Optimized Overlay Metrology Marks: Theory and Experiment

Mike Adel, Mark Ghinovker, Boris Golovanevsky, Pavel Izikson, Elyakim Kassel, Dan Yaffe, Alfred M. Bruckstein, Roman Goldenberg, Yossi Rubner, and Michael Rudzsky

Abstract—In this paper, we provide a detailed analysis of overlay metrology mark and find the mapping between various properties of mark patterns and the expected dynamic precision and fidelity of measurements. We formulate the optimality criteria and suggest an optimal overlay mark design in the sense of minimizing the Cramer–Rao lower bound on the estimation error. Based on the developed theoretical results, a new overlay mark family is proposed—the grating marks. A thorough testing performed on the new grating marks shows a strong correlation with the underlying theory and demonstrate the superior quality of the new design over the overlay patterns used today.

Index Terms—Box-in-box marks, Cramer–Rao lower bound, dynamic precision, Fisher information matrix, grating marks, overlay mark, overlay mark fidelity, overlay metrology.

I. INTRODUCTION

ACCURATE and precise overlay metrology is a critical requirement in order to achieve high product yield in microelectronic manufacturing. New challenges become evident as microlithography processes are developed for each new design rule node. A critical link in the overlay metrology chain is the metrology mark which is chosen to be included on the reticle, printed on the wafer, subsequently processed and which is ultimately imaged in the metrology tool in the metrology process. In this publication a theoretical and experimental study is described that shines new light on the limitations of existing mark designs while proposing and validating new designs of superior performance.

In Fig. 1 a standard overlay (BiB)[1] mark is shown schematically. It consists of two "boxes" printed on two subsequent layers—top (grey) and bottom (black)—between which the overlay is measured. By design the centers of symmetry of the inner (grey) and outer (black) boxes coincide. The actual

M. Adel, M. Ghinovker, B. Golovanevsky, P. Izikson, E. Kassel, and D. Yaffe are with KLA-Tencor, Optical Metrology Division, Migdal HaEmek 23100, Israel (e-mail: mike.adel@kla-tencor.com; mark.ghinovker@kla-tencor.com; boris.golovanevsky@kla-tencor.com; pavel.izikson@kla-tencor.com; Elyakim.Kassel@kla-tencor.com; Dan.Yaffe@kla-tencor.com).

A. M. Bruckstein, R. Goldenberg, Y. Rubner, and M. Rudzsky are with the Computer Science Department, Technion, Haifa 32000, Israel (e-mail: freddy@cs.technion.ac.il; romang@cs.technion.ac.il; yossi@rubner.co.il; rudzsky@cs.technion.ac.il).

[1]In the present paper we unify all conventional overlay mark types—box-in-box, bar-in-bar, frame-in-frame, etc.—under the generic abbreviation "BiB."
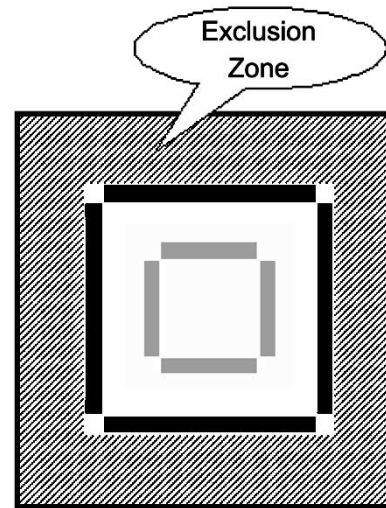

Fig. 1.   Standard BiB mark (schematically).

overlay appears as misregistration between the centers of symmetry of the "black" and "grey" layers.

There are two major use cases in overlay metrology for microlithography. The first and the most obvious is termed lot dispositioning. If measured overlay exceeds some allowable threshold, the lot cannot proceed to the next process step. This generally results in rework, that is the lot is returned to the previous lithography step after the resist is stripped. This is provided the overlay measurements were done immediately after development. Under some circumstances the overlay measurements after development are not viable, and are done after etch. In this case, there is no option for rework, and lots outside of allowable thresholds are scrapped.

The second use case of overlay metrology is for correction of the exposure tool. Usually, the overlay is measured at four corners of the field and over several fields on the wafer, which provides the necessary statistical sampling to enable stepper corrections model to be calculated. This model includes intra-field and inter-field correctibles, such as offset, rotation and scale. These correctibles are fed back to the exposure tool to improve performance on subsequent lots.

Conventional BiB based metrology has been the standard overlay metrology for almost two decades. However, as the overlay budget shrinks together with the lithographic design rules, a number of performance limitations are becoming evident. These shortcomings are addressed in the section below. Application of grating structures to lithography and metrology fields is being extensively studied. One of such applications

is for scatterometry based critical dimension (CD) metrology ([9], [16], [15], [11]). Gratings are also used for phase shift monitoring ([7]). ASML is using grating patterns as alignment marks ([14]). In the current paper, we introduce grating marks for overlay metrology.

A. Device Correlation

As design rules shrink to 100 nm and below, difference in BiB feature size and device feature size have become significant. Both lithographic pattern placement errors (PPE) [4], [10] and influences of other processes (like chemical-mechanical planarization—CMP) are known to be feature size and density dependent [12]. Therefore, overlay metrology results based on BiB marks may suffer from discrepancies compared with device feature overlay.

B. In-Chip to Scribe-Line Discrepancy

Another source of BiB-to-device overlay discrepancy originates from their different spatial location in the exposure tool field. Typically, BiB marks are printed in the scribe lines near the field corners. Optical conditions (aberrations, focus deviations, etc.) near the field edges may differ from those in the field interior, where the device features are printed. Since both overlay budget and process window (as defined by allowed exposure tool focus and exposure) shrink together with the design rules, in-chip to scribe-line discrepancies are becoming critical.

C. Process Robust Marks

Conventional BiB marks are frequently considered design rule violations in modern IC manufacturing processes. BiB marks are generically built of wide lines and require empty surrounding spaces (exclusion zones; see Fig. 1) for successful measurement. Both these facts usually contradict pattern density and feature size design rule requirements commonly in practice today. Such violations make handling of BiB marks by the layout engineer problematic and more importantly have a negative impact on process robustness of the metrology mark since the process is optimized for features and patterns of significantly different dimensions [6].

D. Tool Induced Shifts

Currently overlay measurements are performed on optical imaging based tools. Optical aberrations and illumination imperfections are an unavoidable reality of optical metrology system design and manufacture. A simple and quantitative metric of the quality of the optical metrology tool is Tool Induced Shift (TIS). TIS is defined as the average of the overlay measurements performed on a given overlay mark before and after rotation by 180°:

$$\text{TIS} = (\text{OVL}(0^0) + \text{OVL}(180^\circ))/2. \tag{1}$$

Nonzero TIS is an indication that the metrology tool has induced a systematic discrepancy in the overlay result due to the above system imperfections. TIS is however, by definition, a calibratable error, if measurements are performed at both orientations on a subset of representative marks. A more important
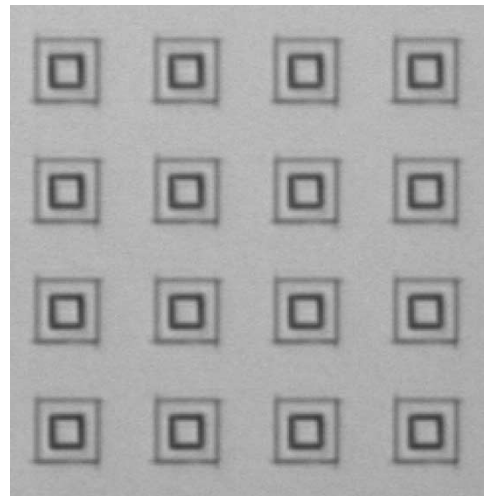

Fig. 2. Schematic arrays of densely printed overlay marks used for OMF calculation.

metrology uncertainty contributor is *TIS variability*, defined as three times the standard deviation of the TIS measured over N sites across the wafer.

E. Information Content

In spite of the large space occupied by the conventional BiB mark, it contains a relatively sparse amount of information for overlay measurement. Generically, each BiB mark consists of four inner and four outer bars only, usually utilizing less than 20% of the occupied real estate. By increasing the informational content of the overlay mark one can minimize the effect of random (both spatial and temporal) noise on overlay measurement. There are two measurable parameters representing overlay measurement uncertainty due to temporal and spatial noise: dynamic precision and overlay mark fidelity (OMF) respectively.

F. Dynamic Precision

Dynamic Precision is defined as three times the standard deviation of the results of a series of measurements of the same overlay mark, when these measurements are done in a dynamic loop (including wafer alignment, mark acquisition and measurement itself). This parameter quantifies temporal noise in the measurement of a given overlay mark.

G. Overlay Mark Fidelity (OMF)

Suppose, one can eliminate the temporal noise in the overlay measurement (by means of averaging over many dynamic loops of measurements on the same mark). This still does not ensure that by measuring two nominally identical marks one will obtain identical results. OMF is defined as three times the standard deviation of the measurement results of the N densely printed identical overlay marks after compensating for dynamic precision [1] (see Fig. 2).

In the present paper, we dwell on the last three shortcomings, that is information content, precision and mark fidelity. We first compare different mark design options from a theoretical perspective, focusing on the sampling, temporal noise, and spatial noise aspects. We then introduce an optimized grating overlay
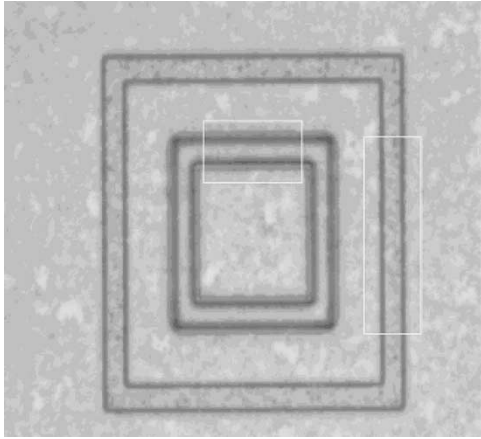
Fig. 3.   BiB overlay mark with right-outer and top-inner regions of interest.

mark, which demonstrates superior performance over the conventional BiB marks. We present experimental data on dynamic precision (as a measure of temporal noise) and OMF (as a measure of spatial noise) for the new grating overlay marks as compared with conventional BiB marks.

II. THEORY: DESIGNING PATTERNS FOR OPTIMAL OVERLAY REGISTRATION AND POSITION ESTIMATION

In this section, we analyze the dependence of the dynamic precision and fidelity of the overlay measurement on various pattern parameters. The overlay measurement is based on measurements of horizontal and vertical positions of known patterns.

Fig. 3 shows an example of the BiB overlay mark with right-outer and top-inner regions of interest.

The frame's edges are corrupted by noise whose character depends on various factors. We shall deal with two types of noise: additive Gaussian noise at the wafer level and additive Gaussian noise at the camera level. The first type of noise is spatial noise whose source is the manufacturing process, and the second type is a dominant source of the temporal noise in the measurement process.

In this paper, we explore how the position estimation error is affected by various pattern characteristics and by the parameters of the measurement process, by deriving the Cramer–Rao lower bound on the estimation error for arbitrary patterns and, then address the question of designing patterns that are optimal in the sense of minimizing the location error.

A. Problem Definition

We shall first deal with the measurement of horizontal position of a known one-dimensional (1-D) pattern $g_0(x)$ in a two-dimensional (1-D) image. Vertical position estimation can be done in a similar way. We assume in this case that the measurement is performed for every image row independently, and then an average pattern location estimate is returned as a result.

A periodic pattern of lines may be represented by

$$g(x,y) = g_0(x) + n_s(x,y) \tag{2}$$

where $g_0(x)$ is a 1-D pattern repeated on every line and $n_s(x,y)$ is the "spatial noise" on the wafer.

Then the pattern at each row of the image acquired by the camera can be described by

$$f(x) = g * h + n_t \tag{3}$$

where h is an overall point spread function, composed of the optical and camera point spread functions, $*$ is the convolution operator, and n_t is the temporal noise at the camera output. All signals are assumed band limited and the noise terms are assumed to be filtered and hence, band-limited, white, and Gaussian.

The task is to find best match locations of the designed pattern $g_0(x)$ given the signals $f(x)$ measured over all image rows. In the analysis below, we derive the distribution of the pattern location estimates over all measurements. The statistical analysis is based on the Cramer–Rao bound, a well-known statistical tool [8], [5].

We can now define the following three problems

1) Find the dependence of the dynamic precision and overlay mark fidelity (OMF) metrics on the general parameters characterizing an one dimensional pattern and the measurement method without going into the detailed structure of the pattern $g_0(x)$.
2) Given the pattern $g_0(x)$, what is a lower bound on the unbiased estimation of the pattern location?
3) What is the optimal pattern $g_0(x)$ in the sense of minimizing the Cramer–Rao lower bound on the estimation error?

B. Dynamic Precision and Fidelity Estimations

In this section, we evaluate the precision and fidelity of the measurement process based on some general physical parameters of the measured signal and the measurement process, such as optical system aperture, wave length, pattern size, signal-to-noise ratio (SNR), and others, without considering the detailed structure of the pattern $g_0(x)$.

1) Single Line Measurement: On every image row we estimate the pattern location by using the optimal Matched Filter or correlation method. Let $\hat{\theta}$ be the estimator of the pattern location θ of 1–D signal $f(x)$ immersed in white Gaussian noise. It is known on the basis of very general statistical principles that the variance of $\hat{\theta}$ is bounded below by the Cramer–Rao bound, which is given by (see, e.g., [8], [5]).

$$\text{var}(\hat{\theta}) = \frac{1}{d^2 \beta^2} \tag{4}$$

where $d^2 = 2E/N$, E is the signal's energy, N is the unilateral spectral density of the noise, and

$$\beta^2 = \int_{-\infty}^{\infty} \omega^2 |F(\omega)|^2 d\omega \Big/ \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega$$

is the square of the effective bandwidth of the signal, where $F(\omega)$ is a Fourier transform of $f(x)$. Using the definitions $E = PT$, $N = P_N/B$ and $\text{SNR} = P/P_N$ where P is the average signal power, T is the signal length, P_N is the noise power and B is the noise bandwidth, we get

$$\text{var}(\hat{\theta}) = \frac{1}{2\beta^2 \cdot T \cdot \text{SNR} \cdot B}.$$

This formula shows that the precision is a function of the signal and noise bandwidths, signal to noise ratio, and the overall length of the signal.

We shall deal with two kinds of noise.

- *Spatial noise* originating from the mark itself. This noise undergoes convolution with the point spread function h and therefore its band width is dictated by the optical system. In this case we assume $n_t(x, y) = 0$.
- *Temporal noise* originating mainly from the camera. This noise is of higher bandwidth since we work in conditions of over sampling and the pixel size is assumed to be roughly 5 times smaller than the optical resolution length. In this case we assume $n_s(x, y) = 0$.

In both cases, the effective bandwidth of the signal is dictated by the optical system. The upper bound for β is dictated by diffraction and can be approximated as the reciprocal of the Rayleigh resolution distance which is given by

$$\delta = \frac{0.61 \cdot \lambda}{\text{NA}}$$

where λ is the optical wavelength and NA is the numerical aperture of the optical system.

Using $\beta_{\max} = (2\pi/\delta) = (2\pi \cdot \text{NA}/0.61 \cdot \lambda)$ we get the following expression for the standard deviation of the pattern location:

$$\text{std}(\hat{\theta}) \geq \frac{0.61 \cdot \lambda}{2\pi \cdot \text{NA}} \sqrt{\frac{1}{2 \cdot T \cdot \text{SNR} \cdot B}}.$$

With all other factors equal, the standard deviation of temporal noise case is smaller because the noise bandwidth is larger. By increasing T while keeping all other factors constant we decrease the variance.

2) Multiple Line Measurement: When we repeat the measurement for multiple image rows we should get the same results with different random jitter or estimation noise. When we take the average of the measurements the precision is increased by the square root of the number of independent measurements. Let us measure L rows with the same method. We shall deal with two cases: Temporal Noise and Spatial Noise.

3) Temporal (Measurement) Noise: The number of independent measurements is L, and therefore, the standard deviation will obey

$$\text{std}(\hat{\theta}) \geq \frac{0.61 \cdot \lambda}{2\pi \cdot \text{NA}} \sqrt{\frac{1}{2 \cdot T \cdot \text{SNR} \cdot B \cdot L}}.$$

This gives us the dynamic precision of the measurement. If we put the following numbers as a concrete (real life) example $\lambda = 0.6\ \mu\text{m}$, $\text{NA} = 0.8$, $T = 5\ \mu\text{m}$, $\text{SNR} = 256$, $B = 6\ \mu\text{m}^{-1}$, $L = 100$ we get

$$\text{std}(\hat{\theta}) \geq 0.06\ \text{nm}.$$

4) Spatial (Pattern) Noise: In this case the number of independent measurements is not L since the optical point spread function blurs the noise. Let us call Δ the vertical sampling interval on the wafer. The number of independent measurements is approximated by

$$L_{\text{eff}} = L \frac{\Delta}{\delta}.$$

In addition the noise bandwidth B is roughly the inverse of the optical spatial resolution δ. The standard deviation of the measurement, interpreted as the Statistical Accuracy (OMF) will hence obey

$$\text{std}(\hat{\theta}) \geq \left(\frac{0.61 \cdot \lambda}{\text{NA}}\right)^2 \frac{1}{2\pi} \sqrt{\frac{1}{2 \cdot T \cdot \text{SNR} \cdot L \cdot \Delta}}.$$

If we use the following numbers for a concrete (real life) example $\lambda = 0.6\ \mu\text{m}$, $\text{NA} = 0.8$, $T = 5\ \mu\text{m}$, $\text{SNR} = 256$, $L = 100$, $\Delta = 0.08\ \mu\text{m}$, we obtain

$$\text{std}(\hat{\theta}) \geq 0.23\ \text{nm}.$$

The expressions for the statistical precision and fidelity (OMF) are very similar, except the dependence on NA and λ. In order to increase both precision and fidelity we need to decrease the wavelength, and to increase the information content of the signal (the effective bandwidth), the signal's spatial region (T and L) and the signal to noise ratio. In order to reach the bound we have to use a signal whose effective bandwidth achieves $\beta_{\max}$. This means that the signal fully exploits the frequency band up to the limit of diffraction.

C. Lower Bound Estimation—Exploring the Detailed Structure of the Pattern

In this section, we derive a lower bound on the estimation error for locating a known one dimensional pattern g_0. Unlike the previous section, this time we wish to get an expression for the Cramer–Rao bound for a *known* pattern, in order to gain some intuition on how optimal patterns should look like.

From (2) and (3), the observed signal $f(x)$ is given by

$$\begin{aligned} f(x) &= ((g_0 + n_s) * h) + n_t \\ &= g_0 * h + (n_s * h + n_t) = g_0 * h + n \end{aligned} \quad (5)$$

where $n = n_s * h + n_t$ is an overall noise added to the designed pattern g_0. We assume that h is a gaussian point spread function (PSF), with standard deviation of σ_s and $n(x) \sim N(0, \sigma_n)$ is a zero-mean white Gaussian noise.

Our measurement vector, $\breve{M} = \{\breve{M}_k\}$ is constructed by sampling the $f(x)$ on a uniform pixel grid. We assume that the first sample, $\breve{M}_0$, has an offset of θ from the origin.

Since $\breve{M}$ consists of statistically independent, due to the whiteness of the noise, measurements $\breve{M}_k$, and since the additive noise is Gaussian, we can write the probability function as

$$p(\breve{M} \mid \theta) = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma_n^2} \sum_{k=1}^{N} (\breve{M}_k(\theta) - \tilde{M}_k(\theta))^2\right\} \quad (6)$$

where N is the number of measurements (pixels), and $\tilde{M}_k$ is the set of samples of $g_0 * h$, taken at the same positions as $\breve{M}_k$.

For convenience we use another form of the (scalar) Cramer–Rao bound [8], [5]:

$$\text{var}(\hat{\theta}) \geq I^{-1}(\theta) \quad (7)$$

where $I(\theta)$ is the (scalar) Fisher information matrix

$$I(\theta) = -\text{E}\left[\frac{\partial^2 \ln p(\breve{M}|\theta)}{\partial \theta^2}\right].$$

For our Gaussian case, using (6) it is easy to derive that

$$
\begin{aligned}
I(\theta) &= \frac{1}{\sigma_n^2} \sum_{k=1}^{N} \left(\frac{\partial \tilde{M}_k(\theta)}{\partial \theta} \right)^2 \\
&= \frac{1}{\sigma_n^2} \sum_{k=1}^{N} \left(\frac{\partial}{\partial \theta} \left[g_0(x) * h_{\sigma_s}(x) \right]_{x=k+\theta} \right)^2 \\
&= \frac{1}{\sigma_n^2} \sum_{k=1}^{N} \left(\left[g_0(x) * h_{\sigma_s}'(x) \right]_{x=k+\theta} \right)^2 \\
&= \frac{1}{\sigma_n^2} \sum_{k=1}^{N} \left(- \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma_s^3} \exp\left\{ -\frac{x^2}{2\sigma_s^2} \right\} \right. \\
&\qquad \left. \times g_0(k+\theta-x)\, dx \right)^2
\end{aligned}
\tag{8}
$$

where $0 \leq \theta < 1$. Notice that the Fisher information term, and therefore the Cramer–Rao bound depend on θ.

Sometimes, more intuition regarding the properties of the pattern $g_0(x)$ can be gained from the following way to rewrite of the Fisher information expression in (8)

$$
\frac{1}{2\pi\sigma_n^2\sigma_s^6} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \exp\left\{ -\frac{x_1^2 + x_2^2}{2\sigma_s^2} \right\}
$$
$$
\sum_{k=1}^{N} g_0(k+\sigma-x_1) g_0(k+\sigma-x_2)\, dx_1\, dx_2.
$$

By limiting our discussion to binary input patterns, we can obtain even simpler expression for the Cramer–Rao bound. Assuming the pattern $g_0(x)$ is composed of B rectangular blocks, whose left end right edges coordinates are $\{(l_i, r_i)|i = 1 \ldots B\}$, the smoothed signal $g_0(x,\theta) * h_{\sigma_s}(x)$ can be easily computed using the erf function as

$$
g_0(x,\theta) * h_{\sigma_s}(x) = \frac{1}{2} \sum_{i=1}^{B} \left(erf\left(\frac{r_i + \theta - x}{\sqrt{2}\sigma_s} \right) \right.
$$
$$
\left. - erf\left(\frac{l_i + \theta - x}{\sqrt{2}\sigma_s} \right) \right)
$$

where

$$
\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.
$$

Then the Fisher information vector will be given by

$$
\begin{aligned}
I(\theta) &= \frac{1}{\sigma_n^2} \sum_{k=1}^{N} \left(\frac{\partial \tilde{M}_k(\theta)}{\partial \theta} \right)^2 \\
&= \frac{1}{\sqrt{2\pi}\sigma_s \sigma_n^2} \sum_{k=1}^{N} \sum_{i=1}^{B} \left[e^{-\frac{(r_i+\theta-k)^2}{2\sigma_s^2}} - e^{-\frac{(l_i+\theta-k)^2}{2\sigma_s^2}} \right].
\end{aligned}
\tag{9}
$$

D. Design of an Optimal Pattern $G_0(X)$

In [2](see also [3]) Bruckstein *et al.* showed how to design a overlay mark pattern of N pixels that achieves an estimation of the position with exponential lower bound $\Omega(2^{-N})$ accuracy. Optimality of this pattern was shown in the sense of information theory. Unfortunately, in our model that uses the Cramer–Rao

bound as optimality criteria, and includes PSF smoothing of the pattern and additive Gaussian noise, the BO&O overlay mark designed in [2] is not optimal.

The information theory approach in [2] uses the actual *values* of the pattern, i.e., zeros and ones. For every additional pixel, the uncertainty of the estimation that was achieved with the previous pixels is improved by a factor of two as the pattern is designed so that the last pixel has a value of "0" for half of the uncertainty area and a value of "1" for the other half. The Cramer–Rao bound on the other hand uses the *changes* in the pattern, i.e., its derivative. This means that the optimal pattern should have high derivative for all values of $\theta(0 \leq \theta < 1)$. The BO&O pattern of [2] is not optimal in the Cramer–Rao bound sense because of two main reasons: first, for the lower frequency part of the pattern, the derivative of the pattern is zero over a large "wasted" areas. Second, for the high frequency part, the changes in the pattern are too dense and will be completely smoothed out by the imaging PSF.

1) Designing an Offset Invariant Bound Pattern: As we have already mentioned, the Cramer–Rao bound is not a scalar (number), but rather a function of the offset θ. Hence the same pattern that undergoes different transformations (translations) yields different Cramer–Rao lower bounds for the different values of the transformation parameters. Therefore we would be interested in an integral measure that guarantees that for any transformation, i.e., for all the possible range of translations, we would be able to recover the transformation parameters with high accuracy.

This implies that the pattern is to be designed in such a way that its derivatives are as large as possible over all the x-support of the pattern (other than a small number of points where the derivative must change sign). Shih and Yu [13] deal with similar problem for the case of continuous value signal by solving the Cramer–Rao minimization problem using quadratic programming. In our case of binary signal the optimality is achieved by using a rectangular pulse signal where the distance between the pulses is $\delta = 4\sigma_s$. This distance brings adjacent pulses in the smoothed pattern to touch each other at points where they reach zero.

In order to maximize the minimum of the Fisher information over all the possible θ values, we shall need to distribute the binary blocks edges as evenly as possible.

2) Uniform Fractional Parts Distribution: The problem of even distribution of binary blocks over a given interval can be formulated in the following way: Given a constant δ, place a maximal possible number of points within an interval of length $T, \{X_i; 0 \leq X_i \leq T, \forall i\}$, in such a way that the minimal distance between adjacent points is greater than δ, and the fractional parts of the points coordinates $\{X_i - \lfloor X_i \rfloor\}$ are uniformly distributed in the interval $[0, 1]$.

By uniform distribution, we mean that if the number of points is M, their fractional parts should be $1/M$ apart from each other.

Without loss of generality we can place the first point at position zero. Indeed, if there exists a better mapping that starts from another position Δ, we can always shift all the points to the left by Δ thus retaining the same number of points and not disturbing the uniform distribution of the fractional parts of points coordinates.
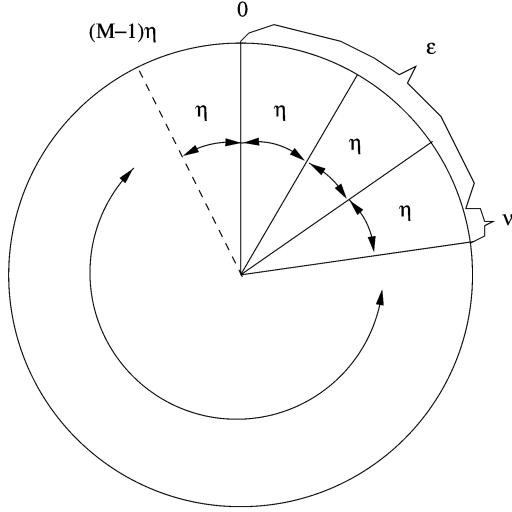
Fig. 4. Fractional parts distribution on unit circle.

If the minimal distance between the adjacent points is δ, then the maximal number of points that can be placed within the interval of length T is $M = \lfloor T/\delta \rfloor + 1$. Following our definition of uniformity, the fractional parts of points positions should be $\eta = 1/M$ apart from each other.

We suggest the following greedy algorithm for choosing M points in the interval $[0, T]$ such that their fractional parts cover the set $\{0, \eta, 2\eta, \ldots, (M-1)\eta\}$ and the points are at least δ apart from each other.

The algorithm:

1) Set Points $= \{0\}$
   Set UsedFractions $= \emptyset$
2) Repeat $M - 1$ times (steps 3–6)
3) Advance δ
4) Advance to the nearest position P_{cur}, whose fractional part $f_{\mathrm{cur}} = P_{\mathrm{cur}} - \lfloor P_{\mathrm{cur}} \rfloor$ is a multiple of η and $f_{\mathrm{cur}} \notin$ UsedFractions.
5) Points $=$ Points $\bigcup P_{\mathrm{cur}}$
6) UsedFractions $=$ UsedFractions $\bigcup f_{\mathrm{cur}}$

Let us analyze the performance of the algorithm above and check whether it can really place M points as required.

We denote the fractional part of the δ by ϵ, i.e., $\epsilon = \delta - \lfloor \delta \rfloor$.

Advancing by δ from the first point placed at 0 position brings us, in general, to a position, whose factional part is not a multiple of η. To get to the nearest multiple of η we have to skip an interval of length $\nu = \lceil \epsilon/\eta \rceil \eta - \epsilon$ (see Fig. 4). This situation emerges every time we advance by a step of size δ from a point aligned to an η-multiple fraction. Hence, the overhead of this η-alignment integrated for M points is bounded by

$$(M - 1)\nu \leq M\eta \leq 1.$$

Now let us estimate the overhead of the search for the position with fractional part that has not been visited yet. As we limit ourselves to fixed positions on the unit circle—the multiples of η—we can now switch to the discrete domain and express everything in terms of steps of size η on the unit circle. Let K be the number of η steps traversed on unit circle when advancing by $\delta + \nu$ on the interval. The K is given by

$$K = \lceil \epsilon/\eta \rceil, \quad 0 \leq K \leq M.$$

For the example depicted in Fig. 4, the $K = 3$. Let $G = gcd(K, M)$. Then by advancing every time by $(\delta + \nu)$ we visit M/G η-aligned locations on unit circle before returning to 0. Ideally, when the K and M are co-prime numbers all $M - 1$ fractions will be visited before returning to 0 position on unit circle, thus yielding a valid placement of M points with zero overhead (besides the $(M - 1)\nu$ mentioned above). In general case, for $G \neq 1$, every time we return to a position that has already been visited, one η-step is to be skipped in order to switch to another fractions chain of length M/G. Clearly, there will be $G - 1$ such chains, which means that the overhead is given by $(G-1)\eta \leq M\eta \leq 1$. Thus, the total overhead of the algorithm is

$$(M - 1)\nu + (G - 1)\eta \leq 2.$$

Therefore, the guaranteed lower bound on the number of points that can be placed by the algorithm within the interval of length T is given by

$$M' = \left\lfloor \frac{T - 2}{\delta} \right\rfloor + 1$$

and the actual maximal number of points can be found as

$$M^* = \max\{m | M' \leq m \leq M, (m - 1)\delta \\ + (m - 1)\nu + (G - 1)\eta \leq T\}.$$

E. The Dense Pattern

Using the binary blocks positions obtained by the algorithm described above we can build an optimal pattern of any length for any given value of σ_s.

Fig. 5 shows a comparative analysis of the newly designed "Dense Pattern" and the BO&O patterns. The left part of the figure shows the BO&O pattern and the right part shows the dense pattern. The original signal, the blurred signal, the blurred signal derivative and the Cramer–Rao bound are shown from top to bottom. Notice that for the BO&O pattern the Cramer–Rao lower bound is different for different values of θ, while for the dense pattern the bound is much lower and almost independent of the offset. Fig. 6 presents the simulation results confirming the theoretical bounds. The simulation was performed by shifting the patterns by various offsets from the interval [0, 1], adding Gaussian noise and looking for maximal correlation score with the original signal. For every offset the experiment was performed 100 times and the average estimation error is calculated. Both BO&O and the dense patterns are of the same length $T = 100$ pixels, the noise variance is $\sigma_n = 0.01$, and the PSF variance $\sigma_s = 0.5$. As predicted, the dense pattern yields
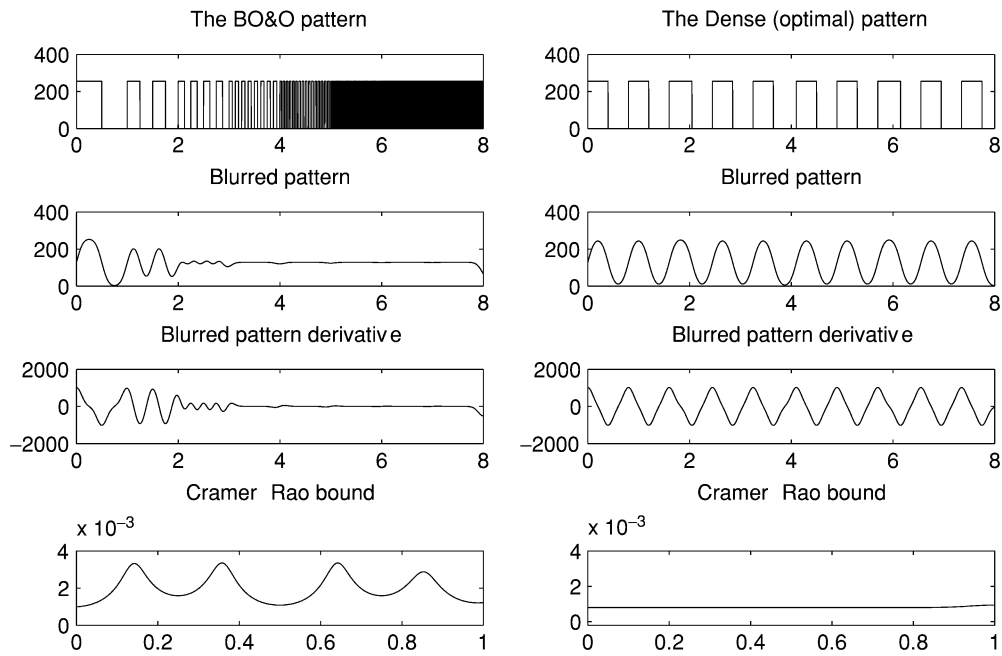
Fig. 5. BO&O and dense (optimal) overlay mark for eight pixels. First row: Original signals. Second row: Blurred signals. Third row: Derivatives. Fourth row: Cramer–Rao bound.
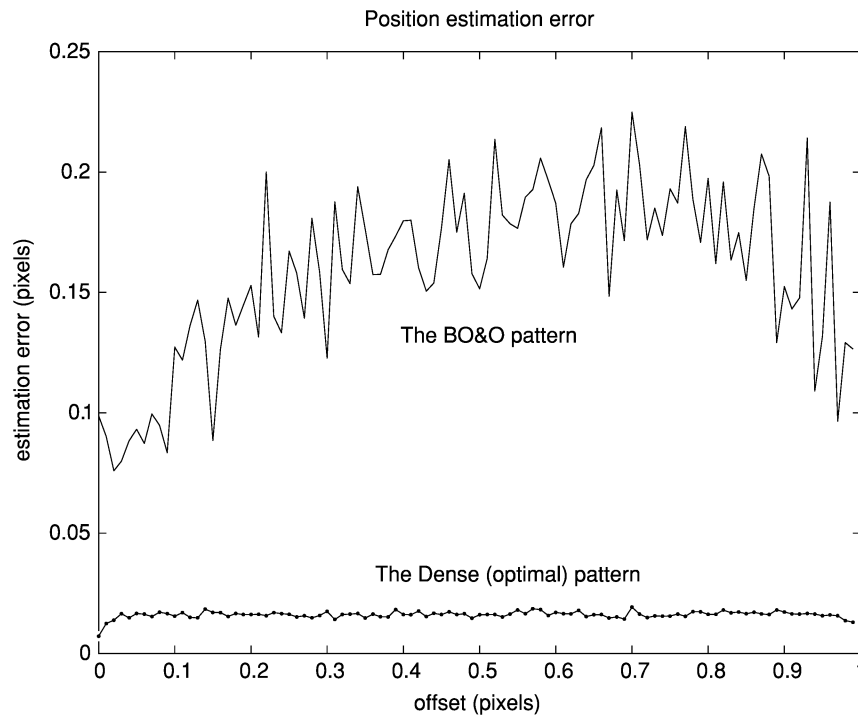

Fig. 6. Position estimation error for BO&O—solid line, and dense (optimal) pattern—dotted line.

lower estimation error, which is more uniformly distributed over the offset range.

III. EXPERIMENT: GRATING MARK—A NEW OPTIMIZED OVERLAY MARK

Motivated by the theoretical results presented above we designed and tested a new family of overlay marks—the grating marks.

Based on the same general principals as the optimal dense pattern described above, grating marks however differ in the following aspects:

- Grating marks, unlike the dense pattern, are periodic. This significantly simplifies the manufacturing process and it turns out that, at the current level of technology, the improvement achieved by uniform edge placement of the dense pattern is negligible compared to other sources of error.
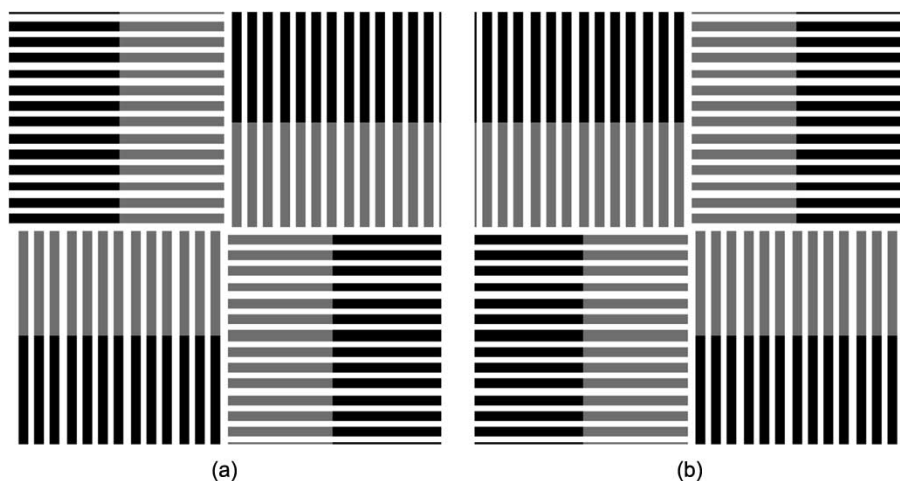
Fig. 7. Grating marks (schematically): (a) clockwise (CW) and (b) counterclockwise (CCW).
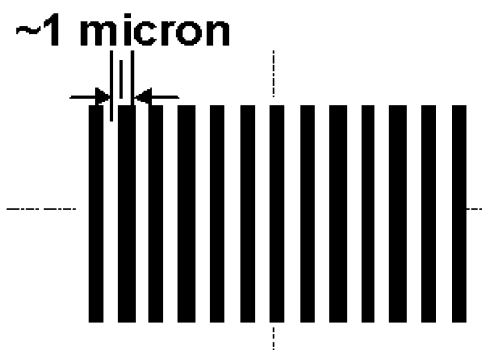

Fig. 8. Zoom to one grating mark octant. It is built of periodic series of lines.

- Grating marks, very much like the BO&O overlay mark, are designed as a multiscale structures. This facilitates application of multiresolution algorithms for faster pattern registration and position disambiguation (anti-aliasing).

In Fig. 7, a grating mark is shown schematically. Similarly to conventional BiB marks, the grating mark consists of inner (grey) and outer (black) structures printed on top and bottom layer. Each of these structures is symmetric with respect to 90^0 rotation. The grating mark consists of eight octants (four "grey" and four "black") and fills an area of $L \times L$, where L is the "grating mark size". Unlike conventional BiB marks, grating marks do not require any exclusion zone around them, i.e., any other structures can be printed in the immediate vicinity of the grating marks. Grating marks are comprised of structures at three scales of design:

- On the largest scale (of the order L) the grating mark consists of two layers ("grey" or "inner" and "black" or "outer") and eight octants. At this scale the grating mark is characterized by its size L, inner and outer layers, and grating mark chirality. There are two possible chiralities of the grating marks: clockwise (CW)—as shown in Fig. 7(a), and counterclockwise (CCW)—see Fig. 7(b).
- On the next spatial scale (termed the "*metrology interaction scale*") one can see that each octant comprises a periodic series of lines and spaces with a characteristic scale about 1 μm (see Fig. 8). This enhances information content and enables new image processing techniques due to
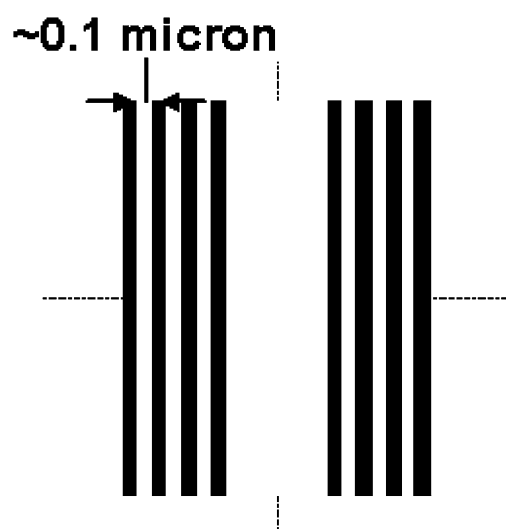

Fig. 9. Zoom to two "coarse" lines from the previous figure. Such "coarse" lines may be finely segmented.

the periodicity of the signal. This concept of a *periodic mark* is conceptually different from the conventional BiB approach. At this scale grating mark is characterized by *pitch* (i.e., period of the line series) and *duty cycle* (or line-to-period ratio).

- Finally, on the third spatial scale (termed "*lithography interaction scale*"), when we zoom in to a single line from Fig. 8, it appears that this line is finely segmented (with the design rule line and space pattern; see Fig. 9). This fine segmentation is typically below the optical resolution limit of the optical metrology tool, and therefore only the "coarse" lines create contrast in the acquired image. However, solid and finely segmented "coarse" lines are known to behave differently both lithographically and in process-related areas. At this scale, the grating mark is characterized by the fine segmentation pattern chosen.

A. Grating Mark Overlay Measurement

Similarly to the conventional BiB, the overlay of the grating mark is calculated as a misregistration between the centers of
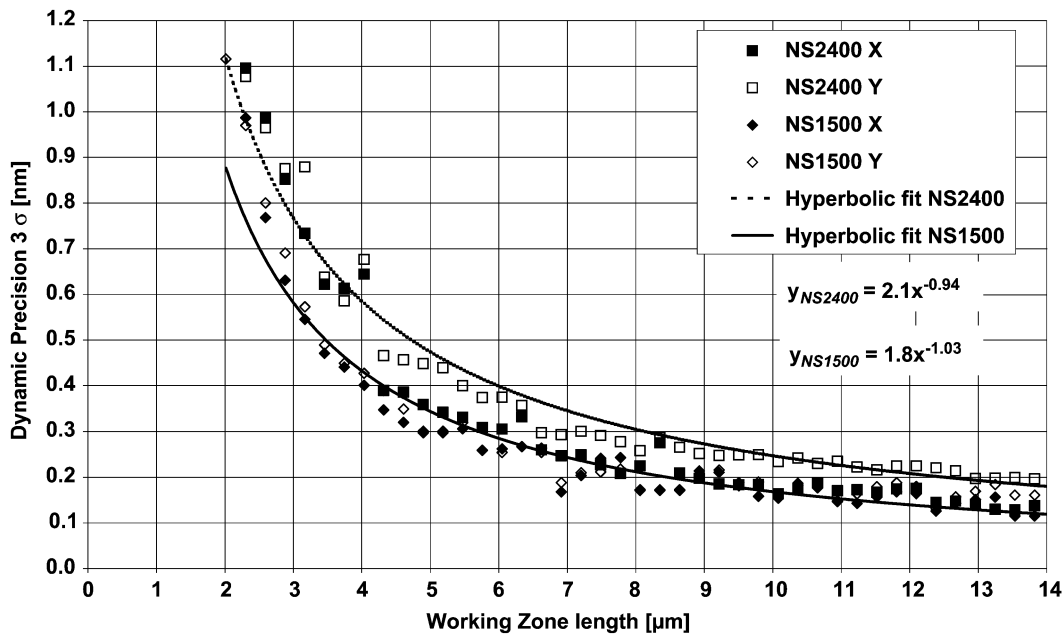
Fig. 10.   Dynamic precision of grating marks as function of the area of the ROIs.

symmetry of the inner ("grey") and outer ("black") patterns. An optimized measurement algorithm may be designed to utilize the pre-defined periodic nature of the grating mark patterns.

B. Improved Dynamic Precision of the Grating Mark

Contrary to conventional BiB marks, grating marks utilize the majority of the mark area for the overlay measurement. Thus information content of the grating mark is significantly higher than that of BiB marks. This increased information content results in improved dynamic precision and overlay mark fidelity.

In the experimental verification of the theory, we present results of measurements performed on two nonsegmented (NS) grating marks with two different grating periods (pitches). We have measured these marks in 10 dynamic loops and calculated dynamic precision as a function of the area of the mark utilized for the overlay calculation. We varied the width (T) and length (L) in a way preserving the same proportion T/L, and observed the dynamic precision of the overlay measurements as a function of T.

Fig. 10 shows graphs of the precision in X and Y directions, for grating marks with pitches of 2400 and 1500 nm. Included are best-fit power law curves of the graphs. It is observed that the experimental data closely follow a hyperbolic relationship between precision and mark area. This is in a good agreement with theory.

The theory provides a lower bound for precision, assuming the maximal effective bandwidth, but it does not give the pitch dependence. The better precision of the 1500-nm pitch mark relative to the 2400-nm pitch mark is in a qualitative agreement with the theory, indicating that the precision improves with the spatial frequency of the grating.

In addition we have performed extensive measurements of the dynamic precision on many different (both conventional BiB and new grating) marks on various layers and wafers, run under different process conditions. Wafers were run in three different

semiconductor manufacturing fabs, identified as *Fab1, Fab2* and *Fab3*, on four different processes described as follows.

- Poly/Active: the first patterning step was an active layer, followed by STI processing and an oxide CMP step. This was followed by a gate oxide process and polysilicon deposition. The second patterning step was at Poly.
- Etched Si: this process is a simplified version of the same sequence of patterning as above. Silicon was etched with the Active pattern. Then a layer of photoresist was spun over etched Silicon and patterned with the Poly reticle.
- Via/Metal: the first patterning step was on a dielectric stack and was processed as a Cu single-damascene metal layer. This was followed by Cu-CMP and deposition of an intermetallic dielectric stack. The second patterning step was at Via.
- Metal/Via: On a dielectric stack intended for Cu-dual damscene, the first patterning step was via. After via etch, photoresist was spun on the same stack and patterned with Metal trenches.

Fig. 11 summarizes dynamic precision performance of the grating overlay marks comparatively to conventional BiB. Nonsegmented and design rule segmented marks are grouped separately.

C. Improved Overlay Mark Fidelity (OMF) of the Grating Mark

In order to verify the anticipated reduction in spatial noise, we have measured an array of closely printed identical overlay marks (both BiB and grating marks). All the marks were measured 10 times in a dynamic loop to separate spatial noise from temporal noise. Fig. 12 shows the dependence of the overlay mark fidelity (OMF; which is experimental measure of spatial noise) on the area of the ROIs.

It can be seen that the graphs do not behave as hyperbolas. By increasing the kernel size we improve OMF up to some limit
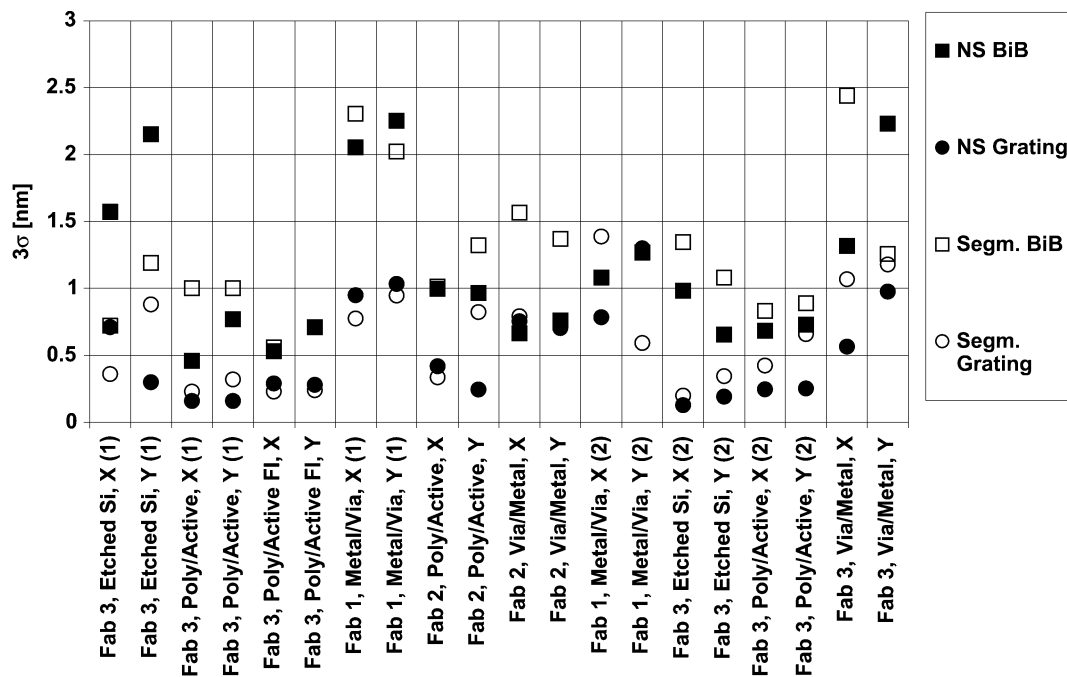
Fig. 11.   Dynamic precision of the old (BiB) and new (grating) overlay marks on many different process layers.
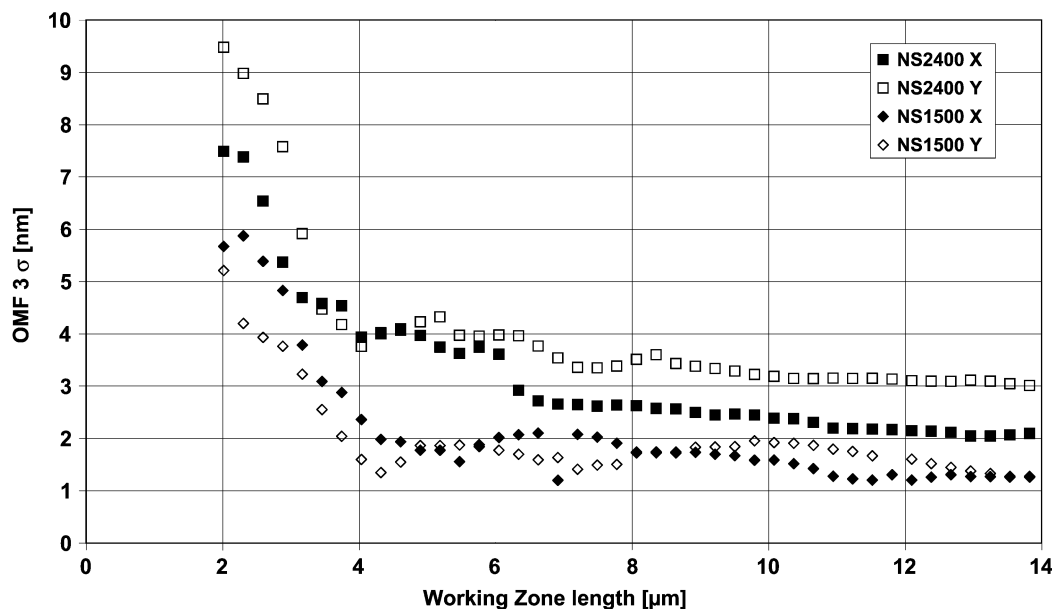

Fig. 12.   OMF of grating marks as function of the area of the ROIs.

(around 4 microns). Beyond this point OMF nearly saturates. This is believed to indicate that spatial noise does not behave as white noise. There are several possible explanations for this hypothesis. Firstly, there are some systematic errors such as reticle errors, which are different from one mark to another, which do however repeat themselves field to field over the wafer [1]. Furthermore, there are some frequency dependent sources of spatial noise due to the nature of wafer processing. Fig. 13 presents the summary of the OMF results of measurements made on many different (both conventional BiB and new grating) marks on various layers and wafers.

Although OMF is an effective metric to estimate the impact of process noise on the metrology uncertainty, there are additional factors that influence overlay metrology performance which should be mentioned. The impact of film stack thickness and composition on overlay metrology performance is twofold. Firstly, they are the key factors determining image contrast. Although there is no fundamental difference in the physics which determines the contrast in images of isolated (BiB like) versus grating structures, as discussed above, multiple edges increase information content and reduce the contrast threshold above which minimum metrology performance is achieved. Secondly, significant topographical differences between process layers may impact metrology tool performance. In this area, no significant differences in performance were detected between BiB and grating targets in the current study.
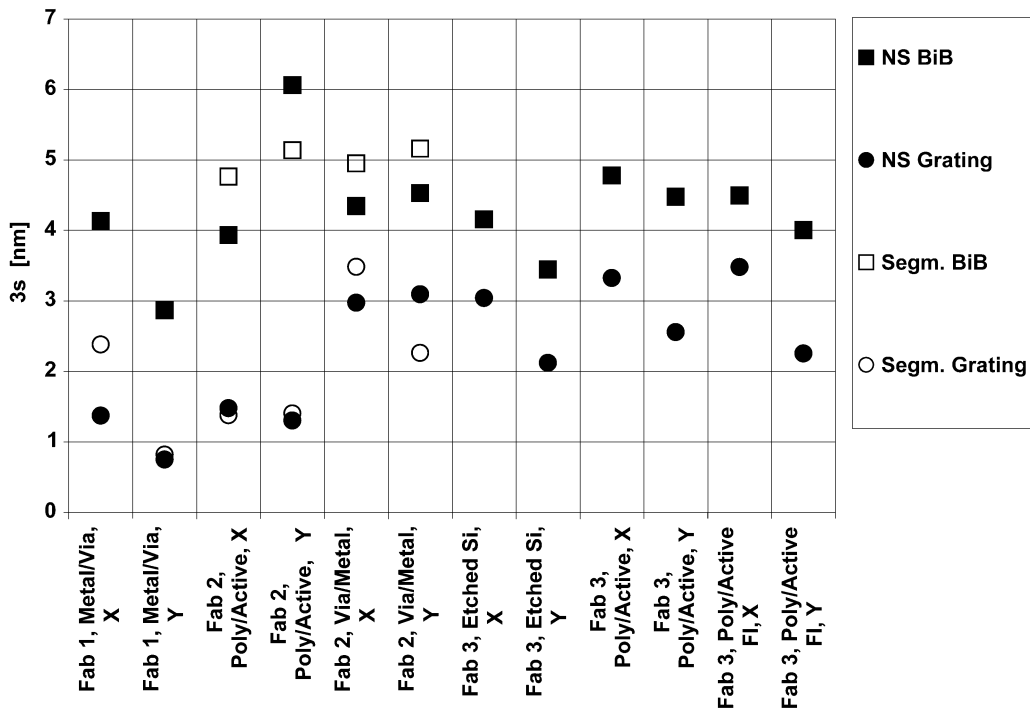
Fig. 13.   OMF of the old (BiB) and new (grating) overlay marks on many different process layers
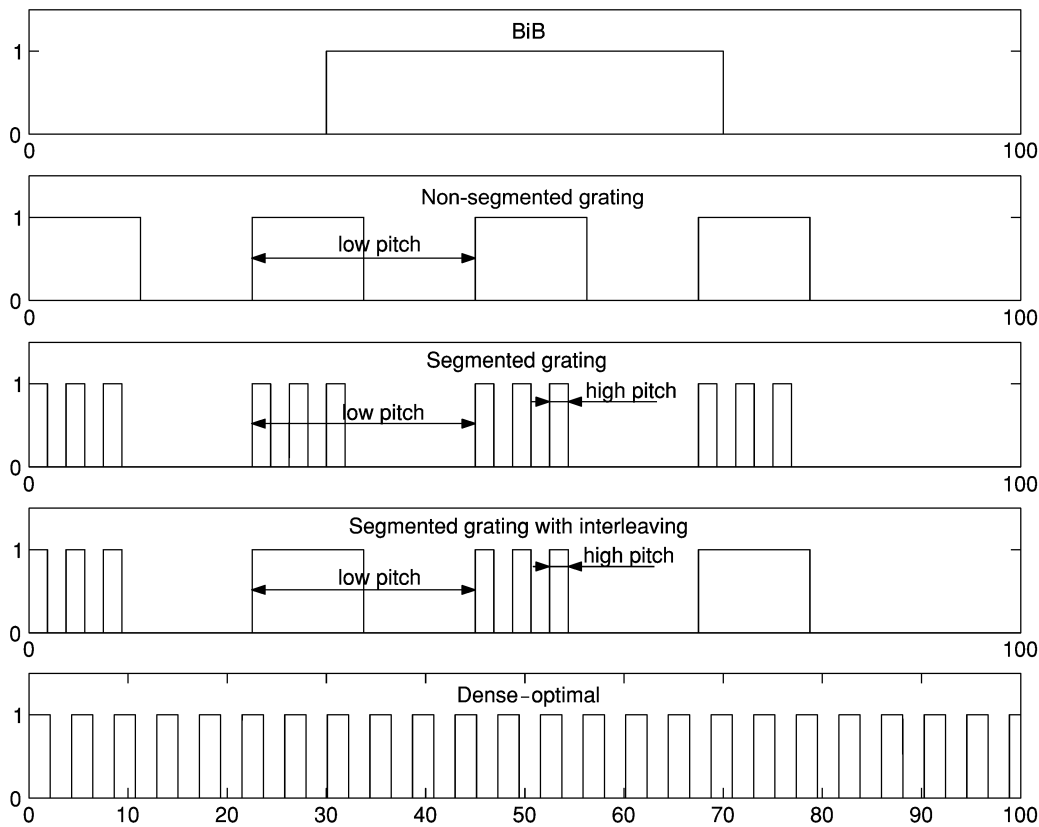

Fig. 14.   Tested patterns. From top to bottom: 1-D BiB, nonsegmented grating, segmented grating, segmented grating with interleaving, and the dense (optimal) pattern.

D. Grating Versus Dense Patterns—Simulation

Finally, let us look once again at the optimal dense pattern and estimate its advantage over the suboptimal grating patterns. Since we do not have a real dense pattern, we may try to answer this question by running a simulation as described in Section II-E. Fig. 14 presents the general form of the tested patterns: the 1-D BiB pattern, three grating patterns and the dense optimal pattern. Three grating patterns were tested using two different low pitches: 1500 and 2400 nm. The high pitch was 300 nm
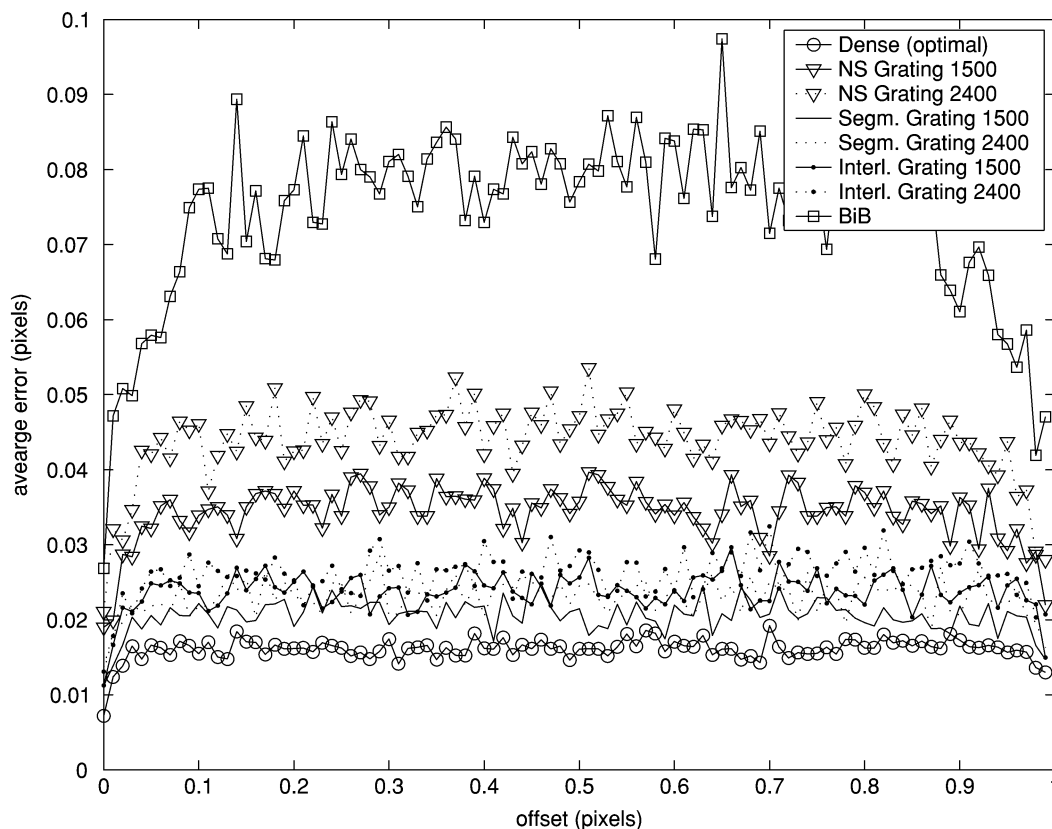
Fig. 15.   Average position estimation error for BiB pattern, three grating patterns with low pitch 1500- and 2400-nm and the dense (optimal) pattern.

in both cases. The simulation was performed in a 100 pixels window, where the pixel size is assumed to be 80 nm.

Simulation results are shown in Fig. 15. One can see that the dense pattern indeed achieves the best accuracy results and the error is approximately the same for the whole range of possible offsets. Another observation is that the precision achieved by the segmented grating mark is very close to that of the dense pattern.

IV. CONCLUDING REMARKS

In this paper, we conduct a thorough analysis of patterns used for overlay metrology and establish the dependence between various pattern properties and the expected dynamic precision and fidelity of the measurements. We show how the Cramer–Rao lower bound on the estimation error can be found for a given pattern.

We formulate a criteria for the offset invariant bound pattern and develop a uniform fractional parts distribution algorithm, which can be used to design an optimal pattern in a minimal Cramer–Rao lower bound sense. We suggest such an optimal design—the dense pattern—and provide a comparative, simulation based performance analysis with the commonly accepted BiB mark.

We then present a new family of overlay mark patterns—the grating marks and provide a detailed analysis of their properties based on real measurements. The measured dynamic precision and overlay mark fidelity of the new grating marks are close to the theoretically predicted values and demonstrate the superiority of the new overlay mark family over the existing BiB

marks. The measurements performed using a computer simulation also show that the grating marks have performance close to that of the optimal dense patterns.

REFERENCES

[1] M. E. Adel, M. Ghinovker, J. Poplawski, E. Kassel, P. Izikson, I. Pollentier, P. Leray, and D. Laidler, "Characterization of overlay mark fidelity," in *Proc. SPIE, Metrology, Inspection and Process Control for Microlithography XVII*, vol. 5, 2003, pp. 5038–5043.
[2] A. M. Bruckstein, L. O'Gorman, and A. Orlitsky, "Design of Shapes for Precise Image Registration," AT&T Bell Labs, Tech. Rep, 1989.
[3] ——, "Design of shapes for precise image registration," *IEEE Trans. Inform. Theory*, vol. 44, pp. 3156–3162, Nov. 1998.
[4] T. A. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. Res. Dev.*, vol. 41, no. 1/2, 1997.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Telecommunications.   New York: Wiley, 1991.
[6] P. Dirksen, C. A. Juffermans, A. Leeuwestein, C. Mutsaers, T. A. Nuijs, R. J. Pellens, R. Wolters, and J. Gemen, "Effect of processing on the overlay performance of a wafer stepper," in *Proc. SPIE, Metrology, Inspection and Process Control for Microlithography XI*, vol. 3050, 1997, pp. 102–113.
[7] B. La Fontaine, M. Dusa, J. Krist, A. Acheta, J. Kye, H. Levinson, C. Luijten, C. Sager, J. Thomas, and J. van Praagh, "Analysis of focus errors in lithography using phase-shift monitors," in *Proc. SPIE*, vol. 4691, 2002, pp. 315–324.
[8] C. W. Helstrom, *Elements of Signal Detection and Estimation*.   Englewood Cliffs, NJ: Prentice Hall, 1994.

[9] J. M. Holden, T. Gubiotti, W. A. McGaham, M. Dusab, and T. Kiersb, "Normal incidence spectroscopic ellipsometry and polarized reflectometry for measurement of photoresist critical dimensions," in *Proc. SPIE*, vol. 4989, 2002, pp. 1110–1121.

[10] A. Luci and E. G. Ballarin, "Optimization of alignment markers to limit the measurement error induced during exposure by lens aberration effects," in *Proc. SPIE, Metrology, Inspection, and Process Control for Microlithography XVI*, vol. 4690, 2002, p. 374.

[11] M. Littau, J. Ch.-Raymond, Ch. Gould, and Ch. Gambill, "Novel implementations of scatterometry for lithography process control," in *Proc. SPIE*, vol. 4689, 2002, pp. 506–516.

[12] B. F. Plambeck, N. Knoll, and P. Lord, "Characterization of chemical-mechanical polished overlay targets using coherence probe microscopy," in *Proc. SPIE, Integrated Circuit Metrology, Inspection and Process Control IX*, vol. 2439, 1995, pp. 298–308.

[13] S.-W. Shih and T.-Y. Yu, "On Designing an Isotropic Fiducial Mark," Multimedia Man-Machine Interface Laboratory, Department of Computer Science and Information Engineering, National Chi Nan University, Tech. Rep. TR-M3LAB-2002-002.

[14] J. Staecker, S. Arendt, K. Schumacher, E. Mos, R. van Haren, M. van der Schaar, R. Edart, W. Demmerle, and H. Tolsma, "Advances in process overlay on 300 mm wafers," in *Proc. SPIE*, vol. 4689, 2002, pp. 927–936.

[15] Y. Toyoshima, I. Kawata, Y. Usami, Y. Mitsui, A. Sezginer, E. Maiken, K.-C. Chan, K. Johnson, and D. Yonenaga, "Complementary use of scatterometry and SEM for photoresist profile and CD determination," in *Proc. SPIE*, vol. 4689, 2002, pp. 196–205.

[16] V. Ukraintsev, M. Kulkarni, C. Baum, K. Kirmse, M. Guevremont, S. Lakkapragada, K. Bhatia, P. Herrera, and U. Whitney, "Spectral scatterometry for 2D trench metrology of Low-K dual-damascene interconnect," in *Proc. SPIE*, vol. 4689, 2002, pp. 189–195.

Mike Adel received the B.Sc. degree in physics from the University of New South Wales, Sydney, Asutralia, in 1983 and D.Sc. degree in solid state physics from the Techion-Israel Institute of Technology, Haifa, in 1989.

Since graduation, he has held positions as an industrial physicist in areas ranging from electrooptical remote sensing, biomedical diagnostics and more recently semiconductor metrology. Since 1999, he has been employed by KLA-Tencor Israel, Migdal HaEmek, where he currently holds the position of Director of Advanced Development for the Optical Metrology Division. He is also an adjunct lecturer to the Technion Industrial Engineering faculty where he teaches hi-tech product development in the MBA program.

Mark Ghinovker received the M.Sc. degree in radiophysics and electronics from the Nizhnii Novgorod (formerly Gorkii) University, Russia, in 1990 and the Ph.D. degree in solid state physics from Bar-Ilan University, Ramat Gan, Israel, in 2000.

He dealt with theoretical and numerical study of superconductivity. Since graduation, he has been employed by KLA-Tencor Israel, Migdal HaEmek, as an algorithm and system engineer in the Advanced Development Group for the Optical Metrology Division. His current research interests span over the fields of superconductivity, signal processing, pattern recognition, numerical methods of solving PDE, optics, microlithography, and semiconductor physics.

Boris Golovanevsky received the M.Sc. degree in industrial heat technology from the Polytechnic Institute of Odessa, Odessa, U.S.S.R. (now Russia), in 1989 and M.Sc. degree in aerospace engineering from the Techion-Israel Institute of Technology, Haifa, in 1995.

Since 2000, he has been employed by KLA-Tencor Israel, Migdal HaEmek, where he currently holds the position of Research Scientist in Advanced Development group for the Optical Metrology Division.

Pavel Izikson received the M.A. degree in statistics from the University of Haifa, Haifa, Israel, in 2000.

Since 2000, he has been employed by KLA-Tencor Israel, Migdal HaEmek, where he currently holds the position of Statistician in Advanced Development for the Optical Metrology Division.

Elyakim Kassel graduated in 1986 from Ecole Nationale Superieure des Mines de Paris (France) majoring in materials science and engineering. He received the Ph.D. degree in physics from the University of Illinois at Chicago om 1991.

He joined National Semiconductor's Fairchild Research Center, Santa Clara, CA, in 1991. He worked there as a senior process engineer in etch. From 1992 to 1995, he was a senior process engineer in lithography at National Semiconductor (later Tower Seminconductor) in Migdal HaEmek, Israel. From 1995 to 2000, he worked at SemiConductor Devices, Misgav, Israel, as a quality engineer and subsequently as acting Quality Engineering manager and Reliability Growth team leader. Since 2000, he has been employed by KLA-Tencor Corporation, Optical Metrology Division, Migdal HaEmek, Israel, where he currently holds the position of project leader.

Dan Yaffe received the Ph.D. degree in physics from the Weizmann Institute of Science, Rehovoth, Israel, in 1971.

He worked at the Israeli Armament Authority (RAFAEL) from 1974 to 1985 on image and signal processing algorithms. Since 1985, he works as an independent consultant for Israeli companies on Computer Vision subjects. His current research interests include Pattern Recognition and Object Recognition.

Alfred M. Bruckstein was born in Transylvania, Romania, on January 24, 1954. He received the B.Sc. and M.Sc. degrees in electrical engineering, from the Technion-Israel Institute of Technology, in 1977 and 1980, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1984.

From October 1984, he has been with the Technion. From 1987 to 2000, he was a frequent visitor at AT&T Bell Laboratories, Murray Hill, NJ. His present research interests are in computer vision, ant robotics, pattern recognition, image processing, and computer graphics. He has also done work in estimation theory, signal processing, algorithmic aspects of inverse scattering, point processes and mathematical models in neurophysiology.

Prof. Bruckstein is a member of SIAM, MAA, and AMS, and presently he is the Dean of the Technion Graduate School.

Roman Goldenberg received the B.A. (*summa cum laude*) and Ph.D. degrees in computer science, from the Technion-Israel Institute of Technology, Haifa, in 1995 and 2003, respectively.

From 1994 to 1996 and 1999, he was with IBM Research Lab in Haifa. Currently, he is a research fellow at the Technion R&D Foundation and the Samuel Neeman Institute for Advanced Studies in Science and Technology. His research interests include video analysis, tracking, motion based recognition, PDE methods for image processing, and medical imaging.
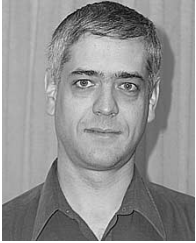
Yossi Rubner received the B.Sc. degree (*summa cum laude*) in computer science from the Technion-Institute of Technology, Haifa, in 1991, and the M.S. and Ph.D. degrees in computer science (with Ph.D. minor in electrical engineering) from Stanford University, Stanford, CA, in 1997 and 1999, respectively.

He is currently an entrepreneur and the principal of Rubner Technology Consulting where he is engaged in projects and consulting to various companies and venture capital firms in the areas of image processing and computer vision. From 2002 to 2003, he was with Applied Materials, Israel, where he managed the Image Processing groups of ETEC's Mask Inspection Product Group (MIPG). Between 1999–2001, he served as Vice President R&D and CTO of Jigami Corp., Israel, where he led the research and development of multimedia acceleration platforms for cellular operators. His previous experience includes consulting for Xerox Corporation in the Palo Alto Research Center (PARC) in the area of image understanding, development of desktop publishing systems at the Scitex Corporation, and leading R&D projects in an Israel Defense Force Technical Unit from which he was honorably discharged with the rank of Captain.

Michael Rudzsky received the Ph.D. degree in physics and mathematics from the Institute of Space Researsh, Moscow, U.S.S.R. (now Russia), in 1980.

He worked in the Scientific and Industrial Association for Space Research, Baku, Azerbajan, till 1990. Since 1991, he was a research fellow in the Physics Department and since 1995 at the Computer Science Department of the Technion, Haifa, Israel. His current research interests include computer vision, pattern recognition, and compression of images.