

Semiconductor sidewall shape estimation using top-down CD-SEM image retrieval*

Jeffery R. Price[†], Philip R. Bingham, Kenneth W. Tobin, Jr., and Thomas P. Karnowski

Oak Ridge National Laboratory, Oak Ridge, TN, USA

ABSTRACT

We present modifications to a feature-based, image-retrieval approach for estimating semiconductor sidewall (cross-section) shapes using top-down images. The top-down images are acquired by a critical dimension scanning electron microscope (CD-SEM). The proposed system is based upon earlier work with several modifications. First, we use only line-edge, as opposed to full-line, sub-images from the top-down images. Secondly, Gabor filter features are introduced to replace some of the previously computed features. Finally, a new dimensionality reduction algorithm – direct, weighted linear discriminant analysis (DW-LDA) – is developed to replace the previous two-step principal component analysis plus LDA method. Results of the modified system are presented for data collected across several line widths, line spacings, and CD-SEM tools.

Keywords: CD-SEM metrology, semiconductor inspection, lithography, linear discriminant analysis (LDA)

1. BACKGROUND

In current fabrication environments, line-width measurements in semiconductor lithography are made almost exclusively using scanning electron microscope (SEM) images. This process – known as critical dimension SEM (CD-SEM) metrology – employs images that are usually acquired in a top-down configuration, i.e., looking down onto the semiconductor line feature. According to the International Technology Roadmap for Semiconductors, continually shrinking line-widths make it increasingly important to know the sidewall shape (e.g., the cross-section profile) of the lines rather than just their width. For example, two pairs of top-down and corresponding sidewall images are shown in Fig. 1. Although CD-SEM may report the same line-widths for the top-downs in (a) and (b), it is evident in (c) and (d) that their sidewall structure is quite different. In fact, the sidewall shape in (d), which is “overcut,” is unacceptable and tends to lead to device defects. To acquire sidewall images like those of Fig. 1(c) or (d), however, the semiconductor device must be physically cleaved; this is a destructive and time consuming process that hampers throughput and/or sampling. Since top-down images can be acquired much more efficiently, we have investigated^{1,2} the possibility of estimating sidewall shape using only features extracted from top-down images and a database of corresponding historical top-down and sidewall images.

In this paper and in our earlier work,^{1,2} 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. We first construct a historical repository of corresponding top-down and cross-section image pairs. Sidewall profiles are extracted from each cross-section image and stored. Features are computed from one or more sub-images 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 present 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 sub-images covered an entire line, like those of Fig. 1(a) and (b), but here we use only the line-edge sub-images. Secondly, we employ Gabor filter³ responses for some features in hopes of capturing two-dimensional, texture-like characteristics of the line-edge sub-images. Finally, we apply a new LDA variant for dimensionality reduction.

The remainder of this paper is organized as follows. In Section 2, we describe the parameterization of the sidewall profiles, the automatic identification of line regions and sub-images in the top-down images, and the computation of

*Prepared by Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

[†]Correspondence: J. Price, pricejr@ornl.gov, (865) 574-5743.

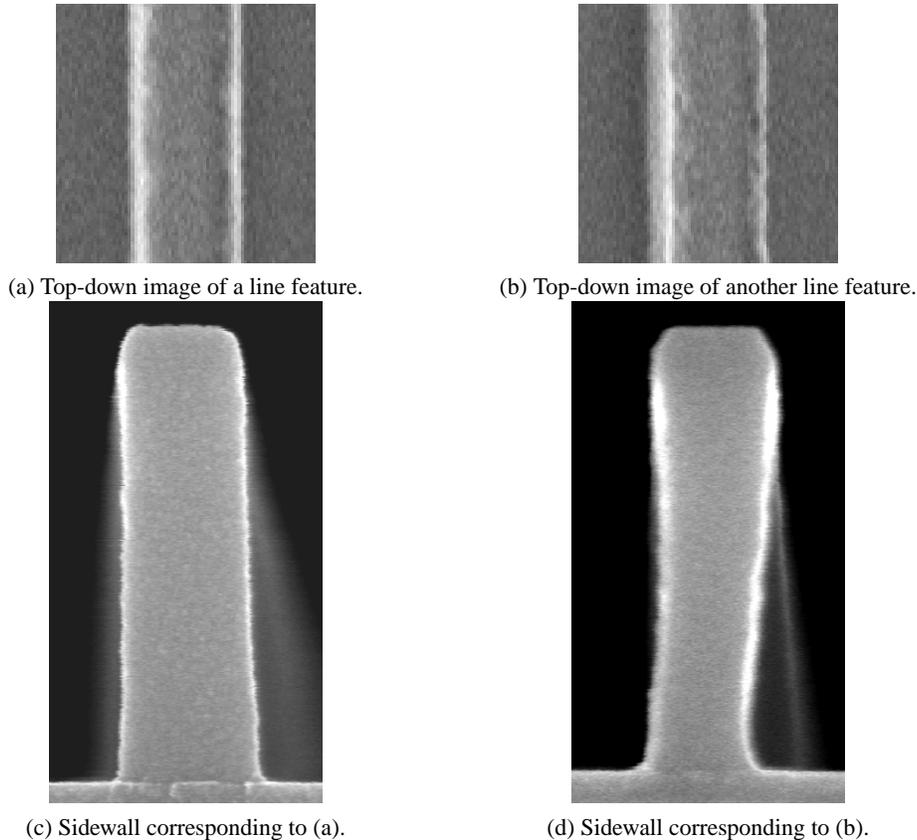


Figure 1. Two top-down and sidewall image pairs from 250nm data. Traditional CD-SEM metrology may report the same line-width for both lines, even though it is clear from (c) and (d) that the sidewall shapes are quite different.

features from these sub-images. We then describe a new approach for dimensionality reduction, which we call direct, weighted LDA (DW-LDA), in Section 3. We present and discuss some experimental results in Section 4 and finally conclude in Section 5 with some summary remarks.

2. FEATURE EXTRACTION

The purpose of feature extraction in this work is to compute features from the top-down images that somehow correlate well with variations in sidewall shape. Towards this purpose, we must define a representation scheme for the sidewall shape – which is the subject of Section 2.1 – so that sidewall variation can be quantified. We then describe the process of finding line features and extracting sub-images from the top-down images in Section 2.2. In Section 2.2, we describe the features that are computed and stored for each sub-image.

2.1. Sidewall Representation

To estimate the sidewall shape, we must first construct a quantifiable representation of it. One goal for the finished system is the capability to estimate sidewall shapes across various design rules – i.e., different line-widths, pitches, and aspect ratios (ratio of line height to width). Towards this purpose, we seek a representation that is invariant to the above mentioned design rule parameters. We therefore define the sidewall shape as the normalized width at 100 equally spaced points from top-to-bottom. These widths (at a few locations) are illustrated by the horizontal dashed lines in Fig. 2(b) and (e), where the sidewall profiles (the outlines of the line structure) are extracted through a semi-automated graphical user interface.² Letting the widths (in nm) be represented by the 101-point vector \mathbf{w} , where $w_n = w(n)$ for $n = 0, \dots, 100$ ($n = 0$ is

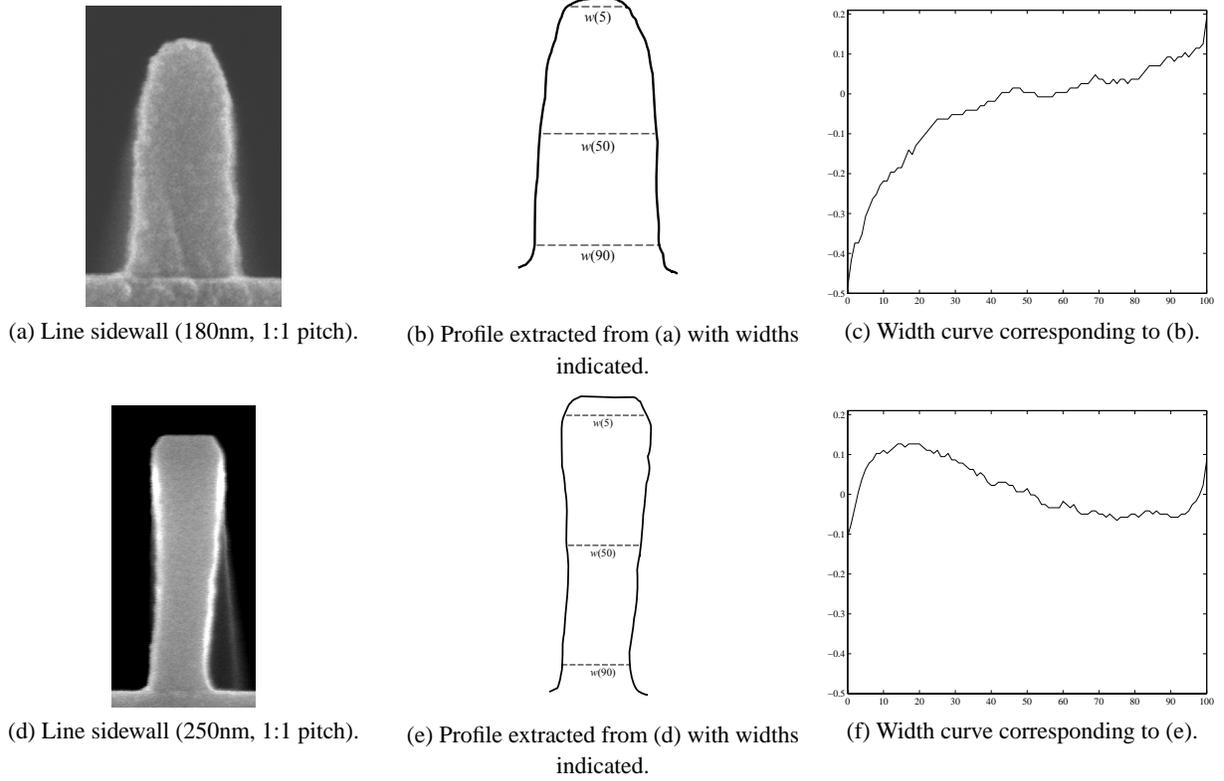


Figure 2. Representation of the line sidewalls. Figures (a) and (d) show sidewall images, (b) and (e) show extracted profiles, and (c) and (f) show the normalized width curves of those profiles sampled uniformly over 101 points from top to bottom.

the top) and letting the design rule (i.e., target) line-width (in nm) be represented by L , the normalized sidewall shape representation (as a vector) is given by

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

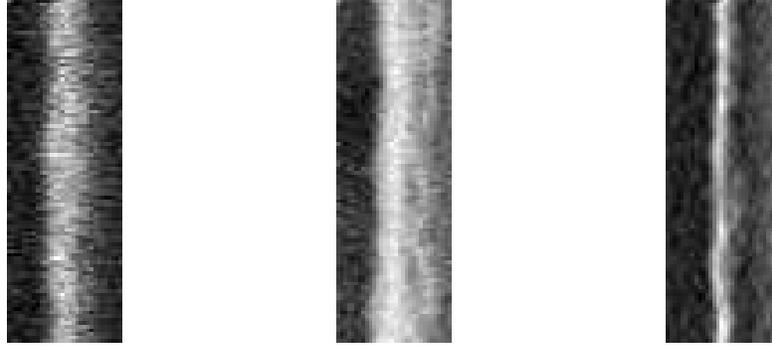
where \bar{w} is the approximate width at the sidewall vertical midpoint as given by

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

Examples of the resulting sidewall width curves are shown in Fig. 2 (c) and (f).

2.2. Top-down Sub-image Extraction

Previously,¹ we extracted one or more sub-images from each line feature; each such sub-image was centered on the line and the sub-image size was three times the design rule on each side. For example, for a 100nm design rule, 300nm \times 300nm sub-images were extracted, and for a 250nm design rule, 750 \times 750nm sub-images were extracted. These sub-images therefore covered the entire line width and were the same size relative to the design rule. The motivation for this approach was to permit 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) line width, 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 200nm-wide line from 250nm design rule and Line 2 is 120nm-wide line from a 100nm design rule. In the extracted sub-images, Line 2 will appear much wider than Line 1 relative to the sub-image 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



(a) Sub-image from 100nm (1:1 pitch) line. (b) Sub-image from 180nm (1:1 pitch) line. (c) Sub-image from 250nm (1:1 pitch) line.

Figure 3. Some example line-edge sub-images.

corresponding sidewall shapes are very much alike. In this work, we remove this implicit dependence by extracting line-edge sub-images rather than full-line sub-images. The process for rotational correction of the entire top-down image and line location within the full-size image is the same as described in earlier work.¹ The extracted sub-images, 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. Sub-images 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 sub-images are shown in Fig. 3

2.3. Feature Computation

For each line-edge sub-image 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 sub-image 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 6 scales and 10 orientations, resulting in a tiling of the discrete-space frequency plane that is illustrated in Fig. 4. The energy of a sub-image 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-based features. We resize each sub-image (using bicubic interpolation) to be 32×96 pixels and raster scan this image to add 3072 more features. We use the actual line width normalized by the design rule line width as the final feature. Although line width is not necessarily an indicator of sidewall shape (recall Fig. 1 for example), in some cases actual width and sidewall shape are somewhat correlated. The net result is a 3193-point feature vector for each line-edge sub-image. We turn our attention in the next section to finding a weighted subset of these features to make database querying more computationally friendly and more robust.

3. 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 related to the large dimensionality of the original space. After the feature extraction process described above, our feature vectors have $p = 3193$ dimensions. Although it is possible to compare top-down feature vectors in this 3193-dimensional 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 a linear discriminant analysis (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. 5 we plot the width curves for two example clusters. The cluster numbers (1-20) are then used as class labels for the top-down images.

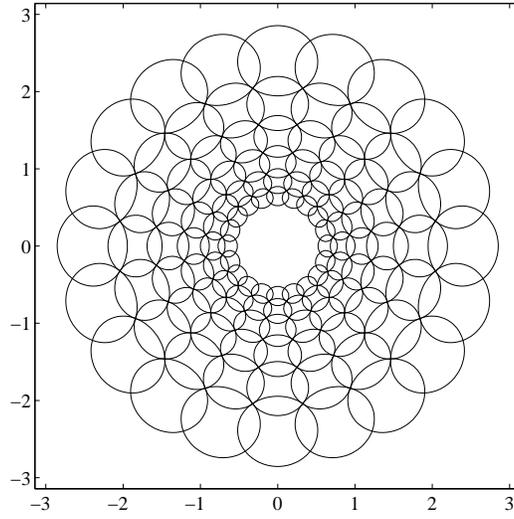


Figure 4. Tiling of the discrete-space $([-\pi, \pi])$ frequency plane with 6-scale, 10-orientation bank of Gabor filters.

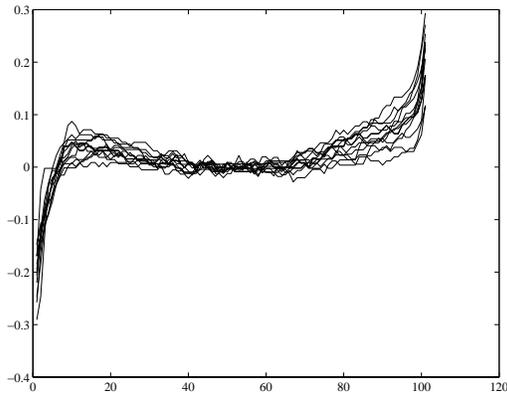
The goal of traditional LDA (T-LDA) is to project high-dimensional feature vectors in \mathbb{R}^n onto a lower-dimensional subspace \mathbb{R}^m , where $m < n$, while preserving as much discriminative information as possible. One formal expression for the corresponding optimization criterion can be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

$$\arg \max_{\mathbf{A}} \frac{\text{tr}(\mathbf{A}^T \mathbf{S}_b \mathbf{A})}{\text{tr}(\mathbf{A}^T \mathbf{S}_w \mathbf{A})}, \quad (3)$$

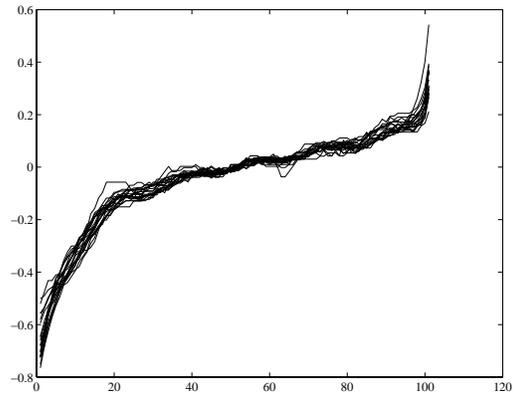
where $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\text{tr}(\cdot)$ is the trace operator, $\mathbf{S}_w \in \mathbb{R}^{n \times n}$ is the *within-class* scatter matrix, and $\mathbf{S}_b \in \mathbb{R}^{n \times n}$ is the *between-class* scatter matrix. The within-class scatter matrix is given by

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

where C is the total number of classes, N_i is the number of samples in class C_i , $\mathbf{x}_j^{(i)} \in \mathbb{R}^n$ is the j^{th} vector of C_i , and



(a) Sidewall width curves of an example cluster.



(b) Sidewall width curves of another example cluster.

Figure 5. Clustering of the normalized sidewall width curves.

$\mu_i \in \mathbb{R}^n$ is the mean of C_i . The between-class scatter matrix is given by

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

where $\mu \in \mathbb{R}^n$ is the ensemble mean. We note that $\text{rank}(S_b) \leq C - 1$ 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 $\text{rank}(S_b) = C - 1$ 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 eigen-problem of S_b and S_w . This solution can be achieved by simultaneously diagonalizing S_w and S_b .⁸ The simultaneous diagonalization process is accomplished (assuming S_w is non-singular) by whitening S_w , diagonalizing the resulting S_b , and then taking the largest eigenvalue eigenvectors of S_b . 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 S_b is generally singular and cannot be whitened.

3.1. Weighted LDA (W-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 *weighted* LDA or W-LDA. To begin, we first note an alternate expression for S_b :⁹

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

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 some weighting function, and setting $\alpha(\cdot) = 1$ makes Eqs. (6) and (5) equivalent. In W-LDA, the Mahalanobis distance is selected for the class separation measure Δ_{ij} :

$$\Delta_{ij} = \sqrt{(\mu_i - \mu_j)^T S_w^{-1} (\mu_i - \mu_j)}, \quad (7)$$

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

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

3.2. Direct LDA (D-LDA)

One problem often encountered with LDA in practice is that the original feature vectors may be of such high dimensionality (they are of dimension 3193 in our case) that the storage and/or eigen-analysis of S_b and 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} – is 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 or 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 \mathbb{R}^n onto the $C - 1$ dimensional column-space of S_b (i.e., discard the nullspace of 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, we have $C = 20$. Recalling that $\text{rank}(S_b) = C - 1$, we can reduce the dimensionality of the problem from n to $C - 1$ by projecting onto the column-space of S_b . By discarding the nullspace of S_b , the between-class scatter matrix in the reduced space is full rank. We may 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 S_w (if it exists), which, according to other research,^{11,12} contains the most discriminative information.

As stated above, the first step in D-LDA is to find a basis for the $C - 1$ dimensional column-space of S_b . Recall that 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 S_b corresponding to the $C - 1$ nonzero eigenvalues can be found by solving a much more tractable $C \times C$ problem.⁸

3.3. Combining 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 S_b for W-LDA, as given by Eq. (6), first requires the computation of S_w , which is a large $n \times n$ matrix (where $n = 3193$ in our case). The matrix S_w is required since Mahalanobis distance is used for Δ_{ij} and, as shown in Eq. (7), S_w^{-1} is needed in the computation. Noting the need for S_w^{-1} leads us to another potential difficulty; one of the primary motivations for D-LDA was the preservation of the nullspace of S_w . If the nullspace of S_w is non-empty, then 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 S_w^{-1} exists (or is replaced with an alternative). In this case, the nullspaces of S_b from Eqs. (6) and (5) are equivalent. Hence we can remove the nullspace by projecting onto the $C - 1$ dimensional column-space of S_b . Recall that the column-space of S_b can be found by eigen-analysis of much more tractable $C \times C$ matrix. Once we've projected to the $C - 1$ column-space, we compute S_w in the reduced space and, if it is non-singular, we simply proceed with W-LDA as described above.

If, however, S_w is indeed singular in the column-space of S_b , we can use a pseudoinverse. We note, however, that S_w is generally never singular in the column-space of S_b so long as we have at least two samples in every class (i.e., 2 top-down sub-images associated with each sidewall cluster). This is always the case in our system, hence the projection of S_w is full rank.

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

1. Let $B \in \mathbb{R}^{n \times r}$ be an orthonormal basis for the column-space of S_b^o , the between-class scatter matrix in the original space. Remove the nullspace of the between-class scatter matrix by projecting all samples onto B .

$$\mathbf{x} \in \mathbb{R}^n \rightarrow B^T \mathbf{x} \in \mathbb{R}^r$$

2. In the reduced space \mathbb{R}^r , compute S_w . If S_w is full-rank, compute S_w^{-1} ; otherwise compute a pseudoinverse, \hat{S}_w^{-1} .
3. Compute S_b using Eq.(6) with α_{ij} given by Eq. (8) and Δ_{ij} given by Eq. (7). If S_w is singular, then use a pseudoinverse, \hat{S}_w^{-1} , when computing Δ_{ij} .
4. Whiten S_b :

$$\begin{aligned} S_b &\rightarrow W^T S_b W = I_{r \times r}, \\ S_w &\rightarrow \tilde{S}_w = W^T S_w W, \end{aligned}$$

where $W = \Psi \Gamma^{-\frac{1}{2}}$ is the whitening transformation of S_b with Ψ being the eigenvectors of S_b and Γ the diagonal eigenvalue matrix.

5. Diagonalize \tilde{S}_w :

$$\tilde{S}_w \rightarrow D_w = V^T \tilde{S}_w V,$$

where D_w is the diagonal eigenvalue matrix of \hat{S} and V contains the corresponding orthonormal eigenvectors.

6. Assume that the eigenvalues and eigenvectors of D_w and V are sorted in ascending order, possibly with some zeros in D_w . To maximize the LDA criterion in Eq. (3) while reducing to dimensionality m , take the first m columns of V which correspond to the m lowest (some possibly zero) eigenvalues. The overall resulting transformation matrix $A \in \mathbb{R}^{n \times m}$ can then be written:

$$A = BWV \begin{pmatrix} I_{m \times m} \\ 0_{(n-m) \times m} \end{pmatrix}. \quad (9)$$

4. 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 a so-called focus/exposure or F/E matrix) 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:

- 100nm dense (2:1 pitch) lines, 47 cross-sections with 126 top-downs;
- 100nm isolated (5:1 pitch) lines, 94 cross-sections with 269 top-downs;
- 180nm dense (1:1 pitch) lines, 70 cross-sections with 201 top-downs;
- 180nm isolated (5:1 pitch), 88 cross-sections with 263 top-downs; and
- 250nm 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 sub-images). From these 972 top-down images, we extracted 9718 sub-images (approximately 10 per full-size top-down) according to the process of Section 2.

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 Section 3. Each of these hold-out top-downs was then submitted as a query. The corresponding sidewall shape was estimated via weighted averaging (described below) 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 neighbor, historical top-downs (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 sub-image feature vectors. In other words, let Q represent the full top-down query image with $q = 1, \dots, S_Q$ sub-images and let H with $h = 1, \dots, S_H$ sub-images be a top-down in the historical (training) database. The distance between Q and H , $D(Q, H)$, is then defined to be

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

where \mathbf{z}_q and \mathbf{z}_h represent the sub-image feature vectors, computing according to Sections 2 and 3, for sub-image 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, $\hat{c}(Q)$ (as a vector), as follows:

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

where $c(H_i)$ 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 $K = 1, \dots, 50$.

We computed the root-mean-square and maximum absolute errors (normalized by the design rule line-width) for the estimated sidewall shape of every top-down hold-out using the four different dimensionality reduction methods mentioned above (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. 6, where the vertical axis represents the error divided by the design rule line-width (i.e., an error of 0.05 for a 100nm design rule implies a 5nm error). We can note from the plots in Fig. 6 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.

In Fig. 7, we show 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 with $\pm 10\%$.

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. 6(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 sub-images are much smaller here, we actually used less total area of each top-down sub-image in order to make the experiments easily implementable with current computational resources. In the previous effort, we had available and used 6629 full-line, sub-images (each three times the line-width design rule square); here we had available 47140 line-edge, sub-images 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 line width is more correlated with 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.

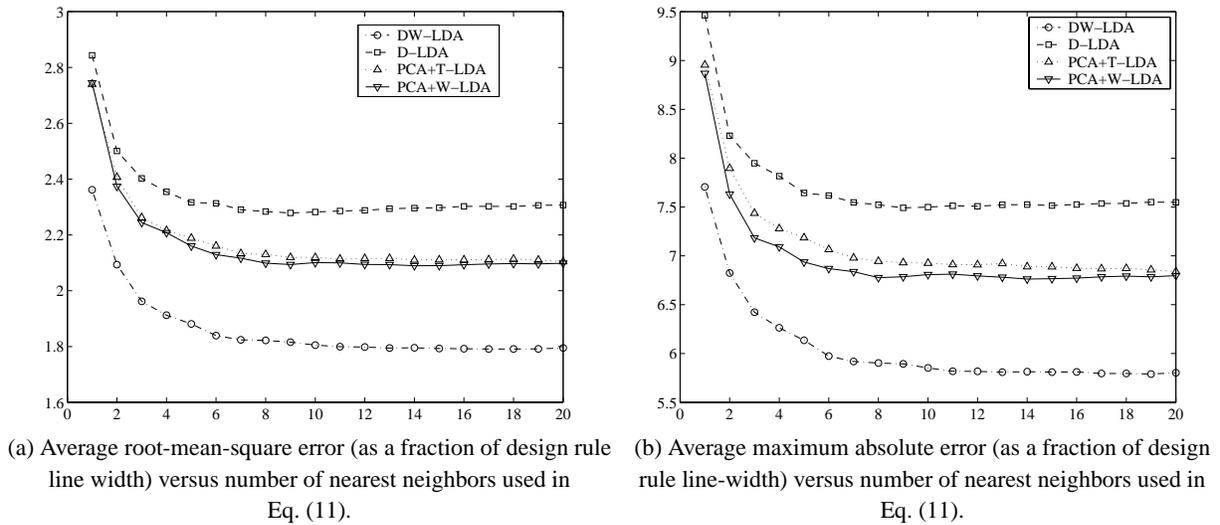


Figure 6. Average root-mean-square (a) and average maximum absolute (b) error over all hold-out data.

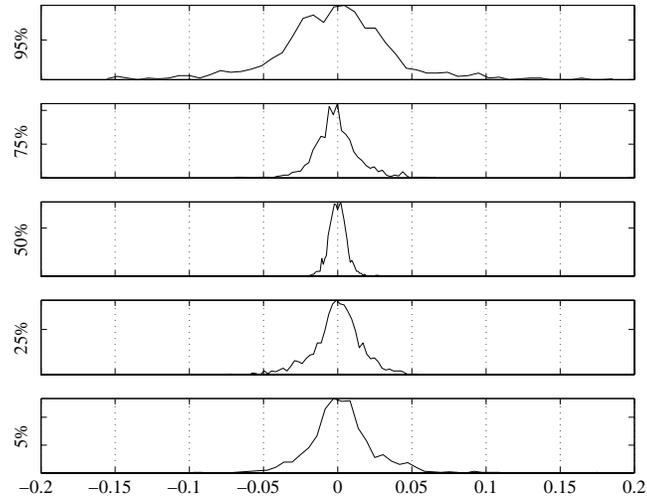


Figure 7. Sidewall shape error distributions (as a fraction of design rule line-width) at various points along the sidewall height (top of line structure is 100%, bottom is 0%).

5. CONCLUSIONS

In this paper, we present an image retrieval system for estimating semiconductor sidewall shapes from top-down scanning electron microscopy images. The proposed system is a modified version of a previously proposed approach. Line-edge sub-images are used in the feature extraction process as opposed to the full-line sub-images used in the previous work. We also employ Gabor filter features to capture two-dimensional, texture-like features of the sub-image that may be correlated with sidewall shape. Finally, we present a new approach for dimensionality reduction called direct, weighted linear discriminant analysis or 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.

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*, **11**, 2003. To appear.
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 II*, SPIE, 2003. To appear.
3. S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Transactions on Image Processing* **11**, pp. 1160–1167, October 2002.
4. D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A* **4**, pp. 2379–2394, December 1987.
5. J. R. Price and T. F. Gee, "Towards robust face recognition from video," in *30th Applied Imagery and Pattern Recognition Workshop (AIPR)*, pp. 94–100, October 2001.
6. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, pp. 711–720, July 1997.
7. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, second ed., 2001.
8. K. Fukunaga, *Statistical Pattern Recognition*, Morgan Kaufmann, 1990.
9. M. Loog, R. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise Fisher criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, pp. 762–766, July 2001.
10. W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Proceedings of the IEEE International Conference 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 Recognition* **34**, pp. 2067–2070, October 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 Recognition* **10**, pp. 1713–1726, October 2000.