Methodologies for Layout Decomposition and Mask Optimization: A Systematic Review

(Invited Paper)

Yuzhe Ma[1], Xuan Zeng[2], and Bei Yu[1]

[1]CSE Department, The Chinese University of Hong Kong, Hong Kong
[2]State Key Laboratory of ASIC & Systems, Microelectronics Department, Fudan University, China
yzma@cse.cuhk.edu.hk, xzeng@fudan.edu.cn, byu@cse.cuhk.edu.hk

***Abstract*—As the transistor feature size keeps shrinking, manufacturability has become an urgent issue in semiconductor industry. In order to improve the manufacturability, various resolution enhancement techniques have been proposed, among which layout decomposition and mask optimization have been considered as the most powerful solutions in advanced technology nodes. Different from many previous survey papers that categorize literatures by type of manufacturing process, we argue that different manufacturing scenarios can share similar mathematical models. This paper carefully summarizes a series of methodologies that have been successfully applied to VLSI layout decomposition and mask optimization problems.**

I. Introduction

Due to the delay of the next generation of lithography techniques, current lithography wavelength is stuck at 193nm. As a result, resolution enhancement techniques (RETs) on layout and mask are of great importance to improve the yield. Multiple patterning lithography (MPL) has achieved great success in pushing forward the technology node. There are two of the most critical stages in MPL process, including layout decomposition and mask optimization. In layout decomposition, the target layout is decomposed into several layouts so that each decomposed layout can be manufactured under the current lithography condition. Two main types of MPL manufacturing process are litho-etch-litho-etch (LELE)-type MPL and spacer-type MPL. Spacer-type MPL typically refers to self-aligned double patterning (SADP). LELE-type refers to conventional double patterning layout decomposition (DPLD) or triple patterning layout decomposition (TPLD), depending on the number of masks that are used to separate the violating patterns. Examples of LELE-type MPL are presented in Fig. 1. SADP deposits a spacer layer over the chip covering all mask features. The covered layer is selectively etched away leaving two sidewalls along any ridge, and then the ridge is removed, as shown in Fig. 2.

E-beam lithography (EBL) is another promising candidates for MPL. Charged electron beams can be easily focused onto nanometer diameter, thus tiny patterns can be manufactured. EBL allows a great flexibility for fast turnaround times and late design modifications to adapt a given layout. Directed self-assembly (DSA) is an emerging technique which is particularly suitable for contact layers which have uniform size. Guiding templates are usually used to form contacts. In sparse layout, each feature can be created using a single-hole template. In a dense layout, two or more features can be grouped and thus manufactured by a multiple-hole template to ensure there is no conflict.

The diffraction effect of the light cannot be ignored since the size of the patterns on layout is comparable with the wavelength of the lithography light source, which may result in low fidelity of the final on-wafer image. In mask optimization, e.g., optical proximity correction (OPC), each mask is refined to compensate the diffraction effect of the light in advance to ensure the high quality of on-wafer image. Finally, all optimized masks go through lithography process separately, then all printed images are combined together to generate the target image.

This work is supported in part by The Research Grants Council of Hong Kong SAR (Project No. CUHK24209017), National Key Research and Development Program of China 2016YFB0201304, and National Natural Science Foundation of China (NSFC) research projects 61376040 and 61574046.

978-1-5386-2880-5/17$ 31.00 © 2017 IEEE

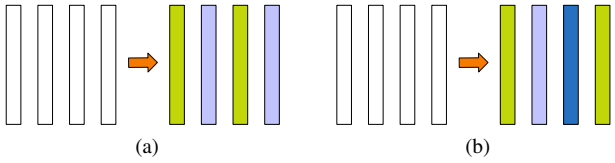


Fig. 1: Examples of (a) DPLD and (b) TPLD.

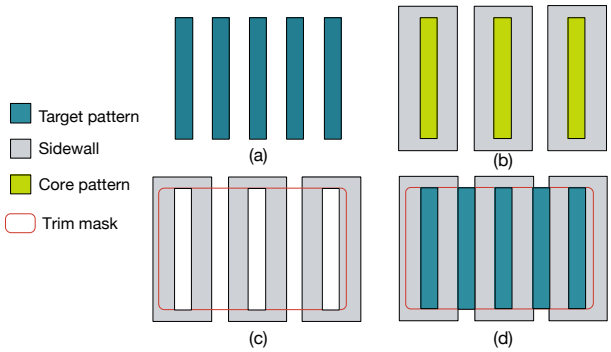


Fig. 2: An example to show SADP decomposition result: (a) Target pattern; (b) Core pattern and spacer; (c) Spacers after core pattern removal and trim mask; (d) Final pattern.

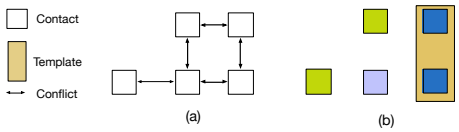
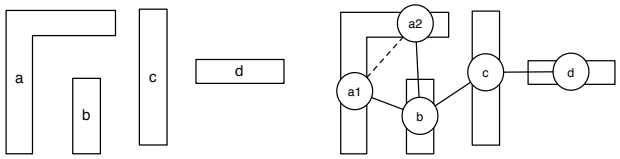


Fig. 3: Example of directed self-assembly: (a) Layout; (b) Mask and template assignment.

Recall that the target of layout decomposition and mask optimization is to improve the manufacturability of the layout. However, each problem is solved independently, which may lose a global view. Basically, the layout decomposition is based on simple design or coloring rules; while the mask optimization is verified by accurate and sophisticated lithography simulation. It is intuitive that the effect of one stage should be taken into consideration in another stage. In other words, layout decomposition and mask optimization may be considered simultaneously.

There are previous surveys introducing the problems of layout decomposition and mask optimization [1]–[4], but all of them categorize



(a) (b)

Fig. 4: (a) Layout; (b) Decomposition graph of LELE-type MPL.

different work by type of manufacturing process. We argue that the layout decomposition problem in different manufacturing processes may share analogous mathematical metrics, and the methodologies applied in different scenarios can be very similar. To avoid ad-hoc layout decomposition solution to a particular manufacturing process, in this paper we summarize a series of methodologies that have been widely used in this field.

The rest of the survey is organized as follows. Section II will introduce various methodologies proposed for solving the problem of layout decomposition. While different techniques for mask optimization, mainly for OPC, will be summarized in Section III. Those work which considered layout decomposition and mask optimization simultaneously will be shown in Section IV, followed by the conclusion in Section V.

II. Layout Decomposition

Definition 1 (Conflict)**.** *A* conflict *is introduced when the distance between two features is less than minimum coloring distance* min_s.

Definition 2 (Stitch)**.** *Some coloring conflict can be resolved by inserting* stitch *to split a pattern into two touching parts, which leads to yield loss due to the potential overlay issue.*

For LELE-type MPLD, a decomposition graph (DG) is an undirected graph with a set of vertices V, and two sets of edges, including conflict edges (CE) and stitch edges (SE). Each vertex in V corresponds to a polygonal shape in the layout. An edge is in CE if there is a conflict between two vertices. An edge is in SE if there is a stitch between the two vertices which are associated with the same polygonal shape. Fig. 4 gives an example of DG of LELE-type MPLD in which there are five vertices. Solid lines represent edges in CE and dash line represents edge in SE.

For other types of MPLD problem, the geometry of the layout can also be represented by a graph. The DG for EBL is very similar to Fig. 4. For MPLD with DSA, stitch insertion is not allowed. However, conflict patterns within grouping distance can be manufactured with the template, thus there is no stitch edge but grouping edge in DG for DSA.

A. Integer Linear Programming

Integer linear programming (ILP) is adopted to solve the problem of layout decomposition, including DPLD [5]–[7] and TPLD [8], [9]. For DPLD, the problem can be formulated as the following ILP:

$$\min_{x} \quad \sum c_{ij} + \alpha \times \sum s_{ij}, \tag{1}$$

$$\text{s.t.} \quad x_i + x_j - 1 \le c_{ij}, \qquad \forall e_{ij} \in CE, \tag{1a}$$

$$1 - x_i - x_j \le c_{ij}, \qquad \forall e_{ij} \in CE, \tag{1b}$$

$$x_i - x_j \le s_{ij}, \qquad \forall e_{ij} \in SE, \tag{1c}$$

$$x_j - x_i \le s_{ij}, \qquad \forall e_{ij} \in SE, \tag{1d}$$

$$x_i \in \{0, 1\}, \qquad \forall i \in V, \tag{1e}$$

where x_i and x_j are binary variables for the colors of vertices v_i and v_j. c_{ij} is a binary variable for conflict edge $e_{ij} \in CE$ and s_{ij} is a binary variable for stitch edge $e_{ij} \in SE$. $c_{ij} = 0$ when $x_i \neq x_j$ and $c_{ij} = 1$ when $x_i = x_j$. A cost α is incurred when v_i and v_j are assigned different colors, i.e., a stitch is introduced.

Xu *et al.* [6] formulated the problem into a maximum-cut problem and utilized an ILP formulation on stitch minimization.

For the TPLD problem, the objective is also to simultaneously minimize the conflict number and the stitch number. A user-defined parameter α is set to define relative importance between conflict number and stitch number. The ILP formulation for TPLD can be given in a similar way as that in DPLD [8]. We skip the detailed formulation here due to the page limit.

In addition to DPLD and TPLD problem, ILP is also widely used in MPL with DSA. In [10], an ILP based approach and a maximum matching based heuristic are proposed to solve the mask assignment problem in DSA. Ou *et al.* [11] apply ILP to assign cuts to different guiding templates, minimizing both conflict and line-end extension. Some speed up techniques are also proposed to improve the scalability of ILP. In [12], [13], ILP is used for simultaneous guiding template optimization and redundant via insertion for DSA.

B. Mathematical Relaxation

Relaxation is a significant approach to some optimization problems which are hard to solve. It can provide an upper bound or a lower bound on the optimal value of the original problem. Considering the layout decomposition is NP-hard which may suffer from runtime overhead, some relaxation techniques are proposed to tackle the issue efficiently.

In [8], a semidefinite programming (SDP) relaxation is proposed, which can be solved in polynomial time.

$$\min_{\mathbf{X}} \quad \mathbf{A} \bullet \mathbf{X}, \tag{2}$$

$$\text{s.t.} \quad \mathbf{X}_{ii} = 1, \qquad \forall i \in V, \tag{2a}$$

$$\mathbf{X}_{ij} \ge \frac{1}{2}, \qquad \forall e_{ij} \in CE, \tag{2b}$$

$$\mathbf{X} \succeq \mathbf{0}, \tag{2c}$$

where $\mathbf{X}_{ij}$ is the entry of the i-th row and j-th column of $\mathbf{X}$. Similarly, we let $\mathbf{A}_{ij}$ denote each entry in matrix $\mathbf{A}$, where $\mathbf{A}_{ij}$ is defined as follows.

$$\mathbf{A}_{ij} = \begin{cases} 1, & \forall e_{ij} \in CE, \\ -\alpha, & \forall e_{ij} \in SE, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Then the coloring solution may be extracted from the optimal solution $\mathbf{X}$. Essentially, if x_{ij} is close to 1, then vertices i and j tend to be in the same color; if x_{ij} is close to 0.5, vertices i and j tend to be in different colors. For those solutions whose x_{ij}s are vague, a partition-based mapping algorithm was used to perform the color assignment. SDP relaxation is also adopted in [14]–[16]. Note that [16] uses randomized rounding proposed in [17] to recover the final coloring solution.

Lin *et al.* [18] formulate the problem of TPLD for contact layer into an ILP. Note that instead of minimizing the total cost from conflicts, the target of our ILP formulation is to seek a feasible color assignment to the variables. Then the ILP formulation is relaxed to linear programming (LP) to avoid the infeasibility issue and find a solution with few conflicts. Since the solution of LP could be non-integer, additional constraints are introduced to prune these native non-integer solutions. For an odd cycle with five vertices each of which is denoted by two bits, if the first bits of the vertices are equal, i.e., $x_{i1} = x_{j1} = x_{k1} = x_{l1} = x_{m1} = 1$ or 0, it is not possible to obtain a solution without conflicts by adjusting the second bit. Instead, the LP relaxation will produce all 0.5 solutions to satisfy the constraints, which is undesired. To avoid the first bits or the second bits being equal,

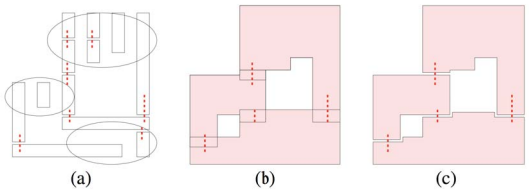


Fig. 5: Example of embedding the stitch graph in a plane [23]: (a) Segments and stitch arcs embedding; (b) Segments expansion; (c) The planar embedding of the stitch graph.

following constraints are added.

$$\begin{cases} x_{i1} + x_{j1} + x_{k1} + x_{l1} + x_{m1} \geq 1, \\ (1 - x_{i1}) + (1 - x_{j1}) + (1 - x_{k1}) + (1 - x_{l1}) + (1 - x_{m1}) \geq 1. \end{cases}$$

In order to push non-integer to integers, objective function keeps changing in original LP formulation in each iteration until no improvement is found.

Li *et al*. [19] propose a discrete relaxation method for TPLD problem. Firstly, the original TPLD problem is relaxed to an ILP by ignoring stitch insertion, whose optimal value can be treated as a lower bound of the optimal value of original TPLD problem. The ILP formulation of the relaxed problem has fewer variables and fewer constraints than the reduced version of [8]. After the relaxation solution is obtained, a feasible solution can be generated by a legalization process, including stitch insertion and backtrack coloring.

C. Satisfiability

Satisfiability (SAT) is a kind of problem which asks whether the variables of a given Boolean formula can be consistently replaced by the values *true* or *false* in such a way that the formula evaluates to *true*.

From the Fig. 2, it can be seen that the final feature is generated in the non-sidewall region which is also covered by trim mask. We can use a Boolean variable $Sidewall = TRUE$ to denote that a sidewall exists in one location and use $Trim = TRUE$ to denote that this location is covered by trim mask. Then the feature generation is expressed a boolean function as $Feature = \neg Sidewall \wedge Trim$.

SAT has been used for solving the layout decomposition problem in SADP [20], [21]. First the layout area is divided into tiles. Boolean variables C_i, S_i and T_i are used to represent whether the tile i is core, sidewall or trim mask, respectively. The featured and non-featured tiles can be expressed by the Boolean clauses as below.

$$F_i = \neg S_i \wedge T_i = TRUE, \tag{4a}$$

$$\neg F_i = S_i \vee \neg T_i = TRUE. \tag{4b}$$

Then, the design rules and geometry constraints are formulated as Boolean expressions in SAT. The objective is to determine whether there is a satisfiable assignment which corresponds to a valid decomposition.

Tian *et al*. [22] adopted SAT for TPLD problem of row-structure layout. Cell boundary constraints and cell inner constraints are captured by SAT clauses. Then coloring solutions of polygons within a cell are enumerated. A solution is a combination of coloring solution for different cells. Illegal combinations are forbidden by extra SAT clause.

D. Optimality for Special Situations

There are some special cases when handling the layout decomposition problem regarding the decomposition graph we derive and the layout structure.

Xu *et al*. [24] prove that the conflict graph that is used to model DPLD problem is planar graph. Then stitching and conflict elimination merge neighboring faces in conflict graph. In this work, a structure called *face graph* is proposed to model face merging. All odd nodes in face graph are paired up optimally to eliminate all odd cycles in

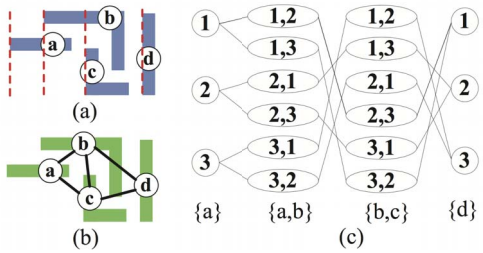


Fig. 6: Example given in [26]: (a) Input layout; (b) Constraint graph; (c) Solution graph.

conflict graph. The complexity of the algorithm is $\mathcal{O}(n^3)$. Tang *et al*. [23] compute a stitch graph from conflict graph and prove that the stitch graph is planar, which can be illustrated with Fig. 5. In (a), the segments and the stitch arcs are embedded in a plane, similar to the original layout. All the segments are expanded by an amount of half of the spacing threshold. Then the segments within one component are unioned into one polygon shape. From the planar embedding of the stitch graph it can be seen that merging the multiple stitch arcs between two components into one stitch arc is equivalent to removing some stitch arcs. And it will not change the property of planarity. Furthermore, they showed the min-cut in the stitch graph gives the decomposition solution to the original layout with the minimum number of stitches. The time complexity of the method for DPLD is $\mathcal{O}(n^{1.5} \log n)$, where n is the number of patterns.

In standard cell-based designs, pre-designed standard cells from a given library are used. A layout consists of multiple rows, and the cells are aligned with power and ground connecting each other in each row. For row-structure layout, the shortest path-based method is proposed in [25] and [26]. With pre-coloring, simple coloring solutions for local patterns are generated, based on which a solution graph is constructed. The shortest path in the solution graph corresponds to the decomposition solution. An example is shown in Fig. 6.

E. Search-based Approach

Kuang *et al*. [27] propose an efficient methodology for TPLD. A graph library is constructed with four, five or six nodes. Given a decomposition graph, some simplification techniques are used to divide it to sub-graphs which will be matched with the graphs in the library. If a subgraph is matched with a 3-colorable one, it can be colored easily. Otherwise, stitches are inserted. Fang *et al*. [28] propose a stitch-aware mask assignment algorithm. Similar to [27], graph reduction is performed first. Then, a heuristic method finds a mask assignment in which the conflicts in the same mask are more likely to be resolved by inserting stitches. Specifically, a weight is assigned to each conflict edge. The larger edge weight the edge has, the more difficult the conflict can be solved by stitch insertion. A modified recursive largest first algorithm is then used for mask assignment. A pairwise coloring (PWC) method was proposed by Zhang *et al*. [29] for solving the DPLD and TPLD. The problem is reduced to sets of bi-coloring problems. The overall solution is refined iteratively by applying a bi-coloring method for pairs of color sets. In [30], a branch-and-bound method is adopted to prune suboptimal nodes in search space. Apart from conventional coloring rules which are minimum spacing rules, there exist other complex coloring rules which make the problem slightly different from conventional one. Chang *et al*. [31] consider more complex coloring rules for TPLD problem. The graph coloring problem is reduced to an exact cover problem which can be solved by DLX. Yang *et al*. [32] apply hybrid lithography which consists of double patterning lithography and E-beam lithography (EBL) to complete TPLD. A planar primal-dual method is proposed for co-optimization of DPL and EBL.

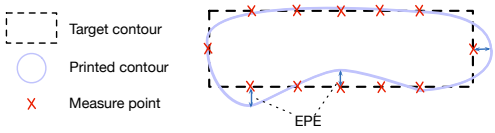


Fig. 7: Illustration of EPE measurement.

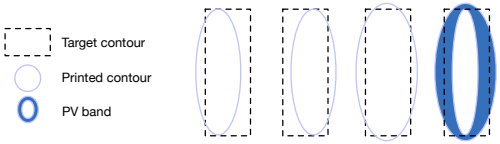


Fig. 8: Illustration of PV band.

A common concern regarding the quality of the layout decomposition solution is the pattern density in each mask. Generally, the desired solution is that the pattern density of each mask is balanced. Therefore, we should not only consider the global objectives such as conflict number and stitch number but also local objective such as pattern density. To do so, some partitioning-based methods are proposed. Yang *et al*. [33] study the density balanced issue of the DPLD problem. The color assignment is completed by ILP and stitch minimization is completed by a min-cut two-way partitioning. Yu *et al*. [14] define the pattern density uniformity locally and optimize conflict number, stitch number and local uniformity simultaneously using SDP relaxation. Different from greedy mapping used in [8], a three-way maximum-cut partitioning-based mapping is proposed to recover the decomposition solution from SDP solution. Chen *et al*. [34] address the balance issue by applying the strategy of maximizing minimum distance of patterns on each mask.

F. Fixed Parameter Tractability

Both DPLD and TPLD are NP-hard that cannot be solved in polynomial time. With fixed-parameter tractable (FPT) algorithm, the exponential runtime is confined to a parameter k, thus with small k values some NP-hard problems can be solved efficiently. Kuang *et al*. [35] apply FPT to three problems, including DPLD, DPLD + EBL hybrid lithography and DPLD + EBL + DSA hybrid lithography. These three problems can be reduced to odd cycle cover (OCC) problem which has FPT algorithms. An odd cycle cover of a graph is a set of vertices whose removal leaves a subgraph without odd cycle, and is therefore 2-colorable. Here we take the problem of DPLD + EBL as an example to show the basic idea. Given the layout, the decomposition graph is built as mentioned above. Each vertex is assigned a weight which is proportional to the area of corresponding pattern. If there is no stitch edge, then the decomposition graph is already an instance of WOCCP. When there exist potential stitches, each stitch edge is replaced by a virtual vertex and two virtual edges. A virtual vertex is in the OCC means the corresponding stitch exists in the decomposition solution.

III. Mask Optimization

Two models are needed to transform mask patterns into printed image: optical lithography model and photo resist model. First, an aerial image $\mathbf{I}$ is generated by convolving the mask $\mathbf{M}$ with a set of optical kernels [36], which is represented as

$$\mathbf{I} = f_{optical}(\mathbf{M}) = \sum_{k=1}^{K} w_k \cdot |\mathbf{M} \otimes \mathbf{h}_k|^2, \tag{5}$$

where $\mathbf{h}_k$ is the k-th optical kernel, w_k is the weight of $\mathbf{h}_k$, and K is the total kernel number.

Then a resist model is applied to the aerial image. In our work a constant threshold resist model is used, which sets an intensity threshold I_{th} to binarize the aerial image, denoted by $\mathbf{Z}$ in the following equation.

$$\mathbf{Z}(x,y) = f_{resist}(\mathbf{I}) = \begin{cases} 1, & \text{if } \mathbf{I}(x,y) \geq I_{th}, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Definition 3 (Edge Placement Error)**.** *Given a target layout and the printed image, the edge placement error (EPE) is defined as the geometric displacement of the image contour from the edge of target image on the layout. A violation is introduced if the perpendicular displacement is greater than an EPE threshold value.*

Definition 4 (Process Variation Band)**.** *The process variability band (PV Band) represents the variability in the position of the contour.,*

Usually, two images are generated at 2 corners which are high dose with nominal focus and low dose with defocus. The PV band is calculated by the XOR of the obtained images. An example of PV band is shown in Fig. 8.

A. Rule-Based OPC

Rule-based OPC requires comprehensive experiments determining design rules to compensate non-desired patterns, thus can only be applied to less aggressive designs. In [37], a rule-based method is proposed to address the issue of gate bridge by performing critical area correction. To search for the critical areas, polygons are generated at corresponding regions that are similar to error markers. Then by monitoring the space critical dimensions while varying the neighboring pattern width, it is easy to extract the OPC rules for bridge elimination.

B. Model-Based OPC

Model-based OPC segments pattern edges into small parts and moves them slightly to make correction for final patterns. However, it is heavily based on lithography simulation which is time-consuming. A regression model for OPC using a hierarchical Bayes model is proposed in [38], whose results can be used the starting point of conventional model-based OPC flow thus can effectively avoid the runtime overhead.

A model-based flow is presented by Awad *et al*. [39] for minimizing EPE and PV band. The whole flow is separated into two phases which target EPE minimization and PV band minimization, respectively. In the first phase, EPE is optimized by shifting two neighboring segments. Hammers are added on the corners of the polygons to improve the printability of the corners. In the second phase, sub-resolution assist features (SRAFs) are inserted to improve the PV band. The width of SRAF is determined by a regression model.

Kuang *et al*. propose a more robust approach in [40]. Different from [39], they perform mask optimization in three steps. The intensity difference is minimized first by iteratively moving the edge segments, during which most of the EPE can be fixed. Next, EPE minimization step conducts edge-moving operation more carefully on the spots where there still exists EPE violation. After that, PV band is minimized by slightly perturbing the width and length of the SRAFs near the segments which have large PV band.

Su *et al*. [41] develop a model-based OPC flow called PVOPC, where a novel dynamic edge fragmentation is used to form segment candidates for correction. To model the process variation, they select representative process corners with different weights, which is more efficient than using all process corners with uniform weight. A forward collision avoidance strategy is to prevent the image contour from open or short, which leads to fast EPE convergence.

C. Inverse Lithography Technique

Under a certain lithography model, inverse lithography technique (ILT) aims to find the ideal mask by solving an inverse problem of the lithography system. The objective of ILT is typically to minimize the difference between printed patterns of the mask and target patterns.

Different from model-based OPC which makes correction on edge segments, ILT is based on pixel-based representation. The ILT can be formulated as in Formula (7), where $\mathbf{Z}_t$ is the target image and $\mathbf{M}$ is the optimized mask.

$$\min_{\mathbf{M}} \quad F = \|\mathbf{Z}_t - \mathbf{Z}\|_2^2, \tag{7}$$

$$\text{s.t.} \quad \mathbf{M}(x,y) \in \{0,1\}, \quad \forall x, y, \tag{7a}$$

$$\mathbf{I} = \sum_{k=1}^{K} w_k \cdot |\mathbf{M} \otimes \mathbf{h}_k|^2, \tag{7b}$$

$$\mathbf{Z} = f_{resist}(\mathbf{I}), \tag{7c}$$

Many numerical optimization methods have been investigated for ILT. Poonawala *et al*. [42] adopt relax the problem to continuous form and solve it by gradient decent. The binary value constraint is first relax to bounded continuous constraint as $0 \leq \mathbf{M}(x,y) \leq 1$, which is further reduced to unconstrained as

$$\mathbf{M}(x,y) = \frac{1 + \cos(\theta(x,y))}{2}. \tag{8}$$

The resist model is relaxed through the $sigmoid$ function as

$$\mathbf{Z}(x,y) = \text{sig}(\mathbf{I}(x,y)) = \frac{1}{1 + \exp[-\theta_Z(\mathbf{I}(x,y) - I_{th})]}, \tag{9}$$

where θ_Z is a user-defined parameter which represents the steepness of $sigmoid$ function, and I_{th} is the threshold in the resist model. Then gradient descent algorithm is used to solve the problem iteratively.

In addition to conventional gradient descent algorithm, stochastic gradient descent (SGD) is also used in ILT [43]. SGD is a variant of gradient descent, which also relies on gradient for optimization. However, the true gradient is approximated by the gradient in a single example. In [43], the process variation is considered. In order to model that, the final printed image is obtained by averaging the intensity of different process corners, which requires more computation when calculating the gradient. With SGD the whole process is accelerated significantly and achieves comparable quality of final masks. Another acceleration technique for ILT is proposed in [44]. Considering that the forward lithography model can be formulated by a series of weighted sum of convolution, the "effective kernel" can be precomputed without loss of accuracy.

$$\sum_{k=1}^{K} w_k \cdot (\mathbf{M} \otimes \mathbf{h}_k) = \sum_{k=1}^{K} \mathbf{M} \otimes (w_k \cdot \mathbf{h}_k) = \mathbf{M} \otimes \sum_{k=1}^{K} w_k \cdot \mathbf{h}_k. \tag{10}$$

What's more, a precise EPE modeling formulation is proposed in [44] which is also differentiable so that gradient descent method can be applied.

Apart from widely used gradient-based optimization technique, level-set-based method has also been studied. Shen *et al*. [45] shows that ILT can be modeled into an image restoration problem and can be solved by level-set method where the boundary of the pattern is iteratively evolved. The mask $\mathbf{M}$ is given a level-set description associated with an unknown function $\phi(x)$ as follows.

$$\mathbf{M}(x,y) = \begin{cases} \mathbf{M}_{int}, & \text{if } \phi(x,y) < 0, \\ \mathbf{M}_{ext}, & \text{otherwise.} \end{cases} \tag{11}$$

Then the boundary of the patterns is governed by the zero level set $\phi(x,y) = 0$. To get the optimal mask $\mathbf{M}$, an appropriate difference scheme is needed. First-order accurate method is used in [45] and the conjugate gradient method is used in [46].

IV. Simultaneous Layout Decomposition and Mask Optimization

It has been seen that layout decomposition and mask optimization are powerful techniques for improving manufacturability by generating optimized masks in two stages. It is intuitive to think that how to combine these two techniques, hoping that the unified flow can be even more powerful. Li *et al*. provided a model-based methodology for single mask OPC which provides several techniques for improving printability, including ensuring sufficient overlap and compensating the overlay [47]. However, the proposed methodology is still applied to two-stage flow. Previous work has presented how to combine ILT with double exposure lithography (DEL) [48], [49]. Compared with general ILT for single mask optimization, two masks are generated in DEL-ILT flow. The final aerial image is equal to the sum of the aerial images obtained from the two individual exposures, which can be formulated as Equation (12) where $\mathbf{M}_1$ and $\mathbf{M}_2$ are two masks.

$$\mathbf{I} = f_{optical}(\mathbf{M}_1) + f_{optical}(\mathbf{M}_2). \tag{12}$$

Basically, this problem is solved by gradient descent method [48], [49]. Besides, cyclic coordinate descent is also introduced as an alternative optimization method [48]. Cyclic coordinate descent optimizes the cost function with respect to only one parameter at a time, and sequentially covering up the whole parameter vector. It is not as accurate as gradient descent in each iteration but it is more efficient. However, none of these work addresses the layout decomposition problem because they only consider the multiple exposures on a single mask.

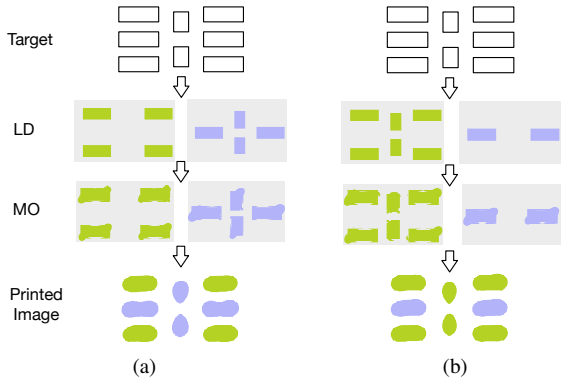


Fig. 9: Same quality layout decompositions (LD) achieve different EPE violation number after mask optimization (MO) [50]: (a) Solution 1 with #EPE violation = 3; (b) Solution 2 with #EPE violation = 1.

The issue of two-stage flow is the inconsistency. Fig. 9 gives an example on such situation. Given the identical target, two different layout decomposition results are found (LD stage in the figures), and both of them satisfy all design rules and coloring rules. After the mask optimization (MO stage in the figures) on each mask, however, it can be observed that the qualities of the printed images are diverse: Fig. 9(a) has three EPE violations, while Fig. 9(b) has only one EPE violation. To address this issue, a unified optimization framework for simultaneous layout decomposition and mask optimization is proposed by Ma *et al*. [50]. A numerical optimization flow and a discrete optimization flow are designed to solve the problem. These two engines are collaborative with each other. The experimental result shows that the proposed framework can find higher quality solutions much more efficiently than conventional two-stage flow.

V. Conclusion

In this paper we have surveyed some commonly used layout decomposition and mask optimization techniques. We first introduce the basic idea of multiple patterning layout decomposition and mask optimization. For each technique, we select some representative previous work and illustrate and compare the ideas behind. Since conventional two-stage flow cannot guarantee the optimality of the printed image, we also introduce some works which try to unify the layout decomposition and

mask optimization stage. As the feature size continues scaling, advanced lithography techniques will be of great significance for improving the manufacturability. We hope this paper will stimulate more systematic studies on layout decomposition and mask optimization.

REFERENCES

[1] D. Z. Pan, B. Yu, and J.-R. Gao, "Design for manufacturing with emerging nanolithography," *IEEE TCAD*, vol. 32, no. 10, pp. 1453–1472, 2013.

[2] D. Z. Pan, L. Liebmann, B. Yu, X. Xu, and Y. Lin, "Pushing multiple patterning in sub-10nm: Are we ready?" in *Proc. DAC*, 2015, pp. 197:1–197:6.

[3] F. G. Pikus and A. Torres, "Advanced multi-patterning and hybrid lithography techniques," in *Proc. ASPDAC*, 2016, pp. 611–616.

[4] B. Yu, X. Xu, S. Roy, Y. Lin, J. Ou, and D. Z. Pan, "Design for manufacturability and reliability in extreme-scaling VLSI," *Science China Information Sciences*, pp. 1–23, 2016.

[5] A. B. Kahng, C.-H. Park, X. Xu, and H. Yao, "Layout decomposition approaches for double patterning lithography," *IEEE TCAD*, vol. 29, pp. 939–952, June 2010.

[6] Y. Xu and C. Chu, "GREMA: graph reduction based efficient mask assignment for double patterning technology," in *Proc. ICCAD*, 2009, pp. 601–606.

[7] K. Yuan, J.-S. Yang, and D. Z. Pan, "Double patterning layout decomposition for simultaneous conflict and stitch minimization," *IEEE TCAD*, vol. 29, no. 2, pp. 185–196, Feb. 2010.

[8] B. Yu, K. Yuan, D. Ding, and D. Z. Pan, "Layout decomposition for triple patterning lithography," *IEEE TCAD*, vol. 34, no. 3, pp. 433–446, March 2015.

[9] B. Yu, S. Roy, J.-R. Gao, and D. Z. Pan, "Triple patterning lithography layout decomposition using end-cutting," *JM3*, vol. 14, no. 1, pp. 011 002–011 002, 2015.

[10] Y. Badr, A. Torres, and P. Gupta, "Mask assignment and synthesis of DSA-MP hybrid lithography for sub-7nm contacts/vias," in *Proc. DAC*, 2015, pp. 70:1–70:6.

[11] J. Ou, B. Yu, J.-R. Gao, and D. Z. Pan, "Directed self-assembly cut mask assignment for unidirectional design," *JM3*, vol. 14, no. 3, 2015.

[12] S.-Y. Fang, Y.-X. Hong, and Y.-Z. Lu, "Simultaneous guiding template optimization and redundant via insertion for directed self-assembly," in *Proc. ICCAD*, 2015, pp. 410–417.

[13] J. Ou, B. Yu, and D. Z. Pan, "Concurrent guiding template assignment and redundant via insertion for DSA-MP hybrid lithography," in *Proc. ISPD*, 2016, pp. 39–46.

[14] B. Yu, Y.-H. Lin, G. Luk-Pat, D. Ding, K. Lucas, and D. Z. Pan, "A high-performance triple patterning layout decomposer with balanced density," in *Proc. ICCAD*, 2013, pp. 163–169.

[15] B. Yu and D. Z. Pan, "Layout decomposition for quadruple patterning lithography and beyond," in *Proc. DAC*, 2014, pp. 53:1–53:6.

[16] T. Matsui, Y. Kohira, C. Kodama, and A. Takahashi, "Positive semidefinite relaxation and approximation algorithm for triple patterning lithography," in *International Symposium on Algorithms and Computation*. Springer, 2014, pp. 365–375.

[17] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[18] Y. Lin, X. Xu, B. Yu, R. Baldick, and D. Z. Pan, "Triple/quadruple patterning layout decomposition via novel linear programming and iterative rounding," in *SPIE Advanced Lithography*, vol. 9781, 2016.

[19] X. Li, Z. Zhu, and W. Zhu, "Discrete relaxation method for triple patterning lithography layout decomposition," *IEEE Transactions on Computers*, vol. 66, no. 2, pp. 285–298, 2017.

[20] H. Zhang, Y. Du, M. D. Wong, R. Topaloglu, and W. Conley, "Effective decomposition algorithm for self-aligned double patterning lithography," in *Proc. SPIE*, vol. 7973, 2011, p. 79730J.

[21] Z. Xiao, H. Zhang, Y. Du, and M. D. Wong, "A polynomial time exact algorithm for self-aligned double patterning layout decomposition," in *Proc. ISPD*, 2012, pp. 17–24.

[22] H. Tian, Y. Du, H. Zhang, Z. Xiao, and M. D. F. Wong, "Constrained pattern assignment for standard cell based triple patterning lithography," in *Proc. ICCAD*, 2013, pp. 178–185.

[23] X. Tang and M. Cho, "Optimal layout decomposition for double patterning technology," in *Proc. ICCAD*, 2011, pp. 9–13.

[24] Y. Xu and C. Chu, "A matching based decomposer for double patterning lithography," in *Proc. ISPD*, 2010, pp. 121–126.

[25] H.-A. Chien, S.-Y. Han, Y.-H. Chen, and T.-C. Wang, "A cell-based row-structure layout decomposer for triple patterning lithography," in *Proc. ISPD*, 2015, pp. 67–74.

[26] H. Tian, H. Zhang, Q. Ma, Z. Xiao, and M. D. F. Wong, "A polynomial time triple patterning algorithm for cell based row-structure layout," in *Proc. ICCAD*, 2012, pp. 57–64.

[27] J. Kuang and E. F. Y. Young, "An efficient layout decomposition approach for triple patterning lithography," in *Proc. DAC*, 2013, pp. 69:1–69:6.

[28] S.-Y. Fang, Y.-W. Chang, and W.-Y. Chen, "A novel layout decomposition algorithm for triple patterning lithography," *IEEE TCAD*, vol. 33, no. 3, pp. 397–408, March 2014.

[29] Y. Zhang, W.-S. Luk, H. Zhou, C. Yan, and X. Zeng, "Layout decomposition with pairwise coloring for multiple patterning lithography," in *Proc. ICCAD*, 2013, pp. 170–177.

[30] B. Yu, X. Xu, J.-R. Gao, Y. Lin, Z. Li, C. Alpert, and D. Z. Pan, "Methodology for standard cell compliance and detailed placement for triple patterning lithography," *IEEE TCAD*, vol. 34, no. 5, pp. 726–739, May 2015.

[31] H.-Y. Chang and I. H.-R. Jiang, "Multiple patterning layout decomposition considering complex coloring rules," in *Proc. DAC*, 2016, pp. 40:1–40:6.

[32] Y. Yang, W.-S. Luk, D. Z. Pan, H. Zhou, C. Yan, D. Zhou, and X. Zeng, "Layout decomposition co-optimization for hybrid e-beam and multiple patterning lithography," *IEEE TCAD*, vol. 35, no. 9, pp. 1532–1545, 2016.

[33] J.-S. Yang, K. Lu, M. Cho, K. Yuan, and D. Z. Pan, "A new graph-theoretic, multi-objective layout decomposition framework for double patterning lithography," in *Proc. ASPDAC*, 2010, pp. 637–644.

[34] Z. Chen, H. Yao, and Y. Cai, "SUALD: Spacing uniformity-aware layout decomposition in triple patterning lithography." in *Proc. ISQED*, 2013, pp. 566–571.

[35] J. Kuang and E. F. Y. Young, "Fixed-parameter tractable algorithms for optimal layout decomposition and beyond," in *Proc. DAC*, 2017, pp. 61:1–61:6.

[36] N. B. Cobb, "Fast optical and process proximity correction algorithms for integrated circuit manufacturing," Ph.D. dissertation, University of California at Berkeley, 1998.

[37] J.-S. Park, C.-H. Park, S.-U. Rhie, Y.-H. Kim, M.-H. Yoo, J.-T. Kong, H.-W. Kim, and S.-I. Yoo, "An efficient rule-based OPC approach using a DRC tool for 0.18 μm ASIC," in *Proc. ISQED*, 2000, pp. 81–85.

[38] T. Matsunawa, B. Yu, and D. Z. Pan, "Optical proximity correction with hierarchical bayes model," *JM3*, vol. 15, no. 2, p. 021009, 2016.

[39] A. Awad, A. Takahashi, S. Tanaka, and C. Kodama, "A fast process variation and pattern fidelity aware mask optimization algorithm," in *Proc. ICCAD*, 2014, pp. 238–245.

[40] J. Kuang, W.-K. Chow, and E. F. Y. Young, "A robust approach for process variation aware mask optimization," in *Proc. DATE*, 2015, pp. 1591–1594.

[41] Y.-H. Su, Y.-C. Huang, L.-C. Tsai, Y.-W. Chang, and S. Banerjee, "Fast lithographic mask optimization considering process variation," *IEEE TCAD*, vol. 35, no. 8, pp. 1345–1357, 2016.

[42] A. Poonawala and P. Milanfar, "Mask design for optical microlithography–an inverse imaging problem," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 774–788, 2007.

[43] N. Jia and E. Y. Lam, "Machine learning for inverse lithography: using stochastic gradient descent for robust photomask synthesis," *Journal of Optics*, vol. 12, no. 4, pp. 045 601:1–045 601:9, 2010.

[44] J.-R. Gao, X. Xu, B. Yu, and D. Z. Pan, "MOSAIC: Mask optimizing solution with process window aware inverse correction," in *Proc. DAC*, 2014, pp. 52:1–52:6.

[45] Y. Shen, N. Wong, and E. Y. Lam, "Level-set-based inverse lithography for photomask synthesis," *Optics Express*, vol. 17, no. 26, pp. 23 690–23 701, Dec 2009.

[46] W. Lv, S. Liu, Q. Xia, X. Wu, Y. Shen, and E. Y. Lam, "Level-set-based inverse lithography for mask synthesis using the conjugate gradient and an optimal time step," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 31, no. 4, p. 041605, 2013.

[47] X. Li, G. Luk-Pat, C. Cork, L. Barnes, and K. Lucas, "Double-patterning-friendly OPC," in *Proc. SPIE*, vol. 7274, 2009.

[48] A. Poonawala and P. Milanfar, "Double-exposure mask synthesis using inverse lithography," *JM3*, vol. 6, no. 4, pp. 043 001–043 001, 2007.

[49] S. Banerjee, K. B. Agarwal, and M. Orshansky, "Simultaneous OPC and decomposition for double exposure lithography," in *Proc. SPIE*, vol. 7973, 2011.

[50] Y. Ma, J.-R. Gao, J. Kuang, J. Miao, and B. Yu, "A unified framework for simultaneous layout decomposition and mask optimization," in *Proc. ICCAD*, 2017.