Alignment of Patterns in Microlithography: General Perspective

Daniel C. Cole

D. Cole is with the Dept. of Manufacturing, Boston University, 15 St. Mary's Street, Brookline, MA 02446. E-mail: dccole@bu.edu.

Abstract

The need to repeatedly align pairs of structures made from separate processes occurs in a number of areas in manufacturing. The fabrication of microchips using microlithography is a key example. This article provides a foundation for discussing this example, as well as more general ones, such as might occur in three-dimensional nanoelectronic constructions, and even more abstract examples such as signal processing. The idea is proposed here that "alignment" scenarios involve repeated matching of pairs of patterns from separate processes, where limited sampling of properties are used to judge "best alignment" for each pair. Ideas such as the effect of the number of degrees of freedom used in an alignment process, changing the number of sampled alignment sites, and altering their location or other properties, are explored here in terms of their effect on alignment statistical properties.

Keywords

Microlithography, overlay, alignment, nanolithography, microelectronics, nanoelectronics, registration marks

I. INTRODUCTION

The present article involves the general subject of the alignment of patterns of structures, particularly in regards to microlithography, but also with an outlook for the eventual needs in nanoelectronics and other related areas that entail the matching of patterns. Technologists who work in these areas are well aware of the importance of concepts such as the "overlay" of patterns and the "error budget" [1] that arises from the metrology tools, the processing steps [2], and the repeated overlaying of pattern on top of pattern, as layered structures are built up in microchips. Projections on alignment and overlay needs are included in the "International Technology Roadmap for Semiconductors," publicly available in Ref. [3]. With gate structures now below 0.1 μm, and with far smaller projections for the future, one needs the alignment of these small structures to be as small a fraction as possible of the other resolution errors inherent in processes such as imaging, etching, and implanting.

Thus, there is clearly a well recognized importance of controlling alignment errors when matching one lithographic mask pattern to another one, and when aligning each underlying chip structure to the next layer of patterns to be constructed. Nevertheless, there does not seem to be a readily available theoretical description of what one wants to do in general to minimize the total "alignment error", that is both accurate and general in

its description, yet not so couched in abstract statistical terms to make it inaccessible to practicing lithographers. The present article is intended to begin to fill that gap. Without a doubt, however, there are so many different alignment systems, from the now familiar "step and repeat" to "step and scan" systems in microlithography, to yet more complex and exotic scenarios, that it would be nearly impossible to attempt to cover all strategies without losing the main message that is intended here. Moreover, whether one is speaking of the comparison of specific full structures within specific patterns, or of only single specific points within patterns, will have a considerable difference in analyzing the impact of different alignment strategies.

Hence, a complete coverage of individual alignment systems will not be attempted here. Instead, a general framework will be presented that with some work can be extended to cover specific cases of interest.[1] The advantages of doing so are several. First, there are a number of subtle, yet important issues typically overlooked in the reported microelectronics literature on "alignment," "registration marks," and "overlay." With the underlying conceptual theory of alignment that will be presented here, it is hoped that more common grounds and subsequent deeper analyses can be reached by technologists involved in such issues in the future.

Second, the approach that will be followed here should yield, besides a general theoretical framework, also results that enable qualitative, conceptional, and "order of magnitude" answers to questions involving the design of alignment systems. More specifically, there are all sorts of issues to take into account when constructing an alignment system, ranging from how many sites to examine and measure, to where to place them, how to weigh their appropriate contributions, what degrees of freedom to employ, etc. We intuitively expect that as the number of alignment sites increase, then our knowledge of the relation of one pattern to another will in some sense increase, which we expect will translate into the ability to make the two patterns match more closely. Likewise, we expect that the more degrees of freedom we employ in our alignment procedure, the more closely we

[1] Without question, depending on whether one is aligning points, lines, or more general structures, and depending on the criteria function used and the types and degrees of freedom employed, then considerable more work might be required to analyze a specific alignment situation. Nevertheless, the material provided here, plus subsequent planned work, should at the very least provide a logical starting point for the analysis of such systems.

should be able to match patterns. For an engineer trying to decide whether to make a "two-point" alignment system, versus one with three, four, or even much higher number of alignment sites, there is much to decide. The cost of making a four-point versus a two-point alignment system can be nearly twice the cost, or twice the alignment time, depending on whether one doubles the number of sensors in the alignment apparatus, or doubles the travel time in going to more sites. Either way can be significant in terms of cost of alignment equipment, or the decrease in throughput if large quantities of wafers are being exposed. Knowing the improvement in alignment that will be gained, at least roughly, can then be extremely helpful in deciding how complex the system should be made.

A third reason for presenting the following article in a fairly general way is that it is technologically important to extend our usual concepts of aligning two physically rigid structures to ones that are "malleable," at least to some degree. Aligning two structures necessarily entails a comparison between the structures. Somehow this comparison is made, and a decision is arrived at on how to transform at least one of the structures to make it conform better with another structure. For physically "rigid" structures, undoubtedly such discussion sounds like a completely needless over-generalization of what can be done, since rigid physical structures can only be translated and rotated. However, in microelectronics, such a restriction is far from what can actually be accomplished, where "images" of structures are what are actually compared; there are controls one employs when detecting the images, in terms of focus,[2] contrast, skew, scaling, orthogonality, etc. These images may be of only the alignment sites themselves, or, the images may be important in the fabrication of the structures, as occurs when exposing photoresist. Moreover, even the structures themselves can be stretched, compressed, or twisted across their entire domain, using either mechanical, thermal, or electromechanical means. This may occur either intentionally or not, but without question, these extra "degrees of freedom" do occur. Alignment structures themselves have process variability in their construction [4],[5],[6] as well as random offsets introduced in viewing them due to variations in film coatings over them.[3] The simple vacuum pressure of a "chuck" holding a wafer [8], the

[2]See, for example, Fig. 2 in Ref. [1], which reports alignment offset versus focus offset.
[3]See, for example, Fig. 2 in Ref. [7].

temperature distribution across a reticle in a projection system, the support structure for a mask in X-ray lithography systems, etc., are sufficient to cause noticeable alignment changes of the degree that are being measured today in microelectronics. Wafer bowing is known to arise from process steps such as chemical-mechanical polishing (CMP), oxidation, and baking [9]. Changes made on the "knobs" that control these optical, mechanical, and thermal systems, without question change the "alignment" of two patterns trying to be made as close in character to each other as is possible [1].

Hence, "alignment of rigid" structures may be what we all think of when we conceive of aligning two physical materials, but such a concept is really only a very limited aspect of what happens in today's technological practices. Consequently, it makes sense to use the concept of translations and rotations to get our bearings on what we mean by aligning two patterns, but a theoretical description that naturally generalizes this concept is needed to address the largely two-dimensional pattern alignment in today's microelectronic industry. Moreover, if done correctly, then this description can readily be extended to the alignment of not just 2-D patterns, such as largely occurs in microelectronics today, but also to 3-D patterns, such as may well become important in the myriad of nanoelectronic devices that are being anticipated today [10],[11],[12]. Such will be our intent here.

Moreover, with very little extra work, the ideas surrounding the alignment of 1-D, 2-D, and 3-D patterns of physical structures, can readily be extended to higher dimensional space. Why do so? The reason is the following, which brings us to the last major goal of this article. The subject of "alignment" can be addressed as nearly a separate entity unto itself. In some sense, "alignment" lies at the very heart of metrology, since it inherently involves a comparison and adjustment (*e.g.*, shift or rotation, or possibly a "stretch") of some structure to another to make their properties match more closely. If we take the idea of "structure" in the most general sense, meaning roughly a system characterized by a set of properties, then comparing two structures means comparing their properties. "Aligning" the two structures then naturally involves utilizing whatever degrees of freedom are accessible to the technologist or scientist to bring the properties of the two systems closer in agreement, according to some criteria.

For example, suppose we have two partially created sets of microchips, where the two

sets were processed similarly, but not exactly the same. Suppose one of the sets had an implant done that was not quite what the designers had intended. Instead of throwing the batch away, perhaps the decision is made to try to bring the batches back into closer agreement by attempting an additional set of implants and anneals on the incorrectly processed set. The degrees of freedom that could be employed here are the dose and energy for each of the additional implants, and the complete temperature cycle for each of the anneals. Each knob adjusted on the equipment used to control the doses, energies, temperature, time, and variances of any of these properties, can be considered as an extra degree of freedom to help bring the two final systems into better agreement with each other in terms of their physical properties.

More simply, but in the same vein, suppose we have two vats of liquid that are slightly different shades of color, and we want to make them the same color. Adding a dye to one of them will adjust the color; the degree of freedom here would be the amount of dye. Perhaps we want one of the liquids to have twice the shade of blue as the other, according to some color scale when the liquids are at the same temperature, or, for one to be ten degrees higher than the other when it achieves the "twice the shade of blue" characteristic, etc. Each additional specification of the properties of the two systems involves the adjustment of one or more degrees of freedom (e.g., amount of dye and the temperature) to better align the properties of the two systems according to some criteria.

Moreover, as metrologists are well aware, comparing the properties of two systems involves all sorts of subtle assumptions. For example, where is the temperature of the two liquids to be taken? If one has a large vat, then having several thermometers at different points in the vat may be important to achieve the desired property. Having one thermometer may of course only yield a rough estimate of the temperature throughout the remainder of the vat. How many thermometers, how many places to sample and measure the color mixture, etc., are all points a careful metrologist would want to consider. These sampled measurements would determine how much extra dye is added, how much more a vat needs to be heated up, or how much more stirring and mixing of the liquid needs to be done. These are all aspects, in some very real sense, of "aligning" the properties of the two systems according to some sampling of the measured properties of the two systems.

Returning now to the alignment of properties of structures in N dimensional space, we see that the N dimensions simply represent the N coordinates that define some property, such as the voltage differences applied to N terminals of a microchip, or the currents flowing out of these terminals. If we compare two microchips made from two similar, but slightly different processes, and we want the same current magnitudes to flow out of the N terminals of each microchip when the same voltages are applied, then changes in terminal impedances might need to be made, or other changes might need to be considered. In a somewhat abstract yet still very real sense, these concerns are closely related to the idea of physically aligning patterns of microchip structures, as they entail the comparison of properties, degrees of freedom, sampling, and adjustment.

The main intent in the present article is to present a framework to better enable microlithographers to analyze their alignment systems of two dimensional patterns. However, in the course of this discussion, it seems interesting to note that many situations in manufacturing, experiments, engineering, and science, can be cast into a similar framework involving ideas of alignment. Key elements of an "alignment system" might be categorized as: (1) The properties of two systems or structures; (2) the comparison of a subset of these properties, (3) an adjustment of some or all of these properties, based on the previously mentioned comparison of the subset of properties; and (4) the analysis of how "best" to make this adjustment based on the limited sampling of the full properties of the two systems.

In some sense, of course, "alignment" is simply a subset of more general optimization problems and techniques, which have been and continue to be investigated extensively within the mathematics and engineering literature [13]. However, in another sense, the manufacturing of aligned structures is an important category unto itself, as it specifically implies the repeated comparison and adjustment of pairs of structures to bring them into conformability, typically under limited sampling of the full properties of each structure. Moreover, a key manufacturing concern is the long term ramifications of repeating this alignment strategy over and over again. In contrast, a pure optimization problem might involve just one system, rather than being concerned about a large ensemble of pairs of systems, each slightly different.

The outline of the remainder of this article will be the following. First, Sec. II will discuss in both physical and mathematical terms elements that enter into characterizing the alignment of members of two separate processes. Section III should be of particular interest to microlithographers, as it contains relevant predictions for sets of two dimensional patterns. An outline and explanation of the mathematical steps is provided, although complete proofs of the more detailed calculations will be published elsewhere. Section IV ends with some concluding remarks.

II. CONCEPTS ON ALIGNMENT OF PATTERNS

A. Main Example

Here we will discuss the main example of "alignment" that will be used as our central vehicle for future developments. This example is a very practical one for microlithographers, yet it is also readily generalizable to many other situations. The example involves two-dimensional planar positions of structures made to conform to other similar structures.

Suppose we have one layer of structure already created by a masking process on a family of microchips. Let us think of this first family as an ensemble of this first layer of structures. Each member of the ensemble will be slightly different from the other members, due to always present processing differences, such as photoresist flow and uniformity, focus and exposure differences, vibrations during the processing, and all of the myriad of variables that enter into whatever physically made the structure after the lithography step (*i.e.*, etching, deposition, oxidation, etc.). Let us label this first full set of processes that made the first layered structure by "b".

Now let us suppose we have a second layer of structure that is to be created on top of the first layer. Again we can think of the family of this second set of structures as arising from a complete set of processes that we will label as "a". As in the first case, each member of this second set of structures is expected to vary somewhat in properties from other members, due to uncontrollable but natural process variations.

Let us take as our set of properties of the structures created in process b as being S positions of certain features of the structures. Likewise, let us assume that process a produces structures with S positions, or sites of interest, that we want to lie on top of

the S features in process b. For example, the layer from process b may be associated with the source and drain structures of FETs, while the layer from process a may be the gate structure associated with the same set of FETs. Thus, both sets of patterns may in general be vastly different in shape, but, they need to be made to lie over each other in a very definite way, within strict tolerances, or else the FETs that are constructed will not function.

Not all of the S sites will typically be examined. In today's microlithography system, only a few points of these S critical positions are specifically forced to lie over each other as closely as possible. The technologist expects that the remaining sites will then also lie very close to each other; exactly how close, and with what probability, are the questions we want to analyze here. Hence we will consider a total of m sites ($m \leq S$) that are examined and used to judge how best to align the two patterns.

B. Process Description

Due to the inherent process variations involved with making the ensemble of structures associated with processes "a" and "b", then a natural way to describe the distribution of structural members from each process is via a probability density function. In what follows, we will assume each positional site is represented by a vector $\mathbf{u}_s$, where s is an index that labels the sites and ranges from 1 to S. For our purposes here, we will assume that $\mathbf{u}_s$ is a vector in three dimensions, so that in a Cartesian coordinate representation, $\mathbf{u}_s = \hat{\mathbf{x}}u_{sx} + \hat{\mathbf{y}}u_{sy} + \hat{\mathbf{z}}u_{sz}$, where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are the three orthogonal unit vectors defining the coordinate system. For the process "a", let the probability of producing a structural pattern in which the first site falls within the infinitesimal volume[4] d^3u_1 of position $\mathbf{u}_1$, the second site falls within the infinitesimal volume d^3u_2 of position $\mathbf{u}_2$, etc., be given by

$$P_3^a(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S)d^3u_1 d^3u_2 ... d^3u_S$$

where $P_3^a(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S)$ is the associated probability density function. We would have a similar expression for the related probability using process "b", but as described by $P_3^b(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S)$, so as to distinguish between the different characteristics of process "a"

[4]The symbol d^3u is simply notational and represents an arbitrary shape of infinitesimal volume in three dimensions. For the specific case of a small rectilinear parallel-piped of sides dx, dy, and dz, then $d^3u = dxdydz$.

from those of "b". For example, if process "a" involves one type of lithographic mask, with specific shapes, and then includes a subsequent etching step, while process "b" involves a different lithographic mask, photoresist processing, and subsequent processing steps, then we expect the probabilities of specific structures falling at various points to be different between the two processes.

In most of present microlithography, only two dimensions would be generally of interest, since structures are presently made largely in near planar layers that are positioned on top of each other. In that case, we will indicate the probability of producing a structural pattern with the first site falling within the infinitesimal area d^2u_1 of position $\mathbf{u}_1 = \hat{\mathbf{x}}u_{1x} + \hat{\mathbf{y}}u_{1y}$, the second site falling within the infinitesimal area d^2u_2 of position $\mathbf{u}_2 = \hat{\mathbf{x}}u_{2x} + \hat{\mathbf{y}}u_{2y}$, etc., as

$$P_2^a(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S)d^2u_1 d^2u_2 ... d^2u_S \quad ,$$

where now we use $\mathbf{u}_s = \hat{\mathbf{x}}u_{sx} + \hat{\mathbf{y}}u_{sy}$.

The coordinates of each pattern produced from such a process may be viewed as random variables. As is often done, the capital letters $\mathbf{U}_1^a, \mathbf{U}_2^a, ..., \mathbf{U}_S^a$ will be used to designate these quantities, while the lower case letters $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S$ will be used for the particular values that $\mathbf{U}_1^a, \mathbf{U}_2^a, ..., \mathbf{U}_S^a$ may acquire.[5] The expectation values of the random variables $\mathbf{U}_1^a, \mathbf{U}_2^a, ..., \mathbf{U}_S^a$ are equal to the first moments associated with the probability density function P^a and will be designated by

$$\boldsymbol{\mu}_s^a \equiv \int ... \int \mathbf{u}_S P_3^a(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_S)d^3u_1 d^3u_2 ... d^3u_S = \langle \mathbf{U}_S^a \rangle \quad , \tag{1}$$

for $s = 1, 2, ..., S$. The notation $\langle\ \rangle$ is used here to indicate the operation of finding the expectation value of the random variable enclosed within the brackets. A similar definition holds for process "b".

Second-order central moments associated with the random variables U_{sx}^a, U_{sy}^a, and U_{sz}^a, where $\mathbf{U}_s^a = U_{sx}^a\hat{\mathbf{x}} + U_{sy}^a\hat{\mathbf{y}} + U_{sz}^a\hat{\mathbf{z}}$ at site s, can be defined via

$$\lambda_{ri,sj}^a \equiv \left\langle \left(\delta U_{ri}^a\right)\left(\delta U_{sj}^a\right)\right\rangle \quad . \tag{2}$$

Here i and j index the three spatial dimensions (or two, if only planar situations are being considered), while r and s denote two different sites.

[5]See, for example, p. 69 in Ref. [14]. Similar notation exists in other standard textbooks.

Higher order central moments could likewise be defined, such as by the notation $\lambda^a_{ri,sj,tk} \equiv \left\langle \left(\delta U^a_{ri}\right) \left(\delta U^a_{sj}\right) \left(\delta U^a_{tk}\right) \right\rangle$. However, for our present purposes, we will not have need of such generalizations, as our calculations in the present work will be confined to the first- and second-order central moments.

Generalizing (1) and (2) to another set of vectorial dimensions, such as the important case of two-dimensions in microlithography, is easily done by simply changing the range over which i and j span, changing the volume elements d^3u_s to d^Nu_s, where N is the number of dimensions, etc. For most of the rest of this article, we will assume $N = 2$, but it is clear that this is readily generalizable.

C. Transformation

Typically in an alignment situation, where one has a list of properties of one structure and a list of properties of a second structure, then one wants to transform at least one of these two structures in a way to try to make the two conform more closely to each other. In microlithography, typically the transformation is done on only one of the patterns, such as the new mask being aligned on top of the structures already made. If we think of the "b" process as giving rise to the site positions $\mathbf{U}^b_s$ already on the underlying structure in a microchip, and the "a" process as giving rise to the new structure to be created on top, then typically the "a" process would involve some degrees of freedom that enable it to be made more in character with the underlying structure. The obvious degrees of freedom are translation and rotation, but, as already discussed, often there are more degrees of freedom available, particularly in optical projection systems.

We can denote the transformation of the site positions $\mathbf{U}^a_s$ into $\mathbf{U}^{a\prime}_s$ by:

$$\mathbf{U}^{a\prime}_s \equiv \hat{\mathbf{x}} f_x \left(\mathbf{U}^a_s, \Delta_1, \Delta_2, ..., \Delta_F\right) + \hat{\mathbf{y}} f_y \left(\mathbf{U}^a_s, \Delta_1, \Delta_2, ..., \Delta_F\right) \quad , \tag{3}$$

where $\Delta_1, \Delta_2, ..., \Delta_F$ are a total of F degrees of freedom available to align the two structures. In the case of translation in one direction, $F = 1$, in two directions, $F = 2$, and if rotation in the plane is included, then $F = 3$, etc. If tilt, focus, skew, curvature, wafer bowing, etc., are taken into account, as well as mechanical stress, and temperature controls and gradients, then the number of adjustments to bring the two patterns into conformity greatly increases. Of course, some adjustments are not done every time two

patterns are compared, but only after large numbers have been compared. In such a case, the treatment of those parameters must be handled differently [9].

For our main example of alignment of a mask to an underlying structure, then typically we would be interested in the differences

$$\mathbf{D_s} = \hat{\mathbf{x}}\mathbf{D_{sx}} + \hat{\mathbf{y}}\mathbf{D_{sy}} \equiv \hat{\mathbf{x}}\left(\mathbf{U_{sx}^{a\prime}} - \mathbf{U_{sx}^{b}}\right) + \hat{\mathbf{y}}\left(\mathbf{U_{sy}^{a\prime}} - \mathbf{U_{sy}^{b}}\right) \quad , \tag{4}$$

for $s = 1, 2, ...N$, and we would want to adjust Δ_1,Δ_2,...,Δ_F to reduce some sampled subset of the $\mathbf{D_s}$ values. Some "measure" of the set of $\mathbf{D_s}$ values needs to be examined, and then minimized.

D. Alignment Criteria Function

A common criteria function chosen is the sum of the squares of the distances between corresponding sampled pattern points, given by

$$\Omega = \sum_{s=1}^{m} (\mathbf{D}_s)^2 \quad , \tag{5}$$

where $m \leq S$, and the sites are assumed to be relabelled such that the first set of indices for $s = 1, .., m$, are taken to be the sites used in the above criteria function. In the language of statistics, this procedure of selecting $m \leq S$ sites, would be described as sampling "m" sites of the total population of "S" sites in order to obtain an estimate of the best alignment parameters.

Alternatively, a more general expression could be assumed, even within the domain of the method of least squares. For example, one could choose $\Omega = \sum_{s=1}^{m} w_s (\mathbf{D}_s)^2$, where $\sum_{s=1}^{m} w_s = 1$, and $0 \leq w_s \leq 1$ for all $s \in \{1, ..., m\}$. Moreover, there are numerous other functions that would serve equally as well as a measure of how close the two patterns are in alignment, such as the sum of the absolute values of the distances between corresponding points, or the number of corresponding sites with distances between them below a particular value. Unfortunately, except for the simplest of cases, it is in general quite difficult to obtain explicit expressions for the parameters that optimize these other functions. Since one of the major aims of this report is to calculate the propagation of the statistical properties of the two processes that produce the arrays to the final statistical properties of the aligned arrays, it is important that explicit expressions for the free

parameters in the optimization procedure be obtained. Consequently, all calculations for aligning the two arrays will be done assuming that (5) is the appropriate expression. The least that this will accomplish is to illustrate by a particular example how to calculate the statistical properties of aligned patterns once a criteria function is chosen. Fortunately, the criteria function chosen yields results of interest in their own right so that the chosen function should serve as more than just a mere example of more general treatments.

III. SUMMARY OF KEY PREDICTED ALIGNMENT RESULTS

A. Alignment Change

As implied earlier, there are a number of subtle assumptions involved in aligning patterns, such as whether one makes changes to one pattern, or to both; or whether one makes comparisons of two patterns directly, or measures patterns separately according to some reference grid and then deduces alignment changes. The consequences of these different alignment procedures are often nonintuitive; the resulting change in propagation of statistical properties can be important. In work to be presented elsewhere, such details will be discussed, and the mathematical development will be provided for handling one- and two-dimensional arrays of points. Here, however, simply some of the more important results that applied to microlithography will be reviewed.

First, if we assume that alignment changes are small, then in most cases one can safely make a Taylor's expansion to first order in the Δ_i parameters in (3):

$$\mathbf{U}_s^{a\prime} \approx \mathbf{U}_s^a + \sum_{f=1}^{F} \mathbf{a}_f(\mathbf{U}_s^a)\Delta_f \quad , \tag{6}$$

where $\mathbf{a}_f\left(\mathbf{U}_s^a\right) \equiv \left.\frac{\partial \mathbf{U}_s^{a\prime}}{\partial \Delta_i}\right|_{\mathbf{\Delta}=0}$. One can then solve for the optimum values of $\mathbf{\Delta}_i$ to minimize (5). Moreover, if all coordinates $\mathbf{U}_s^a$ and $\mathbf{U}_s^b$ are re-expressed in terms of $\delta\mathbf{U}_s^a \equiv \mathbf{U}_s^a - \boldsymbol{\mu}_s^a$ and $\delta\mathbf{U}_s^b \equiv \mathbf{U}_s^b - \boldsymbol{\mu}_s^b$, and if we define $\delta\boldsymbol{\mu}_s = \boldsymbol{\mu}_s^a - \boldsymbol{\mu}_s^b$, then with a fair bit of calculations, one can prove that

$$\mathbf{D}_s \approx \left(\delta\mathbf{U}_s^a - \delta\mathbf{U}_s^b + \delta\boldsymbol{\mu}_s\right) + \sum_{f=1}^{F} \mathbf{a}_f\left(\boldsymbol{\mu}_s^b\right)\Psi_f \quad , \tag{7}$$

where (7) was expressed to first order in the quantities $\delta\mathbf{U}_s^a$, $\delta\mathbf{U}_s^b$ and $\delta\boldsymbol{\mu}_s$. In the last term, the notation $a_f\left(\boldsymbol{\mu}_s^b\right)$ means that the function $\mathbf{a}_f\left(\mathbf{U}_s^a\right)$ is evaluated by setting $\mathbf{U}_s^a$

equal to $\boldsymbol{\mu}_s^b$, and Ψ_f is given by

$$\Psi_f \equiv \sum_{t=1}^{m} \left\{ \left[\frac{\partial \Delta_f}{\partial U_{tx}^b}\right]_{\mu^b} \left(-\delta U_{tx}^a - \delta\mu_{tx} + \delta U_{tx}^b\right) + \left[\frac{\partial \Delta_f}{\partial U_{ty}^b}\right]_{\mu^b} \left(-\delta U_{ty}^a - \delta\mu_{ty} + \delta U_{ty}^b\right) \right\} \quad . \quad (8)$$

In (8), the notation $[\ ]_{\mu^b}$ is meant to indicate that the quantity inside the brackets is evaluated at $\mathbf{U}_t^a = \mathbf{U}_t^b = \boldsymbol{\mu}_t^b$ for all $t \in \{1, ..., m\}$. Although we will not go over the following here, of particular interest to microlithography are the functional forms of $\mathbf{a}_f\left(\boldsymbol{\mu}_s^b\right)$ and Ψ_f for special cases of translation, rotation, skew, scaling, etc.

It should be noted that the first term in (7) of $\left(\delta\mathbf{U}_s^a - \delta\mathbf{U}_s^b + \delta\boldsymbol{\mu}_s\right) = \left(\mathbf{U}_s^a - \mathbf{U}_s^b\right)$, is the value that $\mathbf{D}_s$ possesses before transforming the $\mathbf{U}_s^a$ coordinates to align the two patterns. Hence, the second term in the expression for $\mathbf{D}_s$ in (7) is the change in $\mathbf{D}_s$ due to this transformation. This term is only dependent upon the "m" coordinate values of $\mathbf{U}_t^a$ and $\mathbf{U}_t^b$ that the function Ω in (5) is based upon, as well as the mean coordinates $\boldsymbol{\mu}_s^b$ for site "s", as occurs in $\mathbf{a}_f\left(\boldsymbol{\mu}_s^b\right)$.

B. Ensemble Averages

Upon matching pairs of patterns together, over and over again, as is done in a manufacturing environment, then clearly we want to ascertain how well our alignment scheme performs. Determining the expectation value and variance of $\mathbf{D}_s$ can help provide that insight.

From (7),

$$\langle \mathbf{D}_s \rangle \approx \langle \delta\boldsymbol{\mu}_s \rangle - \sum_{f=1}^{F} \mathbf{a}_f\left(\boldsymbol{\mu}_s^b\right) \sum_{t=1}^{m} \left\{ \left[\frac{\partial \Delta_f}{\partial U_{tx}^b}\right]_{\mu^b} \delta\mu_{tx} + \left[\frac{\partial \Delta_f}{\partial U_{ty}^b}\right]_{\mu^b} \delta\mu_{ty} \right\} \quad . \quad (9)$$

Thus, for small variations in coordinates the expectation value of $\mathbf{D}_s$ is linearly dependent upon the differences in mean coordinates, $\delta\mu_{xt} = \mu_{xt}^a - \mu_{xt}^b$ and $\delta\mu_{yt} = \mu_{yt}^a - \mu_{yt}^b$, of the two processes. The first term in (9) again represents the mean of these expressions if alignment between the two arrays-of-points was not performed, while the second term represents the change in these mean values due to the act of aligning the arrays-of-points.

Higher central moments of $\mathbf{D}_s - \langle \mathbf{D}_s \rangle$ can likewise be obtained, such as the variances of $\left\langle \left(D_{sx} - \langle D_{sx} \rangle\right)^2 \right\rangle$ and $\left\langle \left(D_{sy} - \langle D_{sy} \rangle\right)^2 \right\rangle$, and the covariance of $\left\langle \left(D_{sy} - \langle D_{sy} \rangle\right)\left(D_{sy} - \langle D_{sy} \rangle\right) \right\rangle$. Such details will be presented elsewhere, but, of perhaps more intuitive insight, here let us

examine the situation where variations in different site positions from their mean values are uncorrelated and have a constant value for their variance. Specifically, let us here consider the simpler situation where (2) reduces to $\lambda^a_{ri,sj} = \delta_{rs}\delta_{ij}\left(\sigma^a\right)^2$ and $\lambda^b_{ri,sj} = \delta_{rs}\delta_{ij}\left(\sigma^b\right)^2$, where $\delta_{rs} = 0$ if $r \neq s$ and $\delta_{rs} = 1$ if $r = s$. Thus the case being considered represents the situation where the coordinates of two different sites are uncorrelated, the x and y coordinates for a single site are also uncorrelated, and the variance of the x and y coordinates of all sites for a particular process are equal. Although clearly this simplification does not hold in many important situations, it does in some, and it provides a very nice analytic result that can be used to gain insight into alignment ideas.

With this simplification, one obtains after a fair amount of calculations the following expression

$$\left\langle \left(\delta D_{si}\right)^2 \right\rangle = \left[\left(\sigma^a\right)^2 + \left(\sigma^b\right)^2\right]\left\{1 + A_i\left(s\right)\left[1 - 2\theta\left(s\right)\right]\right\} \quad , \tag{10}$$

where i is an index for either the x or y Cartesian coordinate. In (10), $\theta\left(s\right) = 1$ if $s \in \{1, 2, ..., m\}$, and $\theta\left(s\right) = 0$ when $s \in \{m+1, ..., S\}$. Also,

$$A_i\left(s\right) = \sum_{f=1}^{F} a_{fi}\left(\boldsymbol{\mu}^b_s\right)\left[\frac{\partial \Delta_f}{\partial U^b_{si}}\right]_{\mu^b} \quad . \tag{11}$$

Of particular interest in (10) is that $\left[\left(\sigma^a\right)^2 + \left(\sigma^b\right)^2\right]$ represents the sum of the two variances due to processes "a" and "b" at site "s"; this quantity suggests the "neutral" situation of best alignment, as if one just takes the variances from each process and folds them together. However, for those sites used in aligning the patterns, the factor multiplying $\left[\left(\sigma^a\right)^2 + \left(\sigma^b\right)^2\right]$ in (10) is less than one and equals $\left[1 - A_i\left(s\right)\right]$; likewise, this factor is greater than one and equals $\left[1 + A_i\left(s\right)\right]$ for those sites not used in the alignment determination procedure. Thus, we of course expect better alignment results at the sites used in the sampling of the system information.

Another result that offers some further insight into how the alignment scheme can help improve the average variances of site matching is the following set of results that one can

prove:

$$\begin{aligned} & \frac{1}{S}\sum_{s=1}^{S}\left\langle (\delta D_{sx})^2 + (\delta D_{sy})^2 \right\rangle \\ = \;& 2\left[(\sigma^a)^2 + \left(\sigma^b\right)^2\right]\left[1 + \left(\frac{1}{2m} - \frac{1}{2S}\right)\text{Trace}\left([M]^{-1}[N]\right)|_{\mu^b} - \frac{F}{2S}\right] \quad , \end{aligned} \tag{12}$$

where

$$M_{ij} = \frac{1}{m}\sum_{s=1}^{m}\left[a_{ix}\left(\mathbf{U}_s^b\right)a_{jx}\left(\mathbf{U}_s^b\right) + a_{iy}\left(\mathbf{U}_s^b\right)a_{jy}\left(\mathbf{U}_s^b\right)\right] \quad , \tag{13}$$

and

$$N_{ij} = \frac{1}{(S-m)}\sum_{s=m+1}^{S}\left[a_{ix}\left(\mathbf{U}_s^b\right)a_{jx}\left(\mathbf{U}_s^b\right) + a_{iy}\left(\mathbf{U}_s^b\right)a_{jy}\left(\mathbf{U}_s^b\right)\right] \quad , \tag{14}$$

for $i, j \in \{1, 2, ..., F\}$. The first term of $2\left[(\sigma^a)^2 + \left(\sigma^b\right)^2\right]$ in (12) is again the term that arises if one simply adds the x and y variances together from the two processes. The second term is a positive term that is contributed by the $(S-m)$ "unaligned" sites, while the last term is clearly negative and arises from the m "aligned" sites. As the number of degrees of freedom, F, is increased in terms of making each alignment, then (12) decreases accordingly. Likewise, one can show that the larger the value of m, the number of sites used in forming the alignment criteria, the smaller will be (12).

Equation (12) can be very useful for judging specific alignment situations. The proofs of these results will be shown elsewhere. However, for the present article, where we are trying to gain a more intuitive feel for the significance of these results, then the following should be helpful. If the three conventional degrees of freedom are used in "step and repeat" systems, namely, of translation in the x and y direction, and rotation ϕ about the $\hat{\mathbf{z}}$ axis, then (12) can be shown to reduce approximately to the following[6]

$$\begin{aligned} & \frac{1}{S}\sum_{s=1}^{S}\left\{\left\langle (\delta D_{xs})^2\right\rangle + \left\langle (\delta D_{ys})^2\right\rangle\right\} \\ \approx \;& 2\left[(\sigma^a)^2 + \left(\sigma^b\right)^2\right]\left[1 + \left(\frac{1}{2m} - \frac{1}{2S}\right)\left(2 + \left[\frac{\overline{R^2}^{\text{NA}}}{\overline{R^2}^{\text{A}}}\right]_{\mu^b}\right) - \frac{3}{2S}\right] \quad , \end{aligned} \tag{15}$$

[6]The key approximation from (12) to (15), other than simply restricting (12) to translation and rotation, is the assumption that the square of the means of the coordinates is small compared to the mean of the squares of the coordinates. Otherwise, (15) is a bit more complicated.

where

$$\left[\overline{R^2}^{\mathrm{A}}\right]_{\mu^b} = \frac{1}{m}\sum_{s=1}^{m}\left|\boldsymbol{\mu}_s^b\right|^2 \quad , \tag{16}$$

$$\left[\overline{R^2}^{\mathrm{NA}}\right]_{\mu^b} = \frac{1}{(S-m)}\sum_{s=m+1}^{S}\left|\boldsymbol{\mu}_s^b\right|^2 \quad . \tag{17}$$

The quantities $\overline{R^2}^{\mathrm{NA}}$ and $\overline{R^2}^{\mathrm{A}}$ refer to the average square of the radius of the "not-aligned" and the "aligned" site locations, respectively (*i.e.*, whether the sites are included in the alignment criteria function Ω). The factor of 3 in the last term in (15) is due to the three degrees of freedom in performing the alignment.

Using (15), one can predict the impact on the average site variance in matching the "a" and "b" patterns by doing such things as changing the number "m" of alignment sites, the average radii of the aligned and unaligned sites, and the total of sites S. Hence, two suggestions can immediately be made from (15). First, the larger the value of "m", for $1 \leq m \leq S$, then the smaller becomes the quantity in (15). Likewise, the larger $\overline{R^2}^{\mathrm{A}}$ is compared to $\overline{R^2}^{\mathrm{NA}}$, then the smaller becomes (15), showing quantitatively why one should place "alignment sites" on the outside of large patterns rather than in toward the middle. The effect of changing such alignment placements can be very sizeable, since the effect propagates as the square of the distance.

Of course, the additional time incurred in using additional alignment sites on a production line and the additional cost in improving an alignment system must be weighed against the benefits of smaller overlay variations. Equations (12) and (15) can provide aids in making such decisions

IV. CONCLUDING REMARKS

The present article attempted to cast the idea of the repeated "alignment" of pairs of systems, each made from separate processes, as a special optimization problem that is important in many manufacturing situations. The key example of microelectronics was used here to generate relationships that should prove useful to technologists in microlithography, by showing how changing (1) the number of degrees of freedom used in the alignment process, (2) the number of sites used in deducing the best alignment parameters, and

(3) the positions of alignment sites, can significantly effect the expectation value of the average overlay and the variance of this quantity.

The concept of "alignment," in a sense, applies to most control systems. However, there is a large disconnect between the sophisticated mathematics that have been developed in statistical analysis and the application of this body of mathematics to engineering systems involving the manufacturing of repeated matching of patterns. This article attempted to narrow this gap by emphasizing a particular, but important, optimization problem that occurs in microlithography. This issue will remain a key concern of nanoelectronics, with the expectation that 3-D aspects of devices of the future may well become equally as important as present 2-D concerns. The ideas presented in the present article can be extended to such situations. Moreover, these ideas can be applied to more than just positions of physical structures, as they also hold for the repeated comparison and adjustment of pairs of patterns of other characteristics, such as the properties of voltages, currents, pressures, signals, etc. "Alignment" here would consist of the transformation of at least one of the pairs of patterns, based on a limited sampling of the full characteristics of the net systems. How this process averages out, when performed over and over again, is a key aspect of what this article discussed.

References

[1] K. Simon, O. Vladimirsky, Y. Vladimirsky, and F. Cerrina, "Overlay budget analysis for the 100 nm device generation," *Microelectronics Engineering*, vol. 46, no. 1–4, pp. 457–460, 1999.

[2] P. C. Hinnen, H. J. Megens, M. van der Schaar, R. J. van Haren, E. C. Mos, and S. L. F. Bornebroek, "Advances in process overlay," *Metrology, Inspection, and Process Control for Microlithography XV, Proc. SPIE*, vol. 4344, pp. 114–126, 2001.

[3] ITRS, "International technology roadmap for semiconductors (itrs)," the latest version of the roadmap can be found at http://public.itrs.net . At the time the present article was written, the 1999 release was available, and a nearly completed edited version for 2001.

[4] C. M. Yuan, "Efficient light scattering modeling for alignment, metrology and resist exposure in photolithography," *IEEE Trans. Electron Devices*, vol. ED-39, pp. 1588–1598, 1992.

[5] G. L. Wojcik, J. Mould, F. Leon, Q. D. Qian, and M. A. Lutz, "Laser alignment modeling using rigorous numerical simulations," *Proc. SPIE*, vol. 1463, pp. 292–303, 1991.

[6] Y. Deng, T. V. Pistor, and A. R. Neureuther, "Rigorous electromagnetic simulation applied to alignment systems," *Optical Microlithography XIV, Proc. SPIE*, vol. 4346, pp. 1533–1540, 2001.

[7] A. A. Ghazanfarian, X. Chen, T. Kailath, M. A. McCord, and R. F. W. Pease, "Exploiting structure in positioning of non-symmetric signals," *Proc. IEEE*, vol. 4, pp. 1913–1916, 1998.

[8] H. Stauch, K. Simon, H. U. Scheunemann, and H. L. Huber, "Impact of chuck flatness on wafer distortion and stepper overlay - comparison of experimental and fem-results," *Microelectronics Engineering*, vol. 23, no. 1-4, pp. 197–201, 1994.

[9] A. A. Ghazanfarian, X. Chen, M. A. McCord, W. P. R. Fabian, K. Nguyen, and H. Levinson, "Exploiting structure of wafer distortion in global alignment," *J. Vacuum Science Tech.B*, vol. 16, no. 6, pp. 3642–3646, 1998.

[10] D. C. Cole, S.-Y. Baek, and X. Zhang, "Evolution and integration of optimal ic design: Performance and manufacturing issues," *Design and Process Integration for Microelectronic Manufacturing, Proc. SPIE*, vol. 4692B, 2002.

[11] M. Roco, S. Williams, and P. Alivisatos, "Nanotechnology research directions: Iwgn workshop report – vision for nanotechnology research and development in the next decade," *Workshop report for National Science and Technology Council, by the Interagency Working Group on Nanoscience, Engineering and Technology (IWGN), http://www.wtec.org/loyola/nano/IWGN.Research.Directions*, 1999.

[12] O. Tabata, H. You, N. Matsuzuka, T. Yamaji, S. Uemura, and I. Dama, "Moving mask deep x-ray lithography system with multi stage for 3-d microfabrication," *Microsystem Technologies*, vol. 8, no. 2–3, pp. 93–98, 2002.

[13] O. Nelles, *Nonlinear System Identification*. Berlin, Germany: Springer-Verlag, 2001.

[14] H. J. Larson, *Introduction to Probability Theory and Statistical Inference*. New York: John Wiley and Sons, 1969.