EDA Solutions for Double Patterning Lithography

by

Minoo Mirsaeedi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Expanding the optical lithography to 32-nm node and beyond is impossible using existing single exposure systems. As such, double patterning lithography (DPL) is the most promising option to generate the required lithography resolution, where the target layout is printed with two separate imaging processes. Among different DPL techniques litho-etch-litho-etch (LELE) and self-aligned double patterning (SADP) methods are the most popular ones, which apply two complete exposure lithography steps and an exposure lithography followed by a chemical imaging process, respectively.

To realize double patterning lithography, patterns located within a sub-resolution distance should be assigned to either of the imaging sub-processes, so-called layout decomposition. To achieve the optimal design yield, layout decomposition problem should be solved with respect to characteristics and limitations of the applied DPL method. For example, although patterns can be split between the two sub-masks in the LELE method to generate conflict free masks, this pattern split is not favorable due to its sensitivity to lithography imperfections such as the overlay error. On the other hand, pattern split is forbidden in SADP method because it results in non-resolvable gap failures in the final image. In addition to the functional yield, layout decomposition affects parametric yield of the designs printed by double patterning.

To deal with both functional and parametric challenges of DPL in dense and large layouts, EDA solutions for DPL are addressed in this thesis. To this end, we proposed a statistical method to determine the interconnect width and space for the LELE method under the effect of random overlay error. In addition to yield maximization and achieving near-optimal trade-off between different parametric requirements, the proposed method provides valuable insight about the trend of parametric and functional yields in future technology nodes.

Next, we focused on self-aligned double patterning and proposed layout design and decomposition methods to provide SADP-compatible layouts and litho-friendly decomposed layouts. Precisely, a grid-based ILP formulation of SADP decomposition was proposed to avoid decomposition conflicts and improve overall printability of layout patterns. To overcome the limited applicability of this ILP-based method to fully-decomposable layouts, a partitioning-based method is also proposed which is faster than the grid-based ILP decomposition method too. Moreover, an A*-based SADP-aware detailed routing method was proposed which performs detailed routing and layout decomposition simultaneously to avoid litho-limited layout configurations. The proposed router preserves the uniformity of pattern density between the two sub-masks of the SADP process. We finally extended

our decomposition method for double patterning to triple patterning and formulated SATP decomposition by integer linear programming. In addition to conventional minimum width and spacing constraints, the proposed decomposition method minimizes the mandrel-trim co-defined edges and maximizes the layout features printed by structural spacers to achieve the minimum pattern distortion.

This thesis is one of the very early researches that investigates the concept of litho-friendliness in SADP-aware layout design and decomposition. Provided by experimental results, the proposed methods advance prior state-of-the-art algorithms in various aspects. Precisely, the suggested SADP decomposition methods improve total length of sensitive trim edges, total EPE and overall printability of attempted designs. Additionally, our SADP-detailed routing method provides SADP-decomposable layouts in which trim patterns are highly robust to lithography imperfections. The experimental results for SATP-decomposition show that total length of overlay-sensitive layout patterns, total EPE and overall printability of the attempted designs are also improved considerably by the proposed decomposition method. Additionally, the methods in this PhD thesis reveal several insights for the upcoming technology nodes which can be considered for improving the manufacturability of these nodes.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Mohab Anis for his guidance and support during the course of my studies, particularly during the most challenging times. Also, I would like to thank my co-supervisor Prof. Karim Karim for his help and support. I also would like to thank Prof. Laleh Behjat, Prof. Siddharth Garg, Prof. Bo Cui, and Prof. Eihab Abdel-Rahman for reviewing this work as well as their valuable comments to make the thesis come out in this current form.

As part of this thesis, I was fortunate to work with wonderful people in litho-friendly design group of Mentor Graphics. I am extremely grateful to Dr. Andres Torres for his valuable support, discussion, and comments. I also wish to thank Dr. Javid Jaffari and other members of UW VLSI research group for providing a friendly and joyful environment and for their helpful comments and discussions.

I am also extremely grateful to my M.Sc. supervisor during my studies at the Amirkabir University of Technology, Prof. Morteza Saheb-Zamani, for his endless kindness and support. Additionally, I want to thank the University of Waterloo administration for providing me with the opportunity to pursue my Ph.D. studies. Particularly, I would like to express my thanks to Wendy Boles, Susan King, and Annette Dietrich, of ECE Graduate Office. I would like to acknowledge the support of Ontario Graduate Scholarship (OGS) while I was an international student.

I am always grateful to my husband, Mehdi Saeedi, for his invaluable support and inspiration. His love and understanding played a major role in helping me finish this thesis. I also, would like to thank my family for their endless love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations

Today, the semiconductor industry has been aggressively moving to print ever-decreasing feature size, imposing tremendous pressure on lithography process to extend its present capability. For the past few years, lithography technology for 32-nm half-pitch node and below has been under consideration. By 2018, volume manufacturing of the 14- to 22-nm half-pitch nodes will be needed [16].

The resolution limit of optical lithography is often discussed in the context of the famous Raleigh criterion defined as (1.1).

$$CD = K_1 \frac{\lambda}{NA}, \tag{1.1}$$

where CD, λ, and NA represent the critical dimension, the wavelength, and the numerical aperture of optical system, respectively. The parameter K_1 depends on the process specifications.

To extend lithography resolution, the parameter NA should be increased, and parameters K_1 and λ should be decreased. In addition to using more powerful lenses, the reticle system is immersed in a high-index fluid environment to increases the numerical aperture (NA). Using water as an immersion fluid at 193-nm wavelength has recently emerged as the industry consent for the next major optical lithography extension [17]. The aperture index of water; i.e. 1.44, improves the numerical aperture of an immersed lens to 1.3 [18] and realizes the 45-nm half-pitch node [19]. The spatial response frequency of water-based

systems is limited to 37nm. Moreover, imaging quality degrades in higher NAs because of the lower depth-of-focus (DoF) as:

$$DoF = K_2 \frac{\lambda}{NA^2}. \tag{1.2}$$

Therefore, alternative techniques are needed to enhance the lithography resolution. Low-wavelength lithography methods, e.g. EUV and e-beam lithography methods, have been considered as the primary option for imaging in the 32-nm node. However, they are still lagging behind because of technical development challenges and high cost.

Decreasing the parameter K_1 is another way to improve the lithography resolution. Resolution enhancement techniques (RETs) including PSM[1], OAI[2], and OPC[3] lower the K_1 parameter [1]. These RETs are not mutually exclusive and can be combined to address different image distortion issues. For a single photo-resist layer, the spatial frequency of the optical system is theoretically limited to $2NA/\lambda$, which limits the K_1 parameter to 0.25. In reality, the finest resolution of current lithography systems, which take the advantage of a water immersion lens, a 193-nm light source, OAI, and PSM is 45 nm which is very far from our objective. In other words, existing lithography systems cannot achieve K_1 values lower than 0.3. Therefore, innovative lithography methods are needed to further improve the K_1 parameter and thus CD.

Multiple patterning lithography is an innovative method in which several patterning process are applied to print sub-resolution layout patterns. The applied patterning process can be litho-based, chemical, or mixed litho-chemical ones. Double patterning lithography (DPL) is the simplest forms of multiple patterning which is known as the most promising solution to realize 32-nm half-pitch node [20, 21]. Early test results have demonstrated that key technical challenges of DPL such as tight overlay budget and strict CD-control requirements are solvable. It is widely believed that DPL and higher degrees of multiple patterning can extend the lithography resolution down to 10-nm [16].

Theoretically, double patterning can double the resolution of 193-nm lithography. This means that process factor of K_1 in the Rayleigh equation can be half of the current value. In 193-nm lithography, double patterning has demonstrated a K_1 factor of 0.16, which is beyond both theoretical and practical limits; 0.25 and 0.3 respectively. Therefore, double patterning could provide resolution for a 45 nm half-pitch node using a *0.93-NA* scanner and could resolve 32-nm half pitch using a *1.3-NA 193i* tool [3]. In other words, DPL in

[1]Phase shift masking
[2]Off-axis illumination
[3]Optical proximity correction

combination with 193 nm immersion provides a solution to fill the gap between projection and EUV lithography methods.

There are three popular DPL processes, namely litho-etch-litho-etch (LELE) [22], litho-freeze-litho-etch (LFLE) [23], and self-aligned double patterning (SADP) [24]. LELE method consists of two complete lithography steps each one followed by an etching process. Compared with other DPL methods, LELE is simpler but needs more accurate overlay control [8]. LFLE freezes the latent image of the first exposure, then adds a second resist layer immediately on top for the second exposure. The resist pattern is etched only once after both resists are developed. Although LFLE uses fewer process steps, its required processes are complex and not yet mature. The basic SADP method [24] is similar to the LELE method in patterning steps till the end of the first litho phase. Afterwards, a chemical process replaces the second exposure phase to form required spaces around the patterns of the first exposure and consequently the second group of patterns. SADP is fairly robust against overlay error but only applicable for patterning regular patterns [24]. The basic SADP process followed by a litho-based trimming step may be applied for two-dimensional non-regular layouts [25].

A lot of investigations have been done recently to study different aspects of DPL. For examples, see [10, 26] for layout decomposition methods, [27, 28] for lithography simulation and modeling approaches, and [29] for resolution enhancement schema. However, the state-of-the-art DPL research suffers from serious drawbacks. The proposed decomposition methods have not been fit in physical design flow yet because their large time complexity makes them inapplicable for large layouts. Moreover, there is not an effective collaboration between the proposed lithography models and layout decomposition methods. Therefore, DPL still faces serious concerns especially for high manufacturing cost and demand for high levels of accuracy. In addition to the proposed models, automated techniques for DPL design are very immature.

Motivated by the above facts, automated techniques which deal with analysis and design for multiple patterning lithography are targeted in this thesis. After studying the limitations of LELE methods, we focus on self-aligned double patterning and propose layout design and decomposition methods to provide SADP-compatible layouts and litho-friendly decomposed layouts.

1.2 Structure of the thesis

The basic concepts of traditional RETs, e.g. OPC, PSM, and OAI, as well as principles of double-patterning lithography are reviewed in Chapter 1.2. Chapter 3 outlines the state-

of-the-art DPL methods, divided into decomposition, modeling, and DFM methods, to address different challenges of DPL. Figure 1.1 represents which levels of the DPL-specific physical design flow are studied in this thesis. In Chapter 4, the impacts of overlay error on DPL interconnects are studied and a statistical optimization method is proposed to improve the design yield in the presence of overlay variability. The litho-friendliness requirements of SADP are discussed in Chapter 5, where two litho-friendly decomposition methods are proposed. In Chapter 6, the litho-friendliness and SADP-compatibility requirements are considered to develop an SADP-friendly detailed routing algorithm. In Chapter 7, an ILP-based decomposition method is proposed for self-aligned triple patterning (SATP).

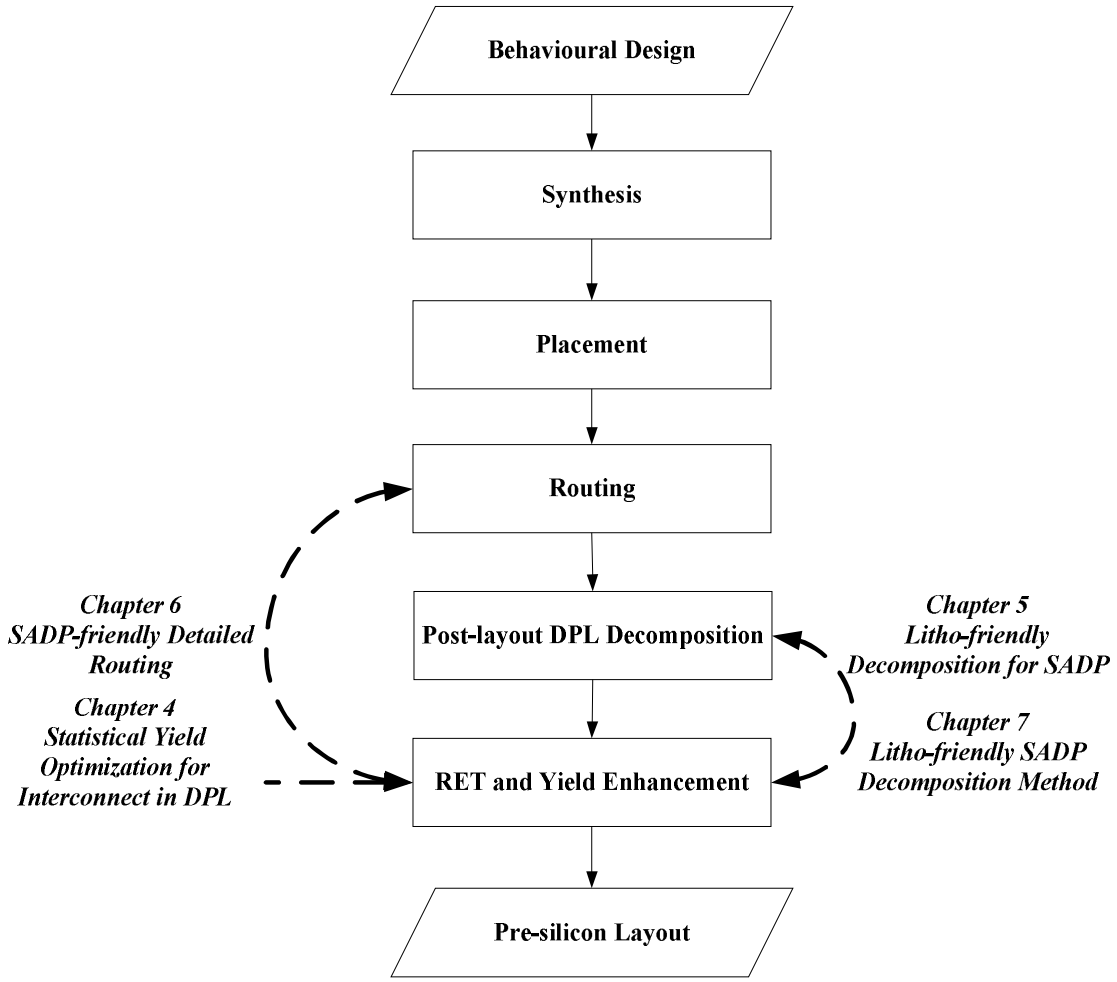


Figure 1.1: Different levels of physical design flow studied in this thesis.

Chapter 2

Basic Concepts of Double Patterning

With pattern dimension decreasing below the illumination wavelength, image quality is degrading rapidly because of diffraction. Accordingly, pattern sizes become increasingly sensitive to fluctuations of the fabrication process and critical dimension control is of paramount importance. Necessary process-level enhancements include reduction of mask CD error with reasonable cost increase, minimization of flare variation across the exposure field, and mitigating aberration in exposure systems. Well-controlled aberrations also benefits pattern placement. Moreover, improved stage precision and alignment systems will be needed to meet the ever stringent overlay budget.

The methods aiming enhancement the raw resolution capability of optical lithography are only part of the task. Despite considerable advancements in the mentioned process control feature, degradation of image quality is becoming more critical as design-level demands are growing faster than lithography process-level advancements. These image imperfections also contribute to line-edge roughness (LER), resulting in poor pattern transfer. The LER phenomenon would result in functional errors; e.g. gap or pinching failures along wires [1], and parametric errors, e.g. performance degradation and leakage increase in transistor channels [30].

Therefore, design of litho-friendly layouts is highly demanded to continue the Moor's law into below 65-nm technology nodes [16]. Litho-friendly design refers to all design techniques that result in mask features print with a low degree of difficulty. The litho-friendly design methods can be performed in different levels of abstraction, from device and circuit design to post-layout mask modifications [31]. However, layout-level litho-friendly design methods are the most effective ones because of accurate knowledge of underlying lithography process and the sources of image imperfection in this level of abstraction. The

layout-level litho-friendly design methods are referred as resolution enhancement techniques (RET), in general.

In this chapter, basic concepts of resolution enhancement techniques are discussed. To this end, conventional methods of RET are introduced briefly. As mentioned before, since conventional RETs do not address all current demands of design technology, innovative lithography methods are under investigation. Thus, we focus on different methods and current challenges of DPL, which is becoming the mainstream of lithography.

2.1 Resolution enhancement techniques

The wavelength of exposure light (λ), numerical aperture of the projection system (NA), and the measurement parameter of lithography aggressiveness (K_1) affect the lithography resolution. As discussed before, K_1 smaller than 0.25 is needed to realize half-pitch nodes under 45nm. However, traditional lithography systems achieve acceptable yield for K_1 greater than 0.75, in practice. For lower K_1 values resolution enhancement techniques are required to improve the imaging quality.

Major RETs are modified illumination schemes, such as annular and quadrupole illumination; and photo-mask techniques, such as optical proximity correction (OPC) and phase shifting mask (PSM). In addition, many other methods such as pupil filtering [32], multiple exposure [33], and top surface imaging [34] are under investigation. These RET methods can possibly reduce K_1 from 0.75 to 0.35, representing more than 50% decrease in CD compared with traditional lithography [1].

2.1.1 Modified illumination

Modified illumination was unnecessary when device CDs were large compared with lithography pitch, $\frac{\lambda}{NA}$. However, when the light strikes in a photo-mask with small pitch, it is diffracted aggressively. If the incident light is parallel to the axis of the optical system, the zero-th diffracted order remains parallel to the axis while the other orders are diffracted sideways. This deviation increases as the pitch and coherency factor (σ) of the light decrease. For sufficiently small pitches, only the zero-th diffraction order passes through the projection lens while others are lost and no pattern is created on the wafer consequently. Therefore, for imaging at $K_1 \leq 0.75$, the light source should be adjusted for a successful lithography.

Using large partial coherency factor (σ_{max}) is an effective way to print dense patterns. However, large σ in general diminishes process latitude for sparse patterns. Moreover, it causes contrast degradation even for dense patterns because a large fraction of illuminated light adds only to the background, which only contains the average of photo-mask pattern, rather than contributing the image formation. The amount of background light can be reduced by OAI. OAI is an optical system set up in which the exposure light strikes the photo-mask at an oblique angle rather than perpendicularly [1]. By making the illumination off-axis, all diffraction orders are titled, which makes it more likely that higher diffraction orders pass through the projection lens and contribute to image formation on the wafer. Figure 2.1 presents the major types of light-source aperture.

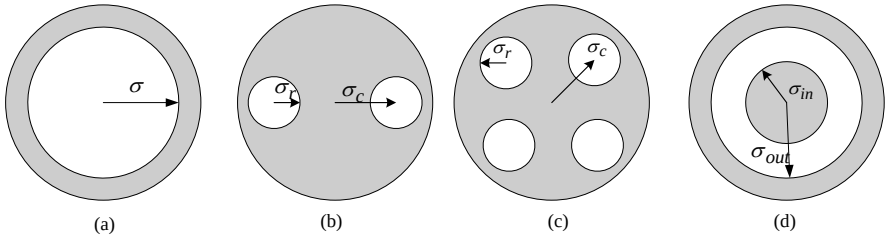


Figure 2.1: Major types of light-source aperture and their parameters: (a) simple circular aperture, (b) circular dipole, (c) circular quadrupole, (d) annular aperture.

In dipole OAI, see Figure 2.1.(b), the depth of focus (DoF) can be increased by setting up a two-beam interference [35]. The optimal locations of poles (i.e. σ_c) where two beams interfere constructively, depend on pitch, wavelength and numerical aperture as shown in (2.1). Moreover, σ_r is preferred to be as small as possible.

$$\sigma_{c,opt} = \frac{\lambda}{2P.NA}. \tag{2.1}$$

However, dipole OAI is poor in printing patterns perpendicular to the poles. Therefore, quadrupole illumination (see Figure 2.1.(c) was proposed to improve DoF in both directions [36]. An optimal aperture of this type has

$$\sigma_{c,opt} = \frac{\lambda}{2P.NA.Cos45^\circ}, \tag{2.2}$$

with σ_r made as small as possible. Although quadrupole OAI provides symmetry in x and y directions, it is not radially symmetrical. Annular OAI is used in applications where circular symmetry is critical. An annular source, shown in Figure 2.1.(d), is characterized

by two parameters: outer annuls σ_{out} and inner annuls σ_{out}. The penalty for resorting radial symmetry is the reintroduction of some DC background in the imaging of one-dimensional dense patterns.

2.1.2 Optical proximity correction

OPC is the technique of predistorting mask patterns so that the printed patterns are as close to the desired shapes as possible. Figure 2.2 shows different image distortions addressed by OPC including proximity effect, line shortening, and corner rounding.

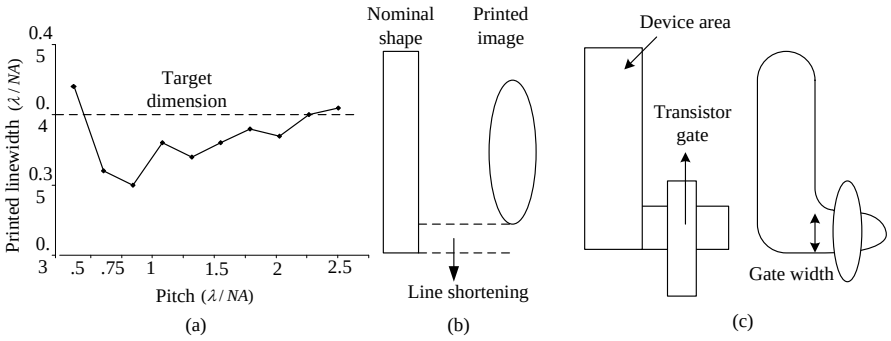


Figure 2.2: Various types of image distortion: (a) proximity effect, (b) line shortening, (c) corner rounding [1].

Proximity effect refers to features with the same CD printing differently due to environment variation, which results in across-chip line-width variation (ACLV). Line shortening, another form of image distortion, occurs mainly due to diffraction at low K_1 imaging. As the CD decreases, line shortening increases dramatically as shown in Figure 2.3. This behavior is very challenging because of its effect on overlay control and circuit density. The third type of error is corner rounding where sharp corners are filtered out by pupil. As shown in Figure 2.2.(c), where the effective channel width degrades due to rounding the elbow corner, corner rounding can cause functional and parametric errors.

Since line width variation occurs mainly due to the proximity errors, OPC resolves the problem by widening a line depending on its nominal dimension and environment, called selective line biasing. For example, inspired by lithography simulation results, dense and sparse lines are made smaller on the mask whereas lines of intermediate periodicities are made larger by OPC. This solution is direct and straightforward; however, its accuracy is

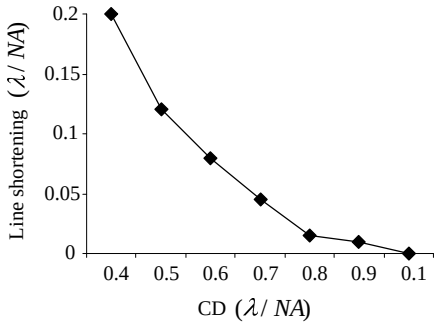


Figure 2.3: Line shortening increases dramatically with CD decreasing.

limited by the pixel size and further improvements need the pixel size reduction, which is unfavorable due to its overhead on the mask fabrication time.

Since OAI improves the imaging quality for dense patterns, extra features can be inserted in both sides of a line to create a dense environment, as shown in Figure 2.4 [2]. These assist-features are smaller than the main line such that they are not printed onto the wafer. Therefore, while the line appears dense to the projection system, it is printed as a sparse line. However, assist-features should be optimized carefully in terms of number, size, and placement of features.

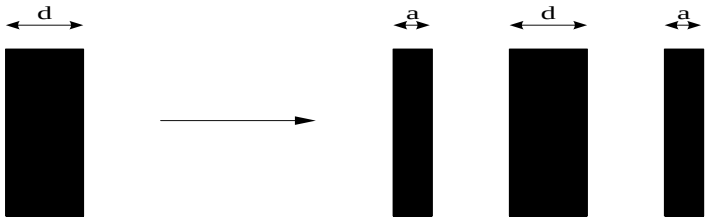


Figure 2.4: Using assist-features and OAI to improves process latitude of sparse features [2].

The simplest and most effective solution of line shortening is its lengthening, as shown in Figure 2.5. However, line lengthening is usually impossible due to very high layout density. In such cases, using serifs, hammer heads, and assist features, shown in Figure 2.6 can be helpful. Corner rounding solutions are shown in Figure 2.6. Although finely shaped auxiliary features, such as those shown in Figure 2.6.(a), are effective in sharpening corners, serif features are usually used because of their lower mask costs.

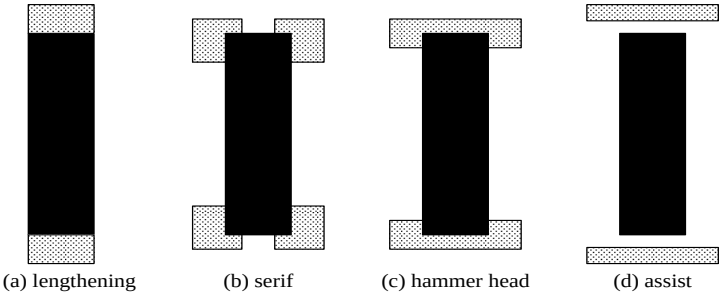


Figure 2.5: Methods of line shortening reduction.

There are two major OPC strategies: rule-based strategy and model-based strategy [1]. In the rule-based method, the design is corrected based on rules extracted previously from simulations, experiments, and etc. Also, the amount of correction applied to a feature or an edge of a feature is determined by predefined tables. Although rule-based methods are efficient for small number of parameters, they cannot be efficiently applied as parameters and rules increase. Therefore, rule-based methods are not suitable for a process with long optical interaction range or assist-feature insertion, where many parameters are effective and many solutions exist.

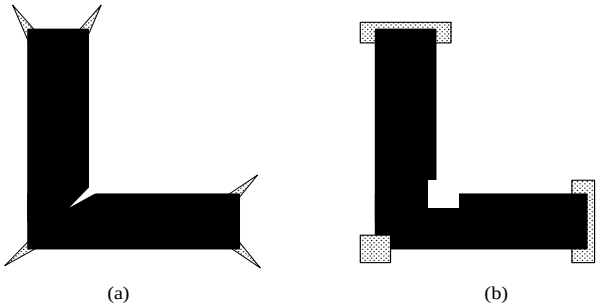


Figure 2.6: OPC solutions for corner rounding distortion.

Rather than applying rules, model-based OPC methods use mathematical models of the fabrication process to determine the correction. In addition to optical imaging, the process model can incorporate other effects such as flare and photo-resist diffusion. There are two flavors of model-based OPC, namely forward and backward OPC. As shown in Figure 2.7.(a), the original layout is iteratively modified in forward OPC until the correction is acceptable in terms of lithography performance and mask manufacturability. However, in backward OPC, shown in Figure 2.7.(b), the desired printed pattern serves as the starting

point. The inverted process model is then used to obtain the optimized layout.

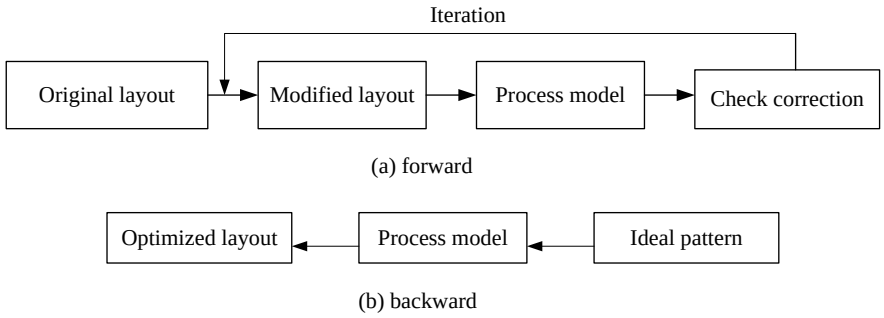


Figure 2.7: Different flavors of model-based OPC.

Regardless of forward or backward approach, the advantage of model-based OPC is its ability to capture a wide range of models and image distortion causes. However, formulation of accurate and efficient models is the primary concern. For example, although models for optical imaging are well established, modeling of photo-resist processing and etching effects are still challenging [37].

2.1.3 Phase shifting masks

Image resolution is improved by taking the advantage of the interference generated by phase differences on phase shifting masks. Alternating PSM [38] and attenuated PSM [39] are the most popular PSM methods.

In alternating PSM, the target line is bordered by transmitting regions with 180° phase difference. As shown in Figure 2.8, this phase shift leads to destructive interference, resulting in a sharp dark image. The reliance on destructive interference between two bright regions to create a dark image makes alternating PSM applicable for imaging only small dark areas.

In a phase shift mask, the regular process is applied to print regions of 0°. Then, 180° phase-shifted regions are etched into the substrate followed by a post-processing for intensity balancing. Although alternating PSM improves imaging quality, it causes CD and placement errors due to imbalanced intensity between regions of 0° and 180° phases. Moreover, in addition to lens aberration, alternating PSM is very susceptible to mask defects not only due to the additional mask processing steps, but also due to the enhanced printing of small defects.

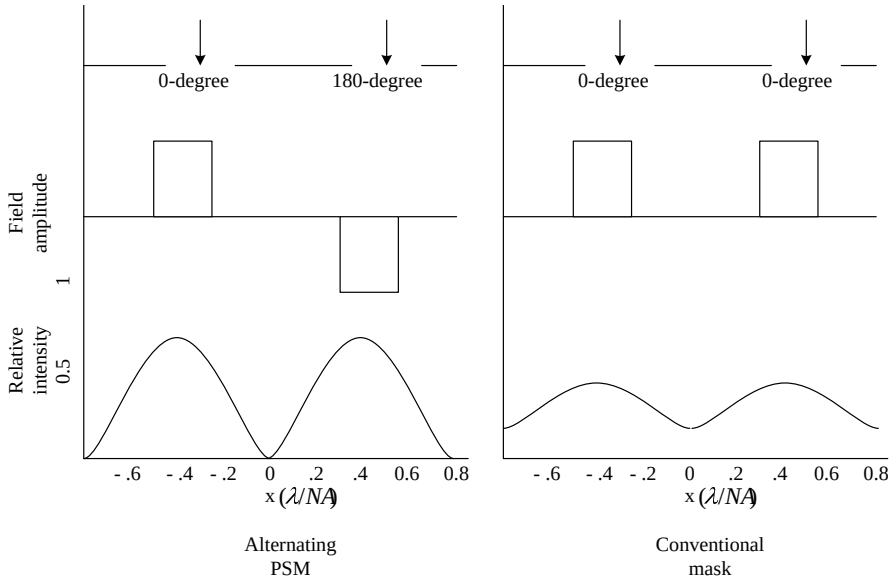


Figure 2.8: Mask schematics and images of a line printed with alternating PSM and conventional mask [1].

The opaque regions of an attenuated phase shift mask transmit the projected light partially with a 180° phase shift relative to the bright regions, as shown in Figure 2.9. The intensity transmission of attenuated PSMs, usually around 7%, does not allow any image formation form the background. However, the destructive interference between the clear area and partially transmitting background enhances the image contrast of the bright region. Compared to alternating PSM, attenuated PSM is suitable to print sparse spaces, e.g. an isolated contact, because of its large DoF and exposure latitude. However, this method cannot compete with alternating PSM in printing narrow dark regions.

To fabricate attenuated PSMs in the past, the chromium layer of a blank mask was tinned until the desired transmission level was achieved. However, because this thinned layer was not 180°, additional phase shift was realized by etching into the mask substrate. Today, this process is replaced by a complicated thin polymeric films to improve transmission and phase shift control.

Since each RET has its concerns and profits, the application demands determine the appropriate RET method to be applied. Table 2.1 summarizes which methods are helpful for different applications. In this table, σ_{small} and σ_{max} represent low-coherency and high-coherency of light source, respectively. In addition, AF, ALT, and ATT denote assist-features, alternating PSM, and attenuated PSM, respectively. Also, $\hat{p} = p\frac{NA}{\lambda}$ represents

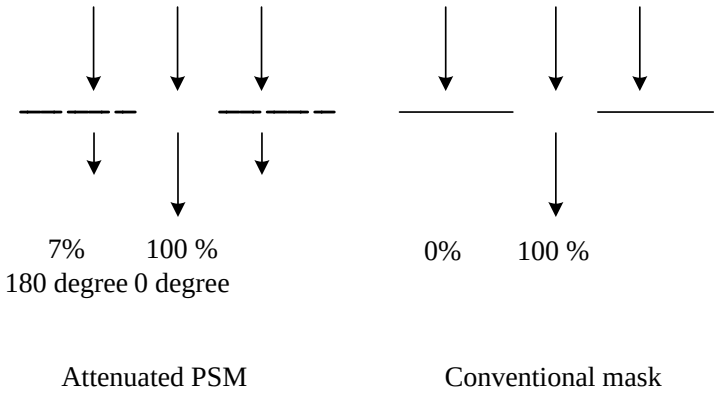


Figure 2.9: An attenuated PSM versus a conventional mask [1].

the normalized pattern spatial period.

Although the discussed methods improve the lithography resolution considerably, they cannot tackle the growing demand for 32-nm half-pitch and beyond resolutions. In the next section, we describe how target technology nodes are achievable using multiple patterning lithography.

2.2 Double patterning lithography

Among different proposed methods for DPL, litho-etch-litho-etch (LELE), self-aligned double patterning (SADP), and litho-process-litho-etch (LPLE) are the most promising ones to be selected as the main manufacturing trend. In this section, these methods are introduced and compared with each other.

2.2.1 LELE double patterning lithography

Either trenches or lines can be patterned using LELE DPL which consists of two complete lithography steps followed by etching. Double trench patterning, shown in Figure 2.10.(a), etches trenches onto a hard mask to form dense lines and space patterns. In this method, a hard mask layer, e.g. SiN, is deposited or coated onto the substrate. This hard mask film should be thin to minimize coating uniformity problems associated with coating of the two resist stacks. Then, a resist stack is coated on the hard mask and the first litho

Table 2.1: General applicability of RETs to common pattern configurations.

Pattern	Appropriate RETs
dense lines $\hat{p} < 1$	OAI ATT+OAI ALT+σ_{small} ATT+σ_{max}
sparse lines $\hat{p} > 1$	ATT+AF+OAI ALT+ σ_{small} ATT
sparse spaces $\hat{p} > 1$	ATT+AF+OAI ATT
mixed pitch	ATT+σ_{small}+OPC
dense contacts	OAI ATT+OAI ALT+σ_{small} ATT+σ_{max}
sparse contacts	ATT ATT+AF+OAI
general two-dimensional patterns	ATT+OPC

process is accomplished to generate semi-dense trenches with a duty ratio (trench/line) of 1:3. The pattern is then transferred to the hard mask layer by an etching process. After stripping, another resist stack, which may be identical to the first stack, is coated onto the hard mask. The second litho step is performed to generate the second set of semi-dense trenches between the previous ones. Then, the pattern is transferred to the hard mask by the second etching process. Finally, the dense pattern with 1:1 trench/line ratio is transferred to the substrate by a final etching step.

Figure 2.10.(b) shows the process of double line patterning. Two different hard masks are needed in double line patterning depicted by orange and yellow layers in Figure 2.10.(b). After patterning of the first group of lines on the first mask and mask etching, resist is developed again over the masks. Now, the second group of lines is patterned on the lower hard mask and a selective etching process is applied to remove unwanted parts of the lower mask but leave the upper one. Finally, substrate etching is done and both hard masks are cleaned from the wafer.

Double trench and double line patterning methods are suitable for printing different types of patterns. Double trench patterning suits the best for dense patterns where spacings are smaller than CD but widths are not much critical, e.g. wires. On the other hand, it is difficult to apply double line patterning to layers that have small area to be opened, e.g. metal layers in logic devices. Therefore, double line patterning is suitable for patterns that their width is less than CD and usually used for contact hole patterning [19]. With respect to cost and throughput, double trench patterning is preferred because it needs only one hard mask layer.

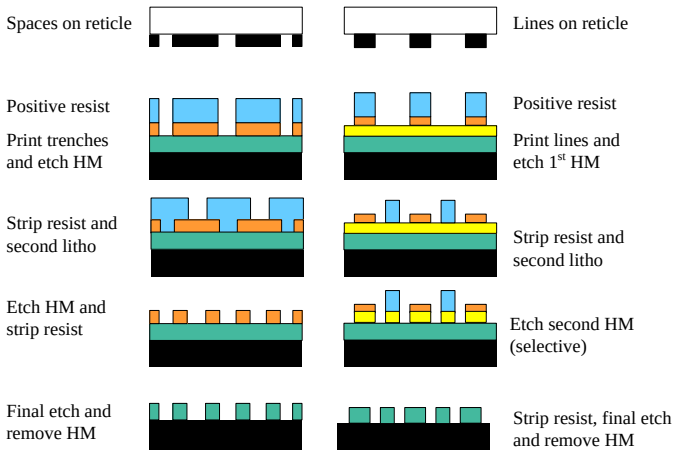


Figure 2.10: Flow of double (dark-field with positive resist) (a) trench double patterning (b) line double patterning [3].

The exposure process in Figure 2.10 requires positive resist and dark-field masks. As well-known, such exposure process window for generating the narrow trenches is too small. Moreover, the dissolution rate of conventional resists increases with the action of acids generated by photo acid generator (PAG). Therefore, it is very difficult to open a narrow trench with conventional positive imaging process. Over-exposure is the only solution to improve the poor image intensity and thus enlarge the process window [4]. In other words, the trenches must be printed larger than the real features on the mask. To compensate for the CD shift due to over-exposure, shrink process should be used to shrink the trenches so that their CDs meet requirements.

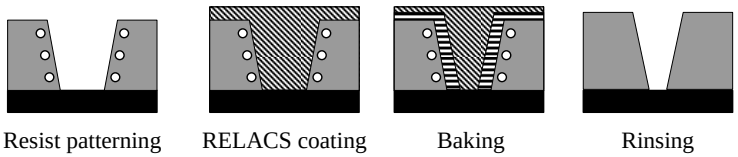


Figure 2.11: RELACS process flow [4].

The resolution enhancement lithography assisted by chemical shrink (RELACS) process [4] has been developed recently to realize narrow trenches with larger depth of focus and lower line edge roughness. In RELACS method, wafer is coated by a RELACS agent after lithography. RELACS agent comprises water-soluble materials that do not cause dissolution damage to the resist pattern during overcoating. Then, the wafer is baked

to accelerate the formation of a water insoluble layer that shrinks the trenches. This insoluble layer is formed by mutual penetration and diffusion of the acid and the cross-linker components of the resist film and the RELACS agent. The final step is rinsing off the non-cross-linked parts of the over-coated RELACS agent. The RELACS process flow is shown in Figure 2.11.

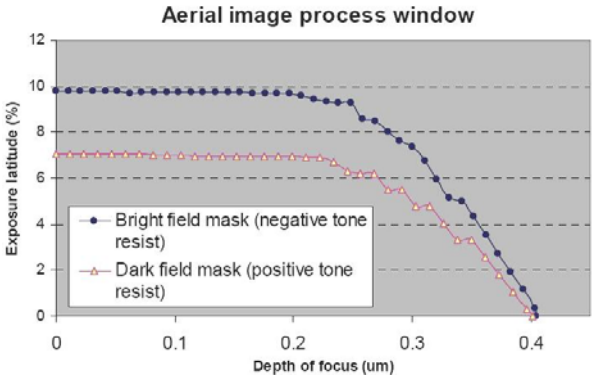


Figure 2.12: Process window bright field mask vs. a dark field mask [3].

The over-exposure method followed by a shrink process for trench patterning further complicates the double patterning process. Instead, we can use a negative-tone resist and a bright-field mask, which provides strong optical intensity and better image contrast. As shown in Figure 2.12, with 5% of the exposure latitude restriction, the use of a bright-field mask gives more than 50-nm of DoF advantage compared with a dark-field mask. As long as the resist resolution supports the extremely aggressive patterning scheme in DPL, the negative- resist/bright-field mask combination is much more beneficial to achieve lower K_1 values. The main concern about the bright-field lithography is that the applied negative resist must have superior lithographic characteristics, such as high resolution, low swelling, and small line edge roughness, which increases the resist cost [3].

2.2.2 Self-aligned double patterning lithography

The self-aligned double patterning, also called spacer DPL, has been proposed to compensate for overlay sensitivity of the LELE method. The main idea behind the SADP methods is replacement of the second exposure phase with a more controllable process, a chemical deposition process in most cases. As a result, the major source of overlay error, which is misalignment between the two sub-masks, is eliminated.

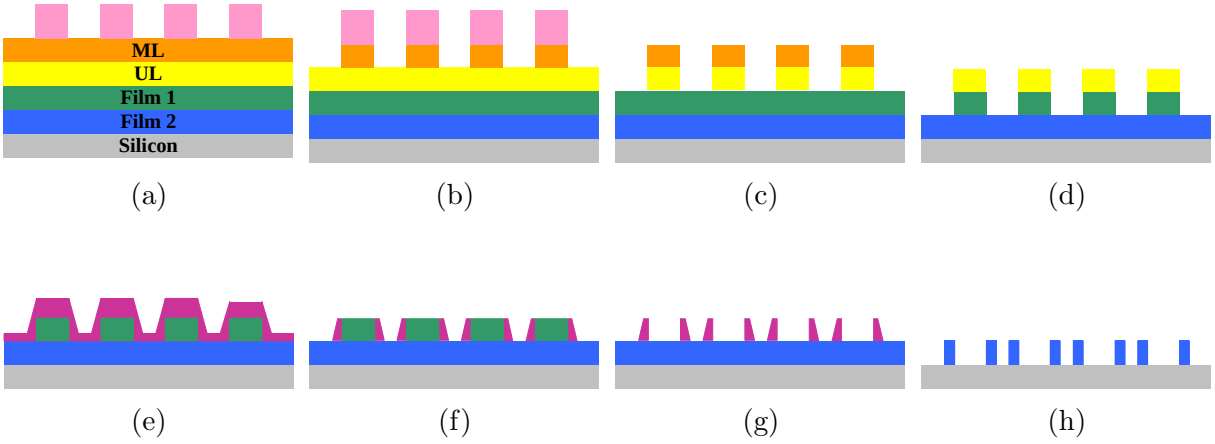


Figure 2.13: LBS Double Patterning [5] (a) Exposure, (b) ML open, (c) UL open, (d) Film1 etch, (e) Spacer deposition, (f) Spacer etch, (g) Sselective etch of film1, (h) Film2 etch.

The basic SADP method can be realized as either a positive- or a negative-tone process. In the positive-tone SADP, also referred as line-by-spacer (LBS) double patterning, the deposited material is used to form the original layout features. Figure 2.13 shows a LBS patterning process. The first phase of LBS process is very similar to the LELE method, where layout trenches are printed by a sacrificial spacer. Then, a hard-mask material is deposited and etched back against the sidewalls of the sacrificial spacers. Afterwards, the sacrificial spacers are removed and target lines are formed instantaneously. Due to suffering from asymmetric etch processes, this method is sensitive to CD uniformity errors [40]. The positive-tone SADP can handle variable spacings; however, the printable line width by positive-tone SADP is limited to the spacer width.

In the negative-tone SADP method, also called line-by-spacer-fill (LBSF) double patterning, the deposited material tends to form critical spacings of the target layout. As shown in Figure 2.14, the negative-tone SADP applies a layer filling step followed by a chemical mechanical polishing (CMP) process instead of non-selective etching in the LBS method. Therefore, negative-tone SADP method is more accurate and more expensive than the positive-tone one. In the negative-tone SADP, different line widths are permitted because a subset of layout features is printed by a regular lithography method.

Compared to the LELE method, the self-aligned double patterning has superior control on CDU and overlay. However, the basic SADP method is applicable only for one-dimensional and strictly regular layout, such as NAND flash memories [24]. To apply

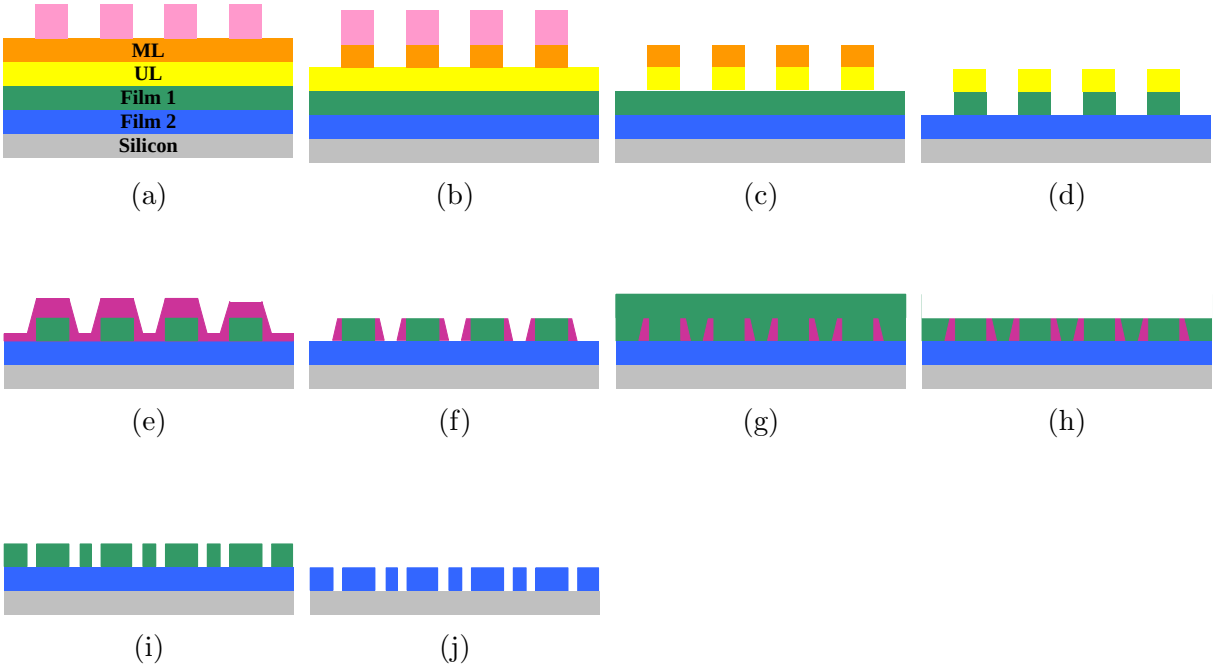


Figure 2.14: LBSF Double Patterning [5] (a) Exposure, (b) ML open, (c) UL open, (d) Film1 etch, (e) Spacer deposition, (f) Spacer etch, (g) Film1 fill, (h) Chemical mechanical polishing, (i) Spacer removal, (j)Film2 etch.

SADP method for general two-dimensional layouts, a litho-based trimming process is appended to the basic process [41]. This process, so-called 2D-SADP, can be done in positive- and negative-tone as well.

Figure 2.15 exemplifies the detailed steps of a negative-tone SADP process, where spacers are used to establish layout trenches. In the first step of the negative-tone SADP, as shown in Figure 2.15.(b), one group of design patterns is printed by the first mask, so-called Mandrel mask. Subsequently, on the sidewalls of the Mandrel patterns, a spacer layer is deposited and etched back for pitch splitting as shown in Figure 2.15.(c). Next, a dielectric layer, demonstrated as yellow grids in Figure 2.15.(d), fills the gaps on the surface using a deposition process followed by a surface smoothing process. Finally, the Trim mask is exposed during the second lithography process to trim away sections of Mandrel and spacer patterns, and to print part of original patterns which have not been formed neither by Mandrel mask nor by spacer patterns. The process is finalized by another etching step followed by filling the formed trenches, which are the actual target patterns.

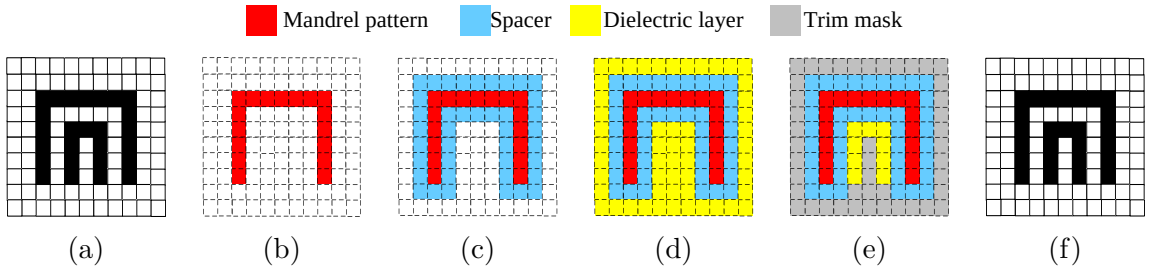


Figure 2.15: Negative-tone SADP process (a) Target pattern, (b) Mandrel lithography, (c) Spacer deposition and etching, (d) Dielectric filling, (e) Trim lithography, (f) Final etch and metal filling.

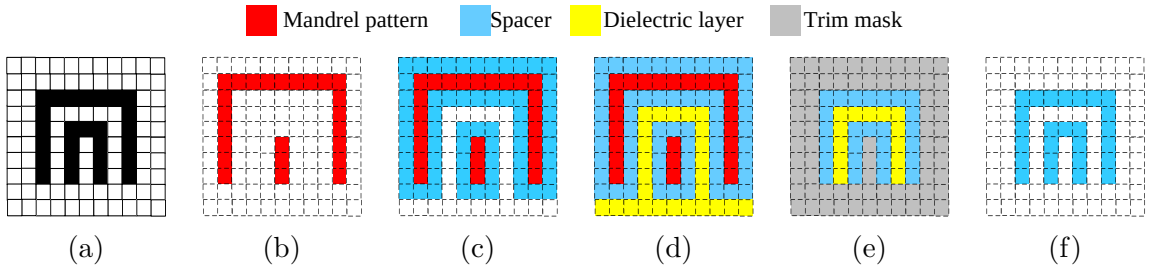
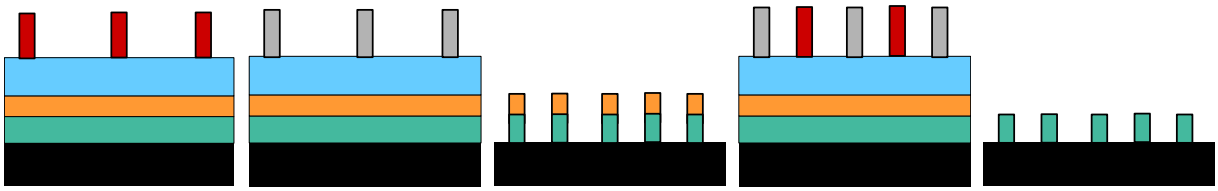


Figure 2.16: Positive-tone SADP process (a) target pattern, (b) Mandrel lithography, (c) Spacer deposition, (d) Dielectric filling and CMP, (e) Trim lithography, (f) Final etch (selective).

Positive-tone 2D-SADP refers to a SADP process in which spacer patterns form the target patterns. Figure 2.16 shows the detailed steps of a positive-tone SADP. In this process, the Mandrel mask, as shown in Figure 2.16.(b), includes part of critical layout trenches and a set of dummy-assist patterns. Deposited spacers around the Mandrel patterns (Figure 2.16.(c)) form the layout patterns. Afterwards, as shown in Figure 2.16.(d), a dielectric layer is deposited and polished to provide a smooth surface for the next lithography step. In positive-tone SADP, Trim lithography is applied to clean Mandrel patterns and undesired parts of spacers. Finally, a selective etching process removes deposited dielectric and leaves spacers as the target patterns. The positive-tone SADP is more limited than the negative-tone approach because the width of layout patterns is restricted to the spacer width in positive-tone method. To make the positive-tone SADP applicable for variable width and pitch layouts, a third lithography step should be affixed to the whole process, which makes it less attractive for layout designers. We will focus on negative-tone SADP in the rest of this thesis.



(a) (b) (c) (d) (e)

Figure 2.17: Process flow for double patterning (double line) using resist freezing technique (a) First exposure, (b) In track process, (c) Etch, (d) Second exposure, (e) Hard-mask removal.

2.2.3 LPLE double patterning lithography

To decrease the cost overhead of DPL, the intermediate etch step is replaced with a curing step by LPLE methods. In these methods, also referred as freezing methods, the first resist pattern is processed to being insoluble during the second exposure and development steps. Preferably, the freezing step is performed in the litho track to save time and cost. The application domain of LPLE methods is the same of LELE methods; i.e. general two-dimensional layouts. The general process scheme is shown in Figure 2.17.

The resist freezing step is the bottleneck of LPLE methods because of damaging challenges associated with this DPL method, such as [3]:

- The first challenge of building resist-on-resist system is how to prevent intermixing of two resists with the same casting solvent.
- The second challenge is how to protect the first imaged resist against distortion during the second exposure. Distortions of the first pattern may arise from any part of the second resist process flow, e.g. resist coating, PAB[1], scanner exposure, PEB[2], and resist development. For example, during the second resist coating step, the first pattern is in contact with the casting solvent that causes many damages, such as pattern collapse, resist swelling, resist lifting, line distortion, and partial or total resist dissolution.
- The third challenge of resist-on-resist system is how to meet lithographic resolution requirements. In the case of pitch splitting approach, we have to print a space that is 3 times the width of a line. Not being a trivial process for conventional resist

[1]Post applying bake
[2]Post exposure bake

systems, this issue becomes even more challenging for resist-on-resist systems owing to a limited number of feasible polymer platforms to choose from.

Different resist freezing methods such as chemical curing, optical freezing, thermal freezing, and ion-beam freezing have been proposed [3]. In the chemical freezing method which is a completely on-track process, a cross-linking material is applied on the resist layer followed by a baking step that enables a reaction between the cross-linking material and the underlying photo-resist.

The UV curing method, has been used for a long time to increase the robustness of resist against etching and implantation steps, can be also applied to prevent pattern deterioration in the first resist layer. The optical freezing needs longer treatment time than the chemical method to provide acceptable level of resist robustness.

Using ion implantation method, where the freezing layer is coated between two resist layers to prevent intermixing, provides valuable by-product benefits, such as resist feature shrinkage and LWR[3] improvements. However, the ion implantation increases the complexity of process because it cannot be done on the track.

Thermal freezing is the simplest available method because it needs only one extra bake step and causes very little distortion on the developed images. Moreover, thermal method is independent from the lithography technology and can be used in 175 nm lithography with no change. However, the resist polymer should be selected carefully such that photo-inactive accurately at a specific temperature.

2.3 Double patterning challenges

In this section, we discuss major challenges for DPL to become the mainstream solution of the next technology nodes.

2.3.1 Entangled CDU and overlay requirements for DPL

The major concern in DPL is the critical dimension uniformity (CDU) errors due to the overlay error between the two masks. The overlay error budget is very tight for DPL, i.e. 7% of the design rules, compared to that for single patterning, i.e. 20% of the design rules [16]. Furthermore, overlay error contributes to CD variability in DPL, which has a very

[3]Line width roughness

tight budget even for single patterning. Therefore, the minimum CD which is achievable by DPL is limited by the overlay and thus CDU errors. In this section, we discuss how the minimum achievable pitch is limitted by overlay errors in single and double patterning methods.

In single exposure lithography, where a pattern is created at once from correlated edges, CDU depends on pattern control. Therefore, design rules are confined by edge placement errors of different process steps, e.g. resist deposition, exposure, and resist etching, which are not correlated necessarily. For example, consider the classical problem of determining the minimum spacing between a contact and a gate in CMOS devices, illustrated in Figure 2.18, is discussed. To avoid short failures, the worst case spacing printed on wafer, $X_{MIN-wafer}$, must be larger than parameter a that depends on device reliability requirements. The worst case spacing happens as:

$$X_{MIN-wafer} = X_{NOM-wafer} - Tol, \tag{2.3}$$

where $X_{NOM-wafer}$ and Tol represent nominal value and tolerance of on-wafer spacing, respectively. The nominal on-wafer spacing differs from the expected spacing, X_{design}, due to the CD control error in poly and contact layers. Assuming that the gate and contact imaging processes are non-correlated, the nominal on-wafer spacing is:

$$X_{NOM-wafer} = X_{design} + \frac{1}{2}Gate_{shrink} - \frac{1}{2}Contact_{increase}, \tag{2.4}$$

where $Gate_{shrink}$ and $Contact_{increase}$ represent inaccuracy in poly and contact layers, respectively.

CD and overlay control errors contribute in tolerance parameter(Tol) as:

$$Tol = \frac{4}{3}\sqrt{\left(\frac{3}{2}\sigma_{CH\_CD}\right)^2 + \left(\frac{3}{2}\sigma_{Gate\_CD}\right)^2 + \left(\frac{3}{2}\sigma_{OL-Gate-to-CH}\right)^2}, \tag{2.5}$$

where σ_{CH-CD} and $\sigma_{Gate-CD}$ represent CD errors occur in imaging of channel and gate respectively, and $\sigma_{OL-Gate-to-CH}$ represents overlay error between gate and channel layers. Therefore, minimum spacing should be greater than worst case one as:

$$\begin{aligned} X_{design} > {} & a - \tfrac{1}{2}Gate_{shrink} + \tfrac{1}{2}Contact_{increase} \\ & + \tfrac{4}{3}\sqrt{\left(\tfrac{3}{2}\sigma_{CH\_CD}\right)^2 + \left(\tfrac{3}{2}\sigma_{Gate\_CD}\right)^2 + \left(\tfrac{3}{2}\sigma_{OL-Gate-to-CH}\right)^2}. \end{aligned} \tag{2.6}$$

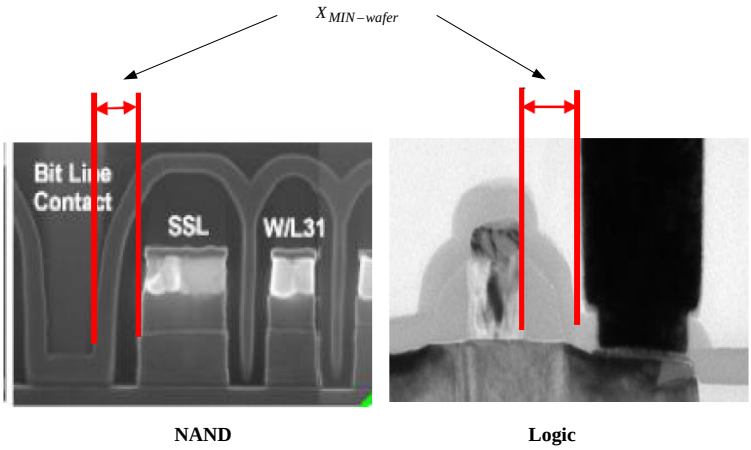


Figure 2.18: The minimum contact to gate edge-to-edge separation [6].

Figure 2.19 shows tolerance budget at 32-nm half pitch. Also, Figure 2.19 demonstrates that CD control and overlay errors are caused mainly by errors rooted in imaging system, such as scanner, exposure, and mask.

To extend the problem to DPL, consider we tend to print a contact between two gate lines each of which is printed by either of the masks. Based on experiments provided by [6], the overlay error between two DPL sub-masks dominates both CD control errors occur in other layers. In the above example, we assumed that edge placement errors do not affect CDU. However, these errors should be considered to model CDU in DPL where a pattern is created from two separate and uncorrelated edges. Consequently, design rules should be defined by twice as many edge placement errors, related to edges created on the same layer by DPL as well as those created by separate process steps.

There are four CD populations in a simple LELE process, two for lines and two for spaces. These CD populations are caused by lithography and etch variations from target at each mask and overlay error between two masks. As shown in Figure 2.20, CDU budgets of spaces and lines are 15% and 10% of design rules, respectively. It is noteworthy that the CD populations of spaces, caused by overlay error, are generally correlated whereas the CD populations of two lines, caused by CD control errors, are uncorrelated. Also, the space variations are uncorrelated from line variations.

Figure 2.21 shows the overlay budget of positive-tone SADP process. As illustrated in this figure, line variations are negligible in self-aligned DPL due to the fine control over film deposition and planarization. On the other hand, spaces vary considerably due to errors in the first sacrificial lithography step. Therefore, there are three CD populations

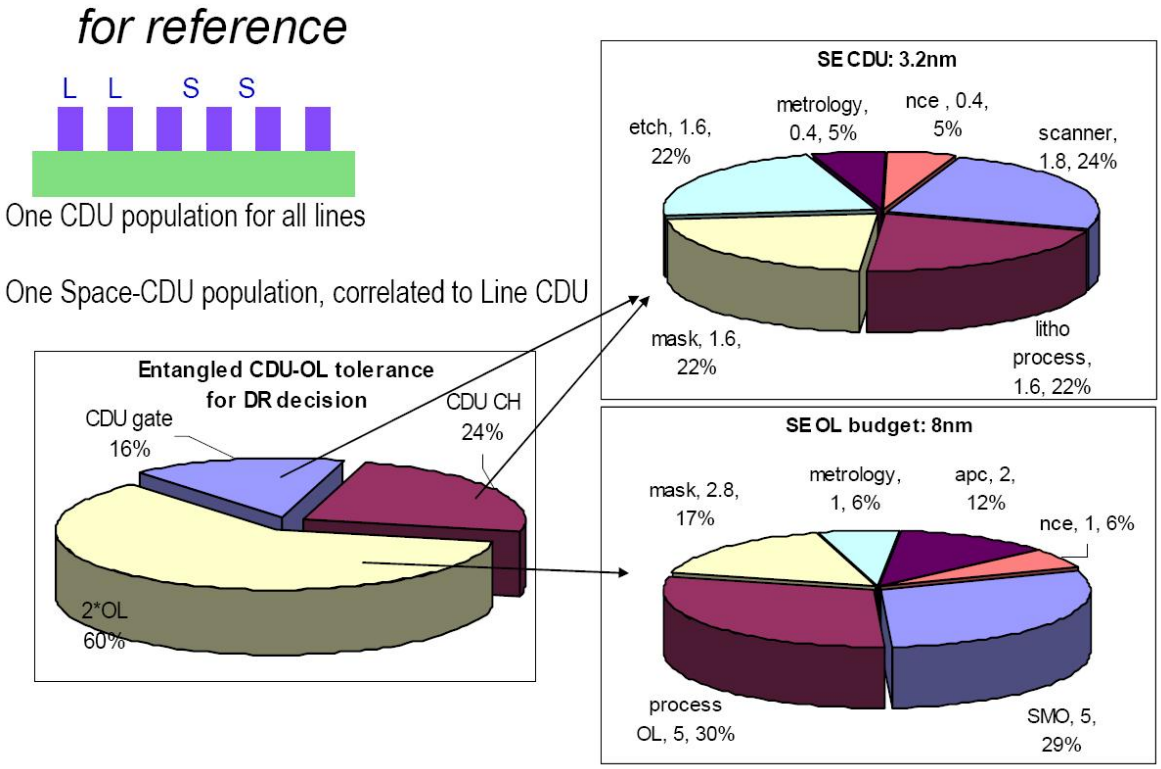


Figure 2.19: CDU and overlay errors dictate the design rule decisions.

in SADP, two for spaces and one for lines.

In addition to functional failures, overlay errors can cause performance degradation. Overlay error between different patterns in the same metal layer can affect metal spacing, related to the interconnect capacitance variability; or metal width, related to the interconnect resistance and capacitance variability. Based on data provided by [42], the overlay error can cause up to 23% variation on the coupling capacitance and 17% variation in the RC delay of Metal1 wires. Therefore, a double patterning design framework should consider the effects of such errors for both functional and parametric accuracies.

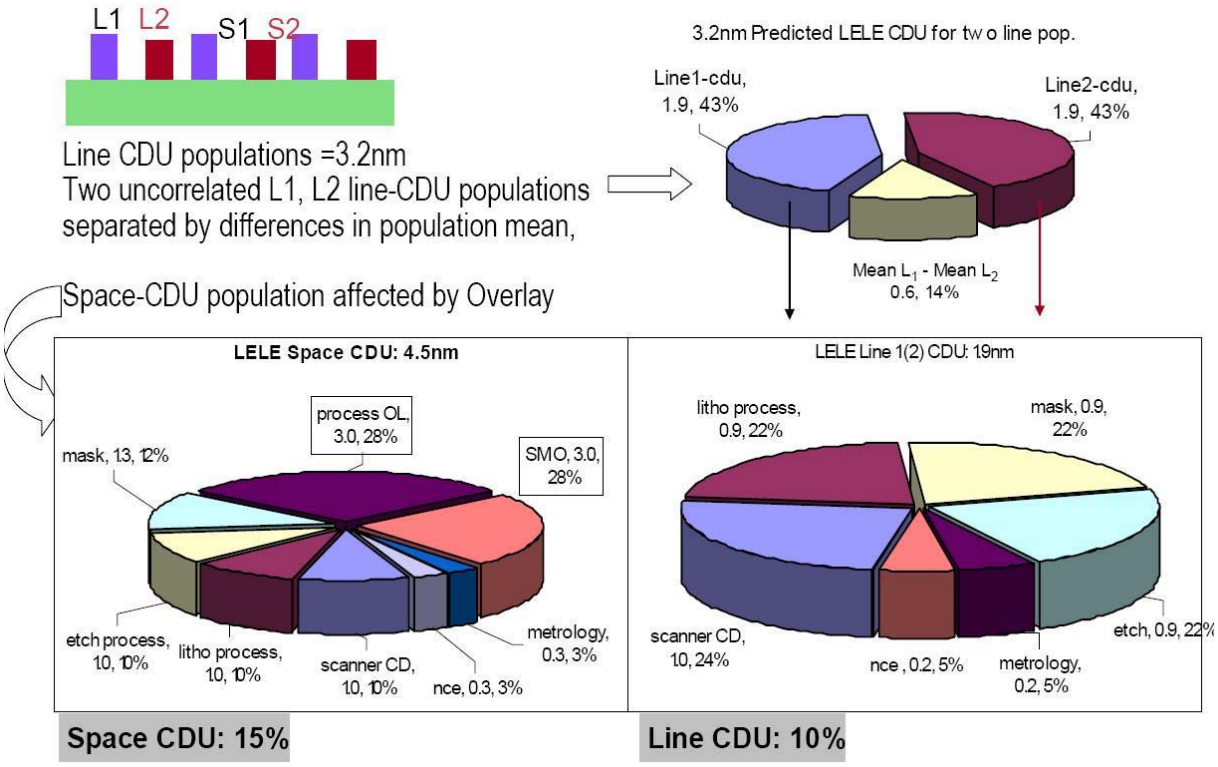


Figure 2.20: The LELE DPL CDU and overlay (OL) budget at 32nm [6].

2.3.2 Layout decomposition

The decomposition of layout for multiple exposure steps is a key issue in an automated DPL design flow. In DPL decomposition problem, layout features should be assigned to either of the sub-masks such that minimum pitch requirements are met in both masks. Layout decomposition can be translated into a graph coloring problem, where the graph nodes and edges represent layout features and pitch conflicts, respectively. Similarly, layout decomposition problem recalls automatic coloring and phase conflict detection and resolution methods for Alt-PSM.

Using the classical graph coloring solutions for DPL layout decomposition is not straightforward because the corresponding graphs are not two-colorable necessarily. In such non-decomposable cases, at least one pattern must be split into two or more parts [20, 7]. These separate patterns are stitched together at the end of the DPL process. However,

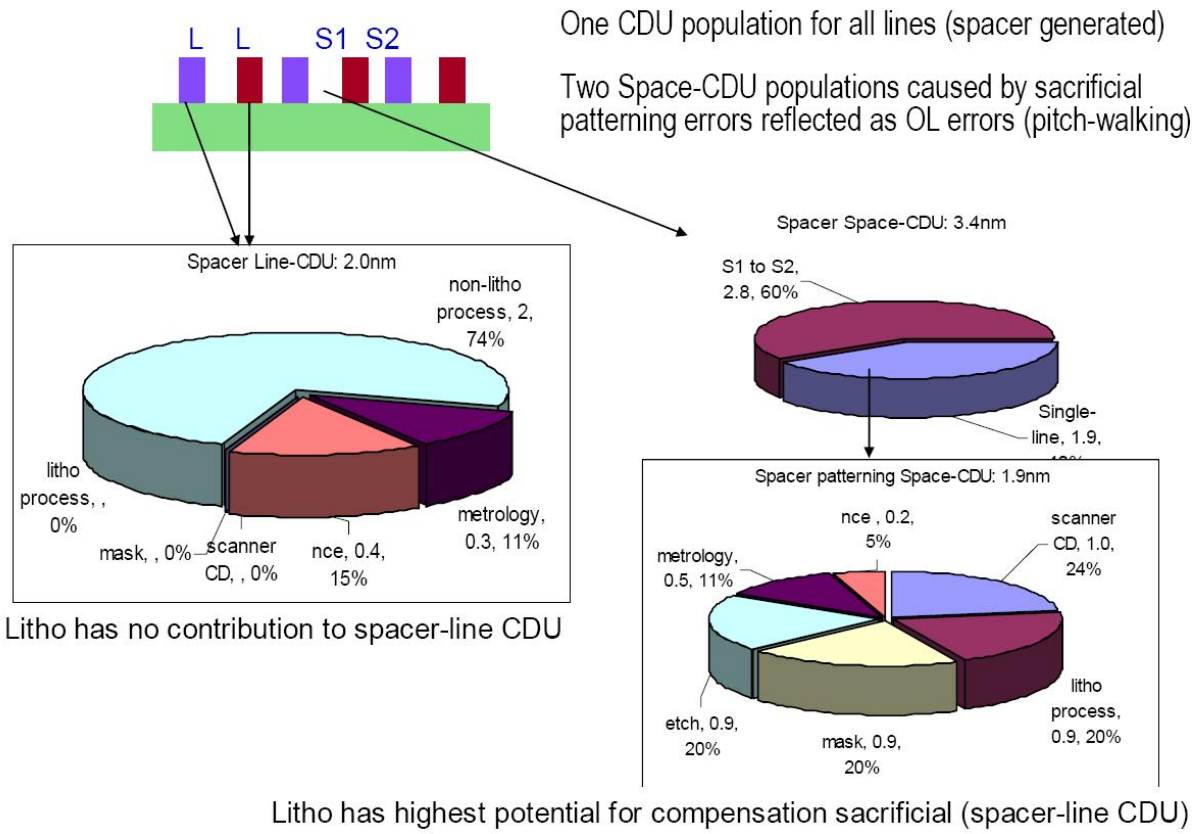


Figure 2.21: Self-aligned CDU and overlay budget at 32nm [6].

being highly susceptible to the overlay error, pattern cuts are not favorable in general. Therefore, a main objective of DPL decomposition methods is to minimize pattern cuts to avoid:

1. Excessive line-ends and yield loss due to the overlay error and line-end shortening under defocus;
2. Over-tightened overlay control requirements, possibly beyond currently envisioned capabilities;
3. Line edge (CD) errors due to the interference mismatch between different masks.

In order to resolve the coloring conflicts it is necessary to reliably identify all conflicting layout configurations on a full chip based on certain commonality. An example for a

decomposition conflict is marked on Figure 2.22 by bold edges. In fact, the challenging problem is not how to find conflicting configurations but is how to identify optimum cut point on the polygon. An optimal cut refers to a cut that resolves coloring conflict with minimum number of cuts, does not generate new conflicts, and meets design rules. One approach for conflict resolution is based on geometric criteria [10, 26]. This approach can be implemented reliably but the solution is guaranteed only if the design is developed with double-patterning technique in mind. In addition, not all legal decomposition solutions are litho-friendly and litho-friendliness objectives complicate the decomposition problem considerably [10, 43]. An effective decomposition method should consider additional factors such as litho-friendly cut insertion [44], mask density balance [45], overlay robustness [10, 11]. Moreover, the complexity of layout decomposition increases in higher levels of patterning, e.g triple patterning [46, 47].

Layout decomposition problem is even more challenging in SADP method, where the solution space is much larger compared with the solution space of LELE decomposition [48]. Moreover, since any pattern split results in a non-resolvable gap in the final image [49, 50], SADP compatibility of layouts is a major concern.

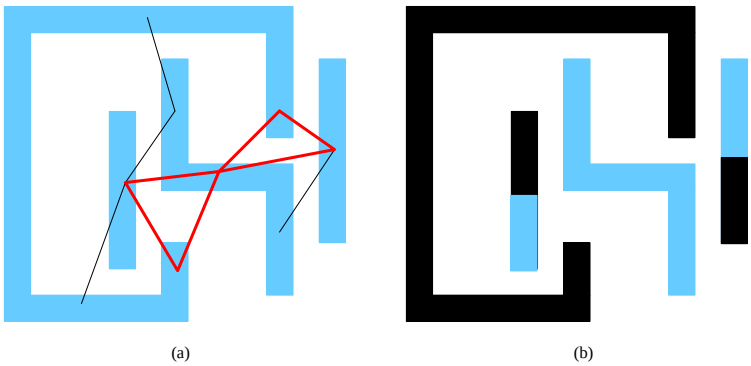


Figure 2.22: Layout decomposition conflict: (a) the original layout (b) a valid decomposition.

Thus, a key optimization goal is to reduce the total cost of layout decomposition, considering the above-mentioned aspects, as well as other concerns about design rule restrictions on each mask and layout density balance across masks.

2.4 Conclusion

In this chapter, principals of traditional resolution enhancement techniques were introduced. Considering the limitations of these methods, DPL was proposed as a promising solution to realize under 32-nm half-pitch nodes. Explaining different methods of DPL, sensitivity to lithography errors and lack of effective EDA solutions were discussed as the main challenges of DPL. Next chapter discusses solutions proposed recently to deal with mentioned challenges.

Chapter 3

Literature Review

To realize double-patterning lithography as the mainstream lithography for below 45-nm technology nodes, there are ongoing research studies in two major fields including process technology [3, 4, 6] and design automation [14, 7, 10]. In the process technology, the major targets are

- to find cost-effective DPL-compliant recipes based on available single exposure imaging facilities
- to improve the process control such that DPL layouts are printed with a reasonable level of yield. The major control parameters mask overlay, dose variation between exposures, and resist etching quality. Moreover, some DPL methods such as litho-freeze-litho-etch (LFLE) need new materials to be used as intermediate resists

According to ITRS [16], current prototype DPL processes are mature enough to be applied for below 45-nm technology nodes. However, higher precision is required to reduce the cost overhead of yield-loss in mass-production scale. The second in demand field of study for DPL, which is considered as the major barrier for industrial DPL solutions [14], is design automation where DPL-specific requirements are considered during layout design. Although the layout design flow does not change considerably in DPL in comparison with single-exposure lithography, DPL-specific lithography effects should be considered in different steps of layout design to gain benefit from the DPL advantages effectively.

Figure 3.1 shows different EDA areas affected by DPL and how each area provides data for other fields. As shown in this figure, EDA-level studies on DPL can be categorized

as DPL-aware lithography simulation and modeling, layout decomposition, characteristics analysis, and DFM and RET methods for DPL. As discussed later in this chapter, DPL-aware lithography simulation and DPL resolution enhancement methods are not new problems and can be answered by extending capabilities of current single-exposure simulation and RET methods. In this thesis, we focus on "layout decomposition" and "characteristic analysis" which are totally DPL-specific.

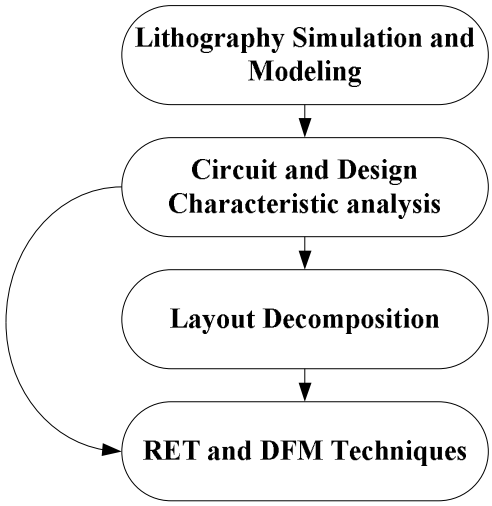


Figure 3.1: Different EDA fields affected by double-patterning lithography.

3.1 Lithography simulation and modeling for DPL

As the critical dimension in optical lithography continues to shrink, it is becoming more important to model non-ideal process conditions and their impacts on the quality of layout images. A library of the most problematic shapes is the core of pattern matching methods to identify error-prone layout features, referred as hot-spots. This hot-spot detection guides layout designers to improve the yield loss due to imperfections of lithography process. Moreover, these suspicious locations can be recorded and later examined at post-silicon time to narrow down the mask inspection regions. To accelerate the hot-spot examination process, candidate spots can be limited to the most vulnerable spots such as points along edges, line ends, inside corners, and/or outside corners.

The pattern matcher is initially developed to quickly scan a layout and locate the regions similar to a given set of test patterns. Match factor (MF) parameter defined as the

convolution of input test pattern and layout is used by pattern matcher as shown (3.1). As can be verified, MF is a number between -1 and +1 representing the degree of similarity between test pattern and layout, where +1 is an exact match and -1 is an exact opposite match.

$$MF = \sum_{i=0}^{ysize-1} \sum_{j=0}^{xsize-1} (pattern[i][j].layout\_matrix[i][j]). \tag{3.1}$$

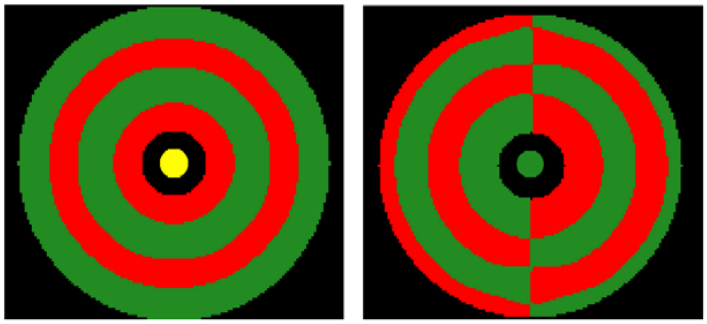
As discussed in Chapter 1.2, the overlay error is the most important accuracy concern for DPL. Since *aberration* errors during the exposure phase contribute considerably in overlay error, they should be modeled accurately in DPL pattern matchers. An *aberration* is any deviation of the real performance of the optical system from its ideal performance. Therefore, aberrations are undesirable intrusions of reality into the attempts to achieve imaging perfection. For current lithography technology, *defocus* and *coma* errors are the most important aberrations [51]. *Defocus* aberration happens when the target resist is not exactly located at the focal point of lens. *Coma* aberration occurs when the light waveform is tilted with respect to the optical surface resulting to a focus area instead of a focal point. Both defocus and coma errors decrease the image intensity and result in pattern distortions as line shortening and corner rounding.

Considering the significant role of the quality of optical lithography on the process window of DPL, some studies have been done recently to simulate and model the non-ideal DPL lithography process [27, 7]. Most of these methods are developed based on existing models of single patterning lithography. For example, pattern matcher engine of Berkeley TCAD EDA tool [52] is extended to examine decomposed DPL layouts across the process window [7]. Major effective parameters in image distortions such as focus-exposure process window, OPC for individual sub-masks, decomposition strategy, and illumination effects are integrated into the TCAD DPL pattern matcher.

The DPL pattern matcher [7] uses maximum lateral test patterns (MLTPs), originally developed by [53], to catch post-decomposition errors due to focus and coma errors. MLTPs are the inverse Fourier transforms of the Zernike polynomials[1] with an additional probe. The amount of light spill over into the central probe is proportional to the amount of aberration present in the system. The focus and coma MLTPs are shown in Figure 3.2, where the central probe for the focus and coma monitors have 90 degrees and 0 degree

[1]The Zernike polynomials are a sequence of polynomials that are orthogonal on the unit disk. Assuming the aperture is circular, the aberration function can be expressed by Zernike polynomials each of which models an specific type of aberration [54].

phase shifts, respectively.



(a) (b)

Figure 3.2: The MLTP patterns: (a) focus MLTP, (b) coma MLTP [7].

Figure 3.3 depicts how a pattern matcher helps to locate a litho-friendly cut point. While Figure 3.3.(a) has a high MF value of 0.11, a pattern split as shown in Figure 3.3.(b) leads to a low MF of -0.005 which improves the reliability consequently.

Inspired by the discussed pattern matcher, the image intensity in DPL is derived theoretically as (3.2) [27]:

$$I \cong \frac{1}{\pi^2}\left[\sum_{x_i,y_i}\int\int e^{\frac{jk}{f_3}[\rho\cos(\theta)x'+\rho\sin(\theta)y']}\rho d\rho d\theta - k^2\left(\frac{2}{3}MF(Z_8)+\frac{1}{3}MF(Z_0)\right)+k^2MF(Z_3)^2\right], \tag{3.2}$$

where the light intensity of each point depends on the Z_0, Z_3, and Z_8 terms of Zernike polynomials which represent position, focus, and spherical focus properties of the exposure system, respectively.

3.1.1 Open research areas

In this section, we introduced DPL-specific lithography modeling methods. In addition to being applied for post-layout detection of lithography hot spots, the developed models can be used for model-based litho-friendly decomposition methods. Such a model-based decomposition improves the lithography yield by taking a wide range of lithography errors into account. However, the time complexity of model-based decomposition methods would

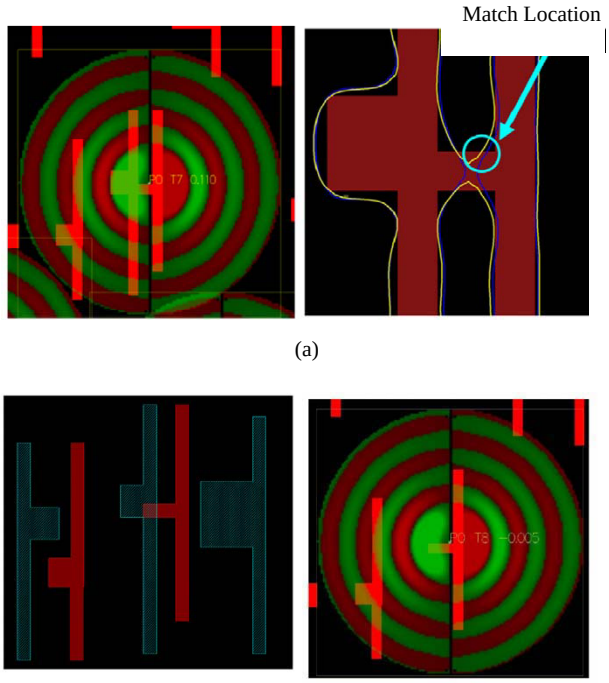


Figure 3.3: Match factor depends on layout split: (a) high MF, (b) low MF [7].

be the matter of concern. In following sections, we discuss rule-based layout decomposition methods that are fed by design rules derived from in-silico or experimental lithography models. These rule-based decomposition methods are less accurate but much faster than the model-based ones.

3.2 Layout decomposition methods

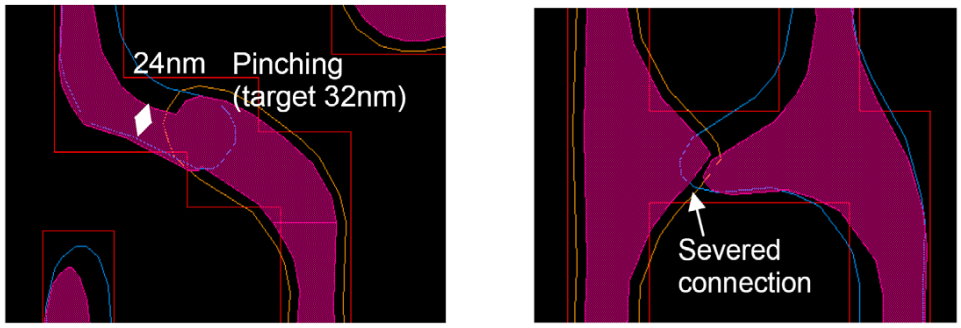
Layout decomposition is a challenging concern for DPL because of its direct impact on DPL efficiency. As discussed before, an efficient layout decomposition can enhance layout printability while an inefficient one may result in functional and parametric errors, such as undesired gaps or bridges, and electrical parameters variation. To address this concern, many decomposition methods have been proposed during the recent years [10, 55, 26, 56, 57, 12, 9].

Assuming the pattern cut as the main source of double patterning lithography errors,

stitch minimization is a common objective between current LELE decomposition methods [10, 55]. There are also decomposition methods which try to avoid error-prone cut insertions [56, 57, 12, 26], such as the stitches close to line-ends and H-shaped patterns. In addition, layout decomposition methods can be divided into post-layout [10, 55, 56, 57, 9] and during-layout methods [12].

Inspired by lithography simulations, some error-prone patterns such as ploy routes and H/U structures were identified by [57]. Moreover, the study suggested some design guidelines to avoid decomposition-related post-lithography errors as:

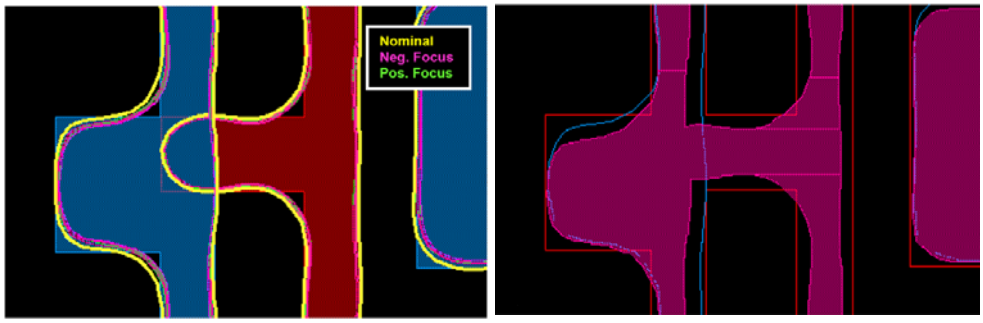
- Avoid small stubs and jogging line-end to make the design robust against defocus errors. Figure 3.4 shows how a careless cut may introduces new line-ends and result-in lithography failures.
- Prioritize the cuts on landing pads, junctions such as T junctions, and long wires due to their low sensitivities to overlay error. Figure 3.5 shows how cut insertion on a landing pad can result in reasonable overall performance.
- As shown in Figure 3.6, insure enough pattern extension on stitching points. The minimum required overlap is a function of overlay error and should be extracted by simulation or measurement for each technology node [8].



(a) (b)

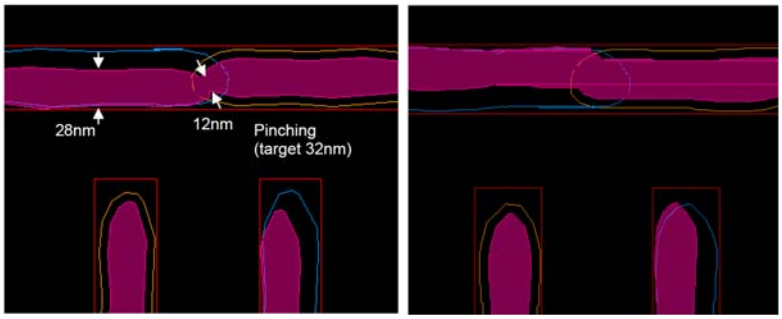
Figure 3.4: Cutting polygons introduces new line-end structures (a) pinching (b) severed connection in worst overlay cases in negative focus and over dose [8].

Afterward, these guidelines were applied to scan a given layout, find pitch conflicts and hotspots, and perform the decomposition. However, the decomposition method leaves



(a) (b)

Figure 3.5: Cut on a landing pad (a) color assignment (b) no pinching due to overlay [8].



(a) (b)

Figure 3.6: Larger cut overlap mitigates risk of pinching (a) 45nm overlap (b) 60nm overlap [8].

many complicated conflicts to be resolved by pitch constraints relaxations and space widening. Experiments show the method of [57] does not deal with at least 20% of the minimum pitch conflicts.

The proposed guidelines by [57] were applied to develop a rule-based decomposition algorithm in [9]. As shown in Figure 3.7, all layout patterns were fragmented by [9] into conflicting segments before the decomposition. Then, the mask assignment was performed to eliminate all error-prone patterns. Finally, a merging process was applied on the colored layout to omit excessive stitches generated by greedy fragmentation. This method was more general compared to the proposed method in [57]. However, such a pattern-based method needs a complete library of un-friendly patterns that should be characterized for each technology node and design style. Therefore, a more comprehensive decomposition

method applicable for various design styles and technologies is required.

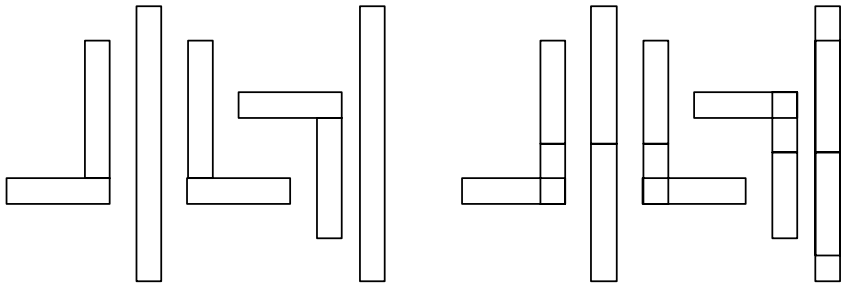


Figure 3.7: Example of layout fragmentation (a) before fragmentation (b) after fragmentation [9].

Translating the layout decomposition problem to a graph bi-coloring problem, Kahng et al. applied the classic graph coloring algorithms to decompose layouts into two masks [55, 10]. The decomposition method proposed in [55] consists of three main phases, namely conflict detection, node splitting, and graph coloring. The conflict detection phase explores the input conflict graph for odd cycles that are not two-colorable. The conflict graph consists of nodes that represent rectangular layout features and edges that show their two end-nodes that do not touch each other but are within the minimum printable distance. A sample layout and its corresponding conflict graph are shown in Figure 3.8.

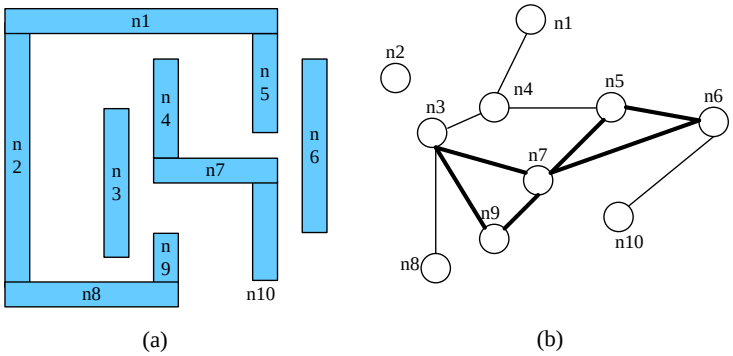


Figure 3.8: Conflict cycles are odd cycles (a) input layout (b) conflict graph.

For each conflict cycle, the node splitting procedure resolves the conflict by cutting one of cycle nodes. To guarantee the newly generated nodes do not cause new conflicts,

the overlap length at the dividing point should be larger than a predefined margin. The overlap length, as depicted in Figure 3.9, is defined as the length that the node can be extended across the dividing point without introducing new edges to the conflict graph. If no acceptable dividing point is found, the corresponding cycle is reported to designer as an inherently unresolvable conflict cycle and is excluded from the layout.

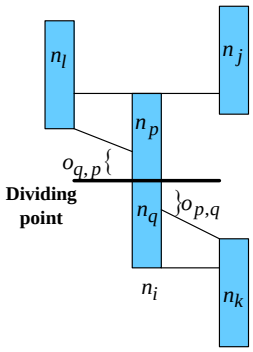


Figure 3.9: Depiction of overlap length.

The color assignment phase is fragmented as an integer linear programming (ILP) problem as (3.3):

$$\begin{aligned} &Minimize \sum c_{i,j} y_{i,j} \\ &Subject\ to: \\ &x_i + x_j = 1, \\ &x_i - x_j \leq y_{i,j}, \\ &x_j - x_i \leq y_{j,i}, \end{aligned} \tag{3.3}$$

where binary variables x_i and x_j represent the colors assigned to rectangles r_i and r_j. The variable $y_{i,j}$ shows whether two touching rectangles are stitched together. The stitching cost is defined as:

$$c_{i,j} = \alpha.f(w_{i,j})/(f(l_i).f(l_j)) + \beta + \gamma/min(o_{i,j}, o_{j,i}), \tag{3.4}$$

where $w_{i,j}$ is width of the rectangles and l_i and l_j are lengths of rectangles. In addition, $o_{i,j}$ and $o_{j,i}$ are the minimum overlap lengths of nodes n_i and n_j, respectively, and α, β, and γ are user-defined scaling parameters. Function f is defined as follows:

$$f(x) = \begin{cases} FS_{\min} & \forall x \geq FS_{min} \\ x & \forall x < FS_{min} \end{cases}, \tag{3.5}$$

where FS_{min} is the minimum feature size. Therefore, higher costs are assigned to pairs of smaller rectangles and those ones are less robust against overlay error, where there is lower possibility for extending line-ends at cut points. Theses patterns are more likely to experience gap or bridging failures in presence of the overlay.

A similar ILP-based approach is proposed in [26] where number of conflicts and stitches are minimized simultaneously using a fine grid. In this method, two groups of neighbors are associated with each occupied grid i:

- Potential stitch grids (PSGs): the occupied grids that touched by the grid i.
- Potential conflict grids (PCGs): the occupied grids within the minimum distance of grid i which doesn't touch it.

Consequently, layout decomposition problem is formulated as an ILP formulation that minimizes number of opposite-colored PSGs and same-colored PCGs, where decision variables are colors of grid points.

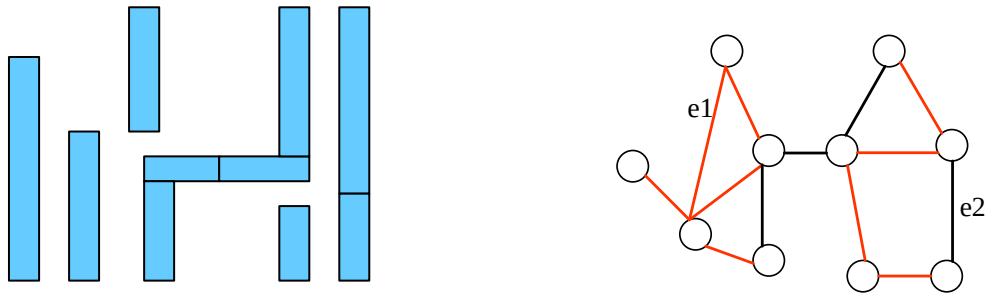
Since the complexity of ILP is NP-hard, the run-times of the discussed ILP methods grow with the layout area. Partitioning the input layout into smaller regions can alleviate the large run-time of these methods. However, partitioning is not an ideal solution because it can lead to conflicts along partition borders. To avoid such conflicts, layout should be partitioned into isolated components and possible conflicts should be resolved by merging process. However, both of these solutions not only increase the run-time but also cannot guarantee a conflict-free layout.

To reduce the time complexity, a heuristic layout decomposition method was proposed by [10]. As shown in Figure 3.10, the input conflict graph is a planar graph that contains two types of edges:

- feature edges that represent the coloring conflicts in the conflict graph;
- touching edges that represent connected but fractured patterns in the layout.

The coloring conflict cycles are the cycles that contain odd number of conflict edges. Therefore, layout decomposition equals to omitting the feature and touching edges so that the graph comes free of conflict cycles. Deletion of a feature edge means moving a layout feature so that the minimum spacing conflict is resolved. On the other hand, deletion of touching edges corresponds to cut insertion. In [10], a phase conflict deletion (PCD) algorithm is applied to delete minimum number of edges and generate optimal bipartite

graph. Experiments show although the proposed method is much faster than ILP method, it generates less optimal results in terms of number of stitches and unresolved conflicts. Moreover, this method did not provide a DRC-aware procedure for moving layout features when a feature edge is moved.



(a) (b)

Figure 3.10: Example of conflict graph for PCD in [10] (a) input layout (b) conflict graph with conflict edges (red) and touching edges (black).

A simultaneous DPL decomposition and layout migration method was proposed in [58]. In this method, a trade-off between cut insertion and space widening is applied to solve post-layout decomposition problem. Large run-time is the main drawback of this ILP-based method that maps the layout into a super-fine grid.

Even an optimal post-layout decomposition method cannot guarantee a fully conflict-free decomposition and some coloring conflicts need layout re-design to be resolved. Fiddling with local patterns cannot solve the problem effectively because it are almost impossible in congested areas, where coloring conflict is very likely. Moreover, they usually lead to an avalanche of new coloring conflicts that makes complete layout re-coloring inevitable and does not guarantee to converge on a fully conflict-free layout. Therefore, it is desirable to keep DPL demands in mind during physical design to generate a DPL-friendly layout.

To improve layout decomposability in metal layers, Cho et al. [11] proposed a DPL-friendly simultaneous routing and decomposition. In this method, the colorability of each grid is tracked by a two-bit variable that takes one of the four states BR, $B\overline{R}$, $\overline{B}R$, $\overline{BR}$, which represent two-colorable, single-colorable, and non-colorable girds, respectively. The initial colorability state of the grid map is determined by a *color shadow* algorithm applied on the pre-colored routing blockages given as the input. In the *color shadow* algorithm, the colorability of free girds are assessed based on their occupied and colored neighbors.

For example, grids near a blue grid will have either $B\overline{R}$ or $\overline{BR}$ states.

The design nets are routed and colored iteratively using a modified A^*-based router. In addition to the conventional wire-length objective, DPL-aware detailed routing tries to avoid non-colorable $\overline{BR}$ grids. As an example, Figure 3.11.(a) shows the current configuration, when the net S-T is to be routed. Figure 3.11.(b) and Figure 3.11.(c) show the routed path after detailed-routing and coloring, respectively. In the path coloring step, grids with $B\overline{R}$ and $\overline{B}R$ are colored as blue and red, and grids $\overline{BR}$ are reported as non-resolvable coloring conflicts. Also, bi-colorable grids are assigned to the nearest color along the path to minimize the number of stitches. After coloring the path, a *color shadow* algorithm updates the state of the grid map, as shown in Figure 3.11.(c).

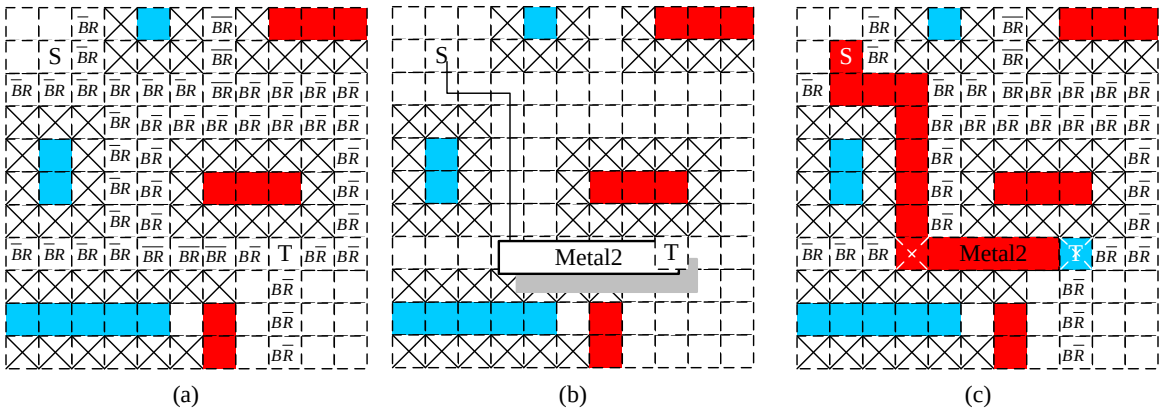


Figure 3.11: The main idea of [11]. The objects are layering in Metal1 by default if not specially notated. The checked boxes are the blockages due to min_{sp}. Except BG, the state is shown in the grid.

Redundant via insertion is integrated with the DPL-friendly detailed routing in [12]. In this method, the feasible locations for redundant via insertion, i.e. those that do not violate the minimum spacing rules, are identified. An example of redundant via candidates are marked with stars in Figure 3.12.(a). Next, all coloring perturbations for each candidate redundant via are generated, e.g. Figures 3.12.(b)-(f); and the perturbations in which a cut occurs along the original layout features, e.g. Figure 3.12.(f), are excluded from the allowed perturbations. Finally, an ILP formulation is developed based on the perturbation set to perform via insertion and coloring simultaneously. The objectives of this ILP optimization are to maximize via insertion rate and to minimize the number of stitches due to redundant vias.

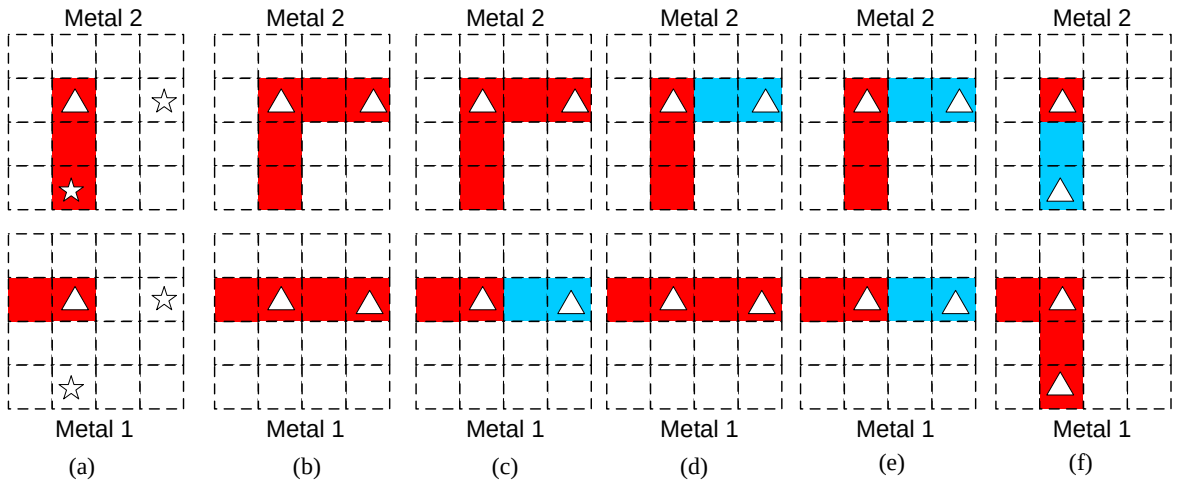


Figure 3.12: Potential configurations for redundant via insertion in DPL [12].

3.2.1 Open research areas

Although there are a lot of studies on LELE layout decomposition, many issues have not been addressed yet. The most challenging problem is how to reduce the time complexity of layout decomposition. The time complexity of current methods depends on number of layout patterns [10] or layout dimensions [26] both of which are very large for current designs. Therefore, some of decomposition methods divide the layout and decompose each part separately. However, as shown by experiments, the accuracy of decomposition degrades considerably due to partitioning [10] because many conflicts occur during the merging process. Therefore, effective divide-and-conquer methods are needed to provide efficient decompositions in a reasonable time. A plausible solution is to develop a DPL-aware standard-cell physical design framework which utilizes a conflict-free cell library and avoid conflicts that may occur between adjacent cells during physical design.

Dividing the layout into many small parts, which are as small as a cell, has two advantages compared with dividing the layout into larger partitions. First, finding the optimal decomposition for a cell is easy because of its small area. With a cell library that contains cells with minimal and robust cuts, the overall number of cuts would be decreased heuristically in the layout. Moreover, one may save significant time by using the same decomposition scheme for all instances of a cell in the layout. The second advantage of using small partitions is that rechecking a large neighborhood is not needed after a local recoloring. However, such conflict-free library is beneficiary only if it is used by DPL-

aware physical design engines. Otherwise, total design time increases during resolving many conflicts which may occur between adjacent cells.

Self-aligned double patterning decomposition is another major area that needs investigations. One-dimensional SADP decomposition is trivia; but the complexity increases for two-dimensional SADP. Although SADP method has attracted much interest because of its higher accuracy, no SADP-specific decomposition method has been proposed so far to address pitch conflict resolution and overlay robustness requirements. It is proved that conflict resolution in SADP decomposition is a NP-hard [48] due to the large search space. This complexity grows considerably when robustness against the overlay between mandrel and trim masks are added to decomposition objectives. In Chapters 5 and 7, we propose EDA solutions for self-aligned double and triple patterning methods.

3.3 Characterization and modeling of DPL designs

Almost all EDA methods, including physical synthesis and design methods, post-layout optimization methods, and DFM methods rely on parametric modeling of underlying process. These parametric models specify the impacts of process technology on different design parameters, e.g. coupling capacitance and RC delay. The accuracy of these models is viable for EDA tools to achieve their objectives. For example, a wire-sizing method would fail to minimize the crosstalk if it uses an inaccurate model of coupling capacitance. Since process uncertainty is significant in deep sub micron regime, different sources of error should be considered by parametric models. As discussed before, impacts of typical error sources, e.g. the overlay error, vary between DPL and single-exposure paradigms. Therefore, conventional modeling methods should be revisited for DPL to take DPL-specific issues into account.

As discussed in Chapter 1.2, lithography errors are becoming highly challenging for DPL because they cause functional errors, such as gap and bridge failures; and change parametric characteristics of both interconnects and transistors. Therefore, it is critical for analysis and optimization methods to characterize design parameters in presence of lithography imperfections. In this section, recent studies for modeling parametric impacts of litho errors on DPL designs are introduced. Most of these studies consider overlay error as the most destructive lithography error in DPL[13].

For the first time, Pan et. al [13] evaluated the effects of different types of overlay error in the performance variation. Considering translation, rotation, and magnification as the main types of overlay, spacing variation between parallel interconnects, ΔS, was

modeled by [13]. Suppose a point (X, Y) in the layout is shifted by (X_T, Y_T), (X_R, Y_R), and (X_M, Y_M) due to translation, rotation and magnification, respectively. Therefore, as shown in Figure 3.13, total spacing variation due to overlay is $\Delta S = \Delta S_T + \Delta S_R + \Delta S_M$ and interconnect spacing changes by $\Delta Y = \Delta Y_T + \Delta Y_R + \Delta Y_M$.

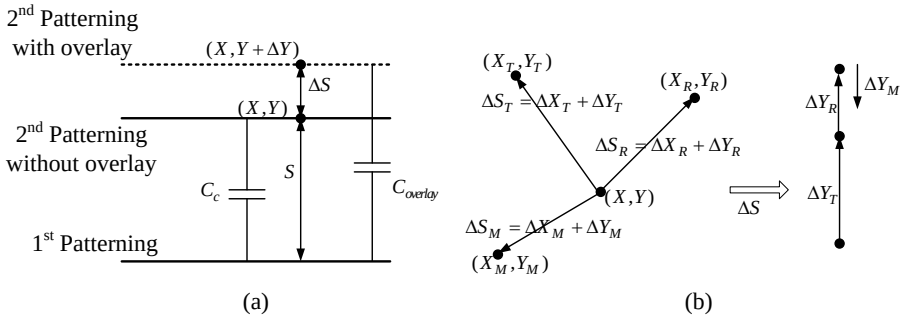


Figure 3.13: Vector expression of overlay: (a)$C_{overlay}$ definition (b)ΔS definition [13].

Using the proposed shift formula, the worst-case overlay is determined for each wire with respect to its within-die location as:

$$\begin{aligned}\Delta S &= \Delta S_T + \Delta S_R + \Delta S_M \\ &= \alpha.\cos(\theta-\gamma) + D.\{\cos(\beta-\gamma-\phi)/\cos\phi - \cos(\beta-\gamma)\} + M.D.\cos(\beta-\gamma),\end{aligned} \tag{3.6}$$

where notations are defined in Table 3.1.

Since coupling capacitance is inversely proportional to metal space, the total shift ΔS can be translated to $C_{overlay}$ as:

$$C_{overlay} = \frac{S}{S+\Delta S}.C_C, \tag{3.7}$$

where ΔS, overall shift due to overlay, is formulated as (3.6) using notations shown in Table 3.1.

In [13], equations (3.6) and (3.7) are fed into a conventional timing analyzer to assess the role of decomposition style in timing variation. As shown in Figure 3.14, a proper decomposition can cause 70% lower variation than an inappropriate decomposition. However, authors did not recommend any guildeline to identify overlay-robust patterings.

Moreover, overlay error was assumed as a fully systematic error in [13]. However, random sources of overlay error dominate the systematic errors specifically for DPL as

Table 3.1: Notations applied during overlay modeling.

Parameter	Description
M	Magnification parameter
D	Distance from the center of mask
α	Translation amplitude
β	position angle
γ	degree between X-axis and orthogonal vector from 1^{st} pattern to 2^{nd} pattern
ϕ	Rotation angle
θ	Translation angle

discussed in Chapter 1.2. Moreover, the proposed model does not take into account the specific issues comprised in different types of DPL.

To model the overlay-induced interconnect capacitance and resistance variations, a simple linear overlay model is used in [42] as:

$$\delta_x = T_x + M_{wx} \times X_w - R_{wx} \times Y_w + M_{fx} \times X_f - R_{fx} \times Y_f + \text{Res}_\text{x}. \tag{3.8}$$

Based on the above overlay model, coupling capacitance between two parallel wires of length L can be calculated as:

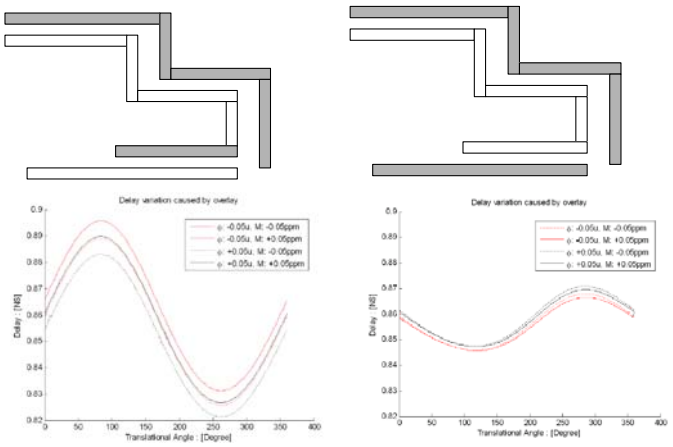
$$C_{LL,2l} = \varepsilon t \int_0^L \frac{1}{s^*} dl, \tag{3.9}$$

where ϵ is the dielectric constant, t is the interconnect thickness, and s^* is the interconnect spacing after the shift due to overlay error. Converting from wafer and field coordinate system to design coordinate system, s^* can be formulated as (4.2) [42]:

$$\begin{aligned} &s^* = s - (T_x + M_{wx}X_o + M_xX_Q - R_{wx}Y_o - R_xY_Q + Res_x) - M_xx + R_xy + sM_x + R_xL, \\ &R_x = R_{wx} + R_{fx}, \\ &M_x = M_{wx} + M_{fx}, \end{aligned} \tag{3.10}$$

where (X_o, Y_o) and (X_Q, Y_Q) refer to the coordinates of field origin in the wafer plane and the origin in the field plane, respectively. Also, (x, y) is the coordinate of the bottom left corner of the line in the design plane. Therefore, the closed form of $C_{LL,2l}$ as a function of structure coordinates in the design is:

$$C_{LL,2l} = \frac{\varepsilon t}{R_x} \ln \frac{s - b - M_x x + R_x y + sM_x + R_x L}{s - b - M_x x + R_x y + sM_x}, \tag{3.11}$$



(a) (b)

Figure 3.14: Delay comparison for different decompositions (a) worst-case delay of 0.895 ns with 9.1% variation (b) worst-case delay of 0.87 ns with 2.7% variation.

where $b = T_x + M_{wx}X_o + M_xX_Q - R_{wx}Y_o - R_xY_Q + Res_x$.

Similar derivation can be performed for a structure of three parallel vertical lines of length L where lines at the edge are printed perfectly and the middle line is printed with an overlay error. The closed form equation of $C_{LL,3l}$ in this case becomes

$$C_{LL,3l} = \frac{\varepsilon t}{R_x}\left[\ln\frac{s-b-M_xx+R_xy+sM_x+R_xL}{s-b-M_xx+R_xy+sM_x} + \ln\frac{s+b+M_xx-R_xy+(s+w)M_x}{s+b+M_xx-R_xy+(s+w)M_x-R_xL}\right]. \tag{3.12}$$

In case of negative-tone process, interconnect width (w) is also affected which leads to the variation of interconnect resistance (R) as well as inter-layer capacitance (C_{LG}) as shown in (3.13) and (3.14).

$$R = \frac{\rho}{tR_x}\ln\frac{w-b-M_xx+R_xy-wM_x+R_xL}{w-b-M_xx+R_xy-wM_x}, \tag{3.13}$$

$$C_{LG} = \frac{\varepsilon L}{2H}[2(w-b-M_xx-M_xw+R_xy)+R_xL], \tag{3.14}$$

where b, R_x and M_x are the same as (3.11), ρ is the wire resistivity, and H is the height of inter-layer insulator.

Also, the coupling capacitance of two ($C_{LL,2l}$) and three ($C_{LL,3l}$) lines structures can be

modeled as (3.15).

$$C_{LL,2l} = \frac{\varepsilon t}{R_x} \ln \frac{s+M_{wx}X_O+M_xX_Q-R_{wx}Y_O-R_xY_Q+M_x(x+w)-R_xy}{s+M_{wx}X_O+M_xX_Q-R_{wx}Y_O-R_xY_Q+M_x(x+w)-R_xy-R_xL},$$
$$C_{LL,3l} = \frac{\varepsilon t}{R_x} \ln \frac{s+M_{wx}X_O+M_xX_Q-R_{wx}Y_O-R_xY_Q+M_x(x+w)-R_xy}{s+M_{wx}X_O+M_xX_Q-R_{wx}Y_O-R_xY_Q+M_x(x+w)-R_xy-R_xL} + \frac{\varepsilon tL}{s}. \tag{3.15}$$

In addition to the overlay error, different decomposed schema of the same target layout result in different delay distributions because of dose and focus variation between the two exposure sub-processes. For example, consider two identical inverters printed by different exposures of DPL, i.e. one inverter is imaged by the first litho-etch process and the other is imaged by the second litho-etch process. Since these inverters have different gate CDs, their electrical characteristics such as delay and power, can also be totally different from each other. Figure 3.15 presents the CD distributions measured from 24 wafers processed by 32-nm half-pitch DPL [14].

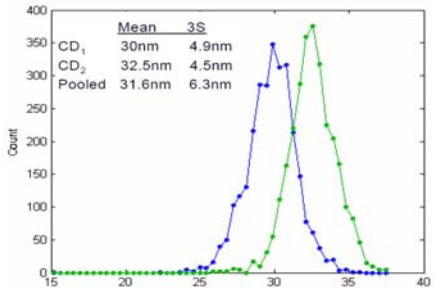


Figure 3.15: Bimodal CD distribution of delay [14].

Therefore, conventional timing analyzers are not accurate enough for DPL layouts and we need to bi- or multi-modal timing analyzers. To catch the timing bi-modality in DPL, [14] proposed a DPL-specific delay distribution. In this model, the expected value of delay is equal to the mean of both distributions. Also, the standard deviation (σ) of the CD distribution is modeled as (3.16):

$$3\sigma^2_{CD,pooled} = \left(\frac{3\sigma_{CD,G1}}{2}\right)^2 + \left(\frac{3\sigma_{CD,G2}}{2}\right)^2 + \left(\frac{3}{2}(\mu_{CD,G1} - \mu_{CD,G2})\right)^2, \tag{3.16}$$

where G1 and G2 are two different groups of CD populations. This unimodal distribution and post-silicon data of 32-nm process reveal that 3σ variation of the delay distribution can be as large as 20% of the mean CD in DPL. Simulations also show the CD variation

grows more drastically as the mean difference between two populations increases [59]. Such large variation cannot be handled as an intra-die variation because conventional statistical timing analysis methods usually cannot handle parameters with large variances.

To deal with bimodal CD distribution in DPL, a bimodal timing analysis and optimization method is proposed in [59]. As exemplified in Figure 3.16, the library cells are characterized assuming only two coloring schema is allowed for each cell.

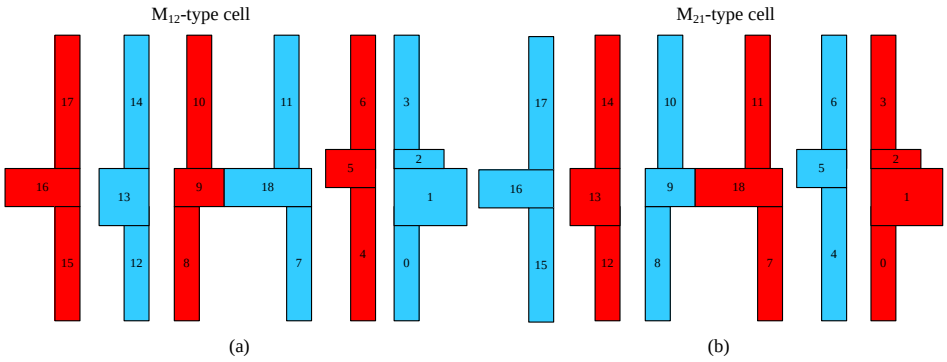


Figure 3.16: Example of two different DPL colorings for a NOR3 cell.

Therefore, the delay variation of a path, consists of m g_i cells of type M_{12} and n q_i cells of type M_{21} can be calculated as (3.17). Figure 3.17 shows an example of the path.

$$\sigma^2_{d-path} = \sum_i \sigma^2_{d-gi} + \sum_i \sigma^2_{d-gi} + 2\sum_{i,j} \mathrm{cov}(g_i, g_j) + 2\sum_{i,j} \mathrm{cov}(q_i, q_j) + 2\sum_{i,j} \mathrm{cov}(g_i, q_j). \tag{3.17}$$

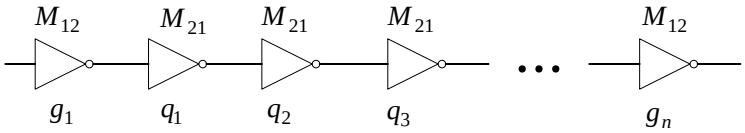


Figure 3.17: An example of a path where cells of the same type are assigned different colorings.

Assume the delays of different cells are not correlated, i.e. $cov(g_i, q_i)$ $approx 0$. Equation (3.17) implies that delay variation of a path would decrease as $cov(g_i, g_j)$

and $cov(q_i, q_j)$ decrease, which means alternative coloring schema are used for the cells of the same type. The proposed model reveals that DPL specifications should be carefully considered in different design steps, e.g. device parameter extraction, cell characterization, placement, and timing optimization. Although (3.17) provides a realistic gage of DPL-related timing variation, it is not comprehensive enough to be effectively used during timing optimization, where we need to consider any possible color perturbation along the path. Moreover, the problem becomes more considerable when note that there are several possible colorings for cells instead of only two colorings. In other words, the real timing distribution is multi-modal instead of bi-modal.

3.3.1 Open research areas

In this section, we introduced the modeling methods that consider particular uncertainty problems which arise in DPL. Although the discussed methods provide useful information about properties of DPL designs, DPL still suffers from high cost overhead due to functional and performance errors which are rooted in lithography errors. It is mostly due to lack of accurate enough DPL-specific models for interconnect and transistor elements. Moreover, the proposed models have not been integrated in design optimization frameworks to improve different characteristics of DPL designs such as performance.

Therefore, it is required to develop design characterization methods using current DPL error models. In addition, these models can be used to derive layout design guidelines to mitigate the undesirable DPL drawbacks.

3.4 DFM methods for DPL

Although DPL allows pitch doubling, resolution enhancement techniques are still needed to improve the litho-friendliness of the decomposed layouts. RETs, OPC in particular, can serve as EDA solutions to identify non-compliant patterns and suggest modifications. Several post-decomposition OPC techniques are proposed in [29] to address OPC challenges in DPL. All techniques focus on ensuring sufficient overlap at stitches to compensate for overlay errors.

To measure how much a stitch is robust against overlap, a *reference line* is defined in [29]that its length determines the minimum CD at overlap location. Therefore, good connectivity can be obtained if the contours from both masks enclose the reference line at the stitch location. Figure 3.18.(a) depicts the reference line, where solid and dashed

lines show target and printed patterns, respectively. The concept of reference line can be extended for different targets, e.g. Figure 3.18.(b) shows different reference lines should be covered in different process steps. To make the design robust against overlay errors, patterns should be extended around a cut point so that the corresponding reference line is covered reasonably. The reference line can be simply considered in the middle of overlap area. However, as shown in Figure 3.18.(c), such primitive extensions result in coloring conflicts. To avoid new coloring conflicts, the reference line should be slid along the overlap area to be far enough from the conflicting neighbors and still being covered adequately.

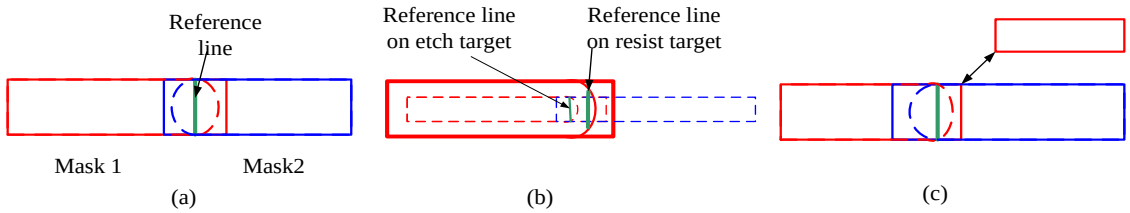


Figure 3.18: (a) Good overlap can be achieved if the contours (in dashed lines) from both masks enclose the reference line (the green line), (b) For each mask, the reference line for the resist target is placed based on the reference line on the etch target and the etch bias, (c) The reference line should be shifted to the left.

The discussed method is very immature although it provides a valuable starting point for LELE-specific OPC. A model-based approach would help to place the reference lines more accurately based on different sources of overlay errors. Also, OPC would perform much more effectively when the applied decomposition method was aware of that because OPC-blind decomposed layouts usually provides few freedoms for OPC application.

3.5 Conclusion

In this chapter, the state-of-the-art DPL automation and DFM methods were reviewed. In conclusion, although DPL is the most promising method to fill the gap between optical lithography and ultra-violet lithography, mature automated design flows are still needed to address different DPL demands, such as functional and performance yields.

As the most challenging step of DPL automation, a lot of studies were done on layout decomposition that mainly focus on the minimization of stitches and coloring conflicts.

However, few of them consider the performance side effects of DPL to mitigate the unfavorable effects. And, interaction of DPL and RETs should be studied more carefully to have a unified flow.

Compared to the LELE decomposition, few studies are done on SADP decomposition. Considering the high accuracy of SADP method, which makes it an attractive option for sub 32-nm nodes, we need to develop automated methods for double and higher degrees of self-aligned patterning. In the following chapters, our DPL automation techniques are suggested and discussed.

Chapter 4

Statistical Yield Optimization for Interconnect in DPL

4.1 Introduction

Amongst various DPL methods, the LELE method has attracted much interest because it is simple and can be realized using existing fabrication facilities. The LELE method consists of two complete lithography steps followed by etching steps. This method can be implemented in both positive and negative-tone imaging systems, where the former approach prints lines and the latter one prints spaces [60].

Overlay error has always been considered as a source of process variation in the single-patterning lithography. However, it should be taken into account more carefully for technology nodes below 45-nm, especially for DPL technology where overlay occurs not only within different layers, but also between two masks of the same layer. According to ITRS [16], the most critical challenge of DPL is to meet overlay error budget (i.e., 7% of design rules), which is much tighter than that for single patterning (i.e., 20% of design rules). Additionally, the overlay error results in critical dimensions (CD) variability in DPL, which has a very tight budget even for single patterning lithography [6].

In general, process variations can lead to functional and parametric yield loss where the desired functionality and performance can be degraded, respectively. Figure 4.1 depicts the impact of overlay error in DPL. As shown in this figure, overlay error between adjacent patterns in the same metal layer can affect metal spacing, which causes interconnect capacitance variability; and metal width, which causes resistance as well as capacitance

variability. Therefore, in addition to the gap and bridge failures, the parametric yield loss should be considered during the tuning of metal width and spacing.

Considering the criticality of overlay error in DPL, the performance consequences of overlay error in DPL have been studied by several papers. Modeling the impact of overlay error on poly line-width variation, [59] proposed a bimodal delay model for different coloring schemes of standard cells. Ghaida et al. [42] provided overlay-aware model for interconnects and discussed the relative impact of different overlay components on electrical parameters of interconnects. Using an analytical model for overly error, Yang et al. [13] analyzed capacitance and delay variation due to overlay in DPL. In [61] and [62], an empirical study was conducted to investigate the impact of overlay on Back-end-of-line (BEOL) layers and its consequences for design timing parameters. The experiments were expanded to develop a variational chip-level simulation methodology for both FEOL and BEOL layers printed by different DPL techniques [63].

Almost all of the proposed overlay analysis and modeling methods considered the overlay error as a pure systematic error [42, 64, 13]. However, experiments show that *random* overlay error is significant for technology nodes smaller than 45-nm [6, 65, 66]. Therefore, the *parametric yield loss* becomes a critical factor in the tuning of metal width and spacing. Experiments provided by [66] confirms the growing randomness of overlay error by showing that it varies in the similar locations of different wafers due to process inaccuracies such as "CPE recipe errors" and "process factors". The variations of interconnect width and space, which are caused by growing overlay error uncertainty, result in uncertainty of electrical characteristics of interconnects and consequently parametric yield loss. Therefore, a statistical approach is required to estimate and mitigate the yield loss due to overlay error.

In this chapter, a statistical method is proposed to determine the interconnect width and space for DPL under the effect of random overlay error. To our best knowledge, this

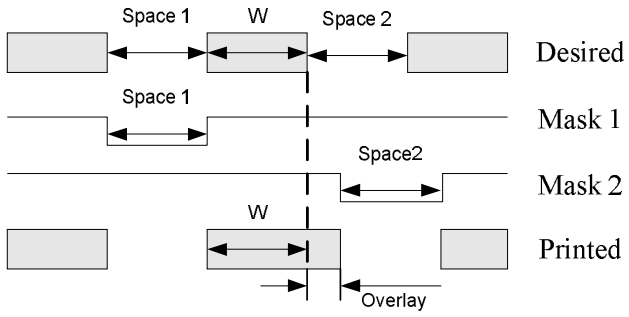


Figure 4.1: Impact of overlay error on CD uniformity.

is the first attempt to use a statistical yield optimization method for DPL. The main contributions of this work are:

1. Overlay error for different DPL methods is formulated and its effects on the variability of electrical parameters of interconnects are analyzed.

2. Using a statistical design methodology, a design center in the metal width-space plane is identified which has the highest probability of meeting the parametric requirements (e.g., total capacitance, RC delay, and resistance variability) in the presence of overlay uncertainty. Using the suggested values for width and spacing as the initial bias can improve both functional and parametric yields effectively.

3. A two-level optimization method is proposed that maximizes the yield, and as a secondary objective, it achieves the best possible (near-optimal) trade-off between different parametric requirements.

4. By using an in-depth analysis, some insightful information about the trend of overlay error and yield loss for present and future technologies is given.

The rest of this chapter is organized as follows: Overlay error formulation for different DPL methods and its translation to the variability of electrical parameters of interconnects are given in Section 4.2. The proposed yield optimization method is presented in Section 4.3. Section 4.4 shows the experimental results and finally, Section 4.5 summarizes and concludes the work.

4.2 Overlay Error Formulation

Between different DPL methods, LELE is more sensitive to overlay error because self-aligned double patterning benefits from a deposition-based process, which is fairly controllable, instead of second exposure phase. However, the LELE method is more likely to be used for printing interconnects due to its simplicity and lower cost compared with the SADP method [3]. Hence, we focus on yield optimization for positive- and negative- tone LELE method in this paper. The extension of the proposed optimization for the SADP method is straightforward by using a SADP-specific overlay model.

Among several proposed overlay error models [64], [67], and [68], the linear one [67] is simple and accurate enough to be applied for interconnect modeling [3] and [42]. Since translation, magnification, and rotation in the wafer and the field are the major overlay

error components [69], a linear model of overlay error along the X direction can be expressed as (4.1).

$$\delta_x = T_x + M_{wx} \times X_w - R_{wx} \times Y_w + M_{fx} \times X_f - R_{fx} \times Y_f + \mathrm{Res_x}, \tag{4.1}$$

where the following notations are used: δ_x for the total overlay error in the X direction, T, M, and R for translation, magnification and rotation errors, w, and f for wafer and field. In addition, (X_w, Y_w) and (X_f, Y_f) refer to Cartesian coordinates in the wafer and the field, and Res is the residual parameter and accounts for un-modeled overlay components and overlay variations.

4.2.1 Electrical Impact in the Positive-Tone DPL

In the positive-tone DPL, the spacing s between two patterns is affected by the overlay error which leads to the change of coupling capacitance (C_{cpl}). Converting from field and wafer coordinate system to the absolute coordinates, the interconnect spacing with overlay error can be shown as (4.2) [42]:

$$\begin{aligned} s^* = s - (T_x + M_{wx}X_o + M_xX_Q - R_{wx}Y_o - R_xY_Q + Res_x) \\ - M_xx + R_xy + sM_x + R_xL, R_x = R_{wx} + R_{fx},\ M_x = M_{wx} + M_{fx}, \end{aligned} \tag{4.2}$$

where (X_o, Y_o) and (X_Q, Y_Q) refer to the coordinates of field origin in the wafer plane and the origin in the field plane, respectively. Also, x and y are the coordinates of the bottom left corner of the line in the design plane and L is the pattern length. Equation (4.2) depends on where the interconnect is located within the design. However, since near interconnects have similar properties, we can use average spacing within a fairly small neighborhood. To this end, design is divided into a set of bins and the average within-bin spacing is used for layout features in each bin as shown in (4.3).

$$\overline{s_b^*} = \tfrac{1}{A_b} \int_{Y_b}^{Y_b+W_b} \int_{X_b}^{X_b+L_b} s^* dxdy = s - k + sM_x + R_xL - \tfrac{1}{2}M_x(L_b + 2X_b) + \tfrac{1}{2}R_x(W_b + 2Y_b),$$

$$k = T_x + M_{wx}X_o + M_xX_Q - R_{wx}Y_o - R_xY_Q + Res_x, \tag{4.3}$$

where (X_b, Y_b), W_b, L_b, and A_b denote bottom left corner, width, length, and area of the target bin, respectively. Clearly, reducing the bin size could improve the optimization accuracy with the penalty of higher computational overhead.

By applying the method of [42] and considering (4.3), the average coupling capacitance between two parallel lines of length L can be expressed as (4.4).

$$\begin{aligned} &\overline{C_{cpl,2l}} = \frac{\varepsilon t}{R_x} \ln \frac{s-k-.5M_xX^b+.5R_xY^b+sM_x+R_xL}{s-k-.5M_xX^b+.5R_xY^b+sM_x}, \\ &X^b = L_b + 2X_b, Y^b = W_b + 2Y_b. \end{aligned} \tag{4.4}$$

Repeating the above steps for three parallel lines leads to (4.5). It is worth noting that in this case, only the middle line is printed by overlay error.

$$\overline{C_{cpl,3l}} = \frac{\varepsilon t}{R_x}(\ln \frac{s-k-.5M_xX^b+.5R_xY^b+sM_x+R_xL}{s-k-.5M_xX^b+.5R_xY^b+sM_x} + \ln \frac{s-k-.5M_xX^b+.5R_xY^b+(s+w)M_x}{s-k-.5M_xX^b+.5R_xY^b+(s+w)M_x-R_xL}). \tag{4.5}$$

For an arbitrary configuration, each line can be considered as a combination of two/three lines to apply the above results.

4.2.2 Electrical Impact in the Negative-Tone DPL

In the case of negative-tone double patterning, further to line spacing, the line width w is affected by the overlay error leading to the variation of interconnect resistance and capacitance [42]. The resistance (R) and inter-layer capacitance (C_{LG}) in the presence of overlay error can be expressed as (4.6).

$$\begin{aligned} &\overline{R} = \frac{\rho}{tR_x} \ln \frac{w-k-.5M_xX^b+.5R_xY^b-wM_x+R_xL}{w-k-.5M_xX^b+.5R_xY^b-wM_x}, \\ &\overline{C_{LG}} = \frac{\varepsilon L}{2H} \left[2(w-k-M_x(.5X^b+w)+.5R_xY^b)+R_xL\right], \end{aligned} \tag{4.6}$$

where k, R_x, and M_x are defined in (4.2), ρ is the wire resistivity, and H is the height of inter-level metal insulator. Applying all steps introduced in Section 2.1 for $C_{cpl,2l}$ and $C_{cpl,3l}$ in negative-tone DPL leads to (4.7).

$$\begin{aligned} &\overline{C_{cpl,2l}} = \frac{\varepsilon t}{R_x} \ln \frac{s+k'+M_x(.5X^b+w)-.5R_xY^b}{s+k'+M_x(.5X^b+w)-.5R_xY^b-R_xL}, \\ &\overline{C_{cpl,3l}} = \overline{C_{cpl,2l}} + \frac{\varepsilon tL}{s}, \end{aligned} \tag{4.7}$$

where $k' = M_{wx}X_o + M_xX_Q - R_{wx}Y_o - R_xY_Q$.

4.3 Yield Optimization Framework

In this section, first we construct the design feasible region and then determine the interconnect width and spacing which result in the desired yield.

4.3.1 Feasible Region Construction

In addition to the required feature size, interconnect capacitance, RC delay, and resistance variability should be considered to determine interconnect dimensions as discussed. Among mentioned parameters, feature size or half pitch is equal to $\frac{1}{2}(w+s)$. Also, resistance variability can be gaged as (4.8).

$$\Delta_R = \frac{R - R_{mean}}{R_{mean}}, \tag{4.8}$$

where R and R_{mean} represent the actual and expected values of interconnect resistance, which the former is calculated by (4.6) and the latter is a user-defined variable.

To calculate the average coupling capacitance within the design, we need to account how it is likely to have two or three parallel lines. Assume that G is the design congestion. The probability of all tracks or only two adjacent ones of three parallel tracks being occupied by a line is equal to G^3 and $2 \times G^2(1-G)$, respectively. Therefore, the average coupling capacitance can be expressed as (4.9).

$$\overline{C_{coupling}} = 2 \times G^2(1-G) \times \overline{C_{coupling,2l}} + G^3 \times \overline{C_{coupling,3l}}. \tag{4.9}$$

Consequently, total interconnect capacitance, required for RC delay calculation, can be calculated by $\overline{C_{total}} = \overline{C_{coupling}} + \overline{C_{LG}}$.

Finally, the feasible region *FR(x)* is formed by design constraints where each point in *FR(x)* satisfies all constraints. The feasible region can be expressed as (4.10):

$$\begin{aligned} &FR(x) = (Pitch < \alpha_1)\text{and}(\overline{RC_{total}} < \alpha_2)\text{and}(\overline{C_{total}} < \alpha_3)\text{and}(\Delta_R < \alpha_4), \\ &x \in \{w, s\}, \end{aligned} \tag{4.10}$$

where x represents the interconnect width and spacing. The parameters $(\alpha_1 \cdots \alpha_4)$ represent user-defined design constraints which are usually adopted from ITRS [16].

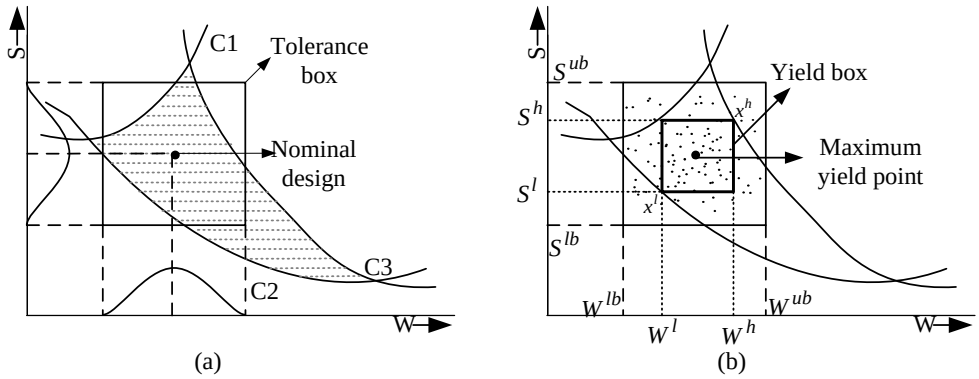


Figure 4.2: Yield maximization method a) parameter space and feasible region b) optimal yield box.

4.3.2 Yield Optimization

The interconnect width and spacing, (w,s), form a 2D parameter space, within which the feasible region (4.10) is defined. This region is shown by dashed area in Figure 4.2.(a) for an arbitrary feasible region defined by a set of arbitrary constraints. Any point inside this region is a potential nominal design which satisfies all design constraints. However, the goodness degree of each point depends on how likely it would stay inside the feasible region in presence of parameter variation.

If we consider w and s as bounded and normal random variables, the real design is located within an imaginary box around the nominal design, called the *tolerance box*. As shown in Figure 4.2.(a), the dimensions of the tolerance box depend on the spread of design parameters. The center of the tolerance box is the nominal design because the design parameters are assumed to have normal distribution, hence symmetrical.

The yield of a nominal design is the area of overlap region between its associated tolerance box and the feasible region. However, it is a computationally-intensive process to calculate the area of a general-shaped feasible region. To reduce the complexity, the rectangular overlap that is attainable between the feasible region and the tolerance box can be used to estimate the yield directly [70]. This rectangular region, shown in Figure 4.2.(b) with bold lines, is referred as *yield box* in the rest of the paper. On the other hand, any random sampling method, e.g., Monte Carlo, can be applied to generate random realizations for a nominal design, shown with small dots in Figure 4.2.(b). The ratio of acceptable realizations to all samples can be considered as yield. We will use this method to verify the accuracy of our yield estimation method.

Now, the yield optimization problem can be specified as: *starting from an initial design, the tolerance box should be moved over the design space to find the best box's location in order to maximize the yield.* Qualitatively, the problem is reduced to finding x^l and x^u, the coordinates of the yield box in Figure 4.2.(b), such that the following conditions are satisfied:

1. the maximum difference between and x^l and x^u is not greater than the maximum spread in design parameters (the yield box should lie within the tolerance box)
2. the yield box lies in the feasible region (all the points lying within the box satisfy all the design constraints that cover within-die variations).

If the aforementioned conditions are met, then for the nominal design placed at the center of the yield box, the probability (yield) that the design constraints are satisfied in the presence of parameter variations is estimated as (4.11):

$$Yield(x^l, x^u) = \prod_{i=1}^{2} \Pr\{x_i^l \leq x_i \leq x_i^u\} = \prod_{i=1}^{2} \left(CDF(x_i^u) - CDF(x_i^l)\right) \tag{4.11}$$

where $x_i \in \{w, s\}$, *CDF(x)* denotes the cumulative distribution function of x, and x^l and x^u represent the bottom left and upper right corners of the yield box, respectively.

The solution of (4.11) involves the evaluation of a multidimensional probability integral by quadrature or MC-based methods, which is computationally expensive [71]. However, the problem is simplified if a closed-form expression for CDF can be used. If design parameters are considered as normal random variables, none of constraints that limits the feasible region has a closed-form CDF. To address the issue, Kumaraswamy's distribution [72], double-bounded probability density function (DB-PDF) defined in (4.12) is employed:

$$\begin{aligned} &f(z) = abz^{a-1}(1-z^a)^{b-1}, \\ &z = \tfrac{x-x^{lb}}{x^{ub}-x^{lb}}, x^{lb} \leq x \leq x^{ub} \end{aligned} \tag{4.12}$$

where z, x^{ub}, and x^{lb} represent normalized value, upper and lower bounds of random variable x, respectively. By assigning different values to a and b, the PDF takes a variety of distributions such as uniform, triangular, and Gaussian. The closed-form CDF of this distribution can be obtained by integrating $f(z)$ and is given by (4.13):

$$F(z) = 1 - (1-z^a)^b. \tag{4.13}$$

Therefore, using (4.11) and (4.12), yield is expressed as a function of the lower and upper bounds of the variables [73], and a reference point (referring to the location of the optimum box), and is obtained by:

$$\begin{aligned} &Yield(x^r, x^l, x^u) = \prod_{i=1}^{2} \Pr\{x_i^l \le x_i \le x_i^u\} = \\ &\left[F\left(\tfrac{w^u - w^r}{t_w}\right) - F\left(\tfrac{w^l - w^r}{t_w}\right)\right] \times \left[F\left(\tfrac{s^u - s^r}{t_s}\right) - F\left(\tfrac{s^l - s^r}{ts}\right)\right], \end{aligned} \tag{4.14}$$

where $x^r = [w^r, s^r]$ is the bottom left corner of the tolerance box, and x^l and x^u are the bottom left and upper right corners of the optimum yield box with respect to the design variables. Also, t_w and t_s represent the range of the distribution of w and s.

Finally, the yield maximization problem is formulated as (4.15):

$$\begin{aligned} &\max Yield(x^r, x^l, x^u) \\ &S.T.: \\ &R(x^l, x^u) \subseteq FR(x), \\ &x^r \ge x^{\min}, x^l \ge x^r, \\ &x^u - x^l \le t, x^r + t \le x^{\max}, \end{aligned} \tag{4.15}$$

where R is the inner optimum tolerance box contained in the feasible region. Therefore, we have:

$$R\left(x^l, x^u\right) \subseteq \{x \in \Re^2 | x^l \le x \le x^u\}. \tag{4.16}$$

In general, the optimal design in the deterministic domain is selected as an initial design. Then, a nominal design with minimum deviation from the deterministic optimum point is found. Since most of design tools and manufacturing facilities are tuned to target ITRS [16], without loss of generality, we consider ITRS suggestions for interconnect width and spacing in each technology node for the initial design. As a result, the reported values by ITRS for different overlay components remain accurate enough within the tolerance box.

The above optimization can be implemented by sequential quadratic programming [74]. To this end, a figure-of-merit function is used to consider the trade-off between maximizing

the objective function and violation of constraints. In this method, the tolerance box, centered at the initial solution, is moved within the feasible region to find the point in which the area of yield box can be improved. Then, the current solution is used as the initial solution for the next iteration. The final location of the tolerance box is a function of the yield and the shape of the feasible region. As a clarification, assume that the feasible region is large enough to surround the entire tolerance box. As long as the tolerance box is inside the feasible region, no matter where the design center is located, the yield is 100%.

Equation (4.15) may have more than one solution. To attain the best possible trade-off between different objectives, it is desired that the final solution, be as close as possible to the deterministic optimum point, i.e. initial solution. Therefore, we consider the distance of the solution from the initial solution as a secondary objective function. By minimizing the objective function λ in (4.17), the tolerance box is moved over the feasible region to maximize the yield and minimize the deviation from the deterministic optimum point.

$$\min \lambda = \left[(x - x^c) (x - x^c)^T \right]^{\frac{1}{2}}, \tag{4.17}$$

where superscript T stands for the transpose of a vector and x is a point on the surface of constraint $g_i(x)$ which has the shortest distance from the center of the tolerance box (x^c).

4.4 Experimental Results

4.4.1 Experiment Settings

The proposed yield optimization problem was solved by using the sequential quadratic programming tool in MATLAB, where design constraints ($\alpha_1 \cdots \alpha_4$); i.e. half-pitch, RC delay, total capacitance, and resistance variability, were chosen from ITRS [16]. Also the spread of overlay error, which determines the dimensions of the tolerance box, was adopted from ITRS. Total overlay was assumed to be equal to 3σ overlay for single-patterning lithography, 20% of design rule. The relative contributions of different sources of systematic and random overlay errors were extracted from the study performed in [65]. Therefore, we assumed only 40% of the total overlay error was caused by systematic sources, each of which contributes to total overlay error as shown in Table 4.1.

4.4.2 Verification of the Proposed Yield Estimation Method

To verify the accuracy of our yield estimation method, we compared our results with the results of a Monte-Carlo simulation method. To this end, we used a layout that was generated by random assignment of wires into different routing tracks with the probability of G, which represents the congestion level. We generated wires of lengths 10um to 1000um randomly using a normal distribution [75]. The width and spacing of routing tracks were chosen as the targets of ITRS for 32-nm technology, (32nm,32nm). Also, the dimensions of wafer, field, and die were chosen similar to the settings in [42]. Using this imaginary layout let us to consider the effects of all possible routing patterns as well as intra-die variation of overlay. Next, we decomposed the layout into two masks using the algorithm proposed in [10]. To consider the impact of random overlay error on the yield, 1000 versions of the original layout were generated using samples of random overlay error which were generated by Monte-Carlo simulation. Finally, parasitic components of each layout were extracted using Calibre-xRC [76] and the yield was determined based on the total number of qualified layouts. Figure 4.3 visualizes our verification method.

For a 75% congested layout in 32-nm technology, Table 4.2 reports runtime and estimated yield for a given layout, when the layout is binned with different binning sizes. In addition, Table 4.2 compares the proposed yield estimation results with the yield calculated by Monte-carlo simulation combined with Calibre-xRC [76]. As shown in this table, the accuracy of our method tends to Monte Carlo simulation for smaller binning size, i.e. larger number of bins, in expense of larger runtime. The reason is that in the binning process, we take the average of overlay error within each bin. Therefore, for a small bin size, the average overlay error is closer to the exact overlay error within the bin. However, as shown in Table 4.2, even with a large bin size, hence a lower precision level, the accuracy of our

Table 4.1: Components of overlay error.

Source	Parameter	Value (nm)	
		32-nm	22-nm
Systematic	Translation	.256	.176
	Wafer Mag. (M_{wx})	.59	.49
	Field Mag. (M_{fx})	.128	.088
	Wafer Rot. (R_{wx})	1.4592	.8976
	Field Rot. (R_{fx})	.128	.088
Non-systematic	-	3.84	2.64
Total		6.4	4.4

yield estimation method is comparable with the Monte-Carlo's one while the runtime of the proposed method is considerably better than the Monte-Carlo's runtime. Since yield estimation is the bottleneck of statistical wire sizing algorithms, the proposed method can be embedded into these tools because of its low runtime and reasonable accuracy.

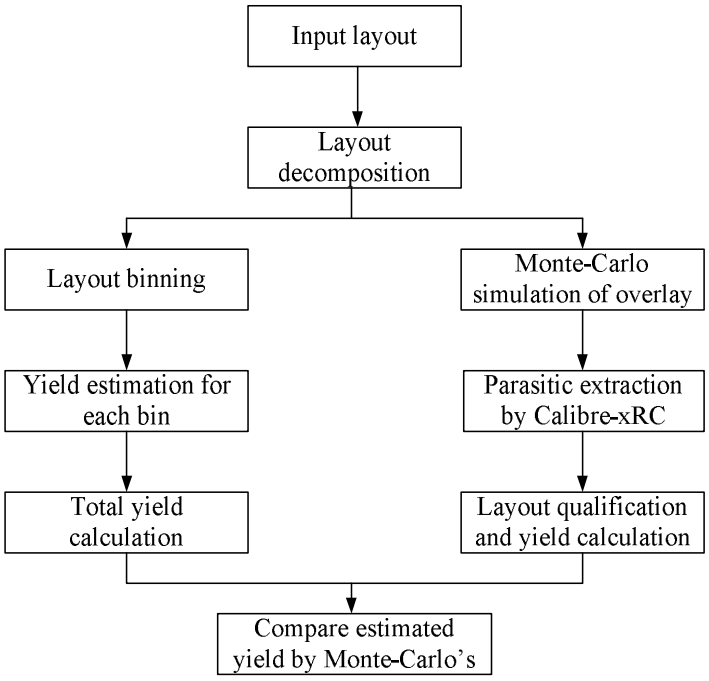


Figure 4.3: The verification procedure for the proposed yield estimation method.

Table 4.2: Accuracy of the proposed yield estimation method.

The proposed method						MC+Calibre-xRC	
# of Bins = 1		# of Bins = 16		# of Bins = 64			
Y (%)	T (s)	Y (%)	T (s)	Y (%)	T (s)	Y (%)	T (s)
75	1.6	78.2	5.5	85	73	85.8	308

4.4.3 Yield Optimization Results

We repeated the yield optimization for both 32- and 22-nm nodes. All the reported experiments in the rest of the paper were obtained for a die of size *8mm*6mm* [42] which

Table 4.3: Optimization results for different technology nodes.

Node	Tone	Corner case	Nominal design Congestion 75% $\overline{Y}$ (%)	Nominal design Congestion 90% $\overline{Y}$ (%)	Optimized design Congestion 75% w (nm)	75% s (nm)	75% $\overline{Y_F}$ (%)	75% $\overline{Y_P}$ (%)	90% w (nm)	90% s (nm)	90% $\overline{Y_F}$ (%)	90% $\overline{Y_P}$ (%)
32nm	Positive	Best	4	5.9	32	32	3.2	.8	32	32	4.5	1.4
		Avg.	6.8	10	32	32	4	2.8	32.5	32.5	7	3
		Worst	12.2	14.5	33	32.7	6.2	3	33.4	33	9.1	3.3
	Negative	Best	10.5	14	32.2	32.2	5.1	1.7	32.4	32.5	6.7	2.6
		Avg.	17	21.8	33	32.6	10.3	3.9	33.7	33	13	4
		Worst	27.5	31	33.5	33	12	4.4	34.5	33.5	14.8	5.8
22nm	Positive	Best	9.8	11.5	22	22	4.8	5	22.5	22.5	5	6.5
		Avg.	26.3	28	22.5	22.7	7	8.5	22.5	23	8.2	9
		Worst	30.7	32.5	22.9	23	7.5	8.5	23.2	23.2	9.5	10.1
	Negative	Best	14.5	19.4	22	22.5	6.3	6.5	22.5	22.7	7.2	8
		Avg.	24.1	30	22.7	23	11	11	23	23.1	13.8	13.2
		Worst	28.5	31.2	23.2	23.2	13.1	13.6	23.1	23.5	14	14.8

was binned into 64 equal bins. The constraints for each technology node were extracted from ITRS [16]. The optimization results for different technology nodes and lithography tones are summarized in Table 4.3. In this table, for bins with different congestion levels the yield loss in the nominal design is reported in columns four and five. Moreover, the optimal solution and resulted yield loss, achieved by the proposed method, are reported in the next columns. Parameters $\overline{Y_F}$ and $\overline{Y_P}$ specify the minimum achievable values for functional and parametric components of yield loss, respectively.

As shown in Table 4.3, the maximum achievable yield and optimal width and space vary for different bins according to the level of overlay error that occurs in the bin. In wire planning, this information can guide the designer to route critical nets via more immune bins. The experimental results demonstrate that our method can resolve 22% and 18% of nominal yield loss for 32-nm and 22-nm nodes on average. In addition, the lowest average yield improvement is achieved for 22-nm node. It means the improvement of overlay accuracy is highly demanded to meet the parametric requirements of future technology nodes. Table 4.3 also shows negative-tone lithography is much more vulnerable to overlay error than positive-tone lithography that conforms with experiments reported by [65].

To examine how congestion affects the yield, the yield optimization is repeated for different congestion levels and the result is demonstrated in Figure 4.4. As shown in Figure 4.4, although parametric yield loss is almost indifferent to congestion in 32-nm node, it is strongly correlated with congestion level and even dominates functional yield loss in some cases in 22-nm node. Therefore, for highly congested designs, although functional errors are main concern now, parametric yield is becoming a critical factor for future technologies.

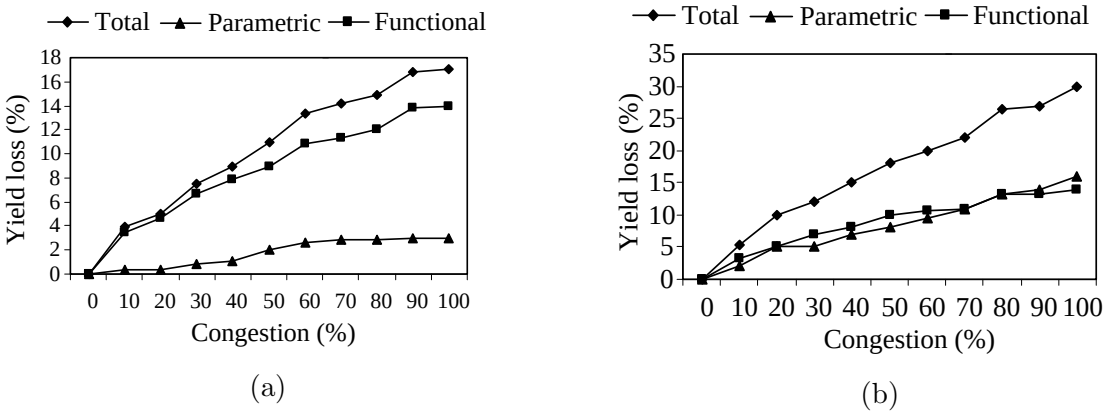


Figure 4.4: Components of yield loss for different congestion levels a) 32-nm node b) 22-nm node.

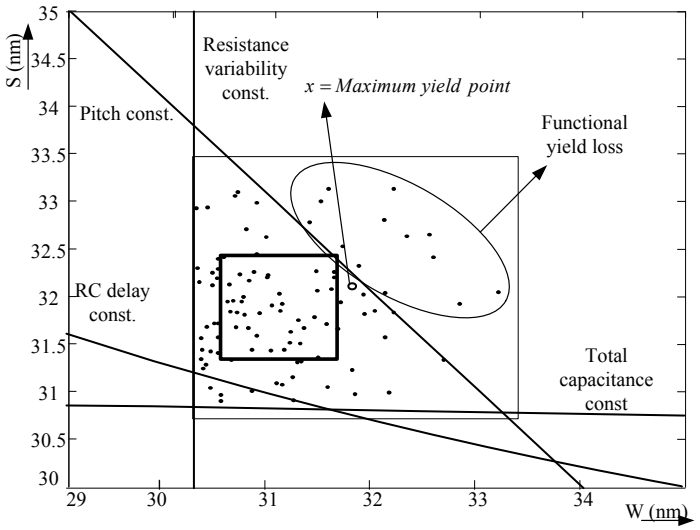
To show the criticality of each design constraint, the parameter spaces and the locations of optimal solutions for an average-case bin printed by negative-tone lithography in 32- and 22-nm nodes are depicted in Figure 4.5. In this figure, the smaller dots within the tolerance box represent possible design realizations obtained by Monte Carlo simulation for wires with length 100um.

Figure 4.5 conveys some insightful information about the trend of overlay error and yield loss for present and future technologies as:

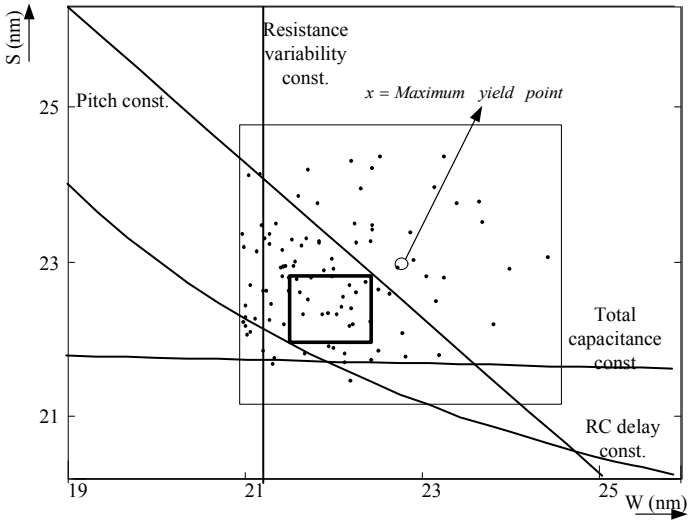
- For 32-nm node, most of the yield loss instances, those which are outside of feasible region, mainly violate the objective feature size, while few cases violate parametric constraints. These violations of feature size can result in undesired gap or connection between wires. In other words, the functional yield loss is more detrimental than the parametric yield loss in 32-nm technology. On the other hand, the parametric yield loss increases considerably in 22-nm technology.

- Among different parametric constraints for 32-nm node, RC delay is the most critical constraint. Moreover, resistance variability gets the second rank because it is met marginally in many of the design realizations and may be violated in case of stronger overlay errors. Additionally, 32-nm node is fairly robust against the violation of total capacitance constraint because none of design realizations violates the constraint.

- The relative contribution of different constraints is different for 22-nm node where a considerable part of parametric yield loss occurs due to the violation of resistance



(a)



(b)

Figure 4.5: Feasible region optimum interconnect dimensions a) 32-nm node: $\overline{RC_{total}} = 2075ps/mm$, $\overline{C_{total}} < 1.9pF/cm$, and $\Delta_R < 32\%$ b) 22-nm node: $\overline{RC_{total}} = 3128ps/mm$, $\overline{C_{total}} < 1.5pF/cm$, and $\Delta_R < 33\%$.

variation. Moreover, total capacitance becomes more critical compared to 32-nm

node.

- The maximum yield point is located outside of feasible region for 22-nm node. It means we should either relax the constraints or improve overlay control; i.e. tighter overlay budget, to gain reasonable yield for 22-nm technology. Moreover, more accurate DP methods such as SADP may be considered to replace LELE for printing critical interconnect layers.

4.5 Conclusion

In this chapter, a statistical method was proposed to determine the interconnect width and spacing for LELE DPL. The method accounts for the systematic overlay error in the interconnect dimensions as well as random overlay errors. In addition, the interconnect dimensions were chosen to satisfy the constraints of RC delay, resistance variability, and total capacitance. The developed method is flexible and it involves a small initial infrastructure in terms of mathematical computations. The proposed method has been applied for positive- and negative-tone LELE DPL methods at different technology nodes. Our analyses reveal that:

- Functional yield loss is more detrimental than parametric yield loss in 32-nm technology.
- For 32-nm node, RC delay is the most critical constraint and resistance variability gets the second rank.
- 32-nm node is fairly robust against the violation of coupling capacitance constraint.
- A considerable part of parametric yield loss occurs due to the violation of resistance variation in 22-nm node.
- Coupling capacitance becomes as critical as RC delay for 22-nm node.
- To gain reasonable yield for 22-nm technology, we should relax the constraints, improve overlay control, or use DFM techniques such as wire spreading. Such constraint relaxation or overlay budget squeezing impose considerable costs on circuit designers and fabrication facilities, respectively. As a future work, a tolerance design method can be applied to improve the expected design yield with the minimum cost overhead.

Chapter 5

Litho-friendly Decomposition for SADP

5.1 Introduction

Amongst all the DPL strategies, LELE method was originally preferred due to its layout decomposition and process simplicity. However, its high sensitivity to overlay has prompted the search for another method that is less sensitive to imaging overlay between the two exposures. Self-aligned double patterning (SADP) has been considered as the most overlay-robust double patterning process to replace LELE. This method had been mainly used in FLASH-specific applications; however, is being considered for M1/M2 interconnects most recently [25]. To this end, the basic SADP method followed by a lithography step, which is used to trim part of spacers and print irregular patterns, has been used to print 2D structures [41]. Based on experiments presented by [41], the achievable pitch resolution of this method depends on how litho-friendly the trim mask is.

Fig. 5.1 exemplifies the detailed steps of a negative-tone SADP process, where spacers are used to establish layout trenches. In the first step of the negative-tone SADP, as shown in Fig. 5.1b, one group of design patterns is printed by the first mask, so-called Mandrel mask. Subsequently, on the sidewalls of the Mandrel patterns, a spacer layer is deposited and etched back for pitch splitting as shown in Fig. 5.1c. Next, a dielectric layer, demonstrated as yellow grids in Fig. 5.1d, fills the gaps on the surface using a deposition process followed by a surface smoothing process. Finally, the Trim mask is exposed during the second lithography process to trim away sections of Mandrel and spacer patterns, and to print part of original patterns which have not been formed neither by Mandrel mask

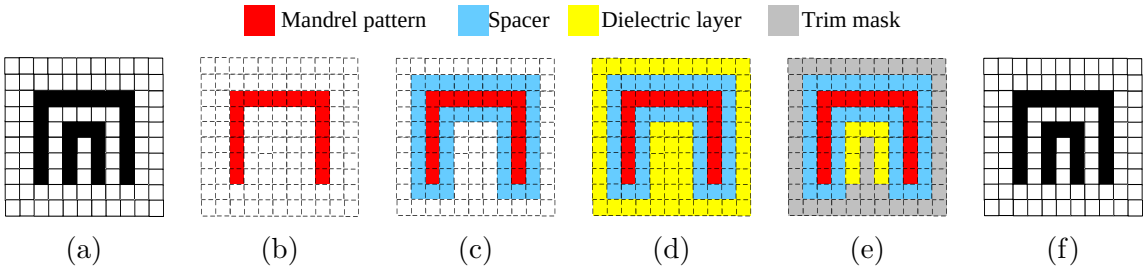


Figure 5.1: Negative-tone SADP process (a) Target pattern, (b) Mandrel lithography, (c) Spacer deposition and etching, (d) Dielectric filling, (e) Trim lithography, (f) Final etch and metal filling.

nor by spacer patterns. The process is finalized by another etching step followed by filling the formed trenches, which are the actual target patterns. As discussed in Chapter 1.2, SADP can also be done in a positiove-tone process in which spacer patterns form the target patterns. Since positive-tone SADP is more limited that negative-tone one with respect to supporting differnt pattern widths in the layout, it is less attractive for layout designers. Therefore, we focus on negative-tone SADP in this work.

Layout decomposition, a major challenge for all DPL methods, is usually translated into a graph coloring problem. Given a layout, a graph-oriented decomposition method generates a pitch-conflict graph, where the nodes and edges represent layout features and pitch conflicts, respectively. For a two-dimensional complex layout, the corresponding pitch-conflict graph is not always two-colorable. To decompose a non-colorable structure, some of its polygons need to be split between two masks. However, since this pattern stitching increases the sensitivity of LELE masks to overlay error, stitch minimization is a primary objective for layout decomposition [57]. In SADP, pattern splitting is not even permitted because it causes unresolvable gap failures.

Several layout decomposition methods have been proposed for LELE double patterning, all of which try to resolve pitch conflicts with minimum number of pattern cuts [10, 77, 26]. However, LELE decomposition methods do not address SADP-specific requirements; i.e.

1. no pattern cut is permitted.

2. the trim mask should be highly litho-friendly.

Although a post-layout decomposition cannot address the first requirement, a smart decomposition can play a significant role in designing litho-friendly trim masks and consequently improving the overall printability of layout. Focused on the litho-friendliness

of mandrel mask, [78] proposed a SADP decomposition method for 2D complex layouts. [49] and [79] formulated the overlay-robustness of both mandrel and trim masks as 3-SAT and ILP problems, respectively. In addition to the NP-hard complexity[1]. of these formulations, both [49] and [79] need ultra-fine grids, where each tile is of the size of overlay. This increases the runtime complexity dramatically. Moreover, these methods were not able to suggest partial layout decomposition for non-decomposable layouts. Partially-decomposed solutions can help designers to modify non-decomposable layouts and generate fully-decomposable ones.

In this chapter, we first discuss the effective parameters in the printability of SADP decomposed layouts. Next, we propose a grid-based ILP formulation of SADP decomposition designed to avoid decomposition conflicts and improve overall printability of layout patterns. The tile size of our grid-based approach is half-pitch. Aside from large complexity, all ILP-based methods are only applicable to fully-decomposable layouts. To overcome this barrier, a partitioning-based method is also proposed. By working on the edges of layout patterns, this decomposition method is faster than ordinary grid-based ILP decomposition methods.

The rest of the chapter is organized as follows: the EDA requirements of litho-friendly SADP decomposition are described in Section 5.2. Section 5.3 introduces the proposed decomposition methods. The details of trim mask synthesis are discussed in Section 5.5. Experimental results are reported in Section 7.4 and finally, Section 5.7 concludes the work.

5.2 Decomposition requirements in SADP method

In this section, the main concerns for SADP decomposition are discussed. The two-dimensional SADP methods are discussed in detail in Section 2.2. Similar to all DPL methods, general-purpose SADP needs to decompose the original design patterns into two groups under the minimum spacing constraint. The major decomposition issues, which make conventional LELE decomposition methods inadequate for SADP, can be categorized as: 1) layout decomposability 2) litho-friendliness of the proposed decomposition.

[1]A problem is non-deterministic polynomial-time hard (NP-hard) if and only if it is at least as hard as an NP-complete problem. General SADP layout decomposition problem is proven to be at least as hard as 3-SAT problem [48] which is NP-complete.

5.2.1 Layout decomposability

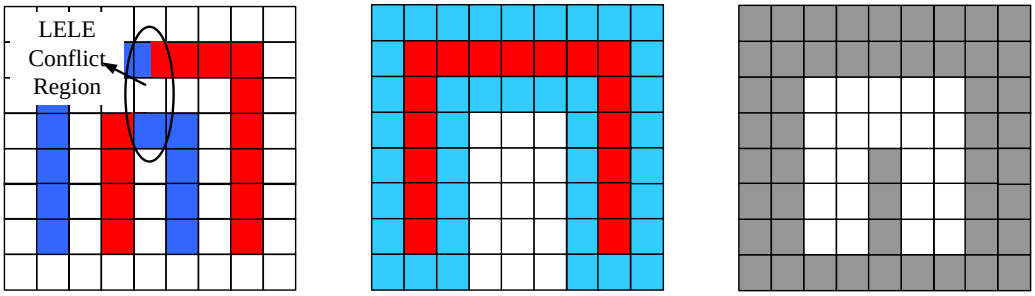
Layout decomposability problem is to determine whether a given layout is manufacturable by the desired double patterning method. Layout decomposability criteria depend strictly on process specifications and thus vary for different double patterning methods.

Including an exact subset of original layout patterns, one of the LELE masks is the complementary of another one. Therefore, in LELE method, odd-cycle test is applied on the pitch-conflict graph to check whether the given layout is two-colorable or not. Odd-cycle test can be performed in polynomial time using breath first search (BFS) on the pitch-conflict graph [77]. In the case of any coloring conflict, simple design rule checks can be used to determine if the detected coloring conflict is resolvable by pattern split or not [24].

Layout decomposability in SADP method is not as straightforward as LELE method because SADP masks are not exact subsets of the original layout. In the SADP method, in addition to the original layout patterns, a set of assist patterns may be added to the mandrel mask and be cleaned later by the trim mask. In other words, a target pattern can be printed in several ways using different schemes on mandrel and trim masks. Although this opportunity allows to print a set of patterns which are not manufacturable by the LELE method, for example, the pattern shown in Figure 5.2 enlarges the search space of SADP decomposability problem drastically. In this figure, the inner U shape cannot be printed either by a single mask, due to minimum pitch conflict between the two legs, or by two LELE sub-masks, due to pitch conflict with the outer shape.

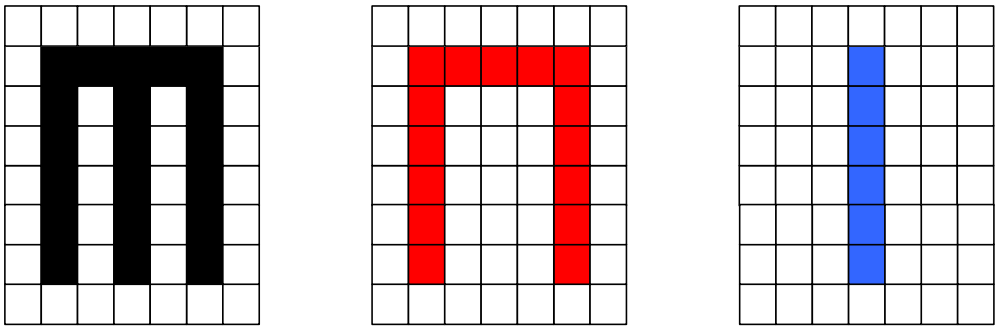
Moreover, the SADP constraint of not being able to split patterns further complicates arriving to an SADP decomposable condition. Due to this restriction, if a layout polygon is assigned to the mandrel mask, it must be assigned to the same mask entirely. Therefore, a set of printable patterns in LELE become non-printable in SADP. Figure 5.3 exemplifies a LELE compliant pattern which is not printable by SADP.

Formulating general SADP decomposition problem as a 3-SAT problem, [48] proved that there is no polynomial solution for SADP decomposability. In this chapter, we show that this large search space can be pruned considerably because a large subset of SADP decomposition candidates are highly sensitive to lithography errors and can be discarded from the search space.



(a) (b) (c)

Figure 5.2: A LELE non-decomposable, SADP decomposable pattern (a) irresolvable LELE conflict, (b) SADP mandrel pattern (red) and spacer patterns (blue), (c) SADP trim mask (gray).



(a) (b) (c)

Figure 5.3: A SADP non-decomposable pattern that is LELE decomposable (a) target pattern, (b) first mask, (c) second mask.

5.2.2 Litho-friendliness challenges in SADP decomposition

In addition to computational intensity of SADP decomposability problem, SADP has extra complications compared to LELE processes. Among possible legal SADP decompositions for a layout, many of them suffer from poor patterning quality because they do not address SADP-specific litho-friendliness demands, as discussed later.

Breaking odd cycles and assigning the nodes of conflict-graph to either of the masks are the final steps of LELE decomposition, which can be performed in polynomial time. Of course, problem complexity increases when extra objectives such as minimum pattern cut [10], decomposition robustness [77], and balanced mask density [45] are desired. However,

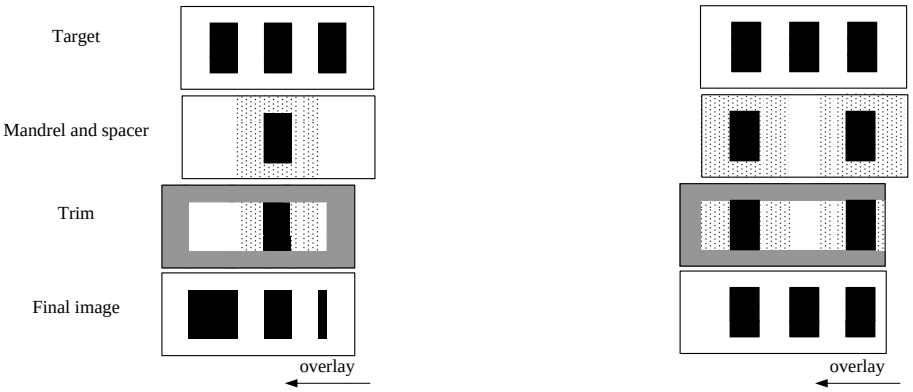
due to inherent differences between SADP and LELE methods, mask decomposition is more challenging for SADP. These distinguishing requirements of SADP can be categorized as:

1. more complex rules describing design rule compliance;
2. decomposition-dependent sensitivity to overlay error;
3. imaging issues in trim lithography.

Preserving design rule compliance of both SADP sub-masks is more complicated than in LELE ones because SADP sub-masks can differ greatly from the original layout. As a result, design rule violations can arise in mask-features of either sub-mask which may not have a one-to-one relation with original layout features due to the addition of necessary dummy patterns. The mandrel mask can be inspected locally to detect such conflicts; however, the trim mask needs a global DRC check.

Another litho-friendliness concern in two-dimensional SADP method is about its susceptibility to the overlay error. Being applicable only for one-dimensional layouts, basic SADP process is highly robust to overlay error because it includes only one lithography step followed by a sequence of accurate chemical processes; i.e. spacer deposition, dielectric fill, and final etching. To expand SADP capabilities for two-dimensional patterns, additional trim lithography step is mandatory after the dielectric fill step. Therefore, the actual level of accuracy depends strictly on input layout and the applied decomposition style. Figure 5.4 shows two valid SADP decomposition schemes for a target layout. As shown in Figure 5.4.(a), a careless decomposition can result in considerable EPE in the final image. As a rule of thumb, overlay sensitivity is larger in regions where the trim mask touches a target pattern.

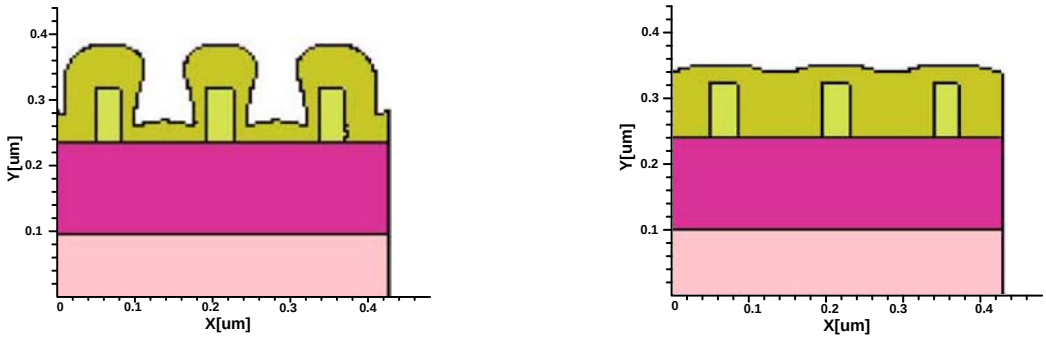
In addition to overlay between mandrel and trim masks, the major source of image imperfections in SADP is the lower accuracy in the trim lithography. According to simulation results reported by [41], the trim mask is highly sensitive to defocus errors greater than 40nm. The criticality of defocus error becomes more challenging when we notice that mandrel lithography and spacer deposition steps cause considerable surface roughness, which increases the DoF variation during the trim lithography. Figure 5.5 shows the result of a topography simulation after spacer deposition, Figure 5.5.(a), and dielectric deposition, Figure 5.5.(b), steps [15]. As shown in this figure, in spite of using a hypothetical highly conformal deposition step and an extra smoothing step [15], some residual surface roughness remains that can cause considerable defocus error. This defocus error is difficult to compensate in the scanner due to the very short length scale in which it occurs.



(a) (b)

Figure 5.4: The role of SADP decomposition on overlay sensitivity (a) sensitive decomposition, (b) robust decomposition.

To alleviate the effects of DoF variation, maximum uniformity of pattern density on both mandrel and trim masks is desirable. Achieving such level of uniformity depends strongly on the characteristics of the original layout and the quality of the applied decomposition method. To illustrate how interactions of mandrel and trim masks affect the quality of the final image, we simulated the image contour of a trim pattern which is located in different neighborhoods with respect to its surrounding mandrel patterns. The simulated image contour (the top figure) and trim mask (the bottom figure) for each test case are shown in Figure 5.6. The optical parameters for these simulations are wave-

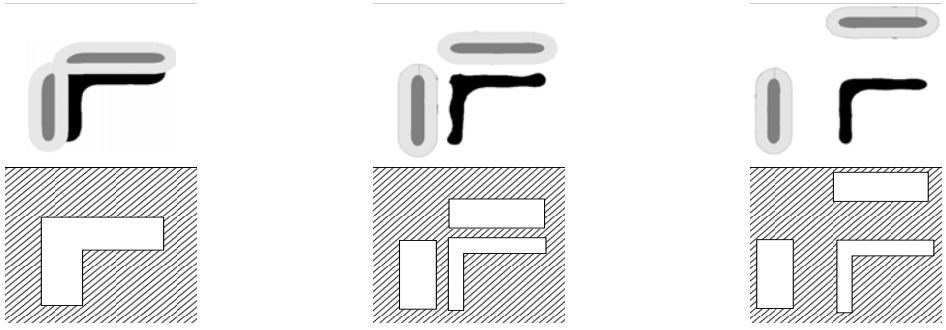


(a) (b)

Figure 5.5: Topography simulation of DPL process [15] (a) Spacer layer deposition, (b) Dielectric deposition.

length=193nm, numerical aperture(NA)=1.35, dipole illumination source, defocus=± 50, and overlay error=5nm. The simulation results convey the following points about the impacts of mandrel-trim mask interactions on the overall printability of the layout:

- Figure 5.6.(a) shows that those edges of trim patterns protected by the sidewalls of spacer patterns are printed more robustly than any other edges.
- The minimum pitch violation is a necessary but not sufficient criterion to perform litho-friendly SADP decomposition. For example, Figure 5.6.(b) represents the inherently non-litho-friendly patterns which cannot be decomposed in a litho-friendly way by any post-layout decomposition method. In this figure, although all conflicting patterns are assigned to different masks, the image quality is lower than other test cases. In addition to complicated target shapes of the trim mask in Figure 5.6.(b), defocus errors due to neighbor mandrel patterns worsen the image resulting from the trim mask.
- Figure 5.6.(c) illustrates that isolated trim patterns which suffer from lower level of pattern distortion compared to the semi-isolated ones, i.e. Fig 5.6.(b). This phenomenon occurs due to smoothness of mandrel layer in sparse regions which decreases the defocus error. Moreover, there is higher freedom in sparse regions to perform aggressive OPC and enhance pattern transfer.



(a) (b) (c)

Figure 5.6: Interactions of mandrel (dark gray), spacer (light gray) and trim (black) patterns affect the quality of final image contour (top) and the complexity of trim mask (bottom). (a) Constructive interaction, (b) Destructive interaction, (c) Neutral interaction.

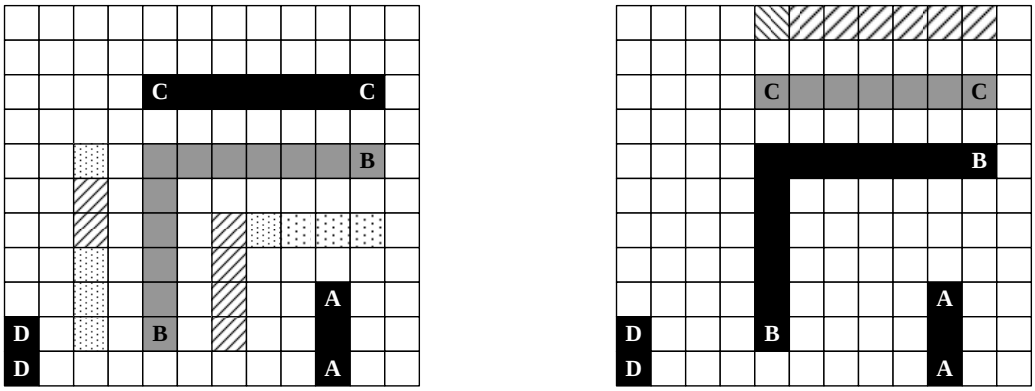


Figure 5.7: Printability improvement of non-mandrel patterns (Gray) using dummy mandrel patterns.

To improve the imaging quality of trim patterns, it is highly recommended to protect them by mandrel patterns. This protection can be performed by either original or dummy mandrel patterns. Dummy mandrel patterns are patterns which are inserted in the layout only to protect the sensitive trim patterns. As long as the minimum pitch and OPC requirements are met within the mandrel mask, these dummy patterns do not impact the functional and electrical characteristics of the design because they will be cleaned by the trim process later on. However, minimum pitch and OPC constraints impose strict limitations on dummy pattern insertion. Moreover, we use only floating dummy mandrel patterns to avoid the trim mask touching original mandrel patterns and consequently increasing the sensitivity to overlay error.

Figure 5.7 shows the challenges of dummy pattern insertion in the mandrel layer. In this figure, the dashed tiles represent the inserted dummy patterns and the dotted ones represent the candidates for dummy pattern insertion which are rejected due to the minimum pitch constraint. The domain of eligible tiles for dummy pattern insertion will be confined further if we consider OPC requirements of mandrel patterns. As shown in Figure 5.7.(b), an alternative decomposition scheme puts lower restrictions on the insertion of dummy patterns.

In summary, in addition to conventional objectives of DPL decomposition, a SADP-specific decomposition should consider two extra objectives:

1. Increase the length of the trim edges which are protected by original mandrel patterns. Original mandrel patterns are preferred to dummy ones for trim protection

because they do not raise extra design rule concerns, mask synthesis and OPC complications.

2. Increase the possibility of dummy pattern insertion for those edges that are not already protected. For example, in Figure 5.7, the pattern C is a better choice to be assigned to the trim mask than the pattern B because some parts of pattern B cannot be protected either by original or dummy mandrel patterns.

5.3 ILP-based litho-friendly SADP decomposition

In this section, we present an ILP formulation for SADP decomposition method.

5.3.1 Terms and definitions

Before formulating the problem, we first define terms which are used in formulation. The decomposition-independent and decomposition-dependent definitions are illustrated in Figure 5.8 and Figure 5.9, respectively. In all of the following definitions, we assume that the layout is fully decomposable which is essential for this SADP decomposition process.

Definition 1 *A segment of the boundary of layout patterns with the length of half-pitch is an* ***occupied fragment (OF)****.*

Each OF must be assigned to either M(andrel) or T(rim) mask. The mandrel fragments are printed by mandrel lithography and the trim fragments are formed either by sidewall patterns or trim lithography.

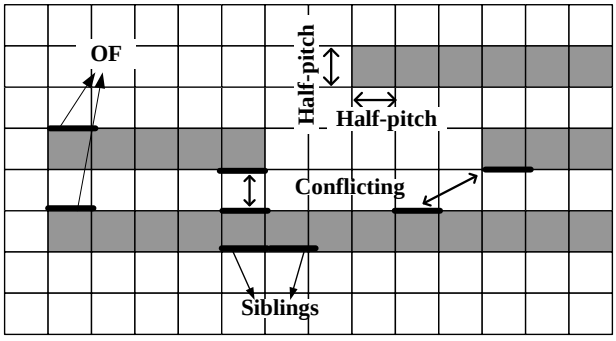


Figure 5.8: Basic definitions

Definition 2 *Two OFs are* ***conflicting*** *if they do not belong to the same net and their orthogonal distance is shorter than minimum pitch.*

Definition 3 *Two OFs are* ***siblings*** *if they belong to the same net and touch each other.*

As discussed before, the major challenge in SADP is providing a conflict-free decomposition. Moreover, the pattern stitching is not acceptable in SADP. Using the above definitions, these two constraints can be translated into the following criteria:

1. Conflicting OFs should be assigned to opposite masks.
2. Sibling OFs should be assigned to the same mask.

Definition 4 *A trim OF which is not assisted by a sidewall pattern is referred as a* ***bare*** *fragment.*

The lack of a parallel mandrel OF next to a trim OF turns it to a bare OF. Since bare OFs experience higher levels of pattern distortion, it is required to assist them by inserting dummy mandrel patterns next to them. However, this dummy pattern insertion is not always possible due to pitch conflicts with original mandrel patterns.

Definition 5 *A trim OF is* ***sensitive*** *if it is bare and cannot be assisted even by a dummy mandrel pattern. Otherwise, it is* ***safe****.*

Definition 6 *A mandrel OF is* ***interfering*** *if it forbids insertion of dummy mandrel patterns for demanding bare OFs.*

The time complexity of determining whether an OF is a potential sensitive fragment or not and finding its interfering neighbors is $O(1)$.

Figure 5.9 exemplifies above definitions, where original mandrel, dummy mandrel, and trim patterns are shown as solid black, dashed black and gray ones, respectively. In Figure 5.9, the interfering mandrel fragments are marked by bold gray edges. These interfering fragments yield sensitive trim fragments which are shown as bold black edges in Figure 5.9. Suffering less sensitive fragments, the decomposition scheme shown in Figure 5.9.(a) is more robust than the one in Figure 5.9.(b). Here, the provided examples are completely on-the-grid for more clarity; i.e. width and height of patterns and spaces between patterns are multiples of half-pitch. However, all of the defined terms and the proposed formulation are applicable for off-the-grid layouts as well.

Definition 7 *An OF hesitates about being assigned to mandrel mask if, as a mandrel fragment, it interferes in one or more trim OFs. This* ***mandrel hesitation (MH)*** *to mandrel mask increases as the fragment causes more sensitive trim OFs.*

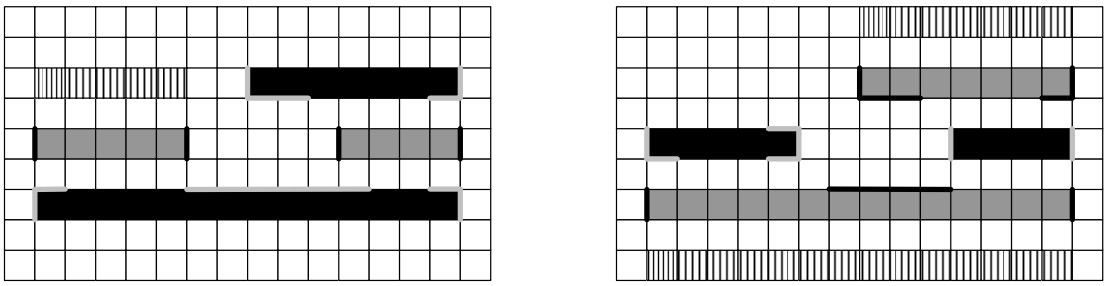
Definition 8 *An OF hesitates about being assigned to trim mask if, as a trim fragment, it is interfered by at least a mandrel OF. Since* ***trim hesitation (TH)*** *is a primary concern, we penalize a trim fragment by the largest penalty even if it is interfered by only one mandrel fragment.*

To improve the overall printability in the final layout, we prefer a solution with lower overall reluctance of fragments to their corresponding masks. Formally, we formulate the SADP layout decomposition as follows:

Problem Formulation: *Given a decomposable layout, decompose it into two parts (mandrel and trim) such that the overall hesitation over both masks is minimized.*

5.3.2 ILP formulation

In this section, we present our SADP layout decomposition based on integer linear programming (ILP). After mapping the input layout to grid, we will process the occupied fragments and formulate the ILP formulation. In grid processing, conflicting, sibling, and potential interfering fragments are identified for each OF initially. We ignored mutually interfering fragments because they increase the ILP complexity without providing a valuable guide for ILP solver. Afterwards, to decrease the size of problem, the mask neutral siblings,



(a) (b)

Figure 5.9: Assessment of robustness for different decomposition schemes (a) shorter sensitive (black bold) edge length, (b) longer sensitive edge length.

whose both MH and TH parameters are zero, are merged and replaced by a representative OF which contains neighborhood data of original fragments. The notation is described in Table 5.1.

Table 5.1: Notation for ILP formulation.

of_i	i^{th} occupied fragment
N	Number of occupied fragments
x_i	Binary variable that shows the selected mask for of_i. $x_i = 0$ if the mandrel mask is selected,otherwise, the trim mask is selected.
CF_i	The set of conflicting fragments for of_i
SF_i	The set of sibling fragments of of_i
IF_i	The set of potential fragments which are interfered by of_i excluding mutual interfering ones.
$\|IF_i\|$	The cardinality of the set IF_i that depends on the neighborhood of of_i.

The litho-friendly SADP decomposition can be formulated as:

$$Minimize \sum_{i=1}^{N} \frac{1}{|IF_i|} MH_{of_i} + TH_{of_i} \tag{5.1a}$$

Subject to:
for of_i in OFs:

$$x_i + x_j = 1 \qquad of_j \in CF_i, \tag{5.1b}$$

$$x_i - x_j = 0 \qquad of_j \in SF_i, \tag{5.1c}$$

$$(1 - x_i) + \sum x_j \leq (|IF_i| - 1)x_i + 1 + MH_{of_i} \quad of_j \in IF_i, \tag{5.1d}$$

$$x_i + (1 - x_j) \leq 1 + TH_{of_i} \quad of_i \in IF_j. \tag{5.1e}$$

The objective function (7.3) is to minimize the total mandrel hesitation for mandrel fragments and trim hesitation for trim fragments. It is noteworthy that mandrel and trim hesitation parameters are zero for trim and mandrel fragments, respectively. As explained later in this section, the upper bound of TH_{of_i} is 1 while the upper bound of MH_{of_i} is

$|IF_i|$. Therefore, the parameter $\frac{1}{|IF_i|}$ is used to normalize the MH_{of_i} with respect to TH_{of_i}.

The first constraint (5.1b) forbids the conflicting fragments from being assigned to the same mask. The second constraint (5.1c) guarantees all fragments of a net are assigned to the same mask; consequently, no pattern stitching occurs. Constraints (5.1d) and (5.1e) are used to calculate the reluctance of the i^{th} fragment to mandrel and trim masks. In constraint (5.1d), fragments x_j belong to the interference set of the fragment x_i, i.e. IF_i as defined in Table 5.1. Accordingly, whenever fragment i is assigned to the mandrel mask, i.e. $x_i = 0$, the parameter MH_{of_i} counts the number of trim fragments which are interfered by the fragment i as a mandrel fragment. On the other hand, by adding the term $(|IF_i| - 1)x_i$ to RHS of constraint (5.1d), it becomes neutral when the fragment i is assigned to trim mask. The set of constraints (5.1e) are issued for all interfering fragments of fragment i. These constraints capture whether the fragment i is a sensitive trim fragment or not, where TH_{of_i} becomes 1 as the fragment i is assigned to the trim mask, i.e. $x_i = 1$, and is interfered by at least a mandrel fragment, i.e $x_j = 0$. Parameter TH_{of_i} is a binary parameter and becomes 1 if the fragment i is interfered by only one fragment.

5.4 Partitioning-based SADP decomposition

Although the proposed ILP formulation can result in the optimal litho-friendly decomposition, it suffers from following limitations.

1. A 0-1 ILP does not converge for non-decomposable layouts. Moreover, it cannot provide an approximate solution which decomposes the decomposable parts of the layout.

2. Any ILP-based method is NP-hard and suffers long runtime. This issue is more disturbing for our problem which needs a fine grid.

3. Handling the balanced density between mandrel and trim masks is not possible in ILP formulation.

To resolve the above issues, we propose a partitioning-based SADP decomposition in this section. Using a weighted conflict graph, this method divides layout polygons to mandrel and non-mandrel partitions such that the total length of sensitive edges is minimized. The key features of the partitioning-based decomposition can be categorized as:

- A weighted directed graph, decomposition graph (DG), is used to consider both pitch conflicts and litho-friendliness demands simultaneously.
- For partially-decomposable layouts, the proposed method suggests a reasonable solution, which provides the lowest pitch conflicts and sensitive edges. Although the provided solution is not completely legal and hence manufacturable, it can provide insightful guide for layout designers to resolve un-manufacturable features of the layout with minimal effort.
- In contrast to ILP-based method which is grid-based, the partitioning-based method is edge-based inherently. Consequently, this method is also applicable for grid-less layouts. In the following, the applied definitions and examples are kept grid-based to be consistent and comparable with the ILP-based method. The extension of the definitions to edge-based ones, which are applicable to grid-less layouts, is straightforward.

5.4.1 SADP-aware conflict graph

In this section, we explain the structure of a decomposition graph that stores both pitch-conflict and litho-friendliness factors for SADP decomposition. To this end, at first, a weighted directed graph is constructed from the input layout, where layout polygons are represented by nodes and pitch-conflict and mask-preference factors are represented by edges. Consequently, the optimal litho-friendly SADP decomposition is equal to minimum cut bi-partitioning of this decomposition graph (DG). Afterwards, we manipulate the preliminary structure of this DG to simplify optimal graph decomposition.

In our preliminary DG, the layout polygon (i) is represented by a node (P_i), which is connected to its neighbors by weighted directed edges. These edges model the interactions between adjacent polygons and show the mask preference of each polygon with respect to its neighbors. Our graph includes two types of edges each of which models the hesitation of their sources to mandrel or trim mask. We call these edges mandrel hesitation edges (MHEs) and trim hesitation edges (THEs), respectively. Moreover, to consider all SADP requirements, each edge in DG, say e, has the following parameters:

1. **Lithography cost parameter (c_e)**: the cost of an edge represents the hesitation of its source node in being assigned to a specific mask due to the interference caused by the destination node. For example, consider two neighbor polygons i and j. The cost of the edge $THE_{i\rightarrow j}$ ($MHE_{j\rightarrow i}$) shows how much the polygon i is interfered by

the polygon j assuming nodes P_i and P_j are assigned to non-mandrel and mandrel partitions, respectively. Here, the interference occurs whenever part of a trim polygon cannot be assisted by either an original or a dummy mandrel pattern. The larger cost means the longer length of the of the trim pattern is left bare due to inappropriate decomposition.

2. **Separation demand parameter (dem_e)**: the separation demand parameter of an edge shows how much its end nodes are likely to be assigned into different masks due to pitch conflict. This parameter depends on the conflict length and the distance between two neighbor polygons. Since total EPE due to pitch violation is proportional to the length of the violation region, the separation demand grows for polygons with longer conflicting edges. On the other hand, separation demand increases as the distance between polygons and consequently the available area for OPC decreases.

For a sample layout shown in Figure 5.10.(a), Figure 5.10.(b) demonstrates the structure of the preliminary DG where MHEs and THEs are demonstrated by solid and dashed lines, respectively. In this decomposition graph, the separation demand between two nodes is calculated based on the maximum pitch conflict that one of the end nodes experiences by another one. For instance, the separation demand between nodes *1* and *4* equals *8* because the lengths of conflicting edges, highlighted as red edges in Figure 5.10.(a), between the nodes *1* and *4* are *7* and *8* grids, respectively. In the Figure 5.10.(a), the grids that determine the $MHE_{2\to3}$ and $THE_{2\to3}$ are colored in yellow and blue, respectively.

5.4.2 Problem formulation

To evaluate different partitioning schemes of DG, we define the cutset parameter as:

Definition 9 *The **cutset** of a partitioning configuration is the summation of the cost of MHEs which leave the mandrel partition and the cost of THEs which leave the non-mandrel partition.*

The cutset defined by Def. 9 is a metric for litho-friendliness assigned to each decomposition candidate because it penalizes the decomposition for interfered trim edges as well as interfering mandrel edges. For the layout shown in Figure 5.10, Figure 5.10.(a) and Figure 5.10.(b) show two partitioned DGs which address all separation demands. In these partitioning solutions, which are equal to decomposition schemes shown in Figure 5.9, those edges that collaborate on cutset are drawn bold. Benefiting from lower cutset,

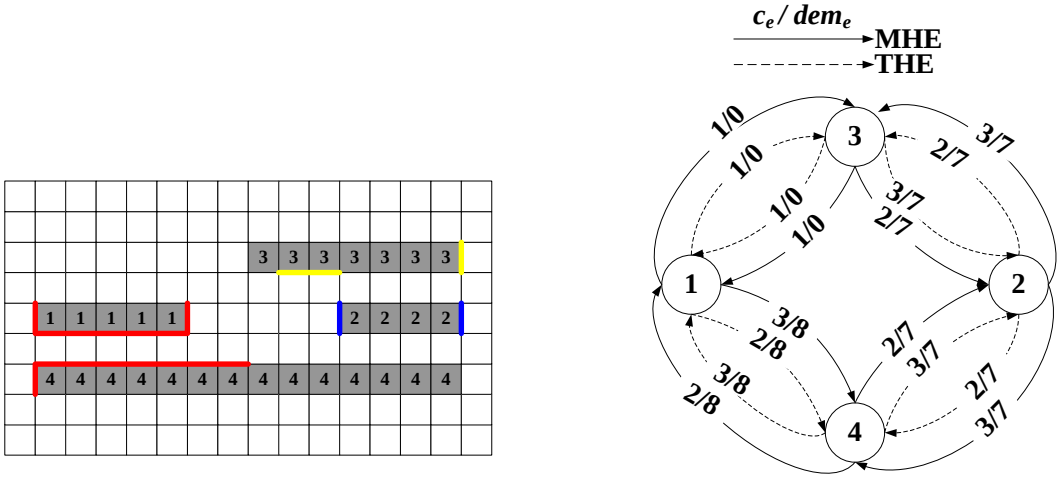
the decomposition scheme in Figure 5.10.(a) is more litho-friendly than the decomposition depicted in Figure 5.10.(b).

Based on the cost metric proposed in Def. 9, the litho-friendly SADP decomposition can be formulated as:

Partitioning-based SADP decomposition: *Let $DG = (V, E)$ be the decomposition graph of a given layout where $E = \{MHEs \cup THEs\}$. Our goal is to divide the graph into two parts mandrel (M) and non-mandrel (NM) $= \{V - M\}$ to minimize the ratio of the total cut cost to the separated demands:*

$$\left(\sum_{\substack{(i,j)\in MHE \\ i\in M}} c_{ij} + \sum_{\substack{(i,j)\in THE \\ i\in NM}} c_{ij} \right) \Bigg/ \sum_{\substack{(i,j)\in E \\ i\in M \wedge j\in NM \\ i\in NM \wedge j\in M}} dem_{ij}.$$

The partitioning-based SADP decomposition is equal to the *bipartite directed sparsest cut* problem which is known as an NP-hard problem [80]. Therefore, some approximation methods were proposed for this problem [81, 82]. However, it has been proved that the problem cannot be approximated within the accuracy interval of $O(n^{\delta})$ in polynomial time [83]. To achieve a reasonable runtime as well as an acceptable level of litho-friendliness, we tweak the DG structure somehow that the new graph can be handled by existing heuristic



(a) (b)

Figure 5.10: Structure of SADP decomposition graph (a) Target layout, (b) Decomposition graph.

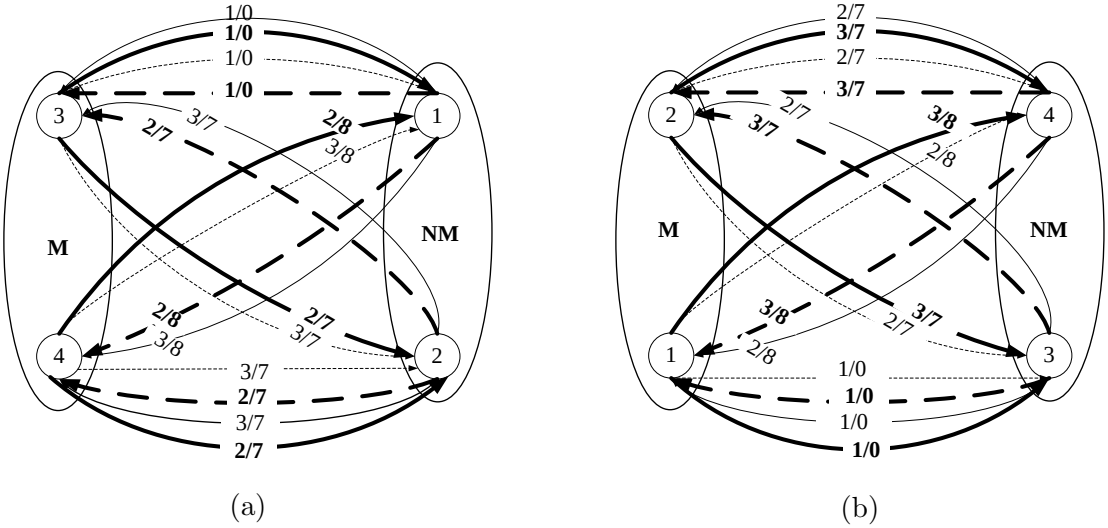


Figure 5.11: The cutset metric represents the level of litho-friendliness of a decomposition. (a) Cutset=14 and Demand=44 (b) Cutset=20 and Demand=44.

methods for graph partitioning. Afterwards, we modify the FM algorithm [84] to perform SADP-aware partitioning.

5.4.3 SADP-aware FM bi-partitioning

In this section, we propose a modified FM partitioning to perform litho-friendly SADP decomposition problem. To this end, we first change the structure of DG graph into a graph which is simpler to handle by FM partitioning.

Since FM partitioning cannot handle two lithography cost and separation demand parameters for edges simultaneously, an additional type of edge is needed in the decomposition graph which accounts for the separation demands, namely pitch conflict edges (PCEs). Mask hesitation edges represent for undesirable printability degradations and thus should be avoided being cut as much as possible. In contrary, we are interested in increasing number of cuts on PCEs because such cuts represent for addressed pitch resolution demands. However, simply using negative weights for PCEs cannot guide the partitioner for finding conflict- free decomposition.

To resolve this issue, each node (P_i) of the DG is split into two nodes, each of which undertakes to represent either lithography cost or separation demand parameters in the

preliminary DG. These nodes, which are always assigned into opposite partitions, are defined as:

1. **Mask node** ($P_{i,m}$): a mask node determines that its corresponding polygon is printed by which mask; i.e., the polygon is printed by the mandrel mask if its mask node is assigned to the mandrel partition; otherwise, it is printed by sidewalls and the trim mask. The mask preference of a layout polygon is evaluated by MHEs and THEs which are connected to its corresponding mask node.

2. **Conflict node** ($P_{i,c}$): a pitch conflict edge (PCE) connects a mask node ($P_{i,m}$) to a conflict node ($P_{j,c}$) to model the separation demand between polygons i and j. The conflict node of a polygon is assigned to the opposite mask partition to which the polygon is assigned. Therefore, pitch conflict occurs between two polygons if their interconnecting PCEs are cut.

Now, the total cut set for a partitioning can be calculated as:

$$\text{cutset} = \sum_{\substack{P_{i,m}\in M \\ P_{j,m}\in NM}} (W_{MHE_{i,j}} + W_{THE_{j,i}}) + \sum_{\substack{\{P_{i,c},P_{j,m}\}\not\subset M \\ \{P_{i,c},P_{j,m}\}\not\subset NM}} W_{PCE_{i,j}}, \tag{5.2}$$

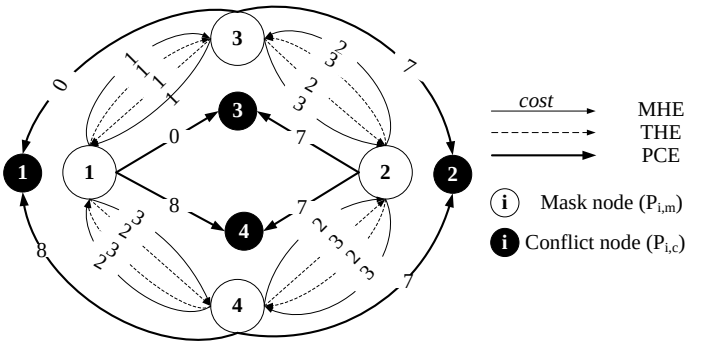
where $W_{MHE_{i,j}}$, $W_{THE_{i,j}}$, and $W_{PCE_{i,j}}$ notify weights of mandrel hesitation, trim hesitation, and pitch conflict edges, respectively.

Figure 5.12.(a) compares the preliminary and modified structures of decomposition graph, where mask nodes and conflict nodes are colored by white and black, respectively. Moreover, MHEs, THEs, and PCEs are represented by solid, dashed and solid bold lines. Those mask hesitations edges which do not impact on the cutset are not shown in Figure 5.12.(b) to help figure readability.

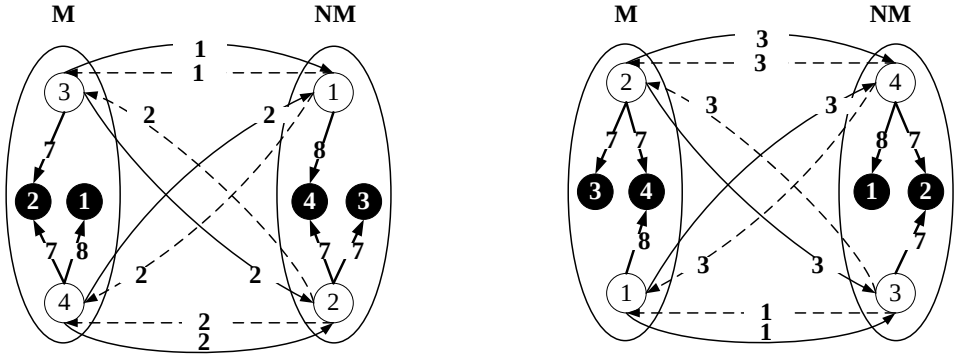
Different objectives of SADP decomposition; i.e. pitch resolution and litho-friendliness, can be prioritized by different weighting policies for different types of edges. Considering the pitch resolution as the primary objective for SADP, we assign edges weights such that the separation demands dominate the mask hesitation costs. The preference of a pattern for different masks is modeled as:

$$W_{MHE_{i,j}} = L_{inf,i\rightarrow j}, \quad P_{i,m} \in M \,\text{and}\, P_{j,m} \in NM, \tag{5.3}$$

$$W_{THE_{i,j}} = L_{sen,i\leftarrow j}, \quad P_{i,m} \in NM \,\text{and}\, P_{j,m} \in M, \tag{5.4}$$



(a)



(b)

Figure 5.12: The modified structure for the DG in Figure 5.10.

where $W_{MHE_{i,j}}$ and $W_{THE_{i,j}}$ denote weights of mandrel hesitation and trim hesitation edges between patterns i and j, respectively. Moreover, $L_{inf,i\rightarrow j}$ and $L_{sen,i\leftarrow j}$ represent the lengths of interfering and sensitive edges of the polygon i in interaction with the polygon j. Sensitive and interfering edges are defined similar to sensitive and interfering fragments defined in Def. 5 and Def. 6.

To prioritize the pitch resolution objective, we assign weights to pitch conflict edges as:

$$W_{PCE_{i,j}} = \sum_{(i,k)\in MHE} W_{MHE_{i,k}} + \sum_{(i,k)\in THE} W_{THE_{i,k}} + \alpha_1 CL_{i,j} + \alpha_2 \frac{1}{D_{i,j}}, \tag{5.5}$$

where parameters $CL_{i,j}$ and $D_{i,j}$ represent the conflict length and the distance between pattens i and j; and, α_1 and α_2 are parameters that allow minor pitch violations based on

the premise that they may become tolerable after OPC. These parameters do not affect the decomposability of the input layout; however, they can help to find a more promising solution for a non-decomposable layout. It is important to choose parameter α_1 large enough so that the minimum pitch constraints are prioritized over other constraints. In this work, α_1 is selected as the maximum of total conflict length of layout patterns with their neighbors; i.e. $\alpha_1 = \underset{i}{Max}\left(\sum_j CL_{i,j}\right)$ where i and j are layout patterns. Parameter α_2 should be relatively small to not mask the first three cost terms in (5.5). We assume $\alpha_2 = 1/\alpha_1$ in this work.

Finally, a minor modification is applied to the classic FM partitioning algorithm [84] to support SADP requirements. A conflict node only moves upon the movement of its associated mask node in SADP-FM. This strategy decreases the number of swap candidates and thus speeds up the swapping phase. Moreover, it does not degrade the optimality of final solution because, as can be seen in (5.2), knowing the current location of mask nodes is enough for cutset calculation. Therefore, being considered as *soft locked* nodes, the gain calculation is not performed for conflict nodes. When a mask node is moved into a new partition, it pushes its associated conflict node to the opposite partition and both nodes are flagged as *hard locked*.

The FM partitioning method can handle a user-defined ratio between the size of partitions. To trade-off between the balanced mask densities and other SADP objectives, we consider upper and lower thresholds for the imbalance between mask densities. The upper threshold can be violated if it helps to improve the overall cutset; however, the lower threshold can be violated if and only if it resolves minimum pitch conflicts. In this work upper and lower imbalance thresholds are considered as 0.45 and 0.3, respectively.

5.5 Synthesis of the trim mask

After partitioning the target layout into mandrel and non-mandrel patterns, the trim mask should be synthesized. For a negative-tone SADP process, the trim mask is mapped to a two-dimensional matrix, where $Trim_{i,j} = 1$ if the underneath pattern should be cleared and $= 0$, otherwise.

To synthesize the trim mask, we note to sensitive non-mandrel patterns which are not already assisted with original mandrel. Then, dummy mandrel patterns are inserted around them. The trim mask is generated in such a way that it cleans the dummy mandrel patterns and prints the non-mandrel patterns which are not formed yet.

Following the rules given below, dummy mandrel patterns are inserted conservatively to avoid the potential negative side-effects on the design rule compliance of the mandrel and trim masks.

1. Since the printability of the original mandrel patterns is a primary objective, no mandrel pattern is allowed within the minimum pitch of an original mandrel pattern. However, the minimum spacing rule is relaxed between two dummy patterns because the resulted printing errors, e.g. bridging between the dummy patterns, will be trimmed from the final image eventually.

2. Original and dummy mandrel patterns are treated by two different OPC recipes, each of which is adjusted to improve the printability for the target patterns. Assuming the deposition rate of spacer layer is constant in all directions; the spacer deposited around the corners of a mandrel pattern is not accurately rectangular, and it cannot protect a non-mandrel neighbor pattern effectively. In other words, all the concave corners of non-mandrel patterns should be trimmed by the trim mask, which increases the mask complexity significantly. To alleviate this issue, the convex corners of dummy mandrel patterns are pushed outside to make the resulted spacer patterns as rectangular as possible. However, a similar strategy is not applicable for an original mandrel pattern because it degrades the image quality of the pattern itself.

After dummy mandrel insertion, we find the subset of spacer patterns which should be preserved during the trim process. These spacer patterns, so-called *critical spacer patterns*, touch the edges of either non-mandrel or original mandrel patterns and are necessary for pitch splitting and overlay robustness, respectively. Now, the trim mask can be calculated as (7.7).

$$trim = Layout\_Area - (M_{org} + NM + SP_{critical}), \tag{5.6}$$

where M_{org}, NM, and $SP_{critical}$ are matrices representing the original mandrel, non-mandrel, and critical spacer patterns, respectively. In these matrices, patterns are represented by ones. Moreover, Layout_Area is a matrix of ones.

The edges of the resulted trim mask can be divided into rigid edges, which touch non-mandrel or mandrel patterns, and lenient edges, which interact with the spacer patterns. The printability of rigid edges is more critical than the lenient edges because any distortion along the rigid edges propagates to the final image. In addition, the lenient edges may be slid within the interacting spacer pattern as long as they do not touch any target pattern in case of worst-case overlay error. An edge can be rigid in one direction and lenient in the other direction. In Figure 5.13, the rigid and lenient edges of a sample trim mask are

bolded as solid and dotted lines, respectively. In the OPC recipe of the trim mask, we tune the lenient edges so that design rule violations are minimized and the rigid edges are printed as accurate as possible.

5.6 Experimental results

In this section, we provide experimental results which evaluate the effectiveness of proposed methods for improving the printability of SADP decomposed layouts. We implemented the ILP-based decomposition in C++ and used glpk [85] to solve ILP equations. Also, the FM code provided in [86] was modified to implement SADP-FM method. The applied OPC recipes are implemented in Calibre nmOPC tool. All experiments were run on a 3.2GHz machine with 4G RAM.

Four 32nm custom industrial designs, which are fully SADP-decomposable, are selected as benchmark designs. The selected custom designs include difficult SATP components such as Π, U, and T shapes. To test the proposed methods for grid-less and non-decomposable layouts, a set of standard cells [87] and ISCAS circuits [88] are scaled down from 45nm to 32nm. Since the Nangate library is not designed for SADP, most of the cells and the ISCAS designs, which are synthesized by the Nangate library, are not fully-decomposable and SADP lithofriendly. However, they are useful to evaluate how partitioning-based method can handle partially-decomposable layouts. The design name, design area, and number of nets are shown in the first three columns, respectively. Next three columns report number of occupied fragments, number of conflicts and potential

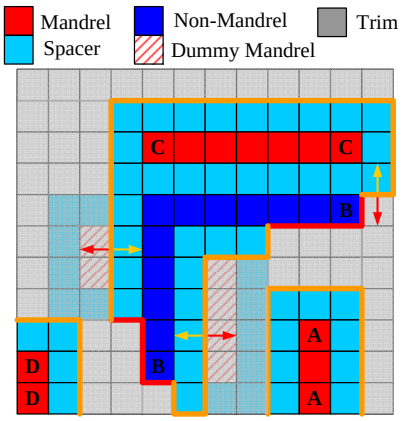


Figure 5.13: Trim mask edges are not equal in criticality.

Table 5.2: Benchmark designs.

Design	Area (um^2)	Nets	OFs	Conflicts	PSOFs
d1	0.6	12	195	216	78
d2	39	52	6581	4689	1641
d3	174	420	71264	90527	18199
d4	319.06	656	19388	16877	5436
AND4_X1	0.64	8	205	321	116
CLKBUF_X1	0.5	5	129	293	47
NAND2_X1	0.43	5	107	85	44
MUX2_X1	0.83	9	301	529	133
c1908	324.53	3524	81196	86484	32252
c7552	2421.81	26056	601809	684584	249286
c5325	1581	16977	392228	454658	164592
s1488	410.292	37434	940576	1307764	465661
s15850	5581.594	37434	940576	1307764	465661
s38417	13192.06	57006	1595774	2384630	784288

sensitive occupied fragments (PSOF), respectively. PSOFs are the fragments that may be *sensitive* if they are assigned to the trim masks, refer to Definition 5. For grid-less layouts, the reported values for the last two columns represent the total edge length and length of potential sensitive segments in terms of half-pitch, rounded if required, respectively. All the lithography simulations are done based on a composite dipole lithography system where λ=193nm, dose variation=10%, and defocus level=0.09. Moreover, we assumed overlay error is equal to 30% of half-pitch [16], i.e. 10nm.

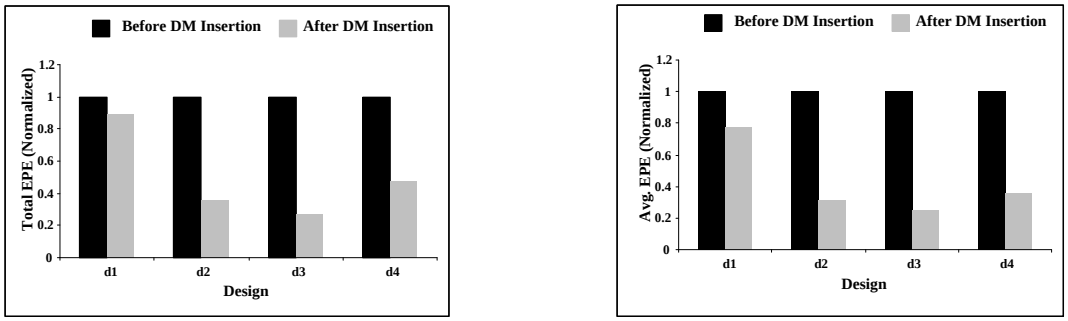
As discussed previously, assuming mandrel-assisted trim edges are printed more accurately, our proposed decomposition avoids *bare* trim edges that cannot be assisted by either original or dummy mandrel patterns. To verify the effectiveness of dummy mandrel insertion, we first decomposed test cases by an SADP-blind method, which does not consider SADP-specific litho-friendliness demands. Next, dummy mandrel patterns were inserted to assist demanding trim edges as much as possible. Finally, we compared rates of EPE for layouts without and with dummy mandrel insertion. As shown in Figure 5.14, dummy mandrel insertion can improve the printability of the layout considerably. The results are only reported for fully-decomposable layouts to make sure the occurred EPEs are due to lithography errors and not due to pitch conflicts.

The results of decomposition are reported in Table 5.3 and 5.4 based on the number of fragments which are assigned to trim mask (T-OFs) and mandrel mask (M-OFs). Columns

Table 5.3: Decomposition results of ILP-based method.

Design	SADP-blind			ILP method (Sec. 5.3)			sensitive T-OF reduction
	safe T-OFs	sensitive T-OFs	M-OFs	safe T-OFs	sensitive T-OFs	M-OFs	
d1	54	37	104	37	30	128	18.9%
d2	2485	905	3191	1343	328	4910	63.7%
d3	26822	7914	36528	28198	4447	38592	43.4%
d4	8764	3230	7394	3999	1303	14087	59.6%
AND4_X1	67	52	166	-	-	-	-
CLKBUF_X1	39	21	70	-	-	-	-
NAND2_X1	24	20	46	-	-	-	-
MUX2_X1	39	75	187	-	-	-	-
c1908	21284	14532	39784	-	-	-	-
c7552	168373	111580	320522	-	-	-	-
c5325	111811	53735	213984	-	-	-	-
s1488	19812	26070	58396	-	-	-	-
s15850	178709	357419	404448	-	-	-	-
s38417	191492	638310	765972	-	-	-	-

"safe T-OF" and "sensitive T-OFs" refer to trim fragments which can be assisted by original or dummy mandrel patterns and to those which cannot, respectively. As shown in Table 5.3 and 5.4, our method decreases the total length of sensitive trim fragments effectively. Among all test cases, all decomposition methods failed to find a legal decomposition for MUX2_X1.



(a) (b)

Figure 5.14: Dummy mandrel insertion mitigates both total and average EPEs. (a) Normalized statistics of total EPEs, (b) Normalized statistics of average EPEs.

Table 5.4: Decomposition results of partitioning-based method.

Design	SADP-blind			FM method (Sec. 5.4)			sensitive T-OF reduction
	safe T-OFs	sensitive T-OFs	M-OFs	safe T-OFs	sensitive T-OFs	M-OFs	
d1	54	37	104	37	30	128	18.9%
d2	2485	905	3191	1527	342	4712	62.2%
d3	26822	7914	36528	26647	4925	39692	37.7%
d4	8764	3230	7394	5803	1498	12087	53.6%
AND4_X1	67	52	166	85	34	166	34.6%
CLKBUF_X1	39	21	70	53	12	65	42.8%
NAND2_X1	24	20	46	21	19	68	5%
MUX2_X1	39	75	187	51	36	214	52%
c1908	21284	14532	39784	24044	11488	47556	20.9%
c7552	168373	111580	320522	199443	82990	361242	25.6%
c5325	111811	53735	213984	134076	3781	237344	26.9%
s1488	19812	26070	58396	17988	20595	65695	21%
s15850	178709	357419	404448	179462	300232	460882	16%
s38417	191492	638310	765972	264898	612778	718098	5%

To evaluate the quality of the proposed method, we compared the printability of our decomposition results with SADP-blind decompositions, which only considers the pitch conflict constraints. As a gauge of printability, we calculated edge placement error (EPE) for both solutions using Calibre LFD tool [89]. The EPE statistics, shown in Table 5.5 and Table 5.6, show our method improves both total and average EPEs compared with the blind decomposition.

Among fully-decomposable and SADP-specific test cases, i.e. designs *d1*-*d4*, the achieved improvement for design *d1* is considerably lower than others. The main reason of this phenomenon is the large number of *PSOF*s in design *d1*. As shown in Table 5.2, the number of *PSOF*s to the number of *OF*s is larger in design *d1*, i.e. 40%, compared with the other test cases', 25% on average. It means, more sensitive fragments were forced to be assigned to the trim mask to meet pitch conflicts; however, could not receive protection by mandrel patterns. Moreover, Table 5.5 and Table 5.6 show that in spite of significant EPE improvement in our proposed decomposition for design *d1*, it still suffers from large EPE relative to its fairly short wire length. Also, a large PSOF results in high image distortion in grid-less/non-decomposable layouts too. As reported in Table 5.6, although MUX2_X1 is non-decomposable and suffers minimum pitch violation, AND4_X1 experiences larger EPE since its relative number of PSOFs is higher than MUX2_X1's. Therefore, we can

Table 5.5: Edge placement error statistics in ILP-based decomposition.

Design	SADP-blind EPE (nm)			ILP-based EPE (nm)			Improvement (%)	
	Avg.	Std.	Total	Avg.	Std.	Total	Avg.	Total
d1	1.79	4.79	412.8	1.26	2.71	288.269	42	43.2
d2	0.69	1.48	3749.9	0.56	1.2	2986.9	23.2	25.5
d3	0.82	4.57	40217.6	0.7	3.88	35289.8	17.1	13.9
d4	0.59	2.43	17968.1	0.45	1.38	13245	31.1	35.6
AND4_X1	4.45	10.21	1120.1	-	-	-	-	-
CLKBUF_X1	2.69	5.51	455.7	-	-	-	-	-
NAND2_X1	6.43	9.4	758.4	-	-	-	-	-
MUX2_X1	2.84	8.37	948.5	-	-	-	-	-
c1908	4.25	2.6	461460.4	-	-	-	-	-
c7552	3.82	9.27	3262595.9	-	-	-	-	-
c5325	3.70	11.3	2098208.3	-	-	-	-	-
s1488	3.22	4.76	138275.172	-	-	-	-	-
s15850	2.74	3.5	3062186.736	-	-	-	-	-
s38417	4.1	8.4	2894022.72	-	-	-	-	-
Average improvement of decomposable designs							28.4	29.6
Total average improvement							-	-

conclude:

1. Relative number of *PSOF*s can be considered as a measure of how much a layout is SADP-friendly.

2. The SADP requirements should be taken into account during the layout design because post-layout decomposition cannot solve trim sensitivities which are imposed by an oblivious-to-SADP layout design.

In addition to total EPE, layout designers desire to minimize or preferably completely avoid critical EPEs to maximize product yield. Figure 5.15 compares the distribution of EPEs in proposed decomposition for the attempted designs versus an industrial SADP layout decomposition. As shown in this figure, the EPE domains in our solutions are smaller than the traditional solutions. Moreover, not only the number of accurate fragments, i.e. with zero EPE, is larger in our decomposition, but also the frequency of large EPEs is lower in our solution compared with the industrial decomposition, which means it suffers from lower fatal EPEs.

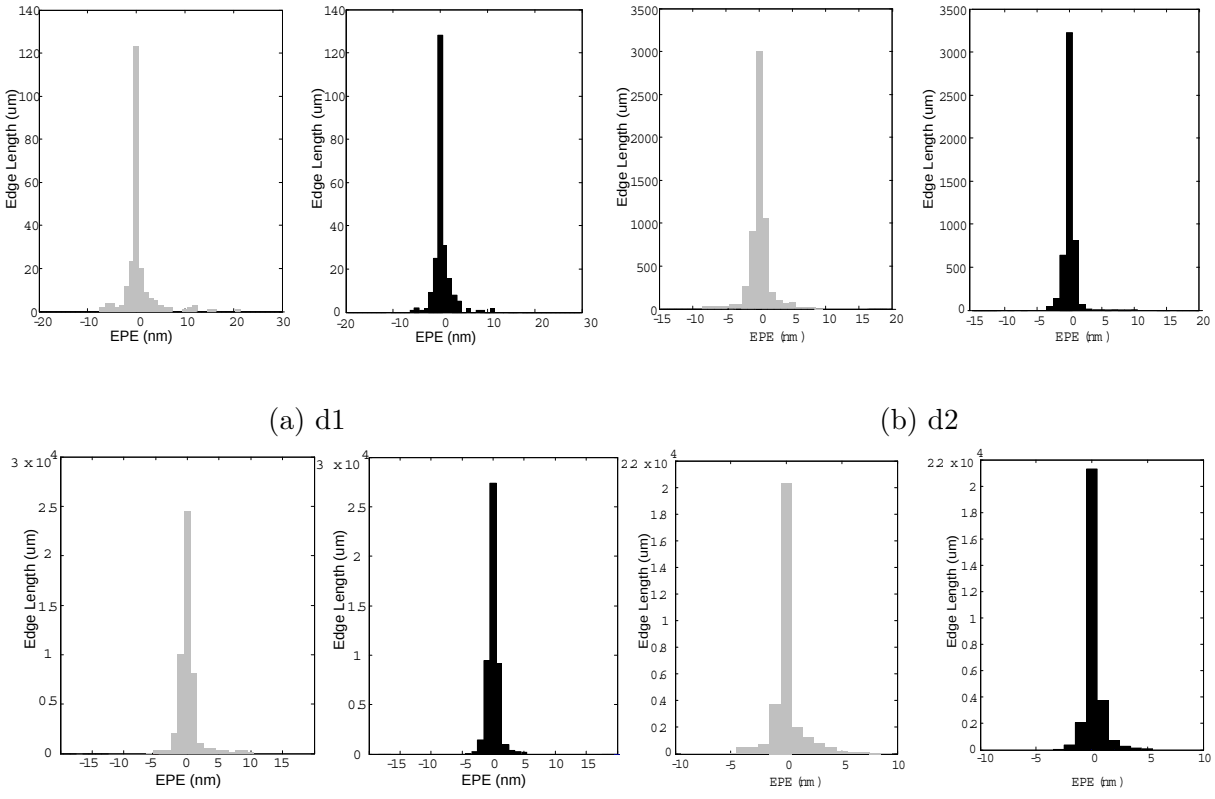
Table 5.6: Edge placement error statistics in partitioning-based decomposition.

Design	SADP-blind EPE (nm)			ILP-based EPE (nm)			Improvement (%)	
	Avg.	Std.	Total	Avg.	Std.	Total	Avg.	Total
d1	1.79	4.79	412.8	1.26	2.71	288.269	42	43.2
d2	0.69	1.48	3749.9	0.56	1.85	3083.2	23.2	21.6
d3	0.82	4.57	40217.6	0.73	4.2	35400.3	12.3	13.6
d4	0.59	2.43	17968.1	0.44	1.42	13612	34.1	32
AND4_X1	4.45	10.21	1120.1	4.19	9.73	1020	9.8	6.2
CLKBUF_X1	2.69	5.51	455.7	1.6	3.2	269.4	69	68.1
NAND2_X1	6.43	9.4	758.4	4.6	7.52	580.4	30	39.7
MUX2_X1	2.84	8.37	948.5	2.2	5.18	757.9	25	29
c1908	4.25	2.6	461460.4	2.95	2.02	373960.8	30	18
c7552	3.82	9.27	3262595.9	2.64	7.12	2295770.4	30	29
c5325	3.70	11.3	2098208.3	2.56	8.7	1408208.8	31	32.8
s1488	3.22	4.76	138275.1	2.75	2.26	113800	14	17.7
s15850	2.74	3.5	3062186.7	2.12	3.2	2223147	22	27.4
s38417	4.1	8.4	2894022.8	3.54	8	2436767	13	16
Average improvement of decomposable designs							27.9	27.6
Total average improvement							30.6	30.3

Runtime for all decomposition methods are reported in Table 5.7. The CPU time of the partitioning-based method is comparable with SADP-blind method for attempted designs. However, partitioning the layout into smaller independent blocks, e.g. the method which was proposed in [26], would be necessary.

5.7 Conclusion and future work

In this chapter, the effective parameters in the printability of decomposed layouts in the SADP method were discussed. Subsequently, a decomposition method for SADP technique was proposed which uses ILP formulation to avoid decomposition conflicts. At the same time, it improves the overall printability of the layout by providing higher edge protection for trim patterns. This edge protection can be performed by either original or dummy mandrel patterns. Finally, seeking to achieve a set of objectives same as the proposed ILP-based method, a partitioning-based method was proposed to improve the time-complexity of SADP decomposition.



(a) d1 (b) d2

(c) d3 (d) d4

Figure 5.15: SADP-aware decomposition improves the worst-case EPE as well as EPE distribution spread. EPE distributions of the decomposed layout by SADP-blind method and the proposed method are colored in gray and black, respectively.

The experimental results show that total length of sensitive trim edges, total EPE and overall printability of attempted designs are improved considerably in the proposed decomposition method. Based on experiments, ignoring the manufacturability demands of SADP during the layout design can result in poor image quality. Therefore, development of a SADP-aware correct-by-construction decomposition method would be considered as the future step.

Table 5.7: Runtime for SADP decomposition methods (sec).

Design	SADP-blind	ILP-based (Sec. 5.3)	FM-based (Sec. 5.4)
d1	0	0.84	0.017
d2	0	11.52	2.306
d3	1	64.8	6.24
d4	0.4	39.78	3.453
AND4_X1	0	-	0.16
CLKBUF_X1	0	-	0
NAND2_X1	0	-	0
MUX2_X1	0.1	-	0.12
c1908	0.7	-	6.21
c7552	1.2	-	9.37
c5325	0.86	-	8.29
s1488	0.9	-	5.2
s15850	1.4	-	9.85
s38417	1.5	-	11

Chapter 6

SADP-friendly detailed routing

6.1 Introduction

As discussed in Chapter 5, self-aligned double patterning (SADP) has attracted much interest among different DPL methods [24]. The basic SADP method [24] is similar to LELE in patterning steps till the end of the first litho phase. Next, a set of overlay-resistant non-litho processes is used to form the required spaces around the patterns of the first exposure and consequently the second group of patterns. Although the basic SADP is fairly robust against overlay error, it has being applied only for printing regular patterns, e.g. FLASH memories. In order to benefit from the advantages of the SADP method, an additional trim lithography, which facilitates printing two-dimensional and irregular structures, is appended to the basic SADP process. Despite the potential robustness of two-dimensional SADP method to the overlay error, its achievable pitch resolution depends strongly on the litho-friendliness of the trim mask [41], which itself depends on the applied policies for layout design and decomposition. Therefore, similar to other DPL methods, layout decomposition is the primary concern in the SADP method.

Pitch conflict resolution is the primary objective of layout decomposition in all DPL methods. However, this objective varies for different DPL methods with respect to the involved process in each method. Recently, several LELE-specific decomposition methods have been proposed. In general, there are two main approaches to achieve double patterning compliant layouts. The first one consists of post-layout methods which consider the DPL-specific manufacturability requirements, i.e. the minimum number of cuts and the maximum overlay robustness, during the decomposition process of a given layout [15, 10, 43, 45]. The second one translates the DPL compliance into a set of cost metrics,

which are considered by the detailed router to find the optimal routing solution in terms of DPL constraints and traditional design objectives such as wire length.

None of these proposed methods considers SADP-compliance requirements; i.e.

1. No pattern cuts are permitted because they cause non-resolving gaps in the final layout.

2. Trim mask should be litho-friendly to achieve appropriate litho quality.

In Chapter 5, the effective parameters in the printability of SADP decomposed layouts were discussed. In summary, it was observed that the trim patterns should be protected by either original or dummy mandrel pattern to improve the litho-friendless and overlay robustness. The obtained experimental results proved that such SADP decomposition method can achieve higher imaging quality for SADP-compliant layouts. However, it was shown in experiments that post-layout SADP decomposition can decompose a minor subset of layouts which were not designed in an SADP-aware style. Moreover, meeting secondary litho-friendliness objectives by post-layout decomposition methods, e.g. balanced mask density, is subject to input layout specifications. Therefore, to benefit from SADP advantages, one needs SADP-aware correct-by-construction methods to generate SADP-compliant layouts.

In this chapter, a SADP-aware detailed routing (SADP-DR) method is proposed which performs detailed routing and layout decomposition simultaneously to avoid litho-limited layout configurations. In addition, the proposed router preserves the uniformity of pattern density between mandrel and trim masks.

The rest of the chapter is organized as follows. A brief review of general-purpose SADP process and its printability challenges is given in Section 6.2. Section 6.3 discusses the proposed SADP-aware detailed routing method. Experimental results are reported in Section 6.4 and Section 6.5 concludes the chapter.

6.2 The printability challenges of SADP method

Figure 6.1 exemplifies how the general-purpose SADP process prints a two-dimensional pattern. In this process, a subset of layout features in conjunction with some dummy patterns are printed by the first lithography phase, so-called the mandrel lithography. Afterwards, the critical trenches are realized by sacrificial spacers deposited around the

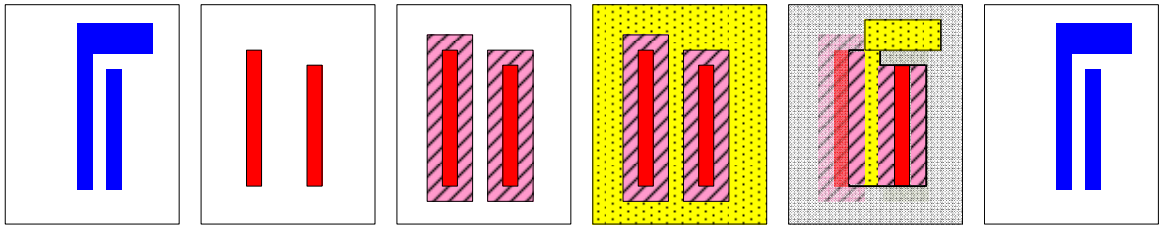


Figure 6.1: General-purpose SADP process: (a) Target image (b) Mandrel lithography (c) Spacer deposition and etching (d) Dielectric filling and CMP (e) Trim lithography (f) Final etching and metal filling.

mandrel patterns. The remainder of layout features are printed by a sequence of di-electric deposition, surface polishing and a litho-based trimming.

Layout decomposability is the primary concern in SADP, where pattern splitting results in unresolvable gaps and thus is prohibited. Therefore, considerable parts of any SADP-oblivious layout are not manufacturable by SADP because of decomposition conflicts. Moreover, post-layout SADP problem has been proven as an NP-complete problem [48].

The second challenge in SADP is the accuracy of trim lithography step. In fact, the trim mask may contain isolated lines and semi-isolated trenches simultaneously which makes the OPC and dose tuning difficult. Moreover, the trim mask experiences considerable defocus error because of surfaces roughness generated during mandrel lithography and spacer deposition. According to simulation results [41], the trim mask is sensitive to defocus errors greater then 40nm. To mitigate the consequences of DoF variation, the uniformity of pattern density on the trim mask is desirable, which strongly depends on the the original layout specifications and the quality of the decomposition method.

Based on the in-silico experiments presented in Figure 5.6, following points are derived about interactions of mandrel and trim masks:

- The most accurate edges are the trim edges which are by the sidewalls of spacer patterns.
- Semi-isolated trim edges are the most undesirable ones because they cannot be assisted by spacers. Also, complicated and irregular shapes are required on the trim mask to print semi-isolated features which increase sensitivity of the trim process to defocus errors.

- Isolated trim patterns suffer from lower level of pattern distortion compared to semi-isolated ones. This phenomenon occurs due to the smoothness of the mandrel layer in sparse regions which reduces defocus errors.

In summary, to improve the imaging quality of trim patterns, they should either be protected by a mandrel pattern or be moved to sparse regions of layout. Each of these solution has its own advantages and disadvantages:

- To improve the printability of sensitive trim patterns, dummy mandrel patterns can be inserted around them. As long as the minimum pitch and OPC requirements are met within the mandrel mask, these dummy patterns do not impact on the functional and electrical characteristics of the design because they will be cleaned by the trim mask. However, minimum pitch and OPC constraints impose strict limitations on dummy pattern insertion. An example for issues of dummy pattern insertion is shown in Figure 6.2.(a), where line-end-shortening and corner rounding errors are not included. In Figure 6.2.(a), the green edges show those trim pattern edges which are assisted by either original mandrel patterns, i.e. solid red grids, or dummy mandrel patterns, i.e. dotted red girds. Also, dummy mandrel insertion is not possible to protect the red trim edges. The domain of eligible grids for dummy pattern insertion will be confined further if we consider the OPC requirements of mandrel patterns.

- Another solution, which increases the total length of protected edges, is ripping up the mandrel patterns and re-routing them next to the sensitive trim patterns. Figure 6.2.(b) shows how rip up and re-route (RAR) achieves higher level of trim protection in expense of longer wire length (WL).

- Moreover, post-decomposition rip up and re-route can be used to move the sensitive trim patterns into sparse areas. For highly congested designs, this solution results in high WL penalty due to the shortage of local sparse areas. Moreover, it is highly probable that the new route generates other sensitive patterns within the trim mask.

Therefore, all of these post-decomposition solutions should be applied carefully due to the discussed trade-offs. Moreover, the effectiveness of each method depends highly on the characteristics of the original layout. Furthermore, no post-decomposition method is useful to solve the decomposition conflicts of the original layout. All of the mentioned problems can be considered during detailed routing with considerable flexibility to find a reasonable trade-off between SADP-specific and conventional design objectives, such as wire length and routability.

6.3 SADP-aware detailed routing method

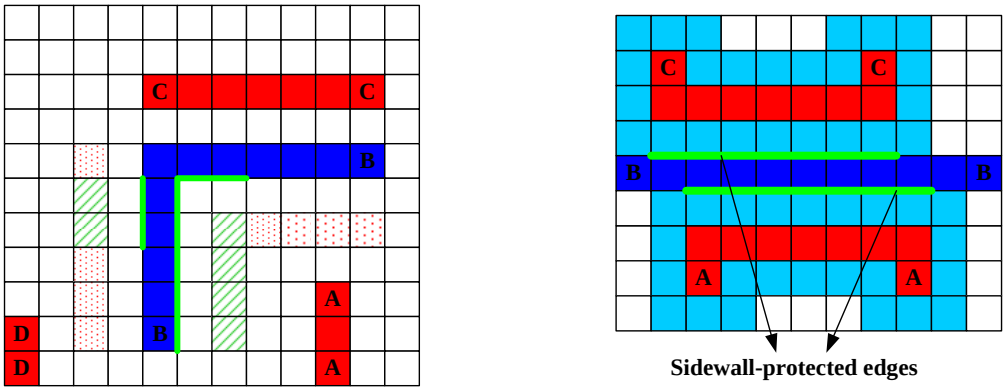
In this section, we propose SADP-aware detailed routing method.

6.3.1 Definition of mask cost parameters

In this section, mask cost parameters are defined to evaluate the litho-friendliness of mandrel and trim masks quantitatively. First, we label each routing grid by three binary variables as:

- **Routing variable (*R*)**: The variable R is set to 1 for already occupied and routing blockage grids.
- **mandrel-blocked variable (*M*)**: This variable is set to 1 if the grid cannot be assigned to the mandrel mask due to existing pitch conflicts.
- **Trim-blocked variable (*T*)**: This variable is set to 1 if the grid is conflicting with an existing trim pattern.

Figure 6.3 shows a grid labeling sample where each label represents R, M, and T variables, respectively. Such grid labeling helps to identify the sensitive and sidewall-assisted edges on the trim mask.



(a) (b)

Figure 6.2: Printability improvement of trim patterns (blue) using mandrel patterns (red), (a) Dummy mandrel patterns should not violate design constraints, (b) Stronger protection is achievable in expense of wire length overhead.

000	000	000	000	000	000	000	000	000
001	001	001	001	001	001	001	001	000
001	111	111	111	111	111	111	001	000
011	111	110	110	110	110	111	011	000
011	111	111	111	111	111	111	011	000
011	111	101	101	101	101	111	011	000
010	111	111	111	111	111	111	010	000
010	010	010	010	010	010	010	010	000
000	000	000	000	000	000	000	000	000

Figure 6.3: RMT labels for detailed routing grid: mandrel pattern (red), trim pattern (blue).

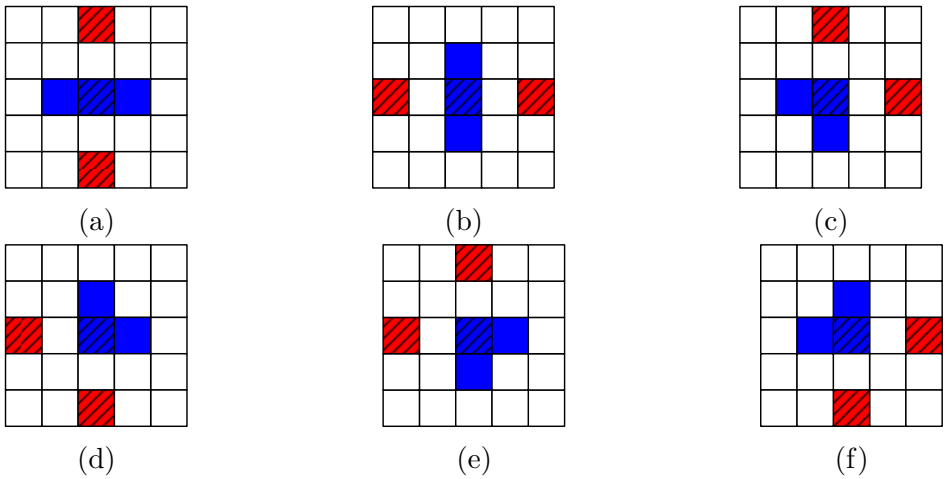


Figure 6.4: Protective grids for different routing directions: trim pattern (blue), protective grids (red).

Definition 10 ***Protective grid****: for any grid* (i, j) *along a trim pattern, its* ***protective grids*** *refer to those ones which can provide assist spacers if they are filled by a mandrel pattern.*

For a trim grid, the locations of its protective grids depend on its routing direction. For example, for a horizontal grid (i, j), its protective grids are grids $(i, j-2)$ and $(i, j+2)$.

Algorithm 1: Trim hesitation calculation

Input: Grid g_{ij} and its routing direction

Protective grids = $\{P_g, P'_g\}$
if $R_g M_g T_g == 1$ **then**
  $TH_g = \infty$
  return TH_g
end if
if $T_g == 1$ **then**
  $TH_g = PVC$
  return TH_g
end if
for each grid $x \in \{P_g, P'_g\}$ **do**
  if $M_x == 1$ **then**
    $TH_g += \alpha$
  end if
end for
$TH_g += 2\beta$
for each grid $x \in \{P_g, P'_g\}$ **do**
  if $(1 - M_x)T_x == 1$ **then**
    $TH_g -= \beta$
  end if
end for
return TH_g

Figure 6.4 shows the protective grids, shown as dashed red ones, for horizontal, vertical and corner trim grids, which are shown as dashed blue grids.

Definition 11 ***Bare grid****: an occupied trim grid is bare if its protective grids are labeled as $X1X$, where X denotes don't care; i.e. no mandrel pattern is permitted.*

Considering above definitions, a free grid hesitates over being assigned to trim (mandrel) mask if it becomes (generates) a bare grid. Consequently, the lithography cost of a candidate path on either masks can be evaluated as a function of the tendencies of its grids to the mask.

Knowing the routing direction, Algorithm 1 calculates the trim hesitation (TH) parameter for a single grid g_{ij}. First, the protective grids are determined in a similar way as shown

Algorithm 2: Mandrel hesitation calculation

Input: Grid g_{ij}, and a path p

1: **if** $R_g M_g T_g == 1$ **then**
2: **return** ∞
3: **end if**
4: **if** $M_g == 1$ **then**
5: **return** PVC
6: **end if**
7: $BB = \{x_{kl} \,|\, i-4 \leq k \leq i+4\,, j-4 \leq l \leq j+4\}$
8: **for** $x_{kl} \in BB$ **do**
9: **if** $R_x M_x (1-T_x) == 1$ **then**
10: $Y = \{BB \cap \{P_x, P'_x\}\}$
11: **if** $g \in Y$ **then**
12: **return** 0
13: **end if**
14: **for** each grid $y \in Y$ **do**
15: **if** $M_y == 0$ and Distance(g,y) $\leq$ 3hp **then**
16: **if** $y \notin p$ **then**
17: $MH_g += \alpha$
18: **else**
19: $MH_g -= \alpha$
20: **end if**
21: **end if**
22: **end for**
23: **end if**
24: **end for**
25: **return** MH_g

in Figure 6.4. The conditions in lines 2 and 6 represent blockage and free trim-blocked grids, respectively. A pitch violation cost (PVC) is assigned to the grid if it conflicts with another trim grid. Since pattern split is not permitted in SADP, PVC parameter should be selected large enough to dominate other types of cost along a path. In lines 10 - 14, the trim hesitation increases for a bare grid, which mandrel patterns are not allowed in its protective grids. Finally, lines 15 - 20 encourage the grids which are already protected by existing mandrel patterns to be assigned to the trim mask.

Table 6.1: Notation for A^*-based detailed routing.

E	Already evaluated set of grids.
Q	A priority queue of tentative girds to be evaluated. The lowest cost grid is the first element of the queue.
$x.cost$	Estimated total cost from source to terminal through grid x.
$x.path$	Tentative optimal path from grid x to source
$dir(x \rightarrow y)$	Routing direction of the grid x when the grid y is its next grid.
$dummy\_mandrel(p)$	Length of required dummy-mandrel patterns to protect a path p on the trim mask.

Similar to trim hesitation parameter, we define a mandrel hesitation (MH) parameter for each grid when it is added to a particular path, showing the number of bare trim grids which will be generated by a mandrel pattern in this grid. The calculation of mandrel hesitation for the gird g_{ij} is shown in Algorithm 2. As stated in line 7, a grid g_{ij} can only interfere in the trim grids which are within its bounding box of length $4 * half\text{-}pitch$. Therefore, the set Y in line 10 includes x_{kl}'s protective grids which are inside the bounding box interfered by the grid g_{ij}, BB. Being included in the set Y, the grid g_{ij} will protect an existing trim grid and is likely to be assigned to the mandrel mask, as stated in line 12. In line 17, we penalize the grid g_{ij} for those interfered grids which are not mandrel-blocked and also are not on the input path. The condition in line 18 is realized whenever grids y and g_{ij} interfere each other. This interference is not effective if both grids belong to the same path; however, the grid y has been already charged for interfering the grid g_{ij}. In line 19, since it lowers the MH parameter for another mandrel grid, we encourage the grid g_{ij} to be assigned to the mandrel mask, and also reimburse path p for the previously applied pessimistic charge.

6.3.2 Detailed routing algorithm

Based on the defined trim and mandrel hesitation parameters, our SADP-aware detailed router is introduced in this section. Table 6.1 defines notation which is used by the proposed algorithms.

As shown in Algorithm 3, the A^*-based detailed routing method was modified to find the optimal paths on trim and mandrel masks for a given net. In line 11, the cost of

unit wire length is considered as 1. Moreover, the A^*cost of each grid is defined as its Manhattan distance to the net's target, t. With respect to the target mask, either trim hesitation or mandrel hesitation of each step is included in the cost of the path in lines 13 and 15, respectively. Since the TH parameter of a grid depends on its routing direction, trim hesitation of the last visited grid is added to the path cost in line 13 to decrease the possibility of having a bare trim grid. For cost on the mandrel mask, the MH parameter of the current grid is added to the cost of the path, as shown in line 15.

Algorithm 3: Path-Finder

Input: A net n and the mask={M,T}

1: {s,t} = source and target grid of n
2: $E = \emptyset$; $Q = \{s\}$
3: $s.cost = A^*cost(s)$
4: **while** $!Q.empty()$ **do**
5: $x = Q.dequeue()$
6: **if** $x == t$ **then**
7: **return** $\{x \cup x.prev.path\}$
8: **end if**
9: add x to E
10: **for** $y \in \{neighbors(x) - E\}$ **do**
11: $tentative\_cost = x.cost + 1 + A^*cost(y)$
12: **if** $mask == T$ **then**
13: $tentative\_cost+ = TH(x, dir(x \rightarrow y))$
14: **else**
15: $tentative\_cost+ = MH(y, x.path)$
16: **end if**
17: **if** $y \notin Q$ **then**
18: enqueue y to Q
19: **end if**
20: **if** $y.cost > tentative\_cost$ **then**
21: $y.cost = tentative\_cost$
22: $y.prev = x$
23: $y.path = \{x.path \cup x\}$
24: **end if**
25: **end for**
26: **end while**

Algorithm 4: SADP-aware detailed routing

Input: A set of blockages B and a set of nets N

1: Decompose blockages and label girds
2: **for** $n \in N$ **do**
3: tp = Path-Finder(n,T)
4: mp = path_finder(n,M)
5: $Cost_{trim} = A * tp.cost + B * WL_{tot}|0.5 - \frac{WL_{trim}+tp.length}{WL_{tot}+dummy\_mandrel(tp)}|$
6: $Cost_{mandrel} = A * mp.cost + B * WL_{tot}|0.5 - \frac{WL_{mandrel}+mp.length}{WL_{tot}}|$
7: select the minimum cost path for n
8: update grid labels
9: update WL_{tot}, WL_{trim}, and $WL_{mandrel}$ parameters
10: **end for**

The SADP-aware detailed routing method is presented in Algorithm 4. First, we find optimal paths on both mandrel and trim masks. Next, in lines 5-7, minimum cost path is selected with respect to wire length, SADP-friendliness, and mask density factors, where A and B are user-defined parameters to distinguish between different objectives. In line 5, the second term evaluates the impact of the candidate trim path on the balance of pattern densities on mandrel and trim masks. In addition to the total wire length on the trim mask, a trim route affects the pattern density on the mandrel mask due to its required dummy mandrel patterns. In line 5, function $dummy\_mandrel(tp)$ returns the length of dummy mandrel patterns required to assist those parts of route tp which are not already protected by original mandrel patterns. It is noteworthy that these potential dummy mandrel grids should be included in neither total wire length (WL_{total}) nor mandrel mask wire length ($WL_{mandrel}$) because these may be filled by original mandrel patterns later. Finally, the grid labels and other routing parameters should be updated with respect to the selected path.

6.4 Experimental results

We implemented our SADP-aware detailed router in C++ and tested on a 3.2GHz machine with 4G RAM. Three standard cell designs, which were scaled down from 180nm to 32nm, were used as test cases. The placement legality of test cases should be preserved after the scaling. Therefore, the benchmark set is limited to what our industrial partner provided. For comparison, we used a A^*-based router followed by SADP decomposition [50].

Table 6.2: Performance comparison between SADP-blind approach and SADP-DR.

Design	Area (μm^2)	Nets	Router	Wirelength (nm)	SADP statistics (nm)				Runtime ($sec.$)	Ratio	
					M	PT	BT	Failure		BT	time
s1488	410.292	672	blind-DR	8352	5344	984	1448	576	30.2	0.33	1.59
			SADP-DR	8416	3584	4320	512	96	48.3		
s15850	5581.594	10891	blind-DR	43713	21729	5632	10208	61444	1523.88	0.3	1.65
			SADP-DR	44512	20416	20544	3552	1024	2517.9		
s38417	13192.06	24877	blind-DR	64458	34199	3309	17653	11297	3328.4	0.19	1.4
			SADP-DR	113236	56000	50780	6456	9396	4715		

Table 6.2 shows the results for two approaches. The experiments are performed by the sensitive trim cost of $\alpha = 4$, and the redundant dummy mandrel insertion cost of $\beta = 1$. Since each dummmy mandrel grid can protect up to four sensitive trim girds, parameters α and β are selected such that the trim hesitation parameter dominates the cost of excessive dummy mandrel insertion. In this table, first three columns report test cases' characteristics. Column 5 shows total wire length for each design and Columns 6-11 report the length of mandrel (M), protected trim (PT), bare trim (BT) and non-decomposable patterns (Failure) respectively.

As demonstrated by results, both non-decomposable and bare-trim edges are improved considerably by the proposed SADP-aware detailed routing methodology. The wire length overhead is tolerable for the first two test cases. However, SADP-aware method increased the wire length by 75% for the third test case. Since the same core detailed routing method is applied in both methods, this wire length overhead shows that SADP-friendliness of a design depends on placement specifications too. Therefore, expanding the SADP-friendly design to higher levels of design cycle such as global and detailed placement can be considered as a future work.

6.5 Conclusion

In this chapter, the printability issues in 2D self-aligned double patterning lithography were discussed. Subsequently, to address the litho-friendliness concerns for SADP, a SADP-aware detailed routing method was proposed. In addition to observe decomposability constraints, the proposed method provides protective spacer sidewalls for trim pattern, which are highly sensitive to lithography imperfections such as defocus and overlay errors.

Chapter 7

Litho-friendly SATP decomposition method

7.1 Introduction

In the absence of mature next generation lithography methods, i.e. EUV and e-beam lithography, multiple patterning attracts significant attention as the most promising solutions to print sub-32nm half-pitch process nodes. Double patterning lithography (DPL) combined with 193nm immersion lithography can drive the half-pitch down to 22nm. To meet the scaling time-line, more aggressive multiple patterning processes such as triple patterning lithography (TPL) are required [16, 90].

Triple patterning lithography can be realized by different methods including split methods [91] and self-aligned methods [92, 93]. In the split triple patterning lithography (STPL), sub-resolution patterns of the original layout are decomposed into three masks, each of which will be printed by an independent lithography process. This approach can increase the pitch frequency by three times. However, using three lithography masks limits the overlay budget, which makes the overlay budget a very critical issue. Moreover, in spite of reducing overall pitch, STPL methods cannot print sub-resolution lines and trenches simultaneously [94].

In self-aligned multiple patterning methods, some of the litho-based imaging steps are replaced by chemical deposition processes, which are accurately controllable. Compared to the different split patterning methods, the key benefit of self-aligned patterning is its inherent robustness against the overlay error. In self-aligned patterning, the second lithography

phase is replaced by a chemical process, therfore, the whole process is not affected by sub-masks overlay as it is in split patterning methods. Self-aligned double patterning (SADP) was originally proposed for dense and highly regular layout patterns such as contact layers [24]. In this method, a sacrificial spacer layer was deposited around core patterns, also referred as mandrel patterns, to print critical spaces and to double the line-space frequency. By appending a litho-based trimming step to the basic process, SADP became applicable for two-dimensional layouts as well, such as NAND flash memories [41].

The major challenging issue of SADP is the complexity of layout decomposition. This complexity increases in two-dimensional SADP, where the overlay between mandrel and trim masks should be taken into account [41]. The decomposition method in [95] modeled SADP decomposition as a shortest path problem. [50, 96] proposed decomposition methods to maximize the sidewall-assisted patterns which are more robust against the overlay error. In a correct-by-construction approach, SADP decomposability and overlay robustness requirements were addressed during detailed routing by [97].

Self-aligned triple patterning (SATP) is proposed to improve the printing resolution down to 15nm [92]. In addition to sacrificial spacers, another spacer layer, so-called structural spacer, is applied in SATP to print sub-resolution layout patterns. Consequently, compared to SADP, SATP relaxes design rules and allows critical lines and trenches simultaneously.

In Chapter 5, we proposed an ILP formulation for SADP decomposition, where the constraints depend on pattern formation flow. Because SATP process is different than SADP, we need to revisit the formulation proposed in Chapter 5 to be applicable for SATP decomposition. Moreover, as discussed later in this chapter, sources of pattern distortion in STAP are different than SADP; therefore, we need to modify the objective to consider SATP characteristics in decomposition.

In this chapter, SATP decomposition is formulated by integer linear programming. In addition to conventional minimum width and spacing constraints, the proposed decomposition method tries to minimize the mandrel-trim co-defined edges and to maximize the layout features printed by structural spacers. Therefore, the overlay sensitivity, due to misalignment between mandrel and trim masks, and line-edge roughness issues, due to mandrel lithography imperfections , are mitigated by the proposed method. The rest of the chapter is organized as follows: the general-purpose SATP process and its EDA requirements are described in Section 7.2. Section 7.3 introduces the proposed decomposition method. Experimental results are reported in Section 7.4 and finally, Section 7.5 concludes the work.

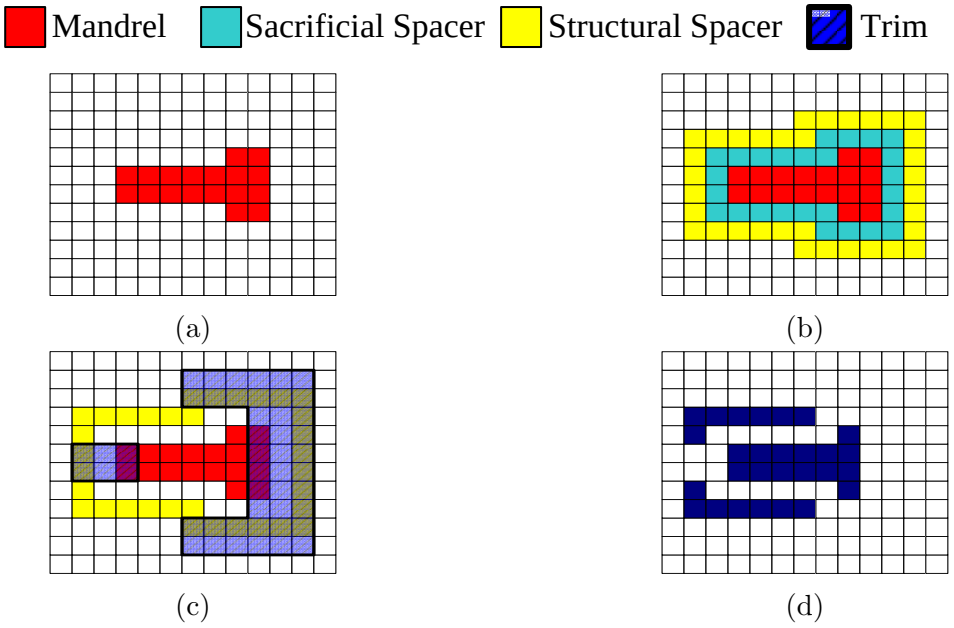


Figure 7.1: Self-aligned triple patterning process (a) Mandrel lithography, (b) Sacrificial and structural spacers deposition, (c) Etch sacrificial spacers and trim lithography, (d) Final image.

7.2 SATP process and requirements

Figure 7.1 demonstrates the process of self-aligned triple patterning. In SATP process, the mandrel mask includes a subset of original layout features and possibly a set of dummy patterns. To fabricate SATP, the mandrel mask is imaged initially. Next, sacrificial spacer (S1) and structural spacer (S2) layers are deposited in sequence to print the critical trenches and the rest of layout features, respectively. Later, sacrificial spacers are etched which leaves mandrel lines and structural spacers unaffected, and results in spatial frequency tripling. Finally, dummy mandrel patterns and excess parts of structural spacers are cleared by a trim mask.

In addition to extending the lithography resolution beyond SADP, SATP lithography relaxes some of the challenging constraints in SADP. First, SATP process allows simultaneous fabrication of both critical lines (by structural spacers) and trenches (by sacrificial spacers). Second, in contrary to SADP, the SATP method is an inherently two-dimensional method because of the structural spacers. Consequently, the application of trim mask, which is the main source of overlay sensitivity in self-aligned patterning process, is much

more limited in SATP as compared to SADP.

However, the applied decomposition method can highly impact on overlay sensitivity and line edge roughness, i.e. two critical issues in multiple patterning lithography methods. In all self-aligned multiple patterning methods, the layout patterns which are co-defined by mandrel and trim masks are the major sources of overlay sensitivity and thus should be avoided. In contrary to SADP decomposition, where trimming is inevitable to print two-dimensional features, a litho-friendly SATP decomposition can effectively increase the overlay robustness by wisely using structural spacers.

Based on SEM images of process development results [92], the CD of layout features printed by the mandrel mask is larger than the ones printed by spacers. Moreover, spacer-printed features experience 40% lower line edge roughness compared to mandrel patterns. Therefore, a litho-friendly SATP decomposition should maximize the critical patterns which are printed by structural spacers.

7.3 ILP-based formulation for litho-friendly SATP

In this section, we present different objectives and constraints of litho-friendly SATP decomposition based on integer linear programming (ILP). Our notation is described in Table 7.1.

Table 7.1: Notation for ILP-based formulation.

F_i	$F_i = 1$ if the i-th tile is occupied by a layout feature, otherwise $F_i = 0$
M_i	$M_i = 1$ if the i-th tile is assigned to the mandrel mask, otherwise $M_i = 0$
$S1_i$	$S1_i = 1$ if the i-th tile is filled by sacrificial spacer, otherwise $S1_i = 0$
$S2_i$	$S2_i = 1$ if the i-th tile is filled by structural spacer, otherwise $S2_i = 0$
T_i	$T_i = 1$ if the i-th tile should be trimmed by the trim mask, otherwise $T_i = 0$

7.3.1 Objectives

Because of the accuracy of deposition process and limited quality of mandrel lithography, it is desired to maximize the use of structural spacers. In (7.1), variable M_i^{orig} accounts for the feature tiles that are printed by the mandrel mask.

$$F_i + M_i - M_i^{orig} \leq 1. \tag{7.1}$$

In addition to the general reluctance to pattern by the mandrel mask, it is preferred to avoid the co-defined mandrel/trim edges to minimize the sensitivity to overlay error. Overlay sensitivity happens where the trim mask touches a feature tile. The corresponding ILP formulations is:

$$F_i + T_j - overlay_{ij} \leq 1, \tag{7.2}$$

where sub-script j refers to all the surrounding tiles of the tile i. To capture the overlay sensitivities fairly, the input layout should be mapped into a grid where the size of each tile is overlay budget.

All together, the objective of litho-friendly SATP can be written as:

$$Minimize \quad \alpha \sum_{i=1}^{N} M_i^{orig} + \beta \sum_{i=1}^{N} overlay_{ij}, \tag{7.3}$$

where N represents the number of tiles and $j \in surroundings_i$. The weighting parameters α and β are applied to trade-off between the CDU[1] error of mandrel lithography and the overlay error within mandrel and trim masks. In this work, we considered equivalent weights for the costs of mandrel patterns and overlay error. However, different α and β parameters would be applied based on the characteristics of the target fabrication facilities. For example, larger β values should be applied for a facility with poor overlay control.

7.3.2 Pattern formation constraints

The first set of pattern formation constraints deals with covering the layout features. In a valid decomposition, all the feature tiles should be covered by either mandrel patterns or structural spacers. The corresponding ILP constraints are:

$$\begin{aligned} &F_i - M_i - S2_i \leq 0, \\ &M_i + S2_i \leq 1, \end{aligned} \tag{7.4}$$

where the first constraint guarantees that a feature tile is printed by one of the patterning processes of SATP. The second constraint guarantees a tile will be assigned to only one of the patterning steps.

Since no pattern split is permitted in self-aligned patterning methods, all the tiles belonging to the same net should be printed by the same patterning process. Therefore,

[1]Critical dimension uniformity

we have:

$$\begin{cases} M_i - M_j = 0 \\ S2_i - S2_j = 0 \end{cases} \tag{7.5}$$

where tiles i and j belong to the same net. To avoid redundant constraints, one can replace M and $S2$ variables for all sibling tiles with two representative variables and eliminate the above constraints.

Spacer deposition processes can be formulated as (7.6a)-(7.6e). (7.6a) states a tile can be occupied by a mandrel pattern, or sacrificial spacer or structural spacer. (7.6b) and (7.6c) specify that sacrificial spacer is deposited in a tile if and only if one of its adjacent tiles are filled by mandrel patterns. Similarly, as shown by (7.6d) and (7.6e), structural spacer forms around the tiles filled by sacrificial spacer.

$$M_i + S1_i + S2_i \leq 1 \tag{7.6a}$$

for all tiles j that touch tile i:

$$M_j - S1_i \leq 0 \tag{7.6b}$$

$$S1_i - \sum_j M_j \leq 0 \tag{7.6c}$$

$$S1_j - S2_i \leq 0 \tag{7.6d}$$

$$S2_i - \sum_j S1_j \leq 0 \tag{7.6e}$$

After patterning all layout features, as the last step of SATP, trim mask should clean dummy mandrel patterns and excess parts of structural spacers. To preserve the printed layout features, trim mask should not interact with feature tiles as formulated by (7.7a). Equations (7.7b) and (7.7c) state that trim mask should cover those mandrel patterns and structural spacers which are not parts of layout features.

$$F_i - T_i \leq 1 \tag{7.7a}$$

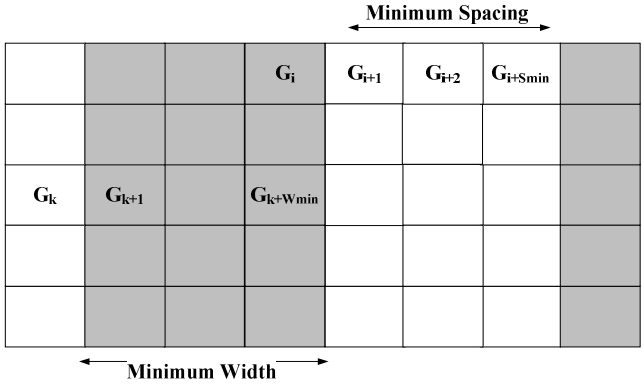


Figure 7.2: Minimum width and spacing constraints for mandrel and trim masks.

$$M_i - F_i - T_i \leq 0 \tag{7.7b}$$

$$S2_i - F_i - T_i \leq 0 \tag{7.7c}$$

7.3.3 Constraints for lithography masks

Since the width of structural spacer is constant, the layout features with the width greater than $2 * W_{S2}$ cannot be printed by structural spacers. Therefore, such patterns must be assigned to the mandrel, and we have:

$$M_{i+1} - F_i + F_{i+1} - F_{i+2} - F_{i+3} \leq 0. \tag{7.8}$$

Moreover, minimum width and spacing design rules should be met in mandrel and trim masks. As shown in Figure 7.2, the minimum width rule within a litho mask implies that if the grid i is clear and the grid $i+1$ is dark, then all the adjacent grids within the minimum width distance; i.e. the grids $\{i+2...i+Wmin\}$, should be dark. Therefore, the corresponding formula are:

$$\begin{aligned} &G_{k+1} - G_{k+2} - G_k \leq 0, \\ &\vdots \\ &G_{k+1} - G_{k+W_{min}} - G_k \leq 0, \end{aligned} \tag{7.9}$$

where G represents the grids of either mandrel or trim masks, $G \in \{M, T\}$.

Table 7.2: Benchmark designs.

Design	Area (um^2)	Nets	WL (um)	CL (um)	Var.	Const.
AND4_X1	0.34	8	2.23	4.8	371	2036
CLKBUF_X1	0.27	5	1.4	4.38	349	1851
NAND2_X1	0.23	5	1.16	1.27	274	913
MUX2_X1	0.44	9	3.27	7.91	431	2961
d1	0.28	12	2.12	3.23	292	1430
d2	18.3	52	71.6	1354.2	22103	41654
d3	81.8	420	775.35	40.14	57019	98960
d4	149.96	656	342.78	252.5	203833	496207

Also, there should be a minimum spacing distance between the external edges of two patterns. In other words, as shown in Figure 7.2, if grids i and $i+1$ are dark and clear respectively, then grids $\{i+2, ..., i+S_{min}\}$ should be clear as:

$$\begin{aligned} &G_i + G_{i+2} - G_{i+1} \leq 1, \\ &\vdots \\ &G_i + G_{i+Smin} - G_{i+1} \leq 1, \end{aligned} \tag{7.10}$$

where G represents the grids of either mandrel or trim masks, $G \in \{M, T\}$.

7.4 Experimental results

We implemented our algorithm in C++ and used glpk [85] to solve integer linear programming equations. All experiments were run on a 3.2 GHz machine with 4G RAM. The complexity of SATP ILP-based decomposition depends strictly on the characteristics of the input layout and fabrication process rules. The number of ILP variables is a function of the layout area, minimum feature size, and overlay budget. The number of constraints depends on number of layout features, total wire length, and minimum pitch conflict length.

The proposed decomposition method was tested on Nangate open cell library [87] scaled down from 45nm to 22nm. We also added a set of custom layouts in our benchmark set, which are larger than the standard cells and include difficult SATP components such as Π, U, and T shapes. It is noteworthy that only a few of the attempted standard cells were decomposable for SATP process which implies the criticality of correct-by-construction layout design for advanced lithography processes. The detailed information of benchmark

designs and specifications of their corresponding ILP formulations are shown in Table 7.2, where the overlay budget is *6nm* [16]. The first five columns show design name, layout area, number of layout patterns, total wire length and total conflict length of test cases and the last two columns represent number of ILP variables and constraints.

To evaluate effectiveness of the proposed litho-friendliness criteria, we compared our decomposition results with the results of an ILP-based decomposition method which only considers minimum width and spacing constraints in Table 7.3. In Table 7.3, columns "M", "SS" and "sensitive" report the number of feature tiles which are printed by the mandrel mask, the structural spacers and touched by the trim mask, respectively. Also, edge placement error (EPE) of the final image and CPU time are reported in Table 7.4. All the lithography simulations are done based on a composite dipole lithography system where λ=193nm, dose variation=10%, and defocus levels=0.09. Moreover, we assumed overlay error is equal to 30% of half-pitch [16].

Experimental results provided in Table 7.3 and Table 7.4 convey the following points:

- Compared to the basic decomposition method, the proposed method improves the overlay sensitivity via decreasing the interaction between the trim mask and layout patterns.
- Due to maximizing the usage of structural spacers, the proposed decomposition method improves the edge placement error in the decomposed layouts.
- The proposed decomposition method impose minor run time overhead compared to the basic decomposition method. As shown in Table 7.2, the number of variables and constraints grow drastically as the layout area increases because the proposed method is a grid-based method. This search space expansion results in relatively large run time for both methods, which clarifies the need to effective heuristic SATP decomposition methods.

Figure 7.3 compares two valid decomposition schemes provided by the basic decomposition method and our decomposition method, respectively. As illustrated in this figure, the proposed decomposition method achieves higher printability because:

- Compared to a non-litho-friendly solution, structural spacers are applied as much as possible in our solution to print the critical layout features.
- In Figure 7.3.(a), most of structural spacers are formed around mandrel patterns with critical dimensions. Therefore, the imaging imperfections of the mandrel mask

Table 7.3: SATP decomposition results.

Design	SATP-blind			Proposed method			Sensitive Improvement (%)
	M	SS	Sensitive	M	SS	Sensitive	
AND4_X1	82	19	26	42	59	17	28
CLKBUF_X1	35	29	10	26	38	6	37
NAND2_X1	32	21	10	10	43	9	9.75
MUX2_X1	92	57	37	25	124	18	44.8
d1	51	45	18	18	78	15	10.8
d2	1578	1677	448	755	2500	169	58.4
d3	18065	17178	3914	13178	22065	2436	31.5
d4	5942	9639	2596	4664	10917	1204	49

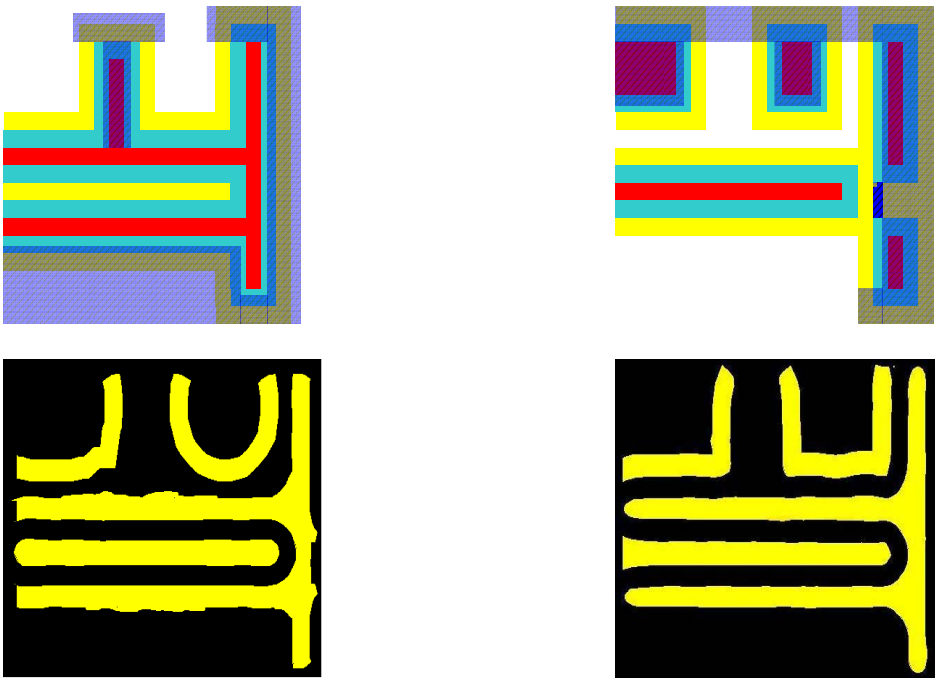
Table 7.4: Edge placement error statistics and runtime of SATP decomposition.

Design	SATP-blind		Proposed method		EPE Improvement (%)
	EPE (nm)	Runtime (sec)	EPE (nm)	Runtime (sec)	
AND4_X1	67.64	38.46	48.65	39.5	34.6
CLKBUF_X1	43.41	32.4	27.3	37	40
NAND2_X1	49.84	25.1	45.1	27	10
MUX2_X1	66.4	70.9	36.67	75.23	51.3
d1	126.496	29.2	112.8	30.1	16.7
d2	91.678	102.42	38.11	108.35	62.3
d3	74.035	284	50.68	305	37.76
d4	111.165	390	56.66	438	53.6

are propagated to the structural spacers. On the other hand, since we used the mandrel mask mainly for dummy patterns, there is higher possibility for aggressive OPC treatments and thus improvement of the overall printability of mandrel mask.

7.5 Conclusion

In this chapter, the effective parameters in the printability of decomposed layouts in the SATP method were discussed. Subsequently, a decomposition method for SATP technique was proposed which uses ILP formulation to avoid decomposition conflicts. The proposed



(a) (b)

Figure 7.3: The printability of SATP decomposed layouts (a) non-litho-friendly solution (b) our solution.

method also improves the overall printability of the layout by using structural spacers efficiently and minimizing the mandrel-trim co-defined layout patterns. The experimental results show that total length of overlay-sensitive layout patterns, total EPE and overall printability of the attempted designs are improved considerably by the proposed decomposition method.

Chapter 8

Conclusion and future work

In this chapter, we summarize our research contributions and discuss future research directions.

As layout feature size is shrinking continuously, there is an aggressive demand for higher resolution of lithography process. Since next-generation lithography systems, e.g. EUV lithography, are lagging behind, we need to continue based on traditional exposure lithography. Multiple patterning lithography (MLP) is the most promising solution to fill the current gap between current technology and the next-generation lithography methods and to realize technology nodes down to 10nm. Double and triple patterning methods are currently the most demanded levels of MLP.

Litho-etch-litho-etch (LELE) and self-aligned double patterning (SADP) are the most promising DPL methods. LELE is the most cost-effective and most comprehensive DPL method; however, it is highly sensitive to overlay error. On the other hand, SADP is fairly robust against overlay error; but, is more complicated in terms of process and layout design requirements. In this thesis, we first study the yield loss of LELE method in presence of overlay variability; and then focus on self-aligned double and triple patterning methods.

To realize DPL in industrial scale, automated design solutions are needed to generate layouts which are friendly to DPL process. Major challenges needed to be addressed in DPL-specific EDA solutions are legal decomposition, overlay-robustness, and litho-friendliness.

In chapter 4, we studied the impacts of systematic and random overlay errors on the functional and parametric yields of LELE interconnects. Then, a statistical method was proposed to determine the interconnect width and spacing for LELE DPL, where the

constraints of RC delay, resistance variability, and total capacitance were satisfied. Our experiments revealed that while functional yield loss is still dominant is 32-nm node, functional and parametric terms of yield loss become comparable in 22-nm node. This finding proves that we need effective parametric yield improvement methods for 22-nm node and below.

In Chapter 5, we introduced two-dimensional SADP process which is more general than the basic SADP process and can print two-dimensional and non-regular layouts. Layout decomposition is complicated in 2D-SADP method because no cut insertion is allowed in SADP. In addition, the decomposition method can contribute to the overlay robustness and litho-friendliness of the final image considerably, where features printed by the trim mask are major sources of image imperfections. Based on performed lithography simulations, spacer-protected trim patterns are printed more accurately than the no-protected patterns. Therefore, we formulated SADP decomposition problem as an ILP constrained by minimum pitch conflicts, where the objective was maximizing the protected trim patterns. Then, using a partitioning-based approach, we improved the run time complexity of our decomposition methods and extended its application to partially decomposable layouts. Experimental results proved that our method can improve the imaging quality significantly compared to a conventional non-litho-friendly decomposition method. However, a considerable subset of attempted designs failed to be decomposed successfully because of their inherent non-compliance to SADP. This shows that SADP-aware layout design methods are highly demanded.

In Chapter 6, we proposed a correct-by-construction detailed routing method which performs detailed routing and SADP-decomposition in parallel. In addition, this method tries to provide maximum spacer protection for trim patterns with minimum overhead on wirelength. Experimental results showed that the proposed method can reduce the number of pitch conflicts and non-protected trim patterns considerably.

In Chapter 7, studying the major sources of pattern distortion in self-aligned triple patterning process, we proposed an ILP-based solution for SATP decomposition problem. The main source of pattern distortion in SATP is mandrel-trim co-defined edges. In addition, the image quality of mandrel patterns is lower than the quality of patterns printed by structural spacers. Therefore, we re-formulated the SADP decomposition ILP formulation in Chapter 5 to adapt SATP process constraints and maximize the length of structural spacer patterns.

The current technology trend shows that the application of multiple patterning lithography will continue towards triple and quadruple patterning. Moreover, both LELE and self-aligned method are likely to survive and being used for different layers or even in a

mixed recipe for one layer.

For LELE approach, higher degrees of patterning mean more limited overlay budget. Therefore, since the precision of overlay control has reached on its limit in the current process, decomposition method should strictly avoid cut insertion. Moreover, the impact of second degree process imperfections such as dose variation will become considerable. To take all these parameters contributing to pattern distortion into account, computationally-effective simulation-based approaches, for example pattern matching and machine learning method, should be used as the core of litho-friendly decomposition methods.

Higher levels of patterning in self-aligned method means more complexity in decomposition and mask synthesis. Because, none of the current SADP decomposition methods can be trivially extended to higher levels of patterning. Effective layout migration methods are also demanded to generate self-aligned compliant layouts. In fact, a major complaint about SADP is that it cannot handle layouts which are designed SADP-blind and complete layout re-design imposes considerable cost overhead. To resolve this issue, we need to effective layout migration methods which resolve SADP non-compatible patterns with minimal tweaks. Moreover, since a SADP-compliant layout is not necessarily compatible to higher levels of self-aligned patterning, layout migration methods would help to transfer layouts into lower technology nodes without complete re-design.

Appendices

Appendix A

Publications from this work

1. Minoo Mirsaeedi, Andres Torres and Mohab Anis. A litho-friendly decomposition method for self-aligned triple patterning," *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*, to appear.

2. Minoo Mirsaeedi, Andres Torres and Mohab Anis. Litho-friendly decomposition method for self-aligned double patterning, submitted to *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*.

3. Minoo Mirsaeedi and Mohab Anis. A atatistical yield optimization framework for interconnect in double patterning lithography. *Microelectronics Journal*, 42(11):1231-1238, 2011.

4. Minoo Mirsaeedi, Andres Torres, and Mohab Anis. A partitioning-based litho-friendly SADP decomposition method. In *IEEE International Workshop on Design for Manufacturability and Yield (DFM&Y)*, 2011.

5. Minoo Mirsaeedi, Andres Torres, and Mohab Anis. Self-aligned double patterning (SADP) layout decomposition. In *IEEE International Symposium on Quality Electronic Design (ISQED)*, pages 103-109, 2011.

6. Minoo Mirsaeedi, Andres Torres, and Mohab Anis. Self-aligned double-patterning (SADP) friendly detailed routing. In *SPIE Advanced Lithography*, volume 7974, page 79740O, 2011.

7. Minoo Mirsaeedi and Mohab Anis. Overlay-aware interconnect yield modeling in double patterning lithography. In *IEEE International Conference on IC Design and Technology (ICICDT)*, pages 138-141, 2010.

Appendix B

Glossary

CD	Critical dimension
NA	Numerical aperture
ACLV	Across-chip line-width variation
APC	Advanced process control
BEOL	Back end of line
CDF	Cumulative distribution function
CDU	Critical dimension uniformity
CMP	Chemical mechanical polishing
CPE	Correction per exposure
DFM	Desgin for manufacturability
DoF	Depth of focus
DPL	Double patterning lithography
EPE	Edge placement error
EUV	Extreme ultraviolet lithography
FEOL	Front end of line
ILP	Integer linear programming
K_1	Emperical lithography resolution parameter
λ	wavelength
LBS	Line by spacer double patterning
LBSF	Line by spacer fill double patterning
LELE	Litho-etch-litho-etch lithography
LFLE	Litho-freeze-litho-etch lithography
LPLE	litho process litho etch double patterning
LWR	line width roughness

Mandrel mask	Fist lithography mask in SADP process
MF	Match factor in pattern matching systems
MLTP	Maximum lateral test pattern
NCE	Non-correctable focus errors
OAI	Off-axis illumination
OPC	Optical proximity correction
PAB	Post applying bake
PDF	Probability distribution function
PEB	Post exposure bake
PSM	Phase shift masking
RELACS	Resolution enhancement lithography assisted by chemical shrink
RET	Resolution enhancement thechnique
SADP	Self-aligned double patterning lithography
SATP	Self-aligned triple patterning
SMO	Source mask optimization
TPL	Triple patterning lithography
Trim mask	Second lithography mask in SADP process

References

[1] A. K. Wong. *Resolution enhancement techniques in optical lithography.* SPIE Press, 2001.

[2] J. F. Chen et al. Optical proximity correction for intermediate-pitch features using sub-resolution scattering bars,. *Journal of Vacuum Science Technology*, 15, November 2001.

[3] Y. Wei and R. L. Brainard. *Advanced processes for 193-nm immersion lithography.* SPIE Press, 2009.

[4] M. Beeck et al. Manufacturability issues with double patterning for 50nm half pitch single damascene applications, using relacs shrink and corresponding opc. In *SPIE Optical Microlithography*, volume 6520, page 65200I, 2007.

[5] W. Shiu et al. Spacer double patterning technique for sub-40nm dram manufacturing process development. In *SPIE Lithography Asia*, volume 7140, page 71403Y, 2008.

[6] W. Arnold. Toward 3nm overlay and critical dimension uniformity: an integrated error budget for double patterning lithography. In *SPIE Optical Microlithography*, volume 6924, page 692404, 2008.

[7] J. Rubinstein and A. R. Neureuther. Post-decomposition assessment of double patterning layout. In *SPIE Optical Microlithography*, volume 6924, page 69240O, 2008.

[8] Martin Drapeau et al. Double patterning design split implementation and validation for the 32nm node. In *SPIE Design for Manufacturability through Design-Process Integration*, page 652109, 2007.

[9] T. Chiou et al. Development of layout split algorithms and printability evaluation for double patterning technology. In *SPIE Optical Microlithography*, volume 6924, page 69243M, 2008.

[10] A. B. Kahng et al. Double patterning lithography aware intelligent layout decomposition. In *SPIE BACUS Symposium on Photomask Technology and Management*, volume 7122, page 712221, 2008.

[11] M. Cho, K. Yuan, and D. Z. Pan. Double patterning technology friendly detailed routing. In *International Conference on CAD (ICCAD)*, pages 506–511, 2008.

[12] K. Yuan, K. Lu, and D. Z. Pan. Double patterning lithography friendly detailed routing with redundant via consideration. In *Design Automation Conference (DAC)*, pages 63–66, July 2009.

[13] J. Yang and D. Z. Pan. Overlay aware interconnect and timing variation modeling for double patterning technology. In *International Conference on Computer-Aided Design (ICCAD)*, pages 488–493, November 2008.

[14] M. Dusa et al. Pitch doubling through dual-patterning lithography challenges in integration and litho budgets. In *SPIE Optical Microlithography*, volume 6520, March 2007.

[15] Kevin Lucas et al. Double-patterning interactions with wafer processing, optical proximity correction, and physical design flows. *Journal of Micro/Nanolithography, MEMS and MOEMS*, 8(3):033002, 2009.

[16] *International Technology Roadmap for Semiconductors*, 2009.

[17] Alex K. Raub, Andrew Frauenglass, Steven R. J. Brueck, et al. Deep-UV immersion interferometric lithography. In *SPIE Optical Microlithography*, pages 306–318, 2004.

[18] John H. Burnett and Simon G. Kaplan. Measurement of the refractive index and thermo-optic coefficient of water near 193 nm. *Journal of Microlithography, Microfabrication, and Microsystems*, 3(1):68–72, 2004.

[19] Abani Biswas and Steven R. J. Brueck. Simulation of the 45-nm half-pitch node with 193-nm immersion lithography—imaging interferometric lithography and dipole illumination. *Journal of Microlithography, Microfabrication, and Microsystems*, 3(1):35–43, 2004.

[20] George E. Bailey, Alexander Tritchkov, Jea-Woo Park, et al. Double pattern eda solutions for 32nm hp and beyond. In *Design for Manufacturability through Design-Process Integration*, page 65211K, 2007.

[21] Jungchul Park, Stephen Hsu, Douglas Van Den Broeke, et al. Application challenges with double patterning technology (dpt) beyond 45 nm. In *Photomask Technology*, page 634922, 2006.

[22] Mircea Dusa, John Quaedackers, Olaf F. A. Larsen, et al. Pitch doubling through dual-patterning lithography challenges in integration and litho budgets. In *SPIE Optical Microlithography*, page 65200G, 2007.

[23] Goji Wakamatsu, Yusuke Anno, Masafumi Hori, et al. Double patterning process with freezing technique. In *SPIE Advances in Resist Materials and Processing Technology*, page 72730B, 2009.

[24] Chris Bencher et al. 22nm half-pitch patterning by CVD spacer self alignment double patterning (SADP). In *SPIE Optical Microlithography*, page 69244E, 2008.

[25] Y. S. Chang et al. Pattern decomposition and process integration of self-aligned double patterning for 30nm node NAND FLASH process and beyond. In *SPIE Optical Microlithography*, page 72743E, 2009.

[26] K. Yuan, J. Yang, and D. Z. Pan. Double patterning layout decomposition for simultaneous conflict and stitch minimization. In *International Symposium on Physical Design (ISPD)*, pages 107–114, 2009.

[27] J. Rubinstein and A. R. Neureuther. Through-focus pattern matching applied to double patterning. In *SPIE Optical Microlithography*, volume 7274, page 72741A, 2009.

[28] G. LukPat et al. Printability verification for double patterning technology. In *SPIE Photomask Technology*, volume 7122, page 71220Q, 2008.

[29] X. Li et al. Double-patterning-friendly OPC. In *SPIE Optical Microlithography*, volume 7274, page 727414, 2009.

[30] Yun Ye, Frank Liu, Sani Nassif, and Yu Cao. Statistical modeling and simulation of threshold variation under dopant fluctuations and line-edge roughness. In *Design Automation Conference (DAC)*, pages 900–905, 2008.

[31] Tejas Jhaveri, Vyacheslav Rovner, Lars Liebmann, Larry T. Pileggi, Andrzej J. Strojwas, and Jason Hibbeler. Co-optimization of circuits, layout and lithography for predictive technology scaling beyond gratings. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 29:509–527, 2010.

[32] H. Fukdau, T. Terasawa, and S. Okazaki. Spatial filtering for depth of focus and resolution enhancement in optical lithography. *Journal of Vacuum Science Technology*, 9, November 1999.

[33] A. Suzuki, K. Saitoh, and M. Yoshii. Multilevel imaging system realizing k1=0.3 lithography. In *SPIE Optical Microlithography*, volume 3679, pages 396–407, 1999.

[34] A. E. Rosenbluth et al. Optimum mask and source patterns to print a given shape. In *SPIE Optical Microlithography*, volume 4346, pages 486–502, 2001.

[35] M. Eurlingsa et al. 0.11 μm imaging in krf lithography using dipole illumination. In *SPIE Conference Series*, volume 4404, pages 266–278, 2001.

[36] V. V. Ivin, T. M. Makhviladze, and K. A. Valiev. Practical aspects of off-axis illumination in optical nanolithography. *Russian Microelectronics*, 33, July 2004.

[37] P. Yu, S. X. Shi, and D. Z. Pan. True Process Variation Aware Optical Proximity Correction with Variational Lithography Modeling and Model Calibration. *Journal of Microlithography, Microfabrication, and Microsystems*, 6, September 2007.

[38] L. Liebmann et al. Alternating phase-shifted mask for logic gate levels, design and mask manufacturing. In *SPIE Optical Microlithography*, volume 3679, pages 27–37, 1999.

[39] B. J. Lin. The attenuated phase-shifting mask. *Solid State Technology*, 35, January 1992.

[40] W. Jung et al. Patterning with spacer for expanding the resolution limit of current lithography tool. In *SPIE Design and Process Integration for Microelectronic Manufacturing*, volume 6156, page 61561J, 2006.

[41] Huixiong Dai et al. Implementing self-aligned double patterning on non-gridded design layouts. In *SPIE Design for Manufacturability through Design-Process Integration*, volume 7275, page 72751E, 2009.

[42] R. S. Ghaida and P. Gupta. Design-overlay interactions in metal double patterning. In *SPIE Design for Manufacturability through Design-Process Integration*, volume 7275, page 727514, 2009.

[43] Mohit Gupta, Kwangok Jeong, and Andrew B. Kahng. Timing yield-aware color reassignment and detailed placement perturbation for double patterning lithography. In *International Conference on CAD (ICCAD)*, pages 607–614, 2009.

[44] X. Li et al. Double-patterning friendly OPC. In *SPIE Conference Series*, volume 7274, page 727414, 2009.

[45] Jae-Seok Yang, K. Lu, Minsik Cho, Kun Yuan, and D.Z. Pan. A new graph-theoretic, multi-objective layout decomposition framework for double patterning lithography. In *Design Automation Conference (ASP-DAC)*, pages 637 –644, 2010.

[46] Bei Yu, Kun Yuan, Boyang Zhang, Duo Ding, and David Z. Pan. Layout decomposition for triple patterning lithography. In *International Conference on CAD (ICCAD)*, pages 1–8, 2011.

[47] Shao-Yun Fang, Yao-Wen Chang, and Wei-Yu Chen. A novel layout decomposition algorithm for triple patterning lithography. In *Design Automation Conference (DAC)*, pages 1185–1190, 2012.

[48] Qiao Li. Np-completeness result for positive line-by-fill sadp process. In *SPIE Photomask Technology*, volume 7823, page 78233P, 2010.

[49] Hongbo Zhang, Yuelin Du, Martin D. F. Wong, Rasit Topaloglu, and Will Conley. Effective decomposition algorithm for self-aligned double patterning lithography. In *Optical Microlithography XXIV*, page 79730J, 2011.

[50] Minoo Mirsaeedi, Andres Torres, , and Mohab Anis. Self-aligned double patterning (sadp) layout decomposition. In *International Symposium on Quality Electronic Design*, pages 1–7, 2011.

[51] C. Mack. *Fundamental principles of optical lithography: the science of microfabrication*. John Wiley and Sons, 2007.

[52] F. Gennari and A. Neureuther. A pattern matching system for linking TCAD and EDA. In *International symposium on quality electronic design (ISQED)*, pages 165–170, 2004.

[53] G. Robins and A. Neureuther. Measuring optical image aberrations with pattern and probe based targets. *Journal of Vacuum Science Technology*, 20:338–343, January 2002.

[54] A. K. Wong. *Optical imaging in projection microlithography*. SPIE Press, 2005.

[55] A. Kahng, C. Park, X. Xu, and H. Yao. Layout Decomposition for Double Patterning Lithography. In *International Conference on CAD (ICCAD)*, pages 465–472, November 2008.

[56] M. Drapeau, V. Wiaux, E. Hendrickx, S. Verhaegen, and T. Machida. Double patterning design split implementation and validation for the 32nm node. In *SPIE Design for Manufacturability through Design-Process Integration*, volume 6521, page 652109, 2007.

[57] V. Wiaux, S. Verhaegen, S. Cheng, F. Iwamoto, P. Jaenen, M. Maenhoudt, T. Matsuda, S. Postnikov, and G. Vandenberghe. Split and design guidelines for double patterning. In *SPIE Optical Microlithography*, volume 6924, page 692409, 2008.

[58] Chin-Hsiung Hsu, Yao-Wen Chang, and Sani Rechard Nassif. Simultaneous layout migration and decomposition for double patterning technology. In *Proceedings of the 2009 International Conference on Computer-Aided Design*, pages 595–600, 2009.

[59] K. Jeong and A. Kahng. Timing analysis and optimization implications of bimodal CD distribution in double patterning lithography. In *Asia South Pacific DAC (ASPDAC)*, pages 486–491, January 2009.

[60] C. Lim et al. Positive and negative tone double patterning lithography for 50nm flash memory. In *SPIE Conference Series*, volume 6154, page 615410, 2006.

[61] R. O. Topaloglu. Variational interconnect analysis for double patterning lithography. In *International VLSI Multilevel Interconnection Conference*, pages 267–270, 2008.

[62] K. Jeong, A. B. Kahng, and R. O. Topaloglu. Is overlay error more important than interconnect variations in double patterning? In *Workshop on System-Level Interconnect Prediction (SLIP)*, pages 3–10, 2009.

[63] K. Jeong and A. B. Kahng and R. O. Topaloglu. Assessing chip-level impact of double-patterning lithography. In *International Symposium on Quality Electronic Design (ISQED)*, pages 122–130, 2010.

[64] B. Eichelberger et al. 32nm overlay improvement capabilities. In *SPIE Conference Series*, volume 6924, page 69244C, 2008.

[65] U. Iessi et al. Double patterning overlay and CD budget for 32 nm technology node. In *SPIE Optical Microithography*, volume 6924, page 692428, 2008.

[66] Won kwang Ma et al. Alignment system and process optimization for improvement of double patterning overlay. In *SPIE Metrology, Inspection, and Process Control for Microlithography*, volume 6922, page 69222T, 2008.

[67] C. Chien and K. Chang. Modeling overlay errors and sampling strategies to improve yield. *Journal of the Chinese Institute of Industrial Engineers*, 18(3):95–103, 2001.

[68] S. Wakamoto et al. Improved overlay control through automated high order compensation. In *SPIE Conference Series*, volume 6518, page 65180J, 2007.

[69] D. Laidler et al. Sources of overlay error in double patterning integration schemes. In *SPIE Metrology, Inspection, and Process Control for Microlithography*, volume 6922, page 69221E, 2008.

[70] Vasudha Gupta and Mohab Anis. Statistical design of the 6T SRAM bit cell. *IEEE Transactions on Circuit and Systems*, 57:93–104, 2010.

[71] S. Director and P. Feldmann. Optimization of parametric yield: a tutorial. In *IEEE Custom Integrated Circuits*, 1992.

[72] P. Kumaraswamy. A generalized probability density function for doublebounded random processes. *Journal of Hydrology*, 46(1/2):79–88, Mar. 1980.

[73] K. Ponnambalam, A. Seifi, and J. Vlach. Probabilistic design of systems with general distributions of parameters. *International Journal of Circuit Theory and Application*, 29:527–536, Dec. 2001.

[74] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999.

[75] M. Wang, X. Yang, K. Eguro, and M. Sarrafzadeh. Multi-center congestion estimation and minimization during placement. In *Proceedings of the 2000 international conference on Computer-aided design*, pages 147–152, 2000.

[76] *Calibre xRC user's and reference manual*. http://www.mentor.com, 2010.

[77] A. B. Kahng, C.-H. Park, X. Xu, and H. Yao. Layout decomposition approaches for double patterning lithography. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 29(6):939–952, 2010.

[78] Yongchan Ban, Alex Miloslavsky, Kevin Lucas, Soo-Han Choi, Chul-Hong Park, and David Z. Pan. Layout decomposition of self-aligned double patterning for 2d random logic patterning. In *Design for Manufacturability through Design-Process Integration V*, page 79740L, 2011.

[79] Hongbo Zhang, Yuelin Du, M.D.F. Wong, and R. Topaloglu. Self-aligned double patterning decomposition for overlay minimization and hot spot detection. In *Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE*, pages 71–76, 2011.

[80] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.

[81] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{logn})$ approximation algorithms for min UnCut, min 2CNF deletion, and directed cut problems. In *ACM Symp. on Theory of Computing (STOC)*, pages 573–581, 2005.

[82] Mohammad Taghi Hajiaghayi and Harald Räcke. An $O(\sqrt{n})$-approximation algorithm for directed sparsest cut. *Information Process Letter*, 97:156–160, 2006.

[83] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Directed metrics and directed graph partitioning problems. In *ACM-SIAM Symp. on Discrete Algorithm (SODA)*, pages 51–60, 2006.

[84] C. M. Fiduccia and R. M. Mattheyses. A linear-time heuristic for improving network partitions. In *Design Automation Conference (DAC)*, pages 175–181, 1982.

[85] *GNU linear programming kit*. http://www.gnu.org/software/glpk.

[86] *Fiduccia-Mattheyses Code*. http://vlsicad.ucsd.edu/UCLAWeb/cheese/codes.html.

[87] *Nangate 45nm Open Cell Library*. http://www.nangate.com.

[88] *ISCAS'85 Benchmark Set*. http://dropzone.tamu.edu/x̃iang/iscas.html.

[89] *Calibre Litho Friendly user's and reference manual (Version 2010.3)*. http://www.mentor.com/, 2010.

[90] Y. Borodovsky. Lithography 2009 overview of opportunities. In *Semicon West*, 2009.

[91] Christopher Cork, Jean-Christophe Madre, and Levi Barnes. Comparison of triple-patterning decomposition algorithms using aperiodic tiling patterns. In *SPIE Photomask and Next-Generation Lithography Mask Technology XV*, page 702839, 2008.

[92] Yijian Chen et al. Self-aligned triple patterning for continuous ic scaling to half-pitch 15nm. In *SPIE Optical Microlithography*, page 79731P, 2011.

[93] Bencherki Mebarki et al. Innovative self-aligned triple patterning for 1x half pitch using single "spacer deposition-spacer etch" step. In *SPIE Optical Microlithography*, page 79730G, 2011.

[94] Yijian Chen et al. Self-aligned triple patterning to extend optical lithography for 1x patterning. In *International Symposium on Lithography Extensions*, 2011.

[95] Yongchan Ban, Kevin Lucas, and David Pan. Flexible 2d layout decomposition framework for spacer-type double pattering lithography. In *Design Automation Conference (DAC)*, pages 789–794, 2011.

[96] Hongbo Zhang et al. Effective decomposition algorithm for self-aligned double patterning. In *SPIE Optical Microlithography*, volume 7973, page 79730J, 2011.

[97] Minoo Mirsaeedi, J. Andres Torres, and Mohab Anis. Self-aligned double-patterning (sadp) friendly detailed routing. In *Design for Manufacturability through Design-Process Integration V*, page 79740O, 2011.