High-Throughput Mapping of Short-Range Spatial Variations Using Active Electrical Metrology

Xu Ouyang, C. Neil Berglund, *Fellow, IEEE,* and R. Fabian W. Pease, *Fellow, IEEE*

***Abstract*—Spatial variations of parameters in semiconductor manufacturing, such as critical dimension (CD) and overlay, have significant impact on the performance and yield of integrated circuits (IC). Among these spatial variations, the variations of parameters between transistors separated by a very short spatial distance such as 1 μm to 100 μm (*intertransistor variations*) can be particularly hazardous for those types of ICs that require exact transistor–transistor matching. To measure these intertransistor variations, both high-throughput and high-spatial-sampling-density beyond the scope of currently available metrology tools are needed. We have thus developed an active electrical metrology method of measuring intertransistor variations using on-chip, active, electrically addressable arrays of test structures to provide the high-throughput (5 μs/data point) and high-density (3 μm/grid spacing) needed. Test chips were designed and fabricated on a HP 0.35-μm process, and the testing configuration was set up to optimize throughput and precision. This method was verified with the measurements of on-chip calibration arrays. The spatial variations of both intertransistor CD (effective gate length) and overlay (between poly/diffusion) within the test chips were mapped with this method. For these circuits, the intertransistor CD variations were found to depend primarily on the layout, whereas the intertransistor overlay variations were found to be dominated by errors of the pattern generator used to fabricate the masks.**

***Index Terms*—Active electrical metrology, addressable array, CD, high-throughput, intertransistor, lithography, metrology, short-range, spatial frequency, spatial variations.**

I. Introduction

URING semiconductor manufacturing, process variations always exist even when the process is within its control limits. The induced parametric variations, such as gate length variations or overlay variations, impact the performance of the chip, and, if sufficiently large, can cause parametric yield loss in some circuits [1]–[6].

The spatial distribution of parametric variation is very important as it determines how the performance and/or yield are affected. It is known that interfield (chip) and intra-field (chip) variations have very different impact on chip performance and yield [2]. Furthermore, the *intertransistor* variations, which are defined as those short-range variations with spatial wavelengths from 1 to 100 μm, are particularly serious for many types of integrated circuits such as most analog circuits that are sensitive to device mismatch problems, or some digital circuits with sense amplifiers or other balanced circuit designs [7]. However, such short-range parametric variations are not systematically measured and characterized mainly due to their metrology difficulties. To study these short-range spatial variations, a large number of data need to be collected on a very dense grid, which demands high-throughput and high-density of the metrology tools.

Memory chips consist of an array of compact memory cells with fast addressing circuits surrounding the memory cell array. Addressing circuits can read out the information stored in each cell very rapidly. By modifying memory cells or replacing memory cells with specially designed test cells, the read-out electrical signal can be made to reflect spatial variations of a specific process parameters such as gate length or overlay. This is the active electrical metrology scheme used in this paper to measure and study intertransistor variations.

The approach of using addressable arrays to study process variations has been previously reported [8], [9]. However, the throughput or measurement time and the spatial variations of the measured parameters, which are the two major advantages of using addressable arrays as metrology tools, have not been discussed.

We have previously used a modified SRAM to study spatial distribution of CD variations and discussed its measurement noise and speed [7], [10]. However, SRAM as a metrology tool has certain disadvantages: the measured gate length variations may have the influence from other error contributors such as variations of doping and gate oxide thickness, and the size of the SRAM cell is not minimized for metrology purpose. Also, the testing of the previous SRAM chips was done on a conventional parametric tester, which significantly limited the throughput. In this paper, we will present a newly designed test chip and testing configuration to provide further verification and significant improvement of throughput over the previous SRAM chips.

This approach of designing test chips with fast addressable arrays and active test structures to measure process or circuit parameters, together with the test configuration to achieve throughput as high as 10^5 measurements/second, is termed active electrical metrology. Active electrical metrology is not meant to replace the existing conventional metrology tools (such as CD-SEM), but to complement them. It has many uses in practice. For example, an obvious use is to probe for intertransistor process variations. In many such cases, it is

Manuscript received January 17, 2001. This work was supported by the Semiconductor Research Corporation under Contract LJ-600.

X. Ouyang is the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA and also with the PDF Solutions, Inc., San Jose, CA 95110 USA (e-mail: xuo@pdf.com).

C. N. Berglund is the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA and also with the Department of Electrical and Computer Engineering, Oregon Graduate Institute of Science and Technology, Portland, OR 97291-1000 USA (e-mail: berglund@ece.ogi.edu).

R. F. W. Pease is with the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 0894-6507(02)01028-X.

either economically unfeasible or technically difficult to use conventional metrology tools to measure such short-range variations. Active electrical metrology has much higher throughput and can be easily implemented with existing test chips or product chips. In these cases, the test chip should be designed to reflect only the process variations, not the design attributes such as layout of the test chip.

On the other hand, active electrical metrology can be applied to measure and quantify some complex process effects that are layout dependent, such as the goodness of optical proximity correction (OPC). It is difficult to use conventional tools such as CD-SEMs to quantify such effects in realistic circuits due to the complexity. However, for active electrical metrology it becomes fairly easy, because the actual electrical performance is the ultimate standard for such measurements, and it is electrical signals that active electrical metrology measures. The problem is then to design the test chip to reflect the actual process effects in real circuits.

In this paper, we will first present the design and layout of the test chip, then explain the testing configuration, error analysis, and verification. Finally we will present the results of intertransistor CD and overlay variations measured with the test chips. Fourier analysis will be used and the results will be shown in spatial-frequency domain [7], [10]–[14].

II. Layout and Design of Test Chip

A. Layout of the Chip

Fig. 1(a) illustrates the layout of the test chip used in active electrical metrology. This test chip was designed at Stanford and fabricated on a commercial HP 0.35-μm process (through MOSIS). It contains three major "active" arrays: a conventional SRAM array to infer CD variations from SRAM current and compare with previous results, a CD array to infer CD variations by transistor saturation current, and an overlay array to infer overlay variations by specially designed overlay test structures [Fig. 1(b)], which will be explained later). The term "active" is used to emphasize the fact that the arrays are made up of active transistors, not passive resistance measurement structures, therefore more closely represents variations in integrated circuits. Addressing circuits have been designed near the arrays to address each cell in the arrays. The test chip also contains a small transistor array to verify and calibrate the relationship between CD errors and the measured currents. In addition, the last column of the overlay array contains overlay calibration test structures.

B. Design of CD Test Structure

The CD test structure in the test cells of the CD array consists of a single transistor. From the accurate transistor model developed for the process, the one-to-one relationship between saturation current and gate length can be established. Therefore, by measuring the variations of saturation current, the variations of gate length can be inferred. Fig. 2 shows the curve of saturation current vs. drawn gate length from the transistor model provided by MOSIS. As we can see, for this process, within the $\pm 10\%$ variation around the drawn gate length, the relationship can be treated linear. This simplifies the mapping between measured current variation and gate length variations to a ratio of $\Delta I_{ds}/\Delta L$.

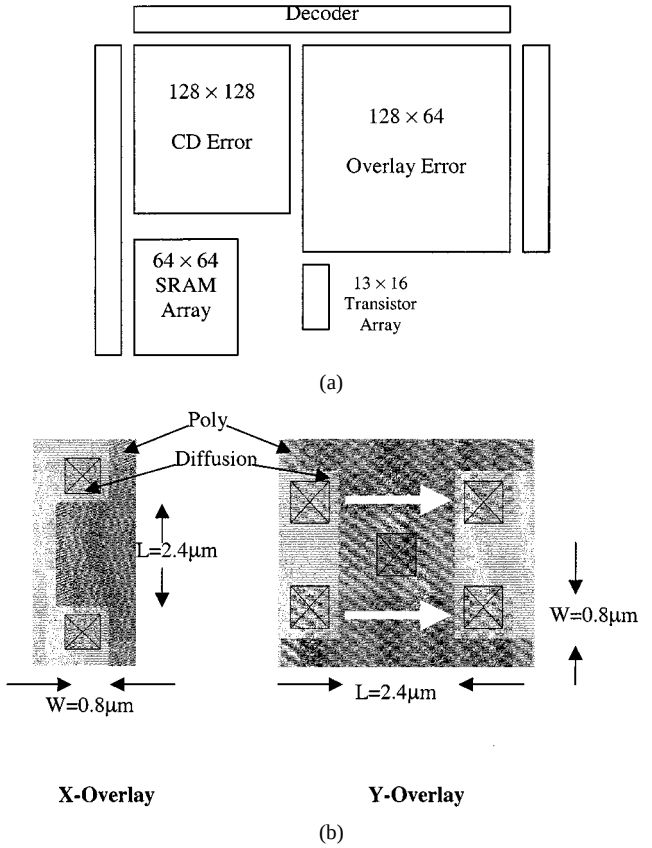


Fig. 1. (a) Layout of the test chip used in active electrical metrology. CD error array consists of identical single transistors with width $= 4 \times$ length; SRAM array consists of standard six-transistor SRAM cells; transistor array consists of transistors with various sizes for calibration purpose; overlay array consists of overlay test structures shown in (b). All the arrays have periphery addressing circuits so that each test structure in the arrays can be rapidly selected. (b) Layout of the x and y overlay test structures. Functions of the test structures are explained in Section II-C.

However, the saturation current is also dependent on some other factors, and the variations of these factors can cause errors in the mapping from saturation current to gate length. Let us now take a look at what these factors are.

To the first order, the saturation current can be modeled as

$$I_{\mathrm{Dsat}} = \frac{\mu_n C_{\mathrm{ox}}}{2} \frac{W}{L} (V_{\mathrm{GS}} - V_T)^2$$

where μ_n is channel electron mobility, C_{ox} is oxide capacitance per unit area, W is transistor width, L is transistor length, V_{GS} is gate to source voltage, V_T is threshold voltage. μ_n and V_{GS} can be regarded as unchanging from cell to cell within the array. The variations of C_{ox}, W, and V_T can affect the transistor saturation current, therefore are the error sources as we try to infer L from saturation current. Their contribution can be reduced by careful design of the transistor, for example, designing the gate width to be large to reduce its percentage variations, and other measurement and data analysis techniques. These will be discussed in greater details in Section III of this paper. From the more detailed analysis presented there, the contribution of these factors to the measurement uncertainty of gate length is about 2.5% in terms of 3-σ percentage variation. It should be noted that this

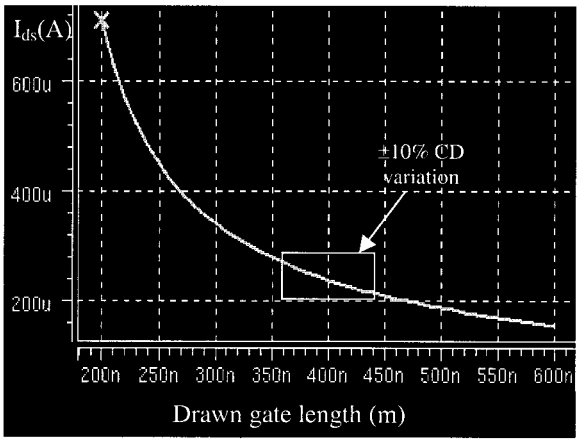


Fig. 2. The curve of saturation current vs. drawn gate length obtained from the transistor model provided by the broker of the fabrication process (MOSIS). Within $\pm 10\%$ CD variation around the nominal drawn CD, the relationship can be treated linear.

estimation is based on the 0.35-μm process used to fabricate the test chips. For more advanced technologies, the contributions by other nongate-length factors may increase. However, some techniques described in Section III can be applied to reduce or cancel their contributions, such as measuring the current at different bias conditions or scaling up the gate width. Further research is needed to verify the practical limits of these techniques.

C. Design of Overlay Test Structure

The layouts of the overlay test structures are shown in Fig. 1(b). The x overlay test structure consists of a vertical transistor with the poly gate between the source and drain diffusion regions. An x overlay error between the poly and the diffusion layers will change the transistor width, thus the transistor saturation current. Therefore, overlay variation can be determined by measuring I_{Dsat} variation. It should be noted that this transistor is different from an ordinary transistor as the poly gate extends on the right side. It can be modeled as two parallel transistors: one ordinary transistor on the left and one transistor on the right formed by the extending poly. Since an x overlay error will only change the width of the left transistor and will not change the current of the right transistor, the ratio of current change vs. x overlay is the same as treating this test structure as an ordinary transistor. The error of inferring overlay variation from saturation current can be estimated in the same way as that of CD metrology, because in both cases the variations are estimated from the transistor saturation current. The only difference is that the gate length in this case needs to be much larger than gate width in order to minimize the influence of gate length variations. Assuming a gate length of 2.4 μm and a gate width of 0.8 μm, the 3-σ error in determining overlay variation can be estimated to be about 13 nm. Note that this 3-σ error is mainly from the contribution of gate length variations. Therefore, by scaling up the gate length, the 3-σ error can be reduced.

Different from the x overlay test structure, the y overlay test structure consists of two symmetric horizontal transistors [Fig. 1(b)]. A y overlay error Δy will change the widths of the two transistors by $+\Delta y$ and $-\Delta y$ respectively, thus causing a difference between the saturation currents of the two transistors.

$$\Delta I_{\mathrm{Dsat}} = \frac{\mu_n C_{\mathrm{ox}}}{2} \frac{W}{L} (V_{\mathrm{GS}} - V_T)^2 \cdot \frac{2\Delta y}{W}.$$

Notice the factor of 2 before Δy, which means that this current difference is twice as sensitive to overlay error as that of the x overlay test structure. Therefore, the 3-σ error in determining y overlay variation is 6.5 nm.

D. Design of Addressing Circuits

The addressing circuits used to rapidly address the arrays mainly consist of decoders commonly used in memory circuits. Given a digital address signal that represents a particular row and column in the array, the cell at the corresponding location can be selected. However, unlike commonly done in a memory chip, we did not break the array into smaller blocks and address each block with separate decoders, as a uniform array is preferred to study the spatial distributions of parametric variations and their spectra. Therefore, large decoders addressing 256 lines have to be used. Two-staged decoders were used and their sizes were optimized in order to obtain fast access speed. The worst-case access time from the pins of the chip to the testing cells was estimated to be 24 ns. In addition, the current flow paths by which the saturation currents of selected transistors are measured have been carefully designed so that the currents are predominantly determined by the selected transistors, not the wires or switch transistors. Wide metal wires and pass transistors with large gate widths are used in the current paths to reduce their resistance. The estimated resistance from the measurement pin to the source terminal of the selected transistor is less than 5% of the resistance of the transistor under test. Furthermore, this parasitic resistance may be estimated by measuring the transistor current at two voltages and this component can be subtracted from the error analysis.

Also, the chip was designed to provide access from the pins of the chip to all gate voltages, substrate voltages, and source voltages of the transistors in both CD and overlay arrays, thus we have the flexibility to test the transistors in different voltage biases. As we will see in Section III-B, for example, we can measure transistor current at two different gate voltages to eliminate the influence from the variations of transistor threshold voltage.

III. Testing Configuration, Error Analysis, and Verification

A. Testing Configuration

Testing was done on packaged chips. Fig. 3 shows the testing configuration. An HP 16505 logic analyzer was used to provide the digital address signals to the chip at a selected clock speed. An operational amplifier chip was electrically connected to the pin of the chip to be measured and converted the current to voltage, and the voltage waveform was recorded with a digital oscilloscope. The value of the resistor determines the current voltage conversion ratio. The operational amplifier also has another function of setting the pin it is connected to at the desired voltage. The chip under test, the operational amplifier, the resistor, and all the necessary electrical connections were located

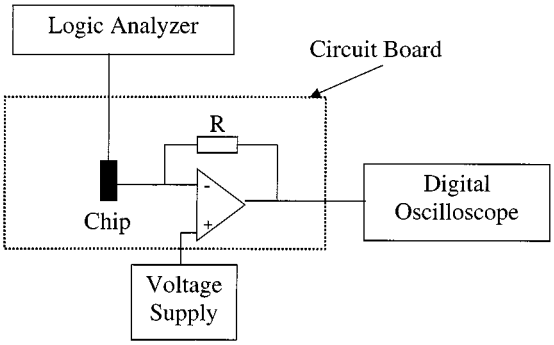


Fig. 3. Testing configuration. The chip is addressed with a logic analyzer and the current is measured with a digital oscilloscope.

on a 3M solderless breadboard. The chip was tested under 3.3 V.

The clock speed of the logic analyzer determines how fast the cells can be consecutively accessed and measured, therefore directly determines the throughput. Fig. 4 shows a segment of the current waveform recorded by the oscilloscope when the CD array was addressed at a clock cycle time of 5 μs. The change of current magnitude within each clock cycle represents CD variations of the addressed transistors, because within each clock cycle the address signal changes and a different transistor is selected. As we can see, however, there is a transient effect taking place at the beginning of every cycle. The current, which is proportional to the measured voltage, is rapidly ramped up and then oscillates with decreasing amplitudes and finally settles down. How fast it settles determines the fastest clock speed that can be set. Since the cell access time determined by the internal addressing circuits has been calculated to be less than 24 ns, it is believed that this transient effect is mainly due to the external testing circuits. Particularly, it may be due to the charging and discharging the parasitic capacitance at the inverting input of the operational amplifier, where the charging and discharging current flows from the output of the operational amplifier through the resistor. In fact, the settle time was found to change with different combinations of operational amplifiers and resistors in the experiments. It can be reduced with a smaller resistor. However, reducing the resistance will decrease the current voltage conversion ratio, thus decreasing the precision of voltage measurements. Because of our very high-throughput, we are able to sacrifice speed and trade it for precision. In the rest of the measurements, a clock cycle time of 80 μs was chosen to provide the assuredness that the current has enough time to settle for all possible conditions, while maintaining the precision of voltage measurements. Even with the clock cycle time of 80 μs, measurement of a 128×128 array such as the CD array can still be finished within 1 to 2 s.

B. Error Analysis

There are two separate and distinct concepts in the error analysis of a metrology method: *accuracy* and *precision* [15]. Take CD metrology as an example, the *accuracy* of a CD measurement is defined as the closeness of agreement between a measured dimension and the true dimension; while the *precision* of a CD measurement is defined as the variability observed in repeated measurements under essentially the same conditions. In this section, both accuracy and precision will be analyzed using the CD array as an example. The other arrays can be analyzed in similar manners.

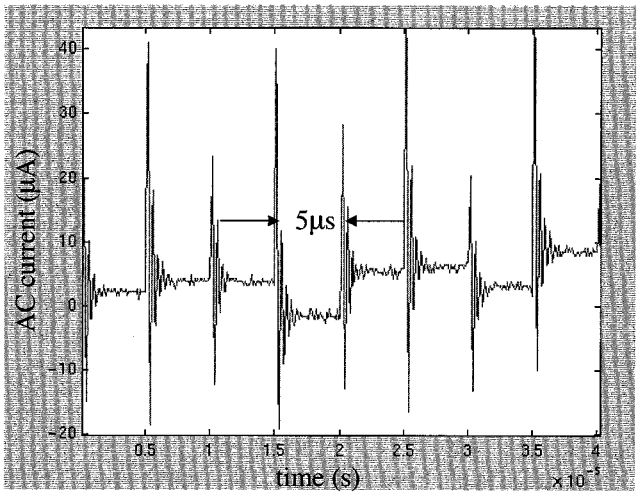


Fig. 4. A segment of the current waveform recorded when the CD array is addressed. It shows that the measurement time per cell can be as short as 5 μs.

The precision, or repeatability of the CD measurements, is mainly determined by the noise in the current (voltage) measurements. In our testing configuration, the current from the chip flows through the resistor and the voltage across the resistor is measured. This voltage is

$$V_o = I_{DS} \cdot R$$

where I_{DS} is the transistor saturation current flowing out from the chip, R is the resistance of the resistor. For the CD array, I_{DS} is about 220 μA, R is 15 kΩ. The 3-σ variation of I_{DS} (stabilized) is about 15 μA across the CD array. Therefore, we are trying to detect the variations of V_o with an average of 3.3 V and a 3-σ variation of 0.2 V. Detailed calculations [16] show that Johnson noise and shot noise of the current, fluctuations of the resistance, and noise caused by the operational amplifier and oscilloscope are all very small compared to the current variations we are trying to measure, thus the precision is mainly determined by other noise sources such as interference with environment and power supply noise. By repeating the measurements on a same part of the CD array, the 3-σ current measurement precision could be calculated to be about 2 μA, which is 0.9% of the average testing current.

It should be noted that precision can always be improved by repeated measurements [15]. This is especially convenient for our active electrical metrology because of its high throughput. In the rest of the tests, each test structures are repeatedly measured 16 times and averaged automatically by the oscilloscope, therefore the precision is conveniently improved to 0.23%. For the CD array, this means a precision of 0.8 nm.

Compared to precision, accuracy, or systematic uncertainty, is more difficult to quantify because the "true CD" is unknown. To calculate accuracy, the conventional metrology methods such as CD-SEM have to deal with many factors such as sidewall angle and irregular shape of the transistor due to corner rounding. In active electrical metrology, because effective gate length is measured directly from current, we do not have these same geo-

TABLE I
SOURCES OF SYSTEMATIC UNCERTAINTIES USING TRANSISTOR SATURATION CURRENT TO DETERMINE CD (GATE LENGTH) VARIATIONS, THEIR SOLUTIONS, AND THEIR CONTRIBUTIONS TO THE TOTAL SYSTEMATIC UNCERTAINTIES IN DETERMINING CD VARIATIONS AFTER ADOPTING THE SOLUTIONS

Sources of uncertainties	Explanation	Solution	Contribution to uncertainties
Gate width	$I_{Dsat}=\frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS}-V_T)^2$	Increase W W=4L	10% ΔW -> 2.5% ΔL
Threshold voltage	$V_T=V_T(N_A,X_{OX})$ $I_{Dsat}\sim(V_{GS}-V_T)^2$	Use difference of square root of I_{sat} at two gate voltages	~0%
Oxide thickness	$V_T=V_T(N_A,X_{OX})$ $I_{Dsat}\sim C_{ox}=\frac{\varepsilon_{ox}}{X_{ox}}$	Averaging among stepper fields	1% X_{ox} -> 1% ΔL

metrical ambiguities. Instead, we should consider the influence on transistor current by the variations of gate width, channel doping, and gate oxide thickness. Their influence can be reduced by careful design of the cell and measuring and data analysis method.

Table I illustrates the solutions to improve accuracy for the CD array and the magnitudes of the uncertainties after implementing the solutions. The influence of transistor width on saturation current can be reduced by designing transistors with a large width. Assuming both gate length and gate width have the same 3-σ variation of 40 nm (which is 10% of the minimum drawn gate length). By increasing gate width to 1.6 μm (four times the drawn gate length), its percentage variation is reduced to 2.5%. From the saturation current equation, this will induce 2.5% current variation, therefore the uncertainty in the determination of transistor length caused by gate width variation is reduced to 2.5%. The influence of V_T can be eliminated by using the difference of $\sqrt{I_{\rm Dsat}}$ at two different gate voltages instead of $I_{\rm Dsat}$ to calculate gate length. The influence of $C_{\rm ox}$ is always coupled with the gate length. It can be reduced by using thicker gate oxide if we have this flexibility. However, even if we do not have this flexibility, the uniformity of the gate oxide layer is believed to be very good within a small area such as several mm^2 [9]. Its variation is typically about 1% of the average gate oxide thickness.

Taking all the above into consideration, the accuracy, in terms of a 3-σ value of the percentage error, of determining gate length is estimated to be

$$\left\{\left[\frac{0.04}{1.6}\right]^2+[0.01]^2\right\}^{1/2}\approx 0.027=2.7\%$$

where the transistor width is 1.6 μm and width variation is assumed to be 40 nm, the percentage variation of gate oxide thickness is assumed to be 1%. In this calculation it is assumed that the error sources are independent thus their variances can be added. It is also assumed that the transistor model which is used to map gate length variations from measured saturation current is accurate.

In the above calculation, a 1% variation of gate oxide thickness is assumed. However, because of the high throughput of active electrical metrology and a difference between the spatial characteristics of the variations of gate length and the variations of gate oxide thickness, the contribution from the variations of gate oxide thickness can be significantly reduced by data averaging. As most of the short-range CD variations are caused by mask and stepper optics, they are predominantly stepper-field periodic; while oxide thickness variations are mostly caused by wafer processing steps, e.g., growth rate variations across the wafer, thus they are predominantly nonstepper-field periodic and varies across the whole wafer. Therefore, with the very high throughput of active electrical metrology of only 1–2 s per chip, we can measure many chips from the same location in different stepper fields and average among the chips. In this way the contribution from oxide thickness variations are significantly reduced. For the CD array, we have measured 23 chips and their average is mainly determined by the variations of gate width. Therefore, by simply designing results. This leads to an accuracy of

$$\left\{\left[\frac{0.04}{1.6}\right]^2+\frac{[0.01]^2}{23}\right\}^{1/2}\approx 0.025=2.5\%.$$

For the CD array with nominal on-wafer CD of 350 nm, this means an accuracy of 9 nm. From the above equation we can see that the accuracy wider transistors, its influence can be further reduced.

To calculate the total error, the accuracy and the precision should be summed up, which is

$$\left\{\left[\frac{0.04}{1.6}\right]^2+\frac{[0.01]^2}{23}+\frac{[0.0023]^2}{23}\right\}^{1/2}\approx 0.025=2.5\%.$$

C. Testing Verification

The functionality of the chip was verified by a number of measurements. First, a subarray of 16×16 test cells in the top-left corner of the CD array was measured with different sets of operational amplifiers, clock speeds, addressing sequences, and settings of the oscilloscope (including different time domain sampling). Fig. 5 shows the resulting gate length variations in this subarray measured under two sets of conditions. It can be seen that the two plots are almost the same. This implies that the measurement results are independent of the test conditions, and suggests that there are no interactions between the test cells or addressing signals.

Then, the calibration array consisting of transistors of different widths and lengths was measured. The calibration array contains 13 columns [Fig. 1(a)]. Each of the 13 columns has 16 transistors with same designed width and length, but each column has different widths or lengths. The measured saturation currents in each column were averaged among the 16 transistors in the column and compared with the transistor model supplied by MOSIS. Fig. 6(a) shows saturation current vs. designed L, while Fig. 6(b) shows saturation current vs. designed W/L. From the figures we can see good agreements with the model. It should be noted that the W's and L's in the figures are the designed values. The actually printed W's and L's can deviate from their designed values due to variations during the

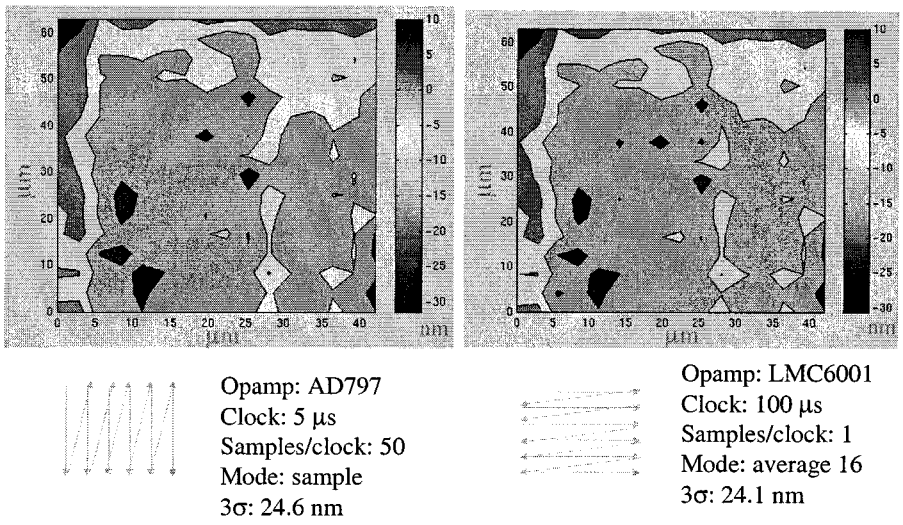


Fig. 5. Measured spatial variations of gate lengths in a corner of the CD array. The different sets of measuring conditions are illustrated below the contours, where the arrows illustrate the scanning sequences. The two different sets of conditions gave the same results.

manufacturing processes, thus causing deviations of the measured currents. For example, in Fig. 6(b) the point with drawn W of 0.6 μm and drawn L of 0.4 μm has relatively large deviation in measured current. This is believed to be caused by the fact that this column of transistors has the smallest gate length and width, and is located at the very left edge of the calibration array. Therefore, their gate lengths and widths are most sensitive to any proximity effects related to their positions in the array, which is commonly observed in transistors or circuits near the edge of an array [10]. Because of their small gate length and width, these dimensional variations caused by the proximity effects have large percentage variations, thus causing the large deviation of the measured current.

In addition to the calibration array for CD measurements, programmed x overlay error of 200 nm was designed into the overlay test structures in the last column of the overlay array. The measurements from the last column and the rest of the array can be used to determine the ratio of overlay variation and current variation. The ratio was calculated to be 57.7 nm/μA. For the double-transistor y overlay test structure, the current difference is twice as sensitive to overlay variation as the single-transistor overlay test structure, therefore its ratio is 28.85 nm/μA.

To further verify the validity of the test chip approach and establish common basis with previous SRAM research, a 64×64 SRAM array was designed into the test chip. This SRAM uses different design rules, has a different layout, uses different addressing and measuring circuits, and was fabricated on a different process than the previous SRAMs. By comparing the results from this SRAM array and our previous results, further confidence about the test chip approach can be gained. In fact, the same type of cell-periodic and layout dependent CD variations was found to dominate the error spectrum of the results of this SRAM array [Fig. 7(a)]. In this plot, the spatial variations are shown in spatial-frequency domain by Fourier transform [7], [10]–[14]. By checking the layout of the SRAM [Fig. 7(b)], the corresponding four-cell and two-cell periodicities due to proximity and overlay can be identified. These results are very consistent with those from the previous SRAMs. Detailed comparisons and explanations of the SRAM results have been discussed in another paper [17].

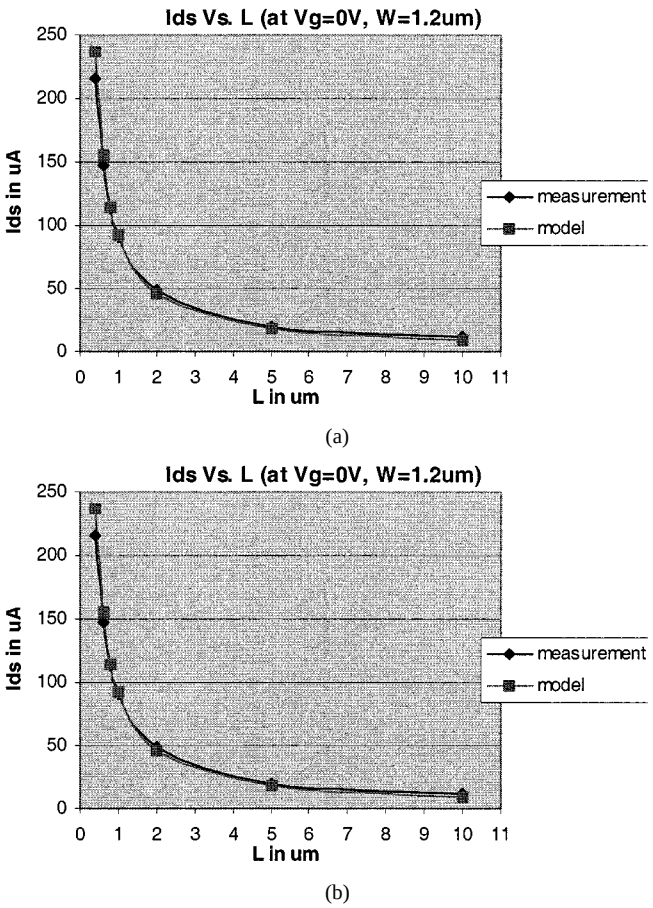
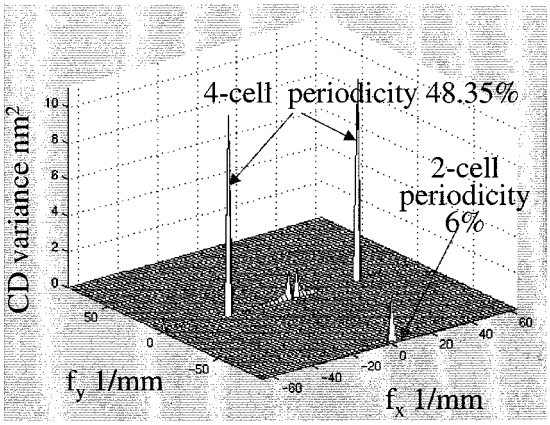
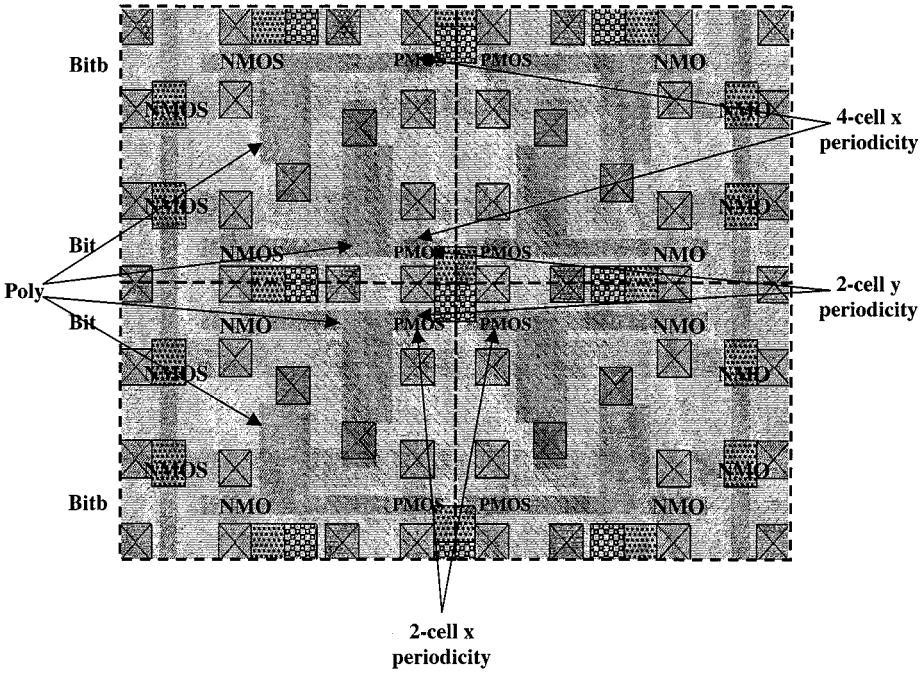


Fig. 6. Measurements from the CD calibration array (PMOS) were consistent with the transistor model, and verified the functionality of the test chip. (a) Measured saturation currents of transistors with different gate lengths. The currents were measured with a fixed gate voltage of 0 V and the transistors plotted have the same widths of 1.2 μm. (b) Measured saturation currents of transistors with variously sized gate lengths and widths. The currents were plotted against the ratio of width and length, W/L. The coefficient of determination, r^2, was calculated to be 0.97, which indicates good agreement with model.



(a)



(b)

Fig. 7. (a) Measurement results from the SRAM array are shown in spatial-frequency domain by Fourier transform. The peaks represent cell-periodic CD variations. (b) The layout of the SRAM two-cell and four-cell periodicities can be found in the layout. The results from the SRAM are very consistent with our previous study of SRAMs although the layouts and processes are significantly different.

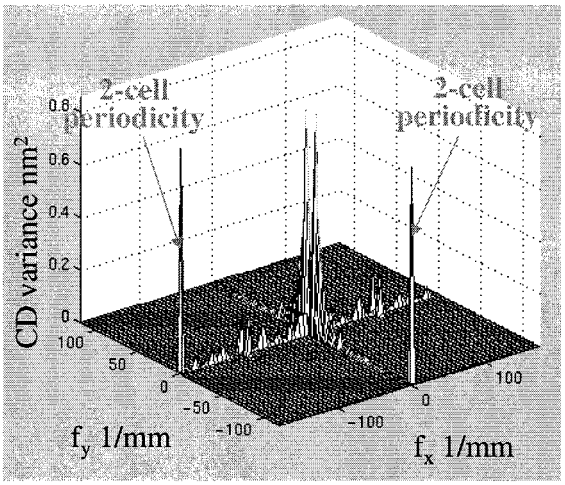
IV. Results

A. Short-Range CD Spatial Variations

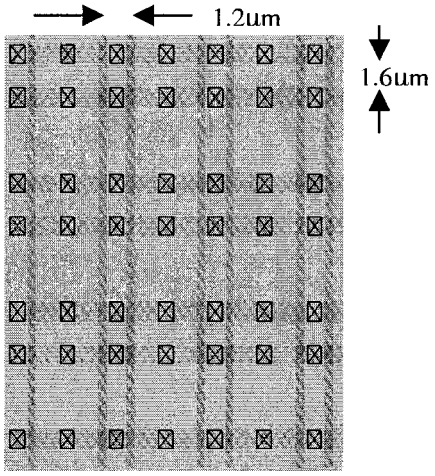
As the functionality of the test chip has been verified and the relationship between CD/overlay and current has been established, the results of the CD and overlay arrays can be analyzed. We will first discuss the short-range CD spatial variations measured from the CD array.

Fig. 8(a) shows the result from the CD array in spatial-frequency domain, where the z axis represents the CD variance σ^2 at different spatial frequencies. The procedures of data analysis are as the following: The measured current variations were translated into gate length variations by the current vs. gate length relationship calculated from the transistor model provided by MOSIS. The gate length variations were then averaged among the 23 chips, and discrete Fourier transform was done. The calculated 3-σ CD variation is 11.9 nm, which is quite significant compared to the usual error budget for a 0.35-μm process (3-σ of 35 nm using the 10% rule [18], [19]).

Although each transistor in the CD array is much more isolated than that in the SRAM array, we still find the two-cell (two-transistor) periodicities in both x and y directions which are represented by the two dominant peaks in Fig. 8(a). Checking the layout of the CD array [Fig. 8(b)], we can see that the poly layer has a corresponding two-cell periodicity in x direction, while the diffusion layer has a corresponding two-cell periodicity in y direction. Similar to previous study of SRAMs [7], [10], these layout differences may have caused the observed current periodicities. We should notice that transistor gate width (determined by the diffusion width) may have affected the transistor saturation current here, thus causing the two-cell periodicity in y direction. This is, however, consistent with our error analysis in



(a)



(b)

Fig. 8. (a) CD variations in the single transistor array in the MOSIS test chip. (b) The layout of the single transistor array. Because the transistor layout is mirrored in order to share contacts therefore reducing the size of the array, the transistors are not in equal spacing, therefore causing two-cell periodicities in both x and y directions.

Section III, which estimates that the impact from gate width is less that 2.5%. From the plot we can see that the variance of the peak representing the two-cell periodicity in y is about 0.8 nm^2, or its 3-σ value is 2.7 nm. This 3-σ value is only 0.7% of the drawn gate length of 0.4 μm, which is much smaller than 2.5%. Also, comparing with the CD variations from the SRAM array in the same test chip (Fig. 7), the CD variations in the CD array are much smaller than those in the SRAM array. This is probably due to the simpler layout of the CD array, thus there are less pattern-sensitive variations. These results also illustrate that our measurement and data analysis techniques have the potential to detect periodic errors with much smaller magnitudes than the estimated error magnitude, although care should be taken to separate them from the impact of other factors, especially that of gate width which can also be spatially periodic.

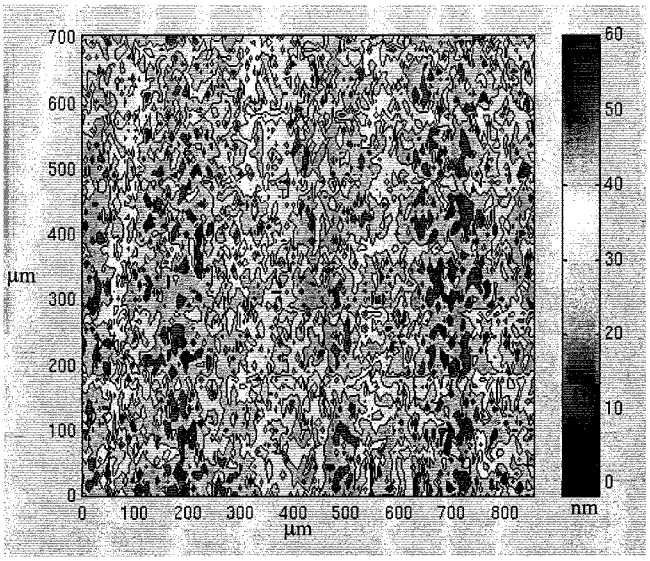
B. Short-Range Overlay Spatial Variations

In this paper, only the y overlay results will be presented. The y overlay test structures form a uniform 127×64 grid within an 856.8 μm $\times$ 705.6 μm area. The grid spacing is 6.8 μm $\times$ 11.2 μm. We have measured 23 chips from the same wafer. The measured current differences were translated into overlay variations by the ratio calculated in Section III using the on-chip overlay calibration test structures. Here we use the calibration result instead of transistor model because the overlay test structure does not consist of regular transistors. Fig. 9(a) shows the typical contour plot of the overlay variations from one of the chips. While the average overlay error is 29.2 nm, the overlay error varies within the array between 0 and 60 nm with a 3-σ value of 25.6 nm. As we can see from the contours, there are significant high-spatial-frequency overlay variations even within this small area of the overlay array. The chip average and 3-σ overlay variations of all the 23 chips are shown in Fig. 9(b). Although the overlay averaged within each chip varies significantly, the 3-σ variation remains almost a constant from chip to chip. This implies that the chip-to-chip average overlay variation and within-chip overlay variation are from two different error sources. The chip-to-chip average overlay variation is probably due to interfield overlay errors such as stepper alignment errors, therefore the chip average can vary significantly from chip to chip as each stepper field is not aligned the same. While the within-chip overlay variation may arise from mask or stepper optics, as these intertransistor variations are unchanging from stepper field to stepper field.

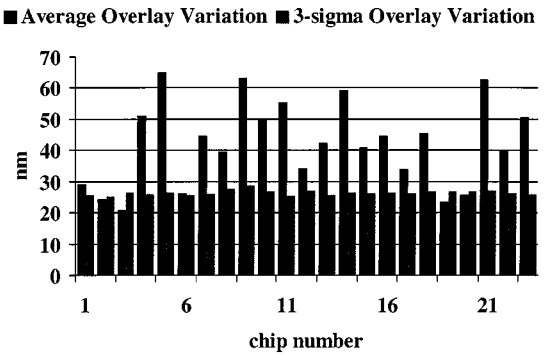
To further study short-range overlay variation, the within-chip overlay variation should be analyzed in greater details. When we take a closer look at the shape of the contours in Fig. 9(a), we can see some stripes running horizontally across the whole array area, and within each stripe the variations seem to have similar distributions. This error spatial signature is similar to that expected from raster-scanned mask writers such as the MEBES series electron-beam lithography systems or the CORE or ALTA series laser lithography systems which are commonly used in the mask industry today. Fig. 9(c) shows the averaged overlay variations among 23 chips, and Fig. 9(d) shows their spatial-frequency spectrum, which is the power spectrum of discrete Fourier transform of (c). From the plots we can see that the stripe periodicity becomes more obvious after the averaging, which further indicates that the stripe periodicity may come from the mask. The averaged overlay variations have a mean of 42.3 nm and a 3-σ value of 16.2 nm. In the spatial-frequency domain, the highest two peaks with a spatial period of 102.4 μm and their high-frequency harmonics represent this stripe periodicity. Thus the stripe length can be determined to be 102.4 μm. Detailed analysis indicates that the stripe-periodic overlay variations may be due to the scan nonlinearity of the beam of the mask writer [20].

V. CONCLUSION

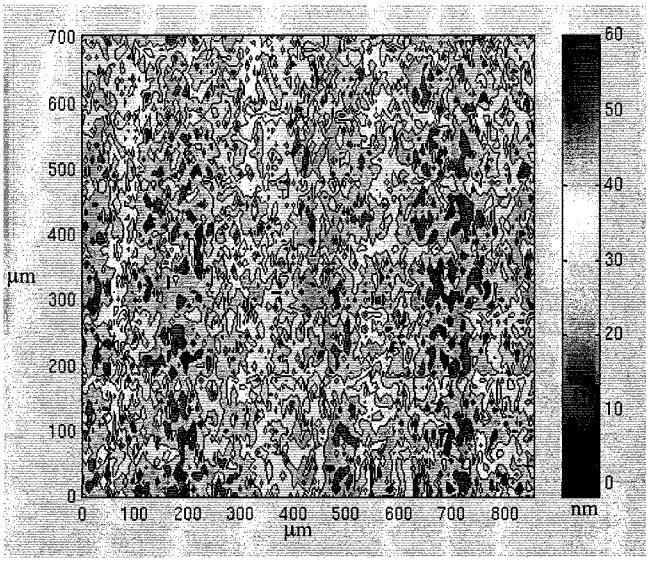
Active electrical metrology using on-chip addressing circuitry and active addressable arrays has been proved as a high-throughput and high-density process metrology method with test chips fabricated from a HP 0.35-μm process. A measurement speed of 5 μs per data point has been achieved with the testing configuration presented. With this improved



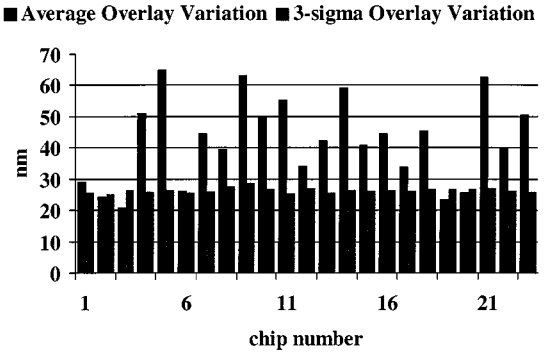
(a)



(b)



(c)



(d)

Fig. 9. Y overlay results. (a) Y overlay variations measured from the overlay array of chip 1. (b) The mean and 3-σ y overlay variations of all the 23 chips. (c) Y overlay variations averaged among the 23 chips. (d) Y overlay variations averaged among the 23 chips shown in spatial frequency domain.

measurement speed, large arrays containing 10^4 test structures can be measured within 1–2 second. Therefore, large amounts of data can be collected very rapidly to study the spatial distribution of intertransistor variations. The intertransistor CD and overlay variations measured with the test chips indicate that they are highly spatial-correlated and consume a significant portion of the overall error budget. These intertransistor CD and overlay variations exhibit periodic variations that can be related to the layout (CD) and the scan stripes of the mask lithography tool (overlay).

ACKNOWLEDGMENT

The authors wish to express their sincere appreciation to the Stanford University Mask Advisory Group for its advice and kind support.

X. Ouyang would like to thank M. McCord and K. Mai of Stanford University, T. Deeter of Intel, G. Burns and C. Hamaker of Applied Materials, and the customer service team of MOSIS for their helpful discussions and assistance.

REFERENCES

[1] J. K. Kibarian and A. J. Strojwas, "IC manufacturing diagnosis based on statistical analysis techniques," *IEEE Trans. Comp., Hybrids, Manufact. Technol.*, vol. 15, no. 3, pp. 317–321, June 1992.

[2] D. G. Chesebro, J. W. Adkisson, L. R. Clark, S. N. Eslinger, M. A. Faucher, S. J. Holmes, R. P. Mallette, E. J. Nowak, E. W. Sengle, S. H. Voldman, and T. W. Weeks, "Overview of gate linewidth control in the manufacture of CMOS logic chips," *IBM J. Res. Develop.*, vol. 39, no. 1/2, pp. 189–200, Jan./Mar. 1995.

[3] M. O'Leary and C. Lyden, "Parametric yield prediction of complex, mixed-signal IC's," *IEEE J. Solid-State Circuits*, vol. 30, no. 3, pp. 279–285, Mar. 1995.

[4] C. N. Berglund, "A unified yield model incorporating both defect and parametric effects," *IEEE Trans. Semiconduct. Manufact.*, vol. 9, no. 3, pp. 447–454, 1996.

[5] B. E. Stine, D. S. Boning, and J. E. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Trans. Semiconduct. Manufact.*, vol. 10, no. 1, pp. 24–41, 1997.

[6] C. Yu, T. Maung, C. Spanos, D. Boning, J. Chung, H. Liu, K. Chang, and D. Bartelink, "Use of short loop electrical measurements for yield improvement," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 150–159, May 1995.

[7] X. Ouyang, T. Deeter, C. N. Berglund, M. A. McCord, and R. F. W. Pease, "High-throughput, high-spatial-frequency measurement of critical dimension variations using memory circuits as electrical test structures," *J. Vac. Sci. Technol. B*, vol. 6, pp. 3655–3660, Nov./Dec. 1999.

[8] B. R. Blaes and M. G. Buehler, "SEU/SRAM as a process monitor," *IEEE Trans. Semiconduct. Manufact.*, vol. 7, no. 3, pp. 319–324, Aug. 1994.

[9] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, pp. 2216–2221, Nov. 1994.

[10] X. Ouyang, T. Deeter, C. N. Berglund, R. F. W. Pease, and M. A. McCord, "High-spatial-frequency MOS transistor gate length variations in SRAM circuits," in *Proc. 2000 Int. Conf. Microelectronic Test Structures*, Mar. 13–16, 2000, pp. 25–31.

[11] C. N. Berglund, N. I. Maluf, J. Ye, G. Owen, R. Browning, and R. F. W. Pease, "Spatial correlation of electron-beam mask errors and the implications for integrated circuit yield," *J. Vac. Sci. Technol. B*, vol. 10, no. 6, p. 2633, 1992.

[12] J. Ye, C. N. Berglund, J. Robinson, and R. F. W. Pease, "A review of errors in masks written on a variety of pattern generators," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 319–325, Aug. 1995.

[13] L. Han, W. Wang, M. A. McCord, C. N. Berglund, R. F. W. Pease, and L. S. Weaver, "Practical approach to separating the pattern generator-induced errors from the blank/process-induced mask CD errors using conventional market measurements," *J. Vac. Sci. Technol. B*, vol. 15, no. 6, pp. 2633–2637, Nov./Dec. 1997.

[14] X. Ouyang, C. N. Berglund, M. A. McCord, R. F. W. Pease, C. Spence, and H.-Y. Liu, "An economical sampling algorithm using Fourier analysis for mapping wafer CD variations," *J. Vac. Sci. Technol. B*, vol. 16, no. 6, pp. 3655–3660, Nov./Dec. 1998.

[15] L. W. Linholm, R. A. Allen, and M. W. Cresswell, "Microelectronic test structures for feature placement and electrical linewidth metrology," in *Handbook of Critical Dimension Metrology and Process Control*, K. Monahan, Ed. Bellingham, MA: SPIE Optical Eng. Press, 1994.

[16] X. Ouyang, "Active Electrical Metrology of Spatial Variations in Integrated Circuits," Ph.D., Stanford Univ..

[17] X. Ouyang, T. Deeter, C. N. Berglund, R. F. W. Pease, and M. A. McCord, "High-throughput, high-spatial-frequency mapping of MOS transistor gate length variations in SRAM circuits," IEEE Trans. Semiconduct. Manufact., submitted for publication.

[18] G. Owen, R. F. W. Pease, N. I. Maluf, J. Ye, and C. N. Berglund, "A rational argument for the impracticability of 1× reticles," *SPIE*, vol. 1809, pp. 39–48, 1992.

[19] The International Technology Roadmap for Semiconductors (1999). [Online]. Available: http://public.itrs.net/files/1999_SIA_Roadmap/Home.htm

[20] X. Ouyang, C. N. Berglund, and R. F. W. Pease, "High-speed mapping of intert-ransistor overlay variations using active electrical metrology," in SPIE Microlithography Conf., Mar. 2001.

Xu Ouyang received the B.S. degree in physics from Beijing University, China, in 1996, the M.S. degree in electrical engineering and the Ph.D. degree in applied physics from Stanford University, Stanford, CA, in 1999 and 2001, respectively.

He was a graduate research assistant in Ginzton Lab and Center for Integrated Systems in Stanford University from 1996 to 2001, where he worked on metrology and characterization of errors in microlithography, with an emphasis on those related to CD control and the error transfer from mask to wafer. He held an internship in KLA-Tencor, Milpitas, CA in the summer of 1997 as a system engineer. He is now a consulting engineer at PDF Solutions, San Jose, CA.

C. Neil Berglund (S'60–M'62–SM'77–F'83) received the B.Sc. degree from Queen's University, Kingston, Canada, in 1960, the M.S.E.E. degree from the Massachusetts Institute of Technology, Cambridge, in 1961, and the Ph.D. degree from Stanford University, Stanford, CA, in 1964, all in electrical engineering.

From 1964 through 1972, he was in the Semiconductor Device Laboratory at Bell Telephone Laboratories in Murray Hill, NJ. From 1972 through 1978, he held several senior technical management positions at Bell Northern Research in Ottawa, Canada. From 1978 through 1983, he was Director of Technology Development at Intel Corporation, and was responsible for setting up the Technology Development group in Oregon. From 1983 through 1987, he was founder, President, and CEO of Ateq Corporation. Ateq, a semiconductor equipment company, developed a laser-based lithography system that currently is used to produce most of the photomasks used by the semiconductor industry. Since 1987, he has been President of Northwest Technology Group, a microelectronics consulting firm based in Tigard, OR. Since 1990, he has also served as a Consulting Professor in Electrical Engineering at Stanford University, and in addition in 1993 joined the Oregon Graduate Institute of Science and Technology as Professor in the Electrical and Computer Engineering Department with a joint appointment in the Management in Science and Technology Department.

Dr. Berglund has held numerous positions within the IEEE Council system. He is co-recipient of the 1995 BACUS Prize (SPIE) for contributions to photomask technology.

R. Fabian W. Pease (M'75–SM'83–F'89) was a Radar officer RAF from 1955 to 1957. He received the B.A. degree in natural sciences in 1960 and the M.A. and the Ph.D. degrees in electrical engineering in 1964, all from Cambridge University, Cambridge, U.K.

He was an Assistant Professor at the University of California at Berkeley from 1964 to 1967. He was with Bell Laboratories, NJ, from 1967 to 1978. He is now Professor of Electrical Engineer, Stanford University, Stanford, CA. His research activities include microlithography and microfabrication technologies and their applications. He spent the academic year 1993–1994 with Affymetrix Inc., applying photomask technology to the synthesis of DNA arrays. He also spent the academic years 1996–1998 at DARPA-ETO managing programs in microelectronics and microprinting.

Dr. Pease is a member of the NAE. The organizations he has served as a consultant include IBM, Xerox, ETEC, LLNL, Ultratech Stepper and Affymetrix.