D. E. Davis
R. D. Moore
M. C. Williams
O. C. Woodard

Automatic Registration in an Electron-Beam Lithographic System

Abstract: In the fabrication of integrated circuits, electron-beam systems are increasingly used to directly expose circuit patterns on resist-covered semiconductor wafers. These systems are made attractive by their inherent capability for writing patterns at high resolution, which eliminates the need for the masks used in optical exposure systems. To compete economically in production, electron-beam systems must attain a high level of wafer throughput. To achieve this, the system must be fully automated and spend the minimum of time in performing overhead functions such as various calibrations, pattern registration, and mechanical positioning of the wafers. This paper describes pattern registration, a process by which the circuit patterns required at a particular level of device fabrication are mapped to those of the preceding level. It discusses the considerations taken as basic in designing an accurate, high-speed registration process for a production-type electron-beam exposure system. This automatic registration system operates in 150 milliseconds per integrated circuit chip, allowing the system to achieve a throughput of 2000 5-mm chips per hour with overlay error of less than 0.75 μm (3σ). The operation of this system, its performance characteristics, and measurements of its pattern-matching accuracy are presented.

Introduction

The recent rapid development of integrated circuits has made possible faster computers, hand-held calculators, and devices to reduce automobile pollution, to name a few of a myriad of applications. Advances in the lithographic systems that form the patterns used to produce circuits have played a large role in these developments.

Integrated circuits are fabricated by successive processing steps applied to a semiconductor wafer. In general, prior to each of these processing steps a pattern is exposed onto a masking resist layer. When this resist is developed, the resulting pattern allows the subsequent process step to act selectively on different areas of the wafer. The required relationship of each succeeding pattern to the others makes it necessary that each one precisely overlie the preceding ones. Errors in mechanical handling, changes in a given exposure tool, and differences among them make it difficult to register a pattern to underlying ones. These errors must be determined and corrected to assure good overlay of the next pattern.

As the dimensions of details in circuit patterns approach the resolution limits of light optics, manufacturers have begun to use micron-sized electron-beam probes to write the patterns directly on wafers [1–4]. Although submicron size probes can easily be produced, the accuracy with which the beam can be deflected to write patterns, and particularly to achieve good pattern overlay, has imposed a major limitation. Where time (throughput) is not a major objective, overlays of 0.1-0.2 μm have been reported [5, 6]. For economy, however, a manufacturing tool must have high throughput and added costs must be kept low. Added process steps and the use of significant wafer areas to accommodate registration marks must be avoided. Setup and calibration times also must be kept to the minimum.

The registration system to be described here is part of an electron-beam lithographic system, EL1 [7], which operates within 150 ms per chip, keeps overlay errors within 0.75 μm (150 ppm over a 5-mm field), and requires less than 15 minutes a day for calibration and setup. Wafers are loaded into the system, all circuit chip sites are registered and written sequentially, and the wafers are unloaded from the system, all automatically. A throughput of 12 82-mm wafers (over 2000 circuit chips) per hour is achieved. The system uses a 2.5 μm beam of 3 μA at 25 kV accelerating potential.

The major components and the signal flow of the registration system are diagrammed in Fig. 1. The deflection memory contains the magnetic deflection definition,

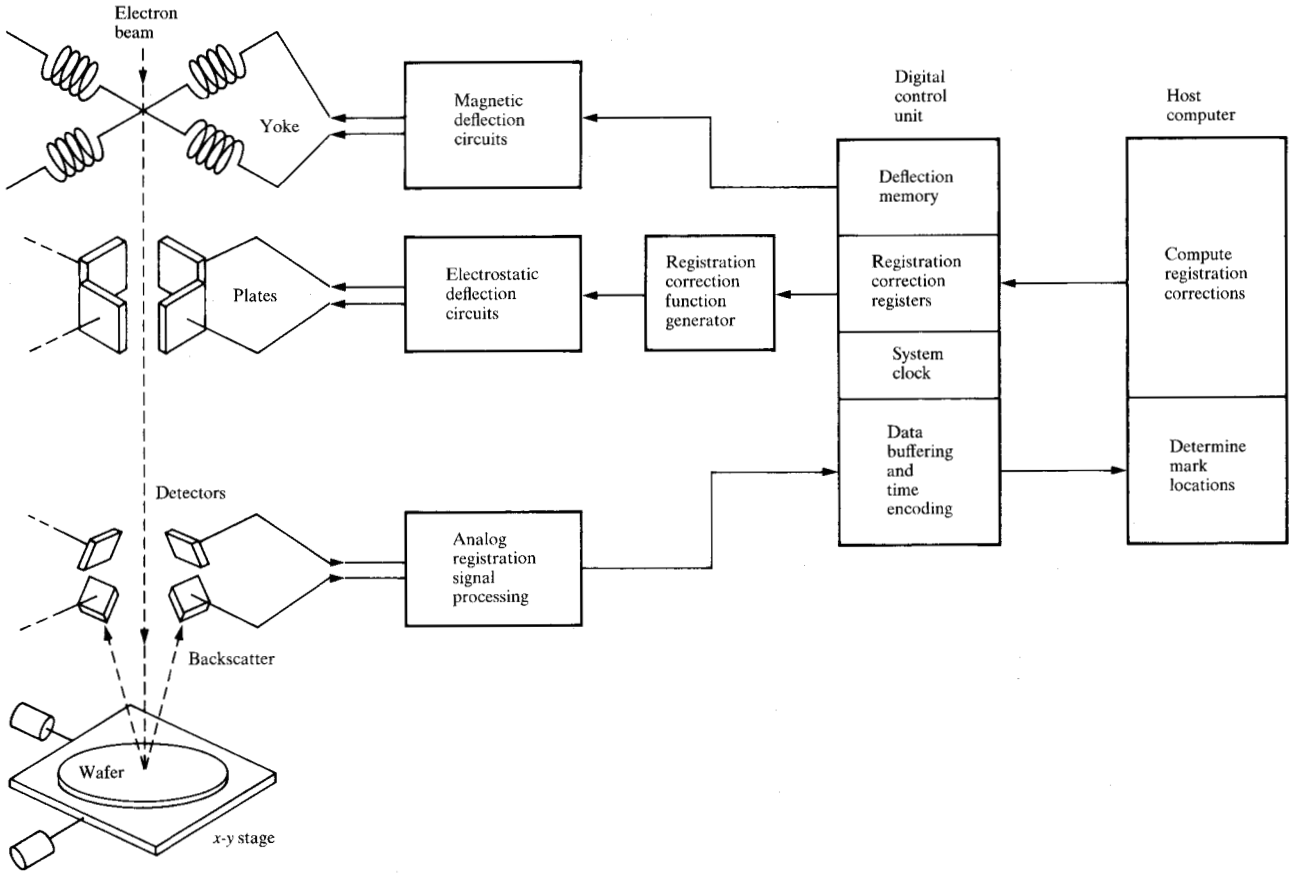


Figure 1 System block diagram.

and is read continuously, causing the system to run in a cycle consisting of registration scans (A), writing scans (B), and move time (C). During the A part of the cycle, the positions of registration marks on the chip to be written are determined by analog, digital, and data processing methods. During the B part, the pattern is written on the chip with appropriate overlay corrections, which are calculated by the computer and applied via digital and analog circuits. During the C part, the *x-y* table positions the next chip underneath the beam. Nonlinear magnetic deflection errors are corrected before the wafers are written, so that only linear corrections are required to achieve good overlay. The differences between the positions and the speeds of the registration and writing scans are measured, and the results are stored in the computer as part A to part B correlation constants. The scale factors for the registration-correction circuits are also measured and stored. These procedures are automated to facilitate calibration and save time.

When a wafer is loaded into the system, specific wafer data are supplied to the computer in what is called the wafer map, which contains parameters for use by the control and data-editing programs in adjusting the system to the unique requirements of the wafer to be written.

To facilitate automation of the registration process without using very large registration marks on each chip, a three-step process is used: mechanical alignment, wafer registration, and chip registration. First, an automated mechanical handler loads the wafer onto a carrier. Then, the wafer carrier is introduced into the vacuum chamber by means of an interlock system, and placed on an *x-y* table, which positions the wafer underneath the electron beam. The overall mechanical handling error at this point is within ± 75 μm. Second, the beam scans wafer registration marks to sense the errors in position, magnification and rotation (one chip site on the wafer is devoted to the wafer marks). Corrections are applied as the *x-y* table moves the first circuit chip underneath the beam with a maximum error of ± 7.5 μm. The third and last step is chip registration, for which eight marks, two in each corner, are used. The chip marks are scanned, then the detected errors are corrected as the pattern is written. Points in the electron-beam writing field which correspond to the expected registration mark locations

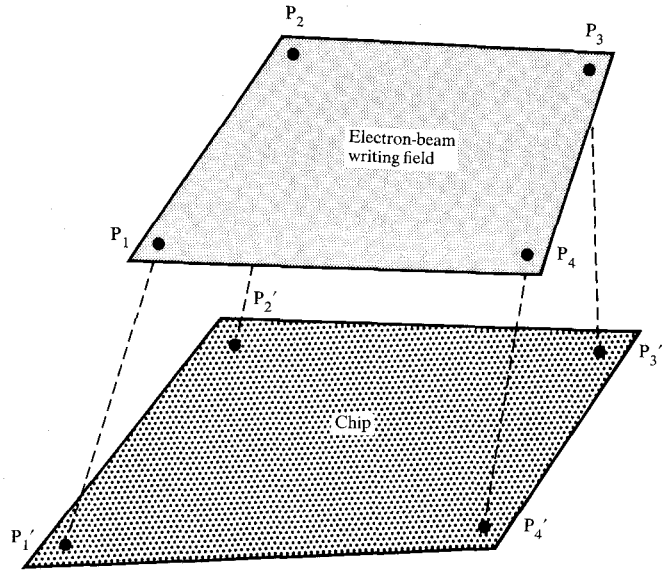


Figure 2 Electron-beam field stretched to match chip. X correction $= A + BX + CY + DXY$; Y correction $= E + FX + GY + HXY$.

(P_1, P_2, P_3, P_4) are adjusted to coincide with the measured mark locations (P_1', P_2', P_3', P_4') (Fig. 2). This ensures matching of the chip fields and writing fields by correcting for linear distortions of the fields. This process is repeated until all the chips on the wafer are written.

Figure 3 illustrates the layout of a typical wafer and the relationships of wafer and chip sites. The rest of this paper concentrates on the details of the chip registration process.

Detection and generation of signals

The registration process begins with the detection of the marks. As the electron beam scans across the marks, a signal is picked up by detectors above the wafer. These detectors are arranged in a square with sides parallel to the sides of the chips. Scanning directions correspond to the same axis. The detectors are rectangular silicon diodes fabricated by a diffusion process. Properly biased, they are fully depleted and capable of detecting electrons incident on the side opposite the diffusion. The current gain produced is proportional to the energy of the incident electrons. Thus, the signal is produced primarily by high-energy electrons that have been backscattered from within the wafer. Low-energy secondary detection would not be satisfactory, since the marks are often covered by thick resist and process layers which do not allow secondary electrons to escape from the mark. Typical backscattering coefficients for integrated circuit wafers are 0.1 to 0.25. The solid angle subtended by each diode detector is 0.43 steradian.

The basic signal detected is a relatively high level of backscattered electrons modulated by the topography and material differences (atomic number) of the mark. The background contains random noise resulting from this backscattered electron current. In this system, shot noise from the electron beam is the largest contributor to random noise; it is greater than detector noise and circuitry noise combined.

Errors caused by random noise are reduced to tolerable values by judiciously filtering the signals and by averaging data for a number of scans. In Fig. 4, the theoretical measurement error of the location of the registration mark is plotted vs the signal-to-noise ratio (s/n), for marks containing two edges and four edges after 30 scans are averaged.

A slow scanning speed of 1.6 μm/μs produces an electrical signal that can be filtered from the broadband noise. However, the scanning speed must be fast enough not to allow localized heating by the electron beam that causes the resist to decompose rapidly and to bubble. The bubbling resist distorts the backscattered signals from the underlying registration mark; the distortion appears as noise, which is correlated to the scanning and is not reduced by the averaging. Furthermore, the heating may make the resist hard to remove.

To further reduce the possibility of causing the resist to bubble, a "dither" method is used for scanning. The dither motion is a 10-MHz triangular signal used to deflect the beam orthogonally to the scan direction. This effectively elongates the beam spot in the direction parallel to the mark edge and perpendicular to the scan direction. Thus the instantaneous energy is dispersed over a larger area, and the peak temperature of the area being scanned is reduced. Thus, the dither method allows the scan to be slow enough for good signal-to-noise ratio without bubbling. Any 10-MHz component that might be added to the signal is removed by low-pass filtering.

The 250-μm wafer registration scans and large wafer marks were designed to accommodate up to a $\pm$75-μm mechanical placement error. Errors in translation, rotation, and magnification, as found from the wafer marks, are used to correct the x-y table position and the beam-deflection field so that the position of subsequent chip marks can be predicted to within $\pm$7.5 μm. Thus the system can have small chip registration marks and short (104 μm) scans.

Fifteen scans forward and fifteen backward are applied in a raster pattern over each chip mark. The raster movement along the mark helps to spread the energy being dissipated in the resist and also averages out local mark defects that might cause errors in the measurement of mark positions. The forward-backward scanning makes it possible to cancel signal propagation delays and

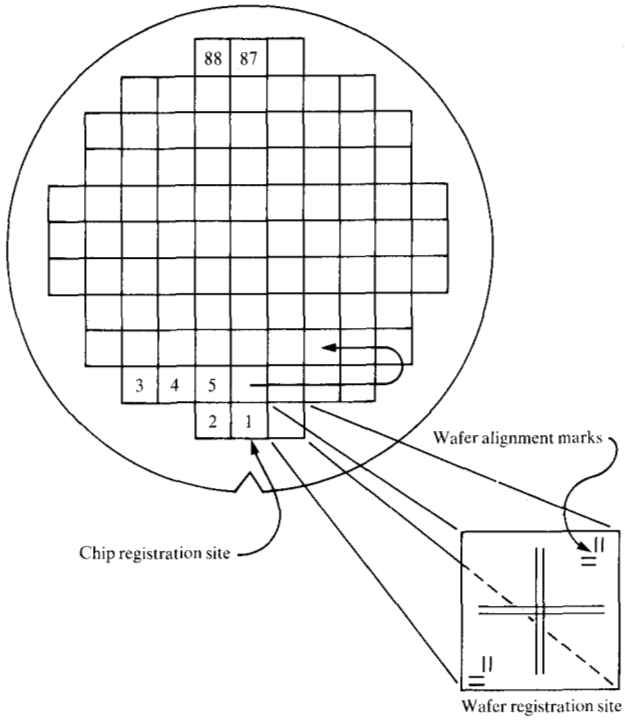


Figure 3 Wafer layout showing individual chips.

time distortion in the registration system, thus eliminating one source of error. A unidirectional scan would require measurement and compensation of these delays. Each mark consists of two parallel bars (see Fig. 5) formed on the wafer surface at some earlier fabrication step. (The bars can be either raised or etched into the wafer surface. When possible, raised bars are preferred, as they protrude upward into the overcoating resist, resulting in a stronger signal than do depressed bars.) These bars are covered with an electron-beam resist and need not be uncovered for the registration process. For a particular resist thickness, beam size, and mark depth, there has been found to be an optimum bar width for the analog signal processing done in this system (see Fig. 6). The bar width (5 μm) was chosen for the worst case resist thickness. It is wider than optimum to account for etch tolerances which reduce the width. The distance between the bars, 30 μm, is wide enough that the signals from the bars are well separated. Two bars are used to provide a known spacing for mark identification. The area of the corner marks on each chip is only about one percent of the wafer surface; thus 99 percent is left for circuits.

Analog signal processing

The backscattered signal, produced as the beam is scanned across a resist-covered alignment mark on the

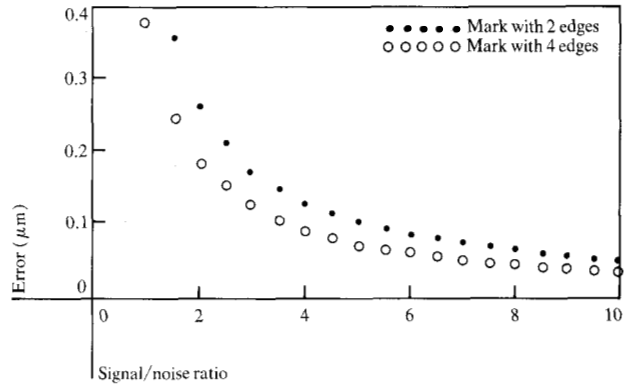


Figure 4 Mark identification error vs signal/noise ratio.

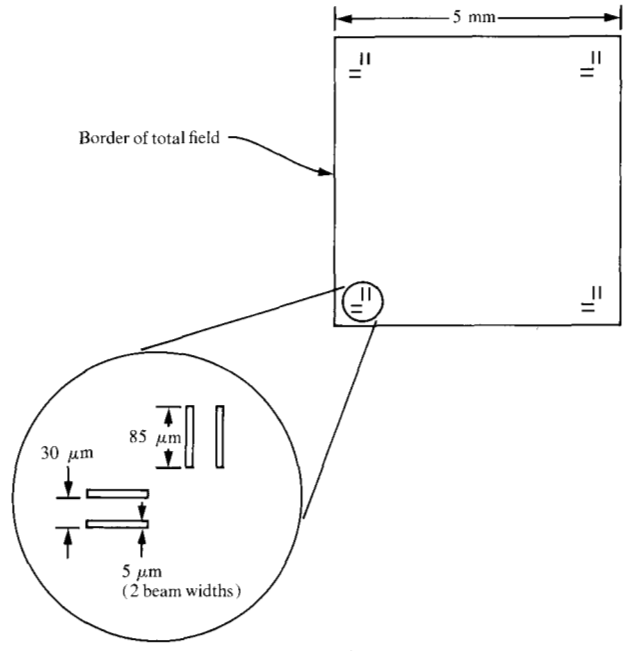
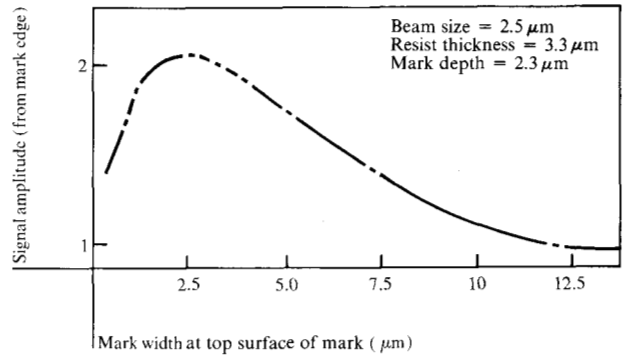


Figure 5 Chip alignment marks.

Figure 6 Registration backscatter signal vs alignment mark width. Mark depth = 2.2 μm. Resist thickness = 3.2 μm.



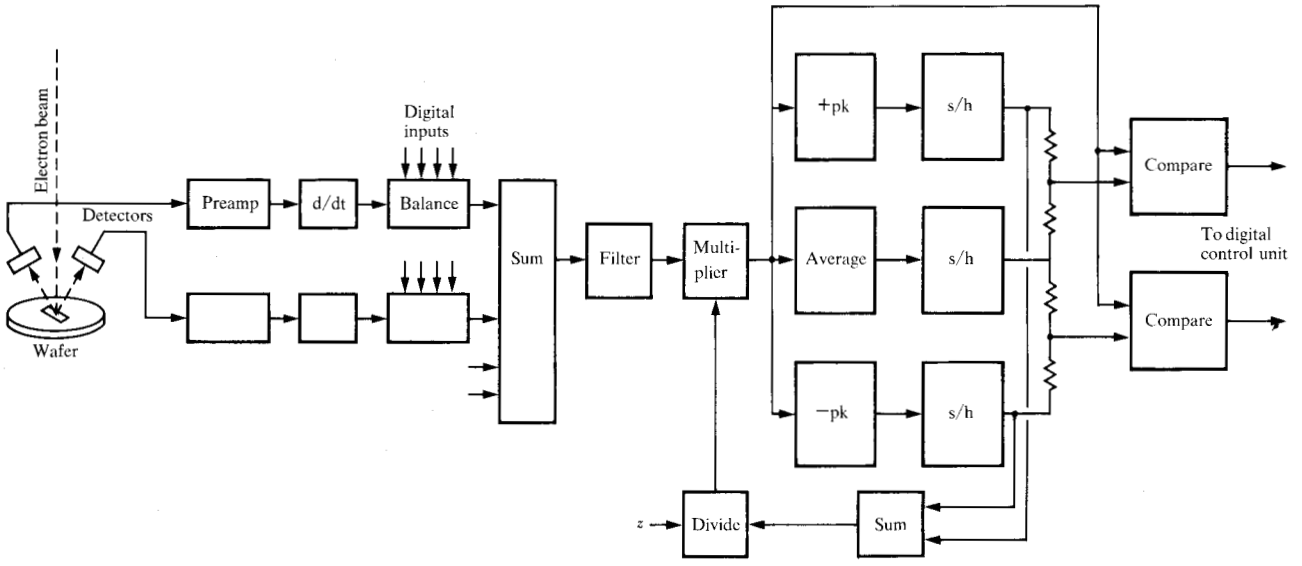


Figure 7 Analog signal processing block diagram.

wafer, first passes through a set of analog circuits. Here it is amplified, filtered, and compensated for variations in amplitude and baseline level. The resulting analog pulses are threshold-detected to produce a digital pulse for each mark edge crossed by the beam. The time values of these digital pulses, determined by sampling the system clock, are used to calculate the locations of the alignment marks.

The analog processing begins with the backscattered signal. The output current from each detector diode is converted to voltage by a preamplifier (Fig. 7). The output voltage from each preamplifier provides the input for a differentiator. Both preamplifier and differentiator are band limited at one MHz to reduce high frequency noise. This sets the bandwidth for the signal detection system. The differentiation process removes the background from each detector signal and also accents the fast transition that occurs in the detected signal as the beam crosses the edge of the registration mark. Each of the four differentiated signals is passed through a balance control circuit that can adjust the signal amplitude of each channel and also furnishes a gain control for all four together. The balance input controls are set digitally, on the basis of values read from the wafer map. Generally all four channels are set with the same balance inputs.

The signals from the balance control circuits are summed to produce a signal that has positive and negative peaks showing when the beam crosses the mark edges, and this summed signal is sent to the AGC loop. During the first scan across the registration mark, the positive and negative peak voltages are detected and held. These values are algebraically summed and divided into a constant z. The resulting quotient is fed back and used to multiply the incoming signals for the rest of the scans to give a constant output. In this way the AGC adjusts the signal amplitude on the remaining scans to a constant peak-to-peak value for a 20 to 1 input range, reducing the dynamic range requirements of the remaining circuits, thereby improving accuracy.

During the second scan across the registration mark, the positive and negative peaks of the analog signal are captured again, and the average value or baseline of the analog signal is determined. A level of 60 percent of the peak value above the average value is used as a comparator reference to clip the next analog signal resulting from a scan in the same direction. As the forward and reverse scans across the marks are alternated, two sets of sample-and-hold circuits are required (one set shown in Fig. 7). The values obtained from each scan are used on the next scan in the same direction, until scanning is complete. The signal-processing circuits continuously adapt to varying signal conditions, both within a wafer and from wafer to wafer.

Digital data processing

Hardware

The digital pulses generated by scanning the beam over the mark edges must be transformed into positional data, to relate these edges to the start points of the scans. This is done in the digital control unit (DCU, Fig. 1), which contains the system clock. A registration data collection begins when the host computer loads the DCU with a data collection program, which it executes in synchronism with the registration deflection. During collection, digital

data are sent to the DCU over two event lines to separate signals from rising and falling mark edges (Fig. 8). Any transition on either of the event lines causes the DCU to record the system clock. It then sums the times representing both transitions of each pulse and gives the sum a sign corresponding to the event line on which it was received. This reduces the data to a single, 16-bit clock value for each mark edge crossed by the beam. The 70-MHz clock and 1.6-μm/μs scan speed correspond to a positional resolution of 0.0115 μm. For each chip registration, about 3000 bytes of data are sent across a selector channel from the DCU to the host computer in about 50 ms.

There are several advantages to retrieving the data in this manner. First, the data are reduced in real time in the DCU by pairing the transitions and eliminating transients. Second, since the DCU executes the data collection asynchronously, the computer is free to edit the incoming data in parallel with the collection; thus, the total registration time is reduced. Finally, since the data are separated into positive values for rising mark edges and negative for falling, the computer program can more easily distinguish the valid edge data points from the noise.

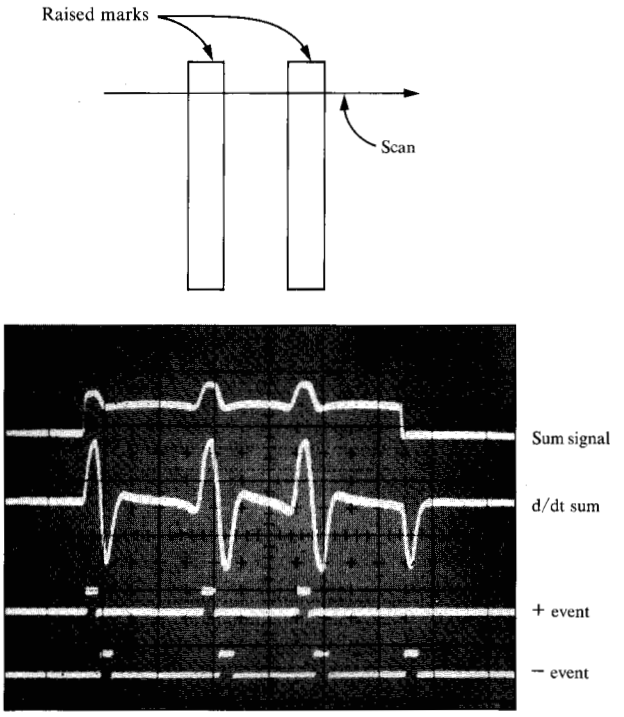


Figure 8 Registration signals.

• *Software*

The host computer is used to edit the edge data rapidly as they are received from the DCU and to determine the displacement of each mark from its expected position in the scan window. This is done by comparing the data from each scan to a mark model read from the wafer map. This single model consists of four signed values representing the mark edges relative to the start of an ideal scan. These values specify the dimensions and location of a mark and whether it is raised or recessed. Thus a mark model of different dimensions, position, or type (raised or recessed) can be specified for different levels of device fabrication or variations of the same level.

The editing program, therefore, uses the specified mark edge spacings to test for valid edge points in the data. To do this, it relies on the fact that the relationship between like edges of the two bars that make up each mark is altered only negligibly by any effects of wafer processing, such as overetching of the bars. Therefore, it can stringently test for a set of leading edges and a set of trailing edges, and then test the relationship of one set to the other. This is done for each line of data. For the final test, all the lines of data for both the forward and backward lists of selected mark center points are used to eliminate points outside of a 2σ limit. The accepted points are then averaged to give a measured mark position for both forward and backward directions. Finally, the mark displacements are determined from the location of the specified model and are modified by the part A to part B correlation constants, which have been determined by an off-line calibration procedure [8]. This ability to determine the relationship between two sets of deflections makes it possible to use the optimum scan for each purpose; that is, the scans for detecting the registration mark can be slow to enhance s/n, and the writing raster can be fast for high throughput.

The displacements determined above are used to calculate a set of correction coefficients, which then map points in the corners of the writing field to the corresponding points on the wafer. The functions that achieve this are the following:

$$\Delta X + A + BX + CY + DXY, \text{ and} \quad (1)$$

$$\Delta Y = E + FX + GY + HXY, \quad (2)$$

where ΔX, ΔY are displacement terms; X, Y are ideal mark positions in the field; and A through H are the correction coefficients that affect translation, magnification, rotation, and trapezoidal distortion in X and Y. Immediately after the corrections are calculated, they are sent through the selector channel to the DCU, which delivers them to the analog correction circuits; they are then applied to the beam by a set of electrostatic deflection plates.

The registration system is equipped to recover automatically from occasional failure situations. When fewer than the expected number of marks are detected, A and

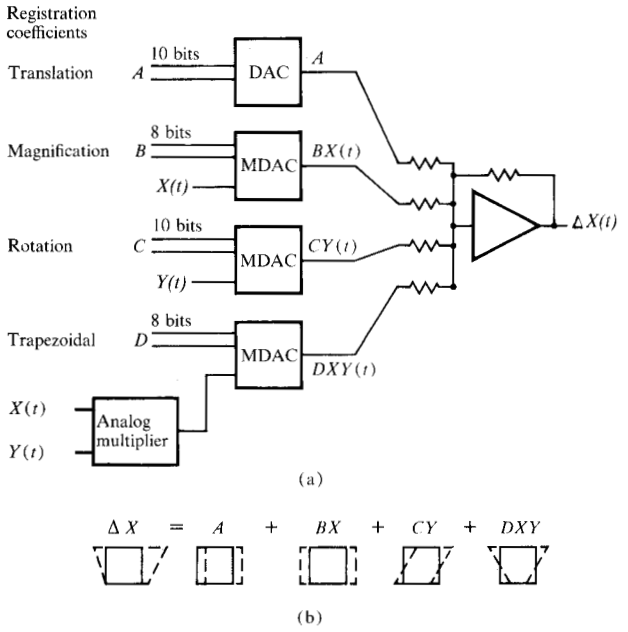


Figure 9 Registration correction. (a) Function generator. (b) Effects on the writing field.

E correction terms are generated from the available mark data. The remaining correction terms used are from the registration of the previous chip. In cases where A and E corrections are excessively large, an x-y table move is made to zero out the large translation error and registration is performed again. If other corrections are too large, the maximum possible values are applied.

The registration software activity is designed for fast execution. The equations that yield the correction coefficients are partially solved (matrix expanded) at wafer start time to avoid a redundant calculation at each site. As was mentioned earlier, the data collection time is exploited by editing the data simultaneously in hardware and software. As a result, registration is complete in 150 ms.

Corrections applied from chip to chip

As the system moves from one chip site to the next to register chips, the scan windows must overlie the marks at each new site. Therefore, the correction terms generated at one site, except A and E (Eqs. 1 and 2), are kept in place throughout the data collection at the next site. In this way, the scan windows are altered to fit the marks on the basis of the preceding site's registration, and are updated at every site. The x-y table address of any site is modified by the translation (A, E) and rotation (C, F) terms determined at the preceding site. This process eliminates cumulative position errors due to any rotation of the wafer axis relative to the mechanical system. Since the values of A and E calculated at one site are incorporated into the address of the next site, the capability of electrostatic translation correction is then used to nullify the sensed x-y table error after a move is made.

These procedures have proved effective in enabling the system to move automatically from site to site so that registration marks consistently appear in the scan windows, well within their ± 7.5-μm capture range.

Correction function generator

The circuitry used to generate the X dynamic correction functions for X is diagrammed in Fig. 9(a). The circuitry used for Y is similar. The A constant is applied to a digital-to-analog converter (DAC), which generates the appropriate correction voltage. Multiplying digital-to-analog converters are used to provide the products BX, CY, and DXY where X and Y are voltage samples from the magnetic deflection circuits (Fig. 2). The XY term is produced by an analog voltage multiplier. The four correction voltages are summed by an operational amplifier and applied to the X electrostatic deflection plates to produce the corrective deflection.

A graphic representation of the effect of each correction term on the field is shown in Fig. 9(b). Overlay errors contributed by the correction circuits are held to less than 0.15 μm, by using very accurate DACs and by limiting the correction range to the minimum required to gain DAC resolution.

Results

The errors in the registration system include errors in determining the position of the alignment marks and in generating and applying registration corrections. These errors, taken together, constitute only one contributor to total overlay error. Others include nonlinearity in the writing field, beam jitter, and imperfect calibration of the beam deflections used for registration scans to those used for pattern-writing scans (part A to part B correlation). There are others, including those that occur when electron-beam patterns must align with patterns previously placed on the wafer by other lithographic methods.

Some typical overlay results for a variety of mark structures and resist thickness are given in Table 1. Both raised and etched marks are represented in the table but have little effect on the overlay (raised are slightly better). The data were collected over a period of a year, with the more recent results somewhat improved due to general system maturation rather than changes in the registration subsystem. These results are presented to show typical results for a cross section of conditions rather than to illustrate the exact error dependence on parameters. In general, the registration component of

Table 1 Overlay results (in microns, 5 mm field).

Date	*Resist thickness*	*Mark relief*	*Number of measurements*	*X Error* (3σ)	*Y Error* (3σ)
12/75	2.6	0.75	220	0.68	0.54
1/76	1.5	0.7	90	0.63	0.48
1/76	3.3	1.5	125	0.53	0.56
10/76	1.2	0.35	300	0.48	0.49
1/77	1.0	0.5	680	0.46	0.28

the overlay error is greater as the resist thickness/mark relief ratio increases. Random component registration error for these system overlays has been shown to be typically less than 0.1 μm. The 3σ-overlay errors as given in Table 1 are seen to be less than 0.68 μm. The system has also demonstrated the ability to register on alignment marks 0.5 μm deep covered by 4.1 μm of electron-beam resist (signal-to-noise ratio of 1.4); the resultant random error in mark location is 0.19 μm.

A typical device on which first-level metal, vias, and second-level metal patterns have been defined by this registration system is shown in a photomicrograph (Fig. 10).

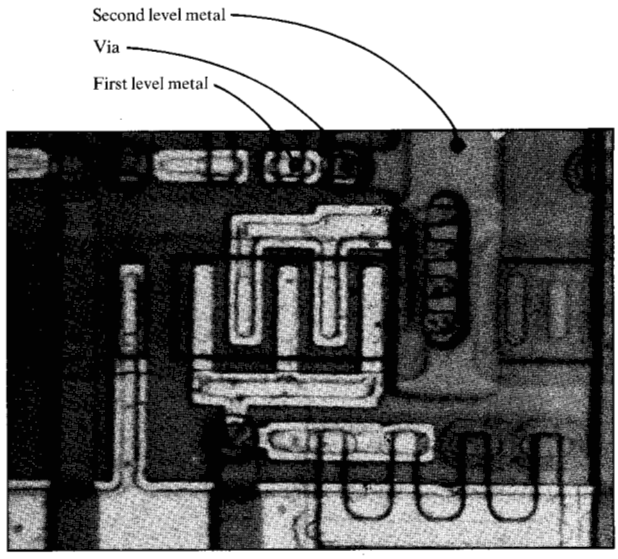


Figure 10 Bipolar device fabricated using EL1.

Summary

The registration system described here is a principal element in the automation that has made it possible to develop a viable electron-beam lithographic method for manufacturing [7]. The adaptability of the signal-processing circuits and software to varying conditions, the off-line calibration methods, the on-line setup procedures, and finally the optimization that has made possible consistent submicron overlays are the major contributors to the success of the registration system.

Acknowledgments

The authors are indebted to the members of the design team for their various contributions. In particular they thank P. A. Cardenia, J. Casulli, S. K. Doran, A. W. Muir, J. J. Stavish, K. Salchow, and J. G. Wilman for their work in hardware design and development; M. A. Habegger, A. Lanzaro, and W. Stickel for backscattered signal studies; and J. Pavick and J. Rogoyski for software design.

References

1. S. Miyauchi, K. Tanaka, and J. C. Russ, "I. C. Pattern Exposure by Scanning Electron Beam Apparatus," *Solid State Technol.* **12,** 43 (1969).
2. G. L. Varnell, D. V. Spicer, and A. C. Rodger, "E-Beam Writing Techniques for Semiconductor Device Fabrication," *J. Vac. Sci. Technol.* **10,** 1048 (1973).
3. D. S. Alles, F. R. Ashley, A. M. Johnson, and R. L. Townsend, "Control System Design and Alignment Methods for Electron Lithography," *J. Vac. Sci. Technol.* **12,** 1252 (1973).
4. T. H. P. Chang, H. Hatzakis, A. D. Wilson, A. J. Speth, A. Kern, and H. Luhn, "Scanning Electron Beam Lithography for Fabrication of Magnetic Bubble Circuits," *IBM J. Res. Develop.* **20,** 376 (1976).
5. N. Saitou, C. Munakata, Y. Miura, and Y. Honda, "Computer-Controlled Electron Beam Microfabrication Machine With a New Registration System," *J. Phys. E* **7,** 441 (1974).
6. A. D. Wilson, T. H. P. Chang, A. J. Speth, and A. Kern, "Automatic Electron Beam Fabrication of Micron-Size Devices," *Proceedings of the Workshop on Microelectronic Device Fabrication and Quality Control with the SEM*, IIT Research Institute, Chicago, April 1976, p. 659.
7. H. S. Yourke and E. V. Weber, "Correction of Field Distortion in Electron-Beam Lithography System, EL1," *IEDM Tech. Digest*, 1976, p. 431.
8. H. Engelke, J. F. Loughran, M. S. Michail, and R. M. Ryan, "A High-Throughput Scanning-Electron-Beam Lithography System, EL1, for Semiconductor Manufacture: General Description," *IEDM Tech. Digest*, 1976, p. 437.

Received March 3, 1977; revised May 12, 1977

The authors are located at the IBM System Products Division laboratory, East Fishkill (Hopewell Junction), New York 12533.