

# A Machine Learning Based Framework for Sub-Resolution Assist Feature Generation

Xiaoqing Xu, Tetsuaki Matsunawa<sup>†</sup>, Shigeki Nojima<sup>†</sup>, Chikaaki Kodama<sup>†</sup>  
Toshiya Kotani<sup>†</sup>, David Z. Pan  
ECE Department, Univ. of Texas at Austin, Austin, TX, USA  
Toshiba Corporation, Semiconductor & Storage Products Company, Yokohama, Japan<sup>†</sup>  
{xiaoqingxu, dpan}@cerc.utexas.edu  
{tetsuaki.matsunawa, shigeki.nojima, chikaaki1.kodama, toshiya.kotani}@toshiba.co.jp

## ABSTRACT

Sub-Resolution Assist Feature (SRAF) generation is a very important resolution enhancement technique to improve yield in modern semiconductor manufacturing process. Model-based SRAF generation has been widely used to achieve high accuracy but it is known to be time consuming and it is hard to obtain consistent SRAFs on the same layout pattern configurations. This paper proposes the first machine learning based framework for fast yet consistent SRAF generation with high quality of results. Our technical contributions include robust feature extraction, novel feature compaction, model training for SRAF classification and prediction, and the final SRAF generation with consideration of practical mask manufacturing constraints. Experimental results demonstrate that, compared with commercial Calibre tool, our machine learning based SRAF generation obtains 10X speed up and comparable performance in terms of edge placement error (EPE) and process variation (PV) band.

## CCS Concepts

•Hardware → VLSI design manufacturing considerations;

## Keywords

Sub-Resolution Assist Feature (SRAF), Machine Learning

## 1. INTRODUCTION

As the technology node continues scaling down, the 193nm wavelength photolithography with low  $k_1$  value is the mainstream technique to achieve smaller feature size. However, low image contrast and complex target pattern shapes make it extremely difficult for low- $k_1$  lithography to obtain acceptable lithographic process windows [1]. Besides the design for manufacturability techniques, like multiple patterning and litho-friendly layout design, mask optimization through resolution enhancement techniques (RETs) remains as the key strategy to improve the lithographic process window

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ISPD '16, April 3–6, 2016, Santa Rosa, California, USA.

© 2016 ACM. ISBN 978-1-4503-4039-7/16/04...\$15.00

DOI: <http://dx.doi.org/10.1145/2872334.2872357>

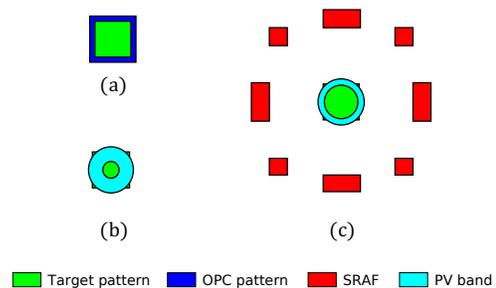


Figure 1: (a) An isolated contact, (b) printing with OPC, (c) printing with SRAF generation and OPC.

and the yield of the volume production in advanced technology nodes [2–6]. Major RETs include source mask co-optimization, sub-resolution assist feature (SRAF) generation and optical proximity correction (OPC). Among them, the SRAF generation is particularly important to improve the lithographic process window of target patterns. The key physical mechanism behind is that, without printing themselves, the small SRAF patterns would deliver light to the positions of target patterns at proper phase so that the printing of target patterns will be more robust to the lithographic variations. The lithographic process window is quantified with the process variation (PV) band area, which should be minimized to obtain a robust mask optimization solution. An example demonstrating the benefit of SRAF generation is shown in Fig. 1. An isolated target contact with the OPC pattern is shown in Fig. 1(a) and the target pattern is optimized only with OPC in Fig. 1(b), while the optimization in Fig. 1(c) is done with both SRAF generation and OPC. It can be clearly observed that much smaller PV band area is achieved in Fig. 1(c). Therefore, fast SRAF generation with high quality is of great importance for the mask optimization.

Multiple SRAF generation approaches, including model-based and rule-based approaches, have been developed and widely used in standard mask optimization flows. The rule-based approach is widely adopted due to its fast execution time and acceptable performance for simple designs and regular target patterns [1, 7, 8]. However, the rule-based SRAF is hard to deal with complex two-dimension (2D) shapes as it requires significant engineering efforts to setup and maintain the rule table [7]. Model-based SRAF generation methods can be divided into two categories based on the lithographic computations involved. One is to use simulated aerial im-

ages to seed the SRAF generation [9–12]. The other is to apply inverse lithography technology (ILT) and compute the image contour to guide the SRAF generation [13, 14]. Despite better lithographic performance and generalization capabilities compared to the rule-based approach, the model-based SRAF is known to be very time-consuming and it is difficult to achieve the same SRAFs around the same layout configurations, i.e. consistent SRAFs [1, 7].

Recently, the machine learning technique has been introduced to the computational lithography domain, with applications to lithographic hotspot detection [15–19] and OPC [20–23]. The machine learning technique calibrates a mathematical model with respect to an objective from the training data set based on accurate lithographic computations. Then, the calibrated model can predict the objective values, like a hotspot or non-hotspot for the hotspot detection and the shifting distance of an edge segment for the OPC, on the testing data. The machine learning technique usually demonstrates a trade-off between computational efforts and lithographic performance, which makes it particularly interesting for the SRAF generation problem. However, to the best of our knowledge, there is no prior art in applying the machine learning technique to the SRAF generation issue. In this work, we propose the first machine learning based framework for the SRAF generation. Our methodology can achieve fast yet consistent SRAFs with high quality in a 2D grid plane. Our main contributions are summarized as follows:

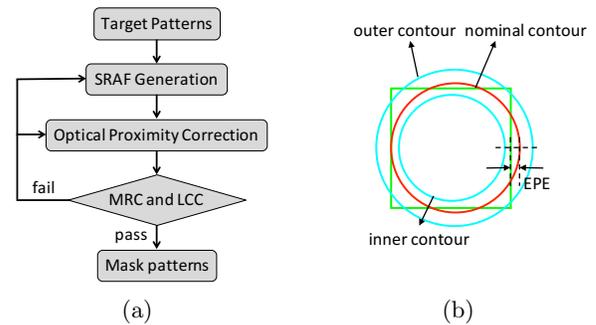
- A machine learning based framework is proposed for the SRAF generation, where a classification model is calibrated for SRAF predictions using model-based SRAFs as the training data.
- We propose a robust feature extraction scheme by adapting the concentric circle with area sampling considering SRAF-specific constraints. We further propose a novel feature compaction technique taking advantage of illumination source symmetry properties to reduce the training data size and improve the SRAF consistency.
- Logistic regression and decision tree models are calibrated for fast SRAF predictions due to the large data set size and high feature vector dimension. Different from conventional label predictions, we propose predictions with probability maxima in the 2D grid plane to generate manufacturing-friendly SRAFs.
- Our machine learning based framework achieves 10X speed-up in layout windows with comparable lithographic performance, compared with an industry strength model-based approach.

The rest of this paper is organized as follows. Section 2 introduces the standard mask optimization flow and related evaluation metrics. Section 3 gives the basic definitions and problem formulations. Section 4 explains the details on the feature extraction/compaction and model calibration. Section 5 shows how to generate SRAFs from the classification model while accommodating the mask manufacturing rules. Section 6 demonstrates the effectiveness of the proposed framework with comprehensive results. Section 7 concludes the paper.

## 2. PRELIMINARIES

### 2.1 Mask Optimization Flow

A standard mask optimization flow consists of several stages, including SRAF generation, OPC, mask manufacturing rule check (MRC) and lithography compliance check (LCC) as shown in Fig. 2(a) [1]. Depending on the outcome of MRC and LCC, iterative optimizations may be applied to achieve legal mask patterns. The MRC will check whether mask patterns satisfy a set of mask manufacturing rules. The LCC means lithography simulations are performed to check whether lithographic constraints are satisfied. In the stage of SRAF generation, small SRAFs will be added and isolated patterns on the mask will become dense patterns as shown in Fig. 1(b). SRAFs will not be printed themselves but will benefit the lithographic process windows of target patterns. In the next stage, OPC will shift the edges of OPC patterns to compensate for the optical proximity effects. Then, for the MRC, we assume the target patterns are MRC-clean and some typical mask manufacturing rules are applied to the SRAFs since this work mainly focuses on the SRAF generation. Typical mask manufacturing rules for SRAFs include maximum width (max\_width) rule, minimum space (min\_space) rule and maximum length (max\_length) rule. The LCC will introduce a lithographic process window involving a set of {focus, dose} conditions [24]. Lithography simulations at various conditions are performed to check whether the metrics, such as PV band and edge placement error (EPE), meet the criteria.



**Figure 2: Mask optimization: (a) mask optimization flow, (b) lithography simulation contours.**

### 2.2 Evaluation Metrics

We introduce several metrics to evaluate the performance of mask optimization results. An example of lithography simulation results is shown in Fig. 2(b). Inner and outer contours are explicitly drawn to demonstrate the lithographic printing variations due to the imposed {focus, dose} conditions. Nominal contour represents the lithographic printing at the best {focus, dose} condition. To quantify the lithographic variations, we define PV band and EPE as follows.

**Definition 1 (PV Band)** *Given the lithography simulation contours at a set of {focus, dose} conditions, the process variation (PV) band is defined as the area between the outer contour and inner contour.*

**Definition 2 (EPE)** *Given the lithography simulation contour at the best {focus, dose} condition, i.e. nominal contour*

and a measurement point, the edge placement error (EPE) is defined as the distance between the target pattern contour and nominal contour.

Thus, in Fig. 2(b), the area between the outer contour and inner contour is the PV band. A measurement point is drawn with a dashed line orthogonal to the vertical edge of the target pattern in Fig. 2(b) and the EPE can be explicitly quantified. The SRAF consistency is an important issue since it is closely related to the process variations on wafer [1]. Consistent SRAFs are preferred around the same target pattern configurations because different SRAFs lead to different OPC results, which potentially introduce extra process variations. We define the consistent SRAF generation as follows.

**Definition 3 (Consistent SRAF generation)** *Consistent SRAF generation means the same SRAF patterns should be generated for the same target layout configurations.*

### 3. PROBLEM FORMULATION

The machine learning based SRAF generation framework works on a 2D grid plane with a specific grid size. The training data consist of a set of layout clips, where each layout clip includes a set of target patterns and model-based SRAFs. With the 2D grid plane and the training patterns, training samples can be extracted at each grid point. To clearly explain the training data, we define the SRAF label as follows.

**Definition 4 (SRAF label)** *Given model-based SRAFs on the 2D grid plane, the SRAF label of a grid is 1 or 0, where 1 denotes an SRAF is inserted at that grid and 0 vice versa.*

Specifically, a training data point includes a feature vector and an SRAF label. The feature vector represents the optical conditions of the grid point with respect to the target patterns. With the training data, we define the classification-based SRAF as follows.

**Problem 1 (Classification-based SRAF)** *Given the 2D grid plane and training patterns with model-based SRAFs, feature vectors and SRAF labels of all grid points are extracted and a classification model is calibrated to predict the SRAF insertion at each grid of testing patterns.*

In the testing phase, the classification model can predict the SRAF label at each grid for testing patterns. Those grids with SRAF labeled as 1 can not directly be treated as the final SRAFs and further simplifications are needed to generate SRAFs accommodating mask manufacturing rules. Thus, we define the SRAF generation as follows.

**Problem 2 (SRAF Generation)** *Given the classification model and test patterns, SRAFs are generated while accommodating the mask manufacturing rules.*

## 4. CLASSIFICATION-BASED SRAF

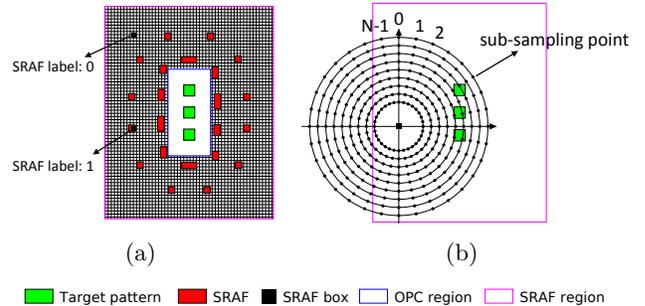
### 4.1 Data Preparation

#### 4.1.1 SRAF Label Extraction

Given training patterns with model-based SRAFs on a 2D grid plane, we need to extract the training data, including

the SRAF label and feature vector for each grid. As shown in Fig. 3(a), a 2D grid plane is imposed on the target patterns and model-based SRAFs. The coordinates of each grid are determined by the pre-set grid size. A SRAF box is introduced at each grid to decide the SRAF label from model-based SRAFs. Specifically, the SRAF box is a rectangle and the size is a parameter, which could be different from the grid size. The SRAF label of the grid is 0 if no model-based SRAF covering the SRAF box on the grid. The SRAF label is 1 when there is a model-based SRAF covering the entire SRAF box area. Therefore, the grid size of the 2D grid plane decides the granularity of the training data extraction while the SRAF box provides an alternative control on the SRAF label extraction accuracy. The SRAF label extraction will give a set of labels for all the grids, denoted as  $\{y_0\}$ .

In addition, an OPC region and an SRAF region are explicitly drawn in Fig. 3(a) to demonstrate SRAF-specific constraints. SRAF generation is not allowed in the OPC region since it is reserved for the OPC stage after the SRAF generation. Since the optical interference happens within some specific lithographic interaction window, the SRAF generation outside of the pre-determined SRAF region can be ignored. Both OPC region and SRAF region are created by expanding the edges of the target patterns by some specific distance. We define the distance of expansion for the OPC region and SRAF region as  $d_{opc}$  and  $d_{sraf}$ , respectively.



**Figure 3: (a) SRAF label extraction and sampling constraints, (b) CCCAS at one grid point.**

#### 4.1.2 Feature Extraction and Compaction

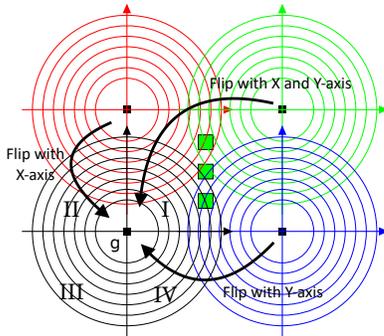
The layout feature extraction plays an important role in the classification model calibration and prediction. The SRAFs benefit the printing of target patterns by delivering light to the positions of target patterns at proper phase. Thus, we need a layout feature that represents this physical phenomenon. The concentric circle with area sampling is an ideal candidate since it represents the information related to the concentric propagation of diffracted light from mask patterns [22]. We adapt it to the constrained concentric circle with area sampling (CCCAS) by incorporating the OPC region and SRAF region constraints discussed in Section 4.1.1. The CCCAS at one grid is illustrated in Fig. 3(b), where each circle centers at the grid and the minimum and maximum radius of the CCCAS are determined by  $d_{opc}$  and  $d_{sraf}$ , respectively. After transforming target patterns into the bitmap on the 2D plane, the CCCAS yields a  $M \times N$  matrix, denoted as  $X_0$ , where  $M$  is the row number and  $N$  is the column number. As shown in Fig. 3(b), the column index of  $X_0$  starts at the positive Y-axis with 0 and increases clockwise to  $N - 1$ . The sub-sampling points, denoted as the

black dots in Fig. 3(b), sharing the same angle to the origin are on the same column of  $X_0$ . The row index of  $X_0$  starts with 0 at the circle with the smallest radius and increases to  $M-1$  as the circle radius becomes larger. The sub-sampling points on the same circle is on the same row of  $X_0$ .

The consistent SRAF generation is an important issue, which means the same SRAFs will be generated surrounding the same target pattern configurations, i.e. the same optical conditions [7]. For example, in Fig. 4, the four grids are on axial symmetric positions of the grid plane with respect to the target patterns. If we assume the annular shape for the illumination source, the optical conditions of these four grids are the same and consistent SRAF generation scheme will give the same SRAF results. However, since the CCCAS at these four grids are different, denoted by different colors in Fig. 4, it is difficult for a classification model to achieve the same SRAF predictions. To achieve better SRAF consistency, we propose a novel feature compaction technique taking advantage of the illumination source symmetry. As shown in Fig. 4, this feature compaction technique transforms the CCCAS of symmetric grids into the CCCAS of the same grid, denoted as  $g$ , in the lower left of the grid plane. The sampling region of the grid ( $g$ ) can be divided into four quadrants, i.e.  $I$ ,  $II$ ,  $III$  and  $IV$ . The target patterns mainly locate at the quadrant  $I$  of the sampling region of the grid ( $g$ ), while target patterns mainly locate at different quadrants for other symmetric grids. For clearer explanations, we define the main quadrant as follows.

**Definition 5 (Main Quadrant)** *The main quadrant for a grid is defined as the quadrant of the CCCAS region where target patterns mainly locate.*

By flipping the CCCAS of other symmetric grids with X or Y-axis as shown in Fig. 4, target patterns will always locate at the quadrant  $I$  of the sampling regions for symmetric grids, which leads to the same CCCAS results. Then, the classification model will give consistent SRAF predictions for axial symmetric grids.



**Figure 4: Feature compaction based on symmetry.**

The details of the feature compaction technique are explained in Algorithm 1. The 2D feature matrix from CCCAS contains the bitmap value at each sub-sampling point. The bitmap values of sub-sampling points within each quadrant of the sampling region correspond to a set of columns in the 2D feature matrix. Thus, the main quadrant should have the maximum summation of bitmap values at the 2D feature matrix. From lines 3 to 9, we scan through the four quadrants of the CCCAS region and decide the main quadrant. The flipping of 2D matrix is performed in line 10 to

---

### Algorithm 1 Feature Compaction

---

**Input:** A  $M \times N$  feature matrix  $X_0$ ;  
**Output:** Optimized feature vector  $x_0$ ;  
1: Define  $main\_quadrant = 1$  as the main quadrant;  
2: Define  $max\_sum = 0$  as the maximum summation;  
3: **for**  $index = 0, index < 4, index++$  **do**;  
4:     Define  $sum =$  summation of  $X_0$  from column  $index \times N/4$  to column  $(index + 1) \times N/4$ ;  
5:     **if**  $sum > max\_sum$  **then**;  
6:          $max\_sum = sum$ ;  
7:          $main\_quadrant = index + 1$ ;  
8:     **end if**  
9: **end for**  
10: Flip  $X_0$  based on  $main\_quadrant$ ;  
11: Flatten  $X_0$  into a vector  $x_0$  and return  $x_0$ ;

---

transform the main quadrant to quadrant  $I$  as demonstrated in Fig. 4. For practical implementation, the flipping of 2D matrix can be achieved with simple column index switching. In line 11, the 2D feature matrix is flattened into a one dimension (1D) feature vector for the classification model calibration. It shall be noted that, the illumination source symmetry-based feature compaction scheme can be easily extended to other symmetric scenarios, such as rotational symmetry. Overall, the feature extraction and compaction will yield a set of 1D feature vectors with  $M \times N$  dimensions.

## 4.2 Model Training

With the SRAF labels and feature vectors, a classification model is calibrated for SRAF predictions. In particular, the size of training data set and feature vector dimension can both be very large because high sampling accuracy is needed for classification-based SRAF. Moreover, the large training and testing data set sizes make the classification-based SRAF problem difficult to adopt support vector machine (SVM) and other advanced classification models from the perspective of calibration and prediction runtime. Thus, we adopt decision tree (DTree) and logistic regression (LGR) models for SRAF predictions with reasonable performance and runtime.

**Decision Tree** The DTree model is simple yet powerful, which partitions the feature space into a set of rectangles and calibrate a simple model (like a constant label for classification model) in each one [25]. Specifically, the DTree model calibration is to construct a decision tree from labeled training data in the top-down manner, using a metric to decide the best splits of set of training data at each step [25]. For this application, the Gini index metric is used to decide the best splits at each step of calibration with the CART (Class and Regression Trees) algorithm. With the Gini index, the DTree classifier can estimate the probability of a label for each data sample by calculating the fraction samples of that label in a leaf [25].

**Logistic Regression** The LGR adopts the logistic function as the probabilistic estimation for each label of the training or testing data. The model calibration is typically achieved with the maximum likelihood method [25]. The LGR model is especially powerful for binary classification, which makes the calibration and prediction scalable to large data set. Due to the large training data set in classification-based SRAF issue,  $L_2$  regularization is added to the LGR model to avoid overfitting. The LGR model provides the direct probabilistic estimation of labels for each data sample.

## 5. SRAF GENERATION

### 5.1 Predictions with Probability Maxima

The typical prediction with a binary classification model will be a label, i.e. 0 or 1, for each testing data. With the label prediction for each grid, clusters of grids will be labeled as 1, denoted as yellow grids, as shown in Fig. 5(a). After the label prediction, clusters of grids in Fig. 5(a) cannot be directly treated as SRAFs because they may violate the mask manufacturing rules or be printed due to large critical dimensions. Instead of using SRAF label for the grid prediction, we propose predictions with probability maxima to simplify the clusters of SRAF grids. When a classification model is calibrated, the probability of the label to be 1, denoted as  $p_1$ , can be calculated for LGR and DTree as explained in Section 4.2. Then, a probability map on the 2D grid plane can be attained as shown in Fig. 5(b). To simplify the clusters of grids for SRAF generation, we only insert SRAFs at grids with probability maxima. A grid with probability maximum means the probability ( $p_1$ ) at that particular grid is larger than that at any other neighboring grids. The idea of predictions with probability maxima originates from the model-based SRAF approach. Model-based SRAFs are generated using the guidance map from lithographic computations [9–12]. A guidance map is also grid based and has intensity assigned to each grid, where SRAFs will only be inserted at those intensity maxima. Thus, we adopt the similar idea during predictions with probability maxima since model-based SRAFs are used as the training data for the classification model calibration.

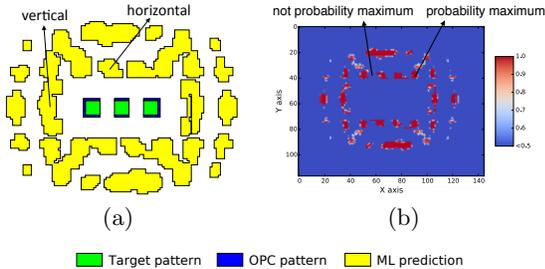


Figure 5: SRAF predictions: (a) label predictions, (b) predictions with probability maxima.

### 5.2 SRAF Simplification

Using predictions with probability maxima, clusters of grids will be predicted as SRAFs on the 2D grid plane but the mask manufacturing and SRAF printing issues are not fully resolved. The SRAF simplification phase aims at simplifying these clusters of grids into SRAFs satisfying the mask manufacturing rules. A greedy simplification scheme is proposed by grid merging and shrinking the SRAFs into rectangular shapes while accommodating mask manufacturing rules.

The overall algorithm of SRAF generation is shown in Algorithm 2. In lines 1-2, we predict SRAFs at grids with probability maxima from the classification model. In line 3, the grids with probability maxima are merged into polygons, which is followed by the spacing rule check and shrinking the polygons to remove violations in line 4. From line 5 to 14, each polygon is processed to generate a rectangular SRAF. Particularly, in line 7, the main direction of SRAF is detected based on the bounding box of target patterns. In

line 8, the bounding box of the polygon is shrunk to achieve a rectangular SRAF parallel to that of target patterns. As illustrated in Fig. 5(a), the main direction of the polygon on the top is horizontal while the main direction on the left is vertical, both of which are parallel to the bounding box of target patterns. With the SRAF simplification, the mask manufacturing-friendly SRAFs can be generated for testing patterns.

---

#### Algorithm 2 SRAF generation

---

**Input:** A 2D grid plane, a classification models, a set of mask manufacturing rules;  
**Output:** The mask manufacturing friendly set  $SRAF$ ;  
1: Compute the probability of label 1 for each grid;  
2: SRAF predictions at grids with probability maxima;  
3: Merge SRAF grids into a polygon set  $SRAF_{pg}$ ;  
4: Spacing rule check and shrink polygons in  $SRAF_{pg}$  to remove violations;  
5: **for** each *polygon* in  $SRAF_{pg}$  **do**;  
6:     Define  $BBox$  as the bounding box of *polygon*;  
7:     Detect the main direction of *polygon* as *direction*;  
8:     Shrink  $BBox$  size based on *direction*;  
9: **end for**  
10: Rule check and shrink rectangles in  $SRAF$  to remove violations;  
11: Return  $SRAF$ ;

---

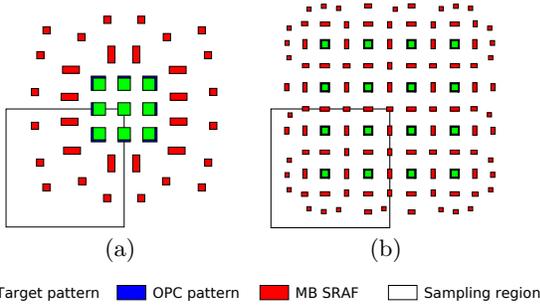
## 6. EXPERIMENTAL RESULTS

We have implemented the machine learning framework in Python and accelerated with Cython [26] and parallel computations. The optical model, model-based SRAF, MRC/LCC recipes and the SRAF simplification are implemented using Calibre script language with the industry-strength setup. All experiments are performed on an 8-core Linux machine with 3.4GHz Intel(R) Core and 32GB memory. For the optical model, the wavelength ( $\lambda$ ) and numerical aperture ( $NA$ ) are set as 193nm and 1.35, respectively. The annular illumination source is used with outer sigma as 0.9 and inner sigma as 0.6. Compact model 1 from Calibre is adopted as the resist model. In the LCC, the outer/inner contours are generated using lithographic process window conditions as a focus variation of  $\pm 30nm$  and a dose variation of  $\pm 3.0\%$ . For model-based SRAF generation, process window conditions above are considered and SRAF manufacturing rules are set as  $max\_width = 40nm$ ,  $min\_space = 60nm$ ,  $max\_length = 90nm$ . We test the SRAF generation framework on two types of contact patterns. One type is dense contact arrays with contact width and space fixed as 70nm. We have dense contact patterns because redundant vias are needed to improve yield during layout design. The other type is sparse contact patterns, where the contact width is 70nm but the space between contact holes is random and the minimum space is 70nm. For CCCAS, the grid size is set as 10nm, the SRAF box size is set as 40nm and radius step size is  $s_r = 15nm$ . For SRAF-specific constraints,  $d_{opc}$  and  $d_{sraf}$  are set as 100nm and 600nm, respectively.

### 6.1 Model Training

A set of training patterns and model-based SRAFs are needed to extract the training data and calibrate the classification model for SRAF predictions on both dense and sparse testing patterns. From the extensive experiments, the training patterns in Fig. 6 yield the best training and testing accuracy. For the dense contact patterns in Fig. 6(a), the width and space are fixed as 70nm. The width of sparse contact patterns in Fig. 6(b) is 70nm, while the space is 350nm.

In particular, since the training patterns are symmetric and feature compaction scheme has been proposed, we only need to sample the lower left part of the layout clip for training data. This is beneficial for the classification model calibration since the training data size can be reduced by 3/4 without losing the critical SRAF information. The training data set statistics are summarized in Table 1. For CCCAS, the number of circle is set as  $M = \lfloor (d_{sraf} - d_{opc}) / s_r \rfloor = 33$ . The number of sub-sampling points in each circle is set as  $N = 32$  to guarantee sampling accuracy for sparse contact patterns. Then, the feature vector dimension is  $M \times N = 1056$ . By combining the training data from dense and sparse patterns, we have 14609 training samples. We have 95412 testing samples from dense patterns and 803756 testing samples from sparse patterns.



**Figure 6: Training layout: (a) dense contact patterns, (b) sparse contact patterns.**

The data set statistics demonstrate the high feature vector dimension and large training data size. The feature vector dimension is difficult to be further reduced since each sub-sampling point contains the information related to the target patterns. Moreover, each training data sample within the lithographic interaction window is considered valuable, so there is little redundancy within the training data set. As discussed in Section 4.2, advanced classification models, such as support vector machine, are not applicable to the classification-based SRAF domain due to large training and testing data sets. In practice, we find that the reasons are twofold. First, the training of advanced classification models has high complexity, which is not as scalable as the simple classification models to large training data set. Second, advanced classification models have more complex prediction model calibrated, which means the testing time for each testing sample would be longer than that of simple models. Since the grid based approach is used here, the runtime overhead accumulates and even becomes unaffordable over the huge amount of testing samples extracted from testing patterns.<sup>1</sup>

We further compare different classification models, including DTree and LGR, for the SRAF generation framework and data statistics are shown in Table 2. The  $F_1$  score is computed as:

$$F_1 \text{ score} = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})} \quad (1)$$

where *precision* is the number of true positive results divided by the number of all positive results and *recall* is

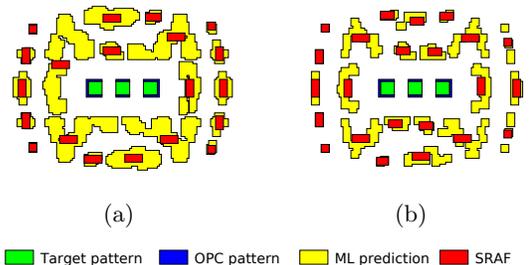
<sup>1</sup>Our implementation of the support vector machine with linear kernels shows much longer runtime than LGR and DTree, which leads to the overall SRAF generation runtime longer than that of the model-based SRAF from Calibre.

the number of true positive results divided by the number of positive results that should have been returned [27]. In particular,  $F_1$  score is best at 1.0 and worst at 0.0 for comparison. Since either label 1 or label 0 can be treated as positive result, we compute the  $F_1$  score for both labels, denoted as  $F_1 \text{ score}(1)$  and  $F_1 \text{ score}(0)$ , respectively. From Table 2, although DTree achieves better  $F_1 \text{ score}(1)$  and  $F_1 \text{ score}(0)$  on training data, the testing  $F_1 \text{ score}(0)$  and  $F_1 \text{ score}(1)$  are worse than LGR on dense patterns by 0.081 and 0.006, respectively. We only see a difference within 0.01 in  $F_1$  score for sparse testing patterns. This means LGR is better than DTree due to its better testing accuracy and LGR is less prone to the overfitting with large training data set and high dimension feature vectors. In addition, the model calibration time and testing time on sparse testing patterns for LGR are less than DTree but the difference is non-significant. Moreover, we observe the  $F_1 \text{ score}(1)$  is much better than  $F_1 \text{ score}(0)$  on all testing data for both classification models. This means that most grids with SRAFs of testing patterns are labeled as 1 correctly but some other grids without SRAFs that should be labeled as 0 are incorrectly labeled as 1. This also proves the necessity of predictions with probability maxima and the SRAF simplification stage later on, which essentially reduces the number of grids labeled as 1 and potentially improves the  $F_1 \text{ score}(0)$  for testing data.

## 6.2 SRAF Generation

### 6.2.1 SRAF Simplification

We add testing layout clips to demonstrate the strength of predictions with probability maxima and SRAF simplification schemes. Since the LGR based approach performs better than the DTree based approach, we only show the SRAFs from LGR predictions for clear explanations. As illustrated in Fig. 7, we compare the SRAFs generated using different machine learning (ML) predictions, i.e. label predictions and predictions with probability maxima, followed by the SRAF simplification phase. Predictions with probability maxima can simplify the clusters of grids labeled as 1, i.e. breaking large clusters into small clusters, which benefits the SRAF simplification stage. Thus, the SRAFs generated using predictions with probability maxima in Fig. 7(b) are much better than those in Fig. 7(a) in terms of PV band from the LCC.



**Figure 7: SRAF generations: (a) label predictions, (b) predictions with probability maxima.**

### 6.2.2 SRAF Consistency

We further demonstrate the benefit of SRAF consistency improvement from the feature compaction technique in Sec-

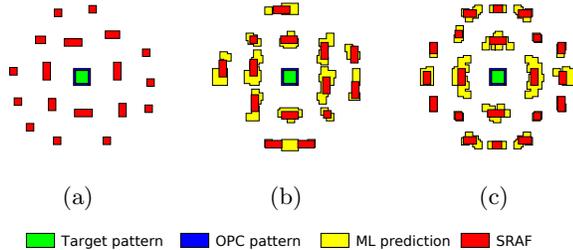
**Table 1: Data set statistics**

feature vector dimension	# of training samples	# of testing samples from dense patterns	# of testing samples from sparse patterns
1056	14609	95412	803756

**Table 2: Comparisons on different classification models**

Model	Training			Dense Testing Patterns			Sparse Testing Patterns		
	$F_1$ score(1)	$F_1$ score(0)	Calibration time(s)	$F_1$ score(1)	$F_1$ score(0)	Runtime(s)	$F_1$ score(1)	$F_1$ score(0)	Runtime(s)
DTree	0.9983	0.9855	5.66	0.9499	0.3319	3.83	0.8787	0.2646	50.44
LGR	0.9938	0.9462	1.06	0.9557	0.4132	3.83	0.8724	0.2629	41.47

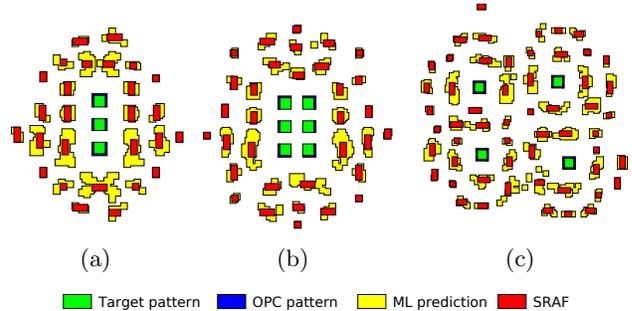
tion 4.1.2. The SRAF generation from model-based method using Calibre [Calibre, v2015.2.36.27], LGR without feature compaction and LGR with feature compaction are shown in Fig. 8(a), (b) and (c), respectively. Since annular illumination source is used, axial symmetric grids share the same optical environment and the consistent SRAF generation should yield the same SRAFs at axial symmetric grids. The feature compaction scheme would transform the feature matrices extracted from axial symmetric grids to the same feature vector. Therefore, the SRAFs in Fig. 8(c) are more consistent than those in Fig. 8(b). Moreover, we have even achieved SRAFs with better consistency than the model-based method shown in Fig. 8(a).

**Figure 8: SRAFs for the isolated contact pattern: (a) model-based, (b) LGR without feature compaction, (c) LGR with feature compaction.**

We further demonstrate the SRAFs from LGR predictions on both dense and sparse testing patterns as shown in Fig. 9. The layout clip of dense contact patterns is defined as  $m \times n$  dense contact patterns, if the layout clip contains a contact array with  $m$  rows and  $n$  columns. Fig. 9(a) and (b) show two cases of redundant vias in real designs. The LGR based SRAF generation can obtain acceptable SRAFs on these dense contact patterns but the degradation of SRAF consistency is observed. The reasons are twofold. First, the training data with model-based SRAFs are not perfectly consistent as shown in Fig. 6. Then, it is difficult to guarantee the consistent SRAF generation with the classification model calibrated with these training data. Second, the CC-CAS results may be slightly different for axial symmetric grids due to the grid error within the 2D grid plane. Fig. 9(c) illustrates a small layout clip of random contact patterns, which proves the capability of the machine learning based SRAF generation on random sparse contact patterns.

### 6.3 Lithography Compliance Check

To evaluate the practical lithographic performance, we combine the SRAF generation with a complete mask optimization flow as shown in Fig. 2, where model-based OPC and LCC are performed using Mentor Calibre tool. We compare the model-based, LGR and DTree approaches in

**Figure 9: Testing contact patterns: (a)  $3 \times 1$  dense contact patterns, (b)  $3 \times 2$  dense contact patterns, (c) sparse contact patterns**

terms of PV band and EPE on both dense and sparse testing patterns. Specifically, we collect the PV band value for each contact and the EPE value at the center of the four edges of each contact at nominal conditions. The mean values are summarized and compared in Table 3. We add the PV band without SRAFs to better demonstrate the benefit from SRAF generation. The model-based approach reduces the PV band from 3.3064 to 2.7705, which is 16.21% reduction compared to no SRAF insertion. Meanwhile, we obtain 13.37% and 14.09% PV band reduction from DTree and LGR, respectively. The LGR based SRAF generation achieves better performance as expected because LGR model is less prone to overfitting compared to DTree model. In particular, there is only 2.12% PV band degradation from model-based approach to LGR based approach. We take the absolute values when calculating the EPE mean to avoid the cancellations between positive and negative values of EPE. The LGR based SRAF generation yields the smallest EPE mean value, which even outperforms the model-based approach. This means there is some trade-off between PV band and EPE because different SRAF results lead to different OPC results. It is very difficult to improve the PV band and EPE simultaneously with a robust mask optimization flow.

**Table 3: PV band and absolute of EPE**

Mean value	No SRAF	Model-based	DTree	LGR
PV band ( $.001\mu m^2$ )	3.306	2.771	2.864	2.841
Absolute of EPE (nm)	3.636	0.539	0.523	0.501

We collect the PV band and EPE values for each contact and further plot the data in histograms as shown in Fig. 10. Fig. 10(a) shows that SRAF insertion significantly improves the PV band and model-based SRAF gives the best performance. The LGR based approach is slightly worse than model-based method but performs better than DTree based method. Fig. 10(b) shows that LGR based SRAF generation achieves the best EPE performance.

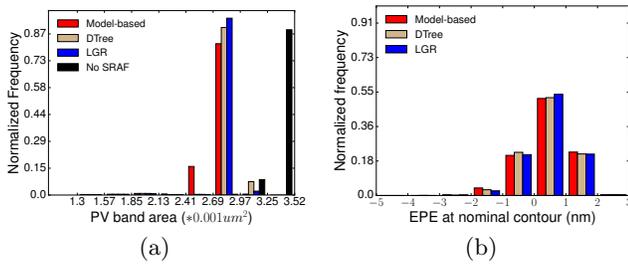


Figure 10: Comparison among different schemes: (a) PV band distribution, (b) EPE distribution.

## 6.4 Run Time

We compare our machine learning based SRAF generation with the commercial Calibre tool, i.e. model-based SRAFs. The mask optimization techniques, including SRAF generation and OPC, usually apply to small layout windows due to the high computational cost [2]. Therefore, we choose small layout windows with dense contact patterns for runtime comparisons between different SRAF generation approaches. In Fig. 11, different dense contact patterns, denoted as  $m \times n$  contact patterns explained in Section 6.2.2, are used for runtime comparisons. The areas of these layout windows considering SRAF regions are in the range from  $1\mu m^2$  to  $2\mu m^2$ . The runtime for the machine learning based approach includes runtime for feature extraction and compaction, predictions with probability maxima and SRAF simplification.

The average runtime for model-based SRAF, LGR based SRAF and DTree based SRAF are 5.14s, 0.41s and 0.41s, respectively. Although we are using a different programming language and database from commercial tool, we still obtain over 10X speed-up from machine learning based SRAF generation compared to the model-based approach in Calibre [Calibre, v2014.4.18.13]. We also check the runtime of the model-based OPC from different SRAF generation approaches and ensure that they are approximately the same.

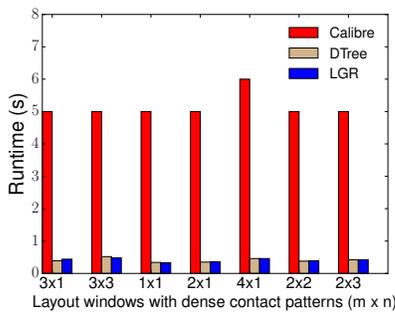


Figure 11: Run time comparison among different schemes on different layout windows.

## 7. CONCLUSION

A machine learning based framework for SRAF generation is demonstrated for the first time. A robust feature extraction scheme is proposed by adapting the CCAS considering SRAF-specific constraints. We further propose a novel feature compaction technique to reduce the training data size and improve the SRAF consistency. Predictions with probability maxima are proposed to achieve mask manufacturing-

friendly SRAFs. Experimental results show that LGR based SRAF generation obtains 10X speed-up in layout windows and better EPE with affordable degradation in PV band, compared to the commercial Calibre tool.

## 8. ACKNOWLEDGEMENT

This work is supported in part by SRC and NSF. The authors would like to thank the memory lithography group (MLG) in Toshiba Corporation for the helpful discussions and feedback on this work.

## 9. REFERENCES

- [1] Y. Ping, S. McGowan, Y. Gong, Y. M. Foong, J. Liu, J. Qiu, V. Shu, B. Yan, J. Ye, P. Li *et al.*, "Process window enhancement using advanced ret techniques for 20nm contact layer," in *Proc. of SPIE*, 2014, pp. 90 521N–90 521N.
- [2] S. Banerjee, Z. Li, and S. R. Nassif, "ICCAD-2013 CAD contest in mask optimization and benchmark suite," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2013, pp. 271–274.
- [3] J.-R. Gao, X. Xu, Y. Bei, and D. Z. Pan, "MOSAIC: Mask optimizing solution with process window aware inverse correction," in *ACM/IEEE Design Automation Conference (DAC)*, 2014, pp. 52:1–52:6.
- [4] Y.-H. Su, Y.-C. Huang, L.-C. Tsai, Y.-W. Chang, and S. Banerjee, "Fast lithographic mask optimization considering process variation," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2014, pp. 230–237.
- [5] A. Awad, A. Takahashi, S. Tanaka, and C. Kodama, "A fast process variation and pattern fidelity aware mask optimization algorithm," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2014, pp. 238–245.
- [6] J. Kuang, W.-K. Chow, and E. F. Young, "A robust approach for process variation aware mask optimization," in *Proc. Design, Automation and Test in Europe*, 2015, pp. 1591–1594.
- [7] J.-H. Jun, M. Park, C. Park, H. Yang, D. Yim, M. Do, D. Lee, T. Kim, J. Choi, G. Luk-Pat *et al.*, "Layout optimization with assist features placement by model based rule tables for 2x node random contact," in *Proc. of SPIE*, 2015, pp. 94 270D–94 270D.
- [8] C. Kodama, T. Kotani, S. Nojima, and S. Mimotogi, "Sub-resolution assist feature arranging method and computer program product and manufacturing method of semiconductor device," Aug. 19 2014, US Patent 8,809,072.
- [9] K. Sakajiri, A. Tritchkov, and Y. Granik, "Model-based sraf insertion through pixel-based mask optimization at 32nm and beyond," in *Proc. of SPIE*, 2008, pp. 702 811–702 811.
- [10] R. Viswanathan, J. T. Azpiroz, and P. Selvam, "Process optimization through model based sraf printing prediction," in *Proc. of SPIE*, 2012, pp. 83 261A–83 261A.
- [11] J. Ye, Y. Cao, and H. Feng, "System and method for model-based sub-resolution assist feature generation," Feb. 1 2011, US Patent 7,882,480.
- [12] S. D. Shang, L. Swallow, and Y. Granik, "Model-based sraf insertion," Oct. 11 2011, US Patent 8,037,429.
- [13] L. Pang, Y. Liu, and D. Abrams, "Inverse lithography technology (ilt): a natural solution for model-based sraf at 45nm and 32nm," in *Proc. of SPIE*, 2007, pp. 660 739–660 739.
- [14] B.-S. Kim, Y.-H. Kim, S.-H. Lee, S.-I. Kim, S.-R. Ha, J. Kim, and A. Tritchkov, "Pixel-based sraf implementation for 32nm lithography process," in *Proc. of SPIE*, 2008, pp. 71 220T–71 220T.
- [15] J. A. Torres, "ICCAD-2012 CAD contest in fuzzy pattern matching for physical verification and benchmark suite," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2012.
- [16] D. Ding, X. Wu, J. Ghosh, and D. Z. Pan, "Machine learning based lithographic hotspot detection with critical-feature extraction and classification," in *IEEE International Conference on IC Design and Technology (ICIDT)*, 2009.
- [17] D. G. Drmanac, F. Liu, and L.-C. Wang, "Predicting variability in nanoscale lithography processes," in *ACM/IEEE Design Automation Conference (DAC)*, 2009, pp. 545–550.
- [18] Y. Yu, G. Lin, I. Jiang, and C. Chiang, "Machine-learning-based hotspot detection using topological classification and critical feature extraction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 34, no. 3, pp. 460–470, 2015.
- [19] T. Matsunawa, J.-R. Gao, B. Yu, and D. Z. Pan, "A new lithography hotspot detection framework based on adaboost classifier and simplified feature extraction," in *Proc. of SPIE*, 2015, pp. 94 270S–94 270S.
- [20] A. Gu and A. Zakhov, "Optical proximity correction with linear regression," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 2, pp. 263–271, 2008.
- [21] R. Luo, "Optical proximity correction using a multilayer perceptron neural network," *Journal of Optics*, vol. 15, no. 7, p. 075708, 2013.
- [22] T. Matsunawa, B. Yu, and D. Z. Pan, "Optical proximity correction with hierarchical bayes model," in *Proc. of SPIE*, 2015, pp. 94 260X–94 260X.
- [23] X. Ma, B. Wu, Z. Song, S. Jiang, and Y. Li, "Fast pixel-based optical proximity correction based on nonparametric kernel regression," *Journal of Micro lithography, Microfabrication and Microsystems*, vol. 13, no. 4, pp. 043 007–043 007, 2014.
- [24] P. Gupta, "What is process window?," *SIGDA NewsL*, vol. 40, no. 8, pp. 1–1, Aug. 2010.
- [25] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.
- [26] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, "Cython: The best of both worlds," *Computing in Science Engineering*, vol. 13, no. 2, pp. 31–39, 2011.
- [27] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in information retrieval*. Springer, 2005, pp. 345–359.