of knowledge systems used in semiconductor manufacturing are P.I.E.S. [37] and SMART [33]. Maly, *et al* [27], recommend using a hierarchical methodology for the interpretation of tester data. Methods such as CART [5] and decision trees [40] would be appropriate in that case. The classifiers constructed in this dissertation can be used by themselves and in conjunction with an expert system.

1.3 Organization of the Dissertation

Chapter 2 provides a brief overview of fabrication processes, and suggests some failure modes leading to various kinds of wafer maps. The focus is narrowed then to look more closely at the specific wafer maps that we attempt to classify. Here, wafer maps are defined in terms of boundaries and failure zones. Boundaries are chosen to be straight or circular, and wafer maps are restricted to one failure zone.

In Chapter 3, the mathematics for creating statistical populations of these wafer maps are developed. As described in detail in Appendix A, the proper definition of random (uniform) population statistics relies on Poincaré’s solution to Bertrand’s paradox. A uniform probability density function is developed for each of the five shapes chosen in Chapter 2. Two to four statistical variables are required to specify a particular sample from one of the shape populations.

Chapter 4 develops the equations by which ‘feature variables’ can be calculated from the statistical variables. Each feature variable is carefully defined, and expressions are developed to calculate features such as Area, width, Location, Center, Centroid, Orientation, and Curve Direction. More features are developed than actually are used in classification. Criteria are established and comparisons made between similar features, such as Center and Centroid, and a subset of the initial feature set is chosen for use in classification.

Chapter 5 describes **wafgen**, a computer program we wrote that generates wafer map failure patterns. **Wafgen** creates synthetic wafer maps by employing the mathematics

developed in Chapters 2–4. Key aspects of the algorithmic methods used in `wafgen` are described first: how wafer map populations are sampled, how failure patterns with similar shapes are made distinct from one another, how a discrete wafer map is created from its continuous description, and how the finished discrete wafer map is re-labeled to ensure accuracy. Following that, the numerical output of `wafgen` is compared with theoretical expectations: a pdf and cdf ⁵ are constructed numerically for A-Disks, A-Annuli, A-Rings, Segments, and Bands. `Wafgen`’s statistical behavior is demonstrated to be correct.

In Chapter 6, a standard 12×12 input format is chosen for the classifier, and methods are established that can transform any wafer map into the 12×12 format. The need to ‘standardize’ originates with the need to accommodate a range of sizes for rectangular IC’s, which vary in size from as tiny as $100 \mu\text{m} \times 300 \mu\text{m}$ for resistors to upwards of 2 cm square for microprocessors. Wafer diameters vary also, from 100–300mm. Hundreds of thousands of tiny IC’s can be fabricated on a large wafer, while perhaps as few as a dozen large IC’s can be fabricated on a small wafer. When probe-test data are assembled, the ‘raw’ wafer map for the tiny IC has hundreds of thousands of entries, but the raw wafer map for the large IC has only a dozen. Since our classifiers require a fixed input format, we must transform raw wafer maps into a standard form. Chapter 6 provides a detailed development of the transformation methods for one raw format. Appendix C contains results from the transformation of five more.

Chapter 7 begins a series of three chapters on classification, providing the foundation upon which the next two chapters are built. It opens with a discussion of how one might divide the area, orientation, location, and curve-direction features into categories. A fundamental idea known as a ‘distance measure’ is presented next and is followed by a description of its application in this dissertation, specifically in nearest neighbor and prototype classifiers. Another fundamental idea is that of the Bayes classifier, which is known to produce the fewest classification errors. Bayes classifiers are explained and compared with distance classifiers, and I report our efforts to construct a Bayes classifier for wafer maps. The chapter closes with an explanation of the making of each of the

⁵‘probability density function’ and ‘cumulative distribution function’

datasets we used to construct and test our classifiers.

Chapter 8 reports the results of three classification experiments conducted using *wafgen*'s 'synthetic' wafer maps. In the first experiment, we establish an upper bound on the Bayes error.⁶ In the second, the prototype and nearest neighbor classifiers are analyzed with respect to the classification of synthetic wafer maps. In the third, various amounts of additive random noise are applied to the synthetic maps, and the classification results are analyzed for both the prototype and nearest neighbor classifiers.

Experiments four, five and six are reported in Chapter 9, where real wafer maps (*i.e.*, re-formatted industrial wafer maps) are classified for the first time. In experiment four, the prototype and nearest neighbor classifiers are used to identify real wafer maps and the classification errors are analyzed in detail. In experiments five and six, real wafer maps are used to build new classifiers. In the fifth experiment, the set of 3,245 real wafer maps is separated into two half-sets, and each half-set is used to classify the other. In the sixth experiment, all of the real wafer maps are used to classify synthetic wafer maps plus noise. After the six experiments are complete, the results are collected and used to assign causes to the classification error of real wafer maps. Error sources such as 'Bayes error,' 'design of classifier,' 'skew statistics,' and 'shape deformation' are identified and quantified. The chapter ends with two demonstrations. First, we show that bit noise in real wafer maps is not random and examine the properties of two of its characteristics. Second, we illustrate how to use the wafer map model from Chapters 2–4 to characterize the statistical properties of a large population of wafer maps.

Finally, Chapter 10 reviews the results of the dissertation, discusses open questions, and provides suggestions for future work.

1.4 Contributions

Among the contributions made in this dissertation are:

- The recognition that wafer maps have the potential to be characterized statistically and geometrically.

⁶The classification error of the optimal Bayes classifier is known as the Bayes error.

- The application of geometrical probability concepts to define the uniform random distribution for each of five populations of geometrical objects.
- The selection of wafer map features for classification and categories for each feature.
- A demonstration that an individual wafer map is uniquely specified by the features selected.
- The creation of wafer map generation software to create synthetic wafer maps as random samples of a specific shape population, to identify the feature categories into which each new wafer map falls, and to label them automatically as they are synthesized.
- The construction of a new mathematical framework to measure and characterize large parts of any foundry's wafer map database by applying the geometrical probability model to measurement.
- A demonstration of the measurement of a population of industrial wafer maps.
- An experimental demonstration that the Bayes error of the synthetic wafer map population is less than 1.5%. Therefore, the synthetic wafer maps comprise an entirely new, statistically well-defined, extremely large, and automatically labeled dataset with a very small Bayes error.
- The construction of several nearest neighbor classifiers and two prototype classifiers.
- The execution and analysis of six classification experiments which examine the accuracy of the classifiers when identifying (1) perfect synthetic wafer maps, (2) synthetic wafer maps with varying amounts of additive random bit-noise, and (3) real industrial wafer maps.
- The invention of a method for re-formatting wafer maps for classification.
- A quantification of the causes of error in the classification of real industrial wafer maps.

