

Correction of Nonlinear Deflection Distortion in a Direct Exposure Electron-Beam System

Abstract: This paper describes a distortion correction technique, used in a high-throughput electron-beam exposure system, which achieves an absolute deflection accuracy of about 30 parts per million (ppm) throughout a 5-mm field. To accomplish this, a cyclic, numerically controlled magnetic deflection with an accuracy of about 1000 ppm and high repeatability is first established. Horizontal and vertical errors in this deflection are measured by sensing the apparent locations of features in a calibration grid that is placed in the exposure field of the beam. The measured error is smoothed by means of a two-dimensional spline-fitting method, and the parameters for the correction surfaces are calculated. Corrections to position and speed, defined by a set of digital tables, are superimposed on the basic deflection, and the process is iterated until acceptable precision is achieved. Parameters relating registration scans to field-writing scans are then calculated for use in the registration of all written fields.

Introduction

Many investigators have applied electron-beam systems to microfabrication, primarily in laboratory environments [1-5]. The distortion correction system described in this paper is part of an electron-beam lithographic system, EL1, designed to expose semiconductor wafers directly in a high-precision, high-volume manufacturing environment [6]. Thus, additional stringent requirements are imposed on this system. The EL1 exposes patterns that accurately overlie structures generated either by optical means or by other EL1 systems, and requires a deflection system capable of 10000 resolved electron optic lines (i.e., 0.5- μm resolution over a 5-mm field) with a deflection accuracy of 30 parts per million over the entire field.

To achieve such resolution and accuracy, the hardware should approach ideal deflection, and any residual errors have to be highly repeatable so that they can be measured and corrected. The hardware, therefore, is designed and adjusted to obtain a deflection with errors of no more than about 0.1 percent. To reduce this error to 30 ppm, a mapping procedure is applied. The actual deflection is measured against a fixed target of known characteristics; the measurements, correlated against the corresponding ideal coordinates for the deflection, serve as source data for generating the corrections. In designing the correction-generation system, several requirements were identified:

Figure 1 The deflection process.

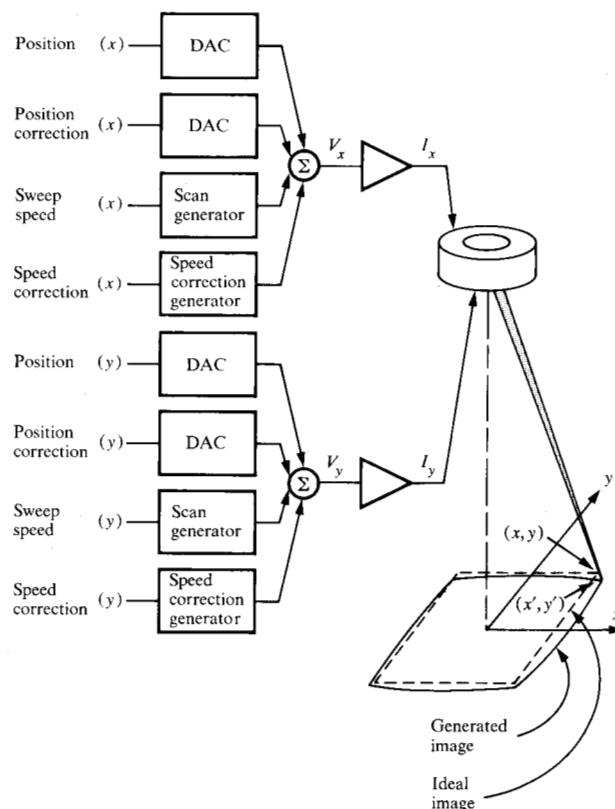


Table 1 Possible sources of deflection error.

<i>Step of deflection chain</i>	<i>Input variables</i>	<i>Output variables</i>	<i>Possible errors</i>
1. Position DACs Sweep speed generation Speed correction	x and y numerical deflection data	Position Position corrections Sweep speed Sweep speed correction	Offset voltage Gain variation Nonlinearities
2. Deflection amplifier (driver)	v_x, v_y	I_x, I_y —deflection currents	Offset currents Gain variation Nonlinearities
3. Deflection yoke	I_x, I_y	(x', y') —coordinates of irradiated image point for ideal object	Variation in deflection sensitivity Yoke-to-object orientation Hysteresis and eddy current effects Nonlinear distortion
4. Target (e.g., wafer) deformation and misalignment in the system	(x', y')	(x_t, y_t) —coordinates of projected image on target	Scale factor variation Displacement error Image distortion

- The collection of field distortion data, computation of corrections, and application of the corrections must be accomplished with a minimum of unproductive time.
- The form of applied corrections must be sufficiently flexible to accommodate a variety of writing field sizes.
- Both the basic deflection cycle and the applied corrections must remain stable for times in excess of a shift of production.

The basic structure within which the correction system operates includes an IBM System/370 computer, a digital control unit (DCU), an analog unit (AU), and an electron optic column. The control computer provides data management, computational power, and high-level supervisory control. The DCU has the master clock for the system and the numerical control tables governing the deflection and the corresponding corrections. The AU contains deflection and correction circuitry to drive the column deflection, as well as signal-conditioning and detection circuitry to sense the location of the beam on the grid. A feedback path is provided from the column through the AU and the DCU to the control computer for measurements of registration and deflection.

Deflection errors

The physical elements controlling the deflection are diagrammed in Fig. 1. The beam is positioned at the starting point of each scan, and then sweeps at nominally constant speed. Digital values for position, sweep speed,

position corrections, and speed corrections determine the desired position of the beam. These values are transformed to currents I_x and I_y , which deflect the beam, causing it to strike the object plane at point (x', y') rather than at (x, y) , the ideal position. Further, since the wafer to be written is misaligned and/or deformed relative to the object plane, the beam strikes a target wafer (not shown) at (x_t, y_t) .

Table 1 identifies some possible sources of error in the deflection. Analysis shows that some of these effects produce linear transformations, which can be minimized by adjusting the hardware during system maintenance. (This off-line maintenance is supported by a variety of computer programs, which direct the adjustment of various manual controls.) Errors due to wafer misalignment, which produce apparent translation, magnification, rotation, and trapezoidal effects, are corrected by a registration process. Registration can also reduce field distortions that appear as such effects, but cannot remove the nonlinear distortions in the deflection field. Such errors are corrected by the procedure to be described in the following section.

Correction process

• Measurement of errors

The repeating deflection pattern has three parts: An A-cycle, during which areas in the four corners of the writing field are scanned; a B-cycle, where a back-and-

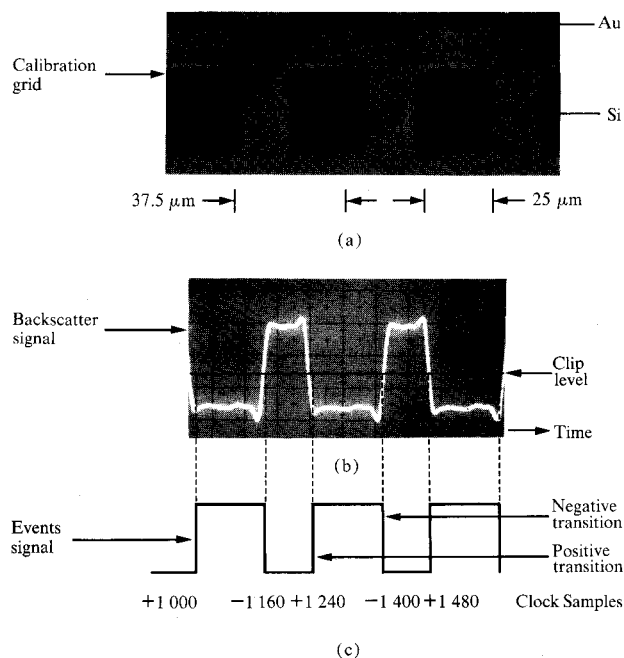


Figure 2 Calibration grid, a representative backscatter signal, and the corresponding events signal.

forth raster-like pattern over the written field is traced; and a C-cycle, a period of time allotted for moving the x-y table from one site to the next, while the deflection scans a path that is occasionally used for focus sensing. The goal of the repeating cycle is to develop a stable history by which the deflection pattern reaches each point in the field. Distortions in the deflection are therefore stable and repeatable. When the system is not actually writing, the beam is blanked but the deflection pattern continues in its fixed cycle to retain stability.

Errors in the deflection scan are determined by periodically scanning a calibration grid with the repeating deflection. The grid, one of several calibration fixtures attached to the x-y table, is made up of an array of square holes in a gold film deposited on a silicon wafer [Fig. 2(a)]. The holes are 25 μm square spaced 37.5 μm apart, center to center. The grid measures about 1 cm square. The configuration of the grid holes represents a compromise among several factors: ensuring that every collected hole is covered by enough scans that errors due to edge roughness become insignificant; ensuring that the distribution of holes is dense enough that all deflection errors can be adequately measured; and achieving a design that can reasonably be fabricated by current manufacturing methods. The grid array need not be perfect, but its characteristics must be known, and are measured by using a set of off-line calibration programs.

For error measurement, a subset of grid holes at points throughout the writing field is selected based on the distortion characteristics of the writing field. About a thousand holes are enough for accurate measurement of distortion in the field scan. The control computer initiates a sequence of programs to move the calibration grid under the beam and to collect data over the selected subset of grid holes.

Three sets of data are collected, one from the A-cycle and two from the B-cycle (one each for horizontal and vertical measurements). For the A-cycle two grid holes in each registration window are scanned. For the horizontal scan in the B-cycle, the writing magnetic deflection is used. The beam is unblanked only over the grid holes selected to be measured. The size of the unblanked collect window is determined by the size of the grid holes and the maximum expected misalignment between the deflection scans and the grid. For the vertical scan in the B-cycle, the writing magnetic deflection is used again, this time with superimposed sawtooth horizontal and vertical electrostatic deflection. The horizontal electrostatic sawtooth deflection causes the beam to dwell periodically along each scan line, while the vertical sawtooth deflection drives the beam vertically at each dwell position. This produces a sequence of "stabs" perpendicular to each scan line and therefore across the upper and lower edges of the holes. The starting positions of the stabs are staggered from line to line to minimize the effect of edge roughness. The superimposed electrostatic deflection makes it possible to measure the vertical coordinates of the grid holes accurately without disrupting the basic magnetic deflection cycle.

When the grid is scanned, the electron backscatter from the gold is greater than that from the silicon; the hole edges are detected from the difference between the two, by using the same diodes that are used for registration. The output signals of these diodes are summed [Fig. 2(b)], and their positive and negative transitions identify the time at which the beam crosses into and out of each grid hole [Fig. 2(c)]. The DCU time-stamps each such transition by sampling the clock, and transmits the samples to the control computer. An editing program reduces the data and determines the hole centers by averaging the times at which the beam crossed the hole edges on successive scan lines. These computed hole centers are compared with a table of times at which the holes should have been traversed by an ideal deflection. In this manner the deflection scan errors, both horizontal and vertical, are determined at the selected array of hole locations across the field.

Software discards data from the edges of holes outside the subset chosen for error measurement, detects and removes some noise-induced data, and identifies malfunctions in the system. Editing the raw data cannot en-

tirely remove error in the measured hole centers. Among the sources of such measurement error are grid characterization error, detector noise, beam jitter, quantization of clock samples, and hole-edge roughness. If the generated correction were based purely on a point-by-point fit of the measured hole centers, the corrections would map the measured data rather than the actual deflection distortion. This would induce random and systematic errors in the corrections, degrading the image. By taking advantage of the low-frequency content of the deflection errors, the spurious effects of measurement noise can be filtered out mathematically. A model of the scan error is established by using a best fit of spline functions to the measured data [7, 8]. Four error surfaces describe deflection distortion: one each for horizontal error, vertical error, forward-going scans, and backward-going scans. Only the horizontal forward error surface will be discussed, since the others are treated similarly.

Let $H(x, y, P)$ be the scan error in the x direction as a function of position in the writing field. Let the horizontal measured hole center errors be K_l , where $l = 1, 2, \dots, N$, and N is the total number of measured holes. The scan error function may then be represented by a two-dimensional spline function of several rectangular meshes. The size and number of meshes is a compromise based on the severity of field distortion and the amount of filtering needed. For EL1, tests showed that a two by three set of meshes, as shown in Fig. 3(a), is satisfactory. In each mesh the scan error is defined by a bicubic polynomial of the following form:

$$H(u, v, P)$$

$$\begin{aligned} = & B_0(u)B_0(v)P_1 + C_0(u)B_0(v)P_2 + B_0(u)C_0(v)P_3 \\ & + B_1(u)B_0(v)P_4 + C_1(u)B_0(v)P_5 + B_1(u)C_0(v)P_6 \\ & + B_0(u)B_1(v)P_7 + C_0(u)B_1(v)P_8 + B_0(u)C_1(v)P_9 \\ & + B_1(u)B_1(v)P_{10} + C_1(u)B_1(v)P_{11} + B_1(u)C_1(v)P_{12}, \end{aligned} \quad (1)$$

where u and v are normalized coordinates of x and y , respectively, running from zero to one within the boundaries of each mesh; P_1 through P_{12} are open parameters; and B_0 , B_1 , C_0 , and C_1 are the following basic functions:

$$\begin{aligned} B_0(w) &= 1 - 3w^2 + 2w^3, \\ B_1(w) &= 3w^2 - 2w^3, \\ C_0(w) &= w - 2w^2 + w^3, \text{ and} \\ C_1(w) &= -w^2 + w^3. \end{aligned} \quad (2)$$

Figure 3(b) gives plots of B_0 , B_1 , C_0 , and C_1 which show that the values and slopes at boundaries $w = 0$ and $w = 1$ are either zero or one. Because of these characteristics, P_1 through P_{12} may be interpreted as follows: P_1 , P_4 , P_7 ,

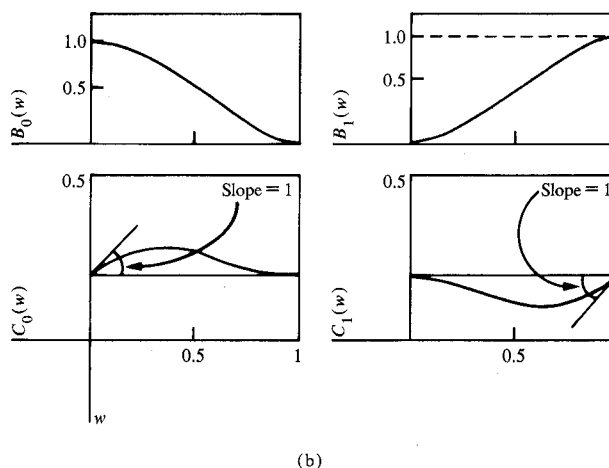
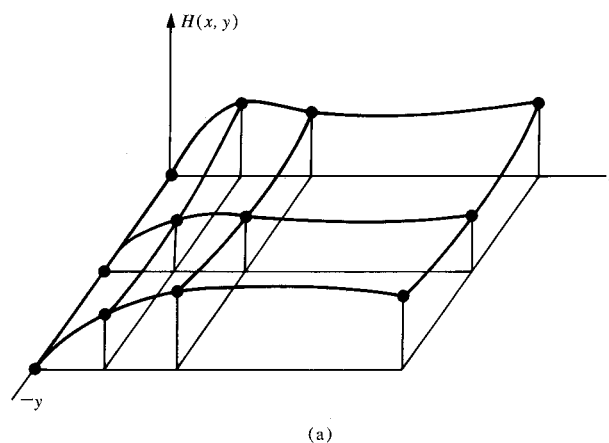


Figure 3 Spline function fitting. (a) Two-dimensional spline function. $N_p = 36$ parameters. (b) Basic functions.

and P_{10} are function values of H at the four corners of the mesh. P_2 , P_5 , P_8 , and P_{11} are the first partial derivatives $\partial H / \partial u$ at the mesh corners. P_3 , P_6 , P_9 , and P_{12} are the first partial derivatives $\partial H / \partial v$.

A deflection in the mesh, then, can be completely determined from the values and slopes at its four corner points. The six meshes are combined to a larger surface by equating values and slopes at the corners. Thus the surface error function and its derivatives become continuous. Thirty-six independent parameters are sufficient to describe this surface of two by three meshes. The objective is to fit the error function for the deflection scan,

$$H(x, y, P_k), \quad k = 1, 2, \dots, N_p,$$

having N_p free parameters, to a set of samples of measured error data K_l , taken at N positions (x_l, y_l) , where $l = 1, 2, \dots, N$.

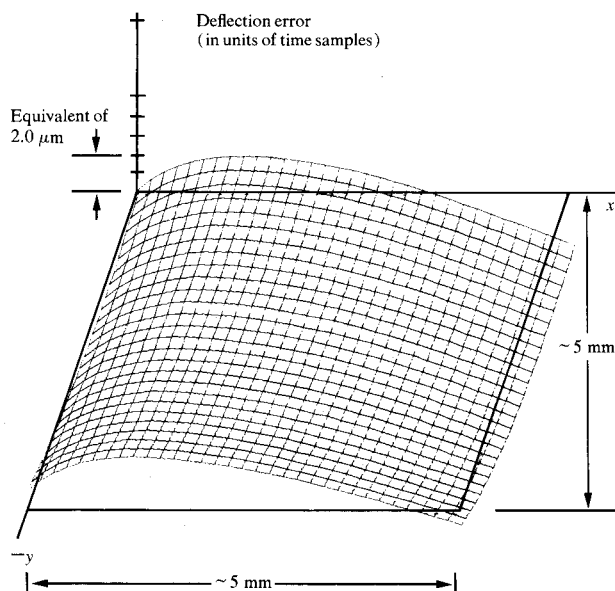


Figure 4 Typical scan errors for EL1.

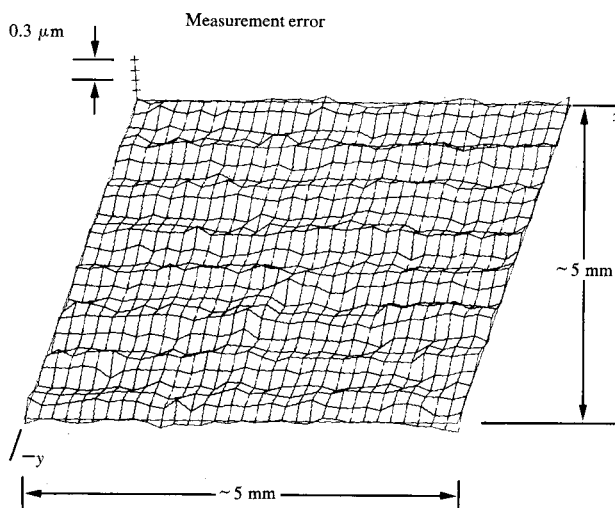


Figure 5 Difference between filtered and unfiltered measured hole centers for horizontal forward scans.

The following equation is satisfied by least-square methods:

$$\sum_{l=1}^N [K_l - H(x_l, y_l, P_k)]^2 = \text{minimum.} \quad (3)$$

By using the spline function in Eq. 1, $H(x, y, P_k)$ can be given the form

$$H_l(x_l, y_l, P_k) = \sum_{k=1}^{N_p} Z_{lk}(x_l, y_l) \cdot P_k, \quad (4)$$

where Z_{lk} is a matrix of N by N_p dimension. Equations (3) and (4) lead to the solution for the parameters P_k :

$$P_k = (K_0)^{-1} \cdot Z_{ml} \cdot K_l, \quad (5)$$

where $K_0 = Z_{ml} \cdot Z_{lk}$, the covariance matrix. Once the values of P_k have been calculated, Eq. (4) can be used to evaluate the smoothed error at any point at the writing field.

Figure 4 is a plot, in units of system clock time samples, of the EL1 scan error without applied corrections. Figure 5 shows the difference between filtered and unfiltered measured hole centers.

• Generating corrections

The process of editing and spline-fitting produces a set of smooth two-dimensional correction surfaces. The corrections required along a given writing line can be approximated by a series of straight lines, giving a continuous, piecewise-linear correction function. The correction at the beginning of the line is subtracted out and treated separately as a position adjustment; the rest of the correction can be viewed as a set of piecewise constant adjustments to the speed of the beam. For a 5-mm field, some 32 such changes in speed, taking place at a set of nonuniformly spaced "breakpoints," suffice to fit the correction function along a scan line.

From one scan line to the next the speed adjustments are approximated by incrementing linearly in the y direction. At 27 to 30 selected scan lines, called breaklines, a new set of increments to the speed adjustments may be applied. This form of approximation makes it possible to construct correction logic that operates independently of the control computer and is not subject to contention for computational or channel resources. Figure 6 shows the basic flow of the speed-correction logic. The correction memory contains a set of 128 speed-correction values, 32 each for horizontal speeds and for vertical speeds, on both forward and backward lines. Each speed correction is read from the correction memory and gated to the appropriate correction register. This process is controlled by a series of programmable synchronizing pulses. Associated with each correction word is an increment, which is added to the correction word before they are both rewritten into the correction memory. During each breakline, new values for every correction increment are read from the new correction increment (NCI) memory. The speed-correction registers feed DACs, which modify the speed of the beam in each axis.

Correction generation software computes both the position and the speed corrections. Thus, it maintains an image of the calculated accumulated corrections throughout the writing field so that errors due to the finite precision of the increment data are not accumulative.

When the calculation is complete, the corrections are transmitted to the DCU during a part of one C-cycle when the deflection memories are inactive. Although the corrections do introduce a change in the magnetic history of the system, this small change is insignificant.

• A-to-B correlation

Ordinarily, each field written by EL1 is registered during the A-cycle and then written during the B-cycle to overlay existing structures on the wafer. The positions of the registration windows must be related to the deflection coordinates for writing. Further, to obtain a good signal-to-noise ratio when the registration marks are sensed, the registration scan is slower than the writing scan [9]. For these reasons a process called A-to-B correlation is necessary to map the registration positions and speeds onto the writing field.

The data collected during the A- and B-cycle deflections are used to map the locations of the holes sensed during the registration (A-cycle) scans, against the locations of those same holes as calculated from the writing (B-cycle) scans. The calculations are based on a mapping of the entire writing field, rather than just a few holes in the corners of the field. This produces a more precise correlation, and also eliminates the need for a separate collect pattern.

For example, consider the horizontal registration window shown in Fig. 7. Two calibration grid holes are scanned by the registration deflection. Their positions, measured relative to the start of the registration scans, are A_1 and A_2 . The expected positions of the same holes with respect to the writing field can be calculated as B_1 and B_2 .

The speed ratio of the registration scans to the writing scans is then given by

$$R = \frac{A_2 - A_1}{B_2 - B_1} \quad (6)$$

The starting location for the horizontal registration window with respect to the writing field, in B scan time units, may be defined as X_s and is given by

$$X_s = \frac{A_1}{R} + B_1 \quad (7)$$

The difference between X_s and its nominal value is then determined. For each registration window, this difference and the speed are saved for use by registration.

• Operational flow

Figure 8 summarizes the operational flow of the correction process. Distortions in the writing field are measured, position and speed corrections are applied, and the distortions are measured again. This process is repeated until the writing is adequately corrected, at which

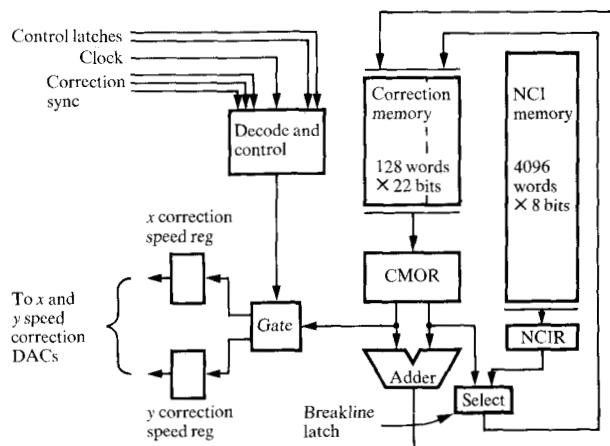


Figure 6 Simplified flow of speed correction logic.

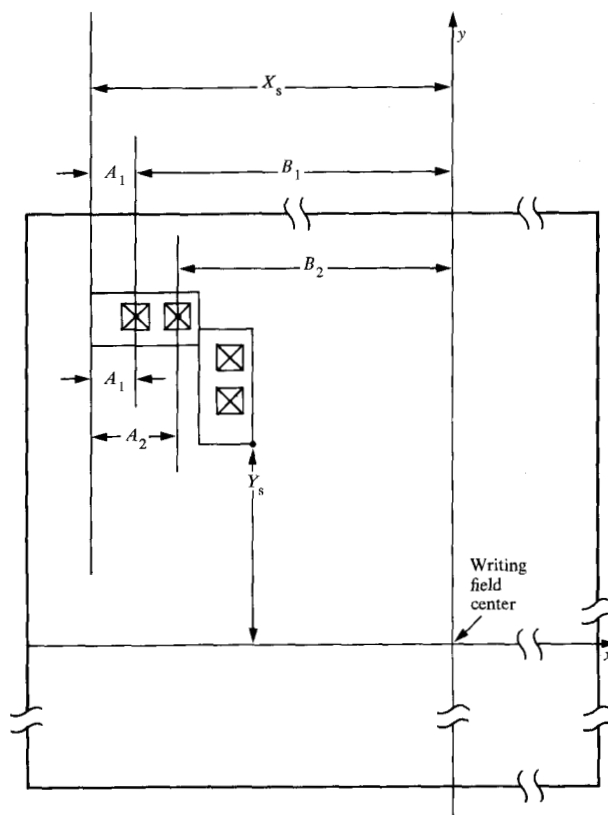


Figure 7 Positional correlation of the registration window to the coordinates for the writing scans. $R = (A_2 - A_1) / (B_2 - B_1)$, and $X_s = (A_1/R) + B_1$.

time the latest A and B collect data are used for A-to-B correlation and the system begins writing wafers. The final results (the deflection tables and the A-to-B correlation data) are retained on disk with the associated deflection pattern.

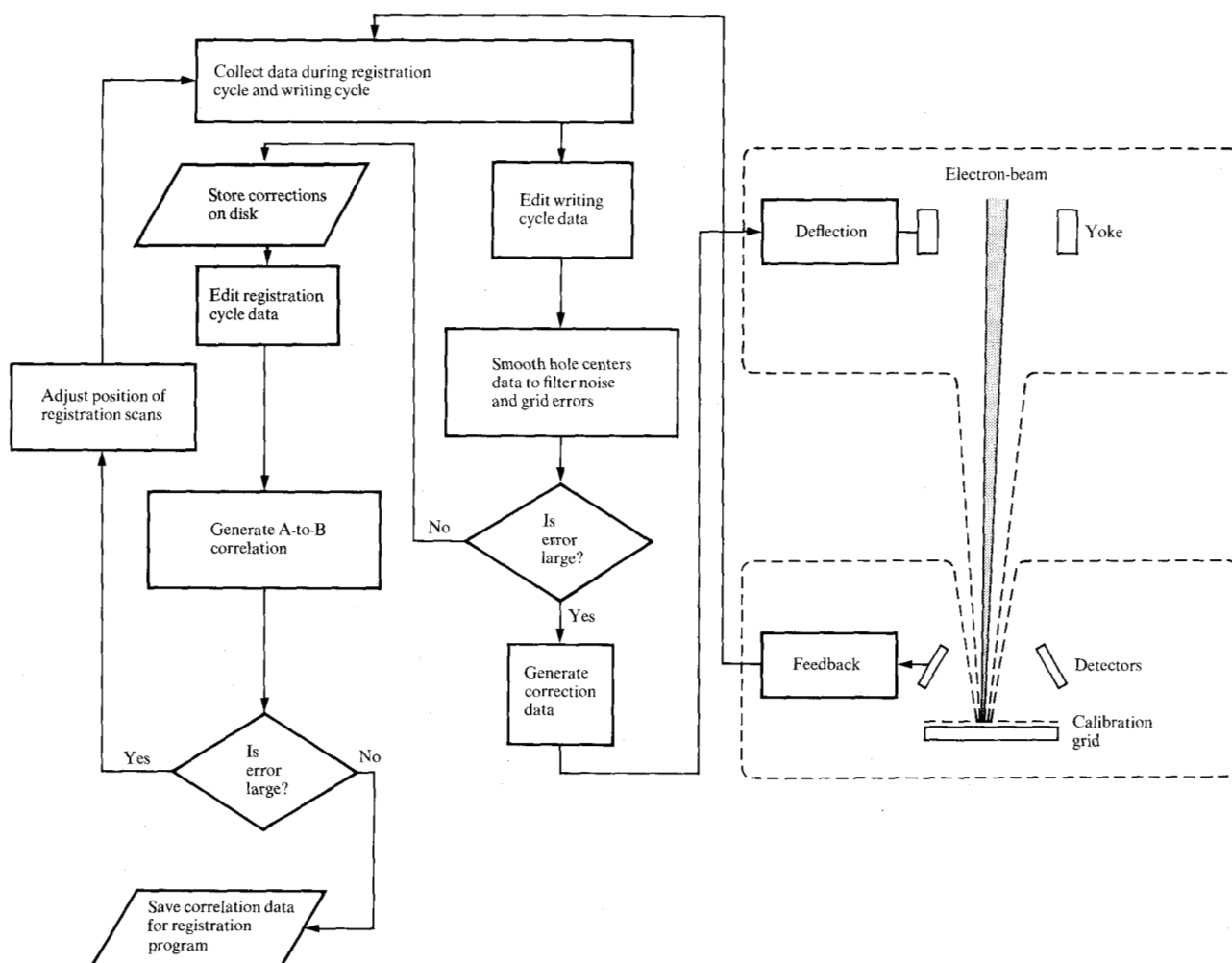


Figure 8 Operational flow of the correction process.

When a different deflection is required, the new deflection and its corrections are retrieved from disk and loaded into the DCU. As soon as the new deflection has stabilized, typically within 30 s, the field is measured for error. Correction updates, if required, are applied before wafer writing starts.

Experience

Once a field has been corrected, the stability of the corrections and the frequency with which the correction and A-to-B correlation processes must be executed become significant considerations in evaluating the throughput and availability of the system. Several studies were conducted on different systems to determine the magnitude and nature of long-term drifts in the scan deflection distortions. Field collects measuring the field deflection error were made over a period of seven weeks. The first data collected were defined as the reference. Comparisons were made between it and collects

taken after the data were modified by different sets of rules. The results were used to evaluate the stability of linear, nonlinear, and A-to-B correlation terms.

To determine the linear drifts (translation, magnification, and nonorthogonal rotation), the collected data for both the forward and the backward fields were analyzed by use of a linear mapping model. The forward and backward values were found to be very similar. All these linear terms can be corrected by registration [9], but to a scale that is limited by the fraction of the registration correction range that can be sacrificed for field correction. Nevertheless, several days elapse before linear-term errors need be corrected other than by registration. At this point a monitoring process can indicate a need to update the correction.

Nonlinear drifts can be corrected only by performing a full field-distortion correction, and for this reason were analyzed separately. These errors were determined by subtracting the linear terms obtained above from filtered

measured errors. They average $0.04\ \mu\text{m}$ and have 3σ of less than $0.07\ \mu\text{m}$. The geometrical errors at five locations in the field due to nonlinear drifts were measured for collects distributed over a period of ten days of normal operation. Results showed that the variation in the nonlinearity at these points was random and the drift in the scan error was insignificant.

Similar experiments were performed to determine the A-to-B correlation stability (the drift of the registration window locations relative to the writing deflection field). The drifts were somewhat larger than the errors due to nonlinearity, but well within acceptable limits.

Additional stability experiments performed with collects half an hour apart, one day apart, and three weeks apart yielded similar results.

Summary

The field distortion correction process described here generates precision deflection scans for a high-throughput electron-beam lithographic system. Typical horizontal scan errors over an uncorrected forward scan are graphed in Fig. 4. After corrections have been applied, the error is reduced to a small fraction of a micrometer ($0.1\ \mu\text{m}$). The system is stable enough that the correction needs to be updated only infrequently. Field size can be changed dynamically with little sacrifice in the availability and throughput of the system.

References

1. D. R. Herriott, R. J. Collier, D. S. Alles, and J. W. Stafford, "EBES: A Practical Electron Lithographic System," *IEEE Trans. Electron Devices* **ED-22**, 385 (1975).
2. H. C. Pfeiffer, "New Imaging and Deflection Concept for Probe-Forming Microfabrication Systems," *J. Vac. Sci. Technol.* **12**, 1170 (1975).
3. G. L. Varnell, D. F. Spicer, A. C. Rodger, and R. D. Holland, "High Speed Electron Beam Pattern Generation," *Proceedings of the Sixth International Conference on Electron Beam Science and Technology*, The Electrochemical Society, San Francisco, 1974, p. 97.
4. A. D. Wilson, T. H. P. Chang, and A. Kern, "Experimental Scanning Electron-Beam Automatic Registration System," *J. Vac. Sci. Technol.* **12**, 1240 (1975).
5. N. C. Yew, "A Modular System for Electron Beam Microfabrication," *Proceedings of the Sixth International Conference on Electron Beam Science and Technology*, The Electrochemical Society, San Francisco, 1974, p. 111.
6. H. S. Yourke and E. V. Weber, "High Through-Put Scanning-Electron-Beam Lithography System, EL1, for Semiconductor Manufacture—General Description," *IEDM Tech. Digest*, 431, (1976).
7. J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, *The Theory of Splines and Their Applications*, Academic Press, Inc., New York, 1967.
8. S. S. Coons and B. Herzog, "Surfaces for Computer-Aided Aircraft Design," *J. Aircr.* **5**, 402 (1968).
9. D. E. Davis, R. D. Moore, M. C. Williams, and O. C. Woodard, *Automatic Registration in the Electron Beam Lithography System*, *IEDM Tech. Digest*, 440 (1976).

Received February 16, 1977; revised April 20, 1977

H. Engelke is located at IBM Deutschland, Tuebinger Allee 49, 7032 Sindelfingen, Germany; J. F. Loughran, M. S. Michail, and P. M. Ryan are located at the IBM System Products Division laboratory in East Fishkill (Hopewell Junction), New York 12533.