# A statistical approach to quality control of non-normal lithographical overlay distributions

by R. M. Booth, Jr.
K. A. Tallman
T. J. Wiltshire
P. L. Yee

**To achieve the high reliability and performance required by integrated circuit (IC) chips in IBM Enterprise System/9000™ processors, lithography tool centerline overlay variations between masking levels were specified at ±0.3 μm, and circuit design images were transferred with 5× step-and-repeat photolithography tools. In contrast to data obtained from 1× lithography tools, the level-to-level overlay data which characterize deviations from circuit design rules did not fit a normal distribution, and quality control was not achieved with traditional statistical procedures. A methodology was empirically developed which transformed measured data into worst-case overlay points and approximated the data by a gamma distribution. More than 80% of the worst-case distributions were fit by the gamma distribution. The transformation of chip worst-case overlay data and the quality control testing applicable to 5× step-and-repeat lithography tool processes are described in this paper.**

## Introduction

Integrated circuits in IBM Enterprise System/9000™ (ES/9000™) processors must meet high reliability and performance requirements. *High reliability* implies designing large devices with increased space between the devices to prevent failure even with the widest process variations, whereas *high performance* implies designing small devices with decreased spaces to reduce parasitic resistance and capacitance components. During an integrated circuit design phase, estimates are made of manufacturing process variations, and a set of design rules [1] are selected that achieve the highest circuit

$BA_{CL}$ = $B$ to $A$ centerline overlay

$A_{IM}$ = $A$ image size change per shape

$B_{IM}$ = $B$ image size change per shape

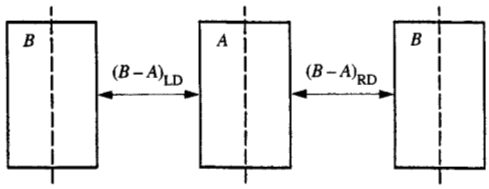
$(B-A)_{LD}$ = Left $B$ space $A$ as designed

$(B-A)_{RD}$ = Right $B$ space $A$ as designed

$(B-A)_{LW}$ = Left $B$ space $A$ on wafer

$(B-A)_{RW}$ = Right $B$ space $A$ on wafer

$(B-A)_{EE}$ = Edge-to-edge variation for $B$ space $A$




$$(B-A)_{LW} = (B-A)_{LD} + \left(-BA_{CL} - \frac{A_{IM}}{2} - \frac{B_{IM}}{2}\right)$$

$$(B-A)_{RW} = (B-A)_{RD} + \left(+BA_{CL} - \frac{A_{IM}}{2} - \frac{B_{IM}}{2}\right)$$

$$(B-A)_{EE} = \pm BA_{CL} - \frac{A_{IM}}{2} - \frac{B_{IM}}{2}$$

**Figure 1**

Edge-to-edge components for a spacing design rule derived from centerline overlay and image size variations between designed and actual size images.

performance and still meet reliability requirements. The design rules define the minimum design distances allowed between various manufacturing masking levels. The difference in alignment between the designed and actual locations of centers of pattern shapes from those in a previous level is defined as a centerline overlay. A centerline overlay tolerance is incorporated into any design rule between two lithography levels.

Earlier chip processes for IBM 3090™ systems used 1× full-field photolithography. Mask and pattern images were the same size, and an entire 125-mm-diameter wafer was exposed during a single exposure. Centerline overlay data obtained from the 1× lithography process were normally distributed. A normal distribution of centerline overlay tolerances simplified design tolerance modeling and allowed product specifications to be maintained using standard statistical quality control methods such as mean ($\mu$) and standard deviation ($\sigma$) control charts, Shewart control charts, and cusum charts [2, 3].

To meet the reliability, density, and performance requirements of bipolar integrated circuits for ES/9000 processors, chips were designed with a centerline overlay tolerance of 0.3 $\mu$m ($3\sigma$). A step-and-repeat lithography tool, in which mask images were five times the printed wafer image size, was used to achieve these tolerances. The mask image was transferred to the wafer in small sections or fields, each of which consisted of four chips. Centerline overlay data obtained from a 5× step-and-repeat lithography process were not normally distributed, and accurate manufacturing pass/fail decisions could not be obtained with standard quality control methods.

A quality control methodology was therefore required to select products with centerline overlay tolerances that are not normally distributed but that conform to design assumptions. The design rule assumptions and requirements are reviewed in the next section. The lithography process overlay data are then presented, and the methodology which ensures that individual manufacturing wafer lots conform to the design rule assumptions is described.

## Design rule requirements

The set of minimum allowable dimensions required to limit interactions between integrated circuit design shapes are the design rules. Single-level design rules specify minimum design shape sizes and spaces for a given design level. Overlay design rules specify minimum design spacing, intersection, or overlap between two different design levels. Single-level and overlay design rules are established to prevent chip reliability or yield failures resulting from variations in manufacturing processes.

Overlay rules are specified as a minimum distance from the edge of a shape on one design level to the edge of a shape on a second design level. An overlay design rule prevents a failure by incorporating the maximum expected manufacturing process variation or tolerance into the design. Differences between the edges of two design levels are described by the centerline overlay and image size variations indicated in **Figure 1**. As noted earlier, the centerline overlay variation is the difference between the designed and actual location of the current level design shapes in relation to the previous level design shapes; it is measured between the shape centers. The image size variation is the difference between the designed and manufactured size of the shapes on each design level.

The difference between the designed and actual distances separating the edges of two design levels, referred to as the edge-to-edge overlay tolerance, is a function of the centerline overlay and image size tolerances [4] (Figure 1). Since $\sigma$ of a distribution is a statistical sum of the $\sigma$s of the components, an edge-to-edge overlay distribution is determined if each overlay component is normally distributed with a known $\sigma$ and $\mu$ of zero. A design rule tolerance is selected from an edge-to-edge distribution as a failure rate objective. If a

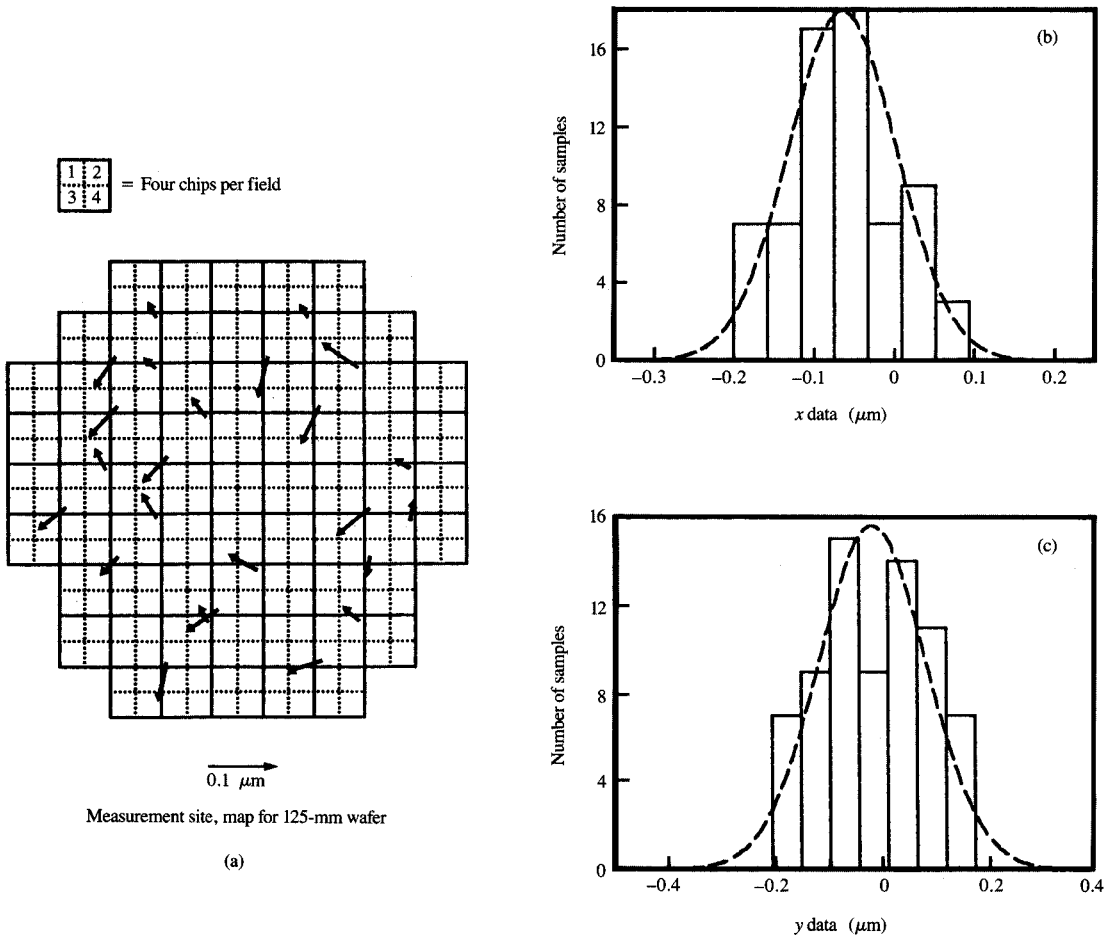



**Figure 2**

Typical distribution of measured overlay data. The map (a) shows the measurement site locations; (b) and (c) are, respectively, histograms for the $x$-component and $y$-component data with normal distribution fits to the data (dashed curves).

design rule tolerance at $\pm 3\sigma$ is selected, the failure rate probability is 0.0027.

Values for an overlay design rule are obtained by adding the edge-to-edge tolerance between two masking levels and the minimum spacing, intersection, or overlap that is needed to prevent a specific failure mode [5]. To meet a specified failure rate objective, each component tolerance must be within the design rule assumptions. If the measured centerline overlay and image size data are normally distributed, standard statistical control methods are used to judge the quality of the data. A manufacturing wafer lot is acceptable if the calculated $\sigma$ and $\mu$ of the lot are within specified limits. A modified methodology is required for data which are not normally distributed.

## Centerline overlay measurement data

Centerline overlay data obtained from measurements on three wafers were used to determine whether a wafer lot conformed to design rule assumptions. Fields on the wafers were exposed, and $x$ and $y$ component vectors of centerline overlays at twelve fields per wafer were measured on an $x$–$y$ stepper with an automated optical measurement tool. Measuring three times as many fields within a wafer showed only marginal changes in the centerline overlay distribution. Wafer-to-wafer overlay variations were negligible compared to overlay variations within a wafer.

Typical centerline overlay sampling data from one manufacturing lot are shown in **Figure 2**. The vectors,

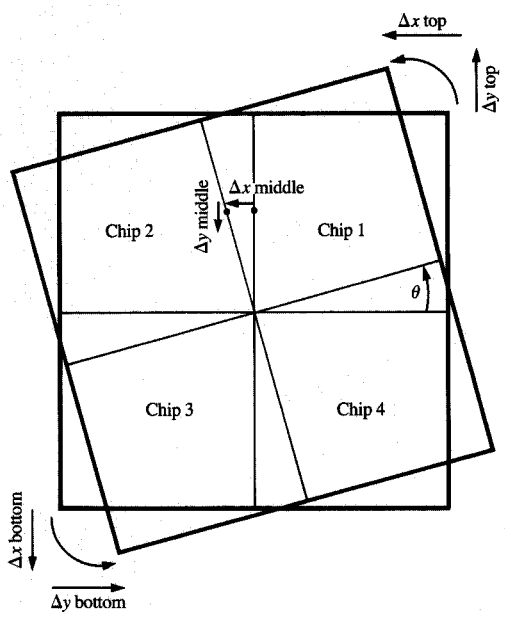



**Figure 3**

Effect of rotation on field overlay data.

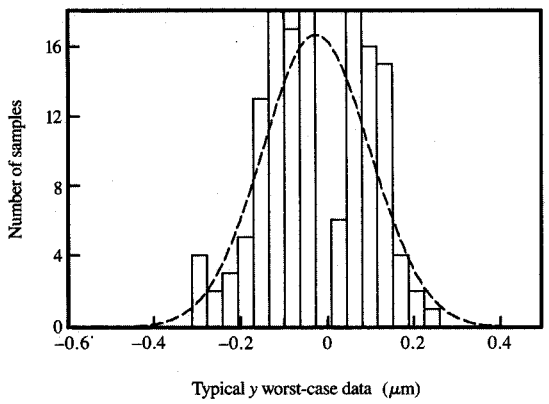



**Figure 4**

Histogram of a typical distribution of worst-case, bimodal $y$-component data and their fit to a normal distribution. A $\chi^2$ goodness-of-fit test yields $\chi^2 = 0$. (Data are not well modeled by the normal distribution for $\chi^2 < 0.05$.)

obtained from two measurements per site in the twelve fields within each wafer, correspond to a three-wafer average of $x$- and $y$-component offset data at each measurement site. Histograms of the $x$ and $y$ data are shown in Figure 2. The data fit to a normal distribution model is indicated by the dashed curves. The quality of this wafer lot was obtained from comparisons of $|\mu| + 3\sigma$ data with design specifications. For the $y$ data, the $|\mu| + 3\sigma = 0.32$ $\mu$m exceeded the 0.3-$\mu$m specification, and the lot was rejected. However, the maximum deviation from zero of data obtained from this lot was 0.21 $\mu$m. Differences $\geq 0.1$ $\mu$m were observed between the $|\mu| + 3\sigma$ and the maximum measurement in $x$ and $y$ data on many manufacturing lots. Many lots were rejected on the basis of the large $|\mu| + 3\sigma$, even though the maximum deviations were within specifications. The high wafer rejection rate forced a reexamination of the methodology used to determine the overlay quality.

Often overlay data distributions, such as the $y$ component data in Figure 2, are bimodal. The cause for bimodality is illustrated in **Figure 3**. Bimodal data inflate $\sigma$ and produce an observed difference between $|\mu| + 3\sigma$ and the maximum raw data measurements, as illustrated in Figure 2. When a field is rotated from its nominal position, the $x$ and $y$ data obtained in the upper right and lower left field corners shift in opposite directions. The bimodal nature of the raw measured data depends on the rotational component of overlay (which varies from field to field) at the measured locations within a field.

With field rotation, a centerline overlay measurement at a field corner differs from a measurement made near the field center (Figure 3). Overlay errors vary across a chip and from chip to chip across a field. Data measurement and overlay error analysis from each wafer lot must account for measurement locations within a chip and from chip to chip. The analysis of data in Figure 2, which compares $|\mu| + 3\sigma$ with the design specifications, does not adequately consider measurement locations and incorrectly characterizes manufacturing lot quality.

## Worst-case methodology

The centerline overlay expected at any point within a field was calculated from a limited number of measurements and a normal distribution model that estimated within-chip and chip-to-chip errors. The centerline overlay at every point in the field was calculated using a linear model [6]. Since four measurements are taken per field (two in $x$, two in $y$), the linear model can be reduced to four parameters: $x$ translation, $y$ translation, rotation, and isotropic magnification. The $x$ and $y$ overlay at each point are then given by

$$dx = T_x - \theta y + Mx,$$
$$dy = T_y + \theta x + My,$$

where

$dx = x$ centerline overlay measurement,
$dy = y$ centerline overlay measurement,

$T_x$ = $x$ translation,
$T_y$ = $y$ translation,
$\theta$ = rotation,
$M$ = isotropic magnification,
$x$ = $x$ coordinate of measurement site,
$y$ = $y$ coordinate of measurement site.

Values of $T_x$, $T_y$, $\theta$, and $M$ are calculated for each field, and $dx$ and $dy$ are calculated for any $(x, y)$ location within a field.

Since centerline overlay varies across a chip, a "good fields rule" proposed by Arnold [7] was adopted to characterize the overlay for each chip. The worst-point (farthest from zero) overlay data were selected within each chip. The chip overlay data were thereby given an upper bound. A worst-point methodology is consistent with a design rule that implicitly assumes that the overlay is within specifications across an entire chip [8].

Typical worst-case $y$-component offset data are shown in **Figure 4**. Each data point in the histogram is a calculated worst-case overlay deviation in the $y$-direction for one chip. The distribution is bimodal, since all points within a chip need zero overlay offsets ("perfect" alignment) for the worst-case overlay data to be located at zero. A bimodal distribution is generated by rotation ($\theta$) and magnification ($M$) errors, which cause chips within the same field to shift in opposite directions (Figure 3). As indicated in Figure 4, a fitted normal curve, a $\chi^2$ goodness-of-fit test [9], and a comparison of the $|\mu| + 3\sigma$ data to the minimum and maximum points indicate that the worst-case $y$-component data from Figure 2(c) are not well modeled by a normal distribution.

The positive and negative worst-case data of the bimodal distribution shown in Figure 4 are reproduced in separate histograms in **Figure 5**. Each histogram is compared with normal and gamma [10, 11] distribution curves of the absolute values of the data. A $\chi^2$ goodness-of-fit test was performed to determine which model provided the best fit to the worst-case data. In Figure 5, the normal and gamma distributions were not rejected for the positive overlay data. However, the normal model was rejected for the negative overlay data. Worst-point data obtained from more than 100 manufacturing wafer lots were selected over a one-year period from different process levels (including device and metallization layers) and exposed on 5× step-and-repeat lithography tools from two different manufacturers. More than 80% of the worst-case data distributions were fit by a gamma distribution to a significance level of at least 0.05. No other distribution (including normal, log normal, beta, and Weibull) provided a comparable fit to these data.

The worst-case data from nine manufacturing wafer lots are plotted in histogram form in **Figure 6** with gamma and normal curves superimposed. These data illustrate the tendency of many lots of worst-point data to be well

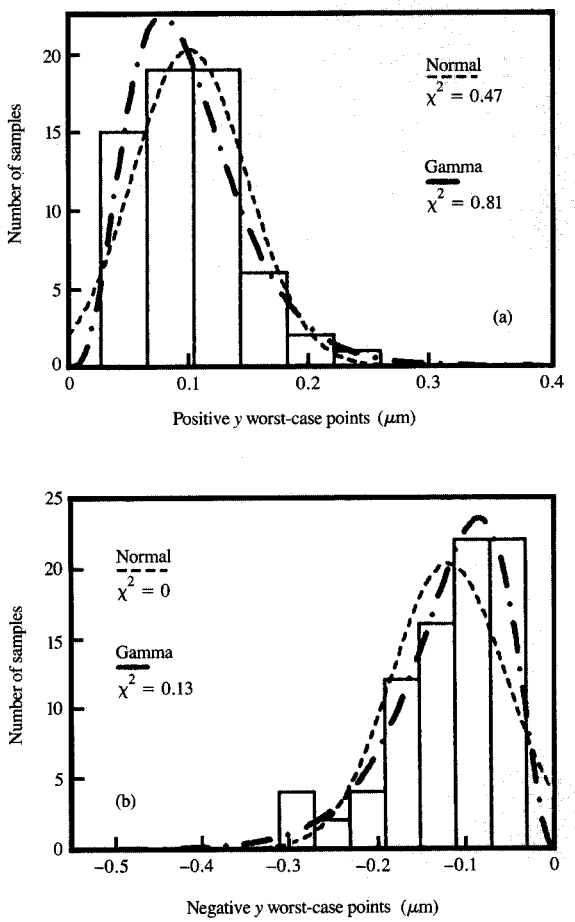



**Figure 5**

Histograms of worst-case bimodal $y$-data from Figure 4 separated into (a) positive and (b) negative modes. Results of $\chi^2$ goodness-of-fit tests for normal and gamma distributions are shown.

modeled by a gamma distribution. On the basis of statistics of extreme values chosen from a normal distribution, this result is not unexpected*. The empirical evidence generated by examining worst-case data from many manufacturing lots is sufficient to justify the use of the gamma model to determine the quality of overlay data.

Once a centerline overlay distribution of worst-case points is generated, a method must be identified for determining wafer lot quality from this distribution. Since design rules are specifically concerned with overlay and image size variations, an edge-to-edge distribution is calculated to evaluate lot quality. An edge-to-edge

*C. Abraham, IBM Research Division, Yorktown Heights, New York, private communication.

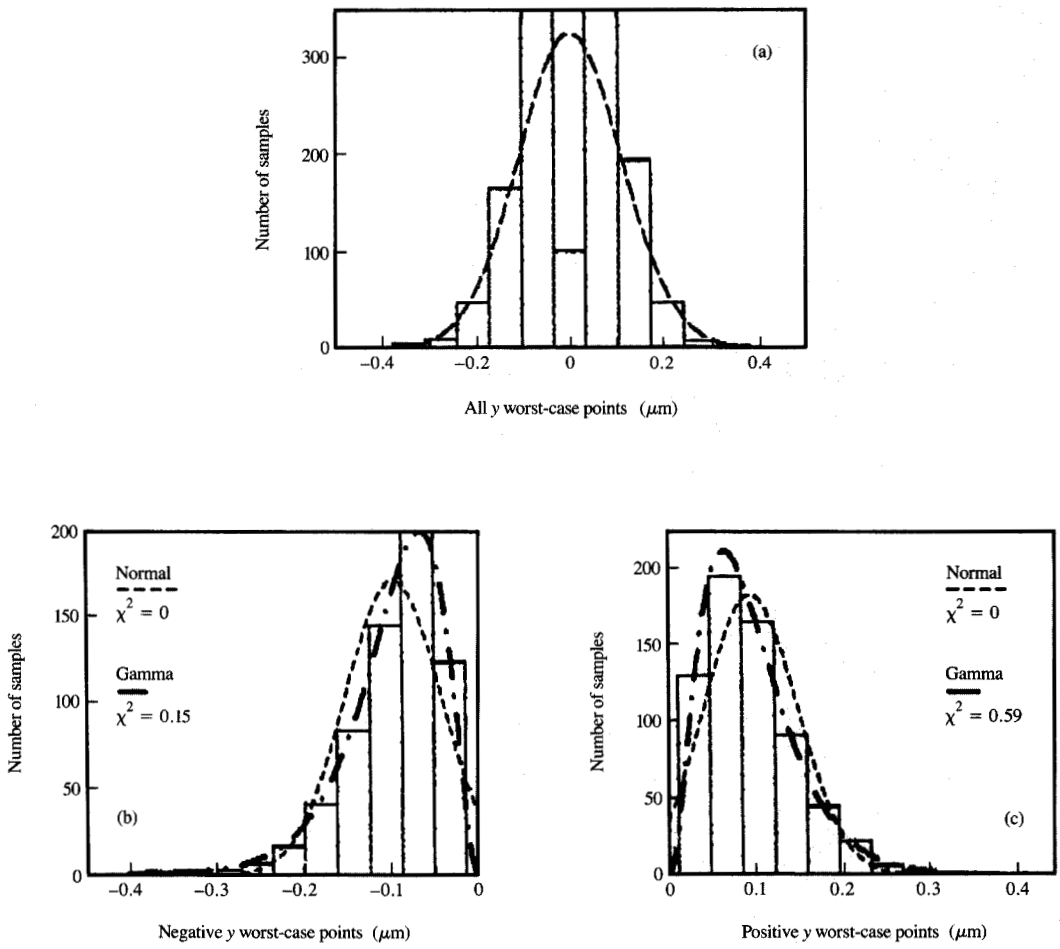



**Figure 6**

Histogram of a bimodal distribution of $y$-component data points obtained from nine wafer lots is shown in (a). Normal and gamma distributions are fitted to the (b) positive and (c) negative mode data.

distribution component is a linear sum of the centerline and image components (Figure 1). In the example illustrated in **Figure 7**, the centerline [Figure 7(a)] and image size probability densities [Figure 7(b)] are convoluted [12] to generate an edge-to-edge distribution [Figure 7(c)]. The image size data are assumed to be normally distributed across a wafer lot (approximately true for optical imaging systems). The probability fraction of points in the edge-to-edge distribution that exceed the specification limit [Figure 7(c), darkened regions] is obtained by numerical integration. The convolution and integration are performed using a Gaussian quadrature method [13].

Tests were made to demonstrate that the failure rates determined from a numerical integration/convolution program to obtain an edge-to-edge distribution are accurate. A Monte Carlo program generated 10 000 points of worst-case overlay data from random values for field offsets, rotations, and magnifications obtained from wafer data analyses. Components were simulated for each field, and worst-case points in the fields were calculated. The simulated centerline overlay worst-case data are plotted in the histogram shown in **Figure 8(a).**

The mean and standard deviations for the positive and negative overlay data were computed and used to derive the $\alpha$ and $\beta$ parameters of gamma distributions for each set

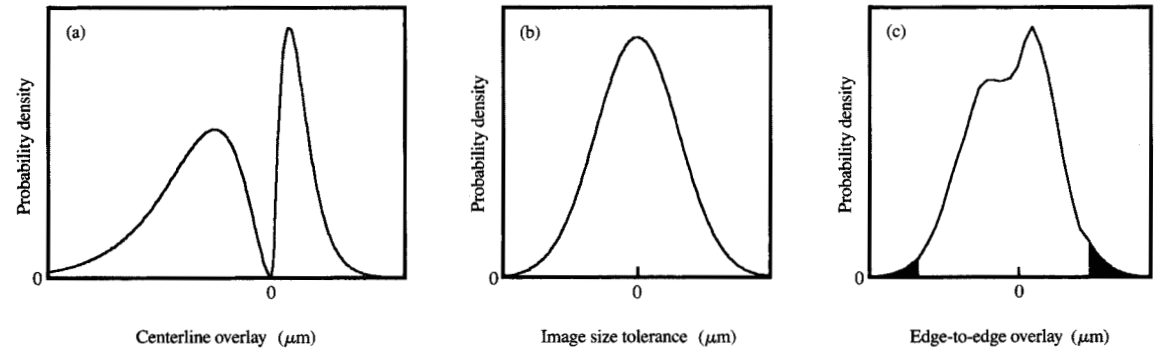



**Figure 7**

Convolution of centerline overlay (a) and image size (b) distribution curves to illustrate the generation of an edge-to-edge distribution (c). The centerline overlay distribution (a) is assumed bimodal, with each mode described by a gamma distribution. The image size distribution (b) is assumed normal. The darkened area in (c) corresponds to the population that exceeds a design tolerance of $\pm 3\sigma$.

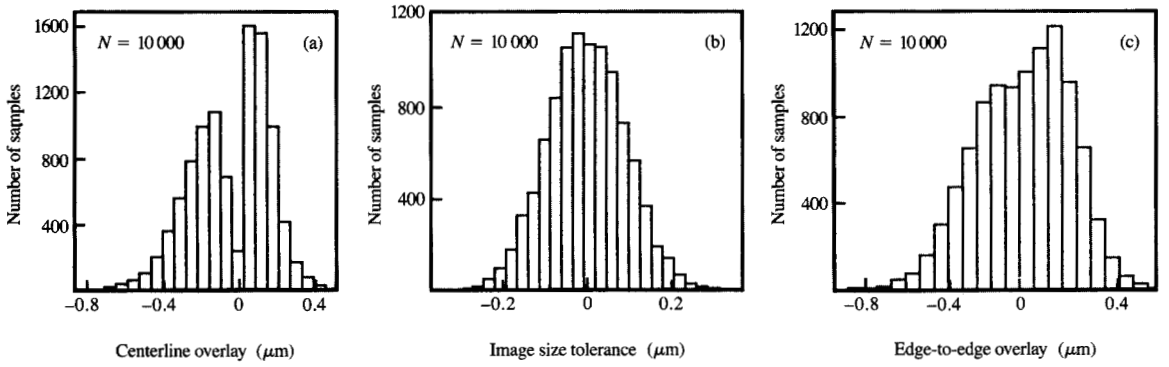



**Figure 8**

Histograms of (a) worst-case overlay data, (b) image size data, and (c) resulting edge-to-edge distribution from randomly generated data.

of data [14]. A simulated set of image size variation data was also generated using a normal distribution assumption **[Figure 8(b)]**. The two data sets were used to calculate the edge-to-edge overlay values plotted in histogram form in **Figure 8(c)**. The edge-to-edge values greater than specified by the design rules were compared to failure rates obtained using the numerical convolution/integration program. The simulated failure rate of 0.0307 compared favorably with the numerical program result of 0.0320. The simulation and numerical program results typically differ by less than 5%, which demonstrates that failure rates can be calculated from the numerical program.

To characterize manufacturing lot quality with this method, a desired lot-by-lot failure fraction is selected. The failure fraction is specified on the basis of the product requirements at each level. A metal layer may require a very small failure fraction specification to avoid possible reliability failure modes. Larger failure fractions are used for other layers in which chip reliability is less sensitive to the mask alignment. Examples of processing steps in which the rejection level was increased from 0.0027 to 0.0124 are shown in **Figures 9** and **10**. Manufacturing lots with failure fractions greater than this specification are reworked or scrapped, while lots with a smaller failure

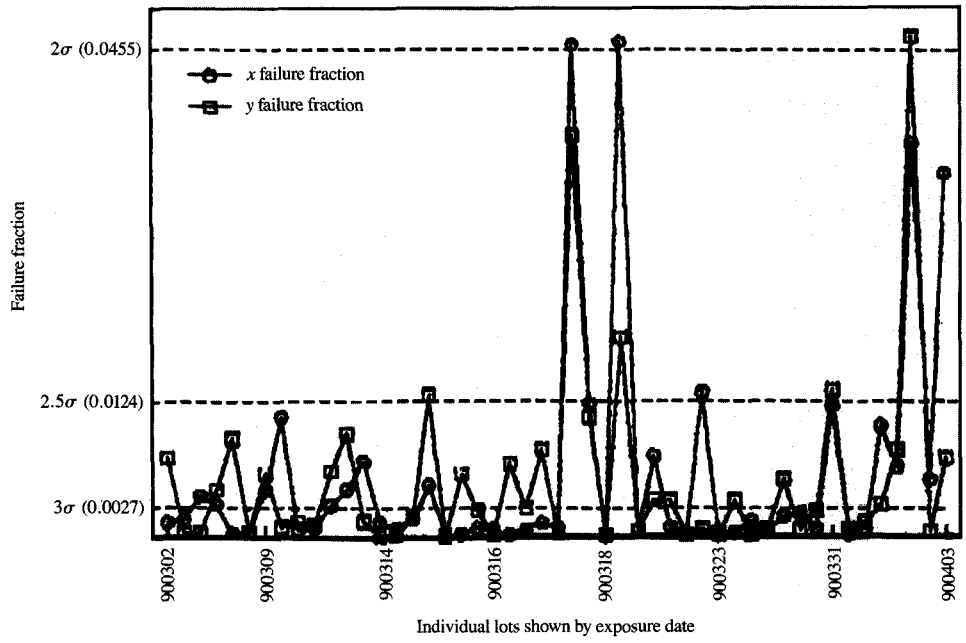



**Figure 9**

Wafer lot failure fractions; $x$- or $y$-component failure fractions greater than 0.0124 were rejected for this processing level. Typical failure fractions (0.0027, 0.0124, and 0.0455) and their normal distribution equivalents ($\pm 3\sigma$, $\pm 2.5\sigma$, $\pm 2\sigma$) are shown.

fraction are passed. A typical failure fraction specification is 0.0027, which is equivalent to the fraction of data points in the tails of a normal distribution that exceeds $\pm 3\sigma$.

A large number of wafer lot failure fractions indicates that there is significant lot-to-lot variation. As indicated in Figure 9, lots with large failure fractions generally have excessive average $x$ or $y$ offset or field rotation. Calculated worst-case points are distributed further from zero, and the failure fraction is large. Some lots were repatterned and reexposed using computed offsets, and failure fractions were greatly reduced. Other lots were discarded after a second exposure showed insufficient improvement.

To identify trends in edge-to-edge failure fractions, the lot failure fractions were regraphed as shown in Figure 10. Each point represents a "running" average of the lot failure fractions; the graph illustrates overlay trends. In practice, a specification limit is placed on the average failure fraction. Should the average exceed the specification, an overlay degradation is indicated, and actions are initiated to correct this condition. The average is calculated from a maximum number of previous lots, where the number is based on processing line loading (number of lots through the lithography sector for a given period of time). A sensitivity to changing trends in edge-to-edge overlay data is enhanced by replacing the oldest data in a fixed number of wafer lots with data from the newest wafer lots.

Using lot-by-lot and running average specifications, an adequate level of statistical quality control of the lithographical overlay is achieved. The lot-by-lot failure fraction is analogous to specifying a standard deviation of a normal distribution. Monitoring variations in the average lot failure fraction allows rapid detection of degrading trends in overlay performance (analogous to Shewart or cusum control charts commonly used for normal distributions). Consequently, conformity to design rule specifications is ensured, and improvements in overlay performance are monitored and quantified.

## Summary

Lithographic overlay data, which characterize the difference in placement of designed and actual wafer images, were not normally distributed after exposure with 5× step-and-repeat lithography tools, and quality control could not be achieved with traditional statistical procedures. A methodology to evaluate device patterning of IC chips for use in ES/9000 systems is described which

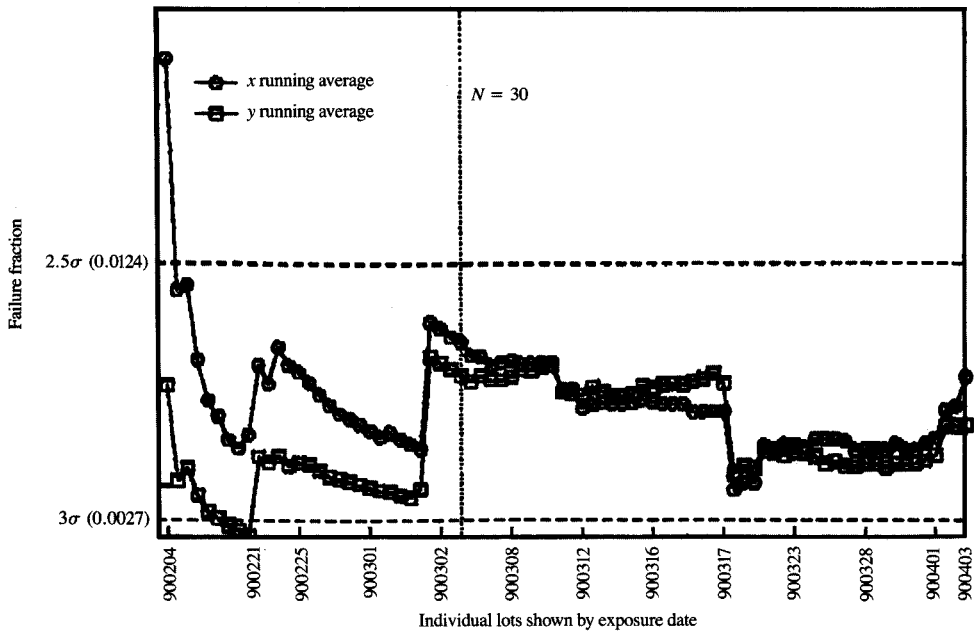



**Figure 10**

Running averages of wafer lot failure fractions; $x$- or $y$-component failure fractions greater than 0.0124 were rejected.

accurately reflected the design assumptions and the nature of the lithography variations. Measured data were transformed into worst-case overlay points. More than 80% of the worst-case data could be fit by a gamma distribution to a $\chi^2$ significance level $\geq 0.05$. Centerline overlay failure rates analyzed from gamma distributions provided a basis for quality control. Improvements in pattern quality due to incremental lithography process enhancements were monitored, and deviations from specifications were readily observed and corrected. Although the transformation and analysis of the experimental overlay data to gamma distributions of worst-case data were computationally intensive, these procedures were necessary to ensure the IC chip quality required for the ES/9000 processors.

## Acknowledgments

The authors wish to thank Steve Bandes for his help with the statistics calculations, and Kirk Kamberg and Phyllis Ng for providing all the overlay data anybody would ever want to look at!



## References

1. S. M. Sze, *VLSI Technology*, McGraw-Hill Book Co., Inc., New York, 1983, pp. 491–494.
2. A. J. Duncan, *Quality Control and Industrial Statistics*, Richard D. Irwin, Inc., New York, 1965.
3. E. L. Grant and R. S. Leavenworth, *Statistical Quality Control*, McGraw-Hill Book Co., Inc., New York, 1988.
4. C. Beck, "Design Rule Verification," *Proceedings of the KODAK Microelectronics Seminar INTERFACE '83*, November 1983, pp. 25–28.
5. W. C. Heavlin and C. Beck, "On Yield Optimizing Design Rules," *IEEE Circuits & Devices Mag.* **1**, 7–12 (1985).
6. D. S. Perloff, "A Four Point Electrical Measurement Technique for Characterizing Mask Superposition Errors on Semiconductor Wafer," *IEEE J. Solid State Circuits* **SC-13**, 436–444 (1978).
7. W. H. Arnold, "Overlay Simulator for Wafer Steppers," *SPIE Optical/Laser Micro Lithog.* **922**, 94–105 (1988).
8. H. J. Levinson and R. Rice, "Overlay Tolerances for VLSI Using Wafer Steppers," *SPIE Optical/Laser Micro Lithog.* **922**, 82–93 (1988).
9. K. V. Bury, *Statistical Models in Applied Science*, John Wiley & Sons, Inc., New York, 1975, pp. 196–203.
10. *Ibid.*, pp. 299–301.
11. N. L. Johnson and S. Kotz, *Continuous Univariate Distributions—1*, Houghton Mifflin Co., New York, 1970, pp. 166–206.
12. A. Papoulis, *Probability, Random Variable, and Stochastic Processes*, McGraw-Hill Book Co., Inc., New York, 1984, pp. 134–141.

13. R. L. Burden, J. D. Faires, and A. C. Reynolds, *Numerical Analysis*, Prindle, Weber, & Schmidt, New York, 1981, pp. 165–171.
14. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole Publishing Company, New York, 1982, pp. 150–151.

*Received August 18, 1991; accepted for publication January 10, 1992*

**Robert M. Booth, Jr.** *IBM Technology Products, Route 52, East Fishkill facility, Hopewell Junction, New York 12533 (BOOTHR at FSHVMBC).* Mr. Booth is a Senior Engineer in the Advanced Lithography Department of the IBM East Fishkill Semiconductor Laboratory. His work is concerned with the evaluation, selection, and optimization of optical lithography tooling for new production lines. He joined IBM in 1968 after serving in the U.S. Army for three years. Mr. Booth received his B.S. in physics from Rensselaer Polytechnic Institute in 1965. He is a member of the Society of Photo Instrumentation Engineers.

**Kurt A. Tallman** *IBM Technology Products, Route 52, East Fishkill facility, Hopewell Junction, New York 12533 (TALLMAN at FSHVMK1, tallman@fshvmk1.vnet.ibm.com).* Mr. Tallman is an Advisory Engineer at the IBM East Fishkill Development Laboratory. He joined the General Technology Division at the East Fishkill facility in 1983, and has since worked on the development and verification of bipolar product design rules. Mr. Tallman received a B.S. in electrical engineering from Iowa State University in 1983, and an M.S. in electrical engineering from Syracuse University in 1987. He is a member of the Institute of Electrical and Electronics Engineers.

**Timothy J. Wiltshire** *IBM Technology Products, Route 52, East Fishkill facility, Hopewell Junction, New York 12533.* Mr. Wiltshire is an Advisory Engineer in the Advanced Photolithography Manufacturing Engineering Department in the VLSI line at IBM East Fishkill; he is currently on temporary assignment at SEMATECH. He received his B.S. in chemical engineering from Clarkson University (Clarkson College) in Potsdam, New York, in 1982. He subsequently joined IBM in East Fishkill, where he has worked on step and repeat lithography in the laboratory and in manufacturing. In 1983, he received an IBM GTD Outstanding Achievement Award for his work in photolithography. Mr. Wiltshire continued his studies and received his M.S. in applied mathematics from Rensselaer Polytechnic Institute in 1986. In October 1992, he was the leader of the Censor Photo Quality Improvement team that won top honors in the East Fishkill site "BEST of the BEST" competition.

**Pui L. Yee** *IBM Technology Products, Route 52, East Fishkill facility, Hopewell Junction, New York 12533 (YEE at FSHVMFK1).* Ms. Yee is an Advisory Engineer at the IBM East Fishkill Development Laboratory. She received a B.A. in sociology from the University of California at Berkeley in 1981, and a B.S. in electrical engineering from the State University of New York at Stony Brook in 1983. Ms. Yee joined the General Technology Division at the East Fishkill facility in 1983. Her first job involved design of test sites used to characterize bipolar devices. Since 1987, she has been working on the development and verification of bipolar and bi-CMOS design rules. Ms. Yee was the team leader for one of East Fishkill's "BEST of the BEST" Quality Improvement teams on lithography used on chips for the ES/9000 machine. In 1991, she received an IBM GTD Outstanding Achievement Award for her work on the ES/9000 machine.