Overlay Simulator for Wafer Steppers

William H. Arnold

Advanced Micro Devices, Inc.
901 Thompson Place, MS 79
Sunnyvale, CA 94088

Abstract

The impact of wafer stepper overlay errors on device yields and design rules are studied. First, the classical Lynch model for normally distributed sizing and overlay errors is reformulated for orthogonal geometries. Then the distribution of overlay errors in the linear Perloff model describing global alignment is derived. Finally, a Monte Carlo program, OVS, for simulating stepper overlay errors is introduced. OVS is used to determine the impact of individual component errors, such as those due to lens distortion or to mask making, on the overall distribution of errors.

Introduction

Pattern overlay tolerance is one of the key design rules for integrated circuit manufacturing. The current method of specifying overlay tolerances is to state, on the basis of simple error budgets, the value of the overlay error at a given high percentage of the total measured data. There are at least three major problems with this[1]. First, most error budgets treat all errors as random when in fact some are systematic. Second, these error budgets rarely take field size or wafer size into account. Finally, rarely are enough data points taken to ensure what is really important: that no point within the imaged array of chips be misregistered by more than the design rule.

An overlay simulation computer program, OVS, has been written to predict the total distribution of errors given as input the characteristic component error distributions in interfield errors (translation, wafer rotation and expansion, orthogonality) and intrafield errors (die rotation, magnification, trapezoid, and distortion). Reticle to reticle stacking errors are taken into account. These are used in a Monte Carlo simulation to generate the expected total error distribution over thousands of individual fields and wafers. The program can be used to predict the performance of a single stepper or of a large group of randomly mixed steppers.

The underlying model for the simulation is the linear interfield model introduced by Perloff[2] combined with the polynomial intrafield model introduced by MacMillen and Ryden[3]. Each error is given a systematic offset and a random component characterized by a Gaussian distribution of a given sigma. The simulation computes the errors found at F points per field, and W points per wafer, where the number of points and their locations are set by the user. If any point within the field exceeds the overlay tolerance it is noted by the program and the field is termed a "bad" field.

The fit of simulation data to experimental data is found to be quite good. One interesting result of the simulations is that many distributions are platykurtic, i.e., have less data in the tails of the distribution than the best fit Gaussian computed from the same data, a tendency noticed by many workers from experimental data. It is also found that as the size of systematic errors grow with respect to random errors, the more the distributions tend to be platykurtic.

Emphasis is placed in this report on determining the overall distribution of overlay errors given the distribution of component errors. First considered is a simple overlay model proposed by Lynch[4] which will allow development of the ideas advanced here against a classical background. Then consequences of the linear model[2] will be explored in terms of its predictions for global overlay error distributions. It is found, for example, that the linear model predicts a semicircular histogram of overlay errors on a given wafer aligned at two points. In the last section the overlay simulator model and details of the computer program are discussed. The effects of individual component errors on the entire error distribution are examined, assuming that the distributions of component errors can be measured and specified[5].

As a result of this work, a new proposal is made for the specification of overlay errors which corresponds more closely to the desires of chip designers. The proposal is called the "good fields rule", in which stepper overlay is quantified in terms of the percentage of individual fields which contain no error greater than the specification.

The good fields rule for total overlay satsfies many intuitions about how overlay affects chip yields and is more stringent than any presently offered by stepper manufacturers. While a large number of points within the usable image field may be properly registered, it only takes one bad point to cause circuit failure, resulting in complete yield loss on single die reticles and fractional yield on multidie reticles.

Lynch Model

Lynch[4] studied the problem of contact window limited yields for LSI devices. In his analysis, the process yield for defining and aligning a circular contact window over a larger circular pad of polysilicon is derived. Yield is defined as the contact window falling completely on the poly pad. The minimum contact window size is set by the aligner's resolution capability. The required size of the poly pad to achieve a given yield is then calculated from contact size, the process standard deviations for edge control for both the contact and the poly pad, and the contact to poly alignment standard deviation. Lynch's model assumes explicitly that the distributions of edge controls for contact and poly as well as the alignment errors are Gaussian.

Device designers typically do not set radial design rules but instead set rules along orthogonal directions. Thus a more germane problem is that of a square contact window falling on a larger square polysilicon feature. If we assume in this case that there is no asymmetry between edge placement or alignment precision in the X and Y directions then Lynch's expression can be recalculated. The geometry of the problem is shown in Figures 1a and 1b. The contact window is free to expand or contract with mean size C and standard deviation s_C. The poly pad likewise has mean size L and standard deviation s_1. The overlay error is assumed to have zero mean error and a standard deviation along the X or Y directions of s_0. Normal distributions are assumed for each.

The problem of the lithographer is to specify the smallest tolerance, T = 1/2(L - C), which will allow the contact to fall completely on the poly pad at a given high percentage level. The tolerance is the nominal distance between the edge of the contact to the nearest poly edge. In a given case, this distance will change to D = 1/2(l - c) where l is the true size of the pad and c the true size of the window. The distribution of D is also Gaussian[6] with mean T and standard deviation $(s_C^2 + s_1^2)^{1/2}$

$$f(D) = (2\pi(s_C^2+s_1^2))^{-1/2} \quad \exp(-(T - D)^2/\ 2(s_C^2+ s_1^2)) \tag{1}$$

The probability that the contact falls completely on poly is then the probability that D - |dx| and D - |dy| > 0. The probability that |dx| (or |dy|) < D for a given D is given by

$$P(|dx|, |dy| < D) = (1 - 2\bar{Q}(D / s_0))^2 \quad \text{where } \bar{Q}(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-t^2/2)dt \tag{2}$$

The probability then that the contact falls on the poly, that it yields at a given tolerance T, is then the convolution of eqn. 1 and eqn. 2:

$$Y(T) = (2\pi(s_C^2+s_1^2))^{-1/2} \int_{0}^{\infty} (1 - 2\bar{Q}(D/s_0))^2 \exp(-(T-D)^2 / 2(s_C^2+s_1^2))\ dD \tag{3}$$

This expression allows one to find how big to make the tolerance given normally distributed process variations in poly and contact sizing, and in overlay. As might be expected, it satisfies the root sum square error budget calculation typically used[7] to calculate the tolerance. This can be seen in Figure 2 in which Y(T) is plotted as a function of T for three different values of overlay precision. The contact sizing precision is 0.05 μm, poly precision is 0.03 μm, and the three overlay precisions are 0.05, 0.1, and 0.15 μm, all numbers at one standard deviation. For the 0.15 μm, one sigma case the root sum square of the sizing and overlay precisions is 0.22 μm. Y(T) = 0.68 at 0.22 μm, which is the percentage included within one standard deviation in a Gaussian or normal distribution, demonstrating the underlying consistency of the Lynch model.

From these simple considerations it can be seen clearly why overlay is such a key parameter for lithography. For example, assume that the smallest contact size which can be reliably defined with 0.05 μm, 1 sigma control is 1.0 μm. The poly width L = C + 2T for this case is 1.48 μm at the 99.7% yield level, and 2.00 μm for the case where overlay control is 0.15 μm, one sigma. The area of the poly pad can be cut by 45% by increasing overlay precision from 0.45 μm, 3 sigma, to 0.15 μm, 3 sigma. Since a CMOS transistor cell size is proportional to this area, it can be seen how sensitive device size is to overlay precision.

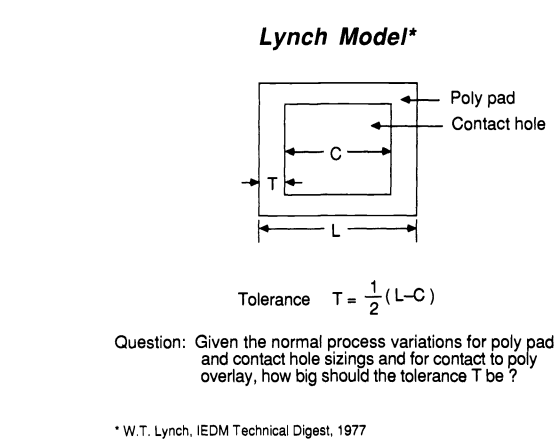


Figure 1a. Definition of the tolerance T in the Lynch problem.

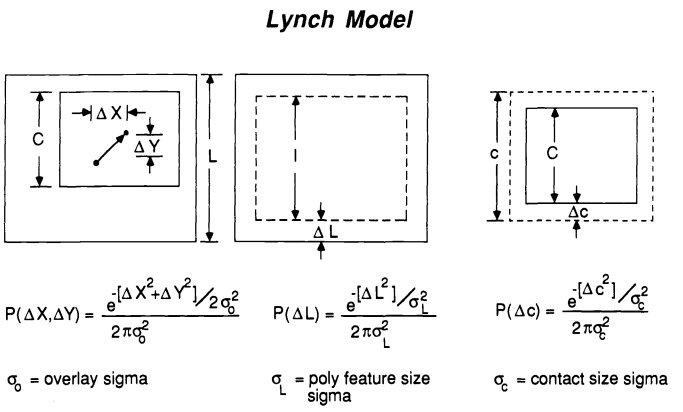


Figure 1b. Probability distributions for contact window and poly pad sizings, and for contact to poly alignment.

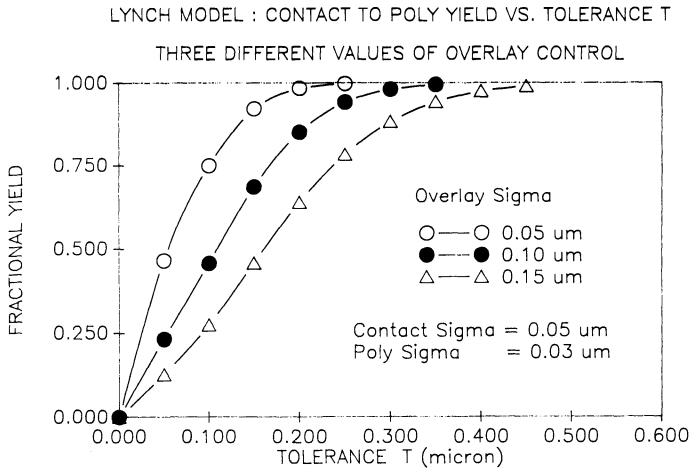


Figure 2. Fra ctional contact to poly yield versus the tolerance T for three different values of overlay precision.

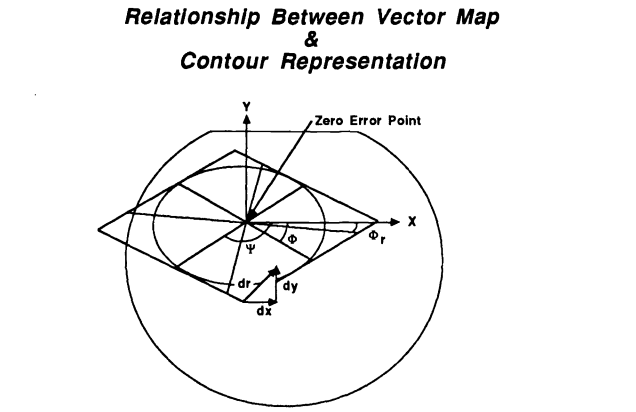


Figure 3. Relationship between the vector and contour representations of linear overlay errors.

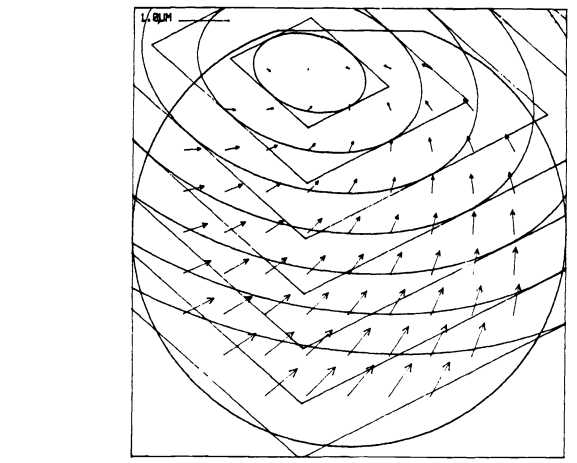


Figure 4. Example of the contour representation of linear overlay errors. T_x = .19 μm, T_y = .30 μm, θ_x = 7.3 ppm, θ_y = 3.9 ppm, E_x = 6.5 ppm, E_y = - 7.5 ppm.

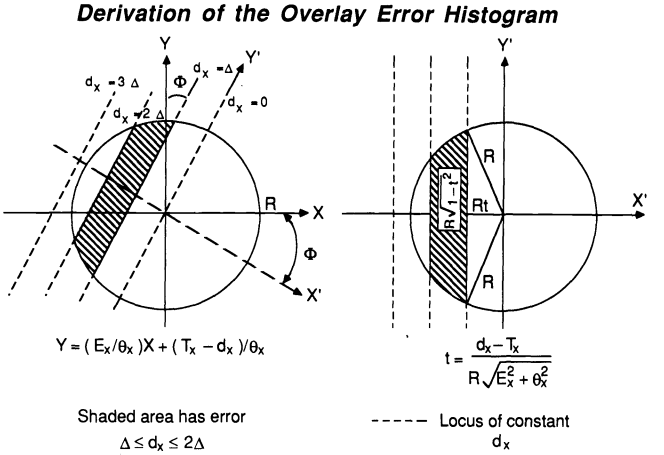


Figure 5. Derivation of the linear overlay error histogram by sectioning the circle and determining the relative area between successive equal error contrours.

Overlay Error Distribution for Global Alignment

Interfield registration errors made by reduction steppers using two-point global alignment can be modelled successfully by linear regression analysis as shown by Perloff and co-workers[2,8]. Kim and Ham[9] showed that assumption of the linear model of registration errors leads to the result that the loci of equal-valued errors in the Cartesian coordinates x and y on the wafer are straight lines which intersect at oblique angles. There is a single point in the xy (wafer) plane at which the registration error is zero, which can fall either on the wafer or outside it. Errors less than a constant absolute value fall within a parallelogram centered on the zero error point. Equal valued vector displacements have elliptical contours.

Registration errors on a globally aligned wafer are estimated independently to first order in the Cartesian coordinates x and y as

$$dx = T_x - \theta_x y + E_x x + s_x \tag{4}$$

$$dy = T_y + \theta_y x + E_y y + s_y \tag{5}$$

where T_x, T_y, E_x, E_y, θ_x, and θ_y are parameters formed by linear regression of a data set consisting of N sets of errors (dx, dy) at wafer locations (x, y). The s_x and s_y are residual errors not fitted and are usually associated with random stage stepping errors. The first order error parameters can be grouped into three simple geometrical classes : translation (T_x, T_y), rotation and orthogonality (θ_x, θ_y), and expansion (E_x, E_y).

The distribution of errors across a wafer predicted by the linear model can be determined by exploring the behavior of eqn (4) and (5) at different wafer locations. One interesting result shown by Kim and Ham is that the loci of equal-valued errors in x and y are straight lines. This can be seen by solving eqn (4) for y (assume $s_x = s_y = 0$) :

$$y = (E_x/\theta_x)x + (T_x - dx)/\theta_x \tag{6}$$

which is of the linear form $y = ax + b$, where the slope $a = (E_x/\theta_x)$ and the y intercept $b = (T_x - dx)/\theta_x$. Similarly, solution of eqn (5) yields $y = -(\theta_y/E_y)\,x + (T_y + dy)/E_y$. Simultaneous solution of eqns. (4) and (5) when $dx = dy = 0$ yields the position of the zero error point:

$$x_o = -\frac{(T_xE_y + \theta_xT_y)}{\theta_x\theta_y + E_xE_y} \; ; \quad y_o = -\frac{(E_xT_y - \theta_yT_x)}{\theta_x\theta_y + E_xE_y} \tag{7}$$

Depending on the wafer radius, the origin of coordinates and the magnitude of the errors in T, θ, and E, the zero error point can either fall on the wafer or outside it (i.e., there is no point on the wafer with perfect overlay). A simple interpretation is that there is a single point in the xy plane at which there is zero overlay error and that error in x and y less than $dx = \pm c$ and $dy = \pm d$ respectively are bounded by a parallelogram with sides described by

$$y = (E_x/\theta_x)x + (T_x \mp c)/\theta_x \tag{8}$$

$$y = -(\theta_y/E_y)x + (T_y \pm d)/E_y \tag{9}$$

The parallelogram is centered on the zero error point (x_o, y_o). This is illustrated in Figure 3. Lines of constant dx are inclined at an angle Φ with respect to the x axis, while lines of constant dy are inclined at an angle Ψ where $\tan\Phi = E_x/\theta_x$ and $\tan\Psi = -E_y/\theta_y$.

If one instead considers the locus of equal valued vector errors ($d_r^2 = (d_x^2 + d_y^2)^{1/2}$), then the contours are ellipses, also centered on the zero error point, with the semi-major axis inclined at an angle Φ_r with respect to the x axis:

$$\tan 2\Phi_r = \frac{2(\theta_yE_y - \theta_xE_x)}{(E_x^2+\theta_x^2)-(E_y^2+\theta_y^2)} \tag{10}$$

A simple illustration of these results is the special case of isotropic expansion error without translation or rotation. Traditional vector maps represent this type of error with radially directed arrows whose lengths grow as the distance from the center increases. This method suggests that the same error can be represented with circular contours, where the circles are centered at the origin and whose radii grow linearly with increasing distance from the origin. Contours of constant and equal dx and dy are squares. The case of pure rotation error without translation or expansion also yields

circular contours centered at the origin. The general case, where translation, rotation, and expansion errors are present, yields ellipses and parallelograms centered on the zero error point which is not the origin of coordinates. Figure 4 shows an example wafer for which both the vector map and contour representations are given. The error parameters are listed in the figure.

What relationship exists between the model parameters and the distribution of overlay errors on a single wafer? Since one can compute the straight line contours of constant dx or dy increments, imagine sectioning the wafer into areas between successive dx or dy increments (see Figure 5). Once this is done the relative size of each area can be calculated using standard relations for the sector of a circle which can then be used to construct the histogram of errors on the wafer. The mathematical details of the derivation are given in reference 12.

The histogram of x overlay errors, H(dx), on a wafer of radius R and with error parameters T_x, θ_x, and E_x is given by

$$H(dx) = \frac{2}{\pi} \frac{d}{R(\theta_x^2+E_x^2)^{1/2}} \left[1 - \frac{(dx - T_x)^2}{R^2(\theta_x^2+E_x^2)} \right]^{1/2} \qquad (11)$$

where H(dx) represents the percentage of wafer area within $\pm$ d/2 of dx and dx ranges between $T_x - R(\theta_x^2+E_x^2)^{1/2}$ and $T_x + R(\theta_x^2+E_x^2)^{1/2}$. The expression for the y histogram, H(dy), is identical to this with all subscripts x replaced by y.

This result explains why overlay errors across a single wafer which is aligned at two points are in general not normally distributed. The form of the histogram is semicircular! It is straightforward to calculate that the standard deviation of this distribution is $s = (1/2)R(\theta_x^2+E_x^2)^{1/2}$. Thus the entire distribution is contained within $\pm$ 2s of the mean T_x.

Figures 6a and 6b show the measured and modeled X axis overlay errors for the example wafer illustrated in Figure 4. The modeled histogram was calculated using equation 11. Random stage errors account for most of the difference between the modeled and measured data.

OVS : Overlay Simulator for Wafer Steppers

Anyone who has studied wafer stepper overlay specifications knows that it is a difficult job to translate the stepper vendor's specifications into numbers which device designers can use. The reasons are that the vendor specification usually refers to an ideal test case, in which wafers with nearly perfect alignment targets are used. Field size is usually restricted to values less than the maximum field. Only a limited number of points across the image field and across the wafer are actually measured. The vendor then guarantees that a given percentage of the total number of measurements will fall below the maximum specified overlay error. For example, one major stepper vendor specifies that overlay be measured at 17 points per image field and at 17 separate fields on each of three wafers. This gives a total of 17 X 17 X 3 = 867 data points per axis. Only 9 points per axis are allowed to exceed the overlay specification.

This type of sampling plan, however, does not guarantee that chips designed with an overlay tolerance equal to the vendor specification will yield at the same high percentage. The major reason is that it only takes one bad point, i.e., one location where the overlay rule is violated, to cause circuit failure and zero yield for that chip. It doesn't matter that 99% or more of the rest of the chip's area is overlaid within specification, the chip still doesn't yield. In the example sampling plan described, assume that matched steppers are used to print only one die per field. In the worst case, one might have a bad corner of the field due to lens to lens distortion differences in which the x overlay error is very close to the spec limit without considering alignment errors. The sampling plan would allow this point to exceed specification in 9 fields out of 17 X 3 = 51 total. A similar situation in the y direction added to this could lead to 18 fields out of 51 containing a point which violated the design rule. The vendor's specification that 99% of the errors were less than the given tolerance is met, yet only about 65% of the chips printed would yield.

One response to this reasoning is that such a situation is highly unlikely. Unfortunately, it is nearly as unlikely that only one field would contain all the overlay errors, so that the yield of image fields without bad points would in general be less than 98% (50 of 51). It must be remembered that modern devices have 15 or more mask layers, 5 or so of which require the tightest overlay specification, so that total overlay yield is roughly the yield of one alignment raised to the power 5 (or greater). Even at 98% single layer yield there is only 90% yield after five critical alignments and

82% after 10.

Given these considerations a new proposal is made for the specification of overlay errors which corresponds more closely to the desires of chip designers. The proposal is called the "good fields rule":

X% of all fields overlaid to a previous field will contain no bad points, i.e., errors greater than the specified overlay error.

X% can be a given high percentage, e.g., 99% or 99.7%, depending on device yield considerations. The good fields rule for total overlay satsfies many intuitions about how overlay affects chip yields and is more stringent than any presently offered by stepper manufacturers.

One serious logistics problem with the good fields rule is the relatively large number of measurements that need to be taken to ensure the specification directly. It is however possible to indirectly estimate the error distribution through computer simulation of the overlay process using as input the characteristic distributions of subcomponent errors such as reduction, die rotation, trapezoid, distortion, and so forth. To meet this need, an overlay simulation program, OVS, has been written. The program incorporates a Monte Carlo routine which simulates the alignment of many reticles and wafers and reports the statistics on all the errors found. The program can simulate the distribution of errors expected from a single stepper and from mixed steppers. Errors taken into account, assuming global alignment, are listed in Table 1.

Table 1 - OVS Input Error Parameters

Intrafield Errors - Single Stepper

1) random magnification error*	M
2) random reticle rotation error*	R
3) random reticle stacking error*	s_l

Mixed Steppers

4) relative distortion between two lenses i and j@	$D_l(i)-D_l(j)$

Interfield Errors

1) random translation offsets in x and y*	T_x, T_y
2) random rotation offset*	θ
3) random orthogonality offset*	$d\theta$
4) random symmetrical expansion (or scaling) offset*	E
5) random assymetrical expansion offset*	dE
6) random variations around offsets for:! translation, rotation, and expansion	
7) random stage errors&	ss_k

Field locations l = 1 to f; wafer locations k = 1 to w

* - constant for one reticle change
! - chosen for each wafer alignment
@ - constant for all reticle changes
& - chosen at the center of each field k

The program starts by reading intrafield distortion errors for lenses i and j from an input data file. The relative intrafield distortion at 9 field locations is then calculated (in principle, any number of field points can be used). This data remains constant throughout the rest of the simulation. A single stepper is simulated by setting all the intrafield distortion errors in the data files to zero.

The program then starts a series of loops to simulate the errors incurred by the change of the reticle and random pressure and temperature variations between reticle changes. For each reticle change, random die rotation, magnification, and reticle stacking errors are chosen by the Monte Carlo routine. All component errors in this model except relative distortion between lenses are assumed to be Gaussian distributed[5]. A random number between 1 and 1663 is chosen by a random number generator. The computer then searches a look-up table which associates the random number with a t-value such that $-3 < t < 3$. The random magnification error is then calcualted as $M = ts_M$, where s_M is the input standard deviation for magnification control. In order for the simulation to be as accurate as possible it is necessary to have good information on the distributions of the component errors. Likewise, random values are chosen for reticle rotation. Reticle stacking errors

are generated at each of the 9 field locations.

Once the magnification, die rotation, and stacking errors are generated, the intrafield errors for the lot of wafers are calculated from (l = 1 to 9)

$$DX_l(i,j) = - R\ y_l + M\ x_l + (D_{lx}(i) - D_{lx}(j)) + s_{lx} \quad (12)$$

$$DY_l(i,j) = R\ x_l + M\ y_l + (D_{ly}(i) - D_{ly}(j)) + s_{ly} \quad (13)$$

The program then enters the wafer alignment loop. It first generates random offsets for x and y translation, wafer rotation, orthogonality ($d\theta = \theta_x - \theta_y$), scaling, and differential scaling ($dE = E_x - E_y$) for the wafer lot. For each wafer the program calculates a translation error in x by generating a random variation around the offset and adds that to the offset determined at the reticle change. The same is done for y translation, x and y rotation, and x and y scaling. The errors at the centers of the fields at all wafer locations k for that wafer are then calculated from

$$WX_k = T_x - \theta_x\ y_k + E_x\ x_k + ss_x \quad (14)$$

$$WY_k = T_y + \theta_y\ x_k + E_y\ y_k + ss_y \quad (15)$$

where ss_x and ss_y are random stage errors, generated at each field separately.

The total overlay error at each field location l and wafer location k is then calculated as the sum

$$VX_{kl} = DX_l + WX_k \quad (16)$$

$$VY_{kl} = DY_l + WY_k \quad (17)$$

For each wafer the VX_{kl} and VY_{kl} form two matrices. In order for a field to be considered good, the corresponding rows in VX_{kl} and VY_{kl} must contain no errors greater than the specification. The computer sorts through the matrices to find bad fields. It also calculates the mean x and y errors and x and y standard deviations for all matrices. Alignment is repeated as often as chosen. In many of the computer runs considered here, 10 wafers are aligned before the next reticle change.

Once the lot of wafers is finished, the computer changes the reticle again and repeats the procedure described above. After the computer runs through all the reticle changes requested, it then calculates the overall statistics for the run, including the mean and standard deviations in x and y for the entire distribution and the percentage good fields.

It is an easy modification to the program to simulate the behavior of steppers which employ field by field alignment rather than global alignment. In this case scaling, orthogonality, and random stage errors are set to zero. Translational and, if appropriate, die rotational alignment errors are generated at each wafer location k.

The program described here does not calculate nonlinear errors due to trapezoid, or due to third or fifth order distortion. Nonlinear intrafield errors are lumped together as unchanging signatures of lenses, and are characterized through the DX and DY input files. This is not an altogether justifiable assumption since trapezoid errors can change and be adjusted. It has also been shown that illuminator defocus can lead to third order intrafield errors which mimic true lens distortion[10]. However, these errors are assumed to be quite small in comparison with the others considered here. It is certainly possible to continue the analysis to cover the case of time varying nonlinear intrafield errors.

As an example of the use of OVS, two different types of 5X reduction wafer stepper were simulated. The first type uses off-axis global alignment and is uncompensated for magnification changes as a function of barometric pressure. The second type represents a stepper with through the lens alignment, barometric compensation, and the ability to adjust magnification to match wafer expansion due to processing. Global alignment is also used on this type stepper. In this example, single stepper performance is simulated by assuming that all intrafield errors except reticle rotation, magnification, and reticle stacking errors are zero.

Table 2 lists the input parameters for the two simulations. Table 3 lists the percentage good fields found at various overlay tolerances and the equivalent $\bar{X}$ + 3 sigma values. Figure 7 is a plot of good fields percentage versus the overlay design rule.

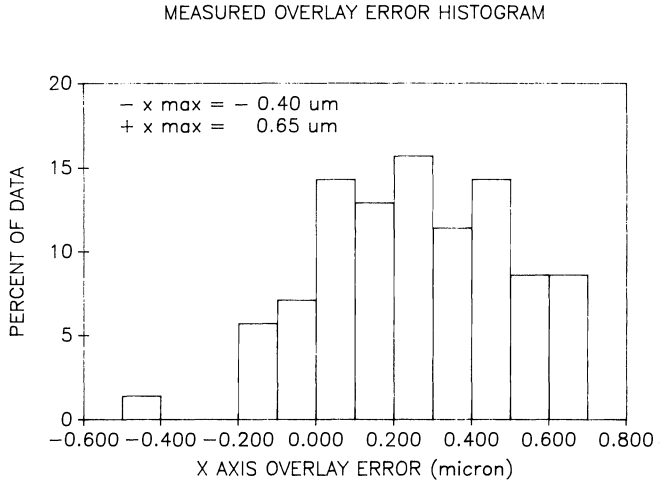


Figure 6a. Measured X overlay error histogram for the example wafer pictured in Figure 4.

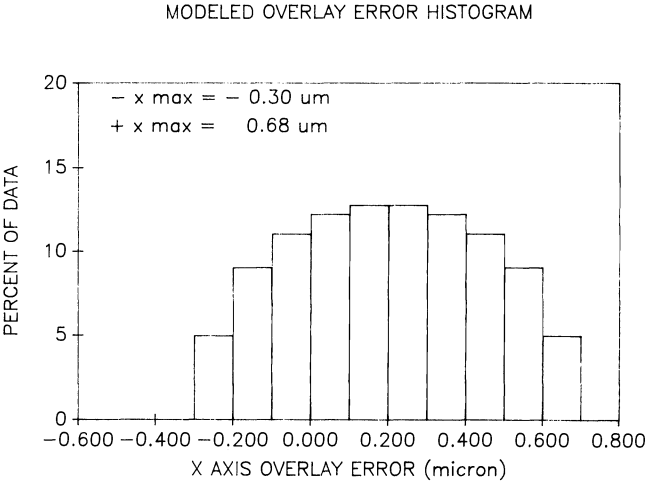


Figure 6b. Modeled X overlay error histogram for example wafer in Figure 4. Histogram calculated using equation (11).

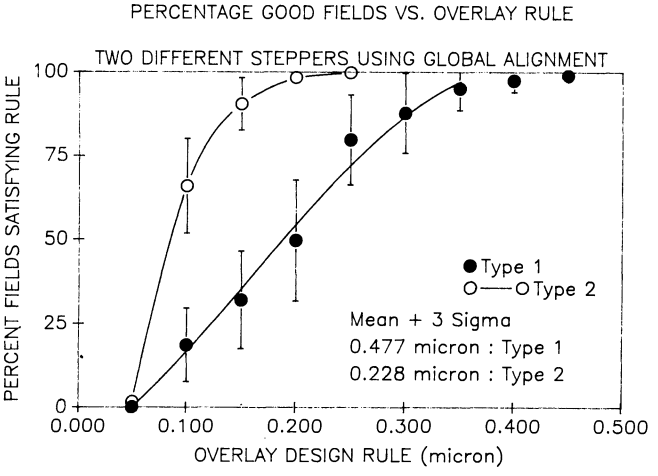


Figure 7. Percentage good fields versus the overlay design rule for two different types of stepper (see Tables 2 and 3).

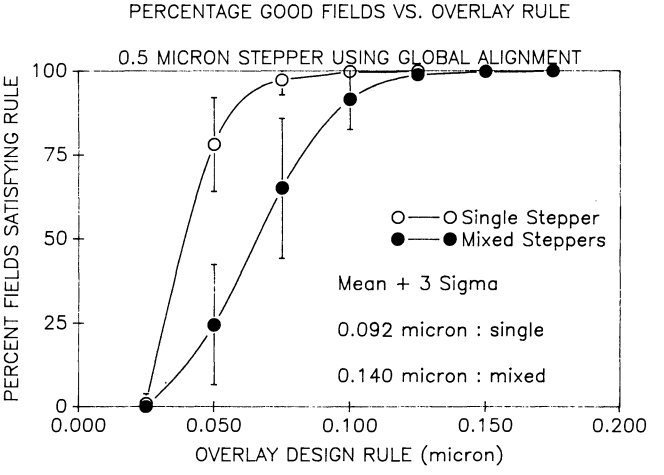


Figure 8. Percentage good fields versus the overlay design rule for a stepper using global alignment to achieve 0.1 μm overlay, single machine, 0.15 μm mixed.

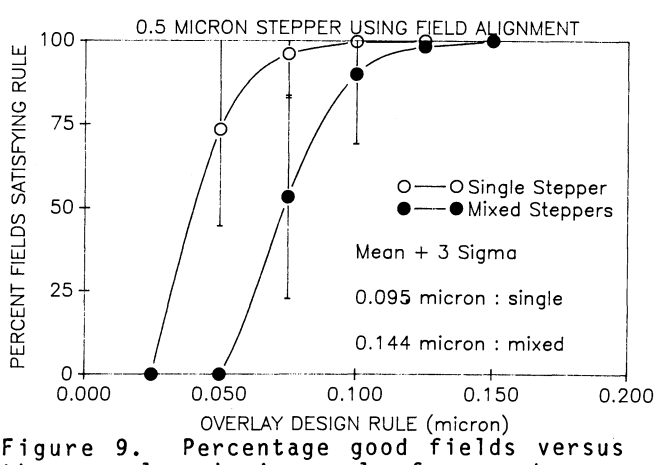
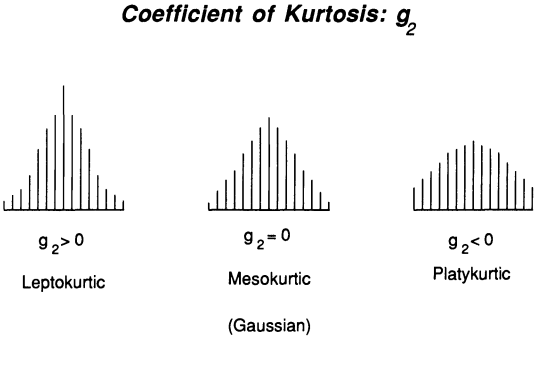


Figure 9. Percentage good fields versus the overlay design rule for a stepper using field alignment to achieve 0.1 μm overlay for single machines, 0.15 μm mixed steppers.



Reference: ASTM STP 15D, Manual on the Presentation of Data and Control Chart Analysis

Table 2 - Characteristic Component Overlay Errors

Interfield Errors	Units	Stepper 1	Stepper 2
Translation offset sigma	µm	.07	.03
Translation sigma around offset	µm	.07	.03
Symmetrical expansion offset sigma	ppm	1.0	0.5
Asymmetrical expansion offset sigma	ppm	0.5	0.5
Expansion sigma around offset	ppm	1.0	0.5
Wafer rotation offset sigma	ppm	1.0	0.5
Orthogonality offset sigma	ppm	0.5	0.5
Rotation sigma around offset	ppm	1.5	0.5
Stage precision sigma	µm	.04	.03
Intrafield Errors			
Magnification sigma around offset	ppm	8.0	3.0
Reticle rotation sigma around offset	ppm	4.0	3.0
Reticle stacking error sigma (@ 1/5X)	µm	.02	.02

Offset sigma refers to run to run variation of the mean. Sigma around the offset refers to variation within the run. In the simulation a run is defined as the number of wafers aligned before the next reticle change.

Table 3 - Overlay Simulation Results

Design Rule (micron)	% Good Fields:Type 1 Mean	% Good Fields:Type 1 Std. Dev.	% Good Fields:Type 2 Mean	% Good Fields:Type 2 Std. Dev.
.05	0.1	0.1	1.7	2.5
.10	18.1	11.1	65.9	14.2
.15	21.7	14.5	90.5	7.9
.20	49.7	18.1	98.4	2.6
.25	79.7	13.2	99.8	0.6
.30	87.7	11.9	99.9	0.1
.35	95.1	6.7	100.0	0.0
.40	97.5	3.2		
.45	98.9	2.1		
.50	99.7	0.4		
.55	99.9	0.1		
.60	100.0	0.0		
$\bar{X}$ + 3 sigma calculated for all data (micron)	0.477		0.228	

Assumes 9 points per field, 14 mm square field size, 9 points per wafer, 150 mm wafer, reticle change every 10 wafers, 100 reticle changes. This gives 81,000 data points per axis and 9000 total fields.

OVS was used to estimate the type of control necessary for individual component errors in order to reach overlay design rules commensurate with 0.5 µm minimum feature size. It is generally agreed that overlay must be in the 0.1 to 0.15 µm, 3 sigma range for 0.5 micron lithography.

First considered is a stepper using global alignment. The input error components used for the simulation are listed in Table 4. The first column in Table 5 lists the good fields percentage and the mean plus 3 sigma for the distribution of all errors for this stepper. The same number of simulation runs and locations as in the previous example were used. 99.85% of all fields meet the 0.1 µm requirement. Note the very high precision necessary for all the component errors, better than is currently available on any stepper system. In particular translation error control to 0.021 µm, 3 sigma, is about 3 to 4 times better than any reported and stage error control to 0.03 µm, 3 sigma, is about 2 to 3 times better. Mask error control of 0.045 µm, 3 sigma, is only possible for reduction reticles. Substantial improvements are necessary in order to reach 0.1 µm single stepper overlay if global alignment is used.

Table 4 - Characteristic Component Overlay Errors; 0.1 μm Stepper

Interfield Errors	Units	Global Stepper	F X F Stepper
Translation offset sigma	μm	0.005	0.015
Translation sigma around offset	μm	0.005	0.015
Symmetrical expansion offset sigma	ppm	0.25	0
Asymmetrical expansion offset sigma	ppm	0.25	0
Expansion sigma around offset	ppm	0.25	0
Wafer rotation offset sigma	ppm	0.25	0
Orthogonality offset sigma	ppm	0.25	0
Rotation sigma around offset	ppm	0.25	0
Stage precision sigma	μm	0.01	0
Intrafield Errors			
Magnification sigma around offset	ppm	1.0	2.0
Reticle rotation sigma around offset	ppm	1.0	2.0
Reticle stacking error sigma (@ 1/5X)	μm	0.015	0.015

Table 5 - Overlay Simulation Results; 0.5 Micron Stepper
Global Alignment

Overlay Design Rule (micron)	Single Stepper % Good Fields Mean	Std. Dev.	Matched Steppers 0.05 um Distortion Mean	Std. Dev.
.025	1.01	2.95	0.00	0.00
.050	78.08	14.04	24.54	17.91
.075	97.33	4.48	65.15	20.86
.100	99.85	0.45	91.56	9.05
.125	100.00	0.00	98.86	1.85
.150	100.00	0.00	99.79	0.57
.175	100.00	0.00	100.00	0.00
$\bar{X}$ + 3 sigma calculated for all data (micron)	0.092		0.140	

OVS was then run for the 0.1 micron system to simulate matching between two systems which have a relative maximum intrafield error of 0.05 micron in both the x and y axes. Figure 8 shows the good fields percentage versus the overlay design rule for a single machine and matched steppers.

Note from the data listed in Table 5 that $\bar{X}$ + 3 sigma for the matched steppers is almost exactly the relative distortion error of 0.05 μm plus the $\bar{X}$ + 3 sigma determined for the single stepper (0.092 + 0.05 = 0.142 μm, as opposed to the simulation result of 0.140 μm). If the root sum square is taken the result is 0.105 μm. This demonstrates that it is incorrect to root sum square a systematic error such as distortion with random alignment errors to arrive at a total overlay budget, because doing so underestimates the true result, as is argued in reference (1).

Next considered is a stepper working in the field by field (FXF) alignment mode. The same field and wafer sizes, and number of points sampled are used. It is assumed that the stepper has a mechanism to adjust magnification to fit symmetrical wafer expansion perfectly, and that asymmetrical expansion is not present. In this case, errors due to wafer rotation, orthogonality, and stage error can be set to zero. The input parameters are listed in Table 5 and the simulation results are given in Table 6. Again it is found that the systematic distortion error adds directly to the result for single machines to arrive at the mean plus 3 sigma total overlay. Figure 9 shows the good fields percentage versus overlay for the FXF case.

The simulation results show that it's possible to relax the alignment error control by about three fold and the reduction error control by two if FXF alignment is used rather than global to reach 0.15 μm total overlay for mixed steppers. The level of control necessary for global alignment seems quite difficult to reach while that for FXF seems possible. The difficulty in employing FXF is usually reduced productivity. In order to achieve reasonable throughput rates, alignment acquisition times have to be quite small at each field. In addition, field alignment targets are typically much smaller than global targets because of the limited real estate available in scribe lines or in the chips themselves. Thus the signal is usually not as strong as with a large global target.

Table 6 - Overlay Simulation Results; 0.5 Micron Stepper
Field by Field Alignment

Overlay Design Rule (micron)	Single Stepper % Good Fields		Matched Steppers 0.05 μm Distortion	
	Mean	Std. Dev.	Mean	Std. Dev.
.025	0.00	0.00	0.00	0.00
.050	73.40	28.90	0.00	0.00
.075	96.10	13.25	53.20	30.46
.100	99.80	1.41	90.00	20.88
.125	100.00	0.00	98.30	1.85
.150	100.00	0.00	100.00	0.00
$\bar{X}$ + 3 sigma calculated for all data (micron)	0.095		0.144	

It is likely that field by field alignment will have to be used to achieve overlay in the 0.1 μm range, but this will require more real estate allotted to targets, more targets per field in order to allow active magnification, reticle rotation, and trapezoid control, and faster, more accurate detection and mechanical adjustment schemes than currently available. In addition, overlay simulation will be necessary to understand in detail the complex interaction of mask making, stepper, and processing variables.

Non-Gaussian Overlay Distributions

It is commonly claimed by aligner manufacturers that overlay distributions are not Gaussian, but rather have less data points in the extreme tails of the distribution than a normal distribution does. The form of a distribution is characterized in statistics as kurtosis. Kurtosis relates to the tendency for a distribution to have a sharp peak in the middle and excessive data in the tails as compared with a Gaussian or conversely to be relatively flat in the middle with little or no tails[11]. The coefficient of kurtosis, g_2, characterizes whether a distribution is leptokurtic (containing more data in the tails, $g_2 > 0$), mesokurtic (Gaussian, $g_2 = 0$), or platykurtic (less data in the tails, $g_2 < 0$). The coefficient of kurtosis, g_2, for a sample of n numbers $X_1, X_2, \ldots, X_n$ is calculated

$$g_2 = \frac{\sum (X_i - \bar{X})^4}{ns^4} - 3 \qquad (18)$$

The coefficient of kurtosis is calculated for each simulated set of alignments (i.e., one reticle change) in OVS. It is found that there is a distribution of g_2 values for each set of error parameters. For example, in the case of the 0.1 μm overlay stepper using global alignment the distribution of g_2 along the x axis for 100 simulated runs is shown in histogram form in Figure 10a. The mean of the distribution is - 0.144, showing that indeed that the average run does have a distribution of overlay errors which is more tightly bunched than a Gaussian. However the distribution of g_2 is very wide, with a standard deviation of 0.294. 22 of 100 runs had positive g_2 values. Thus while the average lot of the 0.1 μm single stepper has a platykurtic distribution of overlay errors, almost a quarter of the runs can be expected to show leptokurtic behavior.

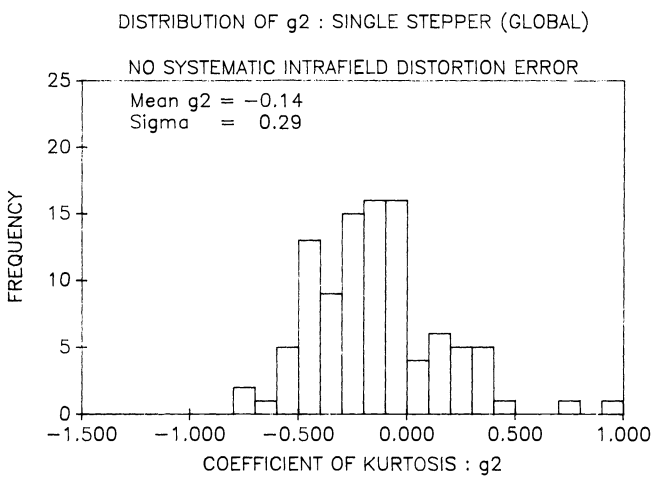


Figure 10a. g_2 histogram;global,single machine.

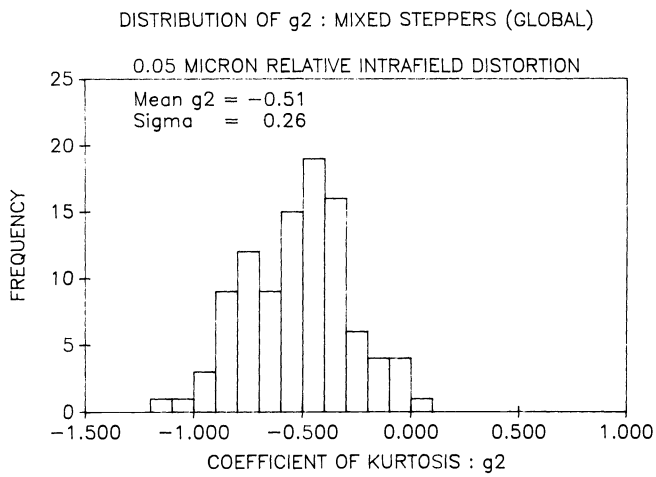


Figure 10b. g_2 histogram;global, mixed.

Table 7 - Coefficient of Kurtosis for Simulator Runs

	Single Stepper	Matched Steppers (0.05 μm distortion)
Global	-0.14 ± 0.29	-0.51 ± 0.26
FXF	-0.35 ± 0.39	-0.62 ± 0.42

For the matched stepper case the distribution of g_2 values along the x axis is shown in Figure 10b. Note that the center of the distribution is more shifted to negative g_2 values than the previous case where there was no systematic intrafield error. The mean is - 0.510 and only two of 100 runs showed positive g_2 values. The mean and one standard deviation values for g_2 for the global and FXF steppers are listed in Table 7.

These simulation results imply that overlay errors for steppers do not in general yield Gaussian distributions. However, neither do they always yield distributions which have less data in the tails than a Gaussian unless systematic errors are of about the same size as the random errors. The reason for the shift to more platykurtic distributions with increasing systematic errors is not completely clear. An intuitive argument can be given. As systematic errors grow larger with respect to the random errors, the overall distribution of errors begins to be dominated by the systematic errors. In the limit the error distribution converges with the systematic error distribution which has no randomness by definition. Thus the error distribution might be expected to be platykurtic. This is an important area for further analysis.

Acknowledgments

I would like to thank a number of people who contributed to the ideas expressed in this paper. First I would like to thank Lucien Nedzi, now with the U.C.S.F. Medical School, for his substantial contribution to the contour representation of linear overlay errors. The section on the contour representation is a condensed version of an unpublished paper[12] written with Anna Minvielle of AMD in 1984. Rory Rice of AMD has been constantly insistent that overlay error budgets take into account systematic errors and in many ways this paper is in response to his searching questions. Alan Levine of Ultratech Stepper and Colin Knight of AMD/Sematech suggested to me the idea of constructing a Monte Carlo simulator for wafer stepper overlay errors. Finally I would like to thank Bill Heavlin and Andy Brown of AMD, and Harry Levinson of Sierra Semiconductor for helpful discussions.

References

(1) R.Rice, H.Levinson,"Overlay tolerances for VLSI using wafer steppers",SPIE Vol.922,1988

(2) D.S.Perloff, "A four point electrical measurement technique for characterizing mask superposition errors on semiconductor wafers", IEEE J.Sol.St.Circ., Vol. SC-13,4,436-444,1978

(3) D.MacMillen,W.D.Ryden,"Analysis of image field placement deviations of a 5X microlithographic reduction lens", SPIE Vol.334,78-89,1982

(4) W.T.Lynch,"The reduction of LSI chip costs by optimizing the alignment yields",IEDM Technical Digest,7G-J, 1977

(5) W.H. Arnold,"Image placement differences between 10:1 reduction wafer steppers and 1:1 scanning projection aligners", SPIE Vol. 394,1983

(6) J.R. Taylor, An Introduction to Error Analysis, University Science Books

(7) W.Heavlin, C. Beck, "On yield optimizing design rules", IEEE Circuits and Devices Magazine, 7-12, March 1985

(8) T.F. Hasan, S.U. Katzman, D.S. Perloff, "Automated electrical measurements of registration errors in step-and-repeat optical lithography systems", IEEE Trans. El. Dev., Vol.ED-27, No. 12, pp2304-2312, 1980

(9) C.S. Kim, W.E. Ham, "Yield-area analysis: part 2 - effects of photomask alignment errors on zero yield loci", RCA Review, Vol. 39, 565-576, 1978

(10) D.W. Peters,"The effects of an incorrect condenser lens set-up on reduction lens printing capabilities", Procee dings of the Kodak Microelectronics Interface, 1985

(11) Manual on the Presentation of Data and Control Chart Analysis, ASTM STP 15D

(12) W.H. Arnold, A.M. Minvielle,"Distribution model and contour representation of systematic registration errors in optical lithography", 1984, unpublished paper