Modeling the Random Component of Manufacturing Yield of Integrated Circuits

Michael E. Long and David L. Farnsworth

College of Science, Rochester Institute of Technology, Rochester, New York, USA

Abstract—**A model is created for the number of integrated circuits that are good from each wafer on which they are fabricated. The goal is to separate the random or common cause loss from the systematic or special loss. The random loss from this type of process is modeled so that false alarms indicating systematic loss are less likely to occur and so that the structure of the systematic loss can be determined.**

Keywords— **Baseline randomness, modeling, skewness, systematic loss, yield curve.**

I. INTRODUCTION

The fabrication of integrated circuits is a large and important industry. Hundreds of integrated circuits are manufactured at the same time on wafers, which are created in lots. Each wafer is more-or-less circular with a diameter of a few inches. The manufacturing process can take many weeks and is quite complicated. Each circuit must pass many tests before it is sold. There is a considerable literature on modeling and minimizing the loss or number of bad die. See [1], [2], [3], [4], and [5] for reviews.

Usually, the integrated circuits are called die when they are on a wafer, chips when they are separated from the wafer, and integrated circuits when installed for use. We prefer the word "die," which we also use for the plural. Optical and electrical measurements are used to test or screen the die. Among the reasons to discard a die are wrong resistance, an open circuit, a missing conductor, and being physically too large or too small in some dimension.

Yield is the fraction or percent of the die on a wafer that passes all tests. Since all wafers in our simulation contain 1000 die, we use the number that pass all tests as the yield. Yield can refer to the fraction or the percent of useable die in a whole process or in a lot as well. Yield can be affected by many events.

We are principally concerned with random loss of yield. Random loss is characterized by the presence or absence of any particular defect in each die being independent of the condition or location of the other die on a wafer. The random loss with respect to each attribute is uncorrelated between die, and the attributes are uncorrelated within the die. Random loss is associated with common causes, such as dust particles in cleanrooms and the manufacturing equipments' imprecision in general. Patterns in which bad die are grouped along the edge, say, of the wafer produce systematic or special effects. Sometimes, systematic loss is caused by equipment failures. Our goal is to model the variability of the lots, wafers, and die, so that random loss can be identified. Then, systematic loss can be revealed as the residual loss after accounting for the random loss.

There are many differing models of the distribution of defects on a wafer. For random effects, a widely used family of models is based on compound or mixed models in which a gamma distribution is used for the density of defects and the yield is an average or integral over that distribution [6], [7], [8]. This produces a generalized negative binomial distribution [9, p. 191] with a parameter, usually designated a, that measures the degree of clustering. There are other ways to reach the same model [10]. We take $a > 10$, which is the Poisson model [6]. This model has many mathematical advantages, in particular it has the additive property that the sum of Poisson random variables is a Poisson random variable [9, p. 146]. Since Poisson random variables and binomial random variables converge in distribution under certain conditions, such as large sample sizes [9, pp. 217–219], and since we assume normal distributions for the variables or attributes that are measured, our analysis is based on binomial and normal distributions. Alternatively, the binomial model can be developed from first principals [3, p. 31], [11, p. 1067], [12].

One tactic is to fit the yield to a prescribed distribution [2], [6], [11], [13], which can be somewhat limiting for this complex process. Another tactic for measuring the random component of yield is called windowing in which neighboring die are grouped as if they were one die and are declared good if all die in the group are good [3, p. 27], [10], [13, pp. 5–6], [14]. Usually these super die are created with two, four, and eight die and sometimes more die under the assumption that random effects are not dependent on area, that is, they are scalable as a Poisson random variable is. This can be an effective method, but it is an art as well as a science; for example, if a grouping contains many die, it will almost certainly be bad and too few groupings will not give significant results.

Usually, the die's yield is represented by the product of a yield limited by random loss and a yield limited by systematic loss [3], [13], [15]. The random and the systematic factors are themselves products. Since each die is affected by various independent effects, yields are determined by products. Our goal is to model the yield from random loss

alone so that the systematic yield can be estimated utilizing the product.

We select the variability among the lots and among the wafers that are in the lots and consider various testing criteria for the die. The parameters of the output are examined with the shape of the yield histogram being of particular interest. The variability of the lots and wafers is modeled with normal distributions. Each test produces a binomial random variable with the probability of each die passing the test being computable from the normal distributions.

Fig. 1 is the histogram, called a yield curve, of the yields for 1500 wafers. The scaling has been removed for reasons of confidentiality for the company, whom we may only thank anonymously for these data and other data that we were allowed to examine and analyze. The histogram does not represent a binomial distribution, and the long left tail might be construed to indicate systematic loss. However, in Sections II and III we create similarly left-skewed yield curves from normally distributed variables that represent measurements on die from a manufacturing process. Yield curves similar to the one in Fig. 1 can be found in [3, p. 30], [4, pp. 70, 85], [11, p. 1068], [13, p. 3], [15, p. 134], and [16, p. 102]. In some of those, the left tail has sufficient structure to indicate that nonrandom effects may be in play, but others display simple left tails like the one in Fig. 1.

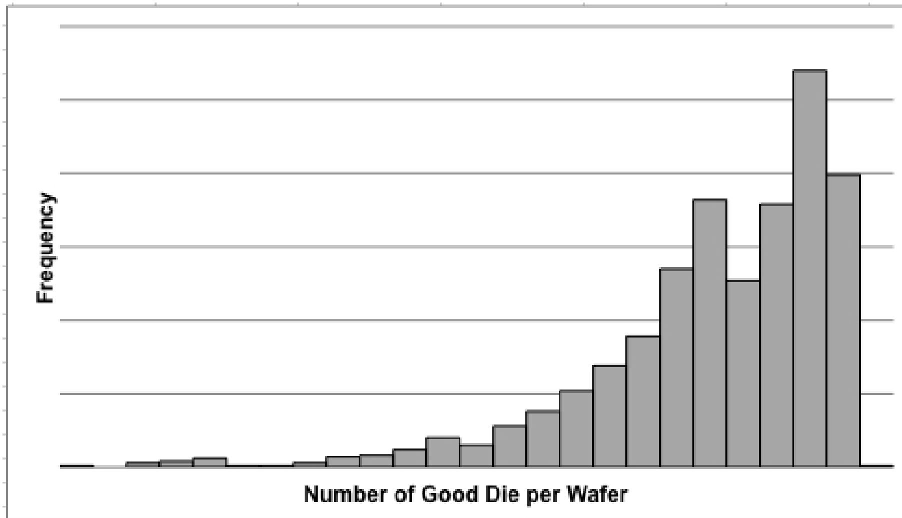

Fig. 1. A yield curve of actual data showing the number of good die per wafer for 1500 wafers. Higher yields are to the right. Scaling has been removed for reasons of confidentiality for the manufacturing company.

2. THE MODEL

For specificity and as an example, in this section we describe the manner in which the simulated data were generated for Fig. 2, 3, and 4, which are presented in Section III. Consider normally distributed random variables, which are the attributes of a die. For this model there are three attributes X_1, X_2, and X_3 that are measured and used for screening each die for acceptance or rejection. In practice, there are many more attributes. There is an acceptable region for each random variable or attribute. The requirements for attributes X_1, X_2, and X_3 are $X_1 \geq 1$, $7 \leq X_2 \leq 11$, and $X_3 \leq 33$. The main features of the model's output that is described in Section III can be created even for only one random variable or attribute. The one-sided left or right and two-sided criteria affect the output in similar ways.

There are ten lots, each containing 20 wafers, and each wafer has 1000 die. For each of the three attributes or random variables, consider process normal distributions with means $m_1 = 3$, $m_2 = 9$, and $m_3 = 27$ and standard deviations $s_1 = 1$, $s_2 = 1$, and $s_3 = 3$, respectively.

For attribute X_1, randomly select ten numbers, written m_{1i}, i = 1, 2, …, 10, from the adjusted process normal distribution with mean $m_1 = 3$ and standard deviation $s_{L1} = a_{L1}s_1 = a_{L1}(1) = a_{L1}$. The parameter a_{L1} is chosen for each run and is between 0.0 and 0.9. These ten m_{1i} are the means for X_1 of the ten lots, where each lot has the standard deviation $s_{W1} = a_{W1}s_1 = a_{W1}(1) = a_{W1}$. The parameter a_{W1} is chosen separately from a_{L1}, is also between 0.0 and 0.9, and represents the wafer-to-wafer variability within each lot. From each of the ten lots' distributions with mean m_{1i} and standard deviation s_{W1}, randomly select 20 numbers, which are the means m_{1ij}, j = 1, 2, …, 20, for the random variable X_1 for the wafers, where each wafer's X_1 has standard deviation $s_1 = 1$. Experience has shown that lot-to-lot variability is less than wafer-to-wafer variability, which is less than within wafer variability, so the parameters are selected to reflect that in Section III. A simple random sample of 1000 measurements of X_1 is taken for each wafer. For wafer j of lot i, the measurements are x_{1ijk}, k = 1, 2, …, 1000, from the normal distribution with mean m_{1ij} and standard deviation $s_1 = 1$.

This model is a 2-stage nested or hierarchical design [17, pp. 571–582], [18, pp. 525–536]. The two stages are the lots and the wafers. There are 1000 replicates, the measurements on each die. It is a balanced design because each run has the same number of lots, each lot has the same number of wafers, and each wafer has the same number of die, that is, measurements or replicates. It is a random effects, not fixed effects, model since there are no constraints that a fixed effects model would have. An example of a constraint is the sum of the ten lots' means has a certain value. No clustering or correlation among the effects is considered, that is, each of the 1000 die on each wafer has the same probability of passing a test.

For attribute X_2, randomly select ten numbers, written m_{2i}, from the adjusted process normal distribution with mean $m_2 = 9$ and standard deviation $s_{L2} = a_{L2}s_2 = a_{L2}(1) = a_{L2}$. These are the means for X_2 of the ten lots, where each has the standard deviation $s_{W2} = a_{W2}s_2 = a_{W2}(1) = a_{W2}$. The

parameters a_{L2} and a_{W2} are selected from the interval (0.0, 0.9), independently of each other and independently of a_{L1} and a_{W1}. From each of the ten lots' distributions of m_{2i}, randomly select 20 numbers, which are the means m_{2ij} of X_2 for the wafers, where each wafer's X_2 has standard deviation $s_2 = 1$. From each wafer, randomly select 1000 numbers x_{2ijk}, which are the realizations of the attribute X_2 on the die. For attribute X_3, repeat the same process, using $m_3 = 27$, $s_3 = 3$, and independently selected a_{L3} and a_{W3}. There is great flexibility in this model, since there are many parameters.

For each of the 10•20•1000 = 200,000 die there are three measurements, (x_{1ijk}, x_{2ijk}, x_{3ijk}), where i is the lot number, j is the wafer number, and k is the die number.

The goal of [17, pp. 571–582] and [18, pp. 525–536] is to use the method of analysis of variance to test for differences in the means in the lots and in the wafers in a nested design. However, in this application there are screening or testing steps, which pass or do not pass each die. Each wafer's yield is the number of the 1000 die that pass all tests, three in this study. There are 10•20 = 200 yields for each run. Our goal is to examine the structure of those 200 numbers.

3. THE MODEL'S OUTPUT

For simplicity, to produce the outputs in Fig. 2, 3, and 4, take the lot-to-lot multipliers to be equal and designate $a_L = a_{L1} = a_{L2} = a_{L3}$, and take the wafer-to-wafer multipliers to be equal and designate $a_W = a_{W1} = a_{W2} = a_{W3}$. Similar outputs are found with varying and unequal values of the parameters.

Although random effects are usually considered to be normally distributed [5, p. 49], the output from this model is not normal or binomial and, indeed, actual manufacturing process' outputs can be non-normal or non-binomial as Fig. 1 shows. Fig. 2, 3, and 4 illustrate sample yield curves, which are typical outputs. For Fig. 2, $a_L = 0.1$ and $a_W = 0.3$; for Fig. 3, $a_L = 0.2$ and $a_W = 0.6$; and for Fig. 4, $a_L = 0.3$ and $a_W = 0.9$. It should be emphasized that the parameters a_L and a_W give relative sizes of the lot-to-lot and wafer-to-wafer variability with respect to the within wafer variability. The skew for the data in Fig. 2, 3, and 4 are –0.687, –0.765, and –0.596, respectively. The kurtosis for the data in Fig. 2, 3, and 4 are 0.416, 0.052, and –0.431, respectively. The normal probability plots show visually that we would reject the three hypotheses that the data sets come from normal populations. For the Anderson-Darling test of normality, the P-value is less than 0.005 for each data set, indicating non-normality. Fig. 2, 3, and 4 show the progression to lower yielding wafers as the lot-to-lot and wafer-to-wafer variability increases.

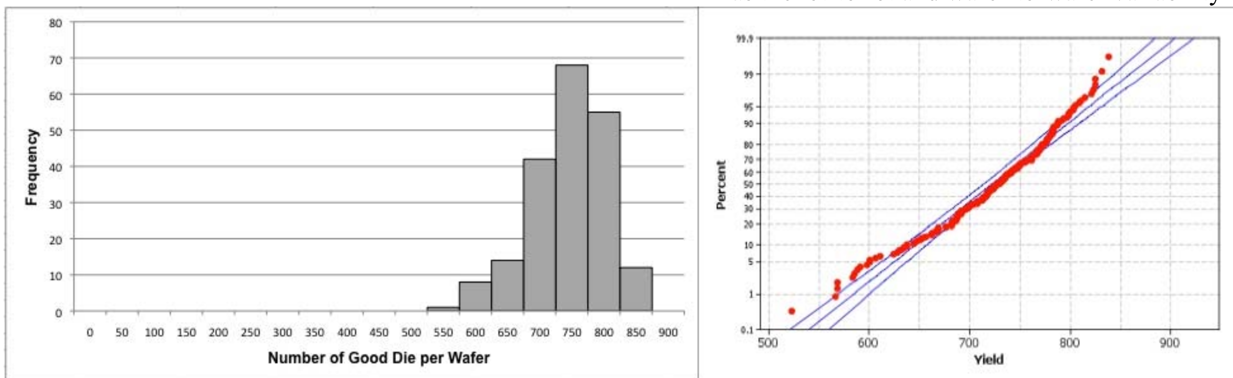

Fig. 2. A yield curve of simulated data (on the left) showing the number of good die per wafer for lot-to-lot standard deviation of 10% of the within wafer standard deviation and wafer-to-wafer standard deviation of 30% of the within wafer standard deviation and a normal probability plot (on the right) of the same data.

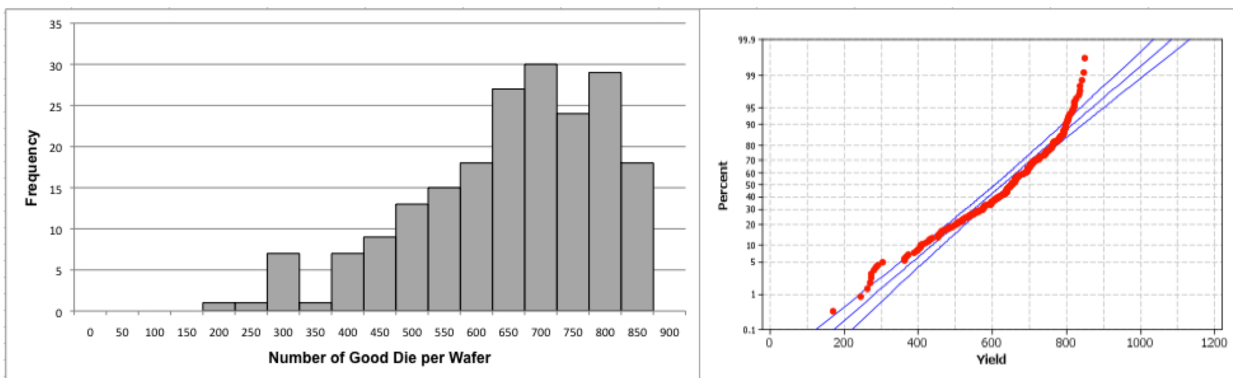

Fig. 3. A yield curve of simulated data showing the number of good die per wafer for lot-to-lot standard deviation of 20% of the within wafer standard deviation and wafer-to-wafer standard deviation of 60% of the within wafer standard deviation and a normal probability plot of the same data.

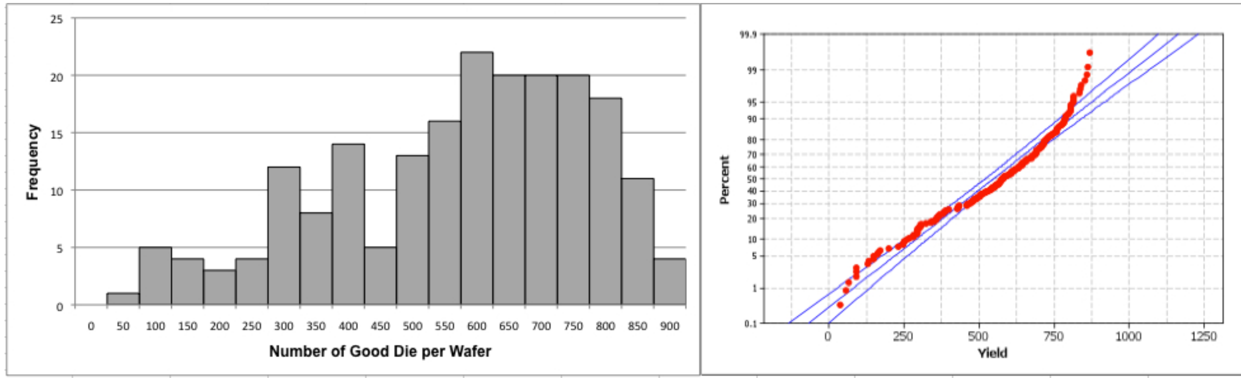

Fig. 4. A yield curve of simulated data showing the number of good die per wafer for lot-to-lot standard deviation of 30% of the within wafer standard deviation and wafer-to-wafer standard deviation of 90% of the within wafer standard deviation and a normal probability plot of the same data.

4. DISCUSSION AND RECOMMENDATION

It is difficult to completely separate with certainty random and systematic loss [3, p. 13]. The skewed distributions of the yields in Fig. 2, 3, and 4 are a cautionary note on assuming that the lower tail of a yield curve contains yields of die subjected to systematic loss. If the tail has some structure such as a large minor mode or if there is a history of yield curves with known random and systematic loss patterns, a reasonable hypothesis might be that the systematic losses are responsible for some of the lower tail [3], [4], [13, p. 3], [15, p. 134].

There are almost always some systematic effects, such as edge effects that arise from the geometry of the process. Those regions of the wafer might be treated separately, since they have well known and distinct problems.

We have treated the decomposition into random and systematic loss as a quality control problem in which the random loss is the baseline. We have investigated this baseline in order to make it less likely that false alarms will be created by yield curves with appearances like those in Figures 1 to 4 and by their accompanying statistics such as skewness.

A yield curve alone can supply only limited information. However, yield curves are commonly used and can be helpful for obtaining some knowledge about the process. As we have shown, practitioners should not expect the baseline yield to be symmetric but can model it as we have done in the simulation, using their own information about their processes.

ACKNOWLEDGMENT

The authors thank Mel Effron of Yield Systems Inc. and Dr. Robert Parody of the Center for Quality and Applied Statistics at Rochester Institute of Technology for many helpful discussions. We would especially thank the SAS Institute Inc., whose generous gift of the software package JMP greatly facilitated the simulations and calculations for this paper.

REFERENCES

[1] M. Baron, A. Takken, E. Yashchin, and M. Lanzerotti, "Modeling and forecasting of defect-limited yield in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 4, pp. 614–624, Nov. 2008.

[2] N. Kumar, K. Kennedy, K. Gildersleeve, R. Abelson, C. M. Mastrangelo, and D. C. Montgomery, "A review of yield modelling techniques for semiconductor manufacturing," *Int. J. Prod. Res.*, vol. 44, no. 23, pp. 5019–5036, Dec. 2006.

[3] R. C. Leachman and C. N. Berglund, "Systematic mechanisms-limited yield assessment survey," Competitive Semiconductor Manufacturing Program, Univ. California, Berkeley, Tech. Rep. CSM-53, 66 pages, Sept. 30, 2003.

[4] A. I. Mirza, "Spatial yield modeling for semiconductor wafers," M.S. thesis, Dept. Elect. Eng. and Comput. Sci., MIT, Cambridge, MA, May 1995.

[5] B. Pinto, "Process control investments support technology, cut costs," *Semicond. Int.*, vol. 28, no. 13, pp. 46–51, Dec. 2005.

[6] R. E. Langford and J. J. Liou, "Negative binomial yield model parameter extraction using wafer probe bin map data," *Proc. IEEE Electron. Device Meeting*, Hong Kong, 1998, pp.130–133.

[7] B. T. Murphy, "Cost-size optima of monolithic integrated circuits," *Proc. of the IEEE*, vol. 52, no. 12, pp. 1537–1545, Dec. 1964.

[8] C. H. Stapper, "On Murphy's yield integral," *IEEE Trans. Semicond. Manuf.*, vol. 4, no. 4, pp. 294–298, Nov. 1991.

[9] R. V. Hogg, J. W. McKean, and A. T. Craig, *Introduction to Mathematical Statistics*, 6th ed., Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[10] C. H. Stapper, F. M. Armstrong, and K. Saji, "Integrated circuit yield statistics," *Proc. of the IEEE,* vol. 71, no. 4, pp. 453–470, Apr. 1983.

[11] S. L. Albin, S. and D. J. Friedman, "The impact of clustered defect distributions in IC fabrication," *Manage. Sci*., vol. 35, no. 9, pp. 1066–1078, Sept. 1989.

[12] C. H. Stapper, "The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions," *IBM J. of Res. and Develop*., vol. 29, no. 1, 87–97, Jan. 1985.

[13] J. Segal, L. Milor, and Y. Peng. (2000, Jan.). Defect and yield analysis: Reducing baseline defect density through modeling random defect-limited yield. *MICRO Magazine*.com, 20 pages. Available: http://www.micromagazine.com/archive/00/01/segal.html

[14] C. H. Stapper, "Large-area fault clusters and fault tolerance in VLSI circuits: A review," *IBM J. of Res. and Develop*., vol. 33, no. 2, 162–173, Mar. 1989.

[15] A. Y. Wong, "A statistical parametric and probe yield analysis methodology," *Proc. IEEE Int. Symp. Defect and Fault Tolerance in VLSI Syst.*, 1996, pp. 131–139.

[16] [16] R. Ross and N. Atchison, "Numerical analysis of equipment yield variance Y_E," *TI Tech. J.*, vol. 15, pp. 97–103, Oct.-Dec. 1998.

[17] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: Wiley, 1978.

[18] D. C. Montgomery, *Design and Analysis of Experiments*, 6th ed., New York: Wiley, 2005.