EXPLOITING STRUCTURE IN POSITIONING OF NON-SYMMETRIC SIGNALS

A. A. Ghazanfarian, X. Chen, T. Kailath, M. A. McCord, R. F. W. Pease

Solid States Laboratory, Information Systems Laboratory
Stanford University
Stanford, California 94305, USA

ABSTRACT

One of the most crucial emerging challenges in Lithography is achieving rapid and accurate alignment under a wide variety of conditions brought about by different processing steps. Current alignment algorithms assume symmetric alignment signals. In this paper, we propose a new algorithm based on subspace decomposition of alignment signals. We assume that the process-induced asymmetries are small enough so that only linear effects need to be considered. We first find the subspace of alignment signals using a set of signals with pre-known positions. The position of a new signal is calculated considering that, if shifted correctly, it will lie in the same subspace of previous signals. Since this method exploits the structure of the signals, it results in more accurate measurement of the position. Simulation results show that the alignment error is about an order of magnitude smaller than that achieved with conventional Maximum Likelihood or phase-fitting approaches.

1. INTRODUCTION

Although the trend towards smaller features has been made possible by the use of higher resolution exposure tools, alignment between different process layers limits the practicality of achieving such small features. The stepper machines have to expose several layers within the required accuracy.

One common alignment method is the scanning/imaging scheme. The marks on the wafer are scanned and the scattered light is detected. The position of the mark is determined from its corresponding alignment signal. In the ideal case, the mark is symmetric, and so is the optical system used for scanning and detecting it. Therefore, the alignment signal will be symmetrical as well, and its center position can be determined easily[1]. In reality, the alignment signals are never symmetrical[2]. The factors causing the asymmetry are:

- Aberration and asymmetry in the optical system, which result in an asymmetric alignment signal even for a symmetric mark. This factor usually causes a constant displacement in the perceived position, and therefore can be corrected through calibration.
- Asymmetry in the mark topography due to asymmetric processing steps, such as resist coating or Chemical Mechanical Polishing(CMP). In particular, CMP is problematic because it also reduces the contrast, resulting in lower signal to noise ratio. A cross-section of an alignment mark asymmetrically covered by another layer, is shown in Fig. 1. Fig. 2 shows a typical alignment signal.

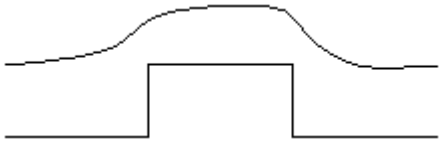

Figure 1. Alignment mark covered with an asymmetric layer. Asymmetric processing steps, such as CMP, cause asymmetric overlying layers.

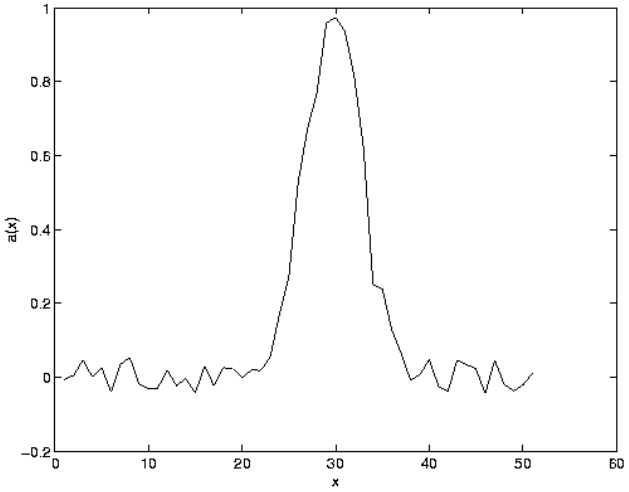

Figure 2. Detected signal, $a(x)$, from scanning the alignment mark. Since the mark is covered with other layers asymmetrically, the resulting signal is also asymmetric.

In this paper, we propose a new algorithm for determining the center position of alignment signals. We assume that the process-induced asymmetries are small so that we need to consider only linear effects. In section 2 we discuss our linear model for the asymmetry effect. In section 3 we discuss the algorithm for constructing the model. In section 4 we then describe our method for measuring the position of alignment signals based on the linear model. In section 5 we present some simulation results

and compare our performance with that of current algorithms. Finally we summarize our major findings and outline some future work.

2. MODELING

The optical system in the scanning and/or imaging based alignment schemes behaves as a low-pass filter with bandwidth proportional to NA/λ [3], where λ is the wavelength of the light, and *NA* is the numerical aperture. Therefore, there is an appropriate sampling rate above which the sampled discrete signal will have no aliasing. In practice, the signals are always oversampled.

The alignment signal is a function of the topography of the mark and all the overlying layers, and therefore is a function of the physical parameters, such as temperature and pressure, of the processing steps. Since the processes are well-controlled, we can ignore all the high order effects of the physical parameter variations. Hence we will have

$$a(x,\mu_1,\ldots,\mu_p) = a(x,\mu_1^0,\ldots,\mu_p^0) + \sum_{i=1}^{p} \Delta\mu_i \left.\frac{\partial a}{\partial \mu_i}\right|_{(\mu_1^0,\ldots,\mu_p^0)}(x) + n(x)$$
$$\equiv \sum_{i=0}^{p} \alpha_i a_i(x) + n(x)$$

where x is the variable along the scanning direction, $n(x)$ is the additive noise, which also includes all the higher order terms, and $\{\mu_i\}$ are the physical parameters, which only have small variations around their set-points, $\{\mu_i^0\}$. Also, we denote $a_0(x) \equiv a(x,\mu_1^0,\ldots,\mu_p^0)$ and $a_i(x) \equiv \left.\frac{\partial a}{\partial \mu_i}\right|_{(\mu_1^0,\ldots,\mu_p^0)}(x)$. By introducing α_0 as a variable, we can also take into account the varying gain of the optical system in comparing different alignment signals.

The $a_i(x)$, i=0,…,p, only depend on the set-points of the physical parameters, i.e. $\mu_1^0,\ldots,\mu_p^0$, and the mark shape, which are fixed in a well-controlled process. Therefore, any alignment signal can be written as

$$a(x;\mu_1,\ldots,\mu_p) = \sum_{i=0}^{p} \alpha_i a_i(x+\delta) + n(x) \qquad (1)$$

where δ is the relative position of the signal. We have also assumed that the additive noise, $n(x)$, is white. The problem can now be stated as:

For a given signal a(x), find minimum p, $a_i(x)$, $i=0,\ldots,p$, α_i, $i=0,\ldots,p$, *and* δ, *to fit the model in (1).*

We divide the problem into two parts. First, we find p and $a_i(x)$, $i=0,\ldots,p$ since they depend on the process and not on a specific alignment signal. Next, for a given alignment signal, we find α_i, $i=0,\ldots,p$ and δ, the latter is the position of the signal.

3. PART I: EXPLOITING THE STRUCTURE IN THE MODEL

Suppose that we have m alignment signals along with their corresponding center positions. The latter data usually come from metrology measurements. Without loss of generality, we can also assume that all the signals are centered at the origin. This could be done by appropriately shifting the data. Therefore, we might represent these m signals as

$$a^{(j)}(x) = \sum_{i=0}^{p} \alpha_i^{(j)} a_i(x) + n^{(j)}(x) \quad , j=1,\ldots,m \qquad (2)$$

Equation (2) implies that $a^{(j)}(x), j=1,\ldots,m$ are m noisy signals of the subspace spanned by $a_i(x), i=0,\ldots,p$

To be more specific, we shall assume that the measurement signals are also sampled, at above the Nyquist rate. Hence, $n^{(j)}(x)$, $a^{(j)}(x)$, $j=1,\ldots,m$, $a_i(x)$, i=0,…,p will be replaced by the vectors $\mathbf{n}^{(j)}$, $\mathbf{a}^{(j)}$, and $\mathbf{a}_i$ respectively and (2) can be written as

$$\mathbf{a}^{(j)} = \sum_{i=0}^{p} \alpha_i^{(j)} \mathbf{a}_i + \mathbf{n}^{(j)}, j=1,\ldots,m \qquad (3)$$

In the following two sections, we first review the *Singular Value Decomposition(SVD)* and then apply it to our problem.

3.1 Singular Value Decomposition

Let $\mathbf{F} \in R^{M\times m}$, $M>m$. Then there exist unitary matrices $\mathbf{U} \in R^{M\times M}$ and $\mathbf{V} \in R^{m\times m}$ such that

$$\mathbf{U}^T\mathbf{F}\mathbf{V} = \mathrm{diag}\{\sigma_1,\ldots,\sigma_m\} \equiv \Sigma \in R^{M\times m} \qquad (4)$$

σ_k, k=1,…,m, arranged in descending order, are known as the *singular values* of **F**. The corresponding decomposition of the matrix **F** is known as the *Singular Value Decomposition*(SVD) of **F**[4]. It has also been shown that **U** is the matrix of eigen vectors of $\mathbf{F}\mathbf{F}^T$, **V** is the matrix of eigen vectors of $\mathbf{F}^T\mathbf{F}$, and σ_k, k=1,…,m are the non-negative square roots of the eigenvalues of $\mathbf{F}^T\mathbf{F}$. If we write $\mathbf{V} = \{v_{rs}\}$, r,s=1,…,m, and $\mathbf{U} = [\mathbf{u}_1,\ldots,\mathbf{u}_M]$, we can write **F** as

$$\mathbf{F} = [\sum_{i=1}^{m} (\sigma_i v_{1i})\mathbf{u}_i, \ldots, \sum_{i=1}^{m} (\sigma_i v_{mi})\mathbf{u}_i] \qquad (5)$$

Note that all the columns of **F** lie in the subspace spanned by $\mathbf{u}_i$, i=1,…,m.

The interesting case for us is when rank(**F**)=l<m, i.e. **F** is not full rank. Then, only the first l singular values are non-zero and the columns of **F** lie only in the subspace of $\mathbf{u}_i$, i=1,…,l. Note that the *SVD* of **F** finds both the dimension and a basis for the

subspace spanned by its columns. In the next section, we use this result to find p and $\mathbf{a}_i$, $i = 0,\ldots, p$ in (3).

3.2 Decomposition Algorithm

Recall from (3) that we have m measurements, $\mathbf{a}^{(j)}$, j=1,…,m, where $\mathbf{a}^{(j)} \in R^M$, j=1,…,m. If $\mathbf{F} \equiv [\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(m)}]$ and assuming that the additive noise terms in (3) are zero, we can write (3) as

$$\mathbf{F} = [\sum_{i=0}^{p} \alpha_i^{(1)} \mathbf{a}_i, \ldots, \sum_{i=0}^{p} \alpha_i^{(m)} \mathbf{a}_i] \tag{6}$$

By comparing (6) and (5) we conclude that

If $\mathbf{F} \equiv [\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(m)}]$ *and* $\mathbf{F} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ *is its Singular Value Decomposition, then* (p+1)=rank(**F**) *and is the number of non-zero singular values of* **F**, *and* $\mathbf{a}_i = \mathbf{u}_{(i+1)}$, i=0,…,p.

In the presence of the additive noise term, we will not necessarily have (p+1)=rank(**F**). In other words, we do not expect to get only (p+1) non-zero *Singular Values.* To go further, suppose that $\mathbf{N} \in R^{M\times m}$ is a random matrix with Gaussian distribution, zero mean with standard deviation σ, and all its elements are uncorrelated to each other. Then **G**=**F**+**N** will represent (6) in the presence of noise and we will have

$$\begin{aligned} \mathbf{E}(\mathbf{G}^T\mathbf{G}) &= \mathbf{E}(\mathbf{F}^T\mathbf{F} + \mathbf{F}^T\mathbf{N} + \mathbf{N}^T\mathbf{F} + \mathbf{N}^T\mathbf{N}) \\ &= \mathbf{F}^T\mathbf{F} + \mathbf{E}(\mathbf{N}^T\mathbf{N}) \\ &= \mathbf{F}^T\mathbf{F} + \sigma^2\mathbf{I} \end{aligned} \tag{7}$$

Therefore if σ_k, k=1,…,m are the *singular values* of **F,** and η_k, k=1,…,m are the *eigenvalues* of $\mathbf{E}(\mathbf{G}^T\mathbf{G})$, we have $\eta_k = \sigma_k^2 + \sigma^2$, k=1,…,m. Also, since $\mathbf{E}(\mathbf{G}\mathbf{G}^T) = \mathbf{F}\mathbf{F}^T + \sigma^2\mathbf{I}$, $\mathbf{a}_i$, i=0,…,p, are the first (p+1) *eigen vectors* of $\mathbf{E}(\mathbf{G}\mathbf{G}^T)$. We can now summarize the algorithm as follows:

1. Perform l measurements of m equivalent marks on the wafer, therefore we have m alignment signals per each measurement. We also acquire their corresponding center positions by metrology. Therefore without loss of generality, we assume that we have shifted all the signal such that they are centered at the origin. Let $\mathbf{G}_q \in R^{M\times m}$ to be the qth measurement set. By choosing l large enough, we can estimate $\mathbf{E}(\mathbf{G}\mathbf{G}^T) \cong \frac{1}{l}\sum_{q=1}^{l} \mathbf{G}_q\mathbf{G}_q^T$ and $\mathbf{E}(\mathbf{G}^T\mathbf{G}) \cong \frac{1}{l}\sum_{q=1}^{l} \mathbf{G}_q^T\mathbf{G}_q$.

2. Find the eigenvalues of $\mathbf{E}(\mathbf{G}^T\mathbf{G})$, η_k, k=1,…,m, arranged in descending order. $\sigma = \eta_m$, and $\sigma_k = \sqrt{\eta_k - \sigma^2}$, k=1,…,m; (p+1) is the number of σ_k, k=1,…,m, that are non-zero.

3. $\mathbf{a}_j$, j=0,…,p are the first (p+1) eigen vectors of $\mathbf{E}(\mathbf{G}\mathbf{G}^T)$ when their corresponding eigenvalues arranged in descending order.

In the next section, we use these basis vectors to find the position of any new alignment signal coming from the same process.

4. PART II: POSITIONING THE SIGNAL

Recall the general model given in (1). This equation can also be rewritten as

$$a(x-\delta;\mu_1,\ldots,\mu_p) = \sum_{i=0}^{p} \alpha_i a_i(x) + n(x) \tag{8}$$

Therefore assuming $n(x)$ to be white Gaussian noise, we can define the error as

$$\begin{aligned} E(\delta,\alpha_0,\ldots,\alpha_p) &= \int \left| a(x-\delta) - \sum_{i=0}^{p} \alpha_i a_i(x) \right|^2 dx \\ &= \int \left| A(f)e^{-2j\pi f\delta} - \sum_{i=0}^{p} \alpha_i A_i(f) \right|^2 df \end{aligned} \tag{9}$$

where $A(f)$ and $A_i(f)$ are the Fourier transforms of $a(x)$ and $a_i(x)$, respectively. When the signal is sampled, with sampling interval Δ, the error will be

$$E(\delta,\alpha_0,\ldots,\alpha_p) = \sum_{k=0}^{M-1} \left| A(f_k)e^{-2j\pi f_k \frac{\delta}{\Delta}} - \sum_{i=0}^{p} \alpha_i A_i(f_k) \right|^2 \tag{10}$$

where M is the length of the sampled signals. Defining $\mathbf{A}_i$, i=0,…,p, and $\mathbf{A}$ as the FFT of $\mathbf{a}_i$, i=0,…,p, and $\mathbf{a}$, respectively, $\mathbf{B} \equiv \mathrm{diag}\{e^{-2j\pi f_0 \frac{\delta}{\Delta}},\ldots,e^{-2j\pi f_{M-1} \frac{\delta}{\Delta}}\}$, $\mathbf{S} \equiv [\mathbf{A}_0,\ldots,\mathbf{A}_p]$, and $\boldsymbol{\alpha} \equiv [\alpha_0,\ldots,\alpha_p]^T$, we will have

$$E(\delta, \boldsymbol{\alpha}) = \left\| \mathbf{B}\mathbf{A} - \mathbf{S}\boldsymbol{\alpha} \right\|^2 \tag{11}$$

and by minimizing E with respect to $\boldsymbol{\alpha}$, we will get

$$\alpha_{\mathrm{opt}} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{B}\mathbf{A} \;\Rightarrow\; E(\delta) = \left\| (\mathbf{I} - (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T)\mathbf{B}\mathbf{A} \right\|^2 \tag{12}$$

Assume that δ is such that $\delta/\Delta << 1$. This is a valid assumption since the position of the signal is first found by a coarse

algorithm. In this case, we can linearize **B** using $e^{-2j\pi f_k \frac{\delta}{\Delta}} \cong 1 - 2j\pi f_k \frac{\delta}{\Delta}$ and therefore $E(\delta)$ will be a 2nd degree polynomial in δ given by

$$E(\delta) = \left\| (\mathbf{I} - (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T)\mathbf{B}_1\mathbf{A} \right\|^2 \quad (13)$$

where $\mathbf{B}_1$ is the linear approximation of **B** as mentioned before. The position of the signal, δ_{opt}, is measured by minimizing $E(\delta)$. We can now summarize the algorithm as follows:

1. Calculate **A** and $\mathbf{A}_i$, i=0,…,p, the Fourier transforms of **a** and $\mathbf{a}_i$, i=0,…,p, where the latter are calculated from part I of the algorithm. Also, create **S.**
2. Initialize δ_{total}=0.
3. Find δ by minimizing $E(\delta)$ given in (13)
4. **A=BA**, where δ in step 3 is used for constructing **B.**
5. $\delta_{total} = \delta_{total} + \delta$, go to step 3.

The iteration stops when there is no significant change in δ_{total}. Also, the iterative part relaxes the constraint $\delta/\Delta << 1$, so the algorithm can be applied even for δ/Δ up to unity.

5. IMPLEMENTATION

The algorithm was implemented in MATLAB©. p=2, l=20, m=10, M=128, Δ=250nm, σ=15nm were used in the simulation. Also, lowpass-filtered Gaussian functions were used as basis vectors in generating the simulated measurement signals. 200 measurement signals were then used in the part I of the algorithm to find the number of singular values, p+1, and the basis vectors. The results of part I of the algorithm were applied to find the positions of 100 new alignment signals. Since the signals are band-limited, only the frequency terms in the bandwidth of the signal were considered in finding the position of signal. This reduces the high frequency noise effects and also the processing time. In all the simulation examples, less than six iteration steps were needed to find the position.

Table 1 shows the mean and standard deviation of the positioning error and compares them to those for two widely used methods, Maximum Likelihood and the phase method. The latter uses the slope of the phase of the Fourier transform of the signal to find its center of symmetry.

The time required to compute the position of an alignment signal directly depends on the order of computation required by the algorithm. There are two sets of computation in our proposed algorithm:

1. Part I of the algorithm, which calculates the order of subspace and the basis vectors in the model. This is done once for a process and does not need to be repeated as long as the set-points of the physical parameters do not change. The computation complexity of this part is $O(M^2m^2)$.
2. Part II of the algorithm, which finds the position of a signal using the model from part I. The computation complexity of this part is $O(M\log(M))$.

	Proposed Algorithm	**Phase Detection**	**Maximum Likelihood**
Mean(*nm*)	0.67	35.64	21.16
STD(*nm*)	5.27	38.71	29.82

Table 1. Comparison between our proposed positioning algorithm and two other common algorithms. One hundred measurement signals were used in calculating the mean and standard deviation of the error. Results show an order of magnitude improvement in the standard deviation of the error.

6. SUMMARY

In this paper, we presented a new algorithm for positioning of non-symmetric signals. We assumed only small changes in the physical parameters contributing to the signals, and exploited this structure in our linear model. We proposed an algorithm to construct the model from a set of signals with pre-known center positions, i.e. training set. We then described an iterative algorithm to find the center position of a new signal from the model. Our simulation results indicated that the performance of this algorithm is about an order of magnitude better than current algorithms, under the same conditions.

The existence of a well-controlled environment so that (1) holds is essential for our approach. We would like to confirm these preliminary simulation results through actual experiments.

7. ACKNOWLEDGEMENT

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force office of scientific research under contract F49620-95-1-0525. It was also supported by DARPA/SRC under contract MDA972-971-0010.

8. REFERENCES

[1] Gallatin G. *Alignment Signal Generation and Analysis.* Short Course notes, SPIE conference on Microlithography, CA, 1995.

[2] Dirksen P., Juffermans C., Leeuwestein A., Mutsaers C., Nuijs A., Pellens R., Wolters R., Gemen J., "Effect of processing on the overlay performance of a wafer stepper". *Proceedings of SPIE conference on Microlithography*, CA, 1997.

[3] Goodman J., *Statistical Optics.* John wiley & Sons, 1985.

[4] Golub G.H., Van Loan C.F., *Matrix Computations*, John Hopkins University Press, MD, 1984.