

Semiconductor sidewall shape estimation

Philip R. Bingham

Jeffery R. Price

Kenneth W. Tobin

Thomas P. Karnowski

Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6010
E-mail: binghampr@ornl.gov

Abstract. For process control, linewidth measurements are commonly performed on semiconductor wafers using top-down images from critical dimension measurement scanning electron microscopes (CD-SEMs). However, a measure of the line sidewall shape will be required as linewidths continue to shrink. Sidewall shape can be measured by physically cleaving the device and performing an SEM scan of the cross section, but this process is time consuming and results in destruction of product wafers. We develop a technique to estimate sidewall shape from top-down SEM images using pattern recognition based on historical cross section/top-down image pairs. Features are computed on subimages extracted from the top-down images. Several combinations of principal component analysis (PCA) and flavors of linear discriminant analysis (LDA) are employed to reduce the dimensionality of the feature vectors and maximize the spread between different sidewall shapes. Direct, weighted LDA (DW-LDA) results in a feature set that provides the best sidewall shape estimation. Experimental testing of the sidewall estimation system shows a root mean square error of approximately 1.8% of the linewidth, showing that this system is a viable method for estimating sidewall shape with little impact on the fabrication process (no new hardware and a minimal increase in process setup). © 2004 SPIE and IS&T. [DOI: 10.1117/1.1763586]

1 Introduction

In current semiconductor manufacturing environments, critical dimension (CD) (i.e., linewidth) measurements are made almost exclusively using scanning electron microscope (SEM) images. This process—known as CD SEM (CD-SEM) metrology—employs images that are usually acquired in a top-down configuration, i.e., looking down onto the semiconductor line feature. Yield management teams currently use the CD-SEM measurements to monitor the lithographic process and make corrections that keep the process within the required operating region. According to the International Technology Roadmap for Semiconductors, continually shrinking linewidths make it increasingly important to know the sidewall shape (e.g., the cross section profile) of the lines rather than just their width. An example of this can be seen in Fig. 1, which shows the top-down images and corresponding cross-section images for two different lines. The top-down images could easily produce similar CD measurements, but the cross-section images

show these are two very different lines with the left line undercut and the right line overcut. While cross-section images such as those shown in Fig. 1 can be used to distinguish between sidewall shapes, the physical cleave of the wafer required to make cross-section images is costly in both time and product. Since CD-SEM tools that capture top-down images are currently being used for measurement of CD, we investigated the possibility of estimating sidewall shape using only features extracted from top-down images and a database of corresponding historical top-down and cross-section images.^{1,2}

In this paper, we propose an image retrieval system to estimate sidewall structure from top-down imagery. Features from a top-down query image are compared to a database of features from other top-down images, each with known corresponding sidewall shapes. The sidewall shape of the query top-down is estimated using the sidewall shapes of the retrieved top-downs. First, a historical repository of corresponding top-down and cross-section image pairs is constructed. Sidewall profiles are extracted from each cross-section image and stored. Features are computed from one or more subimages in each top-down image. As the number of the computed features for each such line region is quite large, dimensionality reduction is performed to make feature storage and feature vector comparisons (i.e., image retrieval) more tractable for large databases. In this paper, we introduce several modifications to our earlier work. First, we use different regions of the top-down images for feature extraction. In the previous work, the top-down subimages covered the entire width of a line, like those shown in Fig. 1, but here only line edges are contained in the subimages. Second, Gabor filter responses for some features are employed in hopes of capturing 2-D, texture-like characteristics of the line-edge subimages. Finally, direct, weighted linear discriminant analysis (DW-LDA) is applied for dimensionality reduction.

The remainder of this paper is organized as follows. Section 2 presents some background information on the development of this estimation approach and how the approach fits in the current process control scheme. Next, we describe the extraction and parameterization of sidewall shapes and the process used to extract features from the top-down images in Sec. 3. Section 4 discusses the use of DW-LDA to reduce the feature vector dimensionality. We

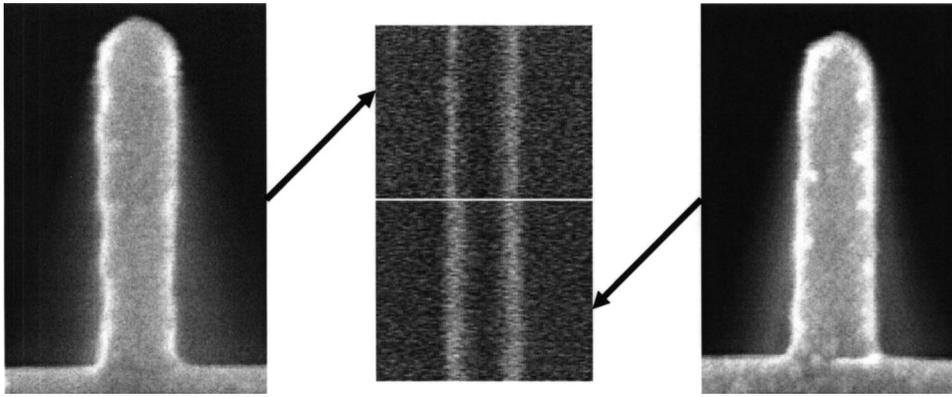


Fig. 1 Two different cross sections with similar top-down linewidths.

present experimental results for the sidewall estimation over a variety of linewidths and pitches in Sec. 5. Finally, we conclude in Sec. 6 with some comments and future paths for the research.

2 Background

In the development of an image from any sensor, many factors determine the response of the sensor to various surface types and shapes. For example, the illumination angle can greatly affect the image captured by a traditional camera. Top-down SEM images are influenced by many factors of both the inspection surface topology and the SEM setup and geometry. Figure 2 shows an example of an electron beam striking the surface of three different line types from overcut to undercut with the electron detector positioned at an angle. A very basic interaction of the electron beam with the line material is shown and clearly indicates that the number of scattered electrons escaping from the material increases at the beam location as the line moves from overcut (left) to undercut (right). However, the backscattered electrons may not reach the detector depending on the detector position.

Even though Fig. 2 shows only a simple model of the electron beam interaction. In this figure, the gray spider web shows the interaction volume. Therefore, if more of this gray region reaches the edges, more electrons will escape and be collected by the detector. This example quickly indicates the complexity of the SEM response to various shapes. The shape and properties of the surface material as

well as the electron-beam shape and power affect the interaction between the electron beam and the surface. A list of variables affecting SEM image formation is

1. electron beam energy
2. material properties
3. geometry of feature (width, height, wall angle, edge sharpness, etc.)
4. relationship to surrounding features (isolated or dense region)
5. charging effects (scanning speed, beam current, and all the preceding factors)
6. detector position, energy selectivity, and amplifier characteristics

Based on variation in SEM interaction with different line shapes, the good news is that the sidewall shapes should affect the top-down SEM image. However, interpretation of sidewall shape from a top-down SEM image is a very difficult modeling problem due to the many factors just listed. For this reason, we have taken a learning approach to determining sidewall structure from top-down SEM imagery. A learning approach takes into account all of the parameters of the SEM and surface materials by using training data of known results taken from the same SEM setup and product material.

Training data is commonly available for this approach since focus/exposure (F/E) matrices are often used to char-

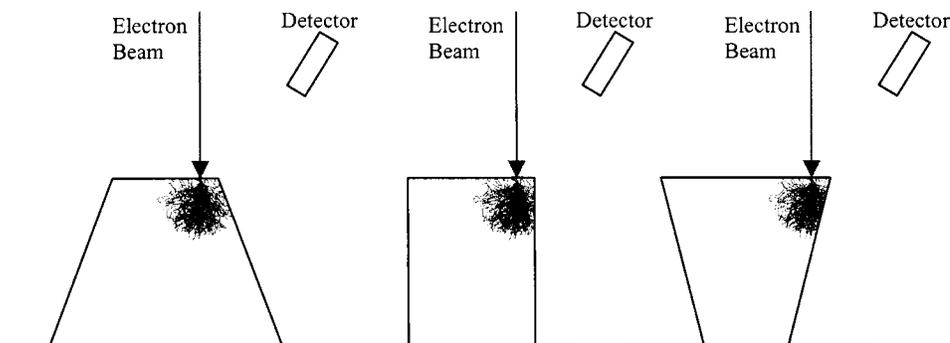


Fig. 2 Electron beam interaction with various line shapes.

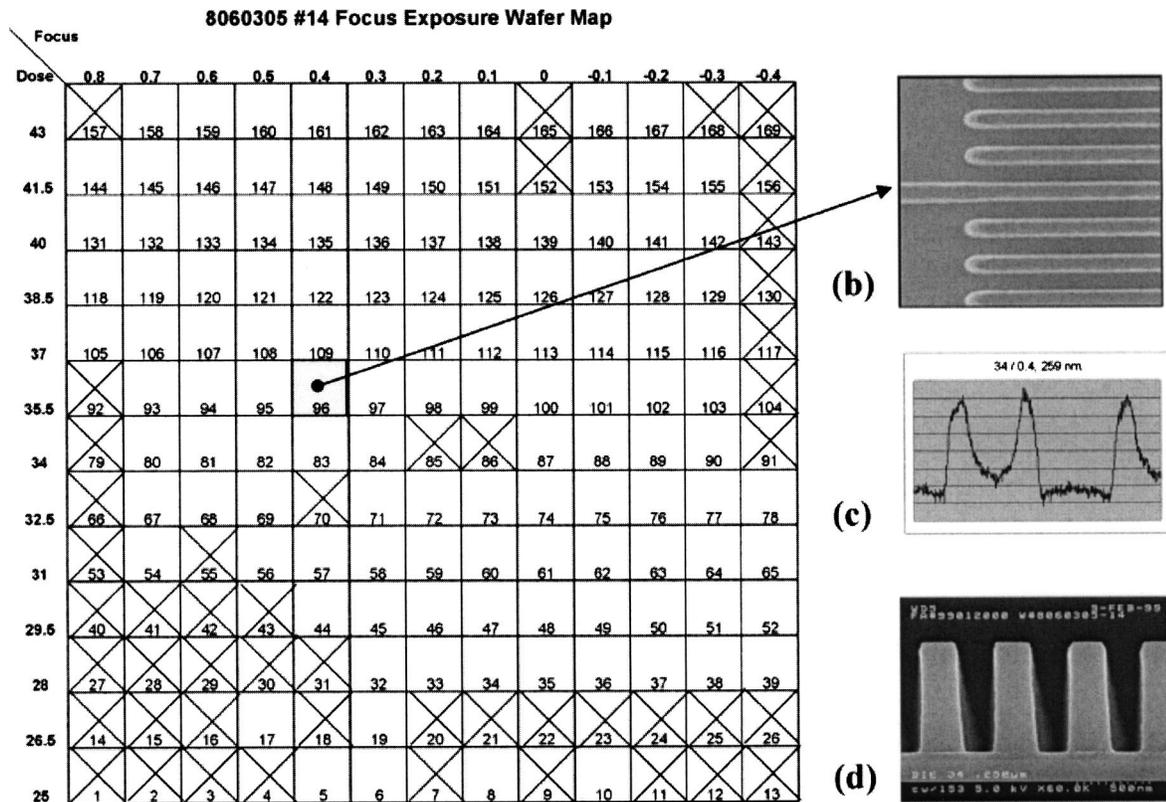


Fig. 3 FEM data: (a) matrix describing focus exposure levels, (b) top-down SEM image, (c) linescan of top-down SEM image, and (d) SEM image of cross section. Squares that have been crossed out indicate data that is unavailable, generally due to physical damage caused in the cleaving process.

acterize the lithographic process before production. An F/E matrix is a wafer on which focus and exposure are varied between die of the wafer to determine the response of the lithographic process with respect to these two parameters. Figure 3 depicts the setup for one of the F/E matrices used in this research. At each of the locations in the matrix, a top-down SEM image is acquired. Next, the F/E matrix wafer is cleaved at those same locations and a cross-section SEM image is captured. Thus, the F/E matrix provides top-down and cross-section data over a range of focus and exposure settings. This results in a good sampling of various sidewall shapes that may be encountered during production. As long as the matrix provides ample representation of all possible sidewall shapes during production, a learning system is possible. During process setup, only a portion of the F/E matrix is cross sectioned. Possible operating regions are selected using the top-down SEM images, and the operation region is further reduced by cross sectioning the regions selected from the top-down imagery. Therefore, using the F/E matrix as a training set may require more effort during process setup, but the gain of being able to accurately predict sidewall shape from top-down SEM images should be well worth the additional setup effort. We expect that historical F/E matrix data can be used to train the system to reduce the dependence on acquiring a new F/E matrix training set for each new product setup, but further testing is necessary. For the development and testing of our sidewall estimation method, International SEMATECH

provided Oak Ridge National Laboratory (ORNL) with a variety of F/E matrix datasets that span several design rules, line pitches, and CD-SEM tools.

As evident from the preceding discussion, top-down SEM images do contain some information that is relative to the sidewall shape. This relationship, however, is difficult to model due to the number of interacting factors. Historical imagery of top-down and cross-section pairs, however, is readily available in the form of F/E matrices already routinely captured in process control. Therefore, a learning-based sidewall estimation approach can provide a low-cost method, requiring no new imaging tools, to control the lithographic process using sidewall shape information.

3 Feature Extraction

Figure 4 shows an overview of the proposed semiconductor sidewall shape estimation system in the form of a block diagram. In addition to a database that holds training data, the two major database interaction components are the build and query components. The build component is used to enter cross-section/top-down pair training data to the system and consists of algorithms to extract information (features) from the top-down images that distinguish between various cross-section types, dimensionality reduction algorithms for determining the feature combinations providing the best discrimination between sidewall shapes, and algorithms for extraction of sidewall shape parameters from

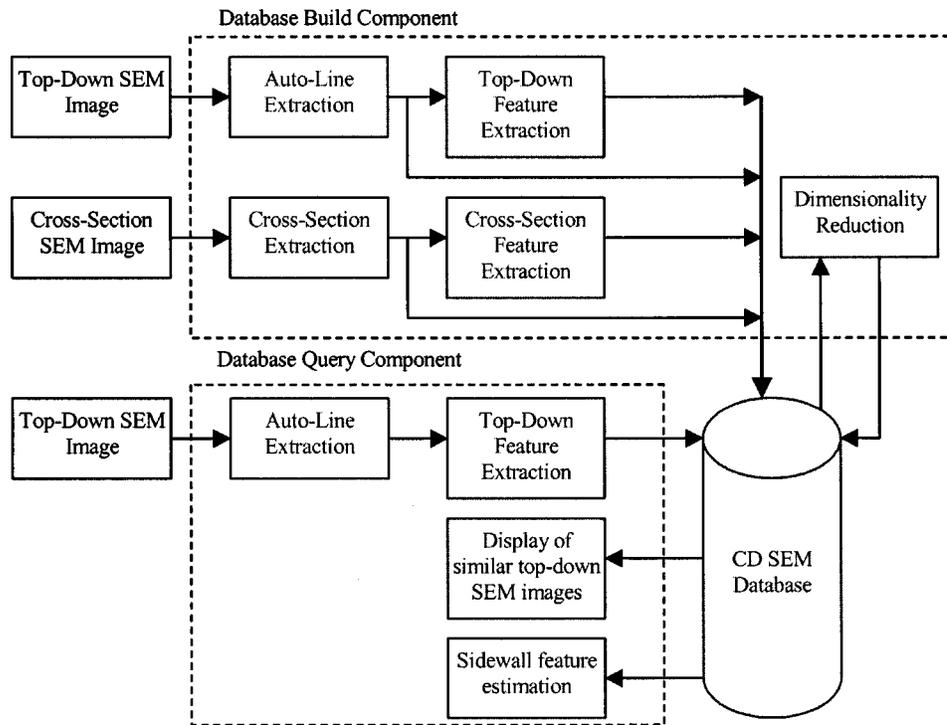


Fig. 4 Sidewall estimation system block diagram.

the corresponding cross-section image. The query component is responsible for submission of top-down images for sidewall estimation, searching the database for similar top-downs, and returning a sidewall estimate. Features must be extracted from a top-down image submitted for sidewall estimation, so the query component uses the same algorithms for extraction of features from top-down images used in the build component. Additionally, the query component contains algorithms for searching and combining similar cross sections to give a sidewall shape estimate. The following subsections detail feature extraction from both cross-section and top-down images and dimensionality reduction for the top-down feature set.

3.1 Sidewall Extraction

Once a wafer has been cleaved and a SEM cross-section image acquired, the sidewall shape must be extracted from the cross-section image. Initial attempts were to automate the extraction, but even with the small number of cross-section images we have encountered (<500 images), the variety in the images has made automation difficult. For the initial implementation, we have developed a semiautomatic algorithm that has been successful. Full automation of the sidewall extraction is not as important as for the top-down feature extraction, because sidewall extraction only needs to be performed during the training process and not during sidewall estimation. Currently, the sidewall extraction process requires the user to measure the scale on the image, rotate the image such that the substrate is on the horizon, select a line in the cross-section image to be extracted, and adjust two threshold levels to properly extract the sidewall. When placed in a more constrained environment (same cross-section SEM tool and setup), the sidewall extraction should be easily automated. As mentioned earlier, the

cross-section extractor is necessary only during the training cycle. Therefore, those using the tool to estimate sidewalls need not learn how to extract the sidewall shape from a cross-section image. This enables the tool to be easily used in the fab once an “expert” has entered the training data.

Once the user has rotationally aligned the cross section and selected a subimage containing the cross section of a single line, histogram normalization is used to alleviate contrast variations experienced in the SEM images, and scaling is performed to ensure cross-section sidewall parameters in the training set are on the same scale. Finally, edges are extracted from the subimage using the algorithm outlined in Fig. 5.

A Gaussian filter is used to remove roughness on the cross-sectioned line material. A first step toward edge extraction uses a Sobel edge filter followed by an edge threshold set by the user. This intermediate binary edge image is multiplied by the original subimage to place the intensity values of the cross section in the edge image. Since many of the line edges are bright, an intensity threshold set by the user enables further isolation of the cross-section edges from other edges in the image. Finally, clustering is performed and clusters smaller than 100 pixels are filtered out removing any spurious edges along the base of the line and those resulting from large rough features within the cleaved line. Notice that two thresholds must be supplied to this edge extraction algorithm. With adjustment of these two thresholds, we were able to employ this sidewall extraction algorithm easily and effectively for all of the cross sections in our data set.

After edges have been isolated, a set of features uniquely describing the cross section is extracted. One goal for the finished system is the capability to estimate sidewall shapes across various design rules (i.e., different line-

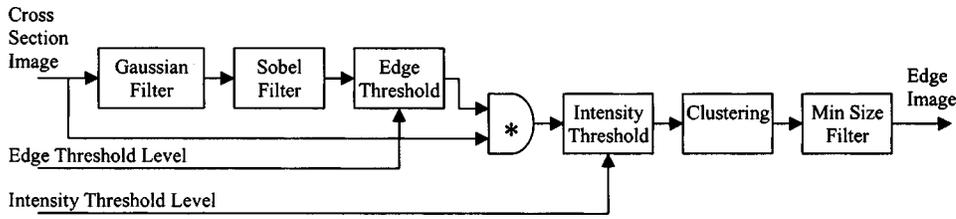


Fig. 5 Edge extraction algorithm.

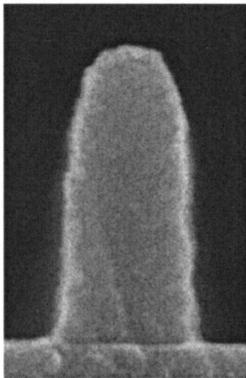
widths, pitches, and aspect ratios). Toward this purpose, we seek a representation that is invariant to the already mentioned design rule parameters. We therefore define the sidewall shape as the normalized width at 101 equally spaced points from top to bottom. These widths (at several locations) are illustrated by the horizontal dashed lines in Figs. 6(b) and 6(e). Letting the widths (in nanometers) be represented by the 101-point vector \mathbf{w} , where $w_n = w(n)$ for $n = 0, \dots, 100$ ($n=0$ is the top) and letting the design rule (i.e., target) linewidth (in nanometers) be represented by L , the normalized sidewall shape representation (as a vector) is given by

$$\mathbf{c} = \frac{1}{L} (\mathbf{w} - \bar{\mathbf{w}}), \quad (1)$$

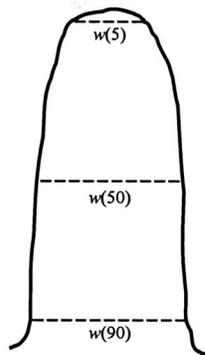
where $\bar{\mathbf{w}}$ is the approximate width at the sidewall vertical midpoint, as given by

$$\bar{\mathbf{w}} = \frac{1}{21} \sum_{n=40}^{60} w(n). \quad (2)$$

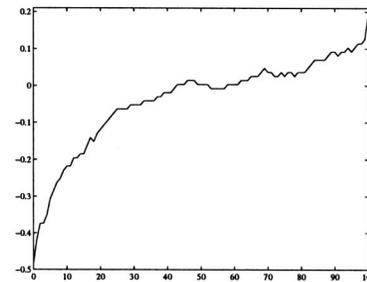
Examples of the resulting sidewall width curves are shown in Figs. 6(c) and 6(f).



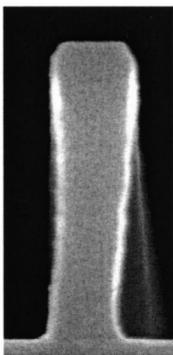
(a) Line sidewall (180nm, 1:1 pitch).



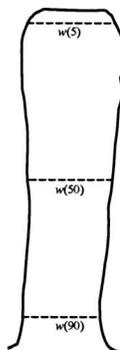
(b) Profile extracted from (a) with widths indicated.



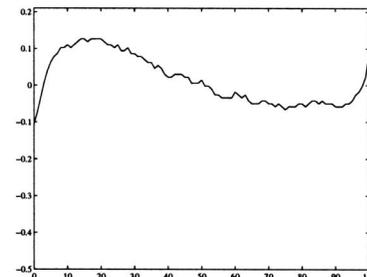
(c) Width curve corresponding to (b).



(d) Line sidewall (250nm, 1:1 pitch).



(e) Profile extracted from (d) with widths indicated.



(f) Width curve corresponding to (e).

Fig. 6 Representation of the line sidewalls: (a) and (d) cross-section images, (b) and (e) extracted profiles, and (c) and (f) normalized width curves of the profiles sampled uniformly over 101 points from top to bottom.

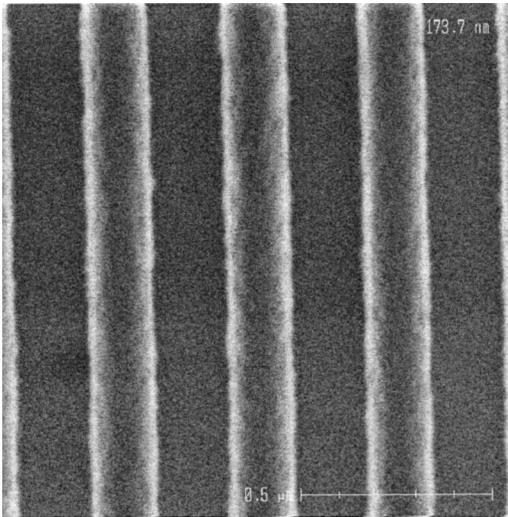


Fig. 7 Example top-down image with three complete lines. This particular image is from a 180-nm (1:1 pitch) design rule.

3.2 Top-Down Subimage Extraction

Since features must be extracted from the top-down images to estimate sidewall shapes; automation of this step is necessary to make the system feasible. Automation of the feature extraction requires rotational alignment of the lines, location of the lines within a top-down image, and extraction of subimages containing the lines. Without loss of generality, we assume henceforth that all lines are vertically oriented. We begin with a top-down image such as that shown in Fig. 7. The first step is to correct for any small angular deviation from vertical. This is accomplished by first rotating the image over a small, fixed set of angles and computing the average column variance at each angle. These variances are then fit to a quadratic and the minimizer of this quadratic is taken as the angle of rotation.

Once minor rotational variation has been corrected, the next step is to locate the lines within the image. A normalized profile of the image is calculated by summing along each column and scaling the resulting curve to lie between 0.0 and 1.0; an example is shown in Fig. 8. From this profile, we first locate peaks by labeling all curve regions above 0.5 as peak regions, dilating those regions to fill any holes, and then labeling as peaks the maximum values in these regions. Peaks near the image edges are discarded. In Fig. 8, the six peaks detected in this manner are indicated with circles. Given the peaks, we next determine which of the between-peak regions, or gaps, represent lines and which represent the space between two lines by applying heuristic rules based on the between-peak average value, the between-peak spacing, and the gradients to the left and right of each peak.

Once the lines have been located in this manner, we extract one or more subimages along each line. Previously,¹ we extracted one or more subimages from each line feature; each such subimage was centered on the line and the subimage size was three times the design rule on each side. For example, for a 100-nm design rule, 300- \times 300-nm subimages were extracted, and for a 250-nm design rule 750- \times 750-nm subimages were extracted. These subimages

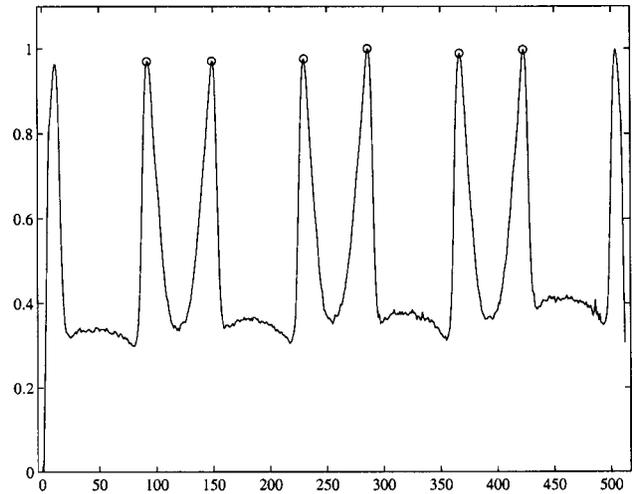


Fig. 8 Normalized horizontal profile from the rotationally corrected version of the image in Fig. 7.

therefore covered the entire linewidth and were the same size relative to the design rule. The motivation for this approach was to enable the use of multiple design rules in the same database since lines from different design rule lines can still have very similar sidewall shapes. The relative (to the design rule) linewidth, however, is still implicitly included in this approach. For example, suppose line 1 and line 2 have the same sidewall shape, but line 1 is a 200-nm-wide line from a 250-nm design rule, and line 2 is a 120-nm-wide line from a 100-nm design rule. In the extracted subimages, line 2 will appear much wider than line 1 relative to the subimage size and this will be reflected in all of the extracted features. Hence, we would not expect the line 1 and line 2 to appear very similar in our historical database even though their corresponding sidewall shapes are very much alike. In this paper, we remove this implicit dependence by extracting line-edge subimages rather than full-line subimages. The process for rotational correction of the entire top-down image and line location within the full-size image is the same as described above and in earlier work.¹ The extracted subimages, however, are centered on line-edges rather than the line center and their size is set to be 0.6×1.8 times the design rule. Subimages are extracted from both right and left line edges and then rotated and/or reflected appropriately so that the line feature is to the right and the gap between lines (substrate) is to the left. Three example line-edge subimages are shown in Fig. 9.

3.3 Feature Computation

For each line-edge subimage extracted, we compute and store a feature vector. To compare structures of different physical dimensions (due to varying design rules), we use features that are invariant to the subimage scale. The first set of features is computed using Gabor filters,^{3,4} which have proven quite useful in analyzing texture-like image properties. We employ a bank of filters that spans six scales and 10 orientations, resulting in a tiling of the discrete-space frequency plane that is illustrated in Fig. 10. The energy of a subimage that is contained in one of these 60 filter bands is used as a single feature. We also use the logarithm of this energy, resulting in a total of 120 Gabor-

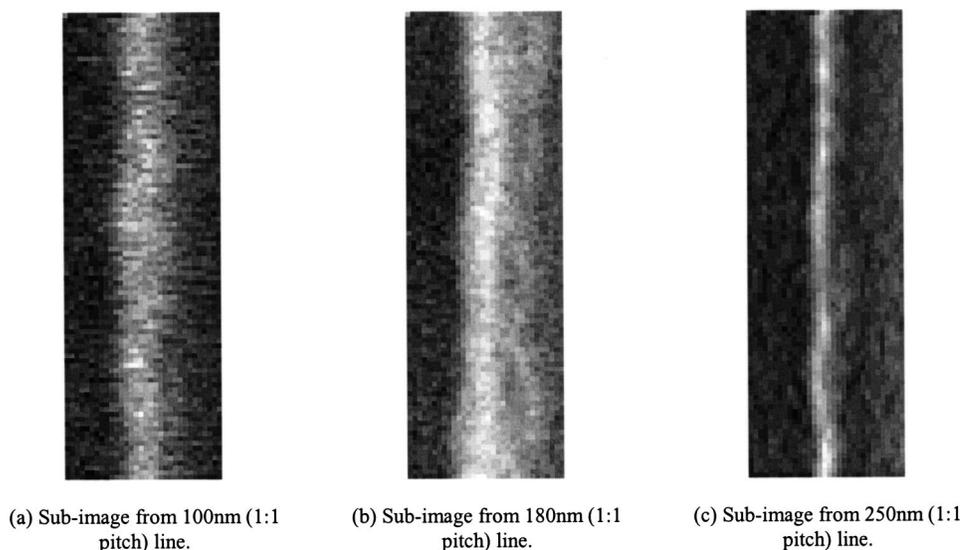


Fig. 9 Examples of line-edge subimages.

based features. We resize each subimage (using bicubic interpolation) to be 32×96 pixels and raster scan this image to add 3072 more features. We use the actual linewidth normalized by the design rule linewidth as the final feature. Although linewidth is not necessarily an indicator of sidewall shape, in some cases actual width and sidewall shape are somewhat correlated. The net result is a 3193-point feature vector for each line-edge subimage. We turn our attention in the next section to finding a weighted subset of these features to make database querying more robust and more computationally friendly.

4 Dimensionality Reduction

The aim of dimensionality reduction is to map feature vectors in a high-dimensional space to some lower dimensional subspace, usually because of computational difficulties re-

lated to the large dimensionality of the original space. After the feature extraction process already described, our feature vectors have $p = 3193$ dimensions. Although it is possible to compare top-down feature vectors in the 3193-D space, it is computationally demanding and unnecessary since there are redundant features as well as features that are not helpful with respect to the sidewall estimation problem. Motivated by techniques that have been successfully applied to template-based face recognition,^{5,6} we adopt an LDA approach for dimensionality reduction. Note that discrimination in our system refers to the ability to differentiate between top-downs associated with different sidewall shapes. To discriminate between different sidewall shapes, however, we must first define groupings of similar sidewalls. We accomplish this by applying the well-known k -means clustering algorithm⁷ to the 101-point normalized sidewall representations defined by Eq. (1). In the current implementation, we employ $C = 20$ clusters. In Fig. 11, we plot the width curves for two example clusters. The cluster numbers (1 to 20) are then used as class labels for the top-down images.

4.1 Direct Weighted LDA

The goal of traditional LDA (T-LDA) is to project high-dimensional feature vectors in \mathcal{R}^n onto a lower dimensional subspace \mathcal{R}^m , where $m < n$, while preserving as much discriminative information as possible. One formal expression for the corresponding optimization criterion can be written as

$$\operatorname{argmax}_{\mathbf{A}} \frac{\operatorname{tr}(\mathbf{A}^T \mathbf{S}_b \mathbf{A})}{\operatorname{tr}(\mathbf{A}^T \mathbf{S}_w \mathbf{A})}, \quad (3)$$

where $\mathbf{A} \in \mathcal{R}^{n \times m}$, $\operatorname{tr}(\cdot)$ is the trace operator, $\mathbf{S}_w \in \mathcal{R}^{n \times n}$ is the within-class scatter matrix, and $\mathbf{S}_b \in \mathcal{R}^{n \times n}$ is the between-class scatter matrix. The within-class scatter matrix is given by

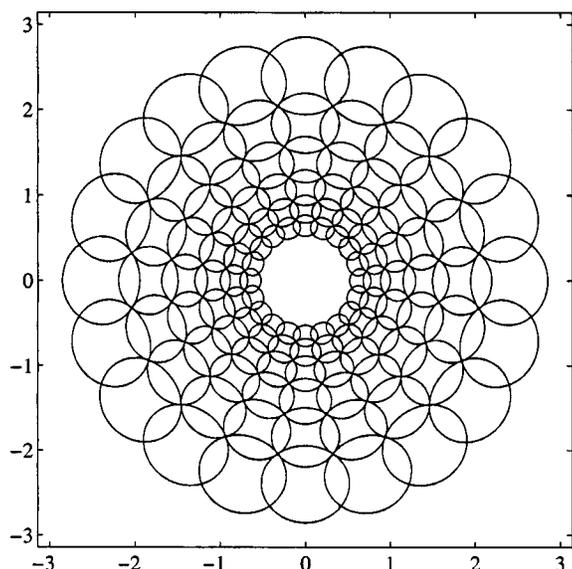


Fig. 10 Tiling of the discrete-space $([-\pi, \pi])$ frequency plane with a six-scale, 10-orientation bank of Gabor filters.

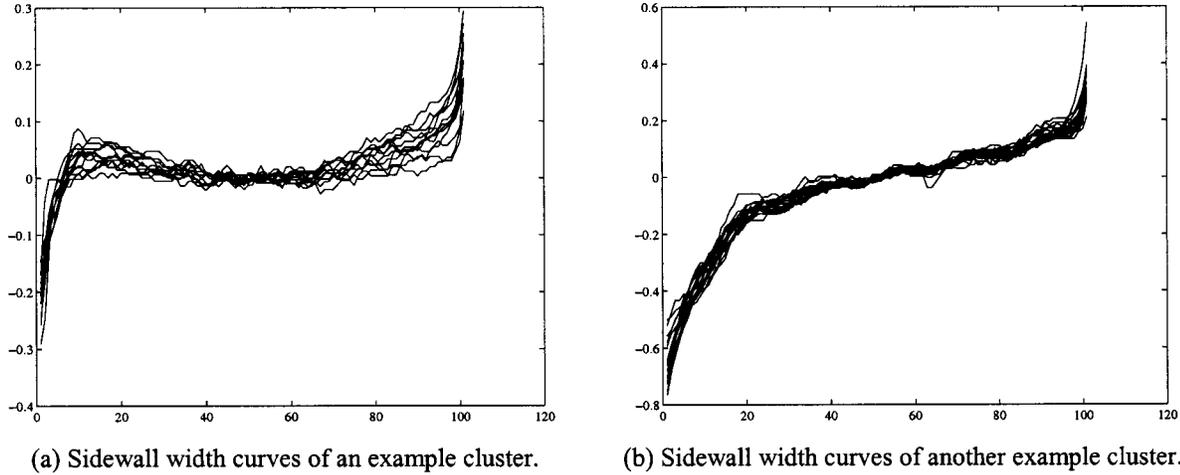


Fig. 11 Clustering of the normalized sidewall width curves.

$$\mathbf{S}_w = \sum_{i=1}^C \sum_{j=1}^{N_i} [\mathbf{x}_j^{(i)} - \mu_i][\mathbf{x}_j^{(i)} - \mu_i]^T, \quad (4)$$

where C is the total number of classes, N_i is the number of samples in class C_i , $\mathbf{x}_j^{(i)} \in \mathcal{R}^n$ is the j 'th vector of C_i , and $\mu_i \in \mathcal{R}^n$ is the mean of C_i . The between-class scatter matrix is given by

$$\mathbf{S}_b = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T, \quad (5)$$

where $\mu \in \mathcal{R}^n$ is the ensemble mean. We note that $\text{rank}(\mathbf{S}_b) \leq C-1$ since it is the sum of C rank one or zero (if $\mu_i = \mu$) matrices, where at most $C-1$ are linearly independent. For convenience, and without loss of generality, we assume that $\text{rank}(\mathbf{S}_b) = C-1$ for the remainder of this paper. The intuitive interpretation of Eq. (3) is that T-LDA attempts to simultaneously minimize the within-class scatter and maximize the between-class scatter. Perhaps the most common approach for solving Eq. (3) is to solve the generalized eigenproblem of \mathbf{S}_b and \mathbf{S}_w . This solution can be achieved⁸ by simultaneously diagonalizing \mathbf{S}_b and \mathbf{S}_w . The simultaneous diagonalization process is accomplished (assuming \mathbf{S}_w is nonsingular) by whitening \mathbf{S}_w , diagonalizing the resulting \mathbf{S}_b , and then taking the eigenvectors of \mathbf{S}_b with the largest eigenvalues. Intuitively, this process can be described as whitening the denominator of Eq. (3) and then maximizing the numerator over a reduced dimensionality. The converse approach of whitening the numerator and minimizing the denominator is equivalent, but recall that \mathbf{S}_b is generally singular and cannot be whitened.

4.1.1 Weighted LDA

The class separability criteria that T-LDA maximizes^{8,9} is the Euclidean distance between the class means. Euclidean distance, of course, is not necessarily representative of classification accuracy, and its use as the separability measure can cause some classes to unnecessarily overlap in the reduced space. One proposed solution for this problem is

known as weighted pairwise Fisher criteria,⁹ which we refer to as the weighted LDA (W-LDA). To begin, we first note an alternate expression⁹ for \mathbf{S}_b :

$$\mathbf{S}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^C \alpha(\Delta_{ij})(\mu_i - \mu_j)(\mu_i - \mu_j)^T, \quad (6)$$

where we have assumed equal class priors, Δ_{ij} is a measure of the separation between classes C_i and C_j , $\alpha(\cdot)$ is a weighting function, and setting $\alpha(\cdot) = 1$ makes Eqs. (5) and (6) equivalent. In W-LDA, the Mahalanobis distance is selected for the class separation measure Δ_{ij} :

$$\Delta_{ij} = [(\mu_i - \mu_j)^T \mathbf{S}_w^{-1} (\mu_i - \mu_j)]^{1/2}, \quad (7)$$

and the weighting function $\alpha(\cdot)$ in Eq. (6) is selected such that the contribution of each pair of classes depends (approximately) on the Bayes error rate between the classes, yielding

$$\alpha(\Delta_{ij}) = \frac{1}{2\Delta_{ij}^2} \text{erf}\left(\frac{\Delta_{ij}}{2\sqrt{2}}\right). \quad (8)$$

4.1.2 Direct LDA

One problem often encountered with LDA in practice is that the original feature vectors may be of such high dimensionality (3193 in our case) that the storage and/or eigenanalysis of \mathbf{S}_b and \mathbf{S}_w may be impractical. In such applications, some other form of dimensionality reduction—usually principal component analysis (PCA) in the face recognition case^{5,6,10}—performed prior to LDA. PCA, however, does not consider class labels and can decrease discriminative capability. Yu and Yang recently proposed¹¹ an LDA algorithm—direct LDA (D-LDA)—that can be directly applied to high-dimensional data.

The critical idea that enables D-LDA is to first project all samples in \mathcal{R}^n onto the $(C-1)$ -dimensional column space of \mathbf{S}_b (i.e., discard the nullspace of \mathbf{S}_b). This is motivated by assuming that directions along which there is no

between-class scatter are not useful for discrimination. Although this assumption is not entirely true, results¹¹ indicate the approach is still effective. In many high-dimensional problems, the number of classes C is much smaller than the dimensionality of the vectors n . Recall that in our case, since we are using 20 sidewall clusters, $C = 20$. Recalling that $\text{rank}(\mathbf{S}_b) = C - 1$, we can reduce the dimensionality of the problem from n to $C - 1$ by projecting onto the column space of \mathbf{S}_b . By discarding the nullspace of \mathbf{S}_b , the between-class scatter matrix in the reduced space is full rank. We can then use the simultaneous diagonalization approach mentioned above, where we whiten the numerator of Eq. (3) and minimize the denominator. This, in fact, permits us to preserve the nullspace of \mathbf{S}_w (if it exists), which, according to other research,^{11,12} contains the most discriminative information.

As stated, the first step in D-LDA is to find a basis for the $(C - 1)$ -dimensional column space of \mathbf{S}_b . Recall that \mathbf{S}_b is an $n \times n$ matrix, which might imply a significant computational burden if n is large. Fortunately, the $C - 1$ eigenvectors of \mathbf{S}_b corresponding to the $C - 1$ nonzero eigenvalues can be found by solving a much more tractable $C \times C$ problem.⁸

4.1.3 Combination of direct and weighted LDA

From the discussion in the previous section, it would seem desirable to exploit the benefits of W-LDA and D-LDA simultaneously. There are, however, a couple of potential issues that must be recognized and overcome. First, we note that the computation of \mathbf{S}_b for W-LDA, as given by Eq. (6), first requires the computation of \mathbf{S}_w , which is a large $n \times n$ matrix (where $n = 3193$ in our case). The matrix \mathbf{S}_w is required since Mahalanobis distance is used for Δ_{ij} and, as shown in Eq. (7), \mathbf{S}_w^{-1} is required in the computation. Noting the need for \mathbf{S}_w^{-1} leads us to another potential difficulty; one of the primary motivations for D-LDA was the preservation of the nullspace of \mathbf{S}_w . If the null space of \mathbf{S}_w is nonempty, then \mathbf{S}_w^{-1} does not exist.

We propose the following approach to address these problems. First, recalling Eq. (6), we make the mild assumption that $\alpha(\Delta_{ij}) > 0$. Note that this assumption implies that no two classes have equal means and that \mathbf{S}_w^{-1} exists (or is replaced with an alternative). In this case, the nullspaces of \mathbf{S}_b from Eqs. (5) and (6) are equivalent. Hence we can remove the nullspace by projecting onto the $(C - 1)$ -dimensional column space of \mathbf{S}_b . Recall that the column space of \mathbf{S}_b can be found by eigenanalysis of a much more tractable $C \times C$ matrix. Once we have projected to the $C - 1$ column space, we compute \mathbf{S}_w in the reduced space and, if it is nonsingular, we simply proceed with W-LDA as already described.

If, however, \mathbf{S}_w is indeed singular in the column space of \mathbf{S}_b , we can use a pseudoinverse. We note, however, that \mathbf{S}_w is generally never singular in the column space of \mathbf{S}_b so long as we have at least two samples in every class (i.e., two top-down subimages associated with each sidewall cluster). This is always the case in our system, hence the projection of \mathbf{S}_w is full rank.

We can now describe the complete DW-LDA algorithm with the following six steps:

1. Let $\mathbf{B} \in \mathcal{R}^{n \times r}$ be an orthonormal basis for the column space of \mathbf{S}_b^0 , the between-class scatter matrix in the original space. Remove the nullspace of the between-class scatter matrix by projecting all samples onto \mathbf{B} : $\mathbf{x} \in \mathcal{R}^n \rightarrow \mathbf{B}^T \mathbf{x} \in \mathcal{R}^r$.
2. In the reduced space \mathcal{R}^r , compute \mathbf{S}_w . If \mathbf{S}_w is full rank, compute \mathbf{S}_w^{-1} ; otherwise compute a pseudoinverse $\hat{\mathbf{S}}_w^{-1}$.
3. Compute \mathbf{S}_b using Eq. (6) with α_{ij} given by Eq. (8) and Δ_{ij} given by Eq. (7). If \mathbf{S}_w is singular, then use a pseudoinverse, $\hat{\mathbf{S}}_w^{-1}$, when computing Δ_{ij} .
4. Whiten \mathbf{S}_b :

$$\mathbf{S}_b \rightarrow \mathbf{W}^T \mathbf{S}_b \mathbf{W} = \mathbf{I}_{r \times r},$$

$$\mathbf{S}_w \rightarrow \tilde{\mathbf{S}}_w = \mathbf{W}^T \mathbf{S}_w \mathbf{W},$$
 where $\mathbf{W} = \Psi \Gamma^{-1/2}$ is the whitening transformation of \mathbf{S}_b with Ψ being the eigenvectors of \mathbf{S}_b and Γ the diagonal eigenvalue matrix.
5. Diagonalize $\tilde{\mathbf{S}}_w$:

$$\tilde{\mathbf{S}}_w \rightarrow \mathbf{D}_w = \mathbf{V}^T \tilde{\mathbf{S}}_w \mathbf{V},$$
 where \mathbf{D}_w is the diagonal eigenvalue matrix of $\hat{\mathbf{S}}$, and \mathbf{V} contains the corresponding orthonormal eigenvectors.
6. Assume that the eigenvalues and eigenvectors of \mathbf{D}_w and \mathbf{V} are sorted in ascending order, possibly with some zeros in \mathbf{D}_w . To maximize the LDA criterion in Eq. (3) while reducing to dimensionality m , take the first m columns of \mathbf{V} , which correspond to the m lowest (some possibly zero) eigenvalues. The overall resulting transformation matrix $\mathbf{A} \in \mathcal{R}^{n \times m}$ can then be written as

$$\mathbf{A} = \mathbf{B} \mathbf{W} \mathbf{V} \begin{pmatrix} \mathbf{I}_{m \times m} \\ \mathbf{0}_{(n-m) \times m} \end{pmatrix}. \quad (9)$$

5 Experimental Results

In this section, we report results obtained using the proposed system on real semiconductor data where different sidewall shapes were produced by varying the focus and exposure (producing the F/E matrices we discussed previously) of the lithographic tool. The available data set tested comprised five design rules, described as follows, with top-down images captured by one or more of three different CD-SEM tools:

1. 100-nm dense (2:1 pitch) lines, 47 cross sections with 126 top-downs
2. 100-nm isolated (5:1 pitch) lines, 94 cross sections with 269 top-downs
3. 180-nm dense (1:1 pitch) lines, 70 cross sections with 201 top-downs
4. 180-nm isolated (5:1 pitch) lines, 88 cross sections with 263 top-downs

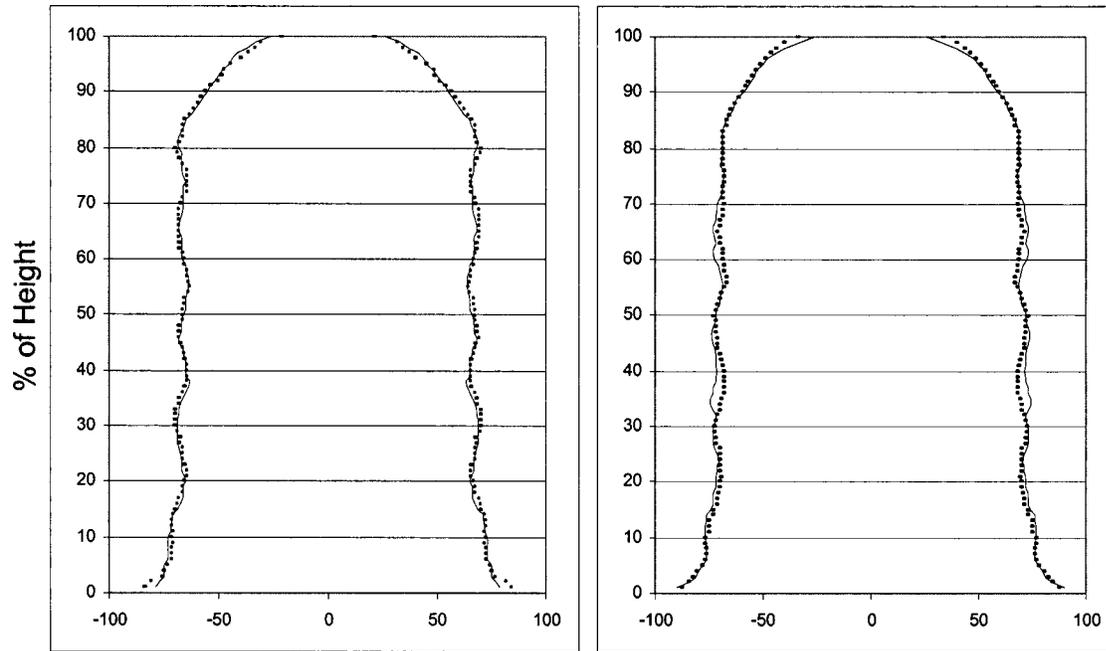


Fig. 12 Examples of sidewall estimates (dotted lines) versus true sidewall shape (solid lines).

5. 250-nm dense (1:1 pitch) lines, 113 cross sections with 113 top-downs

Hence, the complete set of available data comprised 412 sidewalls and 972 top-downs (complete top-down images, not subimages). Having a data set with variations in line widths, line spacing, and imaging tool (CD-SEM) response enables us to test the application over a wide range of circumstances. From the 972 top-down images, we extracted 9718 subimages (~ 10 per full-size top-down) according to the process of Sec. 3.

Hold-one-out type tests were performed by removing a single sidewall and all corresponding top-downs from the training data when computing the transformation matrix for dimensionality reduction, as described in Sec. 4. Each of these hold-out top-downs was then submitted as a query. The corresponding sidewall shape was estimated via weighted averaging (described in the following) and compared to the true sidewall shape. This process was repeated for each of the 412 available cross sections, corresponding to 972 different top-down queries. For comparison to the newly proposed DW-LDA approach for dimensionality reduction, we also tested D-LDA, PCA plus T-LDA (this is the same method from our earlier work¹), and PCA plus W-LDA.

Weighted averaging was employed to estimate the query sidewall shape, where the weighting is determined by the distances between the query and the K nearest top-downs from the historical database (where various values of K were tested). The distance between a full query top-down and a full historical top-down is defined by the closest pair of subimage feature vectors. In other words, let Q represent the full top-down query image with $q = 1, \dots, S_Q$ subimages, and let H with $h = 1, \dots, S_H$ subimages be a top-down in the historical (training) database. The distance between Q and H , $D(Q, H)$, is then defined as

$$D(Q, H) = \min_{\substack{q=1, \dots, S_Q \\ h=1, \dots, S_H}} d(\mathbf{z}_q, \mathbf{z}_h), \quad (10)$$

where \mathbf{z}_q and \mathbf{z}_h represent the subimage feature vectors, computed according to Secs. 3 and 4, for subimage q of full top-down Q and sub-image h of full top-down H , respectively. For the reported experiments, Euclidean distance was used for the distance measure $d(\cdot)$. For a given query image Q , $D(Q, H)$ was computed for every top-down H in the training set and sorted in ascending order. The sidewall width curves corresponding to the closest K historical top-downs were used to estimate the query sidewall shape, $\hat{\mathbf{c}}(Q)$ (as a vector), as follows:

$$\hat{\mathbf{c}}(Q) = \sum_{i=1}^K \alpha_i \mathbf{c}(H_i), \quad (11)$$

where $\mathbf{c}(H_i)$ is the sidewall of nearest-neighbor i and the weighting factors are given by

$$\alpha_i = \left[\sum_{j=1}^K \frac{1}{D(Q, H_j)} \right]^{-1} \frac{1}{D(Q, H_i)}, \quad (12)$$

so that $\sum_i \alpha_i = 1$. The number of nearest neighbors used in the tests was allowed to take on values $K = 1, \dots, 50$. Figure 12 shows two examples of sidewall estimates (dotted lines) and the true sidewall shape (solid lines) corresponding to the submitted top-down image. In this figure, the horizontal axis units are in nanometers and the vertical scale is percentage of line height. Notice the estimates not only provide a good approximation to the top roundness, but the roughness of the sides and the footer shape are also closely estimated. While these figures show the capabilities of this

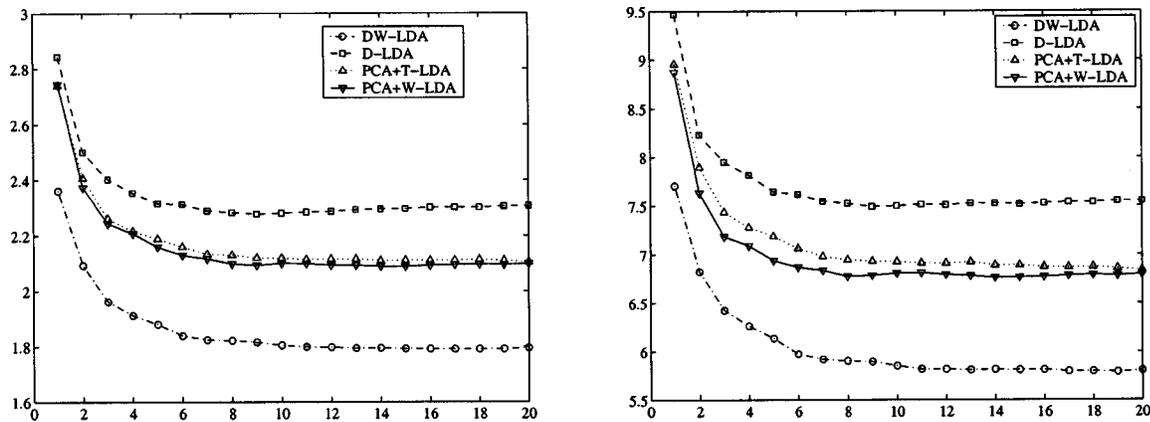


Fig. 13 Average root mean square (left) and average maximum absolute (right) error over all hold-out data (as a percentage of design rule linewidth) versus number of nearest neighbors used.

technique, measures are necessary to quantify the performance of the sidewall estimation technique over the entire set of test data. Root mean square and maximum absolute errors were chosen as the performance measures.

We computed the root mean square and maximum absolute errors (normalized by the design rule linewidth) for the estimated sidewall shape of every top-down hold-out using the four different dimensionality reduction methods mentioned (DW-LDA, D-LDA, PCA+T-LDA, PCA+W-LDA). The average of these errors across all top-down queries is plotted against the number of nearest neighbors in Fig. 13, where the vertical axis represents the error divided by the design rule linewidth as a percentage (i.e., an error of 5% for a 100-nm design rule implies a 5-nm error in width). We can note from the plots in Fig. 13 that DW-LDA performs better than all of the other approaches. Interestingly, D-LDA actually performs worse than all of the other techniques, including PCA plus T-LDA; this is contrary to some results that have been previously reported.¹¹ We believe this is because the D-LDA may tend to preserve noisy features that seem to be discriminative in the training data, but that do not generalize well to the testing data. The PCA first approach, however, would minimize the impact of such noisy features.

Figure 14 shows the error distributions (using all 972 hold-outs) at various positions along the vertical extent of the sidewall using $K = 15$ nearest neighbors with DW-LDA. Note that the largest errors are seen near the top (95%) and bottom (5%) of the sidewall, but that the overwhelming majority of these errors are still within $\pm 10\%$.

Finally, we note that the errors reported here are, in fact, about the same or slightly higher than those reported in the earlier approach.¹ For example, as seen in Fig. 13(a), DW-LDA achieves an average root mean square (rms) error of about 1.8%, while in our previous paper we reported an average rms error of about 1.7%. Similarly, here we achieve an average maximum absolute (MA) error of about 5.7%, while we previously reported an average MA error of about 5.5%. We hypothesize two possible explanations for this. First, since our extracted subimages are much smaller here, we actually used less total area of each top-down image to make the experiments easily implemented with current computational resources. In the previous effort, we

had available and used 6629 full-line subimages (each three times the linewidth design rule square); here we had available 47,140 line-edge subimages but only used 9718. Hence our training data was effectively less diverse. We are currently adapting the existing code to make better use of this larger training set. The second hypothesis is that the actual linewidth is more correlated with the sidewall shape (at least in our currently available data) than we suspected. To test this hypothesis, we intend to append some full-line features to the line-edge feature vectors in future work.

6 Conclusions

We presented an image retrieval system for estimating semiconductor sidewall shapes from top-down scanning electron microscopy images. Since this system is trained with historical data in the form of top-down/cross-section image pairs that are already collected in F/E matrices as a part of process setup, and the system requires no new hardware using top-down SEM images already being taken by CD-SEM tools for process control, the system can be easily introduced to the semiconductor fab to estimate sidewall

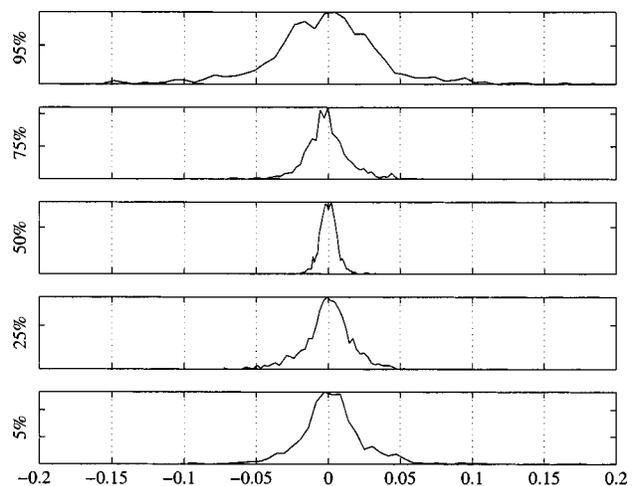


Fig. 14 Sidewall shape error distributions (as a fraction of design rule linewidth) at various points along the sidewall height (top of line structure is 100%, bottom is 0%).

shape. Use of this system would only require the additional effort of building the historical database, which would be done as a part of process setup.

We presented details on the feature extraction portions of this system and the dimensionality reduction methods applied to the problem. We present a new approach for dimensionality reduction called DW-LDA. Experimental results indicate that the proposed system can estimate sidewall shape quite accurately and that DW-LDA is the best of several other linear dimensionality reduction techniques. Hold-one-out testing has shown results of approximately 1.8% rms error between the estimated sidewall shape and the actual sidewall shape. Thus for a 100-nm line, the rms error is 1.8 nm.

Future extensions of this research should include investigation of other discriminating features that may be extracted from the top-down images to improve performance and investigation of whether the training set needs to be rebuilt for various process changes (line width, CD-SEM tool changes, material changes, etc.).

Acknowledgments

This manuscript was prepared by the Oak Ridge National Laboratory, Tennessee, operated by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Funding for this effort was provided by International SEMATECH (ISMT). We would like to thank the member companies of ISMT for their support and guidance. We would like to specifically thank Michael Bishop, Marylyn Bennett, and Hal Bogardus of ISMT.

References

1. J. R. Price, P. R. Bingham, K. W. Tobin, and T. P. Karnowski, "Estimating cross-section semiconductor structure by comparing top-down SEM images," in *Machine Vision in Industrial Inspection XI, Proc. SPIE* **5011**, 161–170 (2003).
2. P. R. Bingham, J. R. Price, K. W. Tobin, T. P. Karnowski, M. Bennett, and H. Bogardus, "Sidewall structure estimation from CD-SEM for lithographic process control," in *Process and Materials Characterization and Diagnostics in IC Manufacturing, Proc. SPIE* **5041**, 115–126 (2003).
3. S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Trans. Image Process.* **11**, 1160–1167 (Oct. 2002).
4. D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**, 2379–2394 (Dec. 1987).
5. J. R. Price and T. F. Gee, "Towards robust face recognition from video," in *Proc. 30th Applied Imagery and Pattern Recognition Workshop (AIPR)*, pp. 94–100 (Oct. 2001).
6. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 711–720 (July 1997).
7. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley-Interscience, New York (2001).
8. K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, New York (1990).
9. M. Loog, R. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise Fisher criteria," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 762–766 (July 2001).
10. W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 336–341 (1998).
11. H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—with application to face recognition," *Pattern Recogn.* **34**, 2067–2070 (Oct. 2000).
12. L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recogn.* **10**, 1713–1726 (Oct. 2000).



Philip R. Bingham is a research and development engineer in the Image Science and Machine Vision (ISMV) group within the Engineering Science and Technology Division of Oak Ridge National Laboratory (ORNL). He earned his BS degree in electrical and computer engineering from the University of Tennessee, Knoxville, and his MS and PhD degrees in electrical and computer engineering from the Georgia Institute of Technology. His doctoral work involved modeling switching mechanisms for parallel computing based on load distributions. Since joining ORNL in December 1999, he has been heavily involved in research on the application of technology to improve yield in the semiconductor industry.



Jeffery R. Price received his BSEE degree from the U.S. Naval Academy in 1993 and his MS and PhD degrees in electrical engineering from the Georgia Institute of Technology in 1997 and 1999, respectively. While at Georgia Tech, he received the Kodak Fellowship. He has been with the Imaging Science Division at Kodak Research Laboratories and the Imaging, Robotics, and Intelligent Systems Laboratory at the University of Tennessee. Dr. Price is currently with the Image Science and Machine Vision Group at Oak Ridge National Laboratory where some of his recent work has included applied computer vision and pattern recognition research in the fields of industrial inspection and defect detection, biomedical imaging, and face recognition.



Kenneth W. Tobin is a corporate research fellow and group leader of the Image Science and Machine Vision Group at the Oak Ridge National Laboratory, Tennessee. The group performs applied computer vision research and development in industrial inspection and metrology, biomedical imaging, and national security. He performs research in nondestructive testing and analysis, image processing, and image-based metrology for automation and process characterization. He has authored and coauthored over 120 publications and he currently holds six U.S. patents with five additional patents pending in the areas of computer vision, photonics, radiography, and microscopy. Dr. Tobin is a fellow of SPIE and a member of the IEEE. He received his PhD degree in nuclear engineering from the University of Virginia, Charlottesville, and his MS degree in nuclear engineering and his BS degree in physics from the Virginia Polytechnic Institute and State University, Blacksburg.



Thomas P. Karnowski is a development associate and research scientist with the Image Science and Machine Vision Group in the Engineering Science and Technology Division of Oak Ridge National Laboratory. Mr. Karnowski has over 12 years experience solving problems in systems integration, software design and development, and hardware development. His work has primarily focused on imaging and signal processing problems in industrial and military environments. Although he is an innovative thinker who always seeks to broaden his technical expertise, he is also very results oriented in his problem solutions. He has authored or coauthored over 15 publications in the areas of digital signal and image processing and pattern recognition applied to industrial environments. He is a coinventor on three patents in the areas of industrial image processing and pattern recognition for solving semiconductor yield problems. Mr. Karnowski received his MS degree in electrical engineering from North Carolina State University and his BS degree in electrical engineering from the University of Tennessee.