Title
Design Enablement and Design-Centric Assessment of Future Semiconductor Technologies

Author
Abou Ghaida, Rani

Publication Date
2012

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Design Enablement and Design-Centric Assessment of Future Semiconductor Technologies

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Rani Abou Ghaida

2012

Abstract of the Dissertation

Design Enablement and Design-Centric Assessment of Future Semiconductor Technologies

by

Rani Abou Ghaida

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2012

Professor Puneet Gupta, Chair

The semiconductor industry is likely to see several radical changes in the manufacturing, device and interconnect technologies in the next few years and decades. One of the most favorable options of manufacturing technologies is multiple-patterning lithography. This novel technology has serious implications on design, however, and its adoption will necessitate the application of "Design Enablement" methodologies to ensure the compatibility of design.

This dissertation contributes to the design enablement of multiple-patterning technology. We propose a general methodology for the automated adaptation of layout to multiple-patterning masking the complexity in dealing with its manufacturing constraints. We also study the impact of this technology on design and show the benefits of bringing the design perspective into making manufacturing-process decisions. Lastly, we propose a novel technique for DP that reduces cost and improve overlay/Critical-Dimension (CD) control in multiple-patterning.

Many technology choices are presented to achieve scaling to every next node and early technology assessment – before the actual development of technologies – has become more necessary than ever as a means to ensure faster adoption and manageable technology/design development costs. Technology assessment is currently a highly unsystematic procedure; it relies on small-scale experiments and manufacturing tests and much on speculations based on technologists/designers experience

with previous technology generations.

This dissertation also addresses the problem of increasing complexity in making technological decisions. It aims at the development of a computation infrastructure for the systematic and early assessment of technologies and their impact on circuit design. The infrastructure is the first of its kind and is expected to have a lasting impact on technology development. The infrastructure allows for true exploration of design and technology choices, thereby redirecting research and development efforts toward options that are more likely to eventually see adoption. Finally, the infrastructure is applied to evaluate multiple-patterning process decisions and study their implications on design.

The dissertation of Rani Abou Ghaida is approved.

Jingsheng Jason Cong

Lei He

Dejan Markovic

Puneet Gupta, Committee Chair

University of California, Los Angeles

2012

To Said, Nabila, Houssam, Rania, and Hikmat ...

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to dedicate this dissertation to my family. To Said, my father, my inspiration and role model of perseverance and willpower. He showed me that no matter how big the dream is, I will achieve it with "enough" determination. I only wish I could take a leaf out of his book. To Nabila, my mother, my *"let it be"* symbol and the source of my creativity. She showed me that no matter how complicated the situation gets, I can always find a way to make things right. To my siblings, Houssam, Rania, and Hikmat, my lifelong best friends and my strength pillar. They were always there to support me and give me much needed advice. I cordially thank all my family, including Hayat, Sara and Zeina, for the endless love and caring they have given me.

I would like to take this opportunity to express my gratitude to my mentors and teachers.

To Prof. Puneet Gupta, my academic advisor and guide throughout my PhD studies. He showed me the way to approach research and helped me develop many of my research skills. His teachings, encouragement, and counseling, were most crucial to my achievement.

To Dr. Lars Liebmann, my mentor during my second internship at IBM and my role model of what an engineer and scientist should be. He was a source of inspiration and enlightenment. The many research discussions I had with Lars resulted in many original ideas, much of which I have incorporated into my research work.

To Dr. Kanak Agarwal, my mentor during my first internship at IBM. Kanak was always ready to discuss research and exposed me to many research ideas which resulted in a considerable part of this work. He also helped me find and convey the value of my ideas.

To Dr. Sani Nassif, my manager during my first internship at IBM. His visionary aptitude inspired me to look at research from a broader angle. I would also like to thank him for giving me the opportunity to start a great chapter in my career

working with the many exceptional researchers at IBM.

I would like to sincerely thank my committee members, Prof. Jason Cong, Prof. Lei He, and Prof. Dejan Markovic, for their valuable feedback on my research and their enlightening teachings.

It is a pleasure to express my thanks to my friends and colleagues at NanoCAD lab of UCLA for the good times we spent together and for the help they have given me in my research work.

I would like to express my sincere appreciation for the insightful discussions and feedback I received from folks in industry regarding my research. From IBM, Dr. Greg Northrop, Dr. Leon Segal, Dr. Dureseti Chidambarrao, Dr. Ruchir Puri, Dr. Shayak Banerjee, Dr. Kafai Lai and Dr. Amith Singhee. From GlobalFoundries, Dr. Luigi Capodieci, Dr. Swamy Muddu, and Dr. Vito Dai. From Samsung, Dr. Chul-Hong Park. From Mentor Graphics, Pradiptya Ghosh and Dr. Andres Torres. From IMEC, Dr. Mustafa Badaroglu and Dr. Arindam Mallik. I would also like to especially thank Dr. Shayak Banerjee for the good times we spent together during my two internship at IBM and for all his help with work logistics.

I would like to thank the staff of the Electrical Engineering Department, especially Deeona Columbia, for their dedication to a better student experience and for handling administrative matters with tact and high level of professionalism.

I would like to cordially thank my friends who made my stay in Los Angeles feel like home, Said AbulHosn, Analuiza Santos, Dr. Mansour Rachid, Dr. Nabil Hajj Chehadeh, and Karim Ismail.

I would like to thank the Integrated Modeling Process and Computation Technology (IMPACT) research center, the National Science Foundation, and the Semiconductor Research Corporation for their generous financial support of my PhD studies.

Lastly, I would like to acknowledge the reprint of copyrighted material and thank my co-authors.

Chapter 2 is in part a reprint of “Layout decomposition and legalization for

double-patterning technology," submitted to *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* and "A novel methodology for triple/multiple-patterning layout decomposition," published in the proceedings of *SPIE Advanced Lithography – Design for Manufacturability through Design-Process Integration 2012*. I would like to thank my co-authors Dr. Kanak B. Agarwal, Dr. Lars W. Liebmann, Dr. Sani R. Nassif, Dr. Xin Yuan, and Prof. Puneet Gupta.

Chapter 3 is in part a reprint of "Within-layer overlay impact for design in metal double patterning," published in *IEEE Transactions on Semiconductor Manufacturing (TSM) 2010*. I would like to thank my co-author Prof. Puneet Gupta.

Chapter 4 is in part a reprint of "Single-mask double-patterning lithography for reduced cost and improved overlay control," published in *IEEE Transactions on Semiconductor Manufacturing (TSM) 2011*. I would like to thank my co-authors George Torres and Prof. Puneet Gupta.

Chapter 5 is in part a reprint of "DRE: a framework for early co-evaluation of design rules, technology choices, and layout methodologies," published in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD) 2012*. I would like to thank my co-author Prof. Puneet Gupta.

Chapter 6 is in part a reprint of "A methodology for the early evaluation and exploration of double-patterning design rules," to appear in the the proceedings of *IEEE/ACM Intl. Conf. on Computer-Aided Design (ICCAD) 2012* and "A framework for exploring the interaction between design rules and overlay control," published in the proceedings of *IEEE Intl. Workshop on Design for Manufacturability and Yield DFM&Y*. I would like to thank my co-authors Tanaya Sahu, Parag Kulkarni, Mukul Gupta, and Prof. Puneet Gupta.

Vita

2000	French Baccalaureate College Protestant Français, Beirut, Lebanon.
2006	B.S., Computer Engineering Lebanese American University, Byblos, Lebanon.
2005–2006	Co-founder and Managing Director Hanaa Company s.a.l., Beirut, Lebanon.
2008	M.S., Computer Engineering University of New Mexico, Albuquerque, NM, USA.
2010	Research Intern IBM Austin Research Lab, Austin, TX, USA
2011	Research Intern IBM T. J. Watson Research Center, Yorktown, NY, USA
2008–present	Research Assistant Electrical Engineering Department, University of California, Los Angeles, CA, USA.

Publications

Papers in Refereed Journals

[J1] R. S. Ghaida, K. B. Agarwal, S. R. Nassif, X. Yuan, L. Liebmann, and P. Gupta, "Layout decomposition and legalization for double-patterning technology," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, submitted.

[J2] R. S. Ghaida and P. Gupta, "DRE: a framework for early co-evaluation of design rules, technology choices, and layout methodologies," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD'12)*, vol. 31(9), pp. 1379-1392, Sept. (2012).

[J3] R. S. Ghaida, G. Torres, and P. Gupta, "Single-mask double-patterning lithography for reduced cost and improved overlay control," *IEEE Trans. On Semiconductor Manufacturing*, vol. 24(1), 10.1109/TSM.2010.2076305, pp. 381-390, Feb. (2011).

[J4] R. S. Ghaida and P. Gupta, "Within-layer overlay impact for design in metal double patterning," *IEEE Trans. On Semiconductor Manufacturing*, vol. 23(3), pp. 381-390, Aug. (2010).

[J5] R. S. Ghaida, K. Doniger, and P. Zarkesh-Ha, "Random yield prediction based on a stochastic layout sensitivity model," *IEEE Trans. On Semiconductor Manufacturing*, vol. 22(3), pp. 329-337, Aug. (2009).

[J6] R. S. Ghaida and P. Zarkesh-Ha, "A layout sensitivity model for estimating electromigration-vulnerable narrow interconnects," *Journal of Electronic Testing: Theory and Apps.*, vol. 25(1), pp. 67-77, Feb. (2009).

Papers in Peer-Reviewed Conferences

[C1] R. S. Ghaida, T. Sahu, P. Kulkarni, and P. Gupta, "A methodology for the early evaluation and exploration of double-patterning design rules", *IEEE/ACM Intl. Conf. on Computer-Aided Design (ICCAD'12)*, Accepted for publication (to appear in Nov. 2012).

[C2] R. S. Ghaida, K. B. Agarwal, S. R. Nassif, X. Yuan, L. W. Liebmann, and P. Gupta, "O(n) layout-coloring for multiple-patterning lithography and conflict-removal using compaction", *IEEE Intl. Conf. on Integrated Circuit Design and Technology (ICICDT'12)*, Jun. (2012) *(Invited)*.

[C3] R. S. Ghaida, K. B. Agarwal, L. W. Liebmann, S. R. Nassif, and P. Gupta, "A novel methodology for triple/multiple-patterning layout decomposition," *SPIE Advanced Lithography – Design for Manufacturability through Design-Process Integration*, vol. 8327, Feb. (2012).

[C4] R. S. Ghaida, K. B. Agarwal, S. R. Nassif, X. Yuan, L. W. Liebmann, and P. Gupta, "A framework for double patterning-enabled design", *IEEE/ACM Intl. Conf. on Computer-Aided Design (ICCAD'11)*, pp. 14-20, Nov. (2011).

[C5] A. R. Neureuther, J. Rubinstein, M. Miller, K. Yamazoe, E. Chin, C. Levy, L. Wang, N. Xu, C. Spanos, K. Qian, K. Poolla, J. Ghan, A. Subramanian, T.-J. K. Liu, X. Sun, K. Jeong, , P. Gupta, A. Kagalwalla, R. S. Ghaida, and T.-B. Chan, "Collaborative research on emerging technologies and design," *SPIE Photomask Japan – Photomask and Next-Generation Lithography Mask Technology*, Apr. (2011).

[C6] T.-B. Chan, R. S. Ghaida, and P. Gupta, "Electrical modeling of lithographic imperfections," *IEEE Intl. Conference on VLSI Design*, pp. 423-428, Jan.

(2010) *(Invited)*.

[C7] R. S. Ghaida and P. Gupta, "A framework for early and systematic evaluation of design rules," *IEEE/ACM Intl. Conf. on Computer-Aided Design (ICCAD'09)*, pp. 615-622, Nov. (2009).

[C8] R. S. Ghaida, G. Torres, and P. Gupta, "Single-mask double-patterning lithography," *SPIE/ BACUS Photomask Technology*, p. 74882J, Sept. (2009).

[C9] R. S. Ghaida and P. Gupta, "Design-overlay interactions in metal double patterning," *SPIE Symp. on Adv. Lithography - DFM through Design-Process Integration*, p. 727514, Feb. (2009).

[C10] R. S. Ghaida and P. Zarkesh-Ha, "Estimation of electromigration-aggravating narrow interconnects using a layout sensitivity model," *IEEE Intl. Symp. on Defect and Fault-Tolerance in VLSI Systems (DFT '07)*, pp. 59-67, Sept. 2007.

Pending Patents

[P1] R. S. Ghaida and K. B. Agarwal, "Resolving double patterning conflict," *filed*, (2011).

[P2] R. S. Ghaida, K. B. Agarwal, L. W. Liebmann, and S. R. Nassif, "Double patterning layout decomposition for ease of conflict removal," *filed*, (2011).

[P3] R. S. Ghaida, K. B. Agarwal, L. W. Liebmann, and S. R. Nassif, "Pin-access maximization under manufacturing constraints," *filed*, (2011).

[P4] R. S. Ghaida, K. B. Agarwal, L. W. Liebmann, and S. R. Nassif, "Mask assignment for multiple patterning," *filed*, (2011).

[P5] P. Gupta and R. S. Ghaida, "Single mask double-patterning lithography," *filed*, (2010).

Conferences/Workshops with no Publicly Available Proceedings

[W1] R. S. Ghaida, M. Gupta, and P. Gupta, "A framework for exploring the interaction between design rules and overlay control," *IEEE Intl. Workshop on Design for Manufacturability and Yield DFM&Y*, Jun. 2012.

[W2] R. S. Ghaida and P. Gupta, "A framework for systematic evaluation and exploration of design rules," *SRC TECHCON Conference*, Sept. (2009).

Software

[S1] R. S. Ghaida, UCLA Design Rule Evaluator, [Online].
Available: http://nanocad.ee.ucla.edu/Main/DownloadForm, (2009–12).

Theses

[T1] R. S. Ghaida, Semiconductor yield analysis and prediction using a stochastic layout sensitivity model, *University of New Mexico*, Master's Thesis, Dec. (2007).

[T2] R. S. Ghaida, A Different Approach to Fabricating Three-Dimensional Integrated Circuits, *Lebanese American University*, Undergraduate Thesis, Oct. (2005).

Chapter 1

Introduction

Microelectronics technology advancement is mainly reflected in the decrease of cost-per-function, which has been following a historical scaling trend known as Moore's Law. Because scaling principally depends on the minimum feature size of Integrated Circuits (ICs), the method by which circuits are patterned, i.e. optical lithography, has been the primary enabler of technology scaling.

Now that feature sizes have reached the nanometer scale and they are approaching physical limits, technology scaling has become tremendously challenging. At present, the cost of tooling a state-of-the-art fabrication plant for a new a technology generation is about 7 billion dollars while R&D cost is about 1.3 billion dollars [Kay, LaPa]. As a result, the number of companies driving the development of new technology has dwindled to the few that are capable of sustaining such a large investment. What is more alarming is that manufacturing, primarily lithography, has been incapable of achieving technology scaling without significantly increasing process variability and imposing massive geometric restrictions on circuit design. Increased process variability has led to more chips failing to meet power and performance specifications (i.e., parametric yield loss) and has become a serious cause of chip failure in today's technology. Geometric restrictions imposed on design – in form of design rules guiding how the layout should look like – have become more and more complex and have grown to an unmanageable number for traditional rule-optimization methodologies. Many of these geometric constraints, essential to enable future technologies, are expected to increase the design effort, leading to a higher design cost, as well as increase the design area, reducing the effective scaling.

Design for Manufacturability (DFM) has been traditionally used as means to enhance manufacturing yield. Today, due to the huge implications of technology scaling on design, DFM has become a critical and major component of the technology development/adoption cycle. It is now responsible for ensuring the adoptability and profitability of technology while maintaining high-standard product quality.

This dissertation contributes to two topics that have recently emerged in the field of DFM. The first is *Design Enablement*, responsible for ensuring timely and cost-efficient deployment of new technologies; the second is *Design Technology Co-Optimization* (DTCO), responsible for optimizing design and manufacturing hand-in-hand by bringing the design perspective to all stages of technology development.

1.1 Scaling in the Sub-wavelength Lithography Regime

For a conventional optical lithography system the minimum resolution (R_{min}) is by Rayleigh's equation

$$R = k_1 \times \frac{\lambda}{NA}, \tag{1.1}$$

where λ is the wavelength of the light source, NA is the numerical aperture of the lenses, and the k_1 *factor* is a manufacturing process-related coefficient. Shrinking of R_{min} is possible by reducing λ and increasing NA. Reducing λ and increasing NA are constrained because they reduce the depth-of-focus given by

$$DOF \simeq k_2 \times \frac{\lambda}{NA^2}, \tag{1.2}$$

where the k_2 *factor* is a process-related coefficient. Because DOF is linearly dependent on λ and inversely dependent on NA^2, the scaling of the minimum resolution has been historically achieved mainly by scaling the wavelength.

Due to technical issues, the scaling of the wavelength of optical lithography was suspended at 193nm. On the other hand, the adoption of Extreme Ultraviolet Lithography (EUV) and other Next-Generation Lithography (NGL) with a wavelength smaller than 14nm has been continually delayed because of its high cost

and its many technical challenges including defect control and the development of high-energy light source and adequate resist material [ITRa].

Despite the suspension of the wavelength scaling, technology scaling has been continued during the last decade in accordance with Moore's Law. Scaling in sub-wavelength regime was only possible with the development of innovative patterning technologies known as Resolution Enhancement Techniques (RETs). Examples of RETs are: immersion lithography, Off-Axis Illumination (OAI), Sub-Resolution Assist Features (SRAFs), and Attenuated Phase-Shift Mask (attPSM). On the down-side, scaling in sub-wavelength operation results in ever-increasing patterning imperfections and variation of printed shapes from drawn geometries. Perhaps the most expressive statement describing manufacturing with sub-wavelength lithography is: *"[it is] a lot like trying to create an intricate oil painting using a broom"* [NN10].

1.1.1 Multiple-Patterning Lithography

Multiple-Patterning Lithography (MP) is a patterning technology that enhances the resolution of a lithographic system by using multiple lithography-exposure steps to form one single layer of the chip stack. The simplest form of MP is Double-Patterning Lithography (DP), which involves only two exposure steps.

DP was first introduced at the 45nm node in a print and cut scheme [ACC08]. In this scheme, lines are formed with a first exposure step and line-ends, which are harder to pattern, are formed afterward with a second exposure step known as a cut (or trim) exposure. DP in this case allowed abrupt line-ends at the polysilicon (poly) layer (as shown in Figure 1.1), which virtually eliminated corner-rounding at poly and its associated impact on device performance and power variability.

A different DP scheme consists of using the two exposures to achieve pitch splitting (as in Figure 1.2). Here, a first set of features is formed with the first exposure and their direct-neighboring features are formed with the second exposure. This scheme allows the pitch (i.e., minimum center-to-center distance of neighboring parallel lines or vias) for each exposure to be double the effective pitch on wafer.

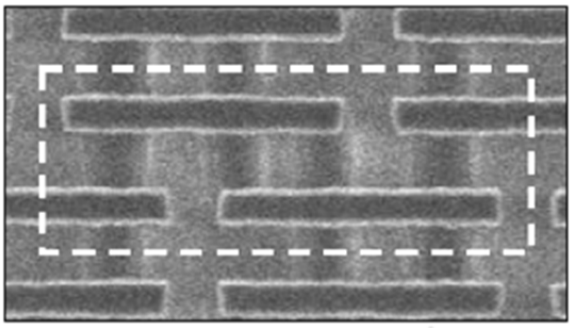


Figure 1.1: Abrupt poly line-ends formed using the trim exposure of a DP process at 45nm technology node (reprinted from [ACC08]).

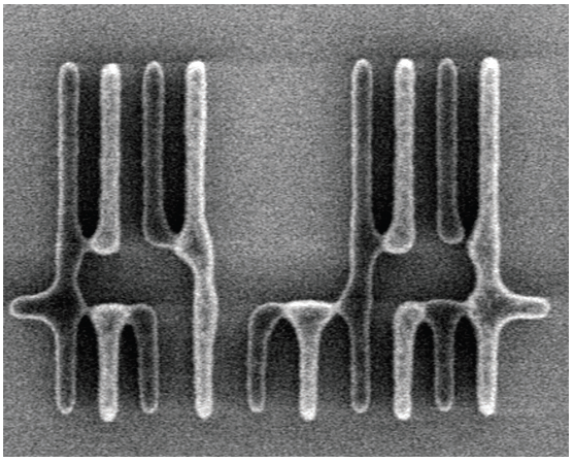


Figure 1.2: Example of pitch-split DP, where dark-colored features were formed with one exposure and light-colored features were formed with a second exposure (reprinted from [Mac08]).

The absolute limit of the k_1 factor from Raylegh's equation (Equation 1.1) with a single-patterning process is 0.28. Consequently, the smallest achievable pitch with state-of-the-art optical lithography systems in a single-patterning process is 80nm [Wal09]. The 20nm technology node (and below) requires sub-80nm pitch and, therefore, single-patterning will not be possible for critical layers. The next-generation of lithography systems (e.g., EUV) will also not be available for high-volume manufacturing at the 20nm node (and possibly at the 14nm node). As a result, pitch-split DP, which can theoretically reduce the k_1 to half that of single-patterning, will be inevitable at this node (at least).

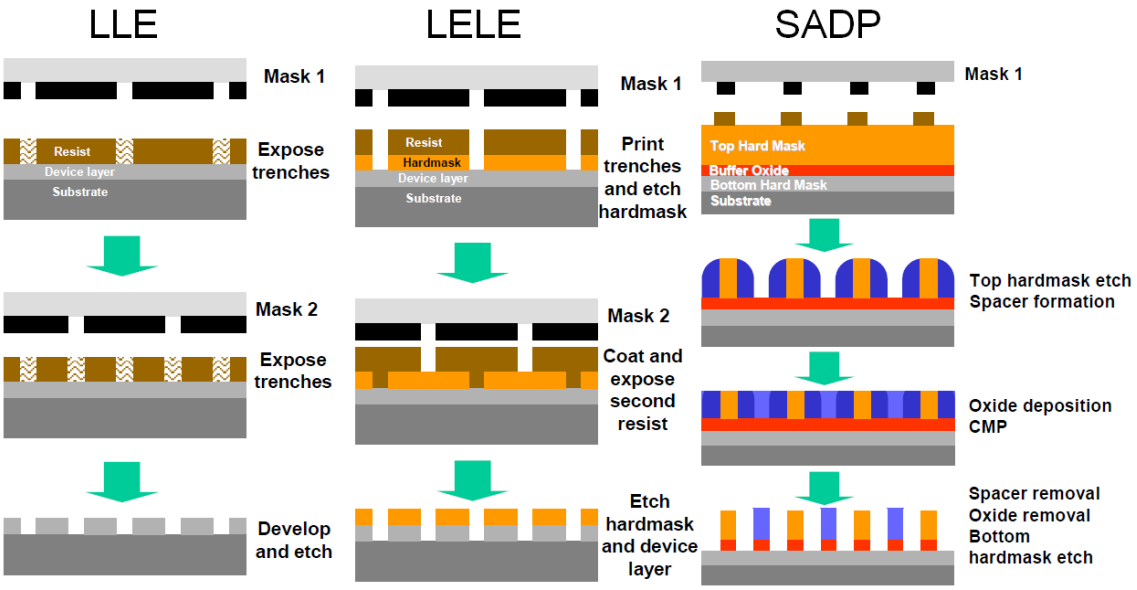


Figure 1.3: Process flows for the different approaches of double-patterning (reprinted from [ITRa]).

Pitch-split DP can be implemented with different manufacturing processes (depicted in Figure 1.3): Litho-Etch-Litho-Etch (LELE), Litho-Litho-Etch (LLE), and Self-Aligned Double-Patterning (SADP), a.k.a. spacer double patterning. In LELE, layout features are split into two sets of features, each getting formed with a separate litho-etch step. This layout decomposition has the objective of relaxing the pitch of each exposure compared with that of the final printed-image on wafer. LLE is similar to LELE with the only difference that one etch step is skipped, mainly, to save on cost. In SADP, sidewall spacer defines either spaces or lines depending on the tone of the process (i.e., positive or negative process). A first set of patterns is formed in a first exposure, a thin film is deposited around the first set of patterns in a spacer-like process, and extra printed features are trimmed away using an additional exposure known as trim exposure (similar to that of the print and cut scheme). In this process, layout features/edges need to be split so that some of the features are formed with the first exposure and the remaining features/edges are formed with the trim exposure.

In all the various processes of pitch-split DP, layout decomposition is constrained so that features assigned to the same exposure meet the pitch/spacing requirements

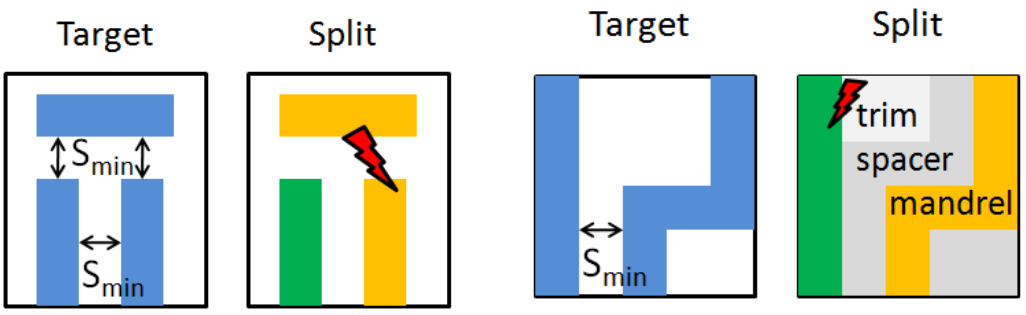


Figure 1.4: Examples of a LELE/LLE forbidden pattern (a) and a SADP forbidden pattern (b). S_{min} denotes the minimum spacing of single-patterning and mandrel denotes layout features that do not appear on the mask of the first exposure.

of single-patterning process. As a result, each of the different flavors of pitch-split DP imposes restrictions on the design layout. For example, the pattern in Figure 1.4(a) cannot be manufactured with LELE or LLE because the pattern cannot be decomposed into first and second exposures without violating the minimum spacing requirement; whereas, the pattern of Figure 1.4(b) cannot be manufactured with SADP because the pattern cannot be decomposed without violating the minimum spacing requirement in the trim exposure. Moreover, the set of shapes that are allowed with SADP is significantly limited compared to that of single-patterning due to the thickness of spacer being uniform across the entire wafer.

One technical challenge of DP is overlay errors between the two patterning steps. Overlay is the positional accuracy with which a pattern is formed on top of an existing pattern on wafer [Mac06]. In traditional single-exposure lithography, overlay errors occur between features of different layers and are handled using low-impact layer-overlap constraints. Because two separate exposures are involved in DP, overlay errors can also occur between patterns of the same layer, effectively translating into CD variability [Arn08, Dus07] and affecting the electrical characteristics of devices and wires.

Triple and Quadruple-Patterning (TP/QP) are also possible options for future technologies. Their processes are similar to that of DP with one or two extra exposure steps. The additional exposure steps are used either to reduce the restrictions

on design or to achieve further scaling beyond the capability of DP.

This dissertation focuses on pitch-split DP/MP and their associated challenges. Hereafter, pitch-split DP/MP will be referred to as DP/MP for brevity.

1.2 Design/Manufacturing Interactions

In todays technology, design and manufacturing has become strongly interactive due to tremendous manufacturing constraints imposed on design.

1.2.1 Design Rules

Design Rules (DRs) are a set of easy-to-verify process-specific geometric constraints that the design layout must obey to ensure a certain level of manufacturability. They reflect the process capabilities and serve as a means to separate process development from design development and a binding agreement between the process team and the design team.

To increase patterns fidelity and mitigate process variability in the sub-wavelength operation, process developers found a need for imposing extra restrictions on the layout. These new restrictions were formulated as design rules and became known as Restrictive Design Rules (RDRs). A classic example of RDR that is widely adopted is single-orientation fixed-pitch polysilicon (poly) gates. As optical lithography continues to be used for future technology generations, RDRs are expected to increase and become more and more restricting.

DRs have a huge impact on the design and productivity. Even small changes in DRs can significantly affect manufacturability [ZCY08] as well as circuit characteristics including layout area, variability, power, and performance [JCS08, She05].

1.2.2 Design Enablement

With new technologies imposing tremendous restrictions on design, design enablement has become crucial to ensure timely and cost-efficient deployment of technol-

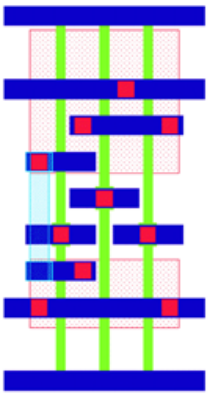


Figure 1.5: Example of a regular fabric, which implements the function !(AB+C) (reprinted from [LPH09]).

ogy.

Design enablement can be achieved in a correct-by-construction fashion. In such an approach, RDRs [LHN04, TY06] are enforced to guarantee a lithography-friendly regular layout [SSD10] and, to ensure an acceptable level of manufacturability, standard-cell libraries and IP modules are required to meet a certain yield-score requirement [Rai06]. In addition to RDRs on poly, RDRs are likely to be adopted to pattern other layers such as metal and contacts/vias [SSD10]. This principle of restricting the layout is pushed to the extreme in [JRL10, JSP09, LPH09] where layout is constructed out of regular fabrics. To build the standard-cell library, pre-characterized regular fabrics (see example in Figure 1.5) are used as a new level of abstraction that replaces design rules. This method not only guarantees extreme regularity but also solves the problem of unmanageable design rules complexity at the 22nm node and below. On the downside, RDRs and regular fabrics are associated with a considerable area overhead as it was shown in [CGK04]. In addition, a correct-by-construction approach can be excessively conservative especially for layouts where patterning imperfections would otherwise be tolerable. Nevertheless, as long as optical lithography is pushed further into sub-wavelength operation and profitability is maintained, scaling requires following the correct-by-construction approach to a certain extent.

An alternative approach for design enablement is through a construct-by-correction

approach. In this case, layouts are allowed to have some lithographic failures known as hotspots. At a later stage, hotspots are detected and fixed using incremental layout modification techniques. For example, a construct-by-correction approach for post-routing hotspot removal was proposed in [MYP05]. Fast lithography simulation was used to detect hotspots and wire spreading and rip up/reroute techniques were performed for hotspot removal. The construct-by-correction approach has had some adoption already through the integration of hotspot-checkers into commercial physical verification tools [Cala, Goe] and hotspot-fixing procedures into commercial routing tools [ICc, RRC07]. The advantage of this approach over the correct-by-construction approach is that it usually results in a tolerable area overhead. On the downside, automated methods are unlikely to remove all hotspots leaving many for costly manual redesign to fix.

1.3 Problem Statement

The semiconductor industry is facing tremendous challenges with the scaling of technology to every new node. Performance targets are no longer realizable simply by feature miniaturization. Power dissipation has become a crucial optimization objective. Scaling profitability is at risk due to growing manufacturing issues such as sub-wavelength lithography imperfections and process variability. To overcome these challenges, the industry is undergoing several radical changes in manufacturing, device and interconnect technologies and is likely to continue to do so in the future. As a result, more and more restrictions are imposed on design increasing the demand and pressure on design enablement to ensure technology profitability and timely adoption.

A large number of technology choices are presented to achieve scaling to every next node (see Figure 1.6). Moreover, these choices will drive and interact with increasing variety in circuit objectives and layout styles. Options in emerging fabrication technologies can be visualized primarily along two axes: (1) a patterning technology axis and (2) a device architecture axis.

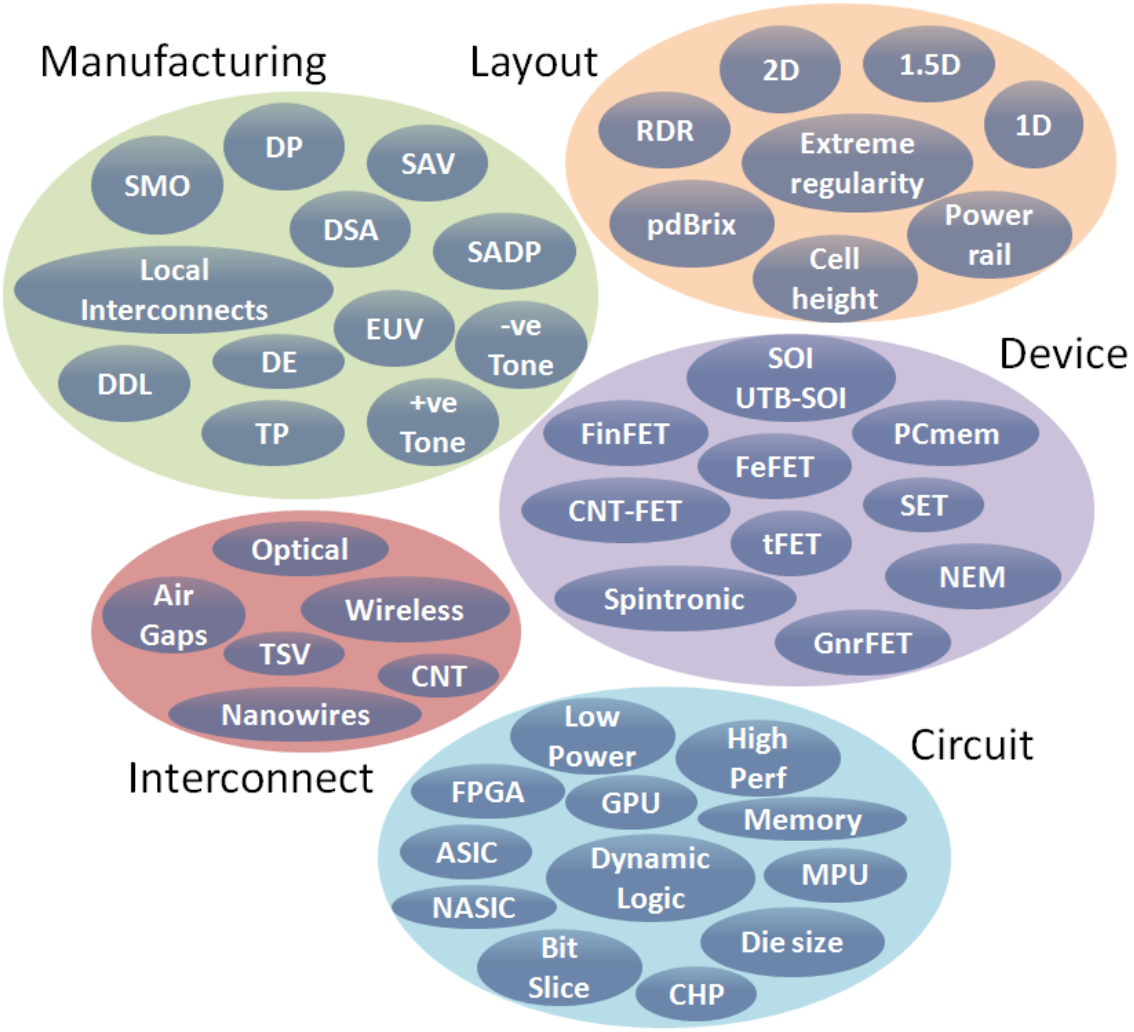


Figure 1.6: Possible technology choices for the scaling to future technology generations.

On the patterning front, technologies that are presented as possible solutions for the scaling to the next few nodes include: double and multiple-patterning with LELE, LLE or SADP process, Source-Mask Optimization (SMO) with extremely regular layouts, local interconnects, and EUV.

In SMO [SSL05, Edn], the illumination source is optimized for a limited set of layout-patterns to ensure manufacturability [JSP09]. The remaining patterns are either forbidden or their use is restrained in the design.

DP/MP could be avoided for certain layers with the use of additional Front-End-Of-Line (FEOL) wiring layers known as local interconnect. Local interconnect

layers are situated between transistors and the first metal layer and they replace the contacts layer. Many local interconnect schemes are presented; they differ by the degree of freedom at each layer, the number of layers needed, and how the local interconnect layers connect to each other and how they connect to transistors and the first metal layer.

Although EUV has a wavelength smaller than 14nm, it brings its peculiar sources of variability including lens flare and mask shadowing [MKN09]. Moreover, the high defectivity of EUV mask blanks may require layout pattern-shifting to avoid defective locations on the reticle [Siv11].

All these patterning solutions impose restrictions on the layout and carry serious challenges and implications on the design cost and quality. Among these patterning technologies, this dissertation focuses particularly on MP since it is one of the most mature and favorable and is expected to have the largest implications on design.

On the device side, many device technologies have been proposed to replace MOSFET. Although significant improvement in performance and leakage power was brought by many techniques such as Silicon-On-Insulator (SOI) and stress engineering, all efforts to extend MOSFET further for future nodes seem to be exhausted. As a result, a radical change in the device technology is very likely. Among the proposed new devices are non-planar multigate devices (FinFET), where the device width is quantized, and ultra thin-body SOI, that may require wiring of the body bias terminal. Each device technology requires unique layout constraints that can have huge implications on the design and, consequently, the overall chip cost.

The accurate assessment of a technology requires the assessment of its implication on the design and density. Moreover, technology assessment needs to be done at early stages of technology development before significant investment in R&D and design enablement had been made. Hence, early assessment of design restrictions imposed by technological choices is absolutely essential.

The eventual purpose of technology is to enable faster, cheaper, less power-hungry and more predictably behaved computational systems. Yet, technology as-

sessment is currently performed on the technology-domain level rather than the design level [ARG09, DCY09]. The evaluation of technology implications on circuit design is traditionally inferred from the evaluation of design rules, which are the biggest design-relevant quality metric for a technology. Unfortunately, even after decades of existence, DR evaluation is largely unsystematic and empirical in nature; it relies on limited and small-scale experiments and manufacturing tests and much on speculations based on technologists/designers experience with previous technology generations [CGK04, ZCY08, DCY09, Cha09].

1.4 Objective and Scope of this Thesis

One objective of this dissertation is to contribute to the design enablement of multiple-patterning lithography by offering methods to enhance the profitability and adoptability of this new technology. In particular, this thesis presents methods for the automated adaptation of layouts for MP technologies, methods for evaluating the impact of MP on design, and a novel manufacturing technique to reduce the overall cost of MP and diminish its impact on design.

This dissertation also aims at the development of a computation infrastructure that enables co-evaluation of manufacturing, device, and interconnect technologies with circuit types and layout methodologies for future generations of computer systems. Specifically, this thesis proposes an infrastructure for design-centric assessment of technology for systematic and qualitative design/technology co-optimization and evaluation of technology choices from a technology/design perspective.

Chapter 2 presents a general compaction-based post-layout methodology for the automatic adaptation of layouts for multiple-patterning technologies. The methodology is applied to adapt 22nm cell and macro layouts from a commercial library for double-patterning lithography and extension of the methodology for triple/quadruple patterning as well as SADP is explained. An important part of the methodology is layout decomposition for which we offer algorithms for the cases of double and triple-patterning.

Chapter 3 studies the relative effects of different overlay sources and the interaction between overlay control and design parameters in context of double-patterning. The electrical impact of overlay errors in double-patterned Back-End-Of-Line (BEOL) is modeled at the design level. The work of this chapter represents an example of design/technology co-optimization and demonstrates its importance. We show that overlay-control requirements in DP can be greatly alleviated when electrical-impact on design, rather than geometric variation, is used to define the requirement. We also compare processing options including positive dual-line and negative dual-trench processes and show the advantages of one over the other from a design perspective.

Chapter 4 proposes a novel technique for DP and studies its implications on density. This technique, which uses a single mask for the two exposures, can bring significant benefits in terms of cost and overlay/Critical-Dimension (CD) control compared to standard LELE/LLE double-patterning.

Chapter 5 presents a computational infrastructure to evaluate design rules, technology choices, and layout methodologies systematically and in a quantitative manner. The evaluation is in terms of area, manufacturability, and variability. By using first-order models of variability and manufacturability and layout topology/congestion-based area estimation, the infrastructure can evaluate big decisions *before* exact process and design technologies are known.

Chapter 6 extends the infrastructure from the previous chapter and applies it for the evaluation of multiple-patterning rules. First, we model the layout compatibility to multiple-patterning using probabilistic routing-estimation and a machine-learning approach. In this chapter also, the infrastructure is further extended to study the interaction between design rules and overlay control. Specifically, methods from previous chapters are combined with a model of overlay yield-loss to evaluate the overall design impact of rules, overlay characteristics, and overlay control options.

Finally, Chapter 7 concludes this dissertation and proposes directions of future research.

Part I

Design Enablement of Multiple-Patterning Technology

Chapter 2

Layout Decomposition and Legalization for Multiple-Patterning Technology

The use of multiple-patterning optical lithography for sub-20nm technologies has become inevitable with delays in adopting the next generation of lithography systems. The biggest technical challenge of multiple patterning is failure to reach a manufacturable layout-coloring solution, especially in dense layouts. This chapter offers a post-layout solution for the removal of conflicts, i.e., patterns that cannot be assigned to different masks without violating spacing rules. The proposed method essentially consists of three steps: layout coloring, exposure layers and geometric rules definition, and, finally, layout legalization using compaction and multiple-patterning rules as constraints. We offer an O(n) layout-coloring heuristic algorithm for double-patterning (DP), which is up to 200X faster than the ILP-based approach, and extend it for multiple-patterning (MP). The conflict-removal problem is formulated as a linear program (LP), which permits an extremely fast run-time (less than 1 minute in real time for macro layouts). The method is general and can be used for different multiple-patterning technologies including LELE double-patterning, triple/multiple-patterning (i.e., multiple litho-etch steps), and self-aligned double patterning (SADP). For demonstration purposes, we apply the proposed method in this paper to remove conflicts in DP. The method was tested on standard cells and macro layouts from a commercial 22nm library designed without any multiple-patterning awareness; for many cells, the method removes all conflicts without any area increase; for some complex cells and macros, the method still removes all conflicts but with a modest 6% average increase in area.

Introduction

With delays in adopting the next generation of lithography systems, the use of double/multiple-patterning (DP/MP) optical lithography for sub-20nm technologies has become inevitable at critical layers. One of the most favorable MP alternatives is pitch-split DP and pitch-split triple-patterning (TP) where layout patterns are formed with two/three separate exposure and etch (or develop) steps (i.e. litho-etch-litho-etch process). Hereafter, we will use the term DP to denote pitch-split DP and TP to denote pitch-split TP.

For a layout to be MP manufacturable, layout features that violate the minimum spacing of single patterning (a.k.a. minimum same-mask or same-color spacing) must be assigned to different masks. The biggest technical challenge of multiple patterning is failure to reach a manufacturable mask-assignment (a.k.a. layout coloring) solution, especially in dense layouts. Layouts designed with conventional rules are generally incompatible with MP; whereas, designing layouts with MP rules is a burden for the designer and requires enormous manual effort. This chapter offers an automated post-layout solution for adapting conventional layouts to MP technology.

In this chapter, we offer an automated post-layout solution for adapting conventional layouts to MP technology. The proposed method essentially consists of three steps: layout coloring, exposure layers and geometric rules definition, and, finally, layout legalization using compaction and multiple-patterning rules as constraints. The method is general and can be used for different multiple-patterning technologies including LELE DP, tripe/multiple-patterning with multiple litho-etch steps, and self-aligned double patterning (SADP). For demonstration purposes, we apply the proposed method in this chapter for DP in LELE process. This chapter also offers a methodology for DP coloring, which guarantees a manufacturable solution when one exists, and extend it for TP.

Our proposed methodology for designing MP-compatible layouts is depicted in Figure 2.1. Using existing non MP-compatible layouts or layouts designed from scratch using conventional rules, we perform an optional step of layout simplifi-

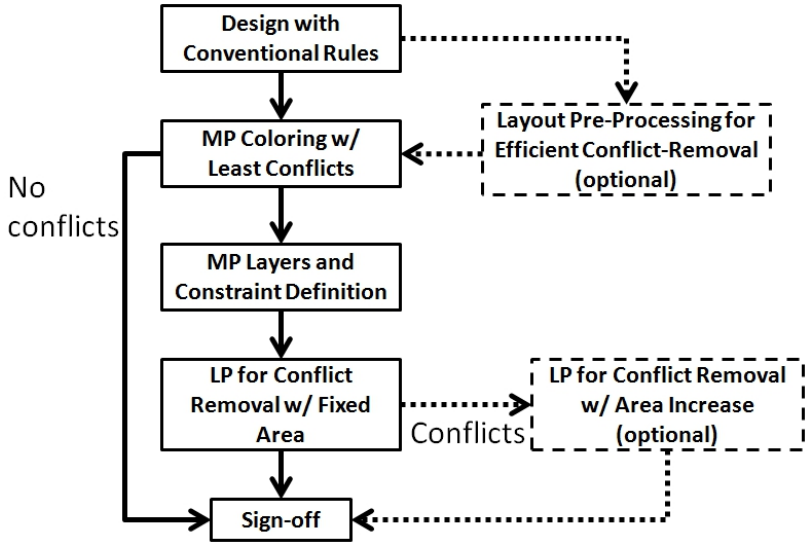
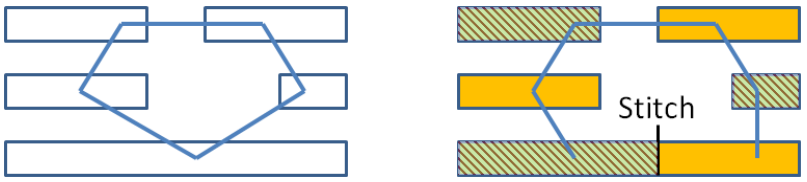


Figure 2.1: The flow for our proposed method to achieve MP-enabled layout design.

cation at MP layers for the possible sacrifice of non-crucial parts as described in Section 2.3.3. We then carry out MP coloring, described in Sections 2.1 and 2.2. If the layout contains MP native conflicts, the conflicts are removed and the layout is legalized simultaneously across all layers while minimizing layout perturbation using a linear-programming (LP) formulation and maintaining the same area as the original layout (described in Section 2.3). Optionally, in case some native conflicts remain unresolved, the LP-based layout legalization is repeated while allowing an area increase to remove more DP conflicts (all conflicts are removed after this step in most cases).

2.1 DP Coloring

DP mask-assignment is essentially a two-color labeling problem [DWH07] and is often referred to as DP coloring or DP layout decomposition. In DP coloring, the layout is represented with a *conflict graph*, where nodes represent layout polygons to be colored and arcs represent coloring constraints. An arc between two nodes denotes a manufacturing constraint on the two corresponding layout polygons to color with two different colors. This constraint is necessary to ensure the printability of non-touching polygons assigned to the same exposure and separated by a distance smaller



(a) (b)

Figure 2.2: Example of a layout with odd cycle in its conflict graph (a) that was broken by introducing a stitch (b).

than the minimum same-color spacing rule. A conflict graph is colorable with two colors and no constraint violations only when the graph contains no odd cycles, i.e. cycles with odd number of arcs; and, an odd cycle is referred to as a coloring conflict.

The difference between DP coloring and the labeling problem of graph theory is that a layout polygon can be a composite of layouts of different masks. The splitting of polygons into multiple parts on different masks is known as *stitching* and the location where the two masks join is called a *stitch*. Although stitching complicates the labeling problem, it is an efficient and almost-free method to conform many, originally DP-unfriendly, layout patterns to DP. In particular, stitching is used to break some odd cycles in the conflict graph getting rid of some coloring conflicts (as illustrated by the example of Figure 2.2). Even with stitching, many patterns cannot be assigned to the two masks without violation of the minimum same-color spacing. Such patterns are called native DP conflicts and resolving these conflicts – with certain layout perturbation – is the biggest challenge facing the deployment of DP.

2.1.1 Prior Art in DP Coloring

Prior works in DP coloring differ mainly by the way stitches are dealt with. Rule-based stitching where polygons are split at certain fixed locations is proposed in [Chi08b, TGK08]. The drawback of this method is that many stitch locations cannot be found by the rules.

In [KPX10, YYP09], the layout is segmented into rectangles, stitches that can

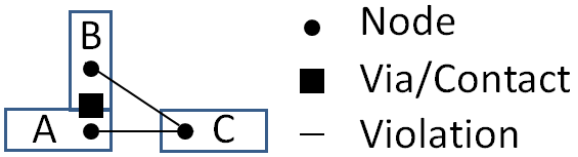


Figure 2.3: Illustration of the drawbacks of segmenting the layout into rectangles, which is performed in prior art of DP coloring and conflict removal.

resolve DP conflicts are determined, and the coloring problem with stitch minimization is formulated as an integer linear program (ILP). Segmentation of the layout into rectangles has many drawbacks. First, it complicates the problem as it forces the consideration a lot of extra stitch locations that should never be used. Consider the example of Figure 2.3. Rectangle C has same-color spacing violations with both rectangles A and B. As a result, A and B must always be assigned to the same mask to avoid a DP conflict and the stitch location at the joint of A and B is never used. The second drawback of segmentation is that it makes the handling of multiple same-color rule values difficult. Because rectangles are mapped into nodes, there is no easy way to distinguish between side-to-side (S2S), tip-to-side (T2S), and tip-to-tip (T2T) same-color spacing rules that may have different values in modern technologies. Another drawback of the methods of [KPX10, YYP09] is that ILPs are very time consuming to solve (NP-hard problem [GJ79]). In addition, the method in [YYP09] can only be applied to gridded layouts with a grid size equal to half the pitch, which is not the case for many layers (e.g., M1).

The work of [XC09] proposes a graph-reduction method to reduce the size of the coloring problem. The method avoids segmentation of the layout into rectangles and its associated drawbacks. On the downside, the method formulates the problem as a maximum-cut problem, which is an NP-complete problem, and solves it using ILP.

The more recent work of [TC11] formulates the coloring problem with stitch minimization as a minimum-cut problem and solves the problem in $O(n^{1.5} \log n)$. In [YLC10], a method for DP coloring with multiple objectives including stitch minimization is proposed. The method is based on min-cut partitioning and the problem is solved in a polynomial time algorithm. These methods are also based

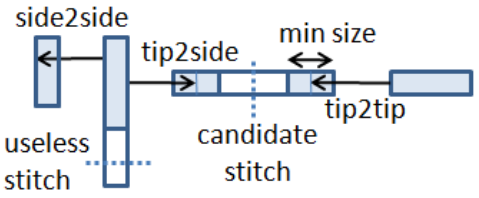


Figure 2.4: DR-dependent projection to identify violating parts and stitch locations. Violating parts are the blue features and non-violating parts are the clear features.

on the segmentation of the layout into rectangles and cannot handle multiple same-color spacing rules. The work in [LH10] offers a method to speed up the coloring process through graph partitioning.

2.1.2 Overview of Our DP Coloring Approach

We follow a different approach for the DP layout coloring than what is presented in the literature. Specifically, we use DR-dependent projection to determine all features that have same-color spacing violations and their actual, possibly non-rectangular, shapes (as in [XC09]). We then formulate the problem as a labeling problem and assign these violating parts to the two masks. In this way, all candidate locations of stitches are automatically defined and can be easily minimized as we show later in this section. In the end, non-violating parts can be assigned to either mask. If a non-violating part touches features of the same mask, we assign it to that same mask to avoid introducing extra stitches; whereas, if a non-violating part touches features of different masks, we assign it to both masks to maximize the overlap region of the masks. Because we use all candidate stitches, our method guarantees a conflict-free coloring solution when the layout has no native conflicts (i.e., conflicts that cannot be resolved with stitching). The way we formulate the problem, allows solving the DP coloring problem with an $O(n)$ heuristic algorithm. The details of this implementation follow.

2.1.3 Multiple Spacing-Rules Projection

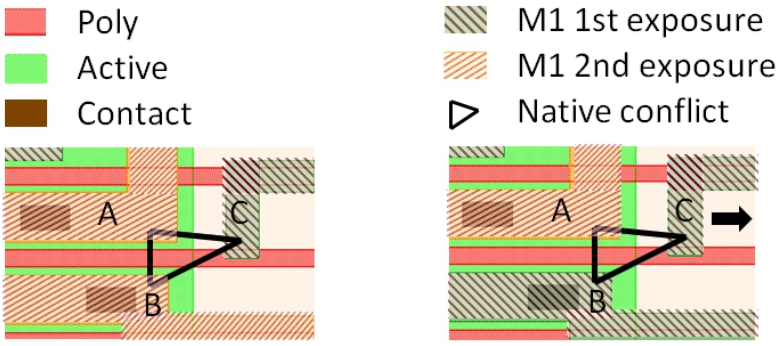
We start with DR-dependent projection to identify violating parts as illustrated in Figure 2.4. From each edge in the layout, side or tip, we project to the neighboring features and determine neighboring edges with which the corresponding same-color spacing rule is violated[1]. From the violating edges and based on the values of the corresponding spacing violation, we determine the exact parts of the layout that violate the same-color spacing with their neighbors. Violating parts that are smaller than the minimum feature size allowed on a single mask are grown within polygons of the original layer to meet the minimum requirement.

Unlike previous works that can only allow a single same-color spacing rule, we allow three same-color spacing rules with different values: side-to-side (S2S), tip-to-side (T2S), and tip-to-tip (T2T). When a single same-color spacing rule is allowed, the largest spacing rule value must be used as the minimum same-color spacing to ensure no DP conflicts are missed. The advantage of allowing multiple same-color spacing rule-values is crucial whenever the values of spacing rules differ, which is the case in latest technologies. The importance of allowing different values for the different rules will be quantified later in this section.

2.1.4 Coloring Objectives

The main objective of DP coloring is to assign features to the two different masks with the minimum number of conflicts. A secondary objective is to minimize stitches, which may increase yield loss due to overlay error between the first and second exposure layers. Because stitches can remove certain conflicts (as illustrated in Figure 2.2), we consider all possible stitch locations during coloring and get rid of stitches that do not affect the number of conflicts. If a stitch is introduced inside any violating part, then one of the stitch's sides will have to be assigned to the same mask as the neighboring part that created the violation, which leads to a new DP conflict (as in Figure 2.3). As a result, stitches should be located only in non-violating

[1]This can be done using existing DRC tools and in a similar fashion as in [KPX10].



(a) (b)

Figure 2.5: Odd cycle coloring can affect the efficiency of conflict removal. In (a), the conflict is on M1 between shapes A and B and can only be fixed if the gates are spaced apart and area is increased; in (b), the conflict is on M1 between shapes B and C and can be fixed by moving C in the direction of the arrow without increasing area.

parts. Since DP conflicts are between violating parts only, stitches are beneficial (i.e. may reduce the number of conflicts) only if placed in non-violating parts that separate two or more violating parts. In other words, a single stitch is sufficient in such non-violating parts and stitches in a non-violating part that connect to a single violating part is useless because we can always assign such non-violating part to the same mask as the connected violating part (see example of Figure 2.4). In addition, stitches that cannot guarantee the minimum overlap length of the two masks are disregarded (by joining the connected violating parts).

Although an odd cycle will always result in a DP conflict no matter the coloring, deciding what features go on the same mask can affect the efficiency of the conflict removal. To see how, consider the example of Figure 2.5. This layout contains an odd cycle between shapes A, B, and C. In Figure 2.5(a), the coloring solution leads to a conflict between shapes A and B that can be resolved only if the gates are spaced apart and, consequently, the layout area is increased; whereas, in Figure 2.5(b), the coloring solution results in a conflict between shapes B and C that can be resolved by moving C to the right without increasing the layout area. To take advantage of this observation, we make violations in the orthogonal orientation of gates (vertical violations for our layouts) more critical than the ones in other orientations

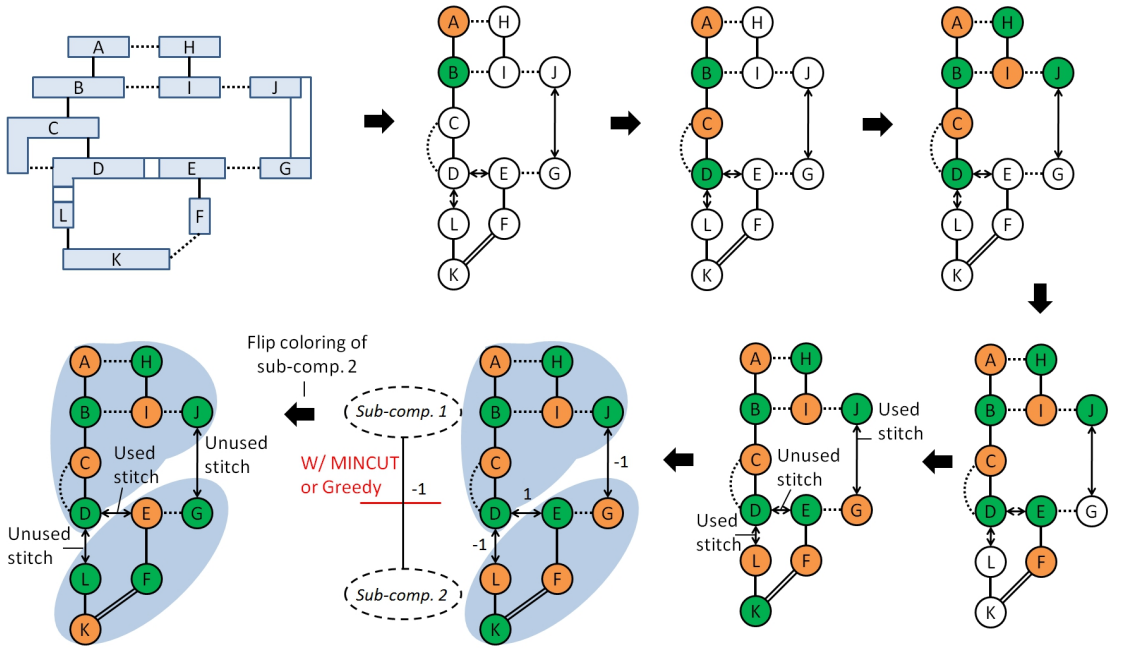


Figure 2.6: An illustrating example showing each step of the coloring process for an isolated region of the layout.

(horizontal and diagonal violations for our layouts). Similarly, we make horizontal violations more critical than diagonal violations because the latter typically require less additional separation to fix.

2.1.5 Implementation Details

The coloring of violating parts is straightforward and its first-stage initial coloring is performed in $O(n)$, where n is the number of violations and candidate stitches. An example that illustrate the coloring steps is given in Figure 2.6 and the details of the algorithm are presented in Figure 2.7.

We start by constructing the conflict graph, where violating parts are represented by nodes and violations and stitches are represented by arcs. We represent vertical violations by solid arcs, horizontal violations by dotted arcs, diagonal violations by double-line arcs, and stitches between two shapes by arcs with two-sided arrows. For each connected component (identifying connected components is $O(n)$), we pick a violation-arc with preference to vertical over horizontal and horizontal over diagonal arcs and assign the two connected nodes to different masks. Whenever a new node

is assigned, its connected arcs get added to first-in-first-out (FIFO) queues of the different types of violations and stitches to be processed next. A new arc (possibly a stitch-arc) is popped from the different queues with preference to violation-arcs over stitch-arcs and the same preference for the different violation-arcs as before. This process is repeated until all arcs in the component are processed. Each node is assigned only once: when a violation-arc is processed, the two nodes are assigned to different masks and, when a stitch-arc is processed, the two nodes are assigned to the same masks.

We perform a second-stage coloring where the initial coloring is possibly flipped to further reduce the number of used stitches. Each part of a component that is connected with violation-arcs only (without stitches) is called a sub-component and stitches connect different sub-components (see Figure 2.6). Each sub-component has a flipping score based on which the coloring of its nodes is flipped or preserved. When a stitch is processed, we record the connection of the two connected sub-components; if the stitch is used, the flipping score is decremented by one (the score being zero initially); if the stitch is unused, the flipping score is incremented by one. So, by flipping a sub-component with a negative score, the number of stitches is reduced by the amount of the score.

We follow two approaches to determine the sub-components where color-flipping is beneficial. The first is a greedy heuristic algorithm where sub-components with negative scores are flipped in a decreasing order of scores. When a sub-component is flipped, the sub-component and its neighbors are prevented from future flipping (i.e., flipping is locked). Although this algorithm may reach a sub-optimal solution, it ensures $O(n)$ running time for the overall coloring procedure. The most suboptimal solution occurs for the case shown in Figure 2.8. Here, the greedy algorithm will only color-flip the center node with the highest flipping-score N. The optimal solution is to flip every other neighbor of the center node with a score of $(N-1)$. Since the center node has N neighbors, the optimal solution results in $N\frac{N-1}{2}$ less stitches than the initial coloring solution and $N\frac{N-3}{2}$ stitches less than the solution obtained with the greedy algorithm. This worst case and similar bad scen-

1: Perform DR-dependent projection.
   Identify violating parts.
   Identify all useful candidate stitches.
2: Construct conflict graph with nodes representing violating parts and four types of arcs representing vertical violations, horizontal violations, diagonal violations, and stitches.
3: Create separate FIFO queues for the different types of arcs.
4: Determine connected components.
5: Determine connected sub-components (i.e. without stitch connections).
6: **for all** Connected components **do**
7: Pick any violation-arc with preference to vertical over horizontal and horizontal over diagonal and assign its nodes to different masks.
8: Push all arcs connected to the assigned node into the different FIFO queues.
9: Pop an arc (possibly a stitch-arc) from the different queues with preference to violation-arcs over stitch-arcs and the same preference for the different violation-arcs as above.
10: Assign the two nodes connected to the arc to different masks if the arc is for a violation and to the same mask if the arc is for a stitch.
11: **if** Arc is a stitch **then**
12: Record the connection of the two sub-components (the two nodes connected to the stitch-arc belong to different sub-components).
13: If the stitch is used (i.e. connected nodes were assigned to different masks), increment the flipping-score of the two sub-components by one.
14: If the stitch is unused (both connected nodes assigned to same mask), decrement the flipping-score of the two sub-components by one.
15: **end if**
16: Repeat steps 9 to 15 until all arcs in the component are processed.
17: **end for**
18: **for all** Sub-component with a positive flipping score sorted by descending score **do**
19: Skip if already processed or marked not to be processed
20: Mark as flipped/processed and mark its neighbors as processed (to prevent future flipping)
21: **end for**
22: **for all** Nodes **do**
23: Flip node if it belongs to a flipped sub-component
24: **end for**

Figure 2.7: $O(n)$ coloring procedure with greedy algorithm for color-flipping[2].

[2]Although the loop of 18 to 21 is theoretically higher than $O(n)$, it takes much less time to execute than the $O(n)$ loop of line 6 to 17 because neighbors of flipped sub-components are skipped and the number of neighbors for a sub-component is limited in practice (at most 10).

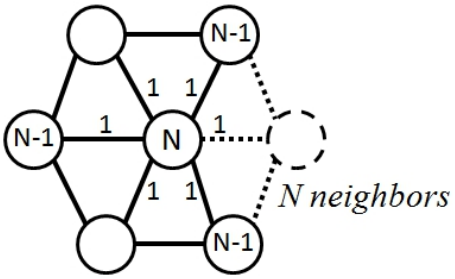


Figure 2.8: Case of most suboptimal solution for greedy-based flipping.

arios are uncommon for actual layouts, however, and the sub-optimality of the greedy flipping is limited in practice as we will show in our experimental results (Section 2.1.7).

In the second flipping approach, the flipping problem is formulated as a minimum-cut problem (inspired by the minimum-cut formulation for graph coloring of [TC11, XC09]). We construct a flipping graph where nodes represent sub-components and every edge represents one or more candidate stitches between two-sub-components. A used stitch is associated with a flipping score of -1 and an unused stitch is associated with a flipping score of 1. The sum of flipping scores of all candidate stitches between any two-sub-components determines the edge's weight as shown in Figure 2.6. Now, the problem is equivalent to partitioning the graph into two parts, one part where the coloring will be flipped and another part where the coloring will be preserved. The problem is solved optimally by partitioning the graph based on the minimum cut with the smallest sum of weights (as it was shown in [TC11]). In case the conflict graph has no DP conflicts (i.e., no odd-cycles), such solution gives the minimum number of stitches; in case the graph has DP conflicts however, optimal flipping may not correspond to optimal overall number of stitches because the way the odd cycle is colored can affect the number of stitches.

For finding the minimum cut, we use Stoer and Wagner's MINCUT algorithm [SW97], which has a running time of $O(n \log(n))$. If the minimum cut has a negative value, the coloring of sub-components of one of the partitions (i.e., from one side of the cut) is flipped to reduce the number of used stitches by the absolute

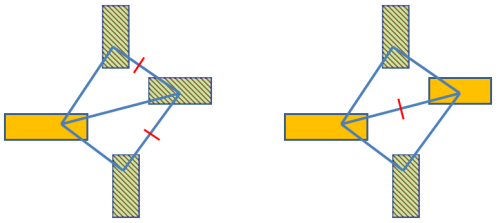
value of the cut.

It is important to note that, for the MINCUT algorithm we used [SW97] to work properly and to enhance the run-time, the algorithm is applied for each connected component of the conflict graph separately since the coloring of connected components can be performed independently.

2.1.6 Stitches vs. Conflicts and Special Cases

Stitches are manufacturable, DP conflicts are not. When the requirement for the minimum mask overlap is met, stitches are safe to manufacture and their minimization is recommended rather than required. Moreover, stitches may occur in millions in large layouts and reducing the number of stitches by few percents does not have a significant impact on the manufacturing yield. On the other hand, a layout with a single DP conflict can never be manufactured. As a result, our primary objective in this work was to achieve a coloring solution with the least number of DP conflicts. Although our method minimizes the number of stitches, it does not guarantee achieving the minimum number of stitches. Most importantly, because we consider all candidate stitch locations, our method guarantees to reach a solution with DP violations only at the locations of *native conflicts* (i.e. conflicts that cannot be resolved with stitching) and a conflict-free solution for layouts without *native conflicts*. A DP native conflict is defined as an odd cycle in the conflict graph that cannot be resolved with stitching. By performing the coloring with a conflict graph that includes all candidate stitches and while ensuring any two nodes with a violation that are not part of an odd cycle are assigned to different masks, our method leads to a solution with zero *non-native conflicts* (i.e. conflicts that are resolvable with stitches).

For some special cases with two or more native-conflict odd cycles share some of their arcs, our method may lead to a solution with non-minimum number of *DP violations* at such native conflicts depending on the propagation order of the coloring. One such special case is shown in Figure 2.9. Because we set a propaga-



(a) (b)

Figure 2.9: Example showing two coloring solutions that our method may give for the same layout (rare case with all diagonal violations) depending on the propagation order of the coloring: (a) with two DP violations and (b) with a single DP violation.

tion preference with purely vertical violations first, diagonal violations second, and purely horizontal violations last, all violations in this four-tip configuration must be diagonal violations for the method to result in the coloring solution with two DP violations (Figure 2.9(a)) for some propagation order; otherwise the method will result in the coloring solution with a single DP violation as in Figure 2.9(b). Besides the peculiarity of this layout, such four-tip configuration may never occur because contacts/vias are on tracks in actual layouts. Furthermore, the number of DP violations may not reflect the amount of effort needed to remove the violations with layout perturbations and, in this special case, the coloring with two DP violations may be easier to fix than that with a single violation depending on the layout (at the same layer as well as the top and bottom layers).

2.1.7 Coloring Results

Our DP-coloring method was implemented using Calibre SVRF code [Calb], for performing projection and forming the final mask-layout, and C++ with OpenAccess database/ API for the actual coloring of non-violating parts of the layout. An implementation from Boost C++ Libraries [Boo] was used for identifying connected components and sub-components and the implementation for the Stoer and Wagner MINCUT algorithm from OGDF library [OGD] was used to find the optimal color-flipping of sub-components.

We test our greedy-based flipping and MINCUT-based flipping coloring methods

Table 2.1: Results of our DP coloring at the M1 layer with two sets of same-color spacing rules and comparison with our greedy and MINCUT-based flipping approaches. The minimum different-color spacing rule is 65nm, while the minimum overlap length is 10nm.

		S2S = 110, T2S = 120, T2T = 130nm					
M1 Layer		**Greedy Flipping**			**MINCUT Flipping**		
Design	**Instances**	**Viol**	**Stitches**	**Seconds**	**Viol**	**Stitches**	**Seconds**
OR1200	3,077	1,007	2,152	0.17	995	2,096	7.20
TV80	6,429	3,692	5226	0.48	3,692	5226	0.52
AE18	10,556	9,053	9,597	1.04	9,053	9,597	1.14
MIPS789	19,868	21,273	26,753	2.35	21,273	26,753	3.27

		S2S = T2S = T2T = 130nm					
M1 Layer		**Greedy Flipping**			**MINCUT Flipping**		
Design	**Instances**	**Viol**	**Stitches**	**Seconds**	**Viol**	**Stitches**	**Seconds**
OR1200	3,077	2,119	1,711	0.15	2,119	1,649	0.54
TV80	6,429	5,667	5,494	0.48	5,667	5,317	9.29
AE18	10,556	14,578	9,376	1.03	14,578	9,376	1.03
MIPS789	19,868	28,582	29,061	2.42	28,578	28,098	405.31

on M1 layouts of designs from [opea] with number of cell-instances ranging from 3K to 20K. We perform this testing with two sets of same-color spacing rules. The first set is with different values for the different rules and the second set is with the same value for all same-color spacing rules[3].

The results, depicted in Table 2.1, show that the greedy-based flipping leads to at most 3.8% larger number of stitches than that achieved with the MINCUT-based flipping, which has the same run-time complexity as [TC11]. And, for many of the cases, both methods lead to the same number of stitches; those are the cases when no color-flipping was necessary.

Previous works on DP coloring allow a single same-color spacing rule-value. Consequently, even if the manufacturing process permits the first set of rules, the coloring

[3]The rule values are assumed and may be different in an actual process. Yet, whenever the rule values are different and no matter the actual values, allowing multiple same-color spacing rules during the coloring and layout legalization leads to a reduced number of violations (or at least the same) compared to when a single same-color spacing rule is allowed.

methods of previous works must be performed with the second set of rules (i.e., all rules are equal to the largest of all rules) to ensure no DP conflicts are ignored during the coloring. The results of Table 2.1 show that, in this case, the number of violation increases by 38% on average and by 53% in the worst case. Hence, allowing multiple rule-values to be used in the coloring is crucial when the rules have different values, which is the case in latest technologies.

Since color-flipping does not affect the number of violations in the conflict graph, the number of violations resulting from both coloring methods, i.e. the greedy-based and MINCUT-based flipping, should ideally be the same. In some cases however, the number of violations obtained with both methods may be slightly different (e.g., OR1200 design with the first set of rules and MIPS789 design with the second set of rules in Table 2.1) due to a coloring problem that will be explained in Section 2.3.

To quantify the benefit of stitches in reducing the number of violations, we repeat the DP coloring while forbidding stitches and compare the results to that from Table 2.1 (case of S2S = S2T = T2T = 130nm). The results, depicted in Figure 2.10, show that allowing for stitches reduces the number of violations by roughly 40%. Therefore, forbidding stitching entirely in DP is expected to significantly restrict the layout. It is important to note that stitching may be forbidden at certain locations as in [TGK08] (e.g., at corners). Forbidding candidate stitch locations can be applied easily with our coloring approach by merging regions of forbidden-stitching with overlapping violating parts as illustrated in Figure 2.11. If the region of forbidden stitching has no overlapping violating part, no measure needs to be taken as stitching will only occur at the interface with a violating part, which is outside the forbidden region in this case.

Different methods for measuring violations can result in significantly different number of violations. For example, corner-to-corner same-color spacing violations may be considered as violations in one method but ignored in another; also, one method may consider same-color spacing violations between vertical segments of any U-shape, but another method may ignore these if one or both vertical segments are shorter than a certain length. Also, since stitches are used to remove DP con-

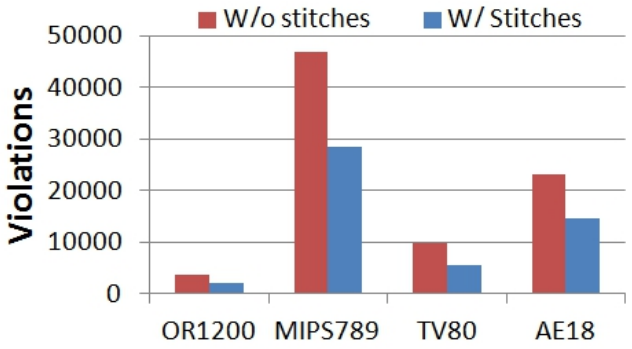


Figure 2.10: Comparison of the number of same-color spacing violations when stitches are forbidden and when they are allowed.

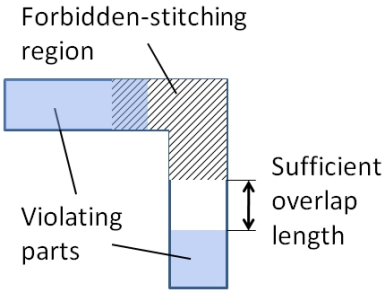


Figure 2.11: Region of forbidden stitching at corner, which will be merged with the overlapping violating part on the left to prevent stitching at the corner. Note if the overlap length in the non-violating part is smaller than the minimum rule, the non-violating part will also be merged with the violating parts and no candidate stitches will be available in this shape.

flicts/violations, if the number of identified violations is different, the number of stitches will also be different. Hence, to compare the number of violations and number of stitches of our coloring approaches with that of a previous work, the same method must be used to detect violations in layouts colored with our methods and layouts colored with methods from the previous work.

Five layout test cases from [KPX10] were available to us: four at the Poly layer in uncolored form and one at the M2 layer in colored form. For the colored test case, we use our violation-measuring method to determine the number of violations in the initial coloring, performed using the pure ILP approach of [KPX10], and the coloring performed using our coloring approaches and the same DP rules (same-color spacing and overlap length). Table 2.2 reports the number of violations and number of stitches for the different coloring methods. Compared with the layout

Table 2.2: Comparison of number of violations and stitches in a M2 layout colored with the pure ILP approach of [KPX10] and the same layout colored with our greedy/MINCUT-based approaches.

	ILP [KPX10]		**Our Greedy**		**Our MINCUT**	
	Viol	**Stitches**	**Viol**	**Stitches**	**Viol**	**Stitches**
AES45	1,793	1,848	887	1,779	887	1,764

colored using the ILP-based method of [KPX10], the layouts colored with our coloring approaches contain roughly half the number of violations, while the number of stitches is proportionate[4]. For the uncolored test cases as well as the colored test case, we compare the running times of our coloring approaches with those reported in [KPX10] for the pure ILP approach and the conflict cycle detection (CCD) approach, which artificially removes odd-cycles prior to solving the ILP. The results are shown in Table 2.3. The layouts are at the Polysilicon (Poly) and M2 layers in 45nm process node for designs with cell-instances ranging from 26K to 300K. The same assumption of minimum same-color spacing (1.2× minimum-spacing relaxation for Poly and 1.1× minimum-spacing relaxation for M2) and overlap length were used for all coloring methods.

Our DP coloring method with greedy-based flipping is the fastest: up to 80× faster than the pure ILP approach, 63× faster than the conflict cycle detection (CCD) approach of [KPX10], and 19× faster than our coloring method with MINCUT flipping. Our MINCUT-based flipping method results in up to 9% smaller number of stitches compared with our greedy-based flipping. It is also up to 9× faster than the pure ILP approach and 7× faster than the CCD approach of [KPX10].

2.2 TP Coloring

For dense bidirectional layers such as the first metal layer (M1), DP typically results in a large unmanageable number of conflicts. Although the number of conflicts

[4]This result may not be generalizable, however, as it is based on the single layout that was available in colored form. Moreover our violation-measuring method that was used to report the results may differ from that of [KPX10] and, hence, many of the violations may not have been considered while performing the coloring of [KPX10].

Table 2.3: Results of our DP coloring at the Poly and M2 layers (minimum-spacing relaxation of 1.2× for Poly and 1.1× for M2) and comparison with methods from previous work of [KPX10]. "Inst", "Viol", "Secs", and "Stch" refer to number of instances, violations, user-time seconds, and stitches respectively, while "Min" refers to minimum overlap length).

				ILP [KPX10]	**CCD [KPX10]**	**Our Greedy**		**Our MINCUT**	
Design	**Inst**	**Min**	**Viol**	**Secs**	**Secs**	**Stch**	**Secs**	**Stch**	**Secs**
ART-A 45(70%)	100K	8	5,976	565	379	29,692	11	27,909	81
ART-B 45(70%)	300K	10	17,912	2,887	2,317	93,262	36	84755	702
ART-A 45(90%)	100K	13	5,976	612	391	33,139	12	30,548	57
ART-B 45(90%)	300K	10	17,912	2,892	2,355	98,053	37	89,734	722
AES45 M2 Layer	26K	20	887	23.5	5.5	1,779	0.8	1,764	0.7

can be significantly reduced by relaxing spacing constraints (such as minimum T2T and T2S spacing rules), DP suffers in this case from poor density scaling as we show later in Chapter 5. As a result, TP in a LELELE process has emerged as a strong alternative for dense bidirectional metal layers. Because of the use of a third exposure/etch, TP can achieve good scaling of rules as well as improved pitch scaling over DP in case further scaling is needed.

As all DP technologies, TP requires the decomposition of the layout into different masks, a.k.a. layout coloring. Coloring is challenging for DP and it is even more challenging for TP. In comparison with DP coloring, the complexity of TP coloring is attributed to the following:

1. Determining whether the layout is DP compatible (i.e., the graph is 2-colorable) can be done very efficiently (in polynomial time) by simply checking if the graph is free of odd cycles; whereas, determining whether the layout is TP compatible (i.e., the graph is 3-colorable) is an extremely hard problem for

which there is no efficient way to find a solution (classified as NP-complete problem in computational complexity theory). Moreover, in the 2-coloring problem, conflicts and their locations can be easily detected by identifying odd-cycles in the conflict graph. In the 3-coloring problem, there are no existing methods for determining the number of conflicts and their locations.

2. In DP, candidate stitch locations can be easily determined prior to the actual coloring (using projection as in 2.4. In TP, however, a stitch can be between the first and second masks, the first and third masks, and the second and third masks. As a result, stitches are color-dependent and candidate stitch locations can be determined only after or during coloring.

2.2.1 Prior Art in TP Coloring

A naive extension of 2-coloring for DP to 3-coloring for TP is to start with a coloring solution for DP and use the third color to resolve DP coloring conflicts. Although this approach is simple and does not require the development of new and sophisticated methods for TP coloring, it may lead to a low-quality solution in terms of the number of conflicts. Consider the example in Figure 2.12. Figure 2.12(a) shows the initial DP coloring solution that contains conflicts; Figure 2.12(b) shows a TP coloring solution resulting from the naive approach and containing a conflict; and Figure 2.12(c) shows that a conflict-free TP coloring solution can be achieved with the use of a stitch.

Although ways for performing TP coloring has been hinted at in some patent disclosures (e.g., [Soc10]), the only complete previous work that covers TP coloring for lines in LELELE process is the work of [YYZ11]. It suggests performing the TP coloring by solving an integer linear program (ILP), which minimizes coloring conflicts as well as stitches. Solving the coloring problem with an ILP was shown to reach good solutions for DP [KPX10, YYP09, YLC10] and the extension for TP is likely to reach good solutions for the 3-coloring problem as well. For design-level layout coloring with large designs, solving the ILP is impractical because it requires a

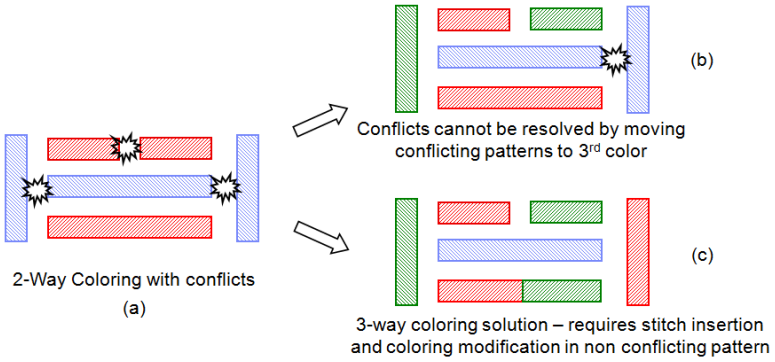


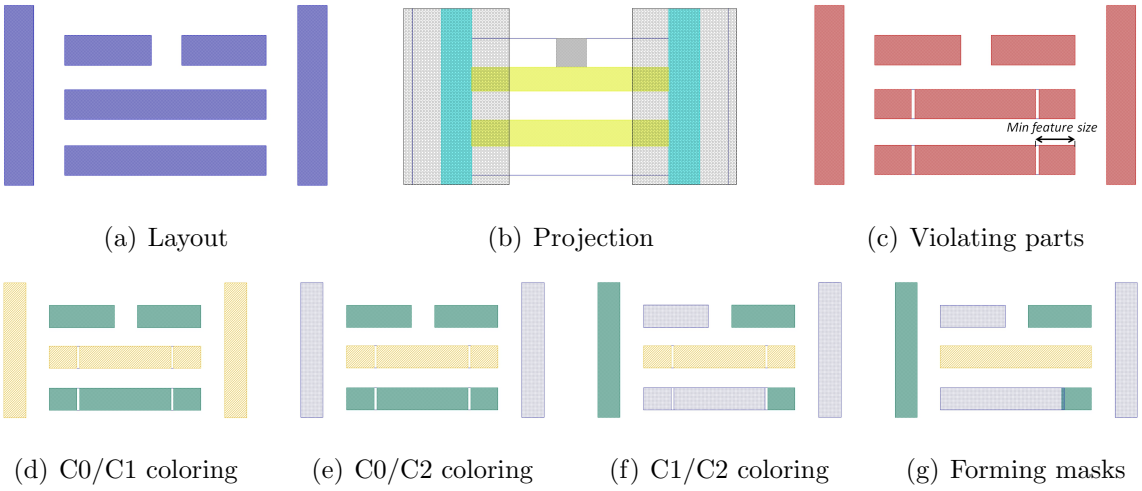
Figure 2.12: (a) Example layout coloring with two colors showing conflicts, (b) TP layout coloring with the ILP approach and the naive approach of using the third color to resolve conflicts, and (c) a conflict-free solution for TP coloring.

very long run-time. The work of [YYZ11] proposes an alternative solution, based on semidefinite programming, that is much faster than the ILP approach but leads to a larger number of coloring conflicts. The biggest limitation of the work of in [YYZ11] is that it uses candidate stitches that are only DP candidate stitches. In reality however, many patterns require the insertion of TP stitches for them to be colorable without violations as in the example of Figure 2.12)(b).

2.2.2 Overview of Our Approach

We offer a novel methodology for TP coloring that leverages TP stitching capability. Rather than simplifying the TP stitching problem by using DP candidate stitches only (as in previous works), the methodology considers additional TP candidate stitch locations to give coloring higher flexibility to resolve coloring conflicts. And, to deal with TP coloring complexity, the methodology employs multiple DP coloring steps, which leverages existing infrastructure developed for DP coloring. We make the following contributions:

- We propose a TP coloring method that uses existing infrastructure and algorithms developed for DP coloring. Our TP coloring method maximizes the use of stitching (among all three masks) to leverage TP stitching capability. The proposed method is also scalable and can be used to perform layout coloring for multiple patterning with k-colors (with k being greater than or equal to



(a) Layout (b) Projection (c) Violating parts

(d) C0/C1 coloring (e) C0/C2 coloring (f) C1/C2 coloring (g) Forming masks

Figure 2.13: Example illustrating the application the different steps involved in our triple-patterning coloring methodology. In (b), T2S violations are expanded and the parts of the target layout that are covered after the expansion are the T2S violating parts. T2T and S2S violating parts are identified similarly.

3) and can handle different values for S2S, T2S, and T2T same-color spacing rules.

- We test the method on different bidirectional layouts styles for 45nm, 32nm, 22nm, and 14nm technologies and report preliminary results.

2.2.3 Description of the Proposed Methodology for TP Coloring

Our TP coloring method uses multiple steps of 2-coloring to achieve the decomposition of the layout into three colors. This allows the re-use of existing DP coloring methods that are already developed and have reached maturity. For the DP coloring, we follow the approach we described in Section 2.1. The application of the different steps involved in our coloring method is illustrated by the example of Figure 2.13 and the different steps of the methodology are depicted in Figure 2.14. We use design rule-dependent projection as illustrated in Figure 2.4 to find all parts of the layout that have DP spacing violations with neighboring features and, then, we assign these violating parts to the two colors. In this way, candidate stitch locations are automatically defined as non-violating parts touching two or more violating

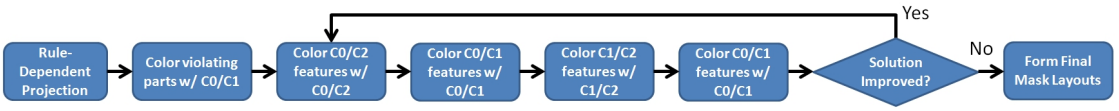


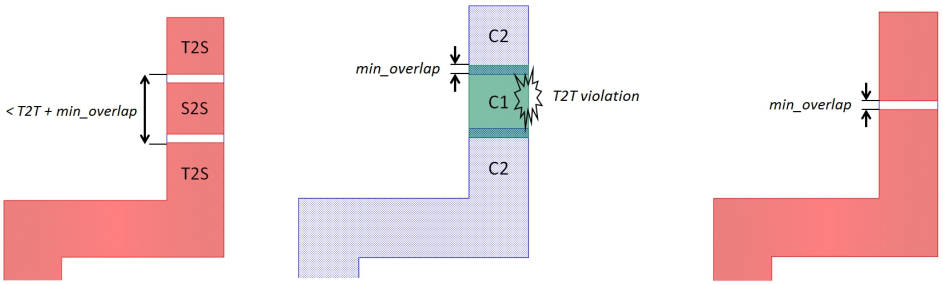
Figure 2.14: The flow for our the proposed TP coloring method that uses multiple steps of DP coloring.

parts. Stitches are then minimized by assigning these violating parts the same color. The TP coloring method involves a minimum of three DP coloring steps. First, we color the layout (violating parts only) with two colors C0 and C1. We then perform an additional coloring of the C0 layout using color C0 as well as the third color C2 to resolve conflicts on C0 (i.e. C0 to C0 spacing violations). We perform a final coloring step for the combination of C1 and C2 features and using C1/C2 colors. Although the method can be used with any number of 2-coloring steps and any order of color combination at each step, we adopt the five 2-coloring steps in the order shown in Figure 2.14. We start with C0/C1 coloring and, whenever a new coloring with the third color is performed, we follow it with a new C0/C1 coloring step effectively enhancing the resolution of conflicts in the initial C0 and C1 layouts. This 2-coloring cycle could be repeated a number of times as long as the solution quality is improved or until a satisfactory solution is achieved.

It is important to note that, because we use multiple steps of 2-coloring to achieve layout coloring for multiple patterning, our coloring method is scalable and can be used to perform layout coloring for multiple patterning with k-colors with any k greater than or equal to 3.

Leveraging TP Stitching Capability

Although our approach is improved over the naive approach in that it resolves conflicts on C0 and C1 in separate steps, it can still fail to achieve a good-quality solution in many cases. Consider the example of Figure 2.12 again. The only way to reach a conflict-free coloring solution as in Figure 2.12(c) is by introducing a TP candidate stitch. Because locating candidate stitches with projection is performed prior to the actual coloring, no candidate stitch locations are found in this layout



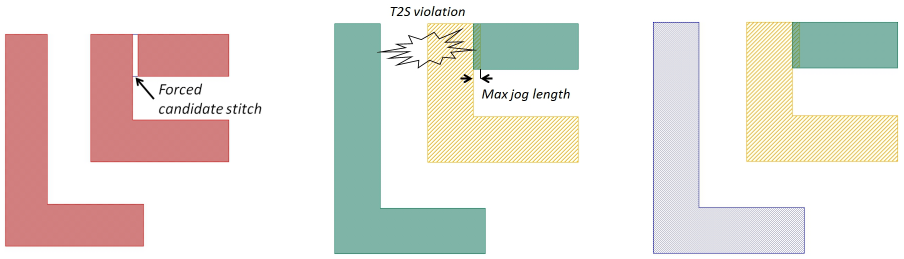
(a) (b) (c)

Figure 2.15: Illustrating example showing (a) candidate stitch at interface of S2S violating parts with other types of violating parts, (b) coloring problem that occurs with insertion of candidate stitch locations at interface, and (c) solving the problem by inserting a single stitch at center of S2S violating parts with length smaller than T2T (or T2S if larger) plus two times the minimum mask overlap length.

and coloring fails no matter what method is used. To avoid this problem, we perform projection for different spacing rules, T2T, T2S, and S2S, separately and introduce a candidate stitch location at the interface between purely S2S violating parts and the other types of violations (i.e. T2S and T2T violating parts). The intuition for inserting a stitch in S2S violating parts and not in other types of violations is that S2S violating parts are typically large and accommodate a split into multiple masks. Moreover, the only place where a candidate stitch is absolutely useless is when it is inserted at a location having violations with two other shapes that also have violations between each other; clearly, this occurs much less often in the case of S2S violating parts than in the case of T2S or T2T violating parts.

When a S2S violating part is short (smaller than the T2T same-color spacing rule plus two times the minimum mask overlap length) and is between two (or more) violating parts from the other types of violations, a coloring conflict may occur in the final solution as illustrated in Figure 2.15. To prevent such conflicts, we insert a single candidate stitch location in the center of the S2S violating part instead of two candidate stitch locations at the two interface regions.

For U-shapes violating parts (U-shapes on violating parts on the violating parts layer)where no stitches can be inserted with the method described earlier, we forcefully introduce a candidate stitch location at one of the U-shape segments, as illus-



(a) (b) (c)

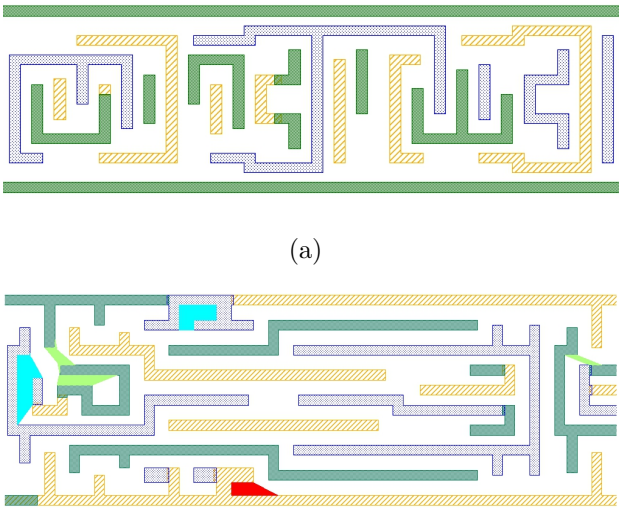
Figure 2.16: Illustrating example of the coloring of an U-shape violating part showing (a) forced candidate stitch insertion to separate the two segments of the U-shape, (b) the coloring problem that may occur once the coloring of violating parts and the mask-layouts are formed, and (c) fixing the coloring with an extra coloring step post coloring of violating parts.

trated in see Figure 2.16(a), so that the coloring tries to assign the two segments different colors. After the coloring of violating parts is complete and the mask-layouts are formed, a coloring violation may occur between one of the U-shape segments and the neighboring features as illustrated in Figure 2.16(b). This can occur because such violation cannot be detected prior to coloring with edge-based projection performed using existing DRC tools. To reduce the occurrence of such violations, we introduce an extra coloring step after the feature coloring is known with projection performed in each mask-layout separately.

2.2.4 Experimental Results

Projection and identification of violating parts were implemented in Calibre SVRF [Calb] and 2-coloring steps were performed using the DP coloring method described in Section 2.1.

We tested our methodology for TP coloring on the M1 and M2 layouts of standard cell libraries as well as complete designs. The layouts vary in size and complexity: from complex bidirectional M1 layouts at 45nm and 22nm to more regular bidirectional M2 layouts at 32nm and M1 layouts at 14nm. In all experiments, we use a pitch relaxation factor close to 2X, a minimum mask overlap length of 10nm, and a minimum feature size equal to the minimum line width. We use three TP coloring



(a)

(b)

Figure 2.17: Layout snippets of results of the proposed TP coloring methodology when applied to standard-cell library designs at (a) 14nm with regular bidirectional M1 and (b) 22nm with more complex bidirectional M1.

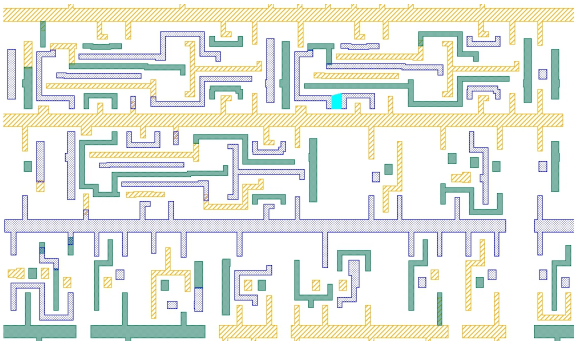
Table 2.4: Summary of results for the proposed TP coloring methodology when applied on library and full-design layouts at 45, 32, 22, and 14nm nodes.

Layout	**Cells**	**Pitch Relaxation**	**Conflicts**	**Stitches**
14nm Library	22	2X	0	33
22nm Library	108	1.9X	285	1,181
32nm M2 Design	26,000	2X	54	5,215
45nm M1 Design A	3,000	1.9X	177	842
45nm M1 Design B	10,000	1.9X	1,322	9,825

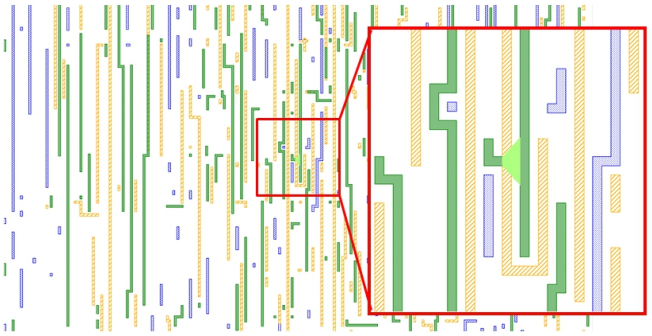
cycles of the flow in Figure 2.14, which correspond to a total of thirteen 2-coloring steps, to enhance the coloring results in terms of the number of coloring conflicts.

TP Coloring Results for 45, 32, 22, and 14nm Nodes

We tested the proposed coloring methodology on the M1 layer of two standard-cell libraries: the first is a 14nm library of 22 cells with bidirectional but simple layout and the second is a 22nm library with extremely dense and irregular patterns in all cells (to maximize M1 pin access). For the 14nm library, the method achieved a conflict-free coloring solution; for the 22nm library, however, the method resulted in



(a)



(b)

Figure 2.18: Layout snippets of results of the proposed TP layout coloring methodology when applied to full-design layouts at (a) 45nm with irregular bidirectional M1 and (b) 32nm with bidirectional but more regular M2.

a large number of TP coloring conflicts. The results are summarized in Table 2.4 and layout snippets showing the level of layout complexity are shown in Figure 2.17. Note that the results correspond to a single layout per library where cell-layouts are abutted to form the complete layout.

We also test the TP coloring methodology on full-design layouts at 45 and 32nm nodes. For testing at 45nm node, we decompose the M1 layer in two designs, one with 3,000 cell instances and the other with 10,000 cell instances. The M1 for both designs is highly irregular and are composed of cells with low as well as high density (unlike our 22nm test case where all cells were very dense). The results, highlighted in Table 2.4, show a reasonable number of conflicts for the smaller design suggesting that fixing the layout with manual incremental coloring or redesign would be possible for small designs (e.g., macros with $<$ 3,000 cells); for large designs, however, fixing

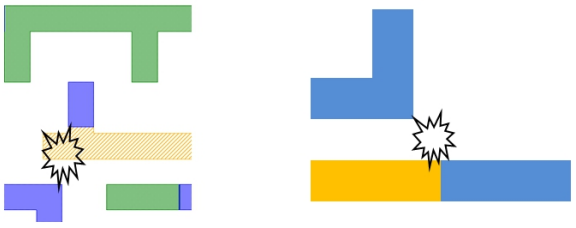


Figure 2.19: Coloring conflicts that may occur due to newly created tips.

Table 2.5: Results comparison between TP and DP coloring.

Layout	**TP Conflicts**	**DP Conflicts**	**TP Stitches**	**DP Stitches**
14nm Library	0	39	33	8
22nm Library	285	1,100	1,181	520
32nm M2 Design	54	3,272	5,215	1,712
45nm M1 Design A	177	4,524	842	1,310
45nm M1 Design B	1,322	27,073	9,825	8,044

the layout seems intractable. It is important to note that many of the coloring conflicts are repeated in certain problematic cells; so, fixing the layouts of such cells can greatly reduce the overall number of conflicts. For testing at 32nm node, we perform the layout coloring for the M2 layer (i.e. first routing layer) of a design with 26,000 cell-instances. Even though the design is fairly large, the coloring resulted in 54 conflicts only. Fixing such a limited number of conflicts is tractable with rip-up and re-route or manual incremental coloring or redesign. Figure 2.18 shows the level of complexity of the 45 and 32nm layouts as well as an example coloring conflict in each case.

When inspecting the final colored layouts, we observed that a considerable portion of the conflicts were caused by newly created tips at stitch locations once the final mask-layouts are formed as illustrated in Figure 2.19. Although this problem occurs in DP coloring when the T2T/T2S rules are larger than the S2S rule, it occurs more frequently in TP coloring because of the use of three masks and the problem may occur even when the rules have the same value because of the stitching at S2S violating parts. In future work, we plan to address this issue by modifying the exact locations of candidate stitches to avoid such conflicts.

Table 2.6: Summary of results for the proposed TP coloring methodology when applied on library and full-design layouts at 45, 32, and 22nm nodes with pushed T2T and T2S same-color rules from 3X to 2X the minimum spacing.

Layout	Initial Conflicts	Conflicts w/ Pushed Rules	Initial Stitches	Stitches w/ Pushed Rules
22nm Library	285	207	1,181	1,143
32nm M2 Design	54	50	5,215	5,566
45nm M1 Design A	177	92	842	809
45nm M1 Design B	1,322	465	9,825	8,918

Comparison with DP Coloring

We compare the TP coloring results to DP coloring for all test cases. As Table 2.5 shows, the number of conflicts with TP coloring is one to two orders of magnitude smaller than with DP coloring. The number of stitches is considerably larger in TP coloring for most cases, however. This is attributed to the facts that our TP coloring methodology makes use of TP stitching capability and that the methodology does not minimize the number of stitches globally between different coloring steps (while the DP coloring minimizes the stitching in the single coloring step).

Pushing T2S and T2T Same-Color Rules

The results in Table 2.4 are for the case where the T2T and T2S same-color spacing rules are equal to the S2S same-color spacing, which is roughly 3X the minimum spacing in the layout. T2T and T2S could be possibly pushed, however. Therefore, we repeat the experiments with pushed T2T and T2S same-color rules to 2X the minimum spacing. The results, reported in Table 2.6, show that pushing the T2T and T2S same-color rules greatly reduces the number of TP coloring conflicts but does not solve the problem entirely.

Effects of Using the Proposed TP Stitching Method

To quantify the impact of using the proposed TP stitching method (i.e., candidate stitch insertion in S2S violating parts), we compare the results with the case when

Table 2.7: TP coloring results with and without employing the TP stitching method (i.e., insertion of candidate stitches in S2S violating parts).

Layout	Initial Conflicts	Conflicts w/o TP stitching method	Initial Stitches	Stitches w/o TP stitching method
14nm Library	0	0	33	0
22nm Library	285	410	1,181	695
32nm M2 Design	54	66	5,215	3,135
45nm M1 Design A	177	366	842	548
45nm M1 Design B	1,322	3,144	9,825	4,044

Table 2.8: Run-time in real time of the proposed TP coloring methodology including projection, graph constructions, and all 2-coloring steps (twelve steps).

Layout	Cells	Pitch Relaxation	Run-time
14nm Library	22	2X	1min 30s
22nm Library	108	1.9X	2min 20s
32nm M2 Design	26,000	2X	7min 10s
45nm M1 Design A	3,000	1.9X	4min 16s
45nm M1 Design B	10,000	1.9X	28min

this method is not used and candidate stitches for TP coloring are candidate stitches for DP. The results summarized in Table 2.7 indicate that the TP stitching method is effective in reducing the number of coloring conflicts (except for the 14nm layout where conflict-free solutions are achieved in both cases).

Run-time Results

Even though the proposed 3-coloring methodology involves several 2-coloring steps (twelve 2-coloring steps), the run-time is not a concern. Table 2.8 reports the run-time of the complete process of projection, graph constructions, and all 2-coloring steps for all test cases.

Final Note on Results

The number of coloring conflicts achieved with TP is greatly reduced compared to the case of DP and TP may allow conflict-free coloring for bidirectional routing layers as

well as irregular M1 layer in small layouts. For extremely dense and irregular layers in large designs, however, legal coloring is unlikely to be possible even when TP is used and, hence, some layout simplification, regularization, and/or legalization may still be necessary.

2.3 Layout Legalization for MP

Whether DP or TP are used, achieving a conflict-free coloring solution for dense bidirectional layers is unlikely as our results and results from previous works suggest. Moreover, the number of conflicts is typically large making manual layout re-design practically impossible. Hence, automated conflict-removal for MP is essential.

In this section, we review previous works on conflict-removal and present the details of our approach. Previous works address conflict-removal for DP only. Our layout legalization naturally applies for DP, TP/MP, and SADP. We focus on the application of the methodology for DP and show how it can be extended for TP and SADP.

2.3.1 Prior Art in DP Conflict-Removal

Prior art in layout perturbation to resolve DP conflicts [HCN09, YP09, CC09] generally formulates the problem as an ILP (except [CC09]). Moreover, all previous works segment the layout into rectangles and move rectangles around to eliminate conflicts.

Working with rectangles has the same drawback discussed earlier and some additional drawbacks. The problem is further complicated because the automated layout perturbation solver (ILP or compaction) needs to maintain the connectivity of rectangles at joints (e.g., L-shape) through additional constraints. Moreover, because the constraints of the solver are defined between rectangles, overlap rules with features from the top and bottom layers cannot be handled correctly. Consider again the example of Figure 2.3 where an L-shape metal overlaps with a via (or contact)

at the corner. If the via movement is blocked, the solver will try to move shapes A and B so that *each* covers the via *completely*. Not only these moves are unnecessary because the via is initially covered, but they can also impact the layout area and the effectiveness of the conflict removal.

In [HCN09], DP requirements are added to the ILP constraints to perform DP-aware layout migration while minimizing area and layout perturbation. In addition to the problems with segmentation, the method leads to unmanageable number of constraints, excessive runtime to solve the ILP, and does not work well when the layout contains DP conflicts initially (i.e. not migrated from a previous generation).

In [YP09], wire spreading is proposed to remove DP conflicts. All wire-spreading options that reduce DP conflicts are pre-computed and conflicts and wire moves are minimized in the ILP. In addition to the problems common to all prior works that are discussed earlier, wire spreading can reduce the number of conflicts by a modest amount (as the results in [YP09] show). Many conflicts can be resolved by edge-location adjustment and wire-width reduction but not with wire spreading. Moreover, to avoid creating new DP conflicts, the method only moves segments when their spacing from all neighboring wires after the movement is at least equal to the same-color spacing. In many actual cases however, we may be able to move the segment to a closer distance from its neighbors – equal to the different-color spacing (typically half the same-color spacing) – and still avoid creating new conflicts[5]. The method of [YP09] cannot detect such cases and unnecessarily limits the wire spreading because, otherwise, the entire graph will have to be checked for newly created conflicts for every wire-spreading option.

Rather than solving the problem with an ILP, the work in [CC09] applies traditional layout compaction – based on minimum-area metric – iteratively as long as DP conflicts are reduced. At each iteration, the process of DP-compliance checking, which includes pattern projection [KPX10], segmentation, conflict graph generation, and odd cycle detection, is performed initially. DP constraints at odd cycles only are then generated and a trial compaction is performed. The DP-compliance check

[5]When the segment is assigned a different color than its neighbors.

is repeated and, if the number of odd cycles is reduced, the DP-constraints are permanently committed. In addition to the problems associated with segmentation into rectangles and the large runtime of iterative compaction and performing the DP-compliance check twice at each iteration, the method is not effective in removing DP conflicts and keeps a large number of conflicts unresolved (as reported in [CC09]). Because DP constraints are generated only at odd cycles, resolving one conflict may create a new conflict in other parts of the layout. As a result, the iterative compaction may stop without removing many DP conflicts that otherwise could have been resolved. In our work, we were able to remove DP conflicts efficiently, effectively, and simultaneously across all layers. This was made possible by essentially defining DP constraints all over the layout in terms of DRs – after an initial coloring that minimizes the number of conflicts – and applying linear programming-based layout compaction once across all layers.

2.3.2 Our Approach – Legalization with Minimum Layout Perturbation

After DP coloring with the least possible number of conflicts is complete (using our coloring methods or any other method), our objective is to make the layout compatible with DP and resolve the conflicts while minimizing layout perturbation.

Layout legalization is then performed using the method proposed in [HCT97]. The problem is formulated as a linear program (LP) with *minimum perturbation* as the objective, unlike [CC09] that uses minimum-area metric for compaction[6]. The layout is represented as a constraint graph, where nodes correspond to the layout edges and arcs correspond to the DRs that need to be met between any two layout edges. Arcs are assigned weights that correspond to the values of rules as illustrated in Figure 2.20. Layer-to-layer connectivity is maintained through the DRs between the layers, which are represented in the graph by arcs between nodes of the different layers.

The two mask-layouts of any double-patterned layer are defined as stand-alone

[6]The advantages of minimum layout-perturbation metric over the minimum area metric for layout compaction are discussed in [HCT97, ZFT05].

layers. Same-color spacing rules, between features of the same mask, including S2S, T2S, and T2T are mapped into arcs between the nodes of the stand-alone mask layer in the constraint graph. DRs that define the interaction between the two mask-layouts (e.g., minimum overlap length) are mapped into arcs between the nodes of the two stand-alone mask layers. For the interactions across different layers in the stack (e.g., M1 and contacts), we define any double-patterned layer as the union of its two mask-layouts and map across-layers DRs into arcs between nodes of the union layers[7].

As in layout compaction, the two-dimensional minimum perturbation problem is simplified by solving the one-dimensional problem successively (in x and y directions). It is important to note that the order in which the two-dimensional problem is solved, i.e. x or y direction first, can give different results; in our experiments, we solve the problem in both possible orders and choose the best solution. The 1D minimum perturbation problem is formulated as a LP as follows.

$$\begin{aligned} Minimize \quad & \sum_i W_i |X_i - X_i^{init}| \\ Subject\ to: \quad & X_j - X_i \geq d_{ij}, \forall A_{ij} \end{aligned}$$

where X_i and X_i^{init} are the current location and the initial location of node i and W_i is the weight for the perturbation of node i from its initial location. W is normally assigned a value of 1. It can be assigned a larger value to penalize the movement of edges that are less desirable (e.g., edges near the cell boundary) or prevent edges at certain layers from moving (e.g., diffusion/poly layers). A_{ij} is the arc between nodes i and j, which represents the DR constraint between the two layout elements, and d_{ij} is the weight of arc A_{ij}, which represent the value of the DR.

Figure 2.20 shows the construction of the constraint graph and the definition of constraints in the x-direction for an example double-patterned layout.

We obtain an equivalent formulation to the original problem with a linear objective function by introducing two new variables L and R for each node i as follows

[7]Rather than using layers of the mask-layouts and have the same problem highlighted in Figure 2.3.

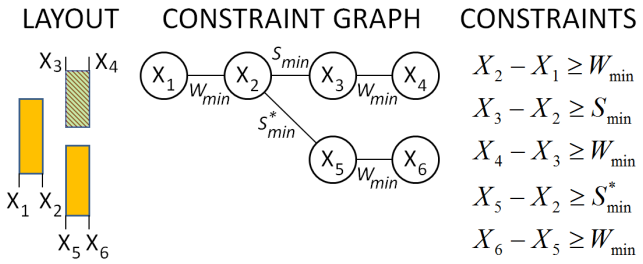


Figure 2.20: Example of x-direction constraint graph construction and constraint definition for a double-patterned layer. W_{min} is the minimum width rule, S_{min} is the S2S different-color spacing rule, and S^*_{min} is the S2S same-color spacing rule.

(details in [HCT97]):

$$\begin{aligned} Minimize \quad & \sum_i W_i(R_i - L_i) \\ Subject\ to: \quad & X_j - X_i \geq d_{ij} \quad \forall A_{ij} \\ & L_i \leq X_i, L_i \leq X_i^{init} \quad \forall i \\ & R_i \geq X_i, R_i \geq X_i^{init} \quad \forall i. \end{aligned}$$

This formulation permits the application of the method for practical layouts that use a discrete manufacturing grid for the coordinates. According to the total unimodularity property [Cul96], when all X_i^{init} and d_{ij} are integers, the solution of the problem consists of integers only. The handling of gridded design rule constraints can be achieved as in [YMH05]. The target on-grid locations are determined and on-grid constraints are relaxed to spacing constraints between the target locations and the cell boundary. After this relaxation, the problem is still formulated and solved as a linear program as detailed in [YMH05].

To handle infeasible constraints, we relax the unsatisfied arc constraints such that all constraints are feasible and a penalty is added in the objective function for the originally infeasible constraint. Section 2.3.3 gives more details about the handling of infeasible constraints and in-depth details can be found in [HCT97].

Our formulation of the problem maintains all inter and intra layer connectivity, which are represented as constraints in the graph. Internal connectivity of double-patterned layers at stitches is maintained through the minimum overlap length constraint. The conflict-removal problem is solved globally for the entire layout using

a single LP with constraints on all layout layers. As a results, our method permits the removal of DP conflicts across all layout layers simultaneously. And, because we formulate the problem as a minimum perturbation problem, our method ensures that the removal of a conflict in one part of the layout will not create a new conflict or design-rule violation in another part of the layout.

The way we formulate the conflict-removal problem permits the application of our methodology for layout legalization for TP and SADP. To apply the methodology for TP, TP coloring is performed instead of DP coloring and the three mask-layouts of any triple-patterned layer are treated as three stand-alone layers. All TP rules that define the interaction between these three mask-layouts (i.e., spacing and overlap rules) are mapped into constraints between the stand-alone layers. And, rules that define the interactions between the triple-patterned layer as a whole and the top/bottom-level layers (e.g., contacts/VIA layers) are mapped into constraints between edges of the union of the three mask-layouts and the edges of the top/bottom-level layers. In a similar fashion, the methodology can be applied for SADP; all that is needed is a SADP-coloring method as [BML11] and a set of design rules to ensure SADP compatibility of the layout as in [MSY12].

2.3.3 Layout Simplification for More Efficient DP Conflict Removal

In actual layouts, we observe that some DP conflicts can be avoided by simple notch removal prior to coloring. In addition, many conflicts on the M1 layer are caused by segments that are added to cover redundant contacts/vias or to maximize the pin-access region. Redundant contacts and vias improve manufacturability, but they are *not absolutely required*. The same is true for pin segments that extend beyond the minimum requirement to ensure pin-accessibility. The addition of these extra segments is considered a good layout practice to maximize the pin-access region and, consequently, improve the routing efficiency. We take advantage of these observations and, *as an option*, we perform notch filling and allow the *possible* sacrifice of redundancy and extra pin segments to improve the results of the DP

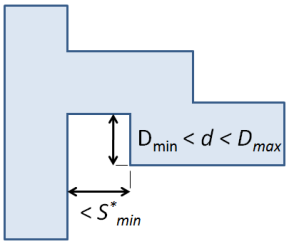


Figure 2.21: Characteristics of notches that will be removed. S^*_{min} is the side-to-side same-color spacing rule, D_{min} is the notch-depth below which the notch is manufacturable with a single exposure, and D_{max} is the depth beyond which notch-filling is not performed to avoid creating fat wires with larger spacing requirements.

conflict removal framework. Specifically, we pre-process the layout prior to the coloring to mark potential sacrificial features. During the legalization and conflict-removal, these features are recovered whenever possible without creating any new violations.

Small-Notch Removal

Small notches, or small-depth U-shapes, with depth less than a certain value, D_{min}, may be manufactured with a single exposure. Deeper notches, however, require the two segments of the notch to be assigned to different exposures (i.e. colors). In some layouts, assigning the two segments of a notch different colors is not possible without creating a coloring violation and, in such case, the notch contributes to the number of DP native conflicts. The removal of DP conflicts caused by notches may not be possible during layout legalization when the layout area is fixed and may lead to an area overhead when the layout area is allowed to increase for legalization. As a result, getting rid of notches that cannot be manufactured in a single exposure prior to coloring is a good practice and makes layout legalization for DP more efficient.

An effective way to remove small notches is by joining their two segments so as to fill the notches. Filling a notch requires little layout modifications: it does not create extra color violations and adding extra material (i.e. metal) does not affect other

[8]Note if these features are necessary to meet a certain requirement on yield/routability score, it would be possible to mark a fraction of these features for possible sacrifice or simply avoid this optional step altogether.

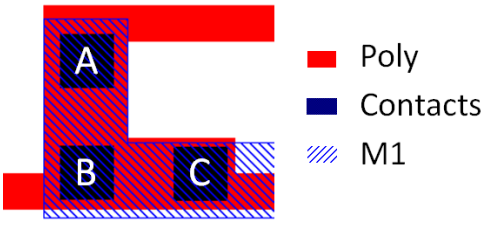


Figure 2.22: Group of redundant contacts connecting to the Poly layer. Contact B has more flexibility of movement than contacts A and C and, thus, we pick B as the required contact and A and C as redundant contacts that may be sacrificed if necessary to resolve conflicts.

layers. To avoid creating "fat wires" that have peculiar spacing requirements with their neighboring features, we fill notches prior to coloring only when the depth of the notch is smaller than a specified value, D_{max}. Figure 2.21 depicts the characteristics of notches that will be filled in our DP-enablement framework.

Sacrifice of Redundant Contacts/Vias When Necessary

The process of identifying redundant vias is similar to the process of identifying redundant contacts and, for brevity, we only describe the latter process. We start by finding overlap regions of the top layer (M1) and the bottom layer (Polysilicon or active). If a single polygon of the overlap region interacts with two or more contacts, these contacts are identified as a group of redundant contacts. Next, we choose one of the contacts from each group to be a required contact/via and add all such required contacts to single contacts to form a new layer of required contacts. The remaining contacts that were not chosen as required contacts are considered redundant. The choice of the required contact among a group is made with preference to the contact with the highest flexibility of movement as illustrated in Figure 2.22. Contacts that were considered redundant are assigned to a new layer.

If M1 is double patterned, the line-end part of M1 that covers a redundant contact is removed, as shown in Figure 2.23, and overlapping redundant contacts with M1 is specified as a recommended, but not required, constraint. The LP of the conflict removal method will meet this recommended constraint only when possible without creating a DP conflict or any DR violations. In other words, redundant contacts

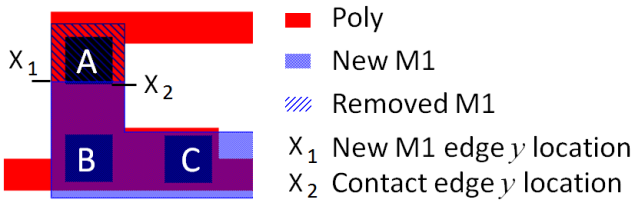


Figure 2.23: Illustration of M1 simplification for possible sacrifice of contacts.

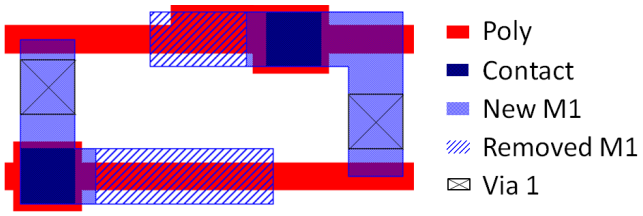


Figure 2.24: Illustration of M1 simplification for possible sacrifice of pin segments.

will be sacrificed only when necessary to resolve conflicts. To ensure recommended contacts still get a chance to be covered by M1 after the layout is perturbed, we add a required constraint to keep redundant contacts at the same spacing and aligned to the corresponding required contact chosen among the group of redundant contacts.

Sacrifice of Pin Segments When Necessary

M1 pin segments that do not connect to any other layer in the layout stack are removed for possible sacrifice as shown in Figure 2.24. To allow the layout perturbation to recover the removed parts when possible without creating violations, the original M1 layer is kept and a recommended constraint is added to the LP problem to minimize the distance between the new M1 edge and the original M1 edge.

The removal of M1 pin segments and M1 parts that cover redundant contacts/vias is performed before the DP coloring. This way, because violations are reduced, extra candidate stitches can be identified and taken advantage of to reduce DP conflicts. When the sacrifice is not necessary to resolve conflicts, these extra stitches will be removed by the coloring algorithm (by coloring violating parts of a stitch with the same color) and the layout perturbation will recover the sacrificed parts as described earlier.

Handling Recommended Constraints During Legalization

Recommended constraints are handled in the LP formulation of the conflict removal framework in a similar way as infeasible constraints are handled, i.e. by introducing a new variable to relax the constraint and minimizing this relaxation variable in the objective function. This is illustrated through the example of Figure 2.23 where M1 covering the redundant contact A is set as a recommended constraint. Here, M1 is shrunk until it slightly overlaps with the redundant contact A. X_1 is the location of the M1 edge overlapping the redundant contact A and X_2 is the location of the bottom edge of the redundant contact (underneath M1). The constraint is then $X_2 - X_1 + r_{12} =$ *contact width rule + M1 overlap past contact rule.* r_{12} is included in the objective function so that it is minimized. The minimization of relaxation variables for recommended constraints is given less priority than the minimization of the relaxation variables of infeasible required constraints (by assigning a smaller weight in the objective function). This way, recommended constraints are met only when possible without creating any DP conflicts, DR violation, or area increase.

It is worth noting that the minimization of the relaxation variables can be weighted according to the importance of what is being sacrificed (e.g., pin-access metric such as in [Tag10]).

2.3.4 DP-Compatible Design

Our DP coloring and legalization framework primarily targets standard-cell-based designs. The framework can also be used, however, for small full-custom macro designs as we demonstrate later in the paper in Section 2.3.5.

To create a DP-compatible standard-cell-based design, the framework is used to build a standard-cell library. The design is then synthesized using the new library and placement and routing are carried out to create the layout of the entire design. This method preserves the abstraction and common practice in the design of state-of-the-art systems, where standard-cell libraries are commonly developed separately from the physical design and must meet all manufacturing constraints before their

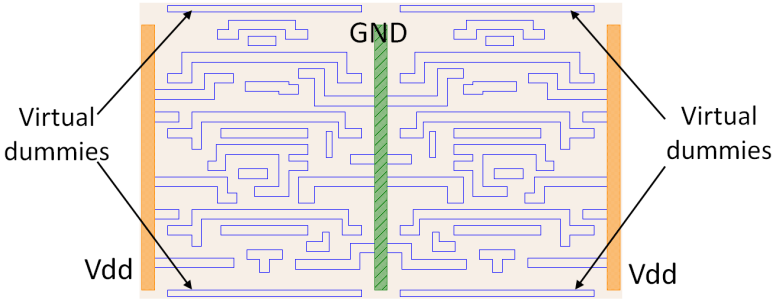


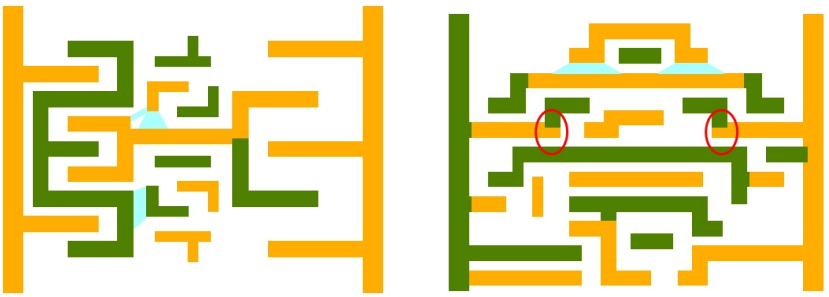
Figure 2.25: Handling cell-boundary conflicts during placement (virtual dummies not on mask).

release.

With our method, DP conflict-free cells are achieved and they are designed so that no DP conflicts can occur at cell boundaries after placement. Consider the example where M1, which is usually the most complex and dense layer, is double-patterned. The coloring of the Vdd and GND power rails is fixed in all cells as shown in Figure 2.25. During placement, cells are possibly flipped with respect to the vertical axis so that cells in the same column have Vdd and GND on the same sides (with fixed coloring) and cells of one column and the next/previous columns share the same power rails (as shown in Figure 2.25). DP conflicts between cells of the same column are prevented as follows. Two dummy M1 wires are added before the coloring process at the cell edges (as in Figure 2.25) so that all features at the cell sides are assigned the same color (similar to [HCN11]). These dummies are removed after the conflict removal flow is complete and they do not appear on the masks.

Two versions of each cell are provided, one with the initial coloring (after conflict removal) and another with flipped coloring (excluding the power rails coloring) that guarantees no conflict with power rails. If the placement of two cells in the same column results in a DP conflict at the cell boundary, we simply use the version of one of the cells where the coloring is flipped.

Timing differences between the coloring-versions of the same cell may occur due to Critical Dimension (CD) variation of the different exposures and different overlay impacts. Such timing variations are expected to be insignificant, however, since



(a) (b)

Figure 2.26: Violation count based on shape movement requirement: (a) layout example with three spacing violations counted as two conflicts only and (b) layout example with two spacing violations between the same two polygons counted as two conflicts. The highlighted regions correspond to possible issues with newly created tips at stitch locations (this will be discussed later).

intra-cell M1 wires are naturally short [GTG11]. Consequently, timing analysis needs not be aware of the exact coloring of cell-instances. If for some types of designs timing analysis is required to model such effects, the two versions of the cell can be dealt with as completely separate cells and their timing can be characterized separately. This is not desirable as it effectively doubles the number of cells in the library and it should only be used for DP-aware timing analysis with high-accuracy requirement.

2.3.5 Experimental Setup and Results

The conflict-removal method was implemented and integrated into the minimum perturbation-based VLSI artwork legalization system [HCT97].

Reporting DP Conflicts

We verify post-coloring DP conflicts (or DP violations) by running design rule check (DRC) on the colored layout using Calibre nmDRC. Same-color spacing violations between features of the same mask-layout of a double-patterned are effectively DP coloring conflicts. In this case, multiple violations may exist between any two polygons.

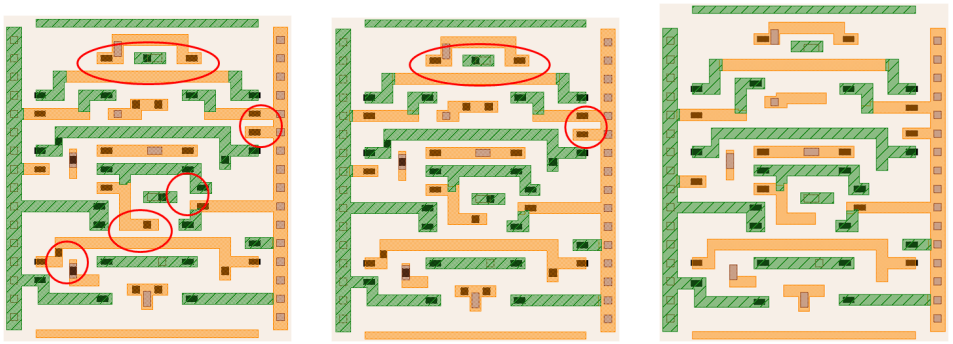
The natural way of counting post-coloring DP conflicts is to count every same-color spacing violation as a DP conflict. Such count is suitable for evaluating the DP coloring; it is not a good metric, however, for quantifying the effectiveness of DP conflict-removal framework. Consider the example of Figure 2.26(a). If every spacing violation (every aqua-blue polygon in the figure) is counted as a DP conflict, the layout would have three conflicts. The conflict-removal framework will only need to fix two conflicts, however, since fixing one of the two spacing violations on the top will most likely fix the other automatically (by moving the bottom edge of the top layout polygon up). Yet, not all spacing violations between the same two layout polygons can be treated as a single violation. In the case of Figure 2.26(b), fixing the two violations will certainly require the movement of two edges (the bottom edges of the top polygon up). Therefore, for better accuracy in reporting the conflict-removal results, we inspect the spacing violations visually to determine the number of conflicts they correspond to.

DP Conflict-Removal and Legalization Results

Our DP conflict removal framework was tested on a commercial 22nm standard-cell and macro layouts. We assume M1 is double patterned and apply the conflict removal method for layouts that have DP conflicts. The M1 minimum spacing in the layout is 40nm and we use a value of 15nm for the minimum overlap length and 80nm for the S2S, T2S, and T2T same-color spacing[9].

In one experiment, we apply our DP conflict removal method to standard cells. The results show that DP conflicts in many cells were completely removed without area increase and any DR violations. For some other cells, few DP conflicts remain unresolvable when the area is fixed. We give two options to deal with such stubborn conflicts. The first option is to keep these conflicts and report their locations so that the layout designer fixes them manually. The second option is to run the

[9]Because DP was assumed for M1 in the process and to avoid making inadequate assumptions on the differences between the spacing values that we cannot justify, we use the same value for the different rules in the experiments. Nevertheless, the benefits of handling multiple same-color spacing rule values were shown through the DP coloring results.



(a) (b) (c)

Figure 2.27: Sample results for a cell layout: (a) before DP conflict removal, (b) after conflict removal with fixed area, and (c) after conflict removal with area increase.

conflict removal framework a second run with non-fixed area so that all conflicts are removed. Figure 2.27 shows an example layout where M1 is double-patterned before and after the layout perturbation to remove conflicts. As Figure 2.27(b) depicts, the conflict removal method with fixed area is able to remove three out of the four conflicts in the original layout of Figure 2.27(a). The remaining stubborn conflict is removed when Poly and active are allowed to move and area is allowed to increase as shown in Figure 2.27(c). In this case, the restrictive DR of Poly on grid are met by modifying the LP program as described in [YMH05]. A summary of the results is given in Table 2.9. For all cells, the runtime for the entire conflict removal flow (coloring plus conflict removal) is less than 10 seconds in real time. In five out of nine cells, all DP conflicts were removed without any area increase or the removal of redundant contacts. In the remaining four cells, few DP conflicts remain after applying our method with fixed area. When we allow the layout area to increase, all conflicts are removed in these cells with an average 6% area overhead (at most 9.1% overhead) and with the sacrifice of a single redundant contacts in just one of the cells.

In another experiment, we apply our DP conflict removal method for two macro layouts, two local clock buffer controllers that consists of multiple latches and inverters with roughly 82 transistors for the first macro and 460 transistors for the

Table 2.9: Results of applying our DP conflict removal method with and without area increase to cells from a commercial 22nm library (area values are normalized, "CF" stands for conflicts, "SRCA" stands for sacrificed redundant contacts, and "Area Inc." stands for area increase).

	Original		**Conflict Removal w/o Area Increase**		**Conflict Removal w/ Area Increase**		
Cell/Macro Layouts	**Area**	**CF**	**CF**	**SRCA**	**CF**	**SRCA**	**Area Inc.**
LCB + latch 1	1	**1**	**0**	0	-	-	-
latch1	1.6	**3**	**2**	0	0	0	9.1%
oai	1.6	**2**	**0**	1	-	-	-
scan latch	2.3	**5**	**3**	0	**0**	0	6.2%
xor	2.4	**2**	**0**	0	-	-	-
latch2	4.3	**19**	**8**	0	**0**	0	3.3%
nand4	4.7	**4**	**0**	0	-	-	-
latch3	5.3	**4**	**3**	0	**0**	0	5.4%
nand3	6.7	**7**	**0**	0	-	-	-
LCB control. 1	13.7	**13**	**7**	4	**0**	4	8.3%
LCB control. 2	50.3	**53**	**31**	1	**0**	2	9.1%

second. The results are given in the bottom two rows of Table 2.9. The method reduces the number of DP conflicts from 13 to 7 for the first macro and from 53 to 31 for the second without increasing the layout area. When the area is allowed to increase, the method removes all remaining conflicts with an average area increase of 8.7% and a total of six sacrificed redundant contacts. The runtime of the entire flow for the largest macro layout is less than one minute in real time ($<$ 2 seconds CPU time).

Effects of Preferred Coloring and Layout Simplification

The use of the preferred coloring method and the possible sacrifice of non-crucial layout features including redundant contacts and pin segments makes the conflict removal more effective. To quantify the impact of these two methods, we run our framework with fixed area and while enabling or disabling the two methods. The number of conflicts results for the different cases are reported in Figure 2.28 and

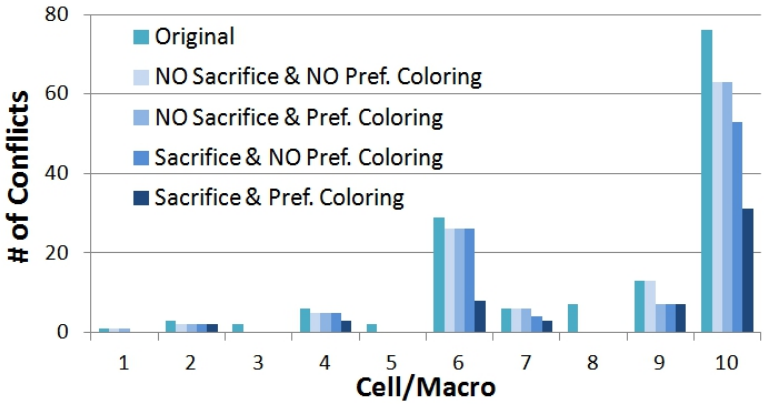


Figure 2.28: Number of conflicts with the fixed-area flow for the different cell and macro layouts showing the effects of using preferred coloring (see Figure 2.5) and the possible sacrifice of redundant contacts and M1 pin segments.

are compared to the original number of conflicts in the layout before applying the legalization framework and with non-preferred coloring. Up to ∼4X smaller number of conflicts can be achieved when the methods are applied for cell layouts and up to ∼2X smaller number of conflicts can be achieved when the methods are applied for macro layouts. It can also be clearly observed from the results that the conflict removal is effective only when both methods are applied at the same time.

A Coloring Problem Solved Automatically with Legalization

When the same-color T2T and/or T2S spacing rules are larger than the different-color S2S spacing rule plus the minimum width rule, unforeseen post-coloring DP violations may be introduced between newly created tips at stitch locations and the neighboring shapes as shown by the highlighted regions in Figure 2.26. This problem may occur with any coloring method that rely on edge-based DRC for checking for color violations and perform projection; it is the reason why we obtained different number of violations with the greedy-based and the MINCUT-based flipping methods in Table 2.1. Our post-coloring legalization approach is advantageous in handling this issue because compaction on already-colored layouts implicitly fixes these newly introduced DP violations.

Conclusions

We proposed a novel methodology to enable MP in the design. The methodology essentially consists of three steps: layout coloring, exposure layers and geometric rules definition, and, finally, layout legalization using compaction and multiple-patterning rules as constraints. For the layout coloring, we propose a $O(n)$ heuristic algorithm for DP that guarantees a conflict-free solution when one exists and extend it for TP, while leveraging TP stitching capability for enhanced the coloring efficiency. Our automated layout legalization for MP is performed across all layout layers simultaneously while minimizing perturbation using a LP. The method enables designing with conventional DRs and masks the designer from the complexity in dealing with DP layers and requirements. The way we formulate the problem allowed us to achieve high-quality results with extremely fast run-time (less than 10 seconds in real time for typical cells). The method targets primarily standard-cell layouts but it can also be applied for small full-custom layouts and interconnect layers in complete designs as we showed in the paper. Although we demonstrate the method on LELE DP, the method is more general and can be naturally extended for other MP technologies including triple/quadruple-patterning in multiple litho-etch steps process as well as SADP.

Acknowledgments

I would like to thank Kevin W. McCullen and Rani R. Narayan from IBM for their help with setting up the minimum perturbation based VLSI artwork legalization system [HCT97]. I would also like to thank Andrew B. Kahng from UCSD and Hailong Yao from Tsinghua University for providing layout test cases that allowed results comparison with their work.

Chapter 3

Overlay Impact for Design in Metal Double-Patterning

In double-patterning lithography (DP), overlay errors between two patterning steps of the same layer translate into critical-dimension (CD) variability. Since CD uniformity budget is very tight, meeting requirement of overlay control is one of the biggest challenges for deploying DP. In this chapter, we electrically evaluate overlay errors for back-end-of-line (BEOL) DP with the goal of studying relative effects of different overlay sources and interactions of overlay control with design parameters. Experimental results show the following: (a) the expected electrical impact of overlay in a path is not significant ($< 6\%$ worst-case RC variation) and should be the basis for determining overlay budget requirement; (b) the worst-case electrical impact of overlay in a single line remains a serious concern (up to 16.6% ΔRC and up to 50mV increase of peak crosstalk noise); (c) translational overlay error has the largest electrical impact compared to other overlay sources; and (d) overlay in y direction (x for horizontal metallization) has negligible electrical impact and, therefore, preferred routing direction should be taken into account for overlay sampling and alignment strategies. Design methods for reducing overlay electrical impact in wires are then identified. Finally, we explore positive/negative process options from an electrical perspective and conclude that positive process is preferred.

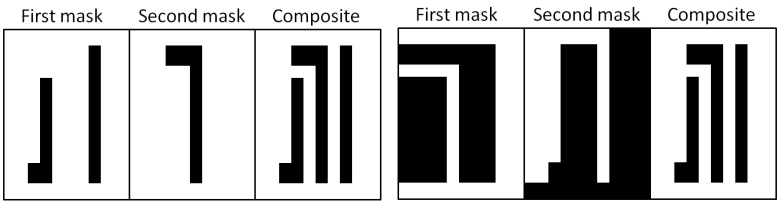
Introduction

Double-patterning lithography (DP) is one of the most likely short-term solutions for keeping the pace of scaling beyond 32nm node [Mac08]. DP consists of printing patterns of the same layer using two separate exposure steps and, thus, allows a smaller pitch between features.

Overlay is the positional accuracy with which a pattern is formed on top of an existing pattern on the wafer [Mac06]. In traditional single-exposure lithography, overlay errors occur between patterns of different layers. Design rules that define interactions between layers (e.g., metal overhang on via rule) make overlay errors less severe and reduce the requirements on overlay control (ITRS [ITRb] estimates the overlay budget in single-exposure lithography to just 20% of the minimum feature size). Because two separate exposures are involved in DP, overlay errors can also occur between patterns of the same layer. Such overlay errors effectively translate into CD variability [Arn08, Dus07], which changes the electrical characteristics of devices and wires. In this case, design rules cannot help reducing this variability problem and, because the CD budget is already very tight (estimated by ITRS to 7% of the minimum feature size), overlay must be very well-controlled. Meeting this requirement for overlay control is seen as one of the biggest challenges for deploying DP technology [ADF06].

In positive dual-line process (with positive photoresist) where the line is the critical feature to be controlled (Figure 3.1(a)), overlay errors translate into metal spacing variation, which affects interconnect capacitance (C). On the other hand, in negative dual-trench process (again with positive photoresist) where the space is the critical feature to be controlled (Figure 3.1(b)), overlay errors translate into line width variation, as illustrated in Figure 3.2, with an impact on interconnect resistance (R) as well as capacitance.

The impact of within-layer overlay on the electrical characteristics of wires has been studied in literature. A method for estimating delay variation due to overlay errors is presented in [YP08]. A compact model to estimate interconnect delay vari-



(a) (b)

Figure 3.1: Example layout with positive DP process (a) and negative DP process (b) both using a positive-tone photoresist.

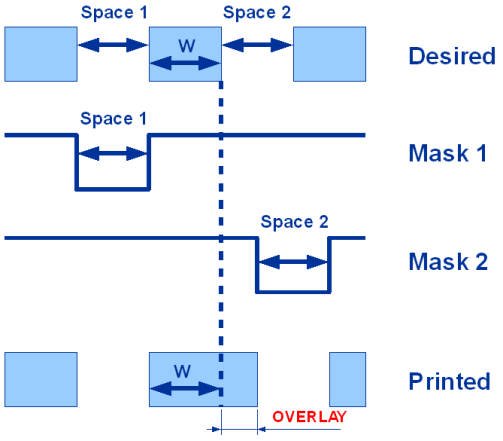


Figure 3.2: Example showing translation of overlay error into CD variation in negative DP.

ation due to overlay and focus variations in DP is offered in [CN09]. A systematic method to compare the effects of overlay to that of CD variability on interconnect delay variation is proposed in [JKT09]. In this work, we electrically evaluate overlay errors for BEOL DP to *study relative importance of different overlay sources and interactions of overlay control with design parameters* and derive methods to alleviate within-layer overlay problem in DP. In addition, we explore processing options including positive dual-line and negative dual-trench processes.

Models of overlay impact on electrical characteristics of wires in positive and negative DP are derived in Section 3.1. In Section 3.2, experimental methodology and results are presented. Observations are discussed and overlay implications on design are analyzed in Section 3.3. In Section 3.4, we explore processing options for next generation technology nodes.

3.1 Electrical Impact of Within-Layer Overlay

In BEOL process implemented with DP, overlay errors between two patterning steps at the same layer affect the electrical characteristics of wires. This section exhibits models for overlay and its electrical impact that are used in our experiments.

3.1.1 Overlay Modeling

Major overlay components are translation, magnification, and rotation in the wafer and field coordinate systems [LLD08, CC01] and are considered in a linear-type overlay model for overlay control and correction. High-order models can also be used to enhance overlay accuracy, but such models require more overlay sampling and excessive alignment [EHO08, Wak07, Lev05]. In our study, we adopt the following widely used linear model [CC01]:

$$\begin{aligned} \delta_x &= T_x + M_{wx} \times X_w - R_{wx} \times Y_w + M_{fx} \times X_f - R_{fx} \times Y_f + Res_x, \quad (3.1) \\ \delta_y &= T_y + M_{wy} \times Y_w + R_{wy} \times X_w + M_{fy} \times Y_f + R_{fy} \times X_f + Res_y, \end{aligned}$$

where δ_x (δ_y) is the total overlay error in the X (Y) direction. T, M, and R refer to translation, magnification, and rotation overlay parameters respectively. Res is the residual parameter, which accounts for un-modeled secondary overlay components such as skewness and trapezoidal overlay. w and f denote the wafer and field respectively. (X_w, Y_w) and (X_f, X_f) refer to Cartesian coordinates in the wafer and field respectively.

Even though the model's parameters are refined continuously during processing, *the model still does not correct for overlay error totally.* This imperfect correction has many reasons: field to field and wafer to wafer overlay variations, limited overlay sampling that does not cover the entire wafer and lot, and un-modeled secondary overlay components.

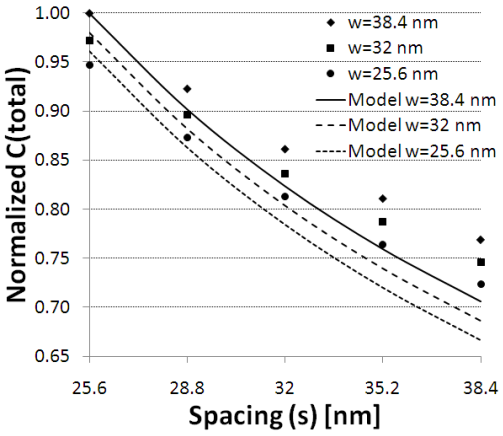


Figure 3.3: Parallel plate capacitor model compared to simulation results for interconnect capacitance with varying interconnect width and spacing. Symbols represent simulated capacitances using Raphael^{TM} [Rap] and lines represent the capacitance evaluated using the parallel plate model. Simulated and modeled capacitances are normalized with respect to their maximum values.

3.1.2 Capacitance Model

Interconnect capacitance can be evaluated using the parallel plate capacitor model or the more accurate models offered in [Sak93, CHA92]. In this work, we use the parallel plate model to derive simple and closed form equations for the impact of overlay in DP. In Figure 3.3, we compare the model to simulation results[1] for varying interconnect width and spacing. It is clear from the figure that the model exhibits a trend similar to simulated data, but it slightly overestimates the overlay impact on capacitance variation.

3.1.3 Electrical Impact in Positive DP

DP can be implemented in a positive process, which prints lines, or negative process, which prints spaces [Lim06, KWK07]. If positive process is implemented for BEOL, interconnect spacing (s) between the two patterns is affected leading to the change of interconnect line-to-line capacitance (C_{LL}).

We derive a closed form equation for C_{LL} between two parallel vertical lines of

[1]Simulations are performed on Raphael^{TM}, a capacitance simulation tool [Rap], for the structure of Figure 3.4(b) and the interconnect characteristics of Table 3.2.

length L, where one line is printed perfectly and the other is printed with overlay error. Using the parallel plate capacitance model, C_{LL} can be expressed as follows:

$$C_{LL,2l} = \epsilon t \int_0^L \frac{1}{s^\star} dl, \tag{3.2}$$

where ϵ is the dielectric constant, t is the interconnect thickness, and $s^\star$ is interconnect spacing with overlay error. Using the overlay model of Equation (3.1) and converting from wafer and field coordinate system to design coordinate system, $s^\star$ is determined by:

$$\begin{aligned} s^\star &= s - (T_x + M_{wx}X_o + M_x X_Q - R_{wx}Y_o - R_x Y_Q + Res_x) \\ &\quad -M_x x + R_x y + R_x L, \\ where \quad & R_x = R_{wx} + R_{fx}, \;\; M_x = M_{wx} + M_{fx}, \end{aligned} \tag{3.3}$$

where (X_o, Y_o) and (X_Q, X_Q) refer to the coordinates of field origin in the wafer plane and die origin in the field plane respectively, and (x, y) are the coordinates of the bottom left corner of the line of interest in the design plane. Consequently, the closed form equation of C_{LL} as a function of structure coordinates in the design is:

$$\begin{aligned} C_{LL,2l} &= \frac{\epsilon t}{R_x} \ln \frac{s - b - M_x x + R_x y + R_x L}{s - b - M_x x + R_x y}, \\ \text{where} \quad b &= T_x + M_{wx}X_o + M_x X_Q - R_{wx}Y_o - R_x Y_Q + Res_x. \end{aligned} \tag{3.4}$$

Line resistance is evaluated using

$$R = \frac{\rho L}{t(w - wM_x)}, \tag{3.5}$$

where ρ is the effective interconnect resistivity and w is the interconnect width. According to Equation (3.5), the effect on R is minor since wM_x is orders of magnitude less than w.

Assuming a single ground plane in the layer below, the capacitance between the line of interest and ground plane is evaluated using

$$C_{LG} = \frac{\epsilon L(w - wM_x)}{H}; \tag{3.6}$$

where H is the height of interlayer dielectric layer. From the equation, we note that the impact of overlay on C_{LG} is again minor.

Similar derivation is 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 overlay error. The closed form equation of C_{LL} in this case becomes

$$C_{LL,3l} = \frac{\epsilon t}{R_x}\left[\ln\frac{s-b-M_x x+R_x y+R_x L}{s-b-M_x x+R_x y} + \ln\frac{s+b+M_x x-R_x y+wM_x}{s+b+M_x x-R_x y+wM_x-R_x L}\right]. \tag{3.7}$$

3.1.4 Electrical Impact in Negative DP

In case of negative process, interconnect width (w) is affected, which causes interconnect resistance (R) and capacitance (C) variations.

Using the parallel plate capacitance model and overlay model of Equation (3.1), closed form equations for R and C are derived in a similar manner to the derivation of R and C equations for positive DP. Considering a structure of two parallel vertical lines where the line of interest is formed by printing one space perfectly and the other with overlay error, R of the line of interest is described by

$$R = \frac{\rho}{tR_x}\ln\frac{w-b-M_x x+R_x y-wM_x+R_x L}{w-b-M_x x+R_x y-wM_x}; \tag{3.8}$$

C_{LL} between the two lines is determined by

$$C_{LL,2l} = \frac{\epsilon t L}{s-sM_x}. \tag{3.9}$$

C_{LG} between the line of interest and ground plane of layer below is modeled by:

$$C_{LG} = \frac{\epsilon L}{2H}[2(w-b-M_x x-M_x w+R_x y)+R_x L]. \tag{3.10}$$

In Equations (3.8, 3.9, 3.10), b, R_x, and M_x are the same as in Equations (3.3, 3.4), ρ is the wire resistivity, and H is the height of inter-level metal dielectric.

For a structure of three parallel vertical lines, R and C_{LG} are calculated using the same equations as for the 2-line structure, i.e. Equations (3.8, 3.10), but another

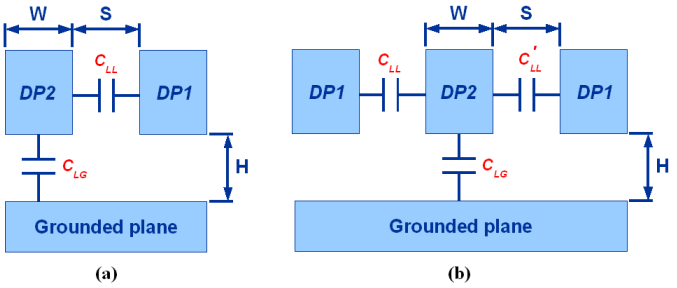


Figure 3.4: Test structures used in the experiments: (a) 2-line structure, and (b) 3-line structure with single grounded plane at the layer below.

equation is used for calculating C_{LL}. Assuming the space between first and second lines is printed perfectly while the space between second and third lines is printed with overlay error, C_{LL} is determined by

$$C_{LL,3l} = \epsilon t L\left(\frac{1}{s - sM_x} + \frac{1}{s}\right). \tag{3.11}$$

3.2 Experimental Methodology and Results

A series of experiments are conducted to evaluate the electrical impact of overlay in BEOL DP. This section describes experimental setup and methodology and presents the results.

3.2.1 Experimental Setup

A 300mm wafer with 63 33x26mm fields *each containing four copies of the same design* is considered. The study is performed for BEOL 32nm technology node (i.e. metal 1 half-pitch) at local interconnect levels with design rules adopted from ITRS [ITRb]. Interconnect length (L) is set to 100μm, which is close to maximum wire length for local interconnect levels where DP is likely to be implemented.

The test structures used in the experiments are the 2-line and 3-line structures depicted in Figure 3.4. In both structures, overlap capacitance (C_{LG}) is assumed to be between the line of interest and a single ground plane at the layer below. Also, lines of the first pattern are labeled as "DP1" and *are assumed to be formed perfectly*, while lines of the second pattern are labeled as "DP2" and *are printed with overlay*

Table 3.1: Estimated overlay breakdown used in the reference experiment.

	% of imperfect correction	**Exact value [nm]**
Translation	5.32%	0.34
Wafer magnification	14.18%	0.91
Field magnification	2.48%	0.16
Wafer rotation	25.53%	1.63
Field rotation	2.48%	0.16
Residual	50%	3.2

error. For the 2-line structure, total capacitance (C) of "DP2" wire is given by:

$$C = C_{LL} + C_{LG} = C_{LL,2l} + C_{LG}; \tag{3.12}$$

as for the 3-line structure, total capacitance (C) of "DP2" wire is given by:

$$C = C_{LL} + C'_{LL} + C_{LG} = C_{LL,3l} + C_{LG}, \tag{3.13}$$

where C_{LL} and C'_{LL} are line-to-line coupling capacitance between the line of interest and left and right lines respectively.

In the experiments, we use *worst-case overlay*, which we assume to be equal to ITRS 3σ overlay for single-patterning lithography in x and y directions (i.e., 20% of the minimum feature size). 50% of the total overlay error is assumed to *originate from un-modeled terms and random errors* and are lumped into *Res* term; the remaining 50% is assumed to *originate from imperfect correction* of the six primary overlay components, i.e. translation, magnification, and rotation in field and wafer. This assumption conforms well to experimental results reported in [Ies08] where, after correction with a linear overlay model and excessive overlay sampling, 58% of overlay is non-systematic error and 42% of overlay is from imperfect correction of systematic error. To study the relative importance of each overlay component, we perform a series of experiments using different scenarios of the overlay-breakdown. A set of experiments involves extreme cases where all error caused by imperfect overlay correction is from a single source: translation, magnification, rotation, field overlay,

Table 3.2: Parameters and corresponding values used in the experiments.

Parameter	**Value**
Wafer diameter	300mm
Number of fields	63
Field dimensions	33x26mm
Number of dies per field	4
w	32nm
s	32nm
t	60.8nm[2]
H	60.8nm
L	100μm
3σ overlay	6.4nm

or wafer overlay. For field and wafer extreme cases, overlay from imperfect correction is split equally among translation, magnification, and rotation overlay components. In addition, we run a reference experiment with the overlay-breakdown of Table 3.1, which is based on estimations of the required precision for overlay measurements offered in [Lec05].

Overlay parameters in *RC* models of Section 3.1 can be inferred from the contributions of overlay components. T is equivalent to *total translation* and *Res* is *equivalent to total residual* because these two components are independent of location; whereas M_w, M_f, R_w, and R_f are *inferred by considering worst-case location that happens to be at the edge of wafer and field*. *Res* is assumed to be in *worst-case direction* across the entire wafer, which is the same direction as T.

All parameters used in the experiments and corresponding values are summarized in Table 3.2.

[2]Based on ITRS prediction of aspect ratio.

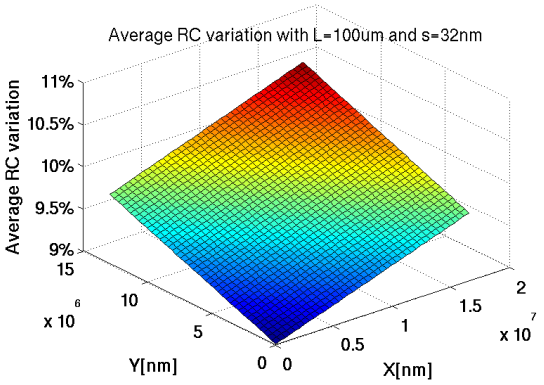


Figure 3.5: Average RC variation for the 2-line structure as a function of its location in the design when overlay components are estimated.

Table 3.3: ΔRC results for the 2-line and 3-line structures in positive DP (in % form).

	2-line structure				**3-line structure**			
	Avg variation		**Worst variation**		**Avg variation**		**Worst variation**	
	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC
Estimated components	11.5-13.6	9-10.7	21.2	16.6	1.5-1.6	1.4	3.2	2.8
Translation extreme	25	19.6	25	19.6	4.2	3.7	4.2	3.7
Magnification extreme	7.9-14.8	6.2-11.6	24.9	19.5	1.5-2	1.3-1.7	4.1	3.6
Rotation extreme	8.6-14	6.75-11	23	18	1.4-1.8	1.2-1.6	3.6	3.2
Wafer extreme	15-15.9	11.8-12.4	21.8	17.1	1.8-1.9	1.6-1.7	3.3	2.9
Field extreme	11.6-19.6	9.1-15.3	23.9	18.7	1.4-2.4	1.2-2.1	3.9	3.4

3.2.2 Evaluation Methodology

Overlay impact on the electrical characteristics of test structures was *evaluated at discrete locations of the structures in the design and for each copy of the design across the entire wafer*. We evaluate absolute worst-case impact as well as average impact over all design copies. For the case of average impact, minimum and maximum impacts for the different locations of the structures in the design are presented. The average and worst-case change of RC_{LL} and RC, which reflect the effect on interconnect delay variation, are reported for positive and negative DP processes.

3.2.3 Results

The first set of experiments is for structures formed with positive DP. Figure 3.5 plots average RC variation for the 2-line structure as a function of its location in the design

Table 3.4: ΔRC results for the 2-line and 3-line structures in negative DP (in % form).

	2-line structure				**3-line structure**			
	Avg variation		**Worst variation**		**Avg variation**		**Worst variation**	
	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC	ΔRC_{LL}	ΔRC
Estimated components	11.5-13.6	9-10.7	21.2	16.6	11.5-13.6	10.1-11.9	21.2	18.6
Translation extreme	25	19.6	25	19.6	25	22	25	22
Magnification extreme	7.9-14.8	6.2-11.6	24.9	19.5	7.9-14.8	7-13	24.9	21.9
Rotation extreme	8.6-14	6.8-11	22.3	18	8.6-14	7.6-12.3	23	20.2
Wafer extreme	15-15.9	11.8-12.4	21.8	17.1	15-15.9	13.2-14	21.8	19.2
Field extreme	11.6-19.6	9.1-15.3	23.9	18.7	11.6-19.6	10.2-17.2	23.9	21

when overlay components are estimated. This figure indicates that ΔRC varies on average from 9% to 10.6% depending on the structure location in the design (all possible locations). Minimum variation occurs when the structure is located at the origin of the design, which is the center of the field in our experiments, and maximum variation occurs when the structure is located at the edge of the design, which is to the edge of the field. This experiment is repeated for all other overlay-breakdown cases for the 2-line and 3-line structures and average and worst-case impacts are reported. Results for positive DP experiments are summarized in Table 3.3.

Similarly for negative DP, experiments for all overlay-breakdown cases are performed. Table 3.4 summarizes the results for negative process experiments.

3.3 Observations and Implications for Design

Experimental results are interpreted and important observations are brought forward in this section.

3.3.1 Results Analysis and Relative Importance of Overlay Sources

Results of Tables 3.3 and 3.4 *for the 2-line structure in positive and negative DP processes are similar.* In fact, with $s = w$ and same amount of line width and spacing variation, ΔRC is, to the first order, the same in positive and negative processes.

In case of a positive process,

$$\Delta RC_{pos} = R \times \Delta C_{LL}. \tag{3.14}$$

In case of negative process,

$$\Delta RC_{neg} = R \times \Delta C_{LG} + \Delta R \times C_{LG} + \Delta R \times C_{LL} + \Delta R \times \Delta C_{LG}. \tag{3.15}$$

To the first order, ΔR is proportional to $1/\Delta w$ and C_{LG} is proportional to Δw causing $(R \times \Delta C_{LG} + \Delta R \times C_{LG})$ and $\Delta R \times \Delta C_{LG}$ in Equation 3.15 to be very close to zero. Since ΔC_{LL} is, to the first order, proportional to $1/\Delta s$, then with $s = w$ and same amount of line width and spacing variation, $\Delta R \times C_{LL} \approx R \times \Delta C_{LL}$ and, consequently, $\Delta RC_{neg} \approx \Delta RC_{pos}$.

For the 3-line structure, results of positive and negative processes are substantially different. For positive DP, ΔRC *is much less in the 3-line structure* (1.4% on average and 2.8% worst-case variation for the experiment of estimated components) than in the case of the 2-line structure ($9 - 10.7\%$ on average and 16.6% worst-case variation for the experiment of estimated components). This huge ΔRC reduction in the case of the 3-line structure is because line-to-line capacitance between the wire in the center and its left and right neighbors change in opposite directions as illustrated in Figure 3.6. Hence, the total capacitance is not significantly affected. For negative DP, Table 3.4 shows that ΔRC *is larger in case of the 3-line structure* ($10.1 - 11.9\%$ on average and 18.6% worst-case variation for the experiment of estimated components) than in case of the 2-line structure ($9 - 10.7\%$ on average and 16.6% worst-case variation for the experiment of estimated components). C and R vary in opposite directions reducing the overall effect on ΔRC. For the 3-line structure, the additional C_{LL} term with the third line is unaffected by overlay resulting in the reduction of overall ΔC. This explains why ΔRC is larger in case of the 3-line structure than in case of the 2-line structure.

Relative importance of different overlay sources can be inferred from the results. For the 2-line structure, translation extreme experiment leads to 19.6% ΔRC; magnification extreme experiment leads to 6.2-11.6% average ΔRC and 19.5% worst-case

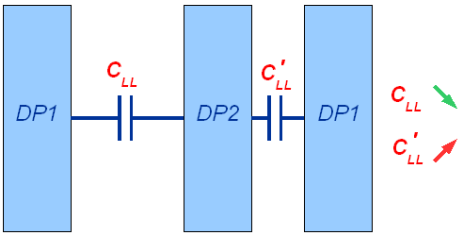


Figure 3.6: Illustration of cancellation effect between line-to-line capacitances in the 3-line structure.

ΔRC; and rotation extreme leads to 6.75-11% average ΔRC and 18% worst-case ΔRC. *Translation impact on average ΔRC is much more important than magnification or rotation impact.* This difference is because magnification and rotation overlay vectors can have opposite directions and their effects are canceled out when averaging over the entire wafer; whereas, translation is actually fairly uniform across wafer[3]. Nevertheless, for worst-case ΔRC, translation, magnification, and rotation are almost equally important. Similar conclusions can be drawn from the results for the 3-line structure in case of a positive and a negative processes. Moreover, in both processes, magnification and rotation have very similar electrical impacts.

Results also show that field overlay has a slightly larger electrical impact than wafer overlay. In addition, field overlay is more dependent on location in the design plane, which is marked by a larger difference between minimum and maximum average variation in Tables 3.3 and 3.4. In practice, however, the amount of field overlay is much smaller than the amount of wafer overlay [Lec05].

3.3.2 Effects of Design Parameters

Effects of wire length (L) and spacing (s) are evaluated by running the "estimated components" experiment for the case of the 2-line structure in positive DP with different values of L and s. Average and worst-case RC variations are reported in Tables 3.5 and 3.6 respectively. These results show that the *effect of L on overlay electrical impact is negligible.* On the other hand, results show a *large effect of s* on overlay electrical impact; e.g., with 20% increase of s, ΔRC is reduced by 22%

[3]In the experiments, translation vector is assumed to have a uniform direction across wafer.

Table 3.5: Average ΔRC across wafer for different values of wire length (L) and spacing (s) in case of the 2-line structure with positive DP.

	s = **25.6nm**	s = **32nm**	s = **38.4nm**
$L = 10\mu$m	13.46%	10.03%	7.91%
$L = 100\mu$m	13.30%	9.84%	7.69%
$L = 1000\mu$m	13.28%	9.82%	7.67%

Table 3.6: Worst case ΔRC across wafer for different values of wire length (L) and spacing (s) in case of the 2-line structure with positive DP.

	s = **25.6nm**	s = **32nm**	s = **38.4nm**
$L = 10\mu$m	22.91%	16.60%	12.81%
$L = 100\mu$m	22.90%	16.60%	12.81%
$L = 1000\mu$m	22.86%	16.57%	12.78%

on average and 23% in the worst case. The effect of s is even larger for smaller dimension of the half-pitch; e.g., with 20% increase of s in 25.6nm half-pitch, ΔRC is reduced by 26% on average and 28% in the worst case.

Similar experiments are run for negative DP. Effects of wire length (L), width (w), and spacing (s) are evaluated by running the experiment for the case of the 2-line structure and estimated overlay components. Average and worst-case ΔRC are reported in Tables 3.7 and 3.8 respectively. Results show that the *effect of L on overlay electrical impact is also negligible*. On the other hand, results show a *large effect of w and a minor effect of s* on overlay electrical impact; e.g., with 20% increase of w, ΔRC is reduced by 22% on average and 23% in the worst case and, with 20% increase of s, ΔRC is reduced by 4% on average and in the worst case. The effect of w is even larger for smaller dimension of the half-pitch; e.g., with 20% increase of w in 25.6nm half-pitch, ΔRC is reduced by 26% on average and 28% in worst case. On the contrary, the effect of s becomes smaller as the half-pitch gets smaller; e.g., with 20% increase of s in 25.6nm half-pitch, ΔRC is reduced by 3% on average and in the worst case.

Referring to the equations describing electrical characteristics of wires, i.e. Equations (3.4-3.11), overlay in y direction (x for horizontal metallization) only affects

Table 3.7: Average ΔRC across wafer for different values of wire length (L), width (w), and spacing (s) in case of the 2-line structure with negative DP.

	w = **25.6nm**			w = **32nm**			w = **38.4nm**		
s =	**25.6nm**	**32nm**	**38.4nm**	**25.6nm**	**32nm**	**38.4nm**	**25.6nm**	**32nm**	**38.4nm**
$L = 10\mu$m	13.80%	13.30%	12.84%	10.28%	9.84%	9.43%	8.09%	7.68%	7.32%
$L = 100\mu$m	13.80%	13.30%	12.84%	10.28%	9.84%	9.43%	8.09%	7.68%	7.32%
$L = 1000\mu$m	13.80%	13.30%	12.84%	10.28%	9.84%	9.43%	8.09%	7.68%	7.32%

Table 3.8: Worst case ΔRC across wafer for different values of wire length (L), width (w), and spacing (s) in case of the 2-line structure with negative DP.

	w = **25.6nm**			w = **32nm**			w = **38.4nm**		
s =	**25.6nm**	**32nm**	**38.4nm**	**25.6nm**	**32nm**	**38.4nm**	**25.6nm**	**32nm**	**38.4nm**
$L = 10\mu$m	23.77%	22.91%	22.11%	17.36%	16.60%	15.91%	13.48%	12.81%	12.20%
$L = 100\mu$m	23.77%	22.91%	22.11%	17.36%	16.60%	15.91%	13.48%	12.81%	12.20%
$L = 1000\mu$m	23.72%	22.86%	22.06%	17.32%	16.57%	15.88%	13.45%	12.78%	12.17%

L term; since the impact of L is negligible as we demonstrated in the experiments, *overlay in the y direction has virtually no electrical effects*. Hence, *preferred routing direction should be taken into account in overlay sampling and alignment strategies*.

The significant effects of s and w in layouts fabricated with positive and negative DP indicates the importance of *wire spreading* (in positive process) and *widening* (in negative process), which are widely used Design for Manufacturability (DFM) techniques. Nevertheless, the use of these methods is limited to non-congested regions of the layout where excess spacing is available.

3.3.3 Expected Worst-Case Overlay Impact in Critical Path

The analysis in previous sections considers a single line suffering a resistance or capacitance increase due to overlay in the 2-line and 3-line structures. Nevertheless, overlay can cause a simultaneous resistance or capacitance decrease in other lines. In this section, we study the expected worst-case impact of overlay in a path with multiple line segments.

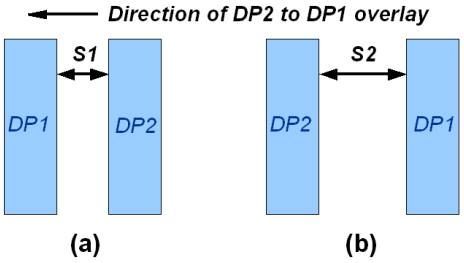


Figure 3.7: Overlay impact dependence on layout decomposition in the 2-line structure with positive DP: reduced spacing between lines for decomposition of (a) and increased spacing between lines for decomposition of (b).

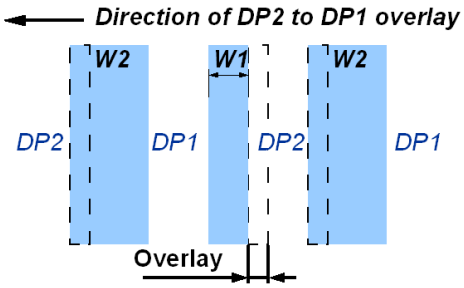


Figure 3.8: Different overlay impact on wire width in negative DP.

Impact of Layout Decomposition

In DP, two features must be assigned to different exposures if their spacing is smaller than the required minimum spacing between features printed with the same exposure. This feature assignment between first and second exposure, a.k.a. layout decomposition and layout coloring, has a significant effect on the overlay impact in the 2-line structure with a positive process. Considering the 2-line structure of Figure 3.7 and assuming the overlay of second exposure to first exposure is in one direction, the different decompositions of Figure 3.7(a) and Figure 3.7(b) lead to reduced spacing between lines in one case and increased spacing between lines in the other case. Layout decomposition has a significant effect on the overlay impact in negative process as well. In this case, spaces that are too close to each other are assigned to different exposures. Depending on the decomposition, some lines will see reduced width and direct neighboring lines will see an increased width as illustrated in Figure 3.8. Results of Tables 3.3 and 3.4 correspond to the line with worsened RC, i.e., reduced spacing in case of a positive process and reduced width in case of

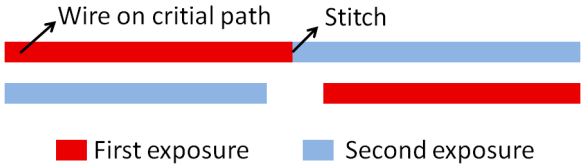


Figure 3.9: Illustration of swizzled decomposition to reduce ΔRC of wires on critical paths.

negative process. We illustrate the magnitude of this decomposition effect on worst-case variation in a path with multiple line segments. Consider a path composed of two line segments where each segment is part of a 2-line structure in positive process. If both line segments have worsened RC, the worst-case interconnect ΔRC of the path is 16.6%; whereas, if one segment has worsened RC and the other segment has a relieved RC, the overall worst-case interconnect ΔRC of the path is only 2.5%.

Layout decomposition effect can be exploited to reduce ΔRC of wires on critical paths in positive process. This can be done using swizzled decomposition, where wires are split into segments and connected segments are assigned to different exposures as shown in Figure 3.9. In this case, stitches need to be inserted with minor concern for manufacturability and wire resistance if enough overlap margin is ensured. The same benefit can also be achieved by actual wire swizzling (as in [GK04]). Actual wire swizzling, however, introduces additional vias with negative impact on wirelength and routability. Moreover, timing-aware layout decomposition strategies (e.g., [GJK09]) can be very useful for improving timing yield.

Impact of Congestion

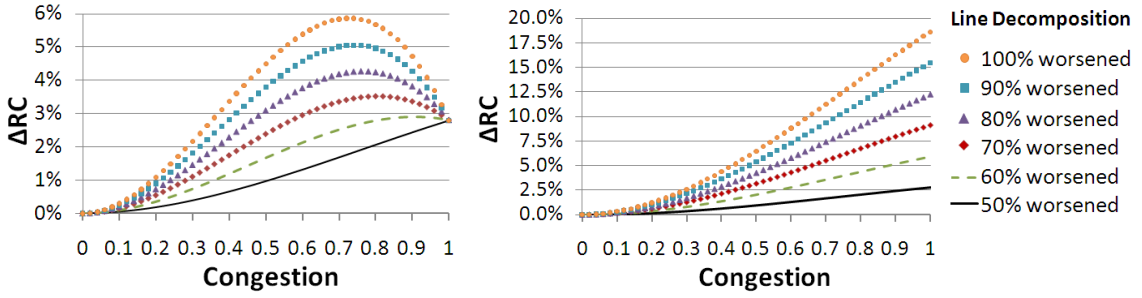
Results of Table 3.3 for positive DP show a much larger overlay impact in case of the 2-line structure than in case of the 3-line structure. On the contrary, results of Table 3.4 for negative DP show a larger overlay impact in the case of the 3-line structure than in the case of the 2-line structure. This dependence of overlay impact on the wire neighborhood gives motivation for considering congestion in electrical evaluation of overlay impact. Given the average congestion G in a layout, we estimate the probability of a line to have two neighboring lines (i.e., 3-line structure),

one neighboring line (i.e., 2-line structure), and no neighbors (i.e., 1-line structure). This is done by considering three channels where each is occupied by a wire with probability equal to G. The probability of 3-line structure, P_{3l}, is G^3, the probability of 2-line structure at minimum spacing[4], P_{2l}, is $G^2 \times (1-G) \times 2$, and the probability of 1-line structure (only middle channel is occupied), P_{1l}, is $G \times (1-G)^2$. Hence, the expected ΔRC of a randomly chosen line from the layout is

$$\begin{aligned} \Delta RC_{avg} &= \Delta RC_{3l} \times P_{3l} + \Delta RC_{2l} \times P_{2l} + \Delta RC_{1l} \times P_{1l} \\ &= \Delta RC_{3l} \times G^3 + \Delta RC_{2l} \times G^2 \times (1-G) \times 2 \\ &\quad + \Delta RC_{1l} \times G \times (1-G)^2. \end{aligned} \tag{3.16}$$

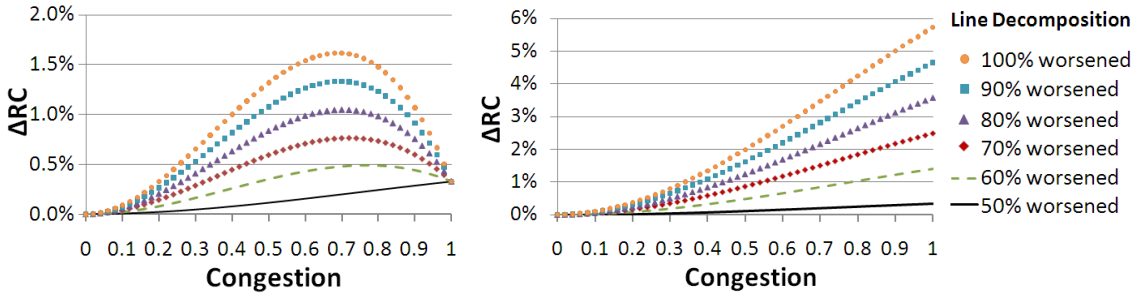
In positive DP process, $\Delta RC_{1l} = 0$ because overlay has no effect on 1-line structure. In negative DP process, overlay can result in line-width variation in a 1-line structure. Nevertheless, R and C_{LG}, which is the only capacitance term in this case, varies in opposite directions rendering the overall ΔRC_{1l} negligible. Using Equation (3.16) and ΔRC_{2l} and ΔRC_{3l} values for the case of estimated overlay components in Table 3.3 (worst-case variation), we plot in Figure 3.10(a) interconnect ΔRC of a path as a function of congestion for the case of a positive process. This is performed for different splits of line segments in the path between worsened and relieved RC, namely, 50 to 100% of lines with worsened RC with 10% intervals and the remaining fraction of lines having a relieved RC. When assuming all line segments in the path have a worsened RC (i.e., most pessimistic worst-case split), interconnect worst-case ΔRC is *at most 5.9%* (for 72% congestion) and is less than 4.7% for highly congested layouts (90% and more). When assuming a 50-50% split of lines between worsened and relieved RC (i.e., most optimistic best-case split), interconnect worst-case ΔRC is *at most 2.8%*, which occurs at a 100% congestion.

The same plots are reproduced in Figure 3.10(b) for the case of negative process. Here, if all line segments in the path are assumed to have a worsened RC (i.e., most pessimistic worst-case split), interconnect worst-case ΔRC *increases monotonically to reach 18.6% at 100% congestion.* When assuming a 50-50% split of lines between worsened and relieved RC (i.e., most optimistic best-case split), interconnect worst-



(a) (b)

Figure 3.10: Plot of path interconnect worst-case ΔRC versus congestion for different line splits between worsened and relieved RC variation in positive DP process (a) and negative DP process (b). The plots assume an overlay budget equal to 20% of half-pitch.



(a) (b)

Figure 3.11: Plot of path interconnect worst-case ΔRC versus congestion for different line splits between worsened and relieved RC variation in positive DP process (a) and negative DP process (b). The plots assume an overlay budget equal to 7% of half-pitch.

case ΔRC *increases monotonically to reach 2.8% at 100% congestion.*

In Figures 3.10(a) and 3.10(b), we assume an overlay budget equal to 20% of half-pitch, i.e., ITRS projected overlay budget for single patterning. In Figure 3.11, we repeat the same experiments with an overlay budget of equal to 7% of half-pitch, which corresponds to ITRS projected CD budget for single patterning. In this case, overlay impact is small indicating that having a double-patterning overlay budget equal to single-patterning CD budget is certainly too conservative.

Congestion benefit can be exploited to reduce ΔRC *in positive process.* This is achieved by *dummy fill insertion*, which is also used to improve planarity.

[4]Multiplication by 2 accounts for the two possible locations of the 2-line structure: occupying

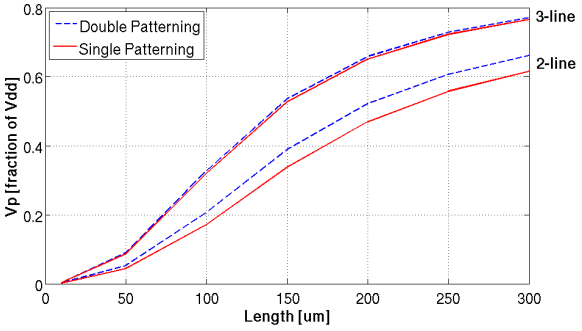


Figure 3.12: Peak crosstalk noise versus interconnect length in positive DP process.

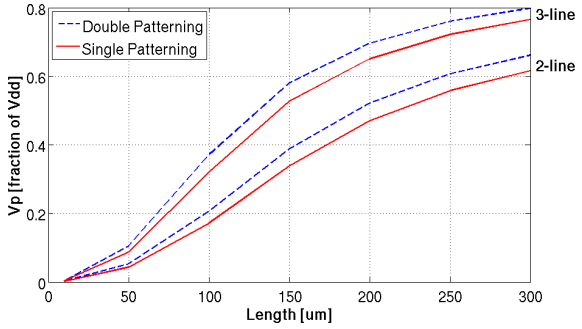


Figure 3.13: Peak crosstalk noise versus interconnect length in negative DP process.

3.3.4 Overlay Impact on Crosstalk Noise

Adjacent lines are susceptible to crosstalk noise, which can lead to signal integrity issues when the noise amplitude exceeds some threshold for a duration long enough to cause a bit flip. A widely used measure of crosstalk noise is the maximum voltage change (a.k.a. peak crosstalk voltage) of the victim line. Using a simplified lumped RC model, the peak crosstalk voltage is determined using the following model [RIX94, CK99]

$$V_p = \begin{cases} V_{dd}\frac{R_v C_{LL,2l}}{t_f}(1 - e^{-(t_f)/R_v C}) & \text{for 2-line struct,} \\ V_{dd}\frac{R_v C_{LL,3l}}{t_f}(1 - e^{-(t_f)/R_v C}) & \text{for 3-line struct,} \end{cases} \tag{3.17}$$

where V_{dd} is the supply voltage, R_v is the resistance of the victim line, t_f is the rise time at the aggressor line, and C is the total capacitance of the victim line.

In Figure 3.12, we plot the peak crosstalk voltage as a function of the length

either first two channels or last two channels.

of adjacent lines for the 2-line and 3-line structures in 32nm half-pitch for positive single and double-patterning processes. Here, we assume a clock frequency of 2GHz, t_f equal to 1/8 of the clock period (i.e., 62ps), and effective resistivity of 48.3Ω-nm and effective dielectric constant of 2.6 that are the values projected by ITRS. Compared to single patterning, V_p in the case of double-patterning is slightly higher *by at most 0.05 of* V_{dd} (e.g., 50mV for V_{dd} =1V) in the 2-line structure and *is less than 0.01 of* V_{dd} (e.g., $<$ 10mV for V_{dd} =1V) in the 3-line structure. The same study is repeated for negative DP process and plots are reproduced in Figure 3.13. For the 2-line structure, we observe the same results as in the case of a positive process; however, results for the 3-line structure are substantially different. Compared to single patterning, V_p in the case of double-patterning is slightly higher *by at most 0.05 of* V_{dd} (e.g., 50mV for V_{dd} =1V). This disparity between the results in positive and negative processes is attributed to the ΔC cancellation effect for the 3-line structure in positive process, which is absent in negative process.

3.3.5 Estimation of Overlay Requirement

Reduction of overlay budget requires challenging and expensive overlay control and alignment strategies. In some cases, this might even necessitate the replacement of scanners by newer ones with better alignment accuracy. As a result, *determining how much overlay is "really" required can avoid unnecessarily tight and costly overlay control.*

Even though overlay error translates into CD variation in DP, our conjecture is that *overlay requirement can be alleviated if electrical variation is the basis for determining the requirement* rather than CD variation, which may lead to excessively constricted budget [Haz08a].

In Figure 3.14, absolute worst-case ΔRC in a single line and worst-case CD variations in positive and negative DP processes are plotted. Even when considering absolute worst-case electrical variation in a single line, *overlay requirement determined from electrical variation tolerance is significantly smaller than that determined*

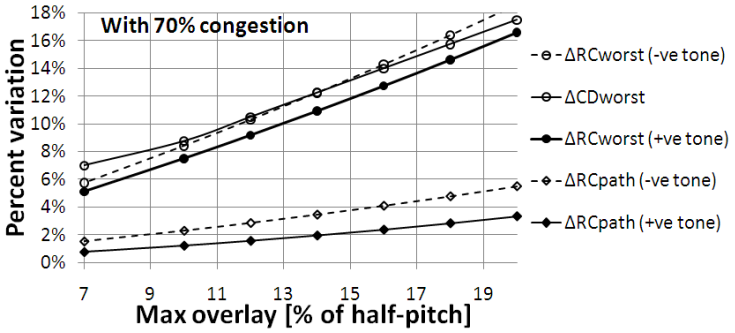


Figure 3.14: Plots of average and worst case CD and C variations versus requirement of maximum overlay with 70% congestion in positive and negative DP.

from CD variation tolerance in positive process; e.g., 10% electrical variation tolerance in case of absolute worst-case ΔRC requires overlay $<$ 4.3nm, while the same CD tolerance requires overlay $<$ 3.8nm; this consists of a 13% reduction of overlay requirement. Figure 3.14 also plots worst-case ΔRC in a path at 70% congestion (i.e., worst-case congestion in positive process) and assuming 70% of line segments of the path have a worsened RC and the remaining segments have a relieved RC. In this case, overlay requirement is greatly alleviated when electrical variation, instead of CD variation, is used to determine the requirement. For positive process, overlay requirement determined using worst-case ΔRC in a path is *5 to 9 times smaller* than that determined using absolute worst-case CD variation. Similarly for negative process, overlay requirement determined using worst-case ΔRC in a path is *3 to 5 times smaller* than that determined using absolute worst-case CD variation.

3.4 Exploring Processing Options

From a *manufacturing perspective*, negative dual-trench process is preferred over positive dual-line process because it requires less processing steps and the first pattern is better protected from the processing steps of the second pattern (i.e., exposure and etch) in negative process [ADF06].

Positive process yield better *patterning quality* than negative process. In particular, positive process has a larger exposure latitude and smaller Mask Error Enhancement Factor and Line-End Roughness than negative process [Lim06, ADF06].

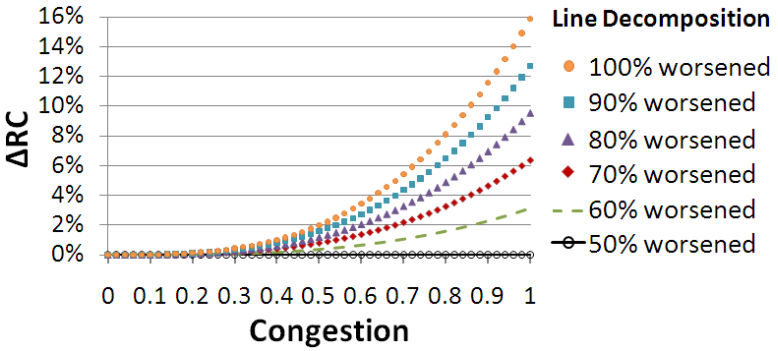


Figure 3.15: Difference between worst-case interconnect ΔRC for a path in positive process and negative process as a function of congestion for different line splits between worsened and relieved RC variation.

Layout decomposition – between first and second exposures – is much more complex in the case of a negative process than in the case of a positive process as can be inferred from Figure 3.1. The complication in negative process is attributed to the fact that the final pattern is formed by trimming unwanted area through either the first or second patterning steps, which substantially increases the number of possible decompositions for the same layout.

From the *electrical perspective*, experimental results of Table 3.3 and 3.4 show that the absolute worst-case electrical variation in positive process (ΔRC for the 2-line structure) is slightly smaller than that in negative process (ΔRC for the 3-line structure), namely, ΔRC of 16.6% in positive process and ΔRC of 18.6% in negative process. Moreover, the expected worst-case electrical variation for a path is much smaller for the case of a positive process than in the case of a negative process. In Figure 3.15, we plot the difference between this variation for positive process and negative process as a function of congestion for different line splits between worsened and relieved RC variation. The difference increases with the level of congestion and more pessimistic assumption on the line-split between worsened/relieved RC variation to reach up to 16% of ΔRC. This large difference between the results of positive process and that of negative process is attributed mainly to the cancellation effect between line-to-line capacitances in the 3-line structure, which only occur in positive DP process.

As for signal integrity issues, we observe identical results in positive process and negative process if the worst case crosstalk noise is assumed to occur in the 2-line structure (i.e., single aggressors); whereas, if the worst case crosstalk noise is assumed to occur in the 3-line structure (i.e., two aggressors), positive process results in a much smaller increase of peak crosstalk noise compared to negative process as demonstrated in Section 3.3.4 (less than 0.01 of V_{dd} versus 0.05 of V_{dd}).

As a result, *a positive process is preferred over a negative process* from an electrical perspective at 32nm half-pitch. *This conclusion is expected to remain valid at future technology nodes independent of scaling* as long as $s = w$ as we have shown in Section 3.3.1.

Conclusions and Future Work

In this work, we electrically evaluate overlay impact in positive and negative DP processes. Experimental results show that the *expected electrical impact of overlay in a path is not severe* especially when congestion and layout decomposition effects are considered. On the other hand, the *absolute worst-case electrical impact of overlay in a line remains a serious problem* (up to 16.6% ΔRC and up to 50mV increase of peak crosstalk noise). Many methods are available for designers to reduce the absolute and expected overlay impacts, especially in critical paths, including *wire spreading and widening*, *swizzled decomposition*, and *dummy fills*. As a result, *overlay requirement can be relaxed* if electrical variation, rather than CD variation, is the basis for determining the requirement. Furthermore, we analyze process options for 32nm half-pitch node and conclude that *positive process is the preferable* process option from an electrical perspective. Our study of the relative importance of different overlay sources reveals that *translation overlay has the largest electrical impact* among all the sources. In addition, overlay in y direction (x for horizontal metallization) has negligible electrical impact. Therefore, *preferred routing direction should be taken into account for overlay sampling and alignment strategies*. In future work, we will extend the results to cover front-end-of-line layers and try to relax overlay

requirements by developing DP-specific and design-aware alignment strategies.

Acknowledgements

This work was generously supported in part by IMPACT UC Discovery Grant (http://impact.berkeley.edu) contract number ele07-10291, Semiconductor Research Corporation contract number 2008-TJ-1816, and NSF CAREER Award number 0846196.

Chapter 4

Single-Mask Double-Patterning Lithography

In this chapter, we propose *shift-trim double patterning lithography* (STDP), a cost-effective double patterning technique for achieving pitch relaxation with a single photomask. The mask is re-used for the second exposure by applying a translational mask-shift. An additional non-critical trim exposure is applied to remove extra printed features. STDP can be used to pattern critical layers and is very suitable for regular and gridded layouts, where redesign effort and area overhead are minimal. The viability of STDP is demonstrated through a design implementation at the poly and contacts layers in bidirectional layouts. Standard-cell layouts are constructed so as to avoid layout decomposition conflicts, which are found to be the limiting factor for the pitch relaxation that can be achieved with double-patterning (STDP as well as LELE double-patterning). 1.8× pitch relaxation is achieved in our implementation while ensuring no layout decomposition conflicts and a small area overhead. Specifically, in comparison to layouts assumed to be feasible with a hypothetical single-patterning process, we observe virtually no area overhead when STDP is applied to the poly layer (<0.3% cell-area overhead) and no more than 4.7% cell-area overhead when STDP is applied at both the poly and contacts layers. Compared to standard pitch-split double-patterning, the proposed method: (1) cuts mask-cost to nearly half, (2) reduces overlay errors between the two patterns, (3) alleviates the bimodal line-width distribution problem in double patterning, and (4) slightly enhances the throughput of critical-layer scanners.

Introduction

Double-patterning lithography (DP) is one of the most likely short-term solutions for keeping the pace of scaling beyond 32nm node [Mac08]. It is one of the many resolution enhancement techniques (RET) that have been introduced to push the limit of optical lithography. DP can be implemented with different manufacturing processes: litho-etch-litho-etch (LELE), litho-litho-etch (LLE), and self-aligned double patterning (SADP), a.k.a. spacer double patterning (SDP). In SADP, sidewall spacer defines either spaces or lines depending on the tone of the process and extra printed features are trimmed away using a cut or block mask. Many patterns cannot be printed using SADP, which make it more suitable for well-structured memory cells than random logic layout [Mac08]. This chapter focuses on LELE and LLE processes referred to as standard-DP processes hereafter.

DP has four major impediments: high mask-cost, low throughput, within-layer overlay errors, and the CD bimodality problem. DP mask-cost is estimated to twice that of single patterning because of the need for two critical masks. The additional processing steps required for double patterning significantly reduce the fabrication throughput. The overlay budget being determined by interactions between different layers in single patterning (e.g., metal overhang on via), 20% of half-pitch estimated by ITRS is considered sufficient. In DP however, overlay budget is much tighter since overlay translates directly into CD variability [ADF06], which has a budget three times tighter than inter-layer overlay according to ITRS [ITRb], and, hence, introduces an extra source of variability [GG09b]. CD typically follows a normal distribution with some σ and μ, which deviates slightly from the target. Since DP has two separate exposure and etch steps, two populations exist: one for features formed by the first exposure/etch step and another for features formed by the second exposure/etch step.

An attempt to use DP with a single photomask and, hence, reduce its cost, is reported in [YRS07]. It consists of splitting the mask area into two regions, each corresponding to a different pattern (similar to a multi-layer reticle). As reported

in [YRS07], this approach renders fabrication throughput even worse than that of standard-DP and does not address other DP technical challenges including within-layer overlay and CD bimodality.

In this chapter, we propose shift-trim DP (STDP), an effective method to use a single mask to achieve pitch-relaxation. Essentially, the method consists of applying a translational mask-shift to re-use the same photomask for both exposures of DP. Extra printed features are then removed using a non-critical trim exposure. STDP can be applied to all layers including, but not limited to, active, polysilicon, contacts, metal, and via layers. Moreover, the method can be used for any type of design as long as some basic layout restrictions (discussed in Section 4.2) are met. We demonstrate the viability of the proposed method when employed to pattern the polysilicon (poly) and contacts (CA) layers in standard-cell based designs. Cell layouts are constructed so that to avoid layout decomposition conflicts. Resolving decomposition conflicts between features of different cells is found to be the limiting factor for the pitch relaxation that can be achieved with double patterning (STDP as well as standard pitch-split DP). As a result, 2× pitch relaxation without conflicts is achieved only at high area overhead. STDP designs show little area overhead, however, while ensuring 1.8× pitch relaxation and no layout decomposition conflicts.

STDP manufacturing process and design requirements are discussed in Section 4.1. In Section 4.2, STDP feasibility at the poly and CA layers are demonstrated by creating a compatible standard-cell library by layout migration of Nangate open cell-library [Nana] and generating compatible real designs. When compared to the original Nangate layouts (assumed to be feasible with a hypothetical single-patterning process), STDP designs show virtually no area overhead for STDP implementation at the poly layer and an affordable area overhead of at most 4.7% for STDP implementation at both the poly and CA layers. In both STDP implementations, the generated trim mask layouts are simple and lead to an easy-to-fabricate photomask. Benefits of the proposed method in terms of cost, overlay control, CD performance, and throughput are discussed in Section 4.3.

4.1 Shift-Trim DP Overview and Layout Restrictions

This section presents an overview of STDP technique and its associated layout restrictions and challenges.

4.1.1 Manufacturing Process

STDP involves the following steps:

1. print the first pattern as in standard DPL processes;
2. shift the photomask of step (1) by a predetermined nanoscale amount X (equal to minimum gate pitch for poly-layer STDP) and print the second pattern;
3. apply a non-critical trim (a.k.a. block) exposure to remove unnecessary features.

The translational mask shift in step (2) is accomplished without any unloading and reloading of the photomask from the exposure tool and no extra requirements on exposure tools. Today's scanners have the capability to perform such translational shift automatically with high precision ($\approx$ 0.6nm) [Lev05].

STDP can be implemented using positive dual-line and negative dual-trench LELE and LLE processes with little modifications as demonstrated in Figure 4.1[1]. We only show the case of positive resist since it is more commonly used in modern lithography. Negative resist can also be used with little changes to the manufacturing process. In this figure, the processes are presented in order of popularity with the first process on the left being the most popular. Although LLE has higher throughput and lower cost than LELE, to best of our knowledge, LLE is currently not production-worthy. For our design implementation of STDP, we use a positive dual-line LELE process. Nevertheless, STDP implementation with the other less popular/realistic alternatives can be performed with little modifications.

STDP requires an extra step on top of standard-DPL. It consists of an inexpensive and non-critical trim-exposure cycle (resist coat-expose-develop) and removal of

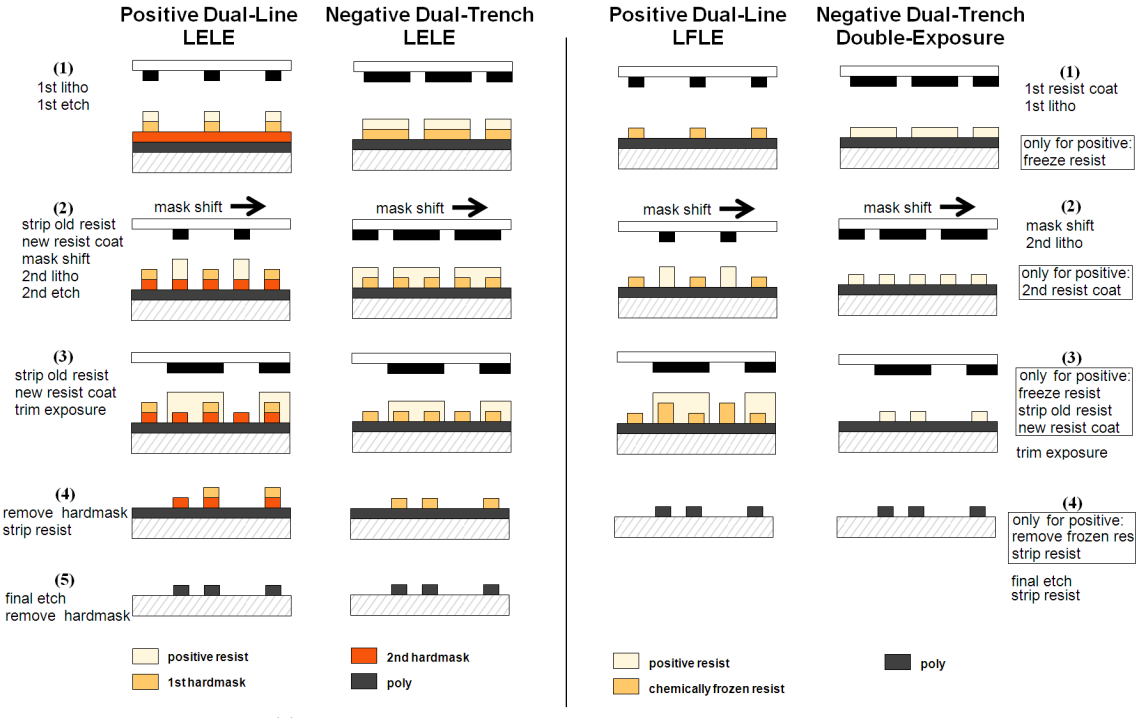


Figure 4.1: Proposed manufacturing processes for STDP: positive dual-line and negative dual-trench LELE, positive dual-line LLE, a.k.a. litho-freeze-litho-etch (LFLE) process, and negative dual-trench double-exposure processes.

hardmask corresponding to extra printed features before the final etch. The trim exposure is a mature and well-known method used in many patterning techniques such as SADP [Chi08a, Shi08], alternating phase-shift mask [LLH01], and subtractive-litho patterning [Mis07, VBR08]. It was recently employed to trim-away printing assist features (PrAF) introduced to enhance the resolution of conventional single patterning [HMB07]. A second hardmask layer is necessary in case of positive LELE process, but this does not represent an extra requirement because many standard DPL implementations favor the use of a second hardmask [ADF06, TYY08]. Relaxed CD and overlay requirements of the trim exposure, which are demonstrated by the results of our implementation, make process control an easy task. Consequently, the cost of trim mask is minor compared to the cost of conventional masks and the trim exposure can be realized using second-tier scanners if this is desirable to enhance throughput.

4.1.2 STDP Challenges and Downsides

In addition to the downside of the need for three exposures (two critical and one non-critical trim exposure) and the extra processing steps associated with the trim exposure, STDP has two other challenges.

Although the same features with exactly the same surroundings are on the mask of the first and second exposures, features of different exposures printed on wafer may vary due to process-differences (e.g., resist thickness, hardmask characteristics, etch-interference, etc...). One way to compensate for this variation consists of using different OPC features for the different patterns [Bee07]. In STDP, this method is no longer possible since the same mask is used for the first and second exposures. As a result, other means to correct for processing differences between the two patterns must be employed (e.g., dose-mapping [Jee06]).

Because the STDP mask-shift is performed just uniformly across the design, the minimum gate-pitch must be set to the contacted gate-pitch (typically equal to the amount of the mask-shift) and all gates in the design must follow the same orientation. When memory and logic are integrated, this gate-pitch limitation may impose restrictions on the allowed contacted pitch for memory if it is not the same as the logic contacted pitch.

4.1.3 Layout Restrictions at Poly-line Layer

Basic layout restrictions are imposed for implementing STDP at the poly layer. X being the amount of mask shift and X_0 being the minimum gate pitch on the mask[2], the following restrictions apply.

1. For every gate, the pitch to the neighboring gates from one side (subsequent gate to the right or left side) must be either X or $\geq X_0$ and the pitch between the left and right neighboring gates must be $\geq X_0$. This is illustrated by the example of three gate-poly lines shown in Figure 4.2.

[1]BARC layers are not shown for brevity.

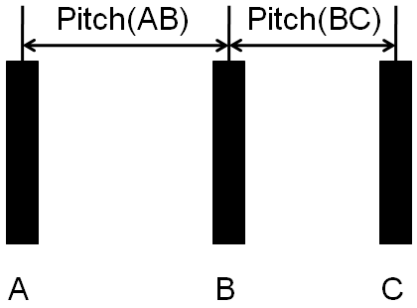


Figure 4.2: Example illustrating gate-pitch restriction. In case $Pitch(AB)$ is $< X_0$ but different than X, then $Pitch(BC)$ must be either X or $\geq X_0$ and $Pitch(AC)$ must be $\geq X_0$. Similarly, if $Pitch(BC)$ is unrestricted, $Pitch(AB)$ is restricted to X or $\geq X_0$ and $Pitch(AC)$ to $\geq X_0$.

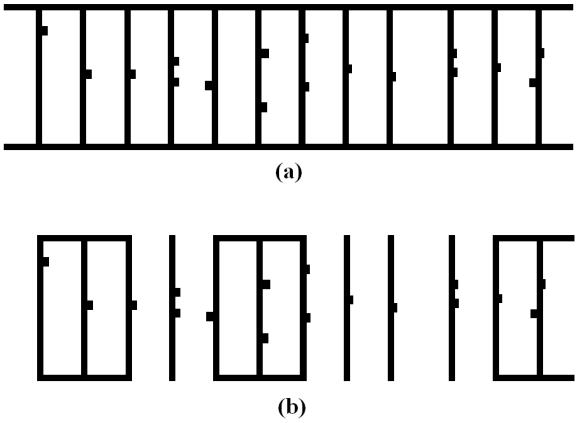


Figure 4.3: Poly layer STDP critical mask snippet corresponding to a flip-flop cell with two structure-options (a) and (b).

2. In light of (1), minimum gate spacing is equal to contacted-gate spacing (equal to X minus poly-line width).

3. "Wrong-direction" (horizontal) poly routing is restricted to top and bottom routing channels of the cell (i.e. poly-routing in the center of the cell is not allowed).

In addition, some design rule restrictions (especially line-end to field-poly spacing and line-end gap) may be necessary to guarantee a simple trim-mask as we show later in this chapter.

[2] X_0 is typically $2X$.

STDP implementation for fixed pitch poly grating is straightforward and requires no redesign effort. In this case, STDP critical mask still consists of fixed-pitch grating but with a perfect 2× pitch relaxation. STDP for unidirectional-poly designs with non-fixed pitch requires small redesign effort. In particular, adjustment of the pitch between some lines might be necessary to enforce restriction (1). This restriction is easily met in real designs, however, because most gates are at contacted-pitch (equal to X) from at least one of its two neighbors. The critical mask for this type of designs consists of simple unidirectional lines with twice the minimum pitch of single patterning. The most challenging type of designs is conventional logic and sequential circuits that involve bidirectional-poly. To handle such designs, two lines in the opposite direction are added at the top/bottom of the critical mask of the cell leading to the ladder-like shapes illustrated in Figure 4.3[3]. This permits the use of "wrong-way" poly to connect gates internally within the cell in the top/bottom routing channels. Both critical-mask options of Figure 4.3 are possible without any effects on the complexity of the trim-mask. Option (a) has wrong-way lines whether they are needed or not. On the other hand, option (b) has these lines only when needed. As a results, option (b) leads to less corner-rounding than option (a). Yet, we assume option (a) in our implementation because it is very regular and, consequently, more favorable for lithography [SLC08a, PSS03]. In these structures, gate-pitch is twice the minimum pitch of single patterning, which ensures pitch-doubling, and small notches that appear on vertical lines correspond to contact-landing pads, which are avoidable in processes in some processes (e.g. Intel's 45nm process [Mis07]).

For all types of designs, layout decomposition of the poly-line layer into critical and shifted exposures is trivial as we show in Section 4.2.2. It is worth noting that the use of the trim exposures allows the elimination of layout decomposition conflicts that may occur when using standard DP. This is because all vertical spacings are formed with the trim exposure and do not appear on the critical mask. Figure 4.4 illustrates some examples on how the use of a trim exposure resolves the decomposition conflicts.

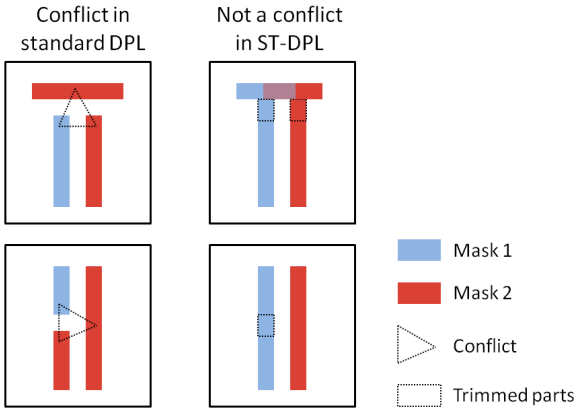


Figure 4.4: Examples showing how the trim exposure resolves decomposition conflicts.

STDP steps applied at the poly-line layer of a 4-input OAI (OR-AND-Invert) CMOS standard-cell from Nangate 45nm open library [Nana] are illustrated in Figure 4.5. In this example, the cell-layout is made compatible with STDP without any area overhead. STDP application at the poly-line layer of most standard-cells is straightforward and introduces no or little area overhead as we show in the next section.

4.1.4 Layout Restrictions at Contacts Layer

STDP implementation at the CA layer imposes more restrictions on the layout than in the case of poly-line layer. X being the amount of mask shift and X_0 being the minimum contact pitch, the following restrictions apply.

1. The contact pitch is restricted similarly to gate-poly pitch. For every contact, the pitch to the neighboring contact from one side (subsequent contact to the right or left side) must be either X or $\geq X_0$ and the pitch between the left and right neighboring contacts must be $\geq X_0$.

2. Assuming the mask shift is to the right, subsequent contacts at pitch equal to X where the left contact is assigned to the first exposure and the right contact is assigned to the second exposure must be aligned to the same vertical

[3]The shape shown in this figure is for illustration purposes and do not include RET-related features.

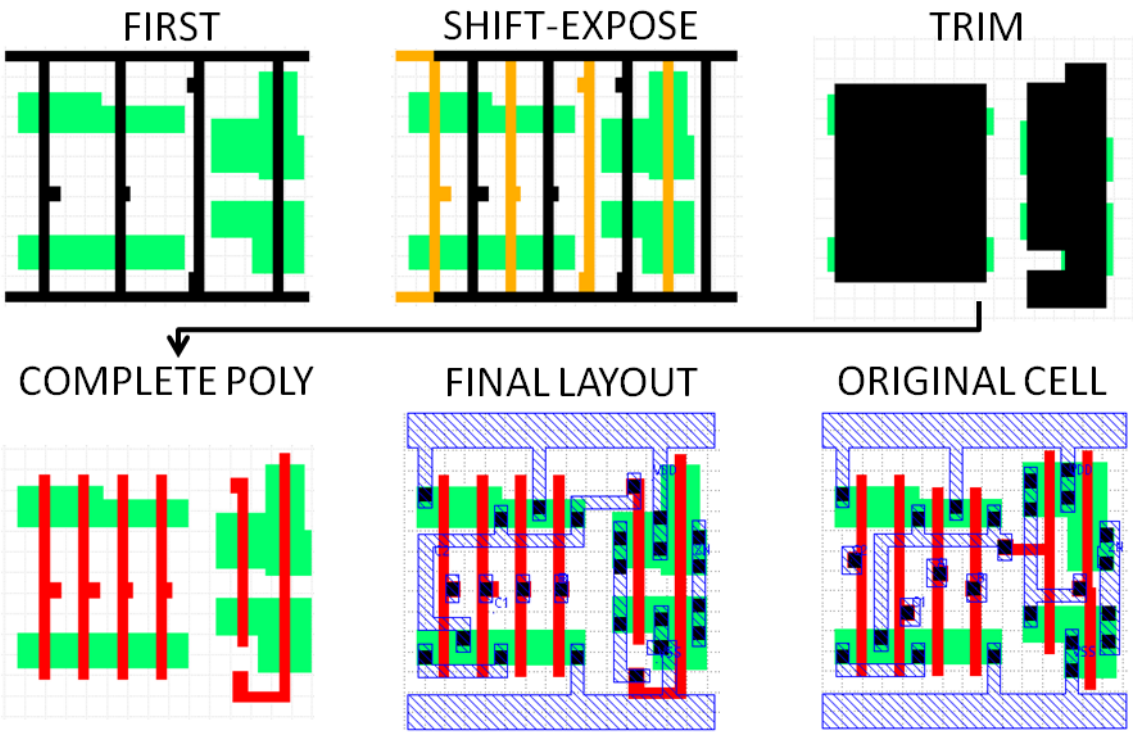


Figure 4.5: Example of 4-input OAI cell layout migrated for the application of STDP at the poly-line layer.

location. Alignment of contacts is not a requirement specific to STDP; it is currently performed to improve patterning quality in state-of-the-art process technologies.

3. The pitch of double-contacts is restricted to the minimum pitch X_0 on the mask (i.e. single-patterning pitch). This restriction makes diffusion double-contacts possible only for very large transistors and poly double-contacts possible only in non-condensed cells with small transistors. This restriction may not be a problem for layout methodologies that avoid double contacts because they worsen channel strain and, hence, device performance.

It is important to note that STDP implementation for fixed pitch grating of contacts (e.g., similar to [Yeh09]) is trivial and requires no extra layout restrictions or redesign efforts.

Figure 4.6 illustrates a 4-input OAI standard-cell layout migration for combined compatibility with STDP at the poly-line and CA layers. Due to the proximity

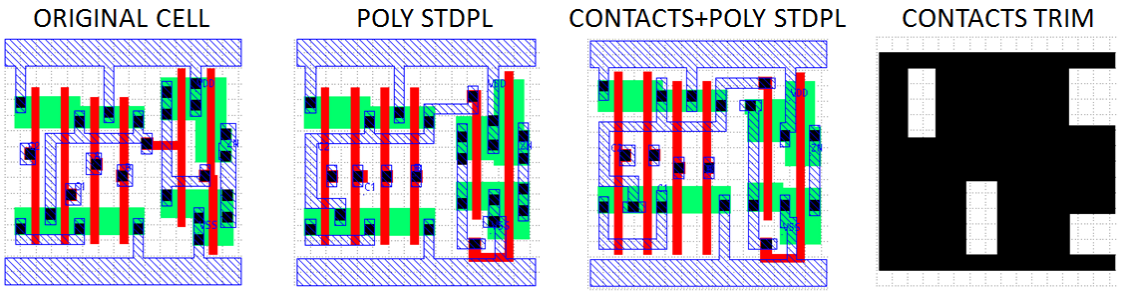


Figure 4.6: Example of 4-input OAI cell layout migrated for the application of STDP at the poly and CA layers with 1.8× pitch relaxation.

of diffusion and poly contacts in the layout and the difference of their depths, it is impractical to form both types of contacts with a single exposure in sub-32nm technologies. As a result, we assume in our STDP implementation that poly and diffusion contacts are formed with two separate exposures. In the example of Figure 4.6, 1.8× pitch relaxation at the poly layer as well as the CA layer are achieved without any area overhead. 2× pitch relaxation at the CA layer, however, leads to a large area overhead as we will show later in Section 4.2.3. In both cases, STDP design implementation at the CA layer leads to a reduced number of double contacts and a negligible change of the diffusion regions of certain transistors. In our implementation, a single trim mask is used for both poly and diffusion contacts patterning as illustrated in Figure 4.6. Unlike the trim mask of the poly layer, the trim mask of the CA layer is very basic and requires no additional changes to the design rules to simplify it.

4.2 STDP Design Implementation

In this section, we demonstrate the application of STDP at the poly layer for standard cell-based designs and extend STDP implementation to the CA layer.

4.2.1 Limitation on Pitch Relaxation of Double Patterning

For cells to be compatible with double patterning, layouts are constructed so that no conflicts occur during decomposition of same-cell features between first and second

exposures. Yet, decomposition conflicts can still occur between features of different cells whenever two cells are placed close to each other. Inter-cell conflicts are handled either by constructing cell-layouts that ensure no conflicts can occur no matter how cells are placed (e.g., as in [HCN09]) or by detailed placement perturbation to resolve conflicts as in [GJK09]. In this work, we follow the former correct-by-construction approach and generate cells that guarantee a conflict-free chip layout. In this approach, the limiting factor for the pitch relaxation that can be achieved in STDP *as well as standard DP* is the spacing between features of different cells assigned to the same exposure. By increasing the spacing between features and the left/right edges of the cell, we can trade area for pitch relaxation. Because the cell width is quantized, a small increase of this spacing can cause a considerable area increase in most cells. In STDP implementation at the poly layer, 1.8× pitch relaxation is achieved without the need to modify the spacing between poly-lines and the cell edge. Whereas, 2× pitch relaxation requires a 40nm increase of this spacing and, thus, causes a significant area overhead. A larger increase of the spacing between contacts and the cell edge is necessary for STDP implementation at the CA layer. Despite this fact, 1.8× pitch relaxation is achieved with a reasonable cell-area overhead as we will show in Section 4.2.3.

4.2.2 Poly-line STDP Standard-Cell Library and Mask Layout Generation

We develop poly-line STDP compatible standard-cell library by manual layout migration of Nangate open cell library [Nana] using FreePDK [Fre] 45nm process design rules. Details on STDP cell library are presented in Table 4.1. Most standard-cells have fairly simple layouts and are made compatible with STDP technology with little or no redesign effort. However, layout migration of large cells, which have a lot of poly landing pads and use poly to route gate signals in the horizontal direction, requires more layout modifications and effort. The primary reason for this complication comes from contact landing pads being printed in the shifted exposure whether they are needed or not. So, unless the part of the line containing the landing pad is

Table 4.1: Poly-line STDP compatible standard-cell library and associated area with 1.8× pitch relaxation (area overhead is quantized because of cell-pitch restrictions).

	Original	**STDP**	**Area**
Cell	**Area** $[um^2]$	**Area** $[um^2]$	**overhead** $[um^2]$
AND2{X2, X4}	1.064/1.064	1.064/1.064	0
AND3{X1}	1.33	1.33	0
AOI211{X1}	1.33	1.33	0
AOI21{X1}	1.064	1.064	0
AOI221{X2}	1.596	1.596	0
AOI222{X2}	2.128	2.128	0
AOI22{X1, X2}	1.33/1.33	1.33/1.33	0
BUF{X1, X2}	0.798/0.798	0.798/0.798	0
CLKBUF{X1, X2, X3}	0.798/1.064/1.33	0.798/1.064/1.33	0
INV{X1, X2}	0.532/0.532	0.532/0.532	0
INV{X4}	0.532	0.532	0
INV{X8}	0.798	1.064	0.266
INV{X16}	1.33	1.596	0.266
NAND2{X1, X2, X4}	0.798/0.798/1.33	0.798/0.798/1.33	0
NAND3{X1}	1.064	1.064	0
NAND4{X2}	1.33	1.33	0
NOR2{X1, X2}	0.798/0.798	0.798/0.798	0
NOR4{X2}	1.33	1.33	0
OR2{X1, X2}	1.064/1.064	1.064/1.064	0
OR3{X2}	1.33	1.33	0
OR4{X2}	1.596	1.596	0
OAI21{X1, X2}	1.064/1.064	1.064/1.064	0
OAI22{X1}	1.33	1.33	0
OAI33{X1}	1.862	1.862	0
OAI211{X1, X2, X4}	1.33/1.33/2.128	1.33/1.33/2.128	0
XOR2{X1, X2}	1.596/1.596	1.596/1.596	0
DFF{X1}	5.054	5.054	0
SDFF{X2}	6.916	6.916	0

trimmed away in the shifted version, enough room must be available so that poly-to-active spacing design rule is not violated. This requires location adjustment of active regions in some cases. For a process enabling trench contacts (e.g. Intel's 45nm process [Mis07]), this complication is eliminated and layout migration can be easily automated. Layout modifications are performed so that transistors width and length are untouched. Yet, few diffusion regions in complicated cells (i.e. flip-flops) have to be increased/decreased by more than 2× to align PMOS and NMOS transistors. Alternatively, diffusion gaps can be introduced/removed and limit the change of the size of diffusion regions.

As discussed in Section 4.2.1, the amount of pitch relaxation on the mask is limited by the spacing between features and the cell edge. The maximum poly-line pitch on the mask that is achieved without changing this spacing is 340nm, which corresponds to a 1.8× pitch relaxation. However, 2× pitch relaxation could still be achieved by increasing this spacing rule from 40nm to 60nm and bearing the associated area overhead.

Layout decomposition into first and second exposures is automated (C++ program based on OpenAccess 2.2 API [opeb]). Since wrong-way poly (horizontal lines of Figure 4.3(a)) is printed in both exposures, the decomposition problem is reduced to assigning gate-poly lines (vertical lines of Figure 4.3(a)) to the two exposures. Traversing each cell in the library from left to right, the following decomposition rules apply:

- if the pitch with the previous line is X, the line is assigned to the shifted-exposure (i.e. second exposure) and the previous line is assigned to the first exposure;
- if the pitch with the previous line is $< X_0$ and different than X, the line is assigned to the first exposure and the previous line is assigned to the second exposure;
- if the pitch with the previous line is $> X_0$, the line can be assigned to either of the two exposures.

Because the trim-mask covers the entire poly-layer in our STDP proposed process, we start by the poly-layer as the base structure of the trim-mask and apply a series of expansions to simplify the mask. Trim-mask structures of two successive gates with pitch $< X_0$ are joined. For gates with larger pitch and gates at the cell-edge, trim-mask structures of each gate are expanded by $S_{min}/2$, where S_{min} is the minimum separation between gates (i.e. X minus gate line-width). This large trim-mask coverage of gates is to have a large resist thickness at sidewalls after development preventing etch interference with gate features under imperfect overlay and etch control (see process details in Figure 4.1). Trim-mask coverage of field-poly is limited to 20nm on all sides to maximize spacing between trim-mask features. Here, sidewall resist thickness requirement is much smaller than in the case of gate-poly because CD control is much less important. Since poly line-ends are formed by printing a long line in one exposure and cutting its ends in another exposure (i.e. trim-exposure), line-end tapering [GJK08] and pull-back (a.k.a. shortening) are substantially reduced [VBR08]. Hence, we assume line-end extension rule, which only addresses trim-to-STI overlay error and possible damage of line-end by etch in STDP, can be reduced from 55nm to 35nm. With this setup, the overall margin of trim-mask overlay error is at least 20nm in X as well as Y directions.

To guarantee an easy-to-fabricate trim-mask and quality trimming, we enforce few design rule restrictions.

1. Poly line tip-to-side and tip-to-tip within-cell spacing rules are increased from 75nm to 140nm.

2. Top/bottom "wrong-way" poly lines used for routing are pushed 35nm toward the center of the cell.

3. Line-ends are extended at most up to the starting location of "wrong-way" lines.

Rule (1) is to ensure reasonable dimensions of the holes in the trim-mask (at least 100nm wide) that can occur in such situations within a cell as illustrated in

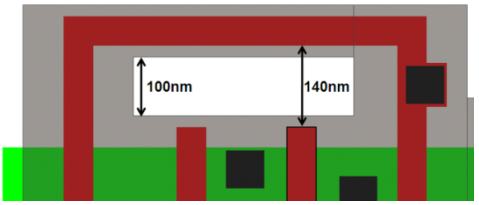


Figure 4.7: Poly line-tip to poly side spacing rule of 140nm to ensure a minimum hole width of 100nm.

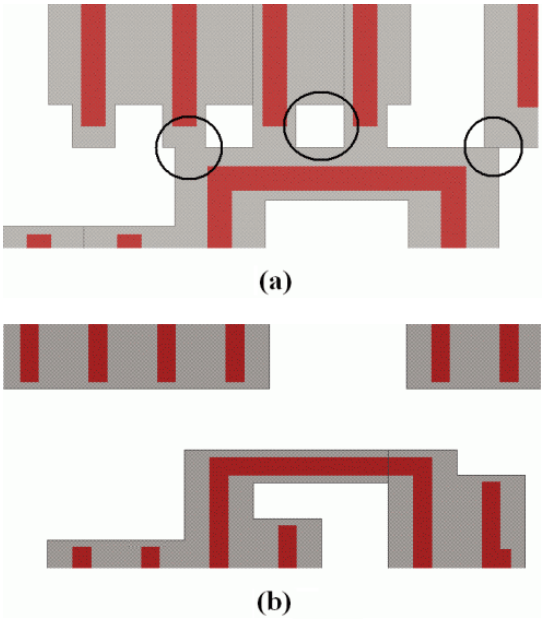


Figure 4.8: Trim-mask complexity at cell-boundaries before (a) and after (b) enforcing rules (2) and (3).

Figure 4.7. Rule (2) and (3) are introduced to avoid small holes in the trim-mask that might occur at cell boundaries, as illustrated in Figure 4.8, resulting in a relaxed separation of at least 100nm between trim-mask features of different cells.

All these rules are specific to FreePDK 45nm process that the cell library is based on and might not be needed for other process technologies. For example, rules (2) and (3) are very likely to be unnecessary (or at least smaller) for commercial processes where line-end gap is considerably larger than the minimum field-poly spacing to meet manufacturability requirements unlike in the case of FreePDK where line-end gap rule is equal to the minimum field-poly spacing. In addition, rules (2) and (3) might be avoided for a cell-library designed for STDP technology rather than migrated from an existing library. In particular, the trim-mask simplification

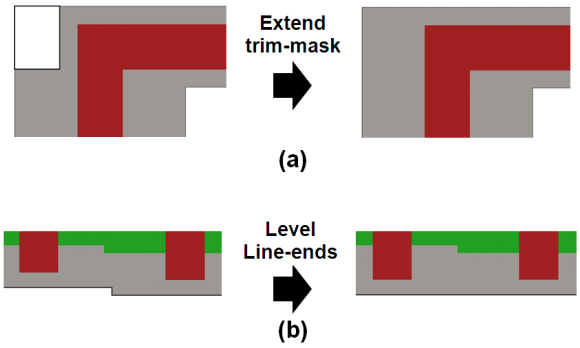


Figure 4.9: Trim-mask notch-fill by (a) trim-mask extension and (b) poly line-end leveling.

at cell-boundaries is better handled during optimization of cell-height and line-end gap rules.

It is important to note that there is a tradeoff between how critical the trim exposure is and the area overhead. Relaxing the values of the design rules listed earlier lead to a less critical trim exposure but might cause an area overhead; whereas, tight design rules lead to a more critical trim exposure but cause no area overhead. In our implementation, we have made a few sensible tradeoff points but others are possible.

A final step of trim-mask simplification is performed to avoid notches wherever possible as illustrated in Figure 4.9. In Figure 4.9(a), notch-filling is performed by extending the trim-mask coverage of field-poly. In Figure 4.9(b), notch-filling is performed by leveling line-ends of neighboring gates (by extension of the shorter line-end).

After generating the different masks for all cells in the library and all possible cell-orientations, mask generation for STDP compatible designs is a simple step. For each cell-instance, cell-type and orientation are determined and mask-features are copied from the corresponding cell in the library to the instance location in the design. The generated mask layout is free of errors at cell-boundaries because critical-mask features outside the cell (or close to the cell-edge) are trimmed away and enough spacing between trim-mask features of different cells is guaranteed by construction.

Table 4.2: Details of poly-line STDP compatible designs showing negligible area overhead.

	Description	**Cell instances**	**Cell-types**	**Flip-flops**	**INV/BUF**	**Cell-area** $[um^2]$	**Area overhead**
mips789	processor core	10529	35	2011	1465	22867.5	0.02%
or1200	combinational logic	3070	35	0	890	3014.8	0.34%
usb	com. controller	478	31	93	52	880.2	0%

(a) (b)

Figure 4.10: Trim-mask layout snippet for poly-lines (a) and contacts (b). In (a), simple large blocks correspond to cells with unidirectional poly-lines and more complex shapes correspond to flip-flops with bidirectional poly-routing. In (b), both flip-flop and combinational logic regions have simple trim-mask features.

Poly-line STDP standard-cell library is implemented with 1.8× pitch relaxation and no area overhead compared to the original Nangate library layouts except for three cells as shown in Table 4.1. This overhead is caused by layout restrictions imposed to simplify the trim-mask. In case these restrictions are avoided for the reasons discussed earlier, none of the STDP compatible cells will have any area overhead. Moreover, if option(b) of Figure 4.3 is used instead of option(a), i.e. having "wrong-way" poly tracks only when needed, STDP implementation of these three inverter-cells results in no area overhead because rule (2) can be avoided.

Three designs from [opea] are synthesized in Cadence RTL Compiler ™v6.2 using the developed poly-line STDP standard-cell library. Designs are placed and routed using Cadence SOC Encounter ™v6.2. Details on the designs and associated *cell-area* overhead are presented in Table 4.2. *Cell-area* overhead for all three designs is negligible (at most 0.34%). The reason is attributed to low utilization of the cells where area overhead occurs (low utilization of large-size inverters is typical).

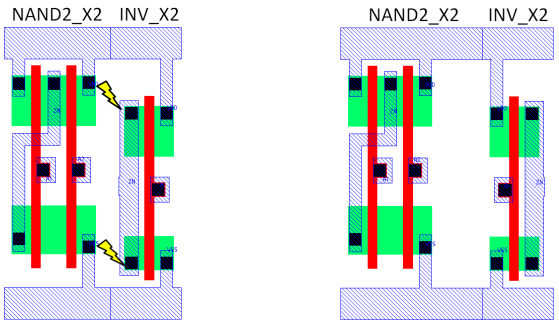
Table 4.3: Details on trim-mask at the poly-line layer for the design of Table 4.2 showing very basic fabrication requirements.

	Line-width [nm]	**Notch Size** [nm]	**Hole dimensions** [nm]	**Overlay margin** [nm]	**Trim-mask fractures**	**Post-OPC poly fractures**
MIPS789	$\geq$ 90	$\geq$ 70	$\geq$ 190 x 145	20	78597	367633
OR1200	$\geq$ 90	$\geq$ 70	$\geq$ 380 x 100	20	5189	43150
USB	$\geq$ 90	$\geq$ 70	$\geq$ 190 x 145	20	2770	14404

Mask layouts are automatically generated for all three designs. A snippet of trim-mask layout at the poly-line layer for the USB design is shown in Figure 4.10(a). In this figure, simple blocks with few vertices correspond to cells with unidirectional poly and more complex shapes correspond to flip-flops involving bidirectional poly-routing. Hence, the trim for purely unidirectional poly designs consists of extremely simple features (large rectangles mostly). Trim-mask complexity is further analyzed. In Table 4.3, we report minimum line-width, notch size, hole dimensions, overlay margin, and number of fractures of the trim-mask. These minimum dimensions are fairly large compared to the minimum feature size of the process (i.e. 50nm) resulting in simple trim-mask for all designs. The dimensions listed in the table are not to be compared directly to dimensions of the critical-mask because trim-mask features do not define patterns but rather protect existing patterns by larger coverage. The number of fractures of the trim-mask (determined using Calibre MDP^{TM}v2008), which affects mask-cost, is 5 to 8 times smaller than the number of feature for post-OPC poly-layer (OPC generated using Calibre OPC^{TM}v2008). In addition, the trim-mask does not require expensive RET features such as OPC and SRAF which substantially increase mask-complexity and cost.

4.2.3 Poly-line Plus Contacts STDP Standard-Cell Library and Mask Layout Generation

For DP implementation (STDP *as well as standard DP*) to be possible at the CA layer without decomposition conflicts, poly contacts need to be well spaced apart from diffusion contacts. In many cells, this can result in a large area increase that makes this approach impractical. Alternatively, poly and diffusion contacts can be



(a) (b)

Figure 4.11: Example of layout decomposition conflicts (marked with the arrows) between features of different cells (a) and our correct-by-construction cell layout that guarantees no conflicts (b).

formed separately with different exposures. In addition to the area benefit, the latter approach has improved printability because the process can be optimized separately for each type of contacts, which have different depths and may even have different shapes and dimensions. The latter approach is assumed in our implementation.

As discussed in Section 4.2.1, to resolve inter-cell layout decomposition conflicts and generate correct-by-construction cell layouts, the spacing between features and the cell edge need to be increased. In case of contacts, this increase is large because contacts can be placed very near the cell edge. To reduce the associated area overhead, we perform the following strategy. We assign the leftmost features in every cell (including cells flipped during placement) to the first exposure. In this way, when abutting two cells side by side, decomposition conflicts can occur only if the rightmost contacts of the left cell is assigned to the first exposure as illustrated in Figure 4.11(a). Therefore, only for those particular cells, the spacing between the rightmost contacts and the right edge of the cell need to be increased as shown in Figure 4.11(b). The amount of this increase to achieve 1.8× pitch relaxation (345nm pitch) at the CA layer is 190nm (or a single cell-width unit in FreePDK design rules).

In STDP implementation at the contacts layer, diffusion contacts are the principal cause of area overhead due to their proximity to the cell edge. To deal with this issue, we introduce hybrid STDP method as a workaround. Only for cells with the rightmost diffusion contacts assigned originally to the first exposure, the method

Table 4.4: Poly-line plus CA STDP compatible standard-cell library and associated area (in $[um^2]$) with 1.8× pitch relaxation. Note that the area overhead is quantized due to cell-pitch restrictions from place and route. O/H stands for overhead.

	Original	**Poly STDP**		**Poly+CA STDP**		**Poly+Hybrid CA STDP**	
	Area	**Area**	**O/H**	**Area**	**O/H**	**Area**	**O/H**
AND2{X2}	1.064	1.064	0	1.064	0	1.064	0
BUF{X2}	0.798	0.798	0	1.064	0.266	0.798	0
CLKBUF{X2}	1.064	1.064	0	1.064	0	1.064	0
INV{X2}	0.532	0.532	0	0.532	0	0.532	0
INV{X4}	0.532	0.532	0	0.532	0	0.532	0
NAND2{X2}	0.798	0.798	0	1.064	0.266	0.798	0
NOR2{X2}	0.798	0.798	0	1.064	0.266	0.798	0
OR2{X2}	1.064	1.064	0	1.064	0	1.064	0
OAI211{X4}	2.128	2.128	0	2.128	0	2.128	0
XOR2{X1}	1.596	1.596	0	1.596	0	1.596	0
DFF{X1}	5.054	5.054	0	5.586	0.532	5.586	0.532
SDFF{X2}	6.916	6.916	0	6.916	0	6.916	0

consists of assigning these contacts to be patterned with the poly contacts in a separate exposure. Consequently, no cell will have the rightmost contacts assigned to the first exposure and there is no need to increase the spacing between diffusion contacts and the cell edge as in the original CA STDP implementation. It is also worth noting that, by preventing the placement of poly contacts between the horizontal location of the first/last gate and the cell edge, STDP can be safely employed to form poly contacts without causing any area overhead.

For the two different styles (STDP and hybrid STDP), we construct a small set of compatible cells by manual layout migration of poly-line STDP compatible cells that were presented in Section 4.2.2. Table 4.4 summarizes the implemented cells and the area overhead associated with each style. Unlike STDP at poly-line layer, layout migration for STDP implementation at the CA layer is a difficult task. The main reason for this complication is that contacts that are shifted replica of one another must be perfectly aligned at the same vertical location and at a pitch equal to the amount of mask shift X. Layout modifications are performed so that

Table 4.5: Cell-area (in $[um^2]$) of three designs synthesized using Poly plus CA STDP compatible cells for the different schemes. O/H stands for overhead.

	Cell instances	Cell-types	Flip-flops	INV/BUF	Original Area	Poly STDP		Poly+CA		Poly+Hybrid CA	
						Area	O/H	Area	O/H	Area	O/H
MIPS789	20192	12	2011	3878	29733	29733	0%	33289	12%	29931	0.7%
OR1200	4240	10	0	889	3759	3759	0%	4436	18%	3759	0%
USB	674	10	93	49	1011	1011	0%	1154	14.2%	1058	4.7%

transistors width and length are untouched. Diffusion length have to be increased by 5nm to ensure contacts are at a distance equal to X. Because double-contacts pitch is restricted to the single-patterning pitch, the use of double contacts was not possible for small-size transistors with our STDP design implementation.

We synthesize the three designs used in the implementation of STDP at the poly-line layer (shown in 4.2 using poly-line plus CA STDP compatible cells of Table 4.4. Cell-area of the three designs after placement and routing for the different STDP flavors are presented in Table 4.5. Implementation of STDP at poly-line and CA layers (Poly+CA STDP) results in 12 to 18% cell-area increase. Whereas, poly-line plus hybrid CA STDP implementation (Poly+Hybrid STDP) results in a negligible cell-area increase except for USB design (4.7% increase) where the area of flip-flops constitute the largest part of the design area. It is clear from the results that the area overhead of CA STDP is mainly caused by the cell-extension rule we enforce to avoid any possibility of inter-cell decomposition conflicts. This is to say that, if such conflicts are left for the placer to handle, the cell area overhead would be the same as the affordable area overhead associated with Poly+Hybrid CA STDP.

Generation of mask-layout at the CA layer is similar to that of mask-layout at the poly-line layer discussed in Section 4.2.2. A snippet of trim-mask layout at the CA layer for the USB design is shown in Figure 4.10(b). Here, the trim-mask layout constitute of simple features with a few number of vertices. Trim-mask complexity is further analyzed for each design. In Table 4.6, we report minimum line-width, notch size, hole dimensions, and the overlay margin. These minimum dimensions are fairly large compared to the minimum feature size of the process (i.e. 65nm) resulting in simple trim-mask for all designs. The dimensions listed in the table are not to be compared directly to dimensions of the critical-mask because trim-mask features do

Table 4.6: Details on trim-mask at the contacts layer for the design of Table 4.2 showing very basic fabrication requirements.

	Line-width [nm]	**Notch Size** [nm]	**Hole dimensions** [nm]	**Overlay margin** [nm]
MIPS789	$\geq$ 100	$\geq$ 165	$\geq$ 155 x 155	65
OR1200	$\geq$ 160	none	$\geq$ 200 x 195	65
USB	$\geq$ 100	$\geq$ 165	$\geq$ 155 x 155	65

not define patterns, but they are rather used to protect existing patterns by larger coverage. Moreover, the trim-mask does not require expensive RET features such as OPC and SRAF, which substantially increase mask-complexity and cost.

4.3 STDP Benefits

In addition to cutting mask-cost to nearly half that of standard DP because of critical-mask reuse for both exposures and a cheap trim-mask as shown in Section 4.2, STDP has many benefits over standard pitch-split DP in terms of overlay and CD control and throughput.

4.3.1 Overlay and Throughput Benefits

The negative dual-trench double exposure process (shown in Figure 4.1) has higher throughput than other processes [Lee08]. This process does not require wafer removal from the exposure tool chuck between the two exposures (as illustrated in Figure 4.1). In case such process becomes feasible, its implementation in combination with STDP allows the second exposure to be performed after a blind translational shift without any alignment of the second exposure. This practically eliminates any overlay error between the two patterns and, also, saves alignment time.

An important source of overlay is reticle metrology errors [SS07], which is caused by reticle mounting and heating as well as particle contamination of the reticle alignment marks. Since mask loading and unloading between both exposures is not necessary in STDP, this source of overlay error is virtually eliminated in all STDP process implementations. Moreover, reticle alignment, which is another source of

Table 4.7: Summary of STDP overlay benefits.

Source	**Benefit**
All sources	almost eliminated in case of negative LLE
Reticle/mask related	eliminated for all STDP processes
Reticle alignment	reduced for all STDP processes
Wafer stage	not affected

overlay, is again eliminated in all STDP processes for the same reason. The time spent on mask loading/unloading as well as reticle alignment is saved.

A major source of overlay is registration error ($\approx 25\%$) [Lee06]. In DP, registration error of the two exposures is observed to be correlated and, as a result, the impact on overlay is greatly reduced. This correlation is mainly attributed to mask-layout similarity [Lee06, BKP09]. In STDP, registration error is expected to have a higher correlation factor than in the case of standard pitch-split DP since mask-layout is exactly the same for both exposures.

A summary of STDP overlay benefits over standard pitch-split DP is given in Table 4.7.

4.3.2 Alleviating CD Bimodality Problem

Whenever two patterns are formed in different exposure and etch steps, lines and spaces have bimodal CD distributions [Fin08] that can have severe implications for the digital design flow [JK09]. Because the same mask is used for both exposures in STDP, mask CDU, which is the second most important contributor to the overall CD variation as reported in [Fin08, Dus07], no longer affects the difference between the two distribution and the bimodal problem is alleviated.

Considering CD of the first (CD_a) and second (CD_b) patterns as random variables, then

$$\begin{aligned} CD_a &= \mu_a + m_a + nm_a, \\ CD_b &= \mu_b + m_b + nm_b, \end{aligned} \tag{4.1}$$

where μ_a and μ_b are the mean of CD_a and CD_b respectively, m is mask CDU random

variable and nm (short for non-mask) is a random variable corresponding to all other contributors to line CDU. Assuming CD_a, CD_b, and all other random variables of Equation 4.1 have independent normal distributions in standard-DP, the covariance of the two CD distributions is zero and CD difference has a normal distribution with $\mu_{diff} = \mu_a - \mu_b$ and $\sigma_{diff} = \sqrt{\sigma_a^2 + \sigma_b^2}$, where σ_a and σ_b are the standard deviations of CD_a and CD_b distributions respectively. In case of STDP, $m_a = m_b = m$ and, consequently, the covariance is

$$\begin{aligned} Cov(a,b) &= E(a.b) - \mu_a \times \mu_b \\ &= \mu_a \times \mu_b + \mu_a(m + nm_b) + \mu_b(m + nm_a) \\ &\quad + m(nm_a + nm_b) + nm_a \times nm_b + m^2 - \mu_a \times \mu_b \\ &= \mu_a(m + nm_b) + \mu_b(m + nm_a) \\ &\quad + m(nm_a + nm_b) + nm_a \times nm_b + m^2. \end{aligned} \tag{4.2}$$

Since m, nm_a, and nm_b have zero mean, Equation 4.2 simplifies to

$$Cov(a,b) = m^2 = \sigma_m^2, \tag{4.3}$$

where σ_m is the standard deviation of mask CDU normal distribution. The distribution of CD difference has $\mu_{diff} = \mu_a - \mu_b$ and $\sigma_{diff} = \sqrt{\sigma_a^2 + \sigma_b^2 - 2Cov(a,b)} = \sqrt{\sigma_a^2 + \sigma_b^2 - 2\sigma_m^2}$ (from Equation 4.3).

Using line-CDU breakdown values for LELE positive dual-line 32nm process from [Fin08] (i.e. 2.7nm 3σ from etch, 1.4nm 3σ from mask-CDU, 0.7nm 3σ from dose, and 0.5nm 3σ from focus), σ_{diff} is 1.49nm in the case of standard-DP and 1.34nm in the case of STDP which corresponds to a 10.3% reduction in standard deviation.

4.3.3 Comparison with Popular Double-Patterning Technologies

In this section, STDP technology is compared to other popular patterning techniques including standard-DP. A summary of attributes is presented in Table 4.8. STDP has advantages over standard-DP as discussed earlier. The drawbacks of STDP in this

Table 4.8: Summary of comparison between standard-DP and STDP methods.

	Standard-DP	**STDP**
Pitch doubling	yes	yes
Mask-cost	high	reduced
Trim exposure	no	yes
Area overhead (for 2D layouts)	small	small
Designing effort (for 2D layouts)	easy	hard
Decomposition conflicts (for 2D layouts)	yes	eliminated
CD bimodality	yes	reduced
Same-layer Overlay	yes	reduced
Throughput of critical scanner	low	slightly improved

comparison are higher redesign effort and the use of a trim-exposure. Because STDP layouts are very regular and successive features are perfectly symmetrical, STDP designs are compatible with self-aligned double patterning (SADP) technology and require little mask-assignment effort. Hence, cell/block reuse from one technology to the other is possible. Trim-exposure non-criticality allows its processing on less expensive fabrication-lines and its use permits the reduction of line-end extension rule and the elimination of layout decomposition conflicts as discussed earlier in this chapter.

Another popular double-patterning technology is subtractive-litho of [Mis07]. Essentially, subtractive-litho consists of printing a grating and removing dummy-poly with a trim-exposure. Subtractive-litho is preferred over DP (standard and STDP) because it has lower cost and less process control requirements. Although the poly grating can cause an area overhead subtractive-litho improves printability due to its imposed regular layout, this method does not achieve pitch-doubling that might be necessary to scale down to future technology nodes (beyond 32nm). Subtractive-litho can also suffer from a considerable area overhead when a poly grating is imposed as reported in [GG09a, MO07, SSW08].

Conclusions

Since extreme ultraviolet lithography (EUV) and other next-generation lithography technologies such as nanoimprint and electron beam direct write [ITRb] will unlikely be ready for volume manufacturing at the 20nm node [Edn, LaPb], there is a trend toward regular and gridded layouts that allow the continuation of scaling using 193nm wavelength optical lithography [Mis07, SLC08a, PSS03, LPH09]. STDP is very suitable for such layout style and the industry can make use of this lithography technique to reduce the mask-cost and improve process control of double-patterning. With ever increasing mask cost [ICK], cutting this cost by almost half that of standard double patterning – achieved using STDP – can significantly reduce the production cost especially for low-volume manufacturing. If a negative dual-trench double exposure process becomes feasible, its implementation with STDP can virtually eliminate within-layer overlay errors. Although it may be argued that good printability can be achieved by good optimization of the illumination source in gridded layouts, we believe that pitch relaxation through double patterning will be inevitable in future technology nodes. In this chapter, we also demonstrate the viability of STDP in conventional bidirectional layouts. While guaranteeing no inter-cell layout decomposition conflicts, design implementation of STDP at the poly-line layer is achieved with 1.8× pitch relaxation with virtually no area overhead; the same pitch relaxation is achieved for STDP combined implementation at poly-line and CA layers with no more than 4.7% *cell-area* overhead.

Acknowledgements

This work was generously supported in part by IMPACT UC Discovery Grant (http://impact.berkeley.edu) contract number ele07-10291, Semiconductor Research Corporation contract number 2008-TJ-1816, and NSF CAREER Award number 0846196.

Part II

Design-Centric Assessment of Technology

Chapter 5

A Framework for Early Co-Evaluation of Design Rules, Technology Choices, and Layout Methodologies

Design rules have been the primary contract between technology developers and designers and are likely to remain so to preserve abstractions and productivity. While current approaches for defining design rules are largely unsystematic and empirical in nature, this chapter offers a novel framework for early and systematic evaluation of design rules and layout styles with the intent of speeding up the rules-development cycle. The framework essentially creates a virtual standard-cell library and performs the evaluation based on the virtual layouts. The evaluation is expressed in terms of major layout characteristics including area, manufacturability, and variability. Due to the focus on the exploration of rules at an early stage of technology development, we use first order models of variability and manufacturability (instead of relying on accurate simulation) and layout topology/congestion-based area estimates (instead of explicit and slow layout generation). Such a framework can be used to co-evaluate and co-optimize design rules, patterning technologies, layout methodologies, and library architectures.

Introduction

The semiconductor industry is likely to see several radical changes in the fabrication and device technologies during this decade. On the patterning front, disruptive changes include the adoption of one or more of candidate next-generation lithog-

raphy techniques such as nanoimprint, electron beam direct write, and extreme ultraviolet [ITRb, Yon08, Mar08, Mei09]. Each of these has challenging implications for layout methodologies and design rules (DRs). Resolution enhancement techniques (RETs) and other patterning solutions such as immersion and double-patterning technology (DPT), off-axis illumination (OAI), sub-resolution assist features (SRAFs), and phase-shift mask (PSM) require additional layout-restrictive DRs [Lai08, Mac08, Bai07, Haz08b, Bal08, Lie03, Lie04]. Therefore, *early assessment of design restrictions imposed by technological choices is absolutely essential.*

DRs are the biggest design-relevant quality metric for a technology. Even small changes in DRs can have significant impact on manufacturability [ZCY08] and circuit characteristics including layout area, variability, power, and performance [JCS08, She05]. Unfortunately, even after DRs have existed for decades, design rule evaluation and exploration is largely unsystematic and empirical in nature. Several published works have done empirical "one-at-a-time" evaluation of design rules [CGK04, ZCY08]. For example, the work in [GJK08] electrically evaluates line-end extension rule and conclude that it may be too conservative. Other recent works [DCY09, Cha09] offer solutions to explore DRs from a pure printability perspective and do not examine the effects of DRs on circuit characteristics. Moreover, none of these methods account for layout topology changes that may happen when the DR values change significantly. They also ignore several practical constraints imposed on layouts by the standard-cell design methodology (e.g., cell width and height quantization). Finally, these approaches are based on explicit layout generation and lithography simulation, which makes them slow and dependent on the models accuracy.

To the best of our knowledge, this work proposes the first framework to systematically and qualitatively explore area-manufacturability-variability tradeoffs in design rules. Rather than fine-tuning DRs, our goal is to make early decisions *before* exact process and design technologies are known. At this stage, accurate evaluation methods and models are unlikely to be available and the return on investment of using them is fairly low. Unlike other approaches that rely on layout generation or

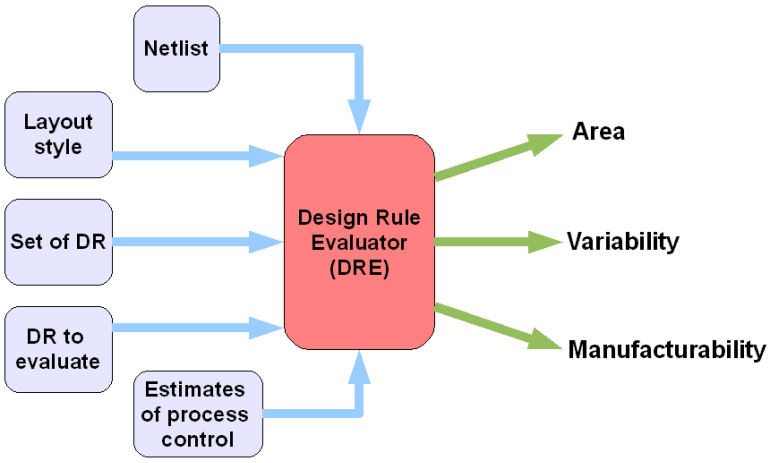


Figure 5.1: Overview-diagram of DRE framework.

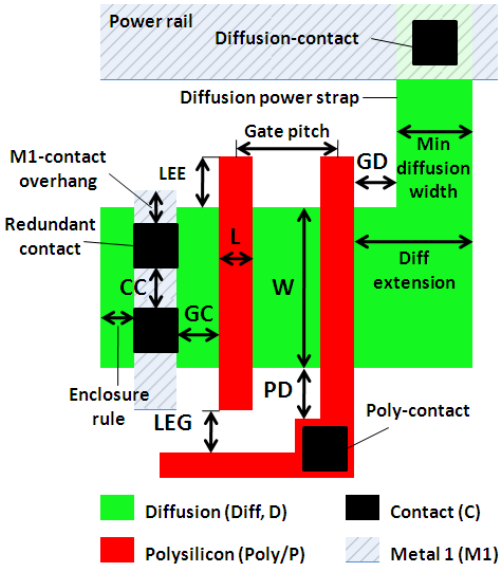
perturbation (e.g., [SSW08, Kob08]), we use simple but justified approximations for manufacturability and variability. Because the search space of DRs is very large, we use fast layout topology generation methods to estimate area as opposed to full-blown layout generation. The former has been shown to produce high accuracy allowing for explicit "layout style" guidelines, as we show later in this chapter.

The structure of the proposed cell-level DR Evaluator (DRE) is depicted in Figure 5.1. The framework takes the following inputs: circuit netlists (e.g., SPICE) of cells (possibly scaled down from a previous technology generation), layout style and preferences (e.g., redundant contacts), DRs and their values (see Figure 5.2), estimates of process control (e.g., overlay error distribution), and benchmark designs (specified as cell usage statistics) to evaluate the rules on. In DRE, only the values of DRs to be evaluated are modified while all other rules remain unchanged. This modified set of DRs is then used to estimate the layout and determine major metrics of area, manufacturability, and variability[1].

We make the following contributions.

- We offer a framework for *fast*, *early* and *systematic* collective evaluation and exploration of DRs, layout styles, and library architectures. The framework

[1]DR choices also affect delay, power, reliability, and designability. Evaluating these aspects of DRs is part of our future work.



DR	Value[nm]
L	$\geq$ 50
M1 1/2 pitch	$\geq$ 65
Gate pitch	$\geq$ 190
Min diff width	$\geq$ 90
LEE	$\geq$ 55
LEG	$\geq$ 75
GC	$\geq$ 35
CC	$\geq$ 75
GD, PD	$\geq$ 50
M1-C overhang	$\geq$ 35
Diff enclosure	$\geq$ 5
Poly enclosure	$\geq$ 5
Diff extension	$\geq$ 70

Figure 5.2: Illustration of major DRs, their notations and values in FreePDK 45nm process [Fre].

makes DR generation and optimization easier and much faster. Rather than exploring the entire search space of DRs with conventional compute-expensive methods, the framework can be used to quickly eliminate poor DR choices.

- We evaluate some major DRs and layout style decisions such as: 1D and 2D poly, multiple and fixed-pitch poly, diffusion and metal 1 (M1) power-straps, and cell height.

- We demonstrate through case studies the use of the framework to explore DRs and compare processes from the design perspective.

The remaining chapter is organized as follows. Section 5.1 describes the methods used for layout topology generation as well as metal-congestion estimation and its impact on the layout area. Sections 5.2 and 5.3 provide details on the models and metrics used for manufacturability and variability. In Section 5.4, comparative evaluations of several DRs are performed in a 45nm process. In addition, we analyze area-manufacturability-variability tradeoffs of a commercial standard and a low power 65nm process and illustrate the use of our framework for the collective exploration of DRs. Finally, Section 5.4.6 summarizes our findings and presents directions of future research.

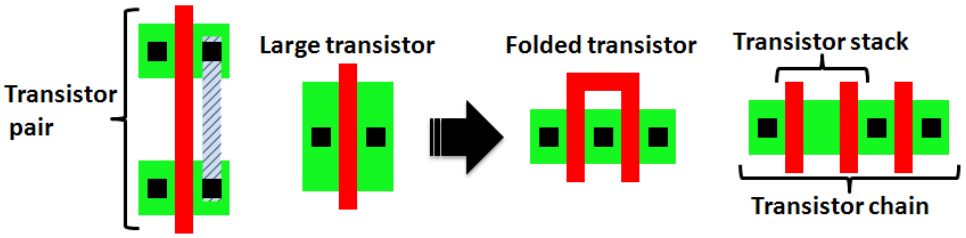


Figure 5.3: Techniques and notations used in layout topology generation.

5.1 Area Estimation

The number of design rules is growing tremendously and design rule manuals (DRM) are becoming unmanageable as we move toward smaller feature sizes [LPH09, YCS06]. In addition, DRs need to be evaluated individually as well as collectively over a wide range of values. As a result, our framework was designed for the *fast* evaluation necessary to enable DR exploration/optimization.

This section describes the methods used in the DRE framework for the fast layout topology generation and metal-congestion estimation.

5.1.1 Layout Topology Generation

Major transistor placement techniques used for layout-area reduction are highlighted in Figure 5.3. Transistor pairing consists of placing two inter-connected transistors, one pMOS and another nMOS, on the same column to minimize wire length and facilitate routing as well as to ensure more layout regularity. The coupled pMOS/nMOS transistors are referred to as transistor pairs. Transistor folding consists of replacing a large transistor by equivalent multiple transistors of smaller sizes connected in parallel. Transistor chaining is the process of abutting transistors of the same type by sharing the same diffusion area. Non-isolated transistors of the same active region form a transistor chain. A transistor stack refers to two transistors sharing a diffusion area that is not connected to any other parts of the circuit (i.e. contact-free diffusion).

Figure 5.4 outlines the flow of transistor placement used in our layout estimation

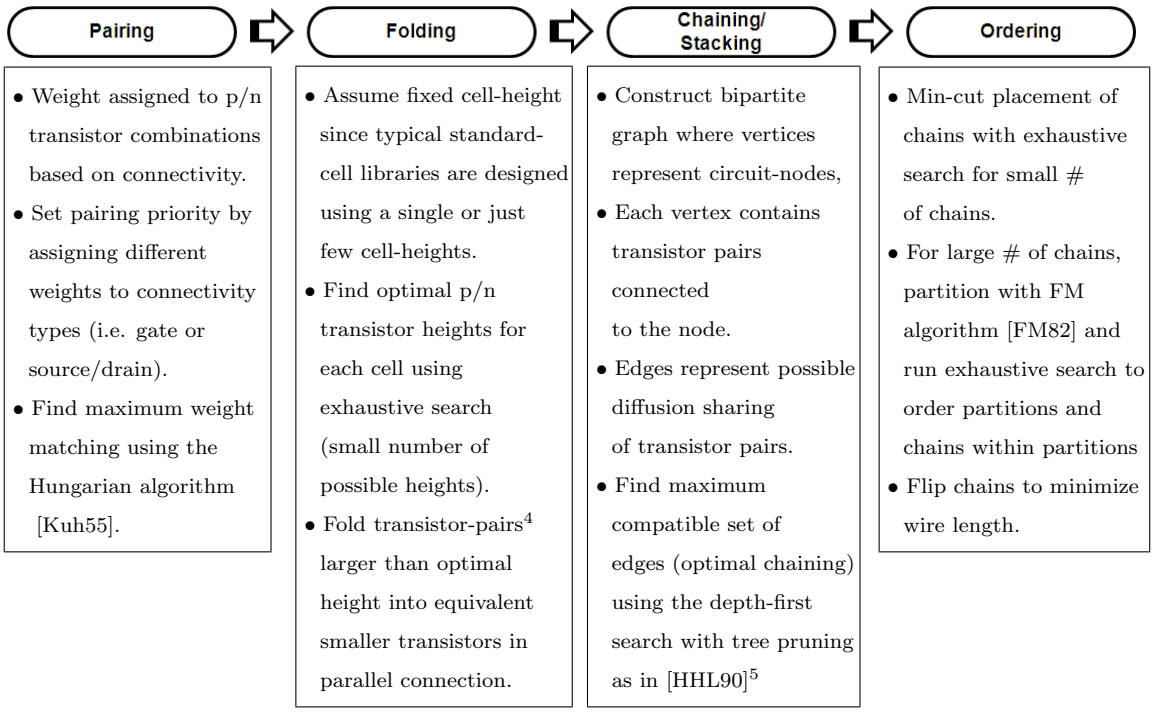


Figure 5.4: Flow of layout topology generation in DRE framework.

and describes the algorithms used at each step. We illustrate the application of these steps on a standard-cell in Figure 5.5.

The first step is transistor pairing. A score is assigned to each pMOS/nMOS transistor combination based on the connectivity and the pairing problem is reduced to finding the matching with the maximum score. According to the layout style, different scores can be associated with different types of connections (i.e. gate or source/drain) to set the pairing priorities. Sharing of the gate signal is typically preferred over source/drain signals to save on contacts and the congestion they induce. This matching problem is solved in DRE optimally using the Hungarian algorithm [Kuh55].

Transistor folding is performed next. A transistor with width larger than its network (pMOS or nMOS) height must be folded into multiple transistors in parallel connection with the same total width. Therefore, the ratio of the pMOS network height to nMOS network height affects the total number of pairs after transistors are folded. Because the cell height is fixed and layout dimensions are quantized to

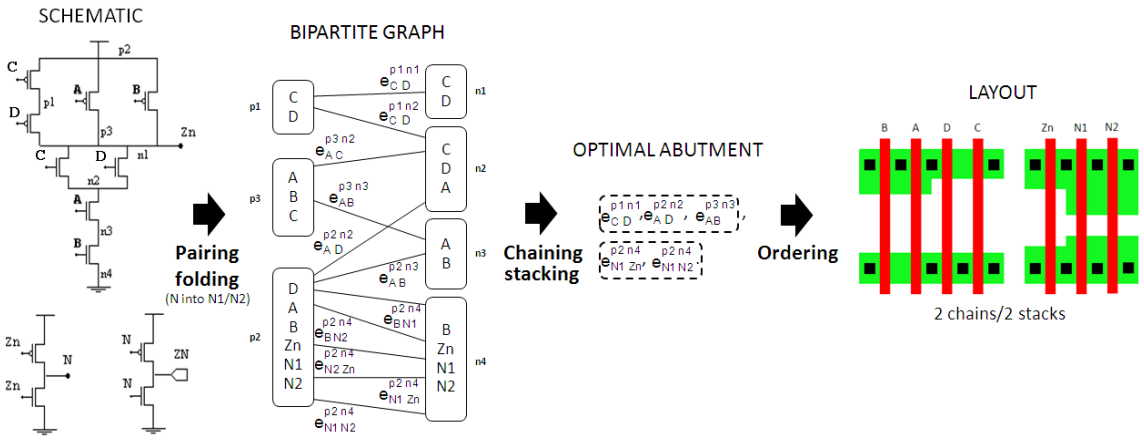


Figure 5.5: Example that illustrates our layout topology generation for a 4-input OAI standard-cell.

the manufacturing grid size, there is a limited number of the possible pMOS/nMOS network-height ratio. So, we determine the total number of pairs associated with each ratio; the ratio leading to the minimum number of pairs (least folds) is set for each cell. After the pMOS/nMOS network heights are decided, wide transistors that exceed the height of the corresponding network are actually folded.

The layout topology generation continues with the step of transistor chaining. The fast method discussed in [HHL90] is implemented to perform the chaining. In this method, the cell circuit is represented as a bipartite graph. Vertices represent nodes in the circuit and each vertex contains all transistor pairs connected to its corresponding node. Edges represent possible abutments of transistor pairs. A depth-first search with tree pruning is used to find the maximum compatible set of edges, which corresponds to the optimal chaining. Solutions with the higher upper bound on the number of realizable abutments are examined first and we found that the optimal solution (i.e. the same chaining as in actual layouts) is reached in almost every case after examining the first few solutions. Thus, we have limited the number of iterations[2] to make the algorithm run faster. Folds of the same transistor are treated as independent transistors and, in some cases, might end up abutted to different transistors and separate from each other to improve the chaining solution.

[2]Twenty eight iterations for cells with more than 20 transistors and six hundred iterations for smaller cells.

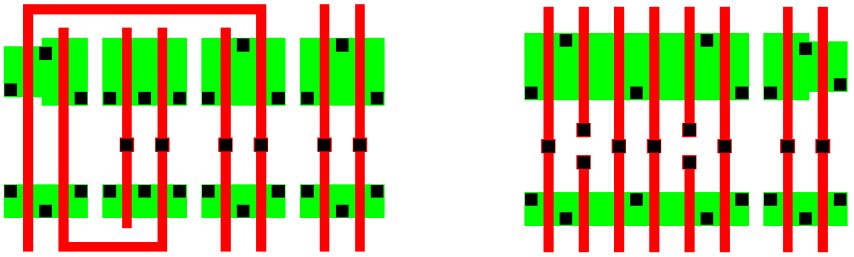
When transistors are folded into large number of folds however, this practice no longer improves the chaining solution and makes the algorithm run much slower. As a result, we cluster large number of folds belonging to the same transistor into groups that we treat as single transistors during chaining. Transistor stacking is considered a special type of chaining. Stacks have an advantage over regular chaining in that they do not need a contact and, consequently, might improve the layout density in some process technologies[3]. Therefore, if multiple chaining solutions have the maximum number of abutments, we pick the one with the maximum number of stacks.

The ordering of transistors within chains is inferred from the abutments associated with the chaining solution that was picked. Chains are then ordered linearly in a row following the familiar 1D placement. The problem is formulated as a *min-cut placement* to minimize the overall wire length. In case the number of chains is small, we run exhaustive search to find the optimal solution; otherwise, we partition the graph of chains using the Fiduccia-Mattheyses (FM) algorithm [FM82] and run exhaustive search to first find the optimal order of partitions and, then, the optimal order of chains within each partition. Once the ordering is complete, chains' orientations are possibly flipped across the Y-axis to reduce the overall wire length further.

The exact transistor and pin locations along the horizontal direction are then determined based on minimum DR dimensions. As for transistor locations along the vertical direction, we consider three possibilities: (a) as near as possible to power rails, (b) exactly in the center of p/n networks, and (c) as near as possible to p/n interface. The choice of vertical location of transistors is regarded as a layout style, which can also be evaluated by the DRE framework[4].

[3]The minimum gate pitch is typically smaller than the contacted gate pitch unless a fixed-pitch poly style is adopted.

[4]This decision has implications on M1-congestion as well as the impact of stress and well-proximity effect on performance [JCS08, She05].



(a) Perfect pairing, area=3.192 (b) Imperfect pairing, area=2.66

Figure 5.6: Example illustrating imperfect pairing and its associated tradeoffs for a DLH_X2 cell layout.

5.1.2 Tweaking Pairing

The pairing step results in pMOS/nMOS pairs with the largest number of shared signals and preference to the sharing of the gate signal. In practice, layout designers may introduce small tweaks on pairing to improve the chaining solution (i.e. reducing the front-end area) as shown in the example of Figure 5.6. Therefore, we introduce an additional step just after the first pairing to perform such tweaks automatically.

Tweaking of the pairing solution is performed using a greedy algorithm. Given the initial pairing solution, we pre-compute for each pair the number of connections with the other pairs that can be performed in both nMOS and pMOS sides, i.e. the number of possible abutments with the other pairs. We then check if the switching of the transistors of any combination of pairs can improve the number of possible abutments. The switch with the best improvement is performed and the involved pairs are prevented from future switching. This process is repeated until all switches with improvement are performed or until all pairs have been switched.

The downside of imperfect pairing is that it requires more spacing between the pMOS and nMOS transistors than in the case of perfect pairing (see Figure 5.6). This extra spacing requirement can result in higher number of folds in some cases. Therefore, we determine, based on the target pMOS to nMOS transistor height ratio of the library, the expected number of folds before and after a transistor switch is made. The switch is prevented if it is expected to cause a larger number of folds. It is worth noting that the extra spacing requirement for imperfect pairing reduces the

available wiring tracks in the top and bottom channels of the Poly (or horizontal local interconnect) layer as shown in Figure 5.6.

5.1.3 Routing Estimation

Once transistor placement is complete, locations of gates and contacts to the gates and transistor source/drain (S/D) terminals are determined. S/D contacts connected to power supply are located as close as possible to the power rail without violating DRs. All other S/D contacts are located near the p/n interface to reduce the length of wires necessary to connect transistors from the nMOS network to transistors from the pMOS network. Contacts to gates (poly contacts) are placed at the p/n interface of the cell (y-coordinate) and the same horizontal locations (x-coordinate) of the gates that they connect to.

Rather than performing actual routing, we estimate the routes and model metal congestion with the goal of considering its effect on layout area. Estimating routes is preferred over performing actual routes for three reasons. First, different automated tools and layout designers can reach completely different routes and a small change in the DRs may result in very different routing solutions. On the other hand, estimated routing is generic, meant to assess the quality of rules, and is not affected by small DR changes. Second, performing actual routing is very time consuming and can be a runtime bottleneck for our automated evaluation. Third, introducing new rules or layout styles may require a significant reimplementation of a router; this problem is much less severe for smart congestion estimation. Hereafter, the term "routing" denotes "estimated routing" and not actual routing.

Transistor interconnections, i.e. intra-cell routes, are assumed to be performed using polysilicon (poly), diffusion for power connections (i.e. power straps) if dictated by the layout methodology, the first metal (M1) layer, and possibly the second metal layer (M2) if accessible for cell design. There are three types of connections: gate-to-gate, S/D-to-gate, and S/D-to-S/D. The way gate-to-gate connections are performed depends on poly-routing restrictions, which are characterized by the lay-

out style.

Three configurations of poly-routing are allowed: no poly-routing (1D), limited poly-routing, and unrestricted poly-routing (2D). In case no poly-routing is allowed, poly is used only to connect dual gates (i.e. gates of same transistor-pair). Connections between any other gates need to be performed with metal layers. In case poly-routing is limited, poly is used to connect adjacent gates in the same network (pMOS or nMOS) in addition to dual gates. In case poly-routing is not restricted, all gate interconnections are performed on the poly layer unless routing is infeasible due to congestion or blocking active layer. Since routing resources are limited, we give priority for routing longer nets to maximize poly utilization. Horizontal wiring on poly uses the available tracks of p/n routing channels at the top and bottom of the cell with the exception of wiring used to connect adjacent gates of folded transistors (a.k.a. fingers), which are assumed to occupy the routing-channel at the p/n interface in the center of the cell. Excluding finger interconnection, there are three cases for gate-to-gate routing involving horizontal wiring not to be possible on the poly layer. The first case is when diffusion power-straps block both the top and bottom routing channels. The second case is when the access to the gate from the top and bottom channels is blocked by nets previously routed on poly. The last case of infeasible poly routing occurs if at any location between source and destination along both top and bottom channels all horizontal poly tracks are occupied.

Hypernets involving S/D-to-gate connections are decomposed into a single metal segment and one or more poly segments depending on poly-routing restrictions. In this case, a single poly-contact is added per poly segment and it is placed along the y-coordinate of the p/n interface at the same x-coordinate of the nearest gate it connects to (see the poly contact in the example of Figure 5.7.

Metal segments of S/D-to-gate and S/D-to-S/D connections are made with metal layers. If all pins have the same x or y-coordinates, the route is performed using a vertical or horizontal wire connecting all pins. For nets involving pins at different x or y-coordinates, we assume they can be routed using a single-trunk Steiner tree as shown in Figure 5.7. Single-trunk Steiner tree routing is common in real layouts

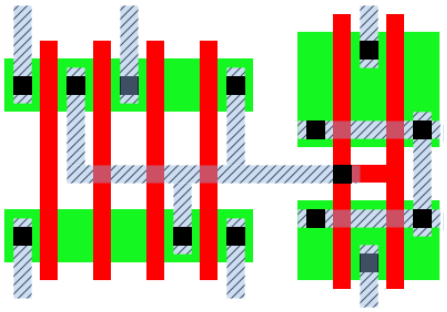


Figure 5.7: S/D-to-gate interconnections may be routed on M1 or poly layers and S/D-to-S/D interconnections may be routed on M1 or M2 layers. We assume a single-trunk Steiner tree for routing.

Table 5.1: Shape count and locations assumptions for nets that cannot be connected with a straight line.

Shape	**Count**	**Location**
Tip	# of pins	Fixed at pin locations with bounding box including the tip in its all possible orientations
L-shape	two	Anywhere in net's bounding box
T-shape	# of pins minus two minus # of crosses	Anywhere in net's bounding box
Cross	Special case of T-shape, detected based on x coordinates of pins	Anywhere in net's bounding box
Line	One horizontal and one vertical if space permits	Anywhere in net's bounding box

and we avoid fixing the trunk to an exact location to keep the routing estimation generic. With this assumption, we can determine what shapes are involved in each route based on the number of pins. The shape count is summarized in Table 5.1 and an illustration example is shown in Figure 5.8. The wire length is estimated as the half-perimeter of the bounding box.

There are three configurations for metal layer assignment:

1. 2D M1 with prohibited access to M2 layer for intra-cell routing;

2. 2D M1 with use of M2 to resolve M1 congestion;

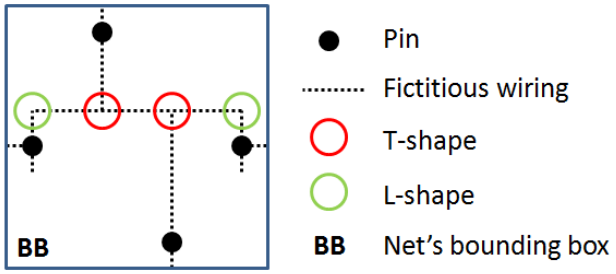


Figure 5.8: Illustrating example showing how the shape count is determined for a 4-pin net based on the assumption of single-trunk Steiner tree routing.

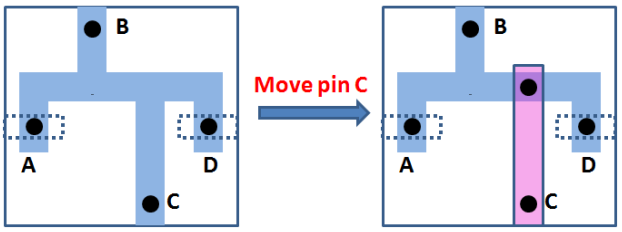


Figure 5.9: Illustrating example showing the move of a pin from M1 to M2 to resolve congestion on M1.

3. and 1D M1 in one direction and 1D M2 in the orthogonal direction.

In case (1), when M1 is congested, the cell-area is increased to accommodate all the wiring. In case (2), certain segments are assigned to M2 to resolve M1 congestion as illustrated in Figure 5.9 and the cell-area is increased only if M1 remains congested after all the available space on M2 is exploited. The number of segments and the utilization of M2 are minimized during the segment assignment to M2. This minimization is done while meeting the maximum utilization allowed on M2 and discounting any segment assignment that introduces congestion in the orthogonal direction. The algorithm used for the layer assignment of segments is described in Figure 5.10.

5.1.4 Congestion Estimation

Once all routes are estimated, we calculate M1/M2 wire length in x and y directions including via/contact-landing pads for the cell I/O pins. Occupied track-length (OTL) in a particular routing direction is then determined as the sum of wire length and blocked track-length from different patterns as well as wires in the orthogonal

Parameters definitions

ΔC^*: needed congestion reduction

$\Delta C_{orth,max}$: allowed increase of congestion in orthogonal direction without an area increase

$L_{M2,max}$: maximum allowed total wire length on M2

ΔC_i: congestion reduction by moving segment i

$L_{M2,i}$: length of the segment when moved to M2

$\Delta C_{orth,i}$: induced change in congestion when segment i is moved to M2

L_{M2}: current total wire length on M2

ΔC_{orth}: current change of congestion in orthogonal direction induced so far

Algorithm

Determine ΔC^*

Determine $\Delta C_{orth,max}$

Determine $L_{M2,max}$

for all segment i sorted from largest ΔC_i to smallest **do**

  Determine $L_{M2,i}$

  Determine $\Delta C_{orth,i}$

  if $L_{M2} + L_{M2,i}$ ¿ $L_{M2,max}$

    OR $\Delta C_{orth} + \Delta C_{orth,i}$ ¿ $\Delta C_{orth,max}$ **then**

    Skip segment i

  end if

  Move segment i to M2

  if $\Delta C + \Delta C_i$ ¿= ΔC^* **then**

    exit loop

  else

    $\Delta C \leftarrow \Delta C + \Delta C_i$

    $L_{M2} \leftarrow L_{M2} + L_{M2,i}$

    $\Delta C_{orth} \leftarrow \Delta C_{orth} + \Delta C_{orth,i}$

  end if

end for

Figure 5.10: Overview of the algorithm used to determine the segments that need to be moved from M1 to M2 to resolve congestion on M1.

direction. Specifically, OTL in y direction is calculated as follows (similar expression

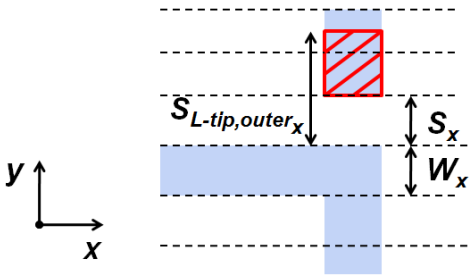


Figure 5.11: Example illustrating blockage model for an instance of L-shape with a single tip facing its outer corner.

for x direction):

$$OTL_y = WL_y + \sum Block_y + \sum_{seg} \left(\left\lceil \frac{WL_{seg,x} + Block_y}{w_y + s_y} \right\rceil - IS_{seg} \right) \times (w_x + s_x). \tag{5.1}$$

The parameters of Equation 5.1 are defined in Table 5.2. IS is included to prevent counting blockage for actual intersections that form corner connections between vertical and horizontal wires. $Block_{x,y}$ models the extra spacing requirement of rules that exceed the minimum spacing (e.g., tip-to-tip). We estimate the number of occurrences of patterns that invoke each of these rules and each occurrence contributes to $Block_{x,y}$ factor by the required spacing minus the minimum spacing. For an illustrating example, consider the pattern of Figure 5.11, which consists of an instance of L-shape with a single tip facing its outer corner. For this pattern, $Block_y$ is the L-shape to tip spacing, $S_{L-tip,outer_x}$, minus the minimum spacing, S_x. The pattern crosses three tracks and has a single intersection. Therefore, the term in the first parenthesis of the second summation of Equation 5.1 evaluates to 2 for this pattern. Estimating the number of occurrences and determining $\sum Block_{x,y}$ are performed using the algorithm in Figure 5.12 and a summary of the spacing rules considered in the DRE framework is given in Table 5.3.

Track congestion in one direction is defined as the ratio of occupied to available track-length (i.e., number of tracks times length of the track).

Table 5.2: Parameter definition for equation 5.1.

Symbol	**Description**
OTL_y	Occupied track-length in y direction
WL_y	Total wire length in y direction
$WL_{seg,x}$	Wire length of a segment in x direction
$Block_y$	Blockage in y direction due to spacing rules that exceed the minimum spacing (e.g., tip to tip)
w_x	Minimum wire-width in x direction
w_y	Minimum wire-width in y direction
s_x	Minimum line-to-line spacing rule in x direction
s_y	Minimum line-to-line spacing rule in y direction
IS_{seg}	Number of actual intersections in a particular segment (i.e. number of L/T-shapes and crosses)

Table 5.3: Spacing rules considered in the DRE framework for both x and y directions.

Tip-to-tip min spacing
Tip-to-line min spacing
L-shape to *outer* tip min spacing
L-shape to *inner* tip min spacing
L-shape to line min spacing
T-shape to *outer* tip min spacing
T-shape to *inner* tip min spacing
T-shape to line min spacing
Cross to tip min spacing rule

5.1.5 Area Increase Due to Congestion

In case congestion (denoted by C) exceeds a certain threshold, the cell-area is increased or M2 layer is used to accommodate all the wiring. This threshold depends on the intra-cell routing efficiency and empty space required on M1 to access the cell I/O pins. Furthermore, routing efficiency is a function of the proportion of non-preferred direction wire length to the total wire length. If wires are mostly in one direction, routing is efficient and increasing the cell-area is only necessary for very high congestion. In contrast, if wires are evenly distributed in the two

Construct list of fixed shapes constituting of tips, power supply wires, input metal pins, and straight-line connections

Construct list of non-fixed shapes as the complementary of the list of fixed shapes

$\sum Block_{x,y} \leftarrow 0$

{*Step 1: check all fixed shapes in the cell against each others*}

for all combination of two fixed shapes that are not processed yet **do**

  if DR^* involving the two shapes is violated **then**

    $\sum Block_{x,y} = \sum Block_{x,y} + DR^* - S_{x,y}$

    Mark both involved shapes as processed

  end if

end for

{*Step 2: check all fixed shapes in the cell against non-fixed shapes*}

for all non-processed fixed shapes **do**

  for all Nets **do**

    if Fixed shape interacts with the net's bounding box **then**

      Find worst DR^* that is violated

      $\sum Block_{x,y} = \sum Block_{x,y} + DR^* - S_{x,y}$

      Mark both involved shapes as processed

    end if

  end for

end for

{*Step 3: check all non-fixed shapes in the cell against each others*}

for all Nets **do**

  for all Nets **do**

    if Bounding boxes of both nets interact **then**

      Find worst DR^* that is violated

      $\sum Block_{x,y} = \sum Block_{x,y} + DR^* - S_{x,y}$

      Mark both involved shapes as processed

    end if

  end for

end for

Figure 5.12: Overview of the algorithm used to determine blockage from rules that exceed the minimum spacing ($\sum Block_{x,y}$).

directions, routing is difficult and increasing cell-area is expected for relatively low M1-congestion. To capture these effects, we model track-congestion threshold as

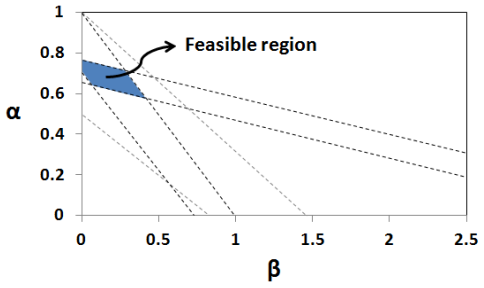


Figure 5.13: Illustrating example for extraction of α and β parameters of Equation 5.2 from M1 congestion data.

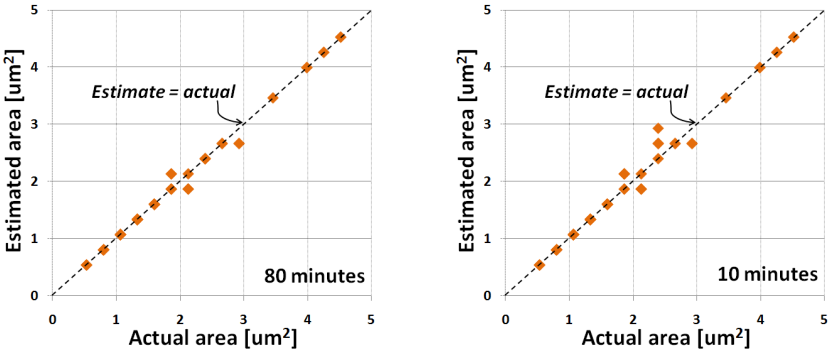
follows:

$$C_{threshold} = \alpha + \left|\frac{U_x - U_y}{U_x + U_y}\right| \times \beta, \tag{5.2}$$

where U_x and U_y are the track utilization in x and y directions. Here, track utilization is defined as the ratio of the occupied track length *without consideration for track blockage from the orthogonal direction wiring* (i.e. Equation (5.1) with $WL_{seg,x} = 0$ for all segments), to the available track-length. α and β parameters, with typical values of 0.6 and 0.2 respectively, are a function of intra-cell routing efficiency. The values of all these parameters are set by the user based on the router specifications.

Figure 5.13 depicts one method to extract α and β parameters either from trial routes of few cells or from cells of a previous generation library. Every single cell implementation adds lower and upper bound lines that narrow down the feasible solution space. Therefore, the more cells are used, the more precise the solution is. If a cell is not congested, we can add one upper bound line derived from $C_{threshold} < 1$ and one lower bound line derived from $C_{threshold} > C$ (by plugging in Equation (5.2) in both cases). If a cell is congested, we can only add one upper bound line that is derived from $C_{threshold} < C$. In the end, exact values of α and β can be approximated by the coordinates of the feasible region's geometric centroid.

Another method to extract α and β is through an automated control loop that runs the DRE framework for a bunch of cells from a previous generation (or cells with trial routes) and fine-tune α and β until the estimated cell-area is very close to the actual cell-area.



(a) (b)

Figure 5.14: DRE estimated cell area versus actual cell area of the Nangate Open Cell Library [Nana] (96 standard cells) with a runtime of 80 minutes and average absolute error less than 1% (a) as well as a runtime of 10 minutes and average absolute error of 2% (b).

5.1.6 Runtime and Validation of Area Estimation

In order to validate our layout estimation method and its efficiency, we use the DRE framework to estimate the topology of the entire Nangate 45nm Open Cell Library [Nana] (96 cells) and estimate cell-area. The comparison between the estimated and the actual areas is depicted in Figure 5.14(a). The results show very good accuracy of the layout estimation method; for 89 out of 96 cells, the estimated areas match exactly with the actual areas and, only for 7 cells, the estimated areas are off by a single poly pitch. This corresponds to an absolute error of less than 1% on average. DRE area estimation has also been validated in [PPL11] by comparison with actual layouts of a commercial 32nm standard-cell library.

The runtime of the evaluation procedure for the entire cell-library is roughly 80 minutes in real time on a single processor of 2GHz clock speed and 2MB cache. This runtime can be reduced to 10 minutes by sacrificing a fraction of the quality of the layout estimation (average absolute error increases to 2%) as depicted in Figure 5.14(b)[5]. In the experiments of this work, we use the DRE setup with the better accuracy.

[5]These area and runtime results are obtained by reducing the maximum number of iterations in the chaining algorithm to 18 iterations.

5.2 Manufacturability Evaluation

Our manufacturability index for evaluating DRs is the functional yield from three sources of failure[6]:

1. overlay error (i.e. misalignment between layers) coupled with lithographic line-end shortening (a.k.a. pull-back);
2. contact-hole failure;
3. random particle defects.

Hence, the overall yield is given by

$$Y = Y_{overlay} \times Y_{contacts} \times Y_{particles}. \tag{5.3}$$

The yield from overlay, $Y_{overlay}$, is equal to the probability of survival (POS) from the overlay error coupled with the lithographic line-end shortening. Overlay vector components in x and y directions are described by a normal distribution with zero mean and process-specific 3σ estimate. We compute POS from overlay causing: failure to connect between contact and poly/M1/diffusion, gate-to-contact short defect, and always-on device particularly caused by poly-to-diffusion overlay error. Connection failure at contacts occurs when the area of overlap with top/bottom connecting layers is smaller than a certain threshold-value. Thus, we consider overlay in both x and y directions in this analysis. In gate-related failure analysis, overlay in just one direction is considered since gates are presumably unidirectional. Moreover, we assume all layers are aligned to a reference alignment mark on substrate[7] and overlay between different layers and the reference layer to be independent[8]. The overall POS from overlay is then calculated as the product of POS from independent overlay errors. If overlay is assumed to be completely a die-to-die variation, then

[6]In this work, we do not model the yield loss from lithography induced systematic failures. In future work, we plan to employ a 2D printability to account for such failures.

[7]This can be modified to conform with the process alignment strategy.

[8]In reality, overlay of different layers with the reference layer have some degree of correlation. This can be dealt with by reducing the amount of overlay (i.e. use smaller 3σ for overlay distribution).

POS of the die is p (equal to POS of the most overlay-critical spot in the layout). On the other extreme, if overlay is completely random within-die variation, then POS of the die is p^n, where n is the total number of critical spots in the design. Reality is closer to the former situation (since field and wafer level components dominate intra-field components [Eic08]), which is our assumption in this work.

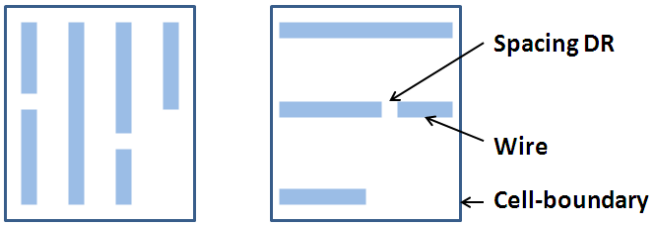
Because contact-hole failure is a random process, we model $Y_{contacts}$ using the Poisson model (as in [eye]). The average number of contact defects (λ) is equal to the number of non-redundant contacts in the layout (N_c) times contact-hole failure rate (D_f). In case contact-redundancy is implemented, duplicated contacts are assumed to always yield since the probability for *two* contacts connected to the *same* pin to fail is negligible. Thus,

$$Y_{contacts} = e^{-\lambda} = e^{D_f \times N_c}. \tag{5.4}$$

To capture failure caused by random particles, we perform critical area analysis for open and short defects at M1/poly/contact layers and short defects between gates and diffusion-contacts. For fast analysis, we use the virtual artwork approach proposed in [Mal85]. Poly and contact layers are represented by strips separated by spacing-DRs; whereas for the M1 layer, this separation corresponds to the spacing that makes the wires as far apart as possible (see example of Figure 5.15). The virtual artwork representation allows quick calculation of critical area as a function of defect size by applying a closed-form model. The average critical area (A_c) for all defect sizes is then determined for each layer while using the following defect size distribution model [Gyv01, Sta83]:

$$f_s(r) = \begin{cases} \frac{2(n-1)r}{(n+1)r_0^2} & \text{if } 0 \le r \le r_0, \\ \frac{2(n-1)r_0^{n-1}}{(n+1)r^n} & \text{if } r > r_0. \end{cases} \tag{5.5}$$

where r is the defect size, r_0 is the defect size with peak density (a.k.a. critical defect size), and n is a parameter related to the cleanliness of the fabrication process and ranges between 2 and 4. Finally, $Y_{particles}$ is calculated using the widely adopted



(a) (b)

Figure 5.15: Virtual artwork representation for (a) horizontal and (b) vertical M1 wires.

negative binomial model [KK98] as follows:

$$Y_{particles} = \prod_{l=1}^{L} Y_{particles,l} \tag{5.6}$$

$$Y_{particles,l} = \prod_{j=1}^{k} \left(1 + \frac{A_{c,j} \times D_0}{\alpha}\right)^{-\alpha}, \tag{5.7}$$

where $Y_{particles,l}$ is the yield from particle defects at layer l, k is the type of defect (e.g., open circuit, short circuit), $A_{c,j}$ is the average critical area for defect type j, D_0 is the average defect density, and α is the defect clustering parameter.

5.3 Variability Evaluation

In sub-wavelength lithography regime, three sources of printing imperfection causing gate-dimension variation are dominant [CGG10] (depicted in Figure 5.16):

- diffusion and poly corner rounding;
- line-end tapering under overlay error and line-end pull-back;
- CD variability associated with different patterning restrictions.

The contribution of each source to gate length and width variations (ΔW and ΔL) is modeled independently. First, we estimate the geometric change in gate length and width from each source. The estimated gates dimensions are then used to determine the overall variability. Our variability index for evaluating and comparing

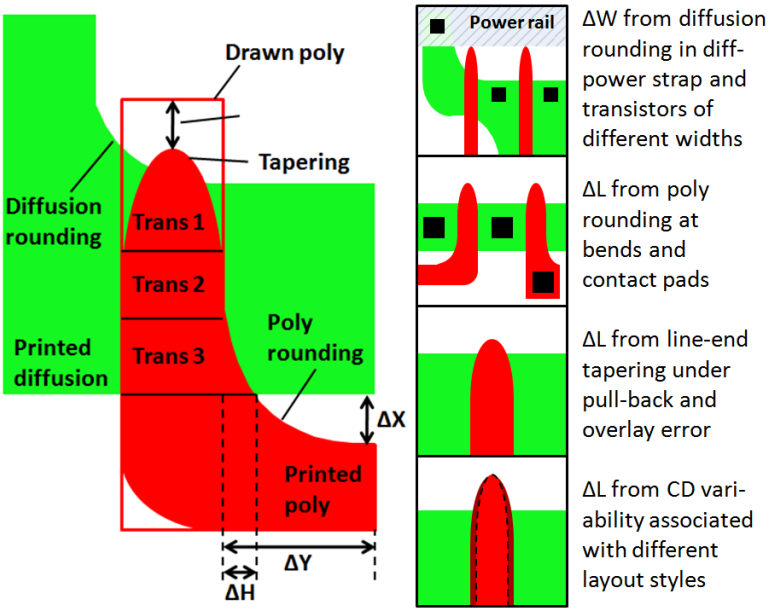


Figure 5.16: Illustration of slicing model, rounding model parameters, and the sources of gate length and width variability considered in the DRE framework. Here, models for tapering and corner-rounding, rather than actual lithography simulation, are used to estimate the contours.

DRs is the total change in drive current, which we calculate using the following equation:

$$\Delta(\frac{W}{L}) = \frac{\sum_{allgates} \left|\Delta(\frac{W}{L})_i\right|}{(\frac{W_{tot}}{L})_{ideal}}, \tag{5.8}$$

where i represents the source of variability[9].

Since the resulting ΔW and ΔL are not across the entire gate, we quantify their contribution to $\Delta(\frac{W}{L})$ by modeling devices as parallel slices of transistors[10].

Diffusion rounding at corners formed by diffusion power-straps and unleveled abutment of transistors (as depicted in Figure 5.16) induces width variation at the gate edge. In addition, poly corner rounding in bends and contact-pads near the gate represents an important source of gate-length variation. The shape of the rounding is a function of the corner dimensions and is modeled as $\Delta H = K_1 \Delta Y \Big/ \sqrt[n]{1 + (\frac{\Delta Y}{K_2})^n}$ where $K_1 = Ce^{D\Delta X}$ and $K_2 = A\Delta X + B$. In this model, ΔX, ΔY, and ΔH

[9]We realize that this estimate is approximate as effects from different sources can interfere. Nevertheless, it is a good indicator of worst-case variability and process control requirement.

[10]More accurate slicing models of [GKK08, SBS07, GKN06, SYP06] can also be embedded in the framework if they are available.

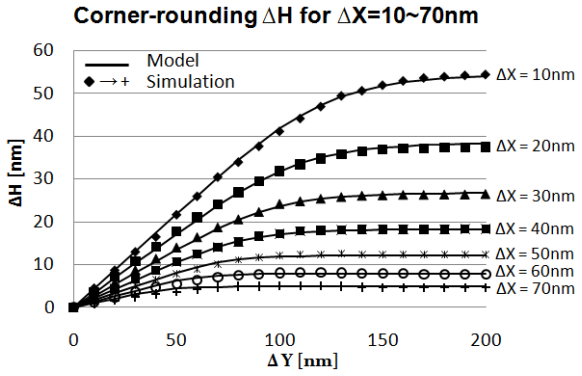


Figure 5.17: Rounding model fitted to give $< 0.8nm$ ΔH error with measured data from printed-image simulations on a fairly wide range of practical corner-dimensions ($\Delta X = 30 \rightarrow 70nm$ and $\Delta Y = 10 \rightarrow 200nm$).

are depicted in Figure 5.16; A, B, C, D, and n parameters are fitted to give $<$ $0.8nm$ ΔH error with measured data from printed-image simulations on a fairly wide range of practical corner-dimensions[11] as shown in Figure 5.17. Simple geometric approximations are then used to infer the gate-length and gate-width variations from the ΔH values caused by the rounding of each corner (in the diffusion and poly layers). It is worth noting that approximate predictive rounding-models fitted from tentative simulation models, which are typically available in early stages of technology development, could be used in lieu of the current model.

Line-end tapering can affect the length of the gate at its edge. This effect becomes pronounced when considering line-end pull-back and poly-to-diffusion overlay error. The tapered shape and gate length at the transistor edge are described using the model offered in [GJK08][12] while accounting for line-end pull-back (mean value) and overlay errors (from distribution). Line-ends are assumed to extend beyond the gate as far as possible unless the user enforces minimum line-end extension (LEE) rule for the entire layout.

[11]The fitting of the model is performed only once per technology node. The model can be fitted to early printed-image simulations or actual silicon data from early testing. Models for our printed-image simulations, which we used to fit the corner-rounding model, were calibrated using CalibreOPC and 45nm OPC models.

[12]$L_i = 2a\big(1 - |\frac{h_i - k}{b}|^n\big)^{\frac{1}{n}}$, where l_i is the gate-length at i location in the line-end extension, h_i is the distance from i to gate-edge, a is half the nominal gate-length, b is the line-end extension, and k and n parameters describe the taper-shape. In our experiments, we use $k = 0$ and $n = 3$.

Table 5.4: BENCHMARK DESIGNS USED IN OUR EXPERIMENTS AND THEIR CORRESPONDING NUMBER OF CELL INSTANCES AND UNIQUE CELL TYPES.

Circuit	**Description**	**Cell instances**	**Cell types**
nova	video compression decoder	43156	81
vga	VGA/LCD controller core	36097	60
mips	processor core	17032	54
ae18	processor core	4358	50

CD uniformity (CDU) is another major contributor to the change in drive current. In our framework, CDU is described by a distribution, which captures the dependency on dose and focus variations. Pattern dependency is captured by using different CDU 3σ values for each poly-patterning style including 1D/2D patterning and multiple/fixed pitch, which can seriously impact CDU [SLC08b, Lie04, PSS03].

After determining all $\Delta(\frac{W}{L})$ terms from different sources, we compute the absolute sum of all terms for the entire layout with the intention of highlighting the actual gate variability. Finally, the drive current variability index is calculated using Equation 5.8.

5.4 Experimental Setup and Results

In this section, we evaluate and analyze major contentious DRs and layout styles for 45nm open-source FreePDK process [Fre]. The DRE framework is also used to compare standard and low power 65nm process from a commercial vendor as well as study the density impact of alternative technologies for the M1 layer at the 14nm node. In another experiment, we collectively explore two gate-spacing related DRs.

5.4.1 Testing Setup

Throughout the experiments, we use four benchmark designs from [opea] synthesized using Nangate 45nm Open Cell Library (scaled for testing with 65nm process). Table 5.4 describes all designs and lists their cell counts and number of unique cell types.

Table 5.5: PROCESS CONTROL PARAMETERS USED IN OUR EXPERIMENTS.

Parameter	**45nm**	**65nm**
Avg defect density [$faults/m^2$]	1395	1757
Critical defect size [nm]	34	45
Max defect size [nm]	250	250
Fab cleanliness parameter	3	3
Clustering parameter (α)	2	2
Contact-holes rate [ppm]	0.00004	0.00004
Overlay (3σ) [nm]	13	15
Line-end pull-back (mean) [nm]	10	14
Gate CDU (3σ) [nm]	2.6	3.3
Critical M1 line-width [nm]	10	15
Critical poly line-width [nm]	15	20
Critical contact-width [nm]	10	15

The experiments were performed using 45nm open-source FreePDK process and 65nm process from a commercial vendor. Estimates of process control parameters associated with each process are summarized in Table 5.5. We use projected values from ITRS technology roadmap [ITRb] and typical values for critical M1 and poly line-width and critical contact-width, which represent the minimum acceptable width for the defect not to be considered a failure. CDU value in the table is for 2D-poly patterning. For 1D fixed-pitch poly, we use CDU 3σ improvement factor of 47% over 2D-poly reported by IBM in [Lie04] and assume that half the improvement is from poly being unidirectional and the other half is from the poly pitch being fixed.

α and β parameters of the congestion threshold model (Equation 5.2) are fine-tuned in a control loop to minimize the error in the estimated area as discussed in Section 5.1.4. Because these parameters model the routing efficiency, the tuning needs to be done just once and only for a small group of cells. We used a couple of cells from the Nangate library that covers the different routing schemes including: highly congested layout with an area increase due to the routing, highly congested layout without area increase, and highly congested layout in a single direction.

Because the area of the benchmark designs is relatively small, we normalize POS values to a $100mm^2$ chip area. We determine for the base case in each experiment the number of design copies that can fit in $10 \times 10mm$ chip size with 80% cell-area utilization and find the corresponding number of contacts and critical areas.

The results of the DR evaluations are a strong function of the base set of rules, layout styles, library architecture, and design type and, hence, they are *not generalizable*. First, we perform studies on 45nm FreePDK process and later we perform studies on a 65nm commercial process as an example.

The number of possible case studies that DRE framework can perform is huge. For brevity, we only show studies of some important DRs and layout styles including: 1D/2D-poly, multiple/fixed pitch poly, diffusion/M1 power-straps, and 8/10/12-track cell heights. Our baseline experiment unless otherwise specified is with the following setup:

- limited routing fixed-pitch poly,
- M1 power-straps,
- and 10-track cell height.

5.4.2 Evaluation of Poly-Patterning Restrictions

Five configurations of poly-patterning styles are investigated:

- unrestricted poly, i.e. 2D-poly,
- limited wrong-way poly, i.e. limited poly routing,
- no poly routing, i.e. 1D-poly,
- limited routing fixed-pitch poly,
- and fixed-pitch 1D poly.

In the cases of 1D poly configuration, poly is used only to connect dual gates (i.e. gates of same transistor-pair). In the cases of limited poly routing, poly is also used

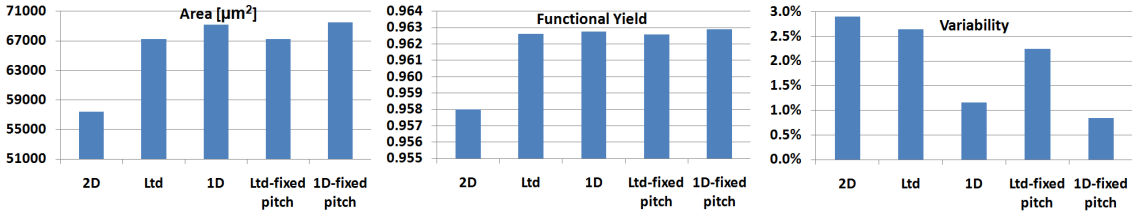


Figure 5.18: Evaluation of restrictive poly-patterning styles on 45nm FreePDK process[13].

to connect adjacent gates in the same p or n network. In the case of 2D poly, poly is used to perform all gate interconnections unless it is blocked by previous routing or diffusion power-straps.

Figure 5.18 shows area, manufacturability, and variability tradeoffs associated with the five configurations of poly-patterning styles on a 45nm process with M1 power-straps and a 10-track cell height.

We observe that 2D poly has a considerable 15% area benefit compared to limited poly routing. On the downside, 2D poly leads to roughly 3× larger variability compared to 1D poly, which is mainly caused by CDU improvement associated with unidirectional patterning. On the other hand, limited poly routing has only 3% area benefit compared to 1D poly and leads to a much larger variability. Thus, allowing small notches on poly (H, U, and Z shapes) with RET complications does not bring much benefits.

Fixed-pitch 1D poly implementation leads to 37% less variability compared to multiple-pitch 1D poly implementation and almost the same area. The area overhead of the fixed-pitch poly restriction is small because the minimum gate pitch (of two stacked gates) is equal to the contacted-gate pitch in FreePDK process and, consequently, a gate-spacing increase is necessary only for isolated gates (with a diffusion gap between the gates).

5.4.3 Evaluation of Layout Styles

Figure 5.19 shows area, manufacturability, and variability tradeoffs associated with M1/diffusion power-straps on 45nm process with limited routing fixed-pitch poly

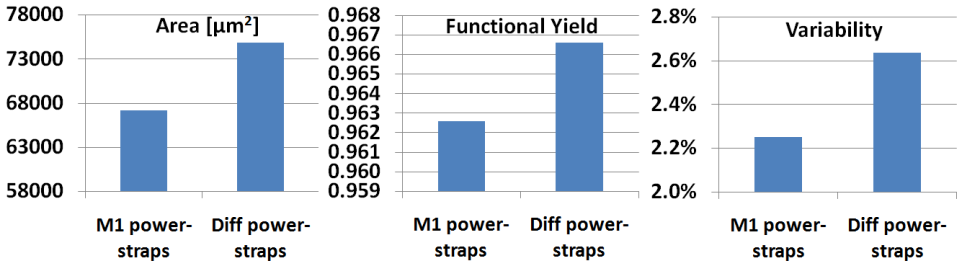


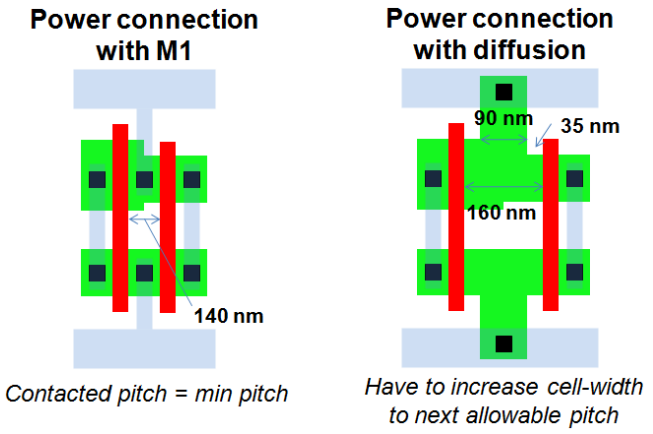
Figure 5.19: Evaluation of M1/diffusion power-strap styles on 45nm FreePDK process[13].

and 10-track cell height. The diffusion power-strap style results in a much larger variability than in the case of the M1 power-strap style (79% larger), which manifests the intensity of the diffusion corner-rounding effect. The reason for this large effect is the fact that cells are packed in the horizontal direction to minimize the cell width and minimum DRs are used. In contrast, poly corner-rounding and line-end tapering effects are usually less important because cells are normally relaxed in the vertical direction (cell-height being fixed).

Furthermore, 11% area overhead is associated with the diffusion power-strap style. This overhead is due to the extra gate separation required to drop the power strap as illustrated in Figure 5.20. The required gate separation at diffusion power straps is even larger when the fixed-pitch poly style is adopted. On the good side, diffusion straps reduce M1 congestion and, consequently, the area of some of the congested cells. In another experiment (not shown in Figure 5.19) with a smaller cell-height (8 tracks), diffusion power-strap style leads to a smaller area overhead (9.6%) than in the case of 10-track cell height, which is because M1 congestion affects the cell area seriously when the cell height is small.

Diffusion power-straps have some manufacturability benefits. Gate-to-contact shorts are reduced and contact redundancy for power connections is implemented at no cost since these contacts are placed on the power-rail in this case.

We also investigate different cell-height decisions. Figure 5.21 shows area, manufacturability, and variability tradeoffs associated with 8/10/12-track cell heights on the FreePDK 45nm process with limited routing fixed-pitch poly and M1 power-



(a) (b)

Figure 5.20: Example of a layout with M1 power-strap (a) and with a diffusion power-strap (b).

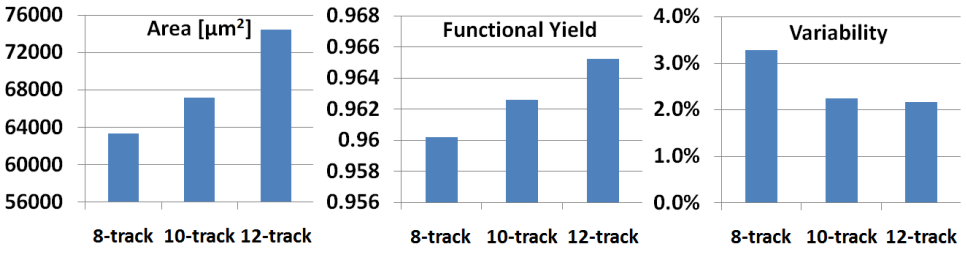


Figure 5.21: Evaluation of 8/10/12-track cell height on 45nm FreePDK process[13].

straps style. The results show a considerable improvement of variability (32%) when the number of tracks is increased from 8 to 10, but only a slight improvement (4%) when the number of tracks is increased from 10 to 12. This is because poly corner-rounding and line-end tapering are aggravated when cells are packed in the vertical direction in the case of a small cell height. The smallest cell-area of the benchmark designs is achieved with 8-track cell height. However, this is not true for all cells as a large cell height is more suitable for cells with wide transistors (as Figure 5.22 shows), i.e. high-performance designs.

[13]The Y-axis showing the functional yield does not start from the zero value to emphasize differences in results (although the differences are tiny).

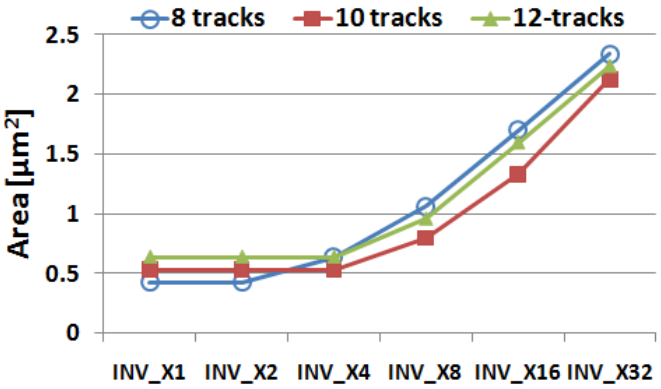


Figure 5.22: Increasing area with increasing transistor width for 8/10/12-track cell height.

5.4.4 Assessment of Technologies and Wiring Schemes

In the previous experiments, we assumed a single metal layer (M1) for the wiring of transistors. Here, we study the effect of allowing an extra metal layer. We run the baseline experiment in the case where M1 is bidirectional and M2 is used only to resolve M1 congestion as well as the case of 1D layout style for M1 and M2. Figure 5.23 depicts the cell area and the M2 utilization associated with each wiring scheme. Allowing M2 in the cell layout reduces the area by 17% on average across all benchmark designs. In case of unidirectional M1 and M2, M2 utilization reaches 27%; whereas in case of bidirectional M1 and M2 used only when M1 is congested, M2 utilization reaches just 10% (2.7× smaller than 1D M1). The downside of higher M2 utilization in the cell layouts is more blockages for the routing at the chip level, which may cause a larger chip area or the need for a larger number of routing layers.

The DRE framework can also be used to assess design implications of patterning technologies. We will show this through an example. Let us consider the patterning for the M1/M2 layers at the 14nm technology node where the alternative technologies are: Single+Trim Exposure and unidirectional M1 (STE) as well as Double-Patterning Technology (DPT) including Pitch-Split Double-Patterning Technology (PS-DPT) and Self-Aligned Double Patterning (SADP), a.k.a. Sidewall Image Transfer (SIT).

In STE, the assumed process consists of forming a grating of unidirectional M1

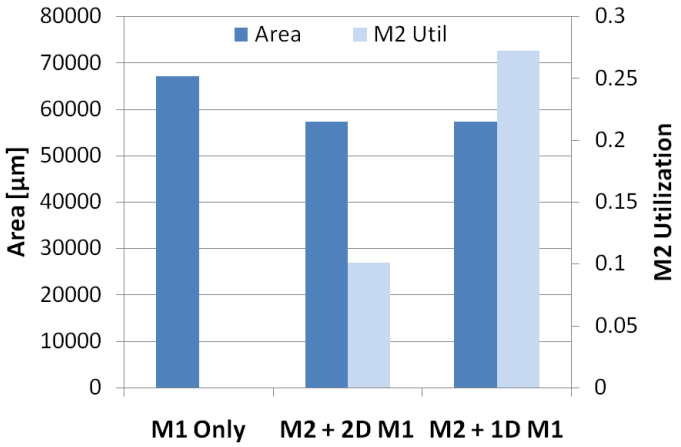


Figure 5.23: Layout area and M2 utilization results when M2 is used only in case of congestion on M1 and when M1 is unidirectional.

at fixed pitch with a single exposure followed by a trim exposure to form line-ends. PS-DPT consists of two separate exposure and etch steps, essentially splitting the layout patterns into two separate masks so that the pitch on the mask is relaxed. SADP consists of forming a first pattern at a relaxed pitch, depositing a sidewall-spacer around the first pattern, and, lastly, defining a second pattern based on the combination of the sidewall-spacer and a trim exposure[14]. On one hand, SADP has typically higher fabrication cost than PS-DPT because it involves more processing steps. On the other hand, when the trim exposure of SADP is allowed to define line-ends but not line-sides (to prevent overlay of trim to mandrel from translating into line-width variation), SADP is more favorable than PS-DPT because of its better overlay performance.

Using immersion lithography and presuming a numerical aperture (NA) equal to 1.35, the limit of bidirectional resolution is at k_1 factor of 0.35 and the limit of unidirectional resolution is at k_1 factor of 0.28 [Wal09]. Therefore, the best pitch that can be achieved with STE and unidirectional M1 is roughly 80nm. With DPT (PS-DPT or SADP), the k_1 limits for bidirectional and unidirectional patterning are roughly one half that of single patterning presented earlier [Wal09]. So, the best achievable pitch with DPT while maintaining bidirectional patterning is 50nm

[14]In a sidewall-is-dielectric process, the first and second patterns are lines; in a sidewall-is-metal process, the first and second patterns are spaces.

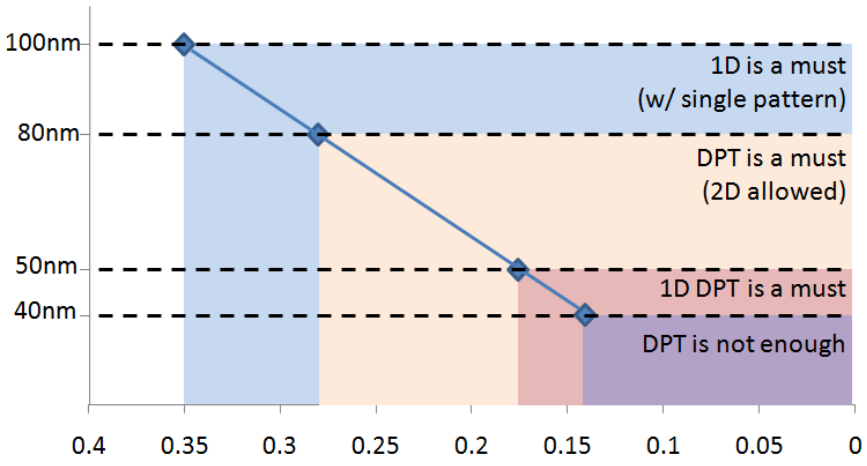


Figure 5.24: Wiring pitch as a function of the k_1 factor and the limits of patterning technologies and directionality (based on [Wal09]).

and the best achievable pitch with unidirectional patterning is 40nm. Figure 5.24 shows the wiring pitch as a function of the k_1 factor and the limits of patterning technologies and directionality.

PS-DPT requires the decomposition of the layout into a first and a second mask layout. Features assigned to the same mask layout must meet the minimum spacing rule of single exposure, which is typically 2× the minimum spacing in the complete layout. For the decomposition to be successful (i.e. without violations), the layout must be adapted for PS-DPT. The layout can be adapted either in a construct-by-correction approach with post-layout perturbations or in a correct-by-construction approach during the design of the layout. One possible method of the correct-by-construction approach is to use conservative spacing rules that, if met, prevent any violations and shield the layout designer from the complexity in dealing with double-patterning violations. For our study, we evaluate the latter method and the minimum spacing of single exposure to be 2× the minimum spacing in the layout. To guarantee *almost zero* double-patterning violations for 2D layouts, we set all rules that involve a tip as well as the L-shape-to-line spacing to the minimum spacing of single exposure. Similarly to PS-DPT, SADP requires layout adaption. Unlike PS-DPT violations, SADP violations are too complex to be prevented with simple geometric rules. 1D layout however, are guaranteed to be SADP decomposable. So

Table 5.6: Summary of patterning styles and rules assumptions made for each technology in our study.

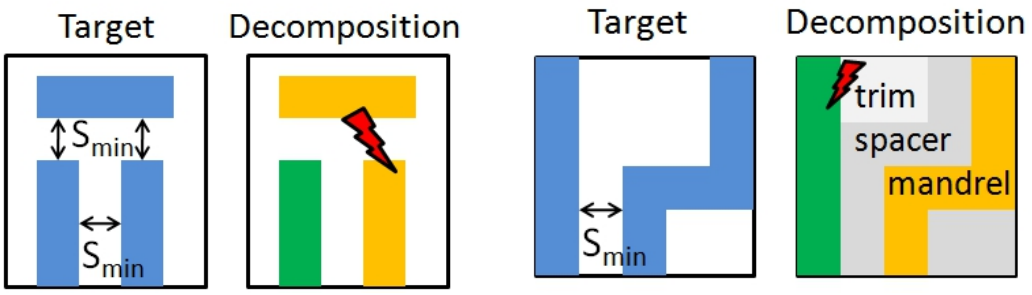
Tech	**Assumptions**
SIT	• Unidirectional patterning only • Pitch = 80nm • All spacing rules = 40nm
PS-DPT	• Bidirectional patterning • Pitch from 80 to 50nm (SADP more favorable below 50nm) • Spacing rules = half pitch except rules involving tips and L-to-L-shape spacing, which are equal to 2× the half-pitch
SADP	• Unidirectional patterning only • Pitch from 80 to 40nm • All spacing rules = half pitch

for SADP, we assume a 1D M1/M2 in our study. Table 5.6 gives a summary list of layout styles and DRs assumptions made for each technology in our study.

We study the density impact of the different alternative technologies available at the limits of the k_1 factor shown in Figure 5.24. At 80nm wiring pitch, we can either have a 1D layout and use STE or enable 2D layout with PS-DPT. At 50nm wiring pitch, we can either have a 2D layout with PS-DPT or a 1D layout with SADP. Finally, for 40nm wiring pitch, only a 1D layout with SADP is possible. For STE, we assume a tip-to-tip spacing rule equal to the minimum spacing in our study[15]. PS-DPT and SADP impose peculiar layout restrictions, however, and many patterns cannot be formed with these technologies (see examples of Figure 5.25).

We run DRE for the designs of Table 5.4 with the three patterning technologies at M1/M2 while assuming all other layers are patterned the same way (i.e. the DRs at all other layers are kept the same in all runs). The results of the evaluation are shown in Figure 5.26. PS-DPT at M1/M2 pitch of 80nm leads to 29% larger area than that achieved with STE at the same pitch. To achieve the same area as with STE at 80nm pitch, the wiring pitch of PS-DPT must be less than 64nm. Hence, a correct-by-construction approach through conservative spacing rules to

[15]Implying that the minimum linewidth of the trim mask is the same as that of the first exposure mask.



(a) (b)

Figure 5.25: Examples of a PS-DPT forbidden patterns because of a coloring conflict (a) and a SADP forbidden pattern because a line-side that cannot be defined except with the trim exposure (b)[16].

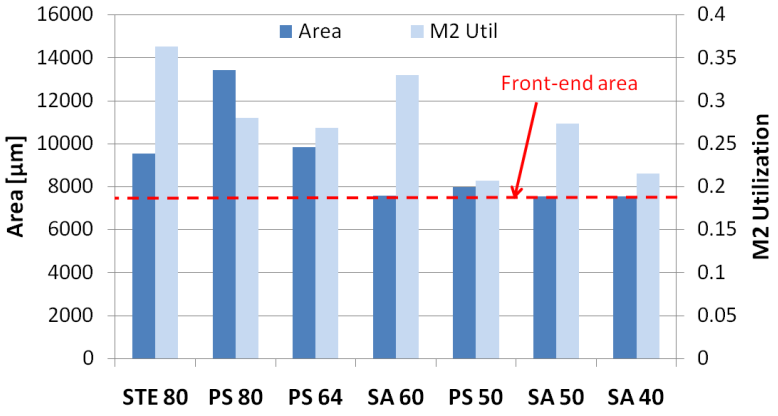


Figure 5.26: Layout area and M2 utilization results for STE, PS-DPT (PS), and SADP (SA) for M1/M2 pitch between 80 and 40nm. Front-end area denotes the area of diffusion, poly, and contacts layers.

enable M1/M2 layouts for PS-DPT may not be satisfactory (given the associated area overhead). On the good side, because PS-DPT allows 2D M1, the M2 utilization with PS-DPT is considerably smaller than that with the other technologies (43% smaller than STE and 24% smaller than SADP). SADP in a correct-by-construction approach with 1D layout seems to be the best alternative in terms of cell area. It can achieve almost the minimum possible area (i.e. the front-end area), which corresponds to 21% smaller area than that of STE, at the wiring pitch of 60nm.

Another example of technology assessment using the DRE framework is the assessment of Shift-Trim Double-Patterning Lithography (ST-DPL), a new double-

[16] S_{min} is the minimum spacing in the layout. Here, we show a sidewall-is-dielectric process for SADP with a trim mask not allowed to define line-sides.

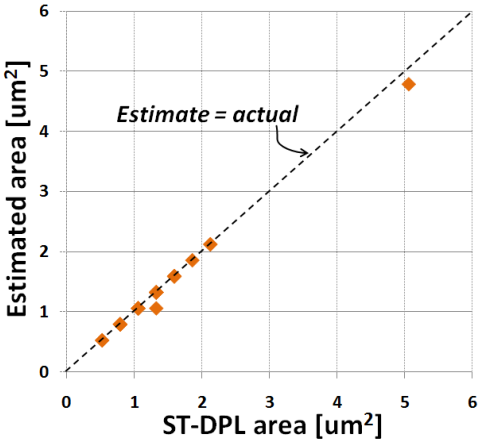


Figure 5.27: DRE estimated cell area versus the actual area of ST-DPL compatible cells designed manually (41 standard cells).

patterning technology that we proposed in Chapter 4. ST-DPL essentially consists of applying a translational mask shift to re-use the same photomask for both exposures of DPT and removing extra printed features using a non-critical trim exposure. To validate the new technique and study its impact on layout density, we migrated in Chapter 4 a small set of standard cells from an existing library so that they become compatible with ST-DPL. Because the automated generation of actual layouts that are ST-DPL compatible is not currently available, the migration of cells was performed manually. Manual layout generation is time-consuming however; only a limited number of layouts can be actually generated and just few layout styles can be tried in practice. Moreover, specific rules are required to simplify the trim mask and exposure in ST-DPL and evaluating the impact of these rules with manual layout generation is practically impossible.

An efficient alternative to manual layout generation is the use of DRE to evaluate the impact of ST-DPL on the design. In Figure 5.27, we compare the area of ST-DPL-compatible cells that are manually generated with the cell area estimated by DRE. For 39 cells, the estimated areas match exactly with the actual areas and, for only two cells, the estimated areas are off from the actual areas by a single poly pitch. The good accuracy of the results imply that DRE can be used instead of manual layout generation to obtain an estimate of the density impact of ST-DPL.

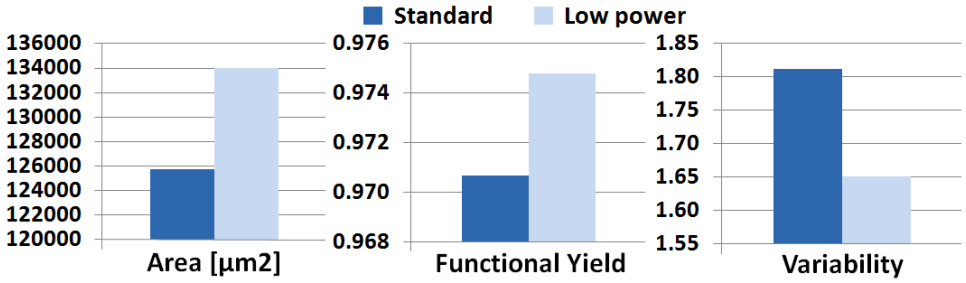


Figure 5.28: Comparison between a standard and a low power 65nm process from the same commercial vendor.

5.4.5 DR Comparison of Different Processes

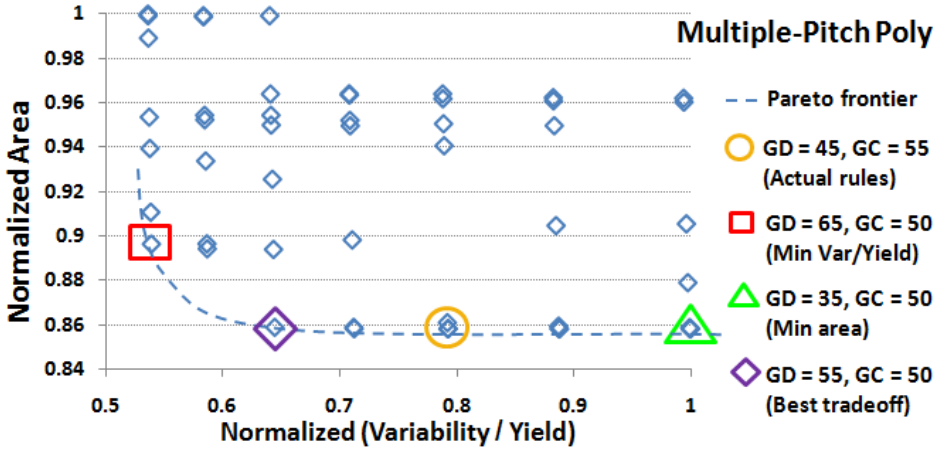
Comparison of DR sets of different processes is another application of the DRE framework. Here, we compare DRs of a standard and a low power 65nm process from the same commercial vendor. We perform this comparison with the layout styles of the baseline experiment. The results depicted in Figure 5.28 show an advantage of low power over standard process in terms of variability and manufacturability; on the other hand, standard process is more area-efficient (7.9% less area).

5.4.6 DR Exploration

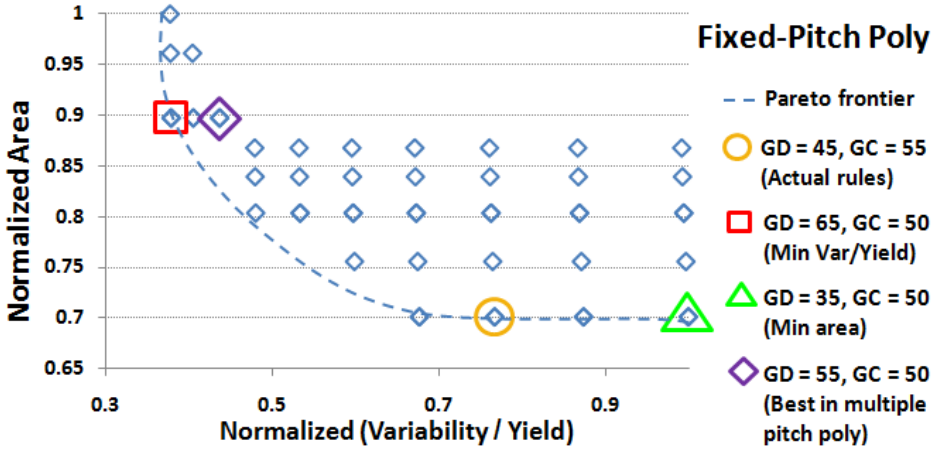
The DRE framework is used for the collective exploration of gate-to-diffusion (GD) and gate-to-contact (GC) spacing rules in the 65nm commercial process. We perform the study for all benchmark designs of Table 5.4 and use diffusion power-straps and limited routing multiple-pitch poly styles that were common at the 65nm node.

The results are depicted in Figure 5.29(a). Each data point correspond to unique combination of GD/GC values. The Y-axis represents the normalized area and the X-axis represents the normalized variability over yield ratio (average values across all benchmark designs). The point corresponding to the process GD/GC actual values falls on the Pareto frontier and very near the solution with the "best tradeoff" (i.e. smallest variability to yield ratio among solutions with almost the smallest area).

We repeat the same experiment with a limited routing fixed-pitch poly style and show the results in Figure 5.29(b). The solution with the "best tradeoff" in the



(a) Multiple-pitch poly



(b) Fixed-pitch poly

Figure 5.29: Co-exploration of GC/GD rules (see figure 5.2) in a commercial 65nm process with diffusion power-straps and limited routing.

previous experiment shifts away from the Pareto optimal frontier and is associated with a large area in this case. Yet, the point corresponding to the process GD/GC actual values falls again on the Pareto optimal frontier and very near the *new* "best tradeoff" solution.

Although quite simplistic, this example provides compelling evidence of our evaluation metrics fidelity and validates our approach. Moreover, the outcomes of this experiment suggest that the optimality of DRs depend strongly on the layout methodology that is in use (layout styles and library architecture) and DR exploration and optimization should be performed across the different layout methodolo-

gies that may be used with the process. This example also shows that the DRE framework can be used as a first-level filter in a DR optimization loop. Rather than exploring the entire search space of DRs with conventional runtime-expensive methods, DRE can be used to quickly eliminate poor DR choices.

Conclusions and Future Work

We proposed a novel framework for *fast*, *early* and *systematic* evaluation and exploration of design rules and technology decisions (**available for download at http://nanocad.ee.ucla.edu/Main/DownloadForm**). By using first order models of circuit characteristics and layout topology and metal congestion-based area estimation, our framework can evaluate big decisions *before* exact process and design technologies are known. In this chapter, we illustrated the potential applications of our framework for the collective evaluation and exploration of DRs as well as the quantitative comparison of DRs from different processes and different technology alternatives. The framework makes DR generation and optimization easier and much faster. Rather than exploring the entire search space of DRs with conventional runtime-expensive methods, the framework can be used as a first-level filter to quickly eliminate poor DR choices. To the best of our knowledge, this is the first work that includes all area, manufacturability, and variability metrics in the evaluation of DRs. Nevertheless, this is just the first step and our ongoing work pursues the following directions:

- address design rule effects on other layout and circuit characteristics including performance, power, reliability, and some notion of designability;
- introduce a 2D printability model (not based on field simulation), for example, derived from [KPX06, CYB08, YS07];
- extrapolate the DR evaluation to the chip level and include intermediate and global metal and via layers;

- study interactions and tradeoffs of variability and area, as in [JKS08] for example.

Acknowledgements

This work was generously supported in part by IMPACT UC Discovery Grant (http://impact.berkeley.edu) contract number ele07-10291, Semiconductor Research Corporation contract number 2008-TJ-1816, and NSF CAREER Award number 0846196. I would like to acknowledge the fruitful discussions with Dr. Lars Liebmann and Dr. Praveen Elakkumanan from IBM and Dr. Chul-Hong Park from Samsung. I would also like to thank Mr. Tuck-Boon Chan from UCSD for providing the corner-rounding model and Mr. Santiago Mok from UCLA for his help with the implementation of the netlist parser.

Chapter 6

Exploration of Design Rules for Multiple-Patterning Technologies

Double/Multiple-patterning (DP/MP) lithography in a multiple litho-etch steps process is a favorable solution for technology scaling to the 20nm node and below. Coloring conflicts represent one of the biggest challenges for MP and limiting them through design rules is crucial for the adoption of MP technology. Also, overlay control is becoming increasingly more important with the scaling of technology. It has become even more critical and more challenging with the move toward MP, where overlay translates into CD variability. Design rules and overlay have strong interaction and can have a considerable impact on the design area, yield, and performance.

In the first part of this chapter, we present a methodology to extend the DRE framework presented in Chapter 5 for the early evaluation and exploration of layout and MP rules. Using a novel wiring-estimation method, we create layout estimates with fine-grained congestion prediction. MP-conflicts are then predicted using a machine-learning approach. We demonstrate the use of the method for double-patterning lithography in litho-etch-litho-etch process. The methodology is more general, however, and can be applied for other multiple-patterning technologies including triple/multiple-patterning with multiple litho-etch steps, self-aligned double patterning (SADP), and directed self-assembly. Results of testing the methodology on standard-cell layouts show a promising 81% accuracy in DP-conflicts prediction. The methodology was then used to explore DP and layout rules and investigate their effects on DP-compatibility and layout area.

In the second part of this chapter, we present a methodology to extend DRE to

explore the interaction between design rules and overlay control. We then use the framework to study this interaction and evaluate the overall design impact of rules, overlay characteristics, and overlay control options. Specifically, we explore the design impact of LELE double-patterning rules for different overlay characteristics (i.e., within-field vs. field-to-field overlay) and different overlay models at the 14nm node.

6.1 A Methodology for the Early Exploration of Design Rules for Multiple-Patterning Technologies

Double/Multiple-patterning (DP/MP) lithography, where layout patterns are formed in multiple separate exposure and etch steps (i.e., litho-etch-litho-etch process), is one of the most favorable solutions for scaling down technology to the 20nm node and below. For the layout to be DP-compatible, layout features must be assigned to two different masks without violating any design rules. Most importantly, features assigned to the same mask, or colored with same color, must obey the minimum same-color spacing rule[1]. Any violation of the same-color spacing is referred to as a DP conflict and achieving a conflict-free assignment is usually impossible for many layouts, especially dense layouts. In fact, it has been shown that layouts typically contain *native conflicts*, which are patterns that cannot be correctly assigned to the two masks without violating the same-color spacing[2].

There are two known approaches to get rid of native conflicts. The first approach is to modify the layout so that it is possible to achieve a correct assignment (as we did in Chapter 2) and this has been investigated extensively in literature [HCN11, FCC12, YP09, GAN11]. These works either fail to achieve a conflict-free assignment or successfully remove all the conflicts in small layouts (cell layouts) at the cost of area increase and considerable layout modifications. The second is a correct-

[1]The minimum different-color spacing rule is equivalent to the minimum spacing rule in the layout and is obeyed during the construction of the layout.

[2]Similar issues exist in other flavors of multiple patterning technologies, such as sidewall image transfer and triple patterning.

by-construction approach. Here, MP rules (i.e., coloring and overlay rules) are accounted for during cell-layout generation and conservative rules are used at the design/cell interface to avoid any possibility of a conflict after placement and routing. Designing with MP rules is believed to be a hassle and conservative rules are expected to have a significant cost in terms of area [LPG11, Ma11].

An alternative to the known approaches, which has not been investigated yet, is to construct layouts with design rules that would bring MP conflicts down to a manageable number, allowing manual or automated legalization of the layout. For this approach to be examined, a method for studying the effect of rules on MP conflicts as well as layout area is needed.

The work in [Den11] presents a flow for DP design rules optimization. The method consists of an optimization loop in which rules are modified, the layout is generated, and printability is analyzed. Because actual layout generation and printability analysis are time-consuming, exploring a wide range of rules and rules combinations is impractical with such approach. Moreover, it is susceptible to the specific layout generator used which makes it tough to measure the inherent "DP-friendliness" of the rules.

We propose a novel methodology for early evaluation and exploration of DP design rules. The overview of the methodology is depicted in Figure 6.1. Given trial design rules and DP rules, the first step is to generate the layout of device layers. Next, wiring-layers layout are estimated and congestion is predicted using a novel fine-grained wiring-estimation method. The presence of DP conflicts in the layout is then predicted using a machine-learning approach. In particular, fine-grained estimates of wiring congestion and estimates of layout features (e.g., lineends and L and T-shapes) and their distribution are given to the machine learning (ML) model, a feed-forward back-propagation Artificial Neural Network (ANN).

To the best of our knowledge, this work is the first to offer a methodology for the exploration of DP rules at early stages of process development. *Although the focus in this chapter is on DP, the methodology is more general and can be applied to explore*

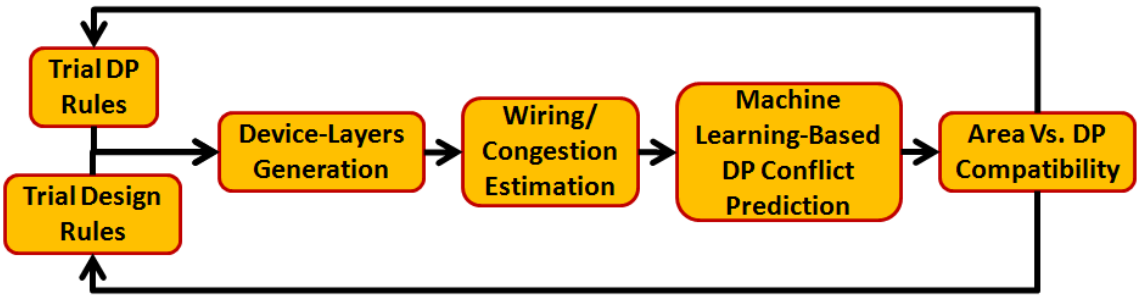


Figure 6.1: Overview of our methodology for exploration of DP design rules.

rules of other layout-restrictive technologies, such as triple patterning, self-aligned double patterning, and directed self-assembly.

We make the following contributions.

- We present the first work on evaluation and exploration of DP rules intended for speeding up the rules-development cycle.
- We propose a novel method for estimating the layout and wiring congestion at a fine-grained level.
- We offer a method that predicts the presence of DP native conflicts based on machine learning, using neural networks, and requires basic layout information like congestion and feature distribution.
- We study the effects of rules and layout styles on DP compatibility of cells and report preliminary results on this topic.

6.1.1 Probabilistic Layout and Congestion Estimation

Our approach relies on layout estimation, rather than actual generation, to enable early exploration of a wide range of design rules. We first estimate the device layers of cell layouts to predict contact-points locations and area of front-end layers. For this estimation, the design-rules exploration framework of [UCL] was used. The framework employs layout-topology generation methods that were shown to be fast and accurate [GG09a], which makes it well suited for our approach. For more details,

the reader is referred to [GG09a]. The next step is to estimate the layout for back-end layers used within the cells (i.e., M1 in our experiments).

Background

At the design level, many techniques exist in literature to probabilistically estimate congestion without performing actual routing. All the approaches of [LTK02, WBG04, LAQ07, Kah03] essentially spread a net across its possible routes and, then, compute congestion for each tile in the design grid based on the probabilistic contributions of nets that pass through the tile. A minimum spanning-tree (MST) is also used to break multi-pin nets into constituent two-pin net-segments.

Probabilistic congestion estimation techniques have been shown to successfully guide design optimization choices at various stages of physical design, including placement [LTK02, WBG04, LAQ07] and logic-synthesis [LM04, SSS05]. Our novel approach extends and improves upon Steiner heuristic called the single trunk Steiner tree (STST) [Sou81], leveraging it for wiring and congestion estimation at the standard-cell level. Due to its linear computation time (as compared to $O(nlogn)$ or more commonly $O(n^2)$ for MST based approaches) the STST has been previously used for wirelength estimation [CQZ92, CNV90, SCK91], with [CQZ92] enhancing it to provide a $O(nlogn)$ approximation, which is on average within 6% of the optimal. We show that our STST based approach allows for fast, yet accurate congestion/wiring estimation as well as prediction of the usage of specific layout shapes and patterns in standard cells. The advantages of using probabilistic STST-based wiring are:

1. The runtime for computing each probabilistic route is $O(N)$, where N is the number of pins on the net. Does not require breaking multi-pin nets into constituent two pen nets, avoiding expensive computations such as MST, RST, RSMT, etc.

2. Each route computation is independent of the other, hence can be easily parallelized if needed.

3. Allows for prediction of local usage of specific patterns and shapes, which is key to predicting multiple-patterning coloring conflicts.

Probabilistic Wiring-Solution Generation

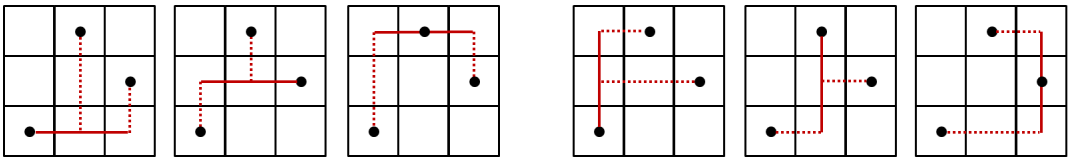
Once the front-end layers are created and contact locations are determined, we generate a number of possible wiring solutions for each net (unlike [GG09a], where the wiring of each net is estimated with a single solution). For each cell layout, the generation of wiring solutions undergoes the following steps.

1. Create a grid.
2. Pick a net that is not yet processed.
3. Enumerate possible wiring solutions following a single-trunk Steiner tree topology.
4. Update track utilization for each involved tile.
5. Repeat steps (2) to (4) until all nets are processed.

The grid consists of tiles sized in terms of the number of horizontal and vertical wiring tracks. The tile size is configurable; a large tile size would lead to fast running time but coarse-grained congestion estimates; while a small tile size would lead to fine-grained congestions estimates at the cost of running time. In our experiments, we used a tile size of two vertical tracks and two horizontal tracks[3].

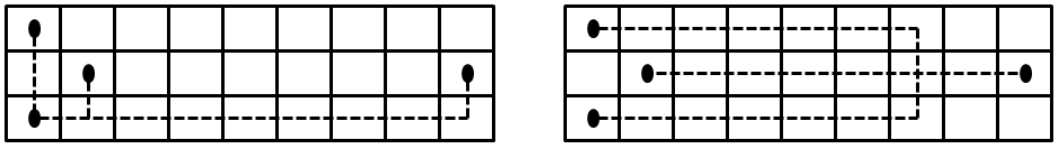
In step (3), we enumerate for each net all possible wiring solutions that follow a single-trunk Steiner tree topology. In general, only wiring solutions within the net's bounding box are considered as in the example of Figure 6.2. When a bounding box is too small only few solutions will be generated. Having just a few solutions, a single one in the extreme case, makes those almost fixed rather than potential solutions.

[3]If the width (height) of the cell is not an exact multiple of the computed tile-width (tile-height), then the tiles in the last column (row) have smaller horizontal (vertical) track capacities as compared to the rest of the grid.



(a) (b)

Figure 6.2: Example of possible wiring solutions with single-trunk Steiner-tree topology for a three-pin net using (a) horizontal trunks and (b) vertical trunks.



(a) Wire length of 12 units. (b) Wire length of 21 units.

Figure 6.3: Example of an unbalanced net and two of its possible solutions: (a) a typical horizontal trunk-based solution and (b) a pessimistic vertical trunk-based solution.

To solve this problem, any bounding box with a number of rows/columns falling below a threshold is expanded by a fixed factor (i.e., we allow detours for such nets). For nets having bounding boxes with very skewed aspect ratio, we ignore wiring solutions with common trunks along the shorter direction as such solutions have very poor wirelength (see Figure 6.3). Each possible solution for a net is assumed to be equally probable (e.g., the probability of each wiring solution in the example of Figure 6.2 is $\frac{1}{6}$).

After the possible wiring solutions for the net are generated, we determine in step (4) the track utilization in the horizontal and vertical directions for all tiles in the bounding box. Here, track utilization is the occupied track length multiplied by the probability of occurrence. Consider the center tile in the example of Figure 6.2. The occupied track length in the horizontal direction is roughly 2.5 times the tile-width and the probability of occurrence for each occupied track segment is $\frac{1}{6}$. The track utilization in the vertical direction is calculated similarly and has the same value in this example.

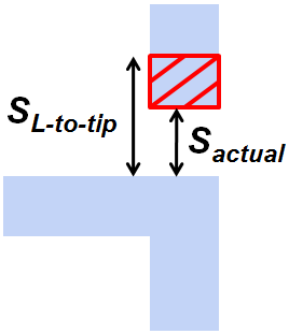


Figure 6.4: Example of DR violation and associated extra track utilization.

Feature Distribution and Design-Rule Violations

Once all nets have been processed (by repeating steps (2) to (4) for all nets), we determine the distribution of the different features including tips, line segments, and L and T-shapes. The probability of a feature occurrence at a particular location is the probability of its corresponding wiring solution.

Given the distribution of features, design-rule violations are identified and their effect is modeled as extra track utilization. To illustrate, consider the example design-rule violation shown in Figure 6.4. Here, the minimum L-shape-to-tip spacing rule, $S_{L-to-tip}$, is violated. To account for this violation, the utilization is updated to include the amount with which the rule is violated (i.e., the shaded region in Figure 6.4) multiplied by the probability of pattern occurrence (i.e., probability of occurrence of the two shapes in Figure 6.4 simultaneously). The spacing rules that are considered for violation checking in our method are given in Table 6.1.

Modeling of Congestion and Its Impact on Area

Congestion is the ratio of occupied to available track-length and is reported for each tile and in the horizontal and vertical directions separately. The occupied track length is modeled as the sum of the utilization and track length blocked by wiring in the orthogonal direction. And, the available track length is inferred directly from the tile size.

Table 6.1: Spacing rules in vertical and horizontal directions considered in our framework of rule violations impact on wiring congestion.

Tip-to-tip min spacing
Tip-to-side min spacing
L-shape *outer corner* to tip min spacing
L-shape *inner corner* to tip min spacing
L-shape to side min spacing
T-shape *inner corner* to tip min spacing
T-shape *side* to tip min spacing
T-bend *side* to side min spacing
Cross *inner corner* to tip min spacing

An over-congested tile (congestion greater than 1) indicates a problematic region for completing the wiring successfully. Such regions are very likely to necessitate a layout-area increase. For cells with over-congested tiles, the total congestion overflow is calculated. The extra track-length requirement is fulfilled by increasing the cell area. Since the cell height is fixed, the cell width is increased with the minimum cell unit-widths to meet the requirement. Lastly, congestion in over-congested tiles is updated to the value 1 and the total congestion overflow is distributed equally among the newly added tiles.

Method Validation

The accuracy of our layout estimation approach is verified by comparison with real cell-layouts. The layout estimation method is run for the entire Nangate 45nm Open Cell Library [Nanb] (110 cells) using the same design rules (i.e., FreePDK 45nm process [Fre]). A comparison of the estimated and actual cell areas, depicted in Figure 6.5, show an absolute error of less than 2% on average and a runtime of 38 minutes in real time (on a single CPU)[4]. A comparison of the number of tiles in each

[4] AlThough the grid-based congestion estimation is fairly accurate, we use the approach of [UCL] for estimating the actual impact of design rules on area since it gave better accuracy for area estimation (1% vs 2% error for Nangate).

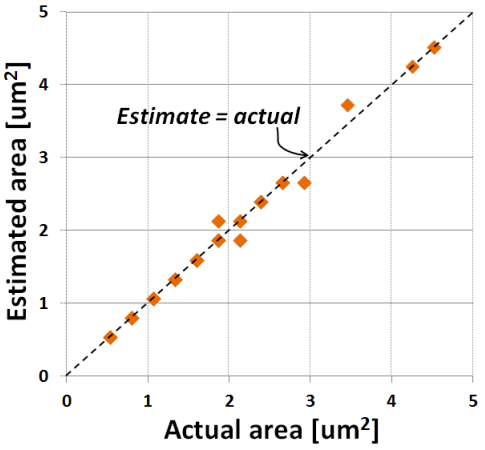
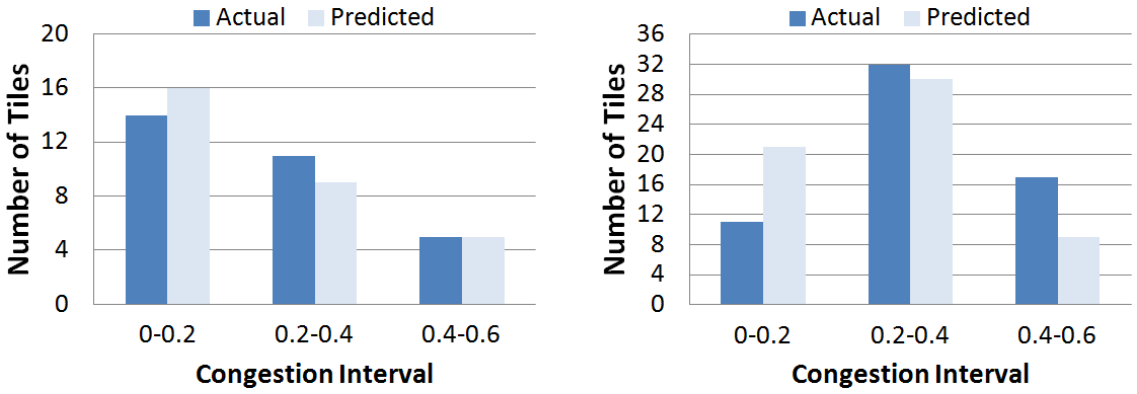


Figure 6.5: Comparison of the estimated cell area of our approach with that of actual cell-layouts using the same design rules.



(a) NOR3_X4 cell.

(b) FA_X1 cell.

Figure 6.6: Comparison between actual and predicted congestion.

congestion-interval for two standard cells[5] is reported in Figure 6.6. Our approach does a good job of identifying areas with high congestion, more specifically the number of tiles with high congestion, which matter the most for our DP conflict prediction.

Our method was also compared to FLUTE [CW08], a rectilinear Steiner minimal tree routing algorithm. Our wirelength estimates, obtained by averaging across all generated solutions, and runtime are compared with that of FLUTE for nets in the Nangate library [Nanb]. Figure 6.7 summarizes the results. Our approach leads to

[5]With 0.2 intervals, i.e., with the same form of input to the machine learning-based conflict predictor.

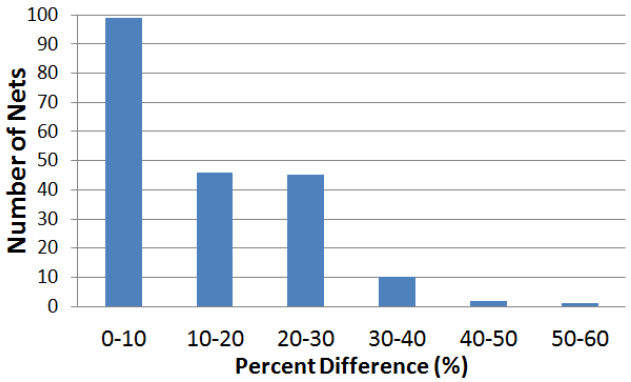


Figure 6.7: Percent difference between our wirelength estimates and that of FLUTE.

11.94% higher wirelength on average and has $44X$ faster runtime compared with FLUTE.

6.1.2 DP Conflict Prediction Using Machine Learning

A machine-learning approach is used to predict DP conflicts in cell layouts as well as at cell boundaries in the design. Artificial neural networks (ANN) have been shown to work well for prediction problems [Mai00]. As a result, we develop an ANN-based classifier, more specifically a feed-forward back-propagation multi-layer perceptron, and use it to predict whether a layout, with a given set of design rules, will have a DP native-conflict (essentially requiring layout redesign).

Overview

An overview of our machine learning-based approach for predicting the presence of DP native conflicts in the layout is depicted in Figure 6.8. The machine-learning model is calibrated (as well as tested) using the estimated layouts and congestion maps obtained from the method described in Section 6.1.1. Characteristics of the estimated layout – as well as DP coloring rules – are given to the model as inputs and are referred to as layout descriptors. These descriptors are carefully chosen so as to correlate well with the existence of DP conflicts.

The model's target data, which is the real-known data that the prediction is com-

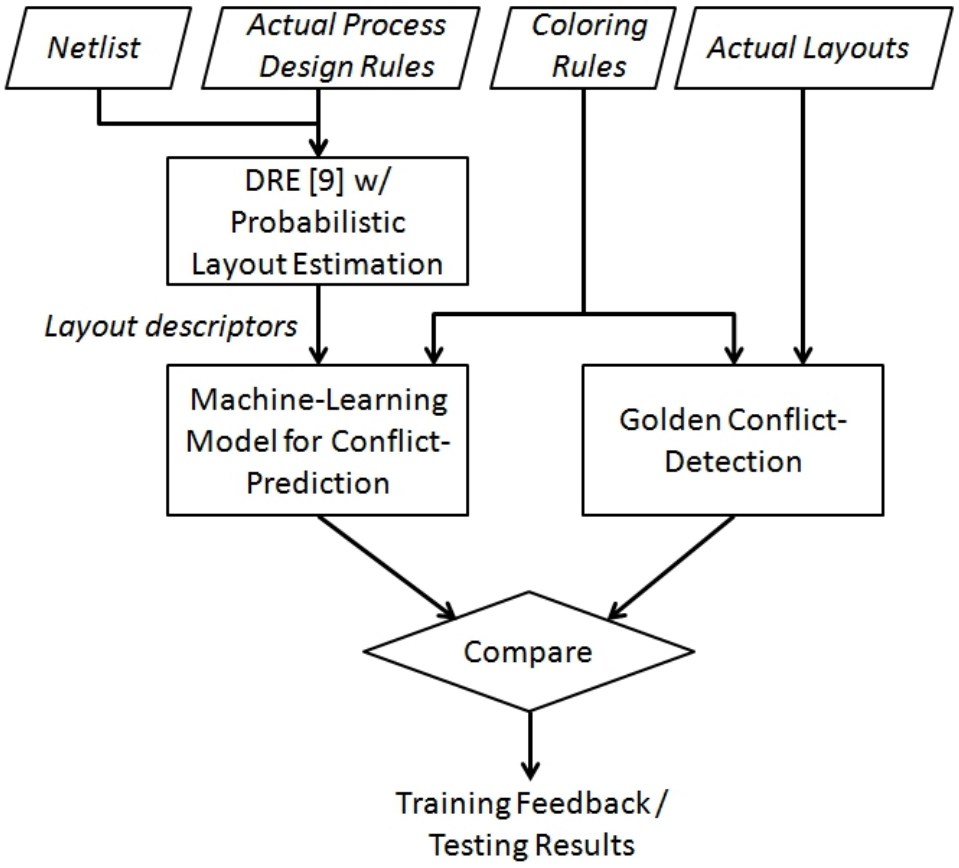


Figure 6.8: Overview of our machine learning based approach for conflict prediction.

pared with in the training and testing of the model, is the actual presence/absence of DP conflicts in *real layouts*. The identification of DP conflicts in real layouts was performed using the golden method of odd-cycle detection in the conflict graph. Specifically, color violations were detected using Calibre SVRF [Calb] and conflict-graph construction and a depth-first search algorithm were implemented in C++ with OpenAccess database to detect odd cycles *in the real layouts*. The Neural Network Toolbox in MATLAB was used in the development of the ANN model [Mat11].

DP conflicts may occur within cell-layouts but also at the interface between cells post-placement. As a result, for the training and testing sets of our model, we use layouts of standard cells as well as layouts of the interface between different cells, which are taken from post-placement benchmark designs and are referred to as *cell-boundary layout snippets*.

The training of the model was performed using cell layouts from Nangate 45nm

Table 6.2: Layout descriptors investigated to build the machine-learning model.

Descriptor	**Details**
Tile horizontal congestion	number of tiles at each congestion-interval
Tile vertical congestion	number of tiles at each congestion-interval
Tile overall congestion	number of tiles at each congestion-interval
Tile wiring property	number of tiles with 1D, 2D, and no wiring
Layout congestion	vertical, horizontal, and overall congestions
Tile tips	number of tiles at each tip-interval
L / T-shapes	number of tiles at each L/T-shape-interval
Highly packed tiles	number of tiles w/ high congestion and number of tips
Min same-color spacing	Side-to-side, tip-to-side, & tip-to-tip rules

Open Cell Library [Nanb] and cell-boundary snippets from benchmark designs synthesized using the same library. The testing of the model was performed on a different set of layouts for cells of the same library and cell-boundary layout snippets from different benchmark designs.

Input Layout Descriptors

Layout descriptors resulting from our layout/congestion estimation method include forms of the following layout characteristics.

- Horizontal wiring congestion.
- Vertical wiring congestion.
- Number and location of tips.
- Number and location of L and T-shapes.

An exhaustive set of layout descriptors, summarized in Table 6.2 and inferred from the above characteristics, is investigated and training is repeated with various combinations of descriptors. The descriptors set was pruned depending on the trend of training accuracy.

The descriptors are fused together into a single input vector for each training/testing layout. Based on training accuracy it was observed that horizontal congestion, vertical congestion and tips had good influence on the prediction accuracy. L and T-shapes were excluded from the feature set because they did not affect the training accuracy[6].

Training the Classifier

Efficient training of an ANN depends on various factors such as the nature of the data set and the topology of the network. The basic ANN structure has three types of layers: input, hidden and output. Using multiple hidden layers, or even a single hidden layer, with too many neurons results in an increased risk of convergence to a local minimum (over-fitting) and, hence, poor performance on unseen samples. The number of neurons was chosen, by first starting with a large number and gradually reducing to a point where best possible results were obtained.

The model is cross-validated during training so that it is generalized across variations in the training patterns and over-fitting is avoided. The Levenberg-Marquardt algorithm [Lev44, Mar63] has been shown to be most efficient in attaining a good mean squared error [ZBR08] and, as a result, was used for training our model.

For robust training, we ensured that all the input features have nearly same order of magnitude to avoid input features with larger magnitudes from dominating the learning process. For the model to perform well on unseen data and to avoid any learning bias, the selection of cells and cell-boundary snippets layouts for use in the training was made at random. And, for the same purpose, almost the same number of positive (layouts with conflicts) and negative (layouts without conflicts) samples were used in the training. It is important to note that, if the frequencies of layouts with/without conflicts differ widely in a certain technology, either the number of samples of each type (i.e., positive/negative) can be modified or samples can be weighted according to the relative frequencies to improve the model's accuracy.

[6]One explanation for this is that including both horizontal and vertical congestion already captures the same information given by L and T-shapes.

Table 6.3: Summary of the Properties of the ANN Model

Number of layers	3
Number of neurons	50
Number of inputs	19
Supported min same-color spacing rules	S2S, T2S, T2T
Training-range of min same-color spacing rules	80-150nm
Training error	8.7%

The tile size also has an impact on DP conflicts. A small tile size implicitly ignores long-range conflicts; whereas, a large tile size will have inaccuracy due to ignoring the inherent locality of conflicts. Different tile sizes were examined and a tile size of two tracks in each direction showed best results.

The properties of the model are summarized in Table 6.3[7]. In practice, the minimum side-to-side (S2S), tip-to-side (T2S), and tip-to-tip (T2T) same-color spacing rules may have unequal values. As a result, the model was designed to support different values for these rules, which are inputs to the model.

Testing and Approach Validation

To obtain the ANN outputs in binary form[8], a thresholding function is applied. The discrimination threshold was varied to obtain the Receiver Operating Characteristic (ROC) curve for the model. Figure 6.9 shows a comparison between ROC-curves of our model, random guess, and a multivariate linear regression-based model. The area under the ROC curve is a measure of the accuracy of the classifier. It can be clearly seen that our classifier performs significantly better than random guess and the regression-based classifier.

The ROC curve manifest the trade-off between true and false positive rates. If the discrimination threshold is chosen so as to operate in P2 region, then the ANN model over-predicts DP conflicts; if the model operates in P0 region, DP conflicts

[7]We use hyperbolic tangent-sigmoid for the hidden-layer activation function and a linear function for the output-layer activation function.

[8]Originally, outputs are decimal numbers because a linear function is used for the output activation function.

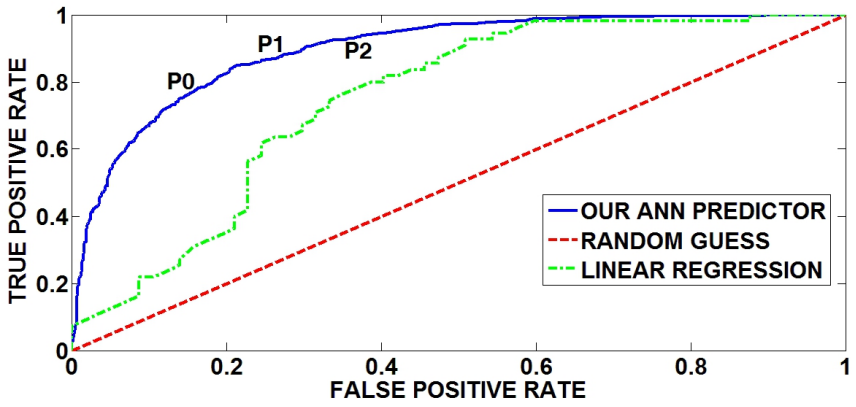


Figure 6.9: ROC curve for our ANN model compared with those of linear regression and random guess.

Table 6.4: Results of testing the ANN model on cell layouts.

True positives	1270
False negatives	271
Positives detection rate	82.5%
False positives	283
True negatives	1138
Negatives detection rate	80%

will be under-predicted. A reasonable trade-off was chosen for the testing, more specifically 82.5% true and 20% false positive rates (i.e., region P1).

The testing results of the ANN model are shown in Table 6.4. The model achieves a positives-detection rate of 82.5% and a negatives-detection rate of 80%.

6.1.3 DP Design-Rule Exploration

In this section, we use our methodology to explore design rules, especially DP rules, and study their effects on DP compatibility of layouts. It is important to note that the results presented in this section show the strength of the methodology but are not necessarily generalizable. *Design rules, layouts, and DP-conflicts have complex interaction, therefore, the results will strongly depend on the precise rule values, layout styles and the library architecture.*

DP-compatibility of the cell library does not guarantee DP-compatibility of the

design after cell placement. To study the general compatibility of layouts with DP, the compatibility metric should involve some notion of DP conflicts at cell-boundaries in addition to DP conflicts in standard-cell layouts. Furthermore, because DP conflicts in cells are typically harder to fix than conflicts at cell-to-cell interfaces (which may be fixed in many cases by cell-placement perturbation), the two types of conflicts should be weighted when forming the metric for the overall DP-compatibility of the design. Hence, we use the following metric:

$$Design\ Conflicts = \alpha \times CC + (1 - \alpha) \times CB, \tag{6.1}$$

where α is the weighting factor, CC is the fraction of conflicting cells in the library, and CB is the fraction of conflicting cell-boundaries in the benchmark design. For our experiments, we use a weighting factor α of 0.5. Nevertheless, the weighting factor can be adjusted in accordance with the relative importance of conflicts within cells and conflicts across cell boundaries.

Testing Setup

Our design-rule exploration experiments are performed on the M1 layer in Nangate 45nm standard-cell library [Nanb] with FreePDK 45nm design rules [Fre] (65nm for the minimum spacing in M1). Although we have used a 45nm setup, we correspondingly scaled the minimum same-color spacing rules for DP to make the experiments realistic.

For our baseline experiment, we use a cell-height of 10 M2-tracks, 1D polysilicon (poly), and local interconnect (LI) to perform gate-to-gate connections[9]. Our methodology, layout/congestion estimation followed by the machine-learning model, is run to predict the presence of DP conflicts in M1 layouts. For evaluating design area, we use four benchmark designs from [opea] synthesized using Nangate 45nm Open Cell Library [Nanb] and the area evaluator of [UCL]. Descriptions, cell-instance counts, and number of cell-types for all four design are given in Table 6.5.

[9]Although LI is not needed at 45nm, we investigate its effect on reducing M1 complexity and, consequently, DP-compatibility.

Table 6.5: BENCHMARK DESIGNS USED IN OUR EXPERIMENTS AND THEIR CORRESPONDING NUMBER OF CELL INSTANCES AND UNIQUE CELL TYPES.

Circuit	**Description**	**Cell instances**	**Cell types**
mips	processor core	19868	73
tv80	processor core	6429	72
or1200	processor core	3077	55
des	cryptography core	1475	28

Prediction of DP conflicts in cell-boundary layout snippets were performed for the same benchmark designs. Since our experiments are performed on the M1 layer and Nangate library has wide horizontal M1 power rails (enough to prevent conflicts between features across the rails), post-placement DP conflicts spanning multiple cells may only occur at the vertical interface between cells. In our experiments, we limit the analysis to cell-boundary snippets at vertical interfaces between cells and we use a snippet size of *Cell Height* $\times$ *8 M1 Pitches*, i.e., four M1 tracks from each side of the interface. For performing experiments on different layers or different types of power rails, however, the layout snippets will need to include snippets at the horizontal interface between cells as well.

DP Design-Rule Exploration

Three spacing rules may have a significant impact on DP-compatibility: side-to-side, tip-to-side, and tip-to-tip minimum spacing rules. With DP, each of these rules has two versions: minimum same-color spacing and minimum design rule in the layout (equivalent to minimum different-colors spacing). Among all three rules, tip-to-side spacing is of particular interest for exploration; tip-to-tip spacing hardly occurs in M1 layouts (for logic) and side-to-side spacing is usually a pre-defined target for the technology.

Our first experiment consists of investigating the minimum tip-to-side same-color spacing rule[10]. In Figure 6.10, *Design Conflicts* (see Equation 6.1) is plotted as a function of the rule value. As expected, we see an increase in conflicts as we increase

[10]This rule affects the coloring of tip-to-side patterns and should not be confused with the minimum tip-to-side design rule, which defines the spacing of tip-to-side patterns in the layout.

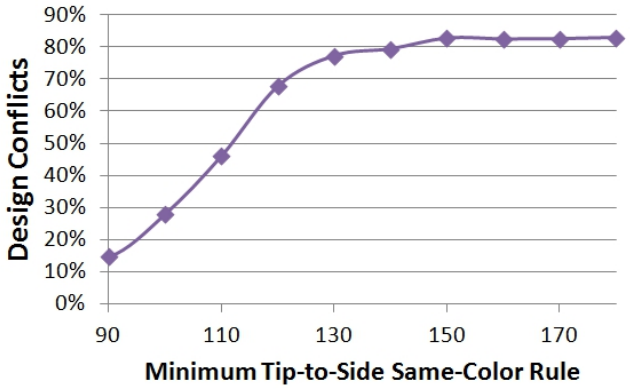
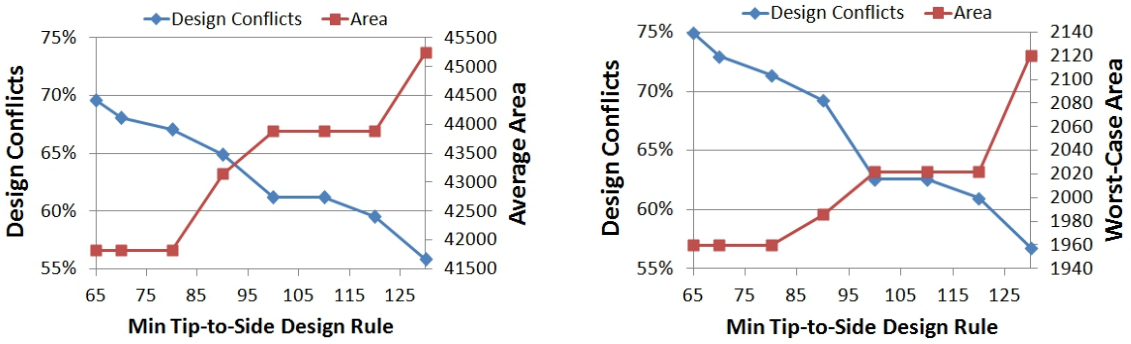


Figure 6.10: *Design Conflicts* for range of tip-to-side same-color spacing rule.

the same-color rule value. The plot identifies a range of rule values with high impact on DP conflicts, namely values from 90 to 120nm, and another range with small impact on DP conflicts, namely values from 120nm to 180nm. Such information can be used to guide efforts on the process-development side to enhance the DP-compatibility of layouts. For example, pushing the rule value from 1.7× to 1.5× the minimum side-to-side spacing design rule (i.e., from 110nm down to 90nm) would more than double *Design Conflicts*. It is worth noting that the plot shows imperfect monotonicity. This limited noise is due to the fact that the machine-learning model does not make perfectly accurate predictions and few incorrect predictions may cause such noise when the total number of cells in the library is limited (110 cells for Nangate).

In another experiment, we study the conflict/area trade-off with changing the minimum tip-to-side design rule. For a same-color spacing fixed at 130nm (2× min spacing for all three same-color rules) and for each rule-value, *Design Conflicts* and the benchmark-designs area are depicted in Figures 6.11. Interestingly, a non-linear trend is observed in both cases for conflicts as well as design areas, which reveals optimization opportunities. For example, increasing the tip-to-side rule from the original value of 65nm to 80nm can reduce *Design Conflicts* by 3% with almost no area increase. Furthermore, 9% reduction in *Design Conflicts* is observed when the tip-to-side design rule is increased to 100nm but the area overhead in this case is 4.7% on average.



(a) Average area across all designs

(b) worst-case design area

Figure 6.11: *Design Conflicts* and area of benchmark designs for range of tip-to-side design rule.

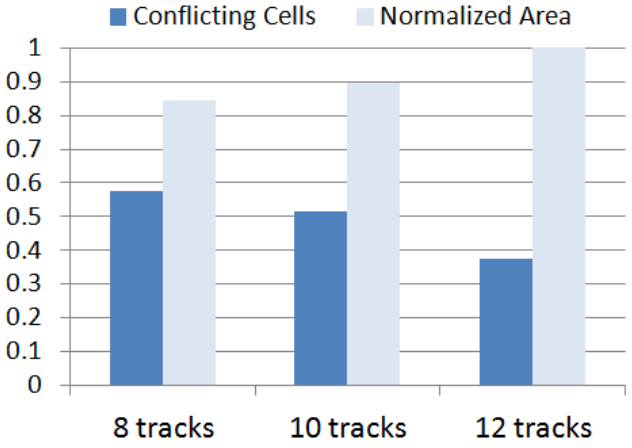


Figure 6.12: *Design Conflicts* and design area for 8-track, 10-track, and 12-track cell-heights.

The proposed methodology is also applied to study the effects of other design rules (layout styles) on DP-compatibility. In one experiment, we vary the cell-height while keeping all other design rules fixed and using a same-color spacing of 130nm. Three cell-heights were investigated: 8, 10 (baseline), and 12 M2 tracks. Results, depicted in Figure 6.12, show that a 10-track cell-height for Nangate library seems to be a good compromise between area and DP-compatibility.

A study of different local interconnect (LI) schemes is also conducted using our methodology. Assuming polysilicon is one-dimensional, three cases were investigated: LI replaces M1 to perform gate-to-gate connections (if possible), LI replaces M1 to perform short connections of neighboring gates only (i.e., "Limited LI"), and

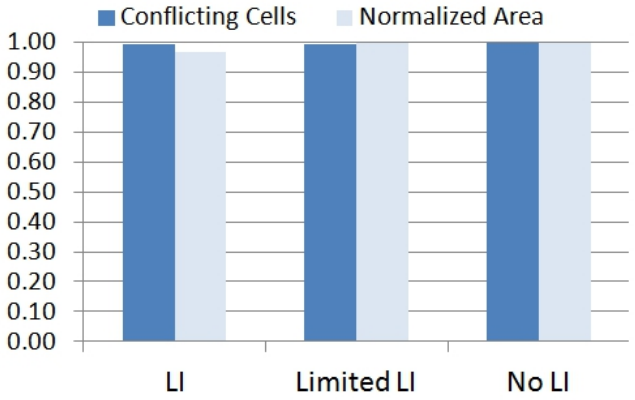


Figure 6.13: *Design Conflicts* and design area by using: local interconnects (LI) to perform poly-to-poly connections, LI to perform connections of only neighboring poly gates, and no LI.

"no LI" where all gate-to-gate connections are performed on M1. The results depicted in Figure 6.13 show that DP-conflicts as well as area are insensitive to the LI scheme. An important conclusion is that adding a horizontal LI layer, to wire poly and relax the M1 layer, does not bring any noticeable benefits with this particular process (i.e., FreePDK 45nm process [Fre]).

6.2 A Methodology for Exploring the Interaction Between Design Rules and Overlay Control

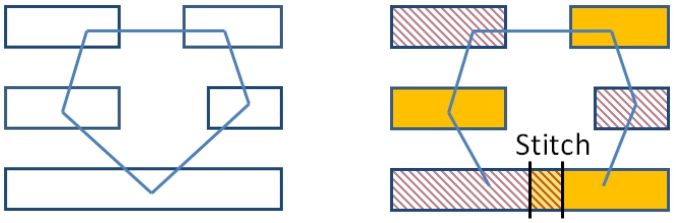
As technology scaling continues, overlay control is becoming more important than ever to allow smaller and smaller feature sizes. Moreover, the introduction of multiple-patterning (MP) lithography, where overlay effectively translates into CD variability [Arn08, Dus07], has made overlay control even more critical and more challenging. Meeting the requirements for overlay control is believed to be one of the biggest challenges for deploying MP technology [ADF06].

Overlay has been traditionally modeled using a linear model with major overlay components of translation, magnification, and rotation in the wafer and field coordinate systems [LLD08, CC01]. This linear model required a simple 2-point alignment. In recent years, the industry has moved toward high-order overlay modeling

and more sophisticated alignment strategies, which requires more overlay sampling and excessive alignment [WCH09, EHO08, WIY07, Lev10, CLB08]. For example, the work in [CLB08] suggests high-order process control by overlay control with one model per lot or one model for every wafer; the work in [WCH09] proposes high-order wafer alignment, while the work in [WIY07] proposes exposure tool characterization using off-line overlay sampling. These improvements in overlay control are capable of reducing overlay errors considerably (by up to 30% [WCH09, WIY07]) when a high-order overlay model is used. On the downside, high-order modeling of overlay requires more advanced exposure scanners, more alignment measurements, and excessive off-line overlay sampling. Hence, the overlay improvement of high-order modeling comes at a huge cost in tool migration and diminished throughput capability due to the additional measuring time.

Design rules that define interactions between different layers (e.g., metal overhang on via rule) or different mask-layouts of the same layer (e.g., mask overlap) effectively serve as guard band for overlay errors. For defining these rules during process development, a prediction of the yield loss due to overlay is needed. If overlay is characterized entirely as a field-to-field error, then the probability of survival (POS) for the die is equal to the POS of the most overlay-critical spot in the layout, say k. On the other extreme, if overlay is characterized entirely as a random within-field variation, then POS of the die is k^n, where n is the total number of critical spots in the design. Hence, depending on the overlay characteristics, rules can either be grown to suppress yield loss or shrank to reduce the layout area.

We offer a general framework for exploring the interaction between design rules, overlay characteristics, and overlay-modeling options. We develop a model for yield loss from overlay that considers overlay characteristics including the residue after overlay correction and the breakdown between field-to-field and within-field overlay. The proposed framework is the first of its kind and it can be *applied during process development to better define overlay-related design rules* and *to project the overlay requirement of the process*. For demonstration purposes, the framework was used in this work to explore DP and overlay-related rules for the M1 layer. The framework



(a) (b)

Figure 6.14: Example of a DP-problematic layout pattern with an odd cycle in its conflict graph (a) that was broken by introducing a stitch (b).

is more general, however, and can be used to explore other inter-layer overlay rules, for different MP technologies, and at other layers.

6.2.1 Design Rules and Overlay Interaction

In this work, we focus particularly on DP-related design rules and their interaction with overlay, namely the mask-overlap length rule and the minimum line-width and spacing design rules.

The overlap-length rule is triggered whenever a stitch is introduced between the different mask layouts of the same layer. Although stitches may be a cause for yield loss, stitching is needed to conform many problematic layout patterns to DP without the need for layout modification (by breaking odd cycles in the conflict graph as in the example of Figure 6.14).

One of the main reasons for yield loss associated with stitches is overlay errors between the first and second exposures in DP. Therefore, the minimum overlap-length rule – a.k.a. overlap margin – has a direct impact on yield. Consider for example a stitch in the center of a vertical line as shown in Figure 6.15. An overlay in the Y direction may result in an insufficient mask overlap and cause an open defect after line-end pullback; an overlay in the X direction may cause the wire to become too narrow at the stitch leading to failure. In addition, the overlap-length rule affects the DP-compatibility of the layout. The larger the overlap length is, the lesser candidate-stitch locations the layout will have. Hence, while a large and

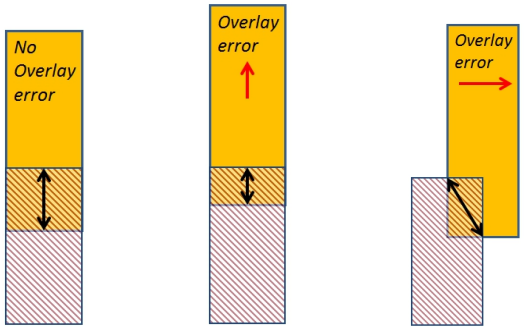


Figure 6.15: Example of a stitch in a vertical line (a), a possible failure with overlay error in Y direction that may occur after lineend pullback (b), and a possible failure with overlay error in X direction due to narrowing (c).

conservative overlap-length rule is likely to inhibit most yield loss of stitches caused by overlay, such overlap length may result in excessive re-design efforts and area overhead to ensure the layout conforms to DP. Another design rule that may affect the yield loss of stitches due to overlay (in x direction for the example in Figure 6.15) is the line-width rule. Clearly, failure from narrowing for initially narrow lines is more severe than such failure in wide lines.

The minimum line-spacing design rule impacts the delay variation of wires caused by overlay errors between the two exposures of DP [GG10, YP08, CN09, JKT09]. Since overlay translates directly into line-spacing variation (with a positive dual-line process), the coupling capacitance between neighboring wires on different exposures will be affected by both overlay and the minimum line-spacing rule. The line-spacing rule has also a direct impact on the layout area. While a large line-spacing rule may confine the wire-delay variation, such spacing rule is likely to induce an area overhead.

6.2.2 Overlay and Yield Modeling

The yield from overlay, $Y_{overlay}$, is equal to the probability of survival (POS) from the overlay error remaining after any overlay correction and referred to as residue[11].

[11] Coupled with the lithographic line-end pullback, which we model as an offset of fixed value.

Overlay-residue vector components in x and y directions are typically described by a normal distribution with zero mean and process-specific 3σ estimate. Therefore, given the fraction, p, of the overlay-residue variance breakdown between field-to-field and within-field components, the probability distribution of each type of overlay error can be calculated as follows:

$$\begin{aligned} f_{field-to-field} &= \frac{1}{\sigma\sqrt{2\pi p}} e^{\frac{-u^2}{2p\sigma^2}}, \\ f_{within-field} &= \frac{1}{\sigma\sqrt{2\pi(1-p)}} e^{\frac{-v^2}{2(1-p)\sigma^2}}. \end{aligned} \tag{6.2}$$

The probability for each type of overlay error to have a value between a and b is then given by

$$\begin{aligned} P_{field-to-field} &= \frac{1}{\sigma\sqrt{2\pi p}} \int_a^b e^{\frac{-u^2}{2p\sigma^2}} \, \mathrm{d}u, \\ P_{within-field} &= \frac{1}{\sigma\sqrt{2\pi(1-p)}} \int_a^b e^{\frac{-v^2}{2(1-p)\sigma^2}} \, \mathrm{d}v. \end{aligned} \tag{6.3}$$

We make the assumption that the field-level overlay residue at every feature of the same layer in the design is independent. Thus, the yield from overlay in one direction, which is equivalent to the probability of all features – say n – in the design surviving such overlay error, is determined as follows:

$$Y_{x|y} = \frac{1}{\sigma\sqrt{2\pi p}} \int_{-r}^{r} \int_{-r-u}^{r-u} \left(\frac{e^{\frac{-v^2}{2(1-p)\sigma^2}}}{\sigma\sqrt{2\pi(1-p)}} \mathrm{d}v \right)^n e^{\frac{-u^2}{2p\sigma^2}} \mathrm{d}u, \tag{6.4}$$

where r can take either the value of the design rule, when an overlay greater than the rule value causes a failure, or $-\infty$, when the overlay in a particular direction effectively increases the overlap at the feature.

Finally, the overall yield from overlay in any direction is approximated as the product of the yield in the x and y directions[12]:

$$(Y)_{overlay} = (Y)_x \times (Y)_y. \tag{6.5}$$

[12]This equation slightly underestimates the yield loss as, in reality, yield loss from overlay is defined by the area of the overlap region, which is influenced by overlay in both x and y directions.

6.2.3 Evaluation of Rules Impact on Design

This section presents the methods we used for evaluating the design impact of overlay-related rules.

Evaluation of Design Area

Our evaluation for the design area associated with DP and overlay rules is achieved using the Design Rules Evaluator (DRE) from [GG09a]. To evaluate area, DRE essentially creates a virtual standard-cell layouts from a set of DRs and transistor-level netlists of standard-cells. Using estimated area of the virtual layouts as well as instances-count of cells in the design, the total cell-area in the design is evaluated.

Evaluation of DP-Compatibility

A layout is said to be DP-compatible, if its features can be assigned to the first and second masks without any spacing violations in each mask-layout. Hence, we choose the number of spacing violations as our metric for DP-compatibility. We use the mask-assignment algorithm of [GAN11], which guarantees to a mask-assignment solution if one exists. To further reduce the number of spacing violations in DP-incompatible layouts, we modify the algorithm to flip the mask-assignment of violating features if the flipping reduces the number of violations.

Evaluation of Overlay-Induced Delay Variation

We use the method described in [GG10] to evaluate the electrical variation of wires formed with DP. In essence, the method consists of modeling the wire resistance and capacitance, which are the main elements of wire delay, as a function of overlay and its different components. Since the method in [GG10] assumes a linear overlay model, we limit our experiments on the minimum line-spacing rules to the case of overlay control with a linear model.

6.2.4 Experimental Results

In this section, we explore DP-related design rules and their interaction with overlay at the 14nm technology node.

Testing Setup

Our experiments were performed using AE18 microprocessor design from [opea], synthesized using Nangate Open Cell-Library [Nanb], and FreePDK open-source process [Fre]. Since the PDK and standard cell-library are for a 45nm process, all rules and layouts were scaled down by a factor of $frac12 \times \sqrt{2}$ to run the experiments for the 14nm node (M1 half-pitch becomes 23nm).

Since the area of the benchmark design is relatively small (10K-cell instances design), we normalize the yield results to a $100mm^2$ die area to have a realistic number of structures that are susceptible to yield loss (i.e., number of stitches in our experiments). We determine for the base case in each experiment the number of design copies that can fit in $10 \times 10mm$ chip size and find the corresponding number of stitches in the design[13].

Projecting the Process' Overlay Capability

In the first experiment, the framework is used to analyze the yield loss of stitches in DP for a variety of overlay-correction residues of associated with different models. First, third, and sixth-order models for field-to-field (i.e., wafer level) and within-field (i.e., field level) overlay correction. Residue values are taken from [EHO08] after scaling to the 14nm node[14]. Figure 6.16 plots the yield of stitches due to overlay errors for the different cases. In this experiment, we assume a minimum mask-overlap length rule of 14nm, which results in a total of 166 million stitches

[13]Note that, in our experiments, we only use the number of stitches and assume half the stitches are in vertical lines and the other half are in horizontal lines to estimate the yield loss as in Equation 6.5. Layout context effects for more accurate modeling is part of ongoing work.

[14]The residue values in [EHO08] correspond to the 32nm half-pitch node, which is equivalent to the 22nm process node. So, we scale them down by a factor of $\frac{1}{\sqrt{2}}$ for 14nm process node.

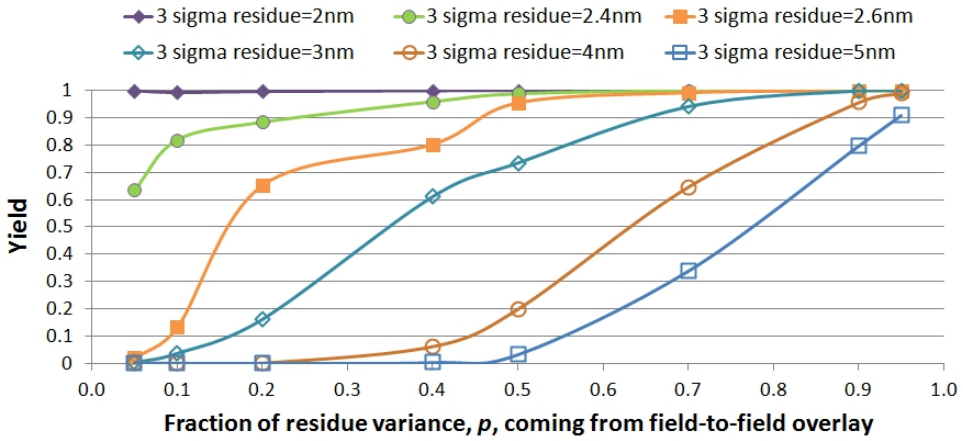


Figure 6.16: Plots showing the effects of the breakdown of overlay among field-to-field and within-field overlay components for different overlay-residue values.

after normalization to a $100mm^2$ die area. The results show that the larger the fraction of within-field overlay component, the larger the yield loss. The plots also identify the value of the residue for which a close to 100% yield can be achieved for a given overlay breakdown between field-to-field and within-field components. Such result can project the overlay capability of the process and serve as early hint for design-rules development.

Interaction between DP-Related Rules and Overlay Control

We also use the framework to study the effects of DP rules on stitch failure and the area and DP-compatibility of the design. In this case, we assume a stitch fails when the overlap is less than 10nm in either directions (see Figure 6.15 for possible failures at stitches). In one experiment, we vary the line-width by few nanometers from the nominal value at 23nm and report the yield loss and the normalized design area for the different overlay-modeling options. The results, depicted in Figure 6.17, show that the line-width has almost no impact on stitch failure. The reason is that the nominal rule value is large enough to avoid stitches failure from overlay in the direction perpendicular to lines. Hence, stitches yield loss may be neglected when deciding on the minimum line-width rule. It can also be clearly seen from Figure 6.17 that the first-order wafer/first-order field-level overlay model, i.e., the linear model, is insufficient for controlling overlay at the 14nm node.

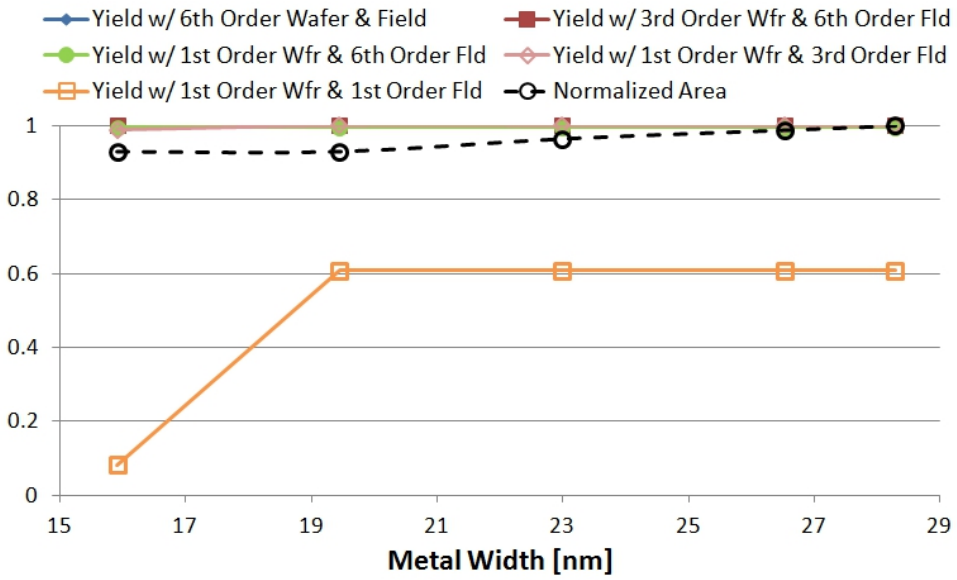


Figure 6.17: Plots showing the interaction between the minimum line-width rule and overlay control and their impact on yield and layout area of the design with minimum overlap-length rule of 14nm.

In another experiment, we vary the minimum mask-overlap length and report the yield loss and number of DP-spacing violations in the design for the different overlay-modeling options. The results, depicted in Figure 6.18[15], show the strong interaction between the rule value and overlay-control options as well as the overall impact on yield and DP-compatibility. Interestingly, few nanometer changes in the rule value may allow the use of a less stringent overlay model without significant impact on DP-compatibility. For example, increasing the minimum mask-overlap length from 13nm to 14nm would allow the use of a 1st-order wafer/sixth-order field-level overlay model instead of a third-order wafer/sixth-order field-level model.

Our last experiment is about studying the effects of the line-spacing rule on wire-delay variation and layout area. We vary the line-spacing rule from the nominal value at 23nm by plus and minus few nanometers. The results, given in Figure 6.19[16], indicates that the impact of this rule on the average RC variation is minor, while its impact on area is considerable. Hence, tweaking the line-spacing rule with the

[15]The number of DP-spacing violations are normalized with respect to the case with the largest number and DP mask-assignment of the layouts was performed using a minimum same-color spacing of 1.5× the half-pitch.

[16]Note that there is always some electrical variation due to overlay errors with any realistic line-spacing rule.

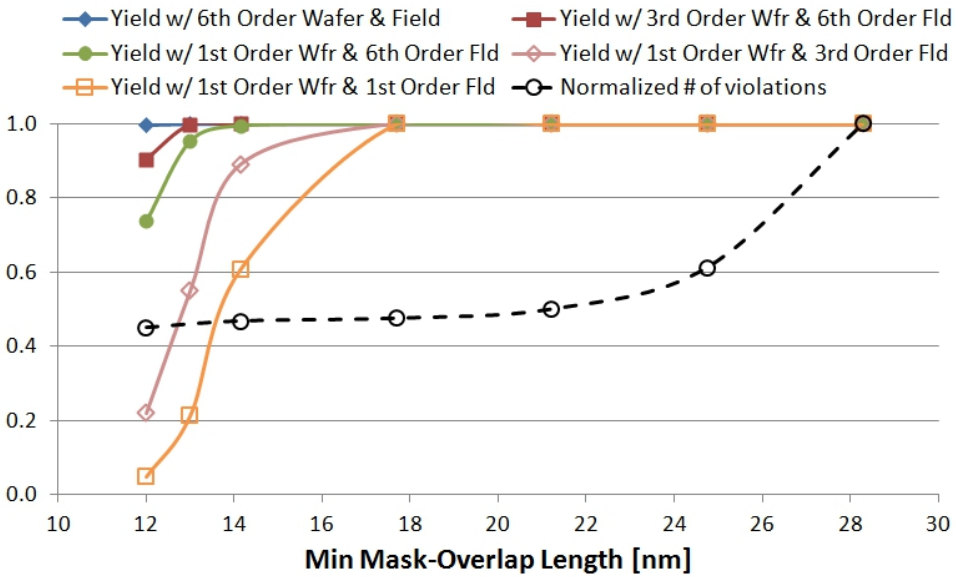


Figure 6.18: Plots showing the interaction between the overlap-length rule and overlay control and their impact on yield and DP-compatibility of the design at the nominal line-width of 23nm.

intention of reducing the electrical variation is ineffective.

Conclusions and Future Research

In this chapter, we presented the first work on early evaluation and exploration of multiple patterning rules intended for speeding up the rules-development cycle. The proposed methodology consists of a novel layout/congestion estimation method and a machine-learning based DP-conflict predictor. The methodology was used to explore DP and layout rules and investigate their effects on DP-compatibility and layout area. *Although the focus of this chapter was on DP, the methodology is more general and can be applied to explore rules of other layout-restrictive technologies, such as multiple patterning, self-aligned double patterning, and directed self-assembly.* Essentially, all that is needed is a layout conflict checker to train the machine learning model. Our ongoing work also explores using the same methodology to assess the impact of forbidding the use of certain layout patterns.

We also proposed in this chapter a general framework to explore the interactions between design rules, overlay characteristics, and overlay modeling options. Yield

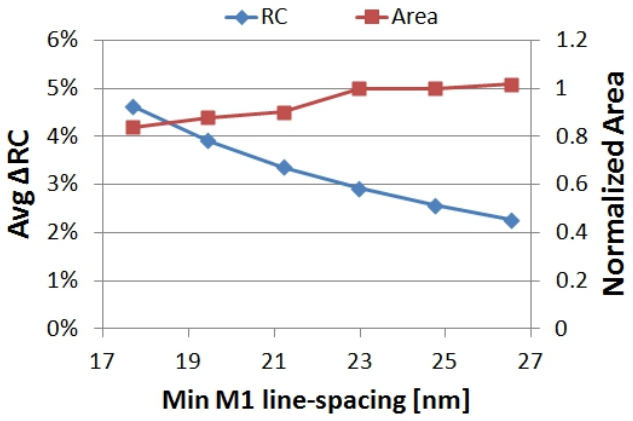


Figure 6.19: Plot for the average ΔRC and the normalized design area for different values of the minimum line-spacing rule.

loss due to overlay is modeled as a function of design-rule values and the overlay characteristics. The proposed framework is the first of its kind and it can be used during process development to better define overlay-related design rules and project overlay requirements for the process. For demonstration purposes, the framework was used in this work to explore DP and overlay-related rules for the M1 layer at the 14nm node. Important conclusions could be drawn from our experimental results. One result shows that increasing the minimum mask-overlap length by 1nm would allow the use of a third-order wafer/sixth-order field-level overlay model instead of a first-order wafer/sixth-order field-level model without a significant impact on design. Another result shows that the minimum line-width and spacing rules have an insignificant impact yield and electrical variation. Although our studies were performed for DP rules at the M1 layer, the framework is more general and can be used to explore other inter-layer overlay rules, for different MP technologies, and for different layers. In future work, we will extend our yield and design-impact analysis to a chip-level analysis across all layers in the design and explore other overlay-related rules including ones that define interaction between different layers. Also, yield modeling for systematic overlay errors is part of our ongoing work.

Chapter 7

Conclusions and Future Research

Semiconductors have fueled wealth creation, making new applications (cost-) feasible with each successive technology generation. Keeping Moores Law alive would require rapid technology changes over the next decade and beyond. Huge capital and human resources investments, essential for integrated circuit design and manufacturing, require accurate projection of design implications of device and technology changes.

The work from this dissertation will help making such projections possible. It will help the design community avoid unforeseen disruptions as well as allow the technology-development community to evaluate alternative technologies from the perspective of improving key circuit-design metrics. The infrastructure has seen a lot of interest from industry and some companies are already making use of it [PPL11].

One disruptive manufacturing technology that is very likely to see wide adoption is multiple-patterning technology. The methods and techniques provided in this dissertation will be help in ensuring the profitability and timely adoption of this technology.

7.1 Research Contributions

This dissertation made important contribution to the topics of design enablement of technology and design/technology co-optimization.

In regards to design/technology co-optimization, we proposed a computational infrastructure for the systematic, hence early evaluation of design rules, technology choices, and layout methodologies (Chapter 5 and 6). The infrastructure is the first

of its kind and will enable fast, systematic, and benchmarked evaluation of design and technology choices, thereby redirecting research efforts toward options that are more likely to get adopted. The infrastructure will allow researchers from different fields – circuits/design/fabrication – to conduct technology exploration studies.

- Device and interconnect technology researchers will be able to evaluate novel device architectures and wiring schemes in a circuit context easily in an automated and standardized fashion for a variety of circuit types (e.g., low power, high-performance, wire-dominated, etc). "What-if" studies to identify preferred device characteristics will trigger new, useful device types.

- Fabrication technology researchers and developers will get quick feedback on the design impact of choices they consider (as opposed to waiting several weeks for feedback from design teams). This will expedite technology development and allow the exploration of a wide range of choices, which otherwise may never be considered because of schedule limitation.

- Circuit and design automation researchers can evaluate novel design/layout optimization algorithms and circuit styles for future technologies.

In regards to design enablement, the contribution of this dissertations is summarized as follows.

- We offered a general methodology for the automatic adaptation of layouts to multiple-patterning technology, which can be applied for DP/TP as well as SADP. This legalization is performed simultaneously across all layout layers while minimizing perturbation. The method enables designing with conventional design rules and masks the designer from the complexity in dealing with MP layers and requirements. The way we formulate the problem allowed us to achieve high-quality results with extremely fast run-time (less than 10 seconds in real time for typical cells). Although the method targets primarily standard-cell layouts, it can also be applied for small full-custom layouts and interconnect layers in complete designs.

- We presented a O(n) heuristic algorithm for the layout decomposition of double-patterning and extend it for triple-patterning (first two sections of Chapter 2). The algorithm is the fastest among the many available layout decomposition algorithms and achieves a violation-free coloring solution when one exists.

- We studied the design-level electrical and yield impacts of overlay in double-patterning and the relative importance of the different components of overlay (Chapter 3 and 6). The methods developed for this study can be used to bring design-awareness into process decision-making, thereby advancing more design-friendly manufacturing and reducing overlay-control requirements.

- We proposed a novel technique for DP that uses a single mask for the two exposures (Chapter 4). This lithography technique is viable for bidirectional layouts and is very suitable for regular and gridded layouts, which are gaining popularity with continued scaling in the sub-wavelength regime. The proposed technique has many benefits over conventional double-patterning in terms of mask-cost, manufacturing throughput, and process control. It is especially attractive for projects of low-volume manufacturing where mask-cost constitutes a large portion of the overall cost.

7.2 Future Research

This dissertation provided an infrastructure for interesting future research.

We envision a more general and modular design-centric assessor of technology (depicted in Figure 7.1) by integrating different frameworks presented in this thesis and extending them.

Due to the complexity of layout/patterning interactions, "pattern-based" rules may sometimes be preferred over simple width and spacing rules (e.g., [DCY09]). Although the Design Rules Evaluator from Chapter 5 is suitable for dimension-based rules, it currently does not support pattern-based rules. Two approaches can be followed to evaluate pattern-based rules.

The first approach is by extending our wiring-estimation method of Section 6.1.1. Specifically, during the enumeration of possible wiring solutions in the probabilistic wiring-estimation method, we can check whether a solution would create a forbidden pattern at any of the tiles (similar to the existing check for design-rule violations) and ignore such wiring solution. This avoidance of certain wiring solutions would translate into higher tile-level and overall congestion, possibly leading to an area increase. The limitation of this approach is that we consider only a limited number of wiring solutions, ones that follow the single-trunk Steiner tree topology only, and these solutions may never trigger the occurrence of certain forbidden patterns (e.g., patterns with a specific tip-to-side spacing).

Another way to block forbidden patterns and evaluate the "cost" of such blocking is by using a real-layout-based approach. In particular, by integrating the layout legalization framework from Chapter 2, we can define a set of design rules that restrict the use of forbidden patterns in the layout. The rules will then be set as constraints in the linear program of compaction. Compaction will in turn legalize rule violations and remove the forbidden patterns from the layout by possibly increasing the area. The area increase will represent the cost for blocking forbidden patterns while the electrical impact can be accurately evaluated by applying existing analysis methods on the final layout (e.g., parasitics extraction).

Although the latter approach is more accurate than the former, it is more time-consuming and assumes the existence of efficient pattern-based compaction methods as well as correct baseline layouts for the technology. Ideally, we would want to use the latter approach only for those patterns that the former approach cannot handle and possibly at advanced stages of technology development.

By adding a device/circuit modeling component to the mix, metrics of power/performance/reliability metrics can be included to the assessment results. The new metrics will allow the use of a unified metric for the assessment, such as the number of functional dies meeting power/performance/reliability constraints.

An accelerated physical-design generation component can be added to the plat-

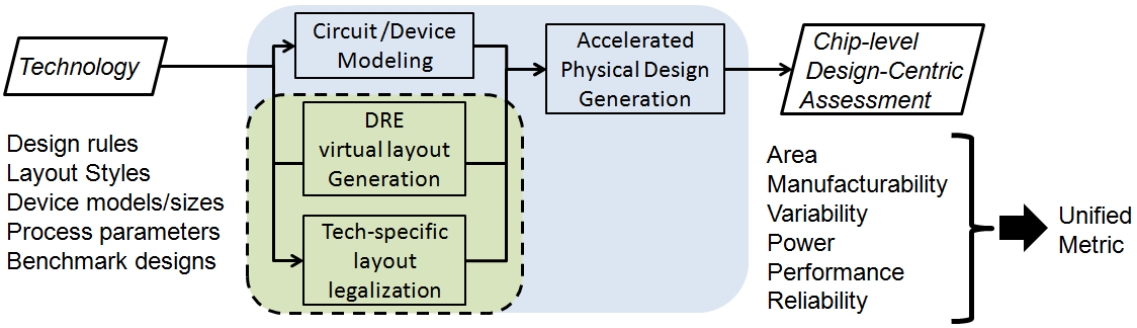


Figure 7.1: Generalized design-centric assessor of technology.

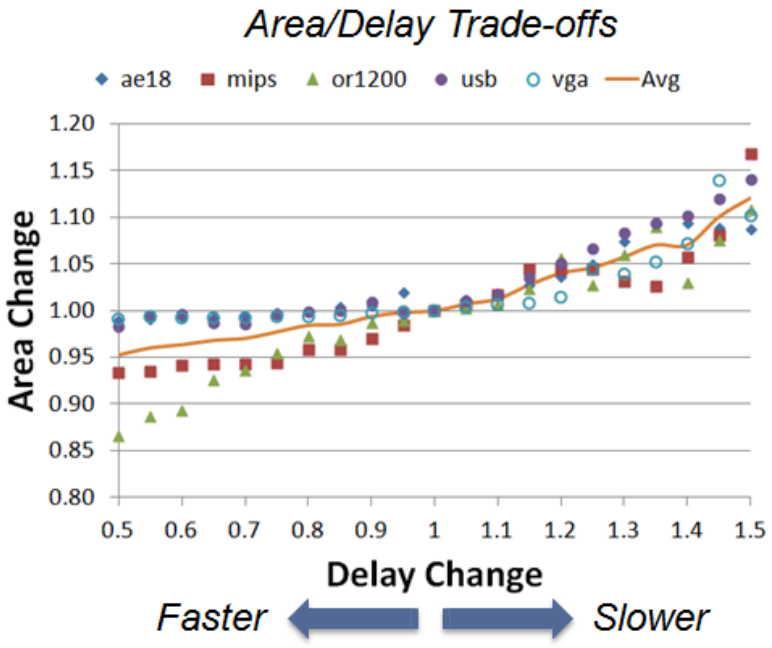


Figure 7.2: Curves are generated based on purely-empirical experiments using existing synthesis/placement/routing tools.

form to perform the assessment at the chip level. A chip-level assessment is more adequate for predicting the impact of technology on design. Such an assessor will allow studying interesting trade-offs like the ones that occur between variability and area (see Figure 7.2). It will also allow the evaluation of a wider range of rules, namely rules and constraints in BEOL layers as well as pin-access and cell-abutment strategies.

References

[ACC08] C. Auth, A. Cappellani, J.-S. Chun, A. Dalis, A. Davis, T. Ghani, G. Glass, T. Glassman, M. Harper, M. Hattendorf, P. Hentges, S. Jaloviar, S. Joshi, J. Klaus, K. Kuhn, D. Lavric, M. Lu, H. Mariappan, K. Mistry, B. Norris, N. Rahhal-orabi, P. Ranade, J. Sandford, L. Shifren, V. Souw, K. Tone, F. Tambwe, A. Thompson, D. Towner, T. Troeger, P. Vandervoorn, C. Wallace, J. Wiedemer, and C. Wiegand. "45nm High-k + metal gate strain-enhanced transistors." In *VLSI Technology, 2008 Symposium on*, pp. 128–129, june 2008.

[ADF06] W. Arnold, M. Dusa, and J. Finders. "Manufacturing challenges in double patterning lithography." In *IEEE Intl. Symp. Semiconductor Manufacturing*, pp. 283–286, Sept 2006.

[ARG09] C. Augustine, A. Raychowdhury, Y. Gao, M. S. Lundstrom, and K. Roy. "PETE: a device/circuit analysis framework for evaluation and comparison of charge based emerging devices." In *Proc. Intl. Symp. Quality Electronic Design ISQED'09*, pp. 80–85, 2009.

[Arn08] W. H. Arnold. "Toward 3nm overlay and critical dimension uniformity: an integrated error budget for double patterning lithography." In *Proc. SPIE*, volume 6924, p. 692404, 2008.

[Bai07] G.E. Bailey et al. "Double pattern EDA solutions for 32nm HP and beyond." In *Proc. SPIE*, volume 6521, p. 65211K, 2007.

[Bal08] A. Balasinski et al. "Patterning techniques for next generation IC's." In *Proc. SPIE Microelectronics:Design, Technology, and Packaging*, volume 6798, p. 679803, 2008.

[Bee07] M. de Beeck et al. "Manufacturability issues with double patterning for 50-nm half-pitch single damascene applications using RELACS shrink and corresponding OPC." In *Proc. of SPIE*, p. 65200I, 2007.

[BKP09] K. Bubke, R. de Kruif, J. H. Peters, M. Dusa, and B. Connolly. "Mask characterization for double patterning lithography." *J. Micro/Nanolith. MEMS MOEMS*, **8**(1):011004, 2009.

[BML11] Y. Ban, A. Miloslavsky, K. Lucas, S.-H. Choi, C.-H. Park, and D. Z. Pan. "Layout decomposition of self-aligned double patterning for 2D random logic patterning." In *Proc. SPIE*, volume 7974, p. 79740L, 2011.

[Boo] "Boost C++ Libraries." http://www.boost.org.

[Cala] "Calibre IC Verification & Signoff." http://www.mentor.com/products/ic_nanometer_design/verification-signoff/.

[Calb] "Calibre Standard Verification Rule Format Manual v2009." http://www.mentor.com/.

[CC01] 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.

[CC09] S.-Y. Chen and Y-W. Chang. "Native-conflict-aware wire perturbation for double patterning technology." In *IEEE Intl. Conf. on Computer-Aided Design*, pp. 556–561, 2009.

[CGG10] T.-B. Chan, R. S. Ghaida, and P. Gupta. "Electrical modeling of lithographic imperfections." In *Intl. Conf. on VLSI Design, VLSID'10*, pp. 423–428, 2010.

[CGK04] L. Capodieci, P. Gupta, A. B. Kahng, D. Sylvester, and J. Yang. "Toward a methodology for manufacturability-driven design rule exploration." In *Proc. IEEE/ACM Design Automation Conference*, pp. 311–316, 2004.

[CHA92] J.-H. Chern, J. Huang, L. Arledge, P.-C. Li, and P. Yang. "Multilevel metal capacitance models for CAD design synthesis systems." *IEEE Electron Device Letters*, **13**(1):32–34, January 1992.

[Cha09] S. Chang et al. "Exploration of complex metal 2D design rules using inverse lithography." In *Proc. SPIE*, volume 7275, p. 72750D, 2009.

[Chi08a] T. Chiou et al. "Full-chip pitch/pattern splitting for lithography and spacer double patterning technologies." In *Proc. of SPIE Lithography Asia*, p. 71401Z, 2008.

[Chi08b] T.-B. Chiou et al. "Development of layout split algorithms and printability evaluation for double patterning technology." In *Proc. SPIE*, volume 6924, pp. 69243M–1, 2008.

[CK99] P. Chen and K. Keutzer. "Towards true crosstalk noise analysis." In *IEEE/ACM Intl. Conf. on Computer-Aided Design*, pp. 132–137, 1999.

[CLB08] D. Choi, C. Lee, C. Bang, D. Cho, M. Gil, P. Izikson, S. Yoon, and D. Lee. "Optimization of high order control including overlay, alignment, and sampling." In *Proc. SPIE*, volume 6922, p. 69220P, 2008.

[CN09] E. Y. Chin and A. R. Neureuther. "Variability aware interconnect timing models for double patterning." In *Proc. SPIE*, volume 7275, p. 727513, 2009.

[CNV90] A.H. Chao, E.M. Nequist, and T.D. Vuong. "Direct solution of performance constraints during placement." In *Custom Integrated Circuits Conference, 1990., Proceedings of the IEEE 1990*, pp. 27.2/1 –27.2/4, may 1990.

[CQZ92] H. Chen, C. Qiao, F. Zhou, and C.-K. Cheng. "Refined Single Trunk Tree: A Rectilinear Steiner Tree Generator for Interconnect Prediction." In *INTL. WORKSHOP ON SYSTEM-LEVEL INTERCONNECT PREDICTION (SLIP)*, pp. 85–89, 1992.

[Cul96] B. D. Cullity. *Combinatorial optimization: algorithms and complexity.* Prentice-Hall, New York, NY, 1996.

[CW08] C. Chu and Yiu-Chung Wong. "FLUTE: Fast Lookup Table Based Rectilinear Steiner Minimal Tree Algorithm for VLSI Design." *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, **27**(1):70 –83, jan. 2008.

[CYB08] M. Cho, K. Yuan, Y. Ban, and D. Z. Pan. "ELIAD: efficient lithography aware detailed router with compact post-OPC printability prediction." In *Proc. Design Automation Conference*, pp. 504–509, 2008.

[DCY09] V. Dai, L. Capodieci, J. Yang, and N. Rodriguez. "Developing DRC Plus rules through 2D pattern extraction and clustering techniques." In *Proc. SPIE*, volume 7275, p. 727517, 2009.

[Den11] Y. Deng et al. "DPT restricted design rules for advanced logic applications." In *SPIE*, volume 7973, p. 79730H, 2011.

[Dus07] M. Dusa et al. "Pitch doubling through dual-patterning lithography challenges in integration and litho budgets." In *Proc. SPIE*, volume 6520, p. 65200G, 2007.

[DWH07] M. Drapeau, V. Wiaux, E. Hendrickx, S. Verhaegen, and T. Machida. "Double patterning design split implementation and validation for the 32nm node." In *Proc. SPIE*, volume 6521, pp. 652109–1, 2007.

[Edn] "IBM Research lays out its approach to 22 nm scaling: it's software, not EUV." http://www.edn.com.

[EHO08] B. Eichelberger, K. Huang, K. O'Brien, D. Tien, F. Tsai, A. Minvielle, L. Singh, and J. Schefske. "32nm overlay improvement capabilities." In *Proc. SPIE*, volume 6924, p. 69244C, 2008.

[Eic08] B. Eichelberger et al. "32nm overlay improvement capabilities." In *Proc. SPIE*, volume 6924, p. 69244C, 2008.

[eye] "EYES Semiconductor Yield Estimation Tool." http://www.icyield.com/yieldmod.html.

[FCC12] S.-Y. Fang, S.-Y. Chen, and Y.-W. Chang. "Native-Conflict and Stitch-Aware Wire Perturbation for Double Patterning Technology." *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, **31**(5):703–716, may 2012.

[Fin08] J. Finders et al. "Double patterning for 32nm and below: an update." In *Proc. SPIE Optical Microlithography*, pp. 692408–1–692408–12, 2008.

[FM82] C. M. Fiduccia and R. M. Mattheyses. "A linear time heuristic for improving network partitions." In *Proc. ACM/IEEE Design Automation Conference*, pp. 175–181, 1982.

[Fre] "FreePDK." http://www.eda.ncsu.edu/wiki/FreePDK.

[GAN11] R. S. Ghaida, K. B. Agarwal, S. R. Nassif, X. Yuan, L. W. Liebmann, and P. Gupta. "A Framework for Double Patterning-Enabled Design." In *Intl. Conf. on Computer-Aided Design*, pp. 14–20, Nov. 2011.

[GG09a] R. S. Ghaida and P. Gupta. "A framework for early and systematic evaluation of design rules." In *Intl. Conf. on Computer-Aided Design*, pp. 615–622, Nov 2009.

[GG09b] R. S. Ghaida and P. Gupta. "Within-Layer Overlay Impact for Design in Metal Double Patterning." *IEEE Transactions on Semiconductor Manufacturing*, **23**(3):381–390, 2009.

[GG10] R. S. Ghaida and P. Gupta. "Within-Layer Overlay Impact for Design in Metal Double Patterning." *IEEE Transactions on Semiconductor Manufacturing*, **23**(3):381–390, Feb 2010.

[GJ79] M. R. Garey and D. S. Johnson. *A Guide to the Theory of NP-Completeness*. Freeman, 1979.

[GJK08] P. Gupta, K. Jeong, A. B. Kahng, and C. Park. "Electrical metrics for lithographic line-end tapering." In *Proc. SPIE*, volume 7028, p. 70283A, 2008.

[GJK09] M. Gupta, K. Jeong, and A. B. Kahng. "Timing yield-aware color reassignment and detailed placement perturbation for double patterning lithography." In *Proc. ACM/IEEE Intl. Conf. on Computer-Aided Design*, pp. 607–614, 2009.

[GK04] P. Gupta and A. B. Kahng. "Wire Swizzling to Reduce Delay Uncertainty Due to Capacitive Coupling." In *Proc. IEEE Intl. Conf. on VLSI Design*, pp. 431–436, 2004.

[GKK08] P. Gupta, A. B. Kahng, Y. Kim, S. Shah, and D. Sylvester. "Investigation of diffusion rounding for post-lithography analysis." In *Proc. Asia and South Pacific Design Automation Conference*, pp. 480–485, 2008.

[GKN06] P. Gupta, A. B. Kahng, S. Nakagawa, S. Shah, and P. Sharma. "Lithography simulation-based full-chip design analyses." In *Proc. SPIE*, volume 6156, p. 61560T, 2006.

[Goe] R. Goering. "How DRC Plus Makes DFM Easy at 28nm." http://www.cadence.com/Community/blogs/ii/archive/2010/10/25/how-drc-plus-makes-dfm-easy-at-28nm.aspx.

[GTG11] R. S. Ghaida, G. Torres, and P. Gupta. "Single-mask double-patterning lithography for Reduced Cost and Improved Overlay Control." *IEEE Transactions on Semiconductor Manufacturing*, **24**(1):93–103, Feb 2011.

[Gyv01] J. P. de Gyvez. "Yield modeling and BEOL fundamentals." In *Proc. IEEE/ACM System Level Interconnect Prediction SLIP'01*, pp. 135–163, 2001.

[Haz08a] A. J. Hazelton et al. "Double patterning requirements for optical lithography and prospects for optical extension without double patterning." In *SPIE Optical Microlithography XXI*, p. 69240R, 2008.

[Haz08b] A.J. Hazelton et al. "Double patterning requirements for optical lithography and prospects for optical extension without double patterning." In *Proc. SPIE*, volume 6924, p. 69240R, 2008.

[HCN09] C. Hsu, Y. Chang, and S. R. Nassif. "Simultaneous layout migration and decomposition for double patterning technology." In *IEEE Intl. Conf. on Computer-Aided Design*, pp. 595–600, 2009.

[HCN11] C.-H. Hsu, Y.-W. Chang, and S. R. Nassif. "Simultaneous Layout Migration and Decomposition for Double Patterning Technology." *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, **30**(2):284–294, Feb. 2011.

[HCT97] F. L. Heng, Z. Chen, and G. Tellez. "A VLSI artwork legalization technique based on a new criteria of minimum layout perturbation." In *ACM/IEEE Intl. Symp. on Physical Design*, pp. 116–121, 1997.

[HHL90] C. Hwang, Y. Hsieh, Y. Lin, and Y. Hsu. "A fast transistor-chaining algorithm for CMOS Cell Layout." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **9**(7):781–786, 1990.

[HMB07] H. Haffner, J. Meiring, Z. Baum, and S. Halle. "Paving the way to a full chip gate level double patterning application." In *SPIE/BACUS Photomask Technology*, p. 67302C, 2007.

[ICc] "IC Compiler." http://www.synopsys.com/Tools/Implementation/PhysicalImplementation/Pages/ICCompiler.aspx.

[ICK] http://www.icknowledge.com/.

[Ies08] U. Iessi et al. "Double Patterning Overlay and CD budget for 32 nm technology node." In *SPIE Optical Microlithography XXI*, p. 692428, 2008.

[ITRa] International Technology Roadmap for Semiconductors, Report 2011.

[ITRb] International Technology Roadmap for Semiconductors, Report 2007.

[JCS08] V. Joshi, B. Cline, D. Sylvester, D. Blaauw, and K. Agarwal. "Leakage power reduction using stress-enhanced layouts." In *Proc. ACM/IEEE Design Automation Conference*, pp. 912–917, 2008.

[Jee06] N. Jeewakhan et al. "Application of Dosemapper for 65-nm gate CD control: strategies and results." In *Proc. SPIE*, p. 63490G, 2006.

[JK09] K. Jeong and A. B. Kahng. "Timing analysis and optimization implications of bimodal CD distribution in double patterning lithography." In *Proc. Asia and South Pacific Design Automation Conference*, pp. 486–491, 2009.

[JKS08] K. Jeong, A. B. Kahng, and K. Samadi. "Quantified impacts of guardband reduction on design process outcomes." In *Proc. Intl. Symp. Quality Electronic Design ISQED'08*, pp. 790–797, 2008.

[JKT09] K. Jeong, A. B. Kahng, and R. O. Topaloglu. "Is overlay error more important than interconnect variations in double patterning?" In *System Level Interconnect Prediction*, pp. 3–10, June 2009.

[JRL10] T. Jhaveri, V. Rovner, L. Liebmann, L. Pileggi, A. J. Strojwas, and J. D. Hibbeler. "Co-optimization of circuits, layout and lithography for predictive technology scaling beyond gratings." *Trans. on Computer-Aided Design of Integrated Circuits and Systems*, **29**(4):509–527, April 2010.

[JSP09] T. Jhaveri, A. Strojwas, L. Pileggi, and V. Rovner. "Economic assessment of lithography strategies for the 22nm technology node." In *Proc. SPIE*, p. 74882Y, 2009.

[Kah03] A. B. Kahng. "Accurate pseudo-constructive wirelength and congestion estimation." In *Proc. ACM Int. Workshop on System-Level Interconnect Prediction*, pp. 61–68, 2003.

[Kay] R. Kay. "Rising Stakes In Semiconductor Game Squeeze Out All But A Few." http://www.forbes.com.

[KK98] I. Koren and Z. Koren. "Defect tolerance in VLSI circuits: techniques and yield analysis." In *Proc. of the IEEE*, volume 86, pp. 1819–1837, 1998.

[Kob08] S. Kobayashi et al. "Yield-centric layout optimization with precise quantification of lithographic yield loss." In *Proc. SPIE*, volume 7028, p. 70280O, 2008.

[KPX06] A. B. Kahng, C. Park, and X. Xu. "Fast dual graph-based hotspot detection." In *Proc. SPIE*, volume 6349, p. 63490H, 2006.

[KPX10] A. B. Kahng, C.-H. Park, X. Xu, and H. Yao. "Layout decomposition approaches for double patterning lithography." *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, **29**(6):939–952, 2010.

[Kuh55] J. P. S. Kuhn. "The Hungarian method for the assignment problem." *Naval Research Logistics Quarterly*, **2**:83–97, 1955.

[KWK07] R. Kim, T. Wallow, J. Kye, H. J. Levinson, and D. White. "Double exposure using 193nm negative tone photoresist." In *Proc. SPIE Optical Microlithography XX*, p. 65202M, 2007.

[Lai08] K. Lai et al. "32 nm logic patterning options with immersion lithography." In *Proc. SPIE*, volume 6924, p. 69243C, 2008.

[LaPa] M. LaPedus. "Foundries Tip Hybrid FinFET Flows." http://www.semimd.com.

[LaPb] M. LaPedus. "Intel pushes 193-nm litho down to 15-nm." http://www.eetimes.com/.

[LAQ07] Z. Li, C.J. Alpert, S.T. Quay, S. Sapatnekar, and W. Shi. "Probabilistic Congestion Prediction with Partial Blockages." In *Quality Electronic Design, 2007. ISQED '07. 8th International Symposium on*, pp. 841 –846, march 2007.

[Lec05] L. Lecarpentier et al. "Overlay measurement accuracy verification using CD-SEM and application to the quantification of WIS caused by BARC." In *SPIE Metrology, Inspection, and Process Control for Microlithography XIX*, pp. 1413–1423, 2005.

[Lee06] D. Lee et al. "Impact of registration error of reticle on total overlay error budget." *J. Vac. Sci. Technol.*, **B24**(6), 2006.

[Lee08] S. Lee et al. "An analysis of double exposure lithography options." In *Proc. SPIE Optical Microlithography*, p. 69242A, 2008.

[Lev44] K. Levenberg. "A method for the solution of certain non-linear problems in least squares." *Quarterly Journal of Applied Mathmatics*, **2**(2):164–168, 1944.

[Lev05] H. J. Levinson. *Principles of Lithography*. SPIE-International Society for Optical Engineering, Bellingham WA, 2 edition, 2005.

[Lev10] H. J. Levinson. *Principles of Lithography*. SPIE Press, Bellingham, WA, 2010.

[LH10] W-S. Luk and H. Huang. "Fast and lossless graph division method for layout decomposition using SPQR-tree." In *IEEE Intl. Conf. on Computer-Aided Design*, pp. 112–115, 2010.

[LHN04] M. Lavin, F. Heng, and G. Northrop. "Backend CAD flows for "restrictive design rules." In *IEEE/ACM Intl. Conf. on Computer Aided Design*, pp. 739–746, November 2004.

[Lie03] L. W. Liebmann. "Layout impact of resolution enhancement techniques: impediment or opportunity?" In *Proc. IEEE Intl. Symp. on Physical Design*, pp. 110–117, 2003.

[Lie04] L. W. Liebmann et al. "High-performance circuit design for the RET-enabled 65nm technology node." In *Proc. SPIE*, volume 5379, pp. 20–29, 2004.

[Lim06] C. Lim et al. "Positive and negative tone double patterning lithography for 50nm flash memory." In *SPIE Optical Microlithography XIX*, p. 615410, 2006.

[LLD08] D. Laidler, P. Leray, K. D'havé, and S. Cheng. "Sources of Overlay Error in Double Patterning Integration Schemes." In *Proc. SPIE*, volume 6922, p. 69221E, 2008.

[LLH01] L. Liebmann, J. Lund, F-L. Heng, and I. Graur. "Enabling alternating phase shifted mask designs for a full logic gate level: design rules and design rule checking." In *Proc. of Design Automation Conference, DAC'01*, pp. 79–84, 2001.

[LM04] Q. Liu and M. Marek-Sadowska. "Pre-layout wire length and congestion estimation." In *Design Automation Conference, 2004. Proceedings. 41st*, pp. 582 –587, july 2004.

[LPG11] L. Liebmann, D. Pietromonaco, and M. Graf. "Decomposition-aware standard cell design flows to enable double-patterning technology." In *SPIE*, volume 7974, p. 79740K, 2011.

[LPH09] L. Liebmann, L. Pileggi, J. Hibbeler, V. Rovner, T. Jhaveri, and G. Northrop. "Simplify to survive: prescriptive layouts ensure profitable scaling to 32nm and beyond." In *Proc. SPIE*, p. 72750A, 2009.

[LTK02] J. Lou, S. Thakur, S. Krishnamoorthy, and H.S. Sheng. "Estimating routing congestion using probabilistic analysis." *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, **21**(1):32 –41, jan 2002.

[Ma11] Y. Ma et al. "Double patterning compliant logic design." In *SPIE*, volume 7974, p. 79740D, 2011.

[Mac06] C. A. Mack. "How to characterize overlay errors." In *Yield Management Solutions*, pp. 46–47, 2006.

[Mac08] C. A. Mack. "Seeing double." *IEEE Spectrum*, pp. 46–51, 2008.

[Mai00] C. Mair et al. "An investigation of machine learning based prediction systems." *Journal of Systems and Software*, **53**(1):23–29, 2000.

[Mal85] W. Maly. "Modeling of lithography related yield losses for CAD of VLSI circuits." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **CAD-3**(3), 1985.

[Mar63] D. W. Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." *SIAM Journal on Applied Mathematics*, **11**(2):431–441, 1963.

[Mar08] T. Maruyama et al. "EBDW technology for EB shuttle at 65nm node and beyond." In *Proc. SPIE*, volume 6921, p. 69210H, 2008.

[Mat11] "MATLAB v. 7.13.0.", 2011.

[Mei09] H. Meiling et al. "EUVL system: moving towards production." In *Proc. SPIE*, volume 7271, p. 727102, 2009.

[Mis07] K. Mistry et al. "A 45nm logic technology with high-k+metal Gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging." In *IEEE Int. Electron Devices Meeting*, pp. 247–250, 2007.

[MKN09] J. Moon, C.-K. Kim, B.-S. Nam, B.-H. Nam, Y.-S. Hyun, S.-K. Kim, C.-M. Lim, Y.-D. Kim, M.-S. Kim, Y.-K. Choi, C.-R. Kim, and D. Yim. "Evaluation of shadowing and flare effect for EUV tool." pp. 727144–727144–10, 2009.

[MO07] H. Muta and H. Onodera. "Manufacturability-aware design of standard cells." *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, **E90-A**(12):2682–2690, December 2007.

[MSY12] Y. Ma, J. Sweis, H. Yoshida, Y. Wang, J. Kye, and H. J. Levinson. "Self-aligned double patterning (SADP) compliant design flow." In *Proc. SPIE*, volume 8327, p. 832706, 2012.

[MYP05] J. Mitra, P. Yu, and D. Z. Pan. "RADAR: RET-aware detailed routing using fast lithography simulations." In *Proc. Design Automation Conference*, pp. 369–372, 2005.

[Nana] "Nangate Open Cell Library v1.2. 2008." http://www.si2.org/openeda.si2.org/projects/nangatelib.

[Nanb] "Nangate Open Cell Library v1.3. 2009." http://www.si2.org/openeda.si2.org/projects/nangatelib.

[NN10] S. R. Nassif and K. J. Nowka. "Physical Design Challenges Beyond the 22nm Node." In *Proc. IEEE Intl. Symp. on Physical Design*, pp. 13–14, Mar. 2010.

[OGD] "OGDF: Open Graph Drawing Framework." http://www.ogdf.net.

[opea] http://www.opencores.org/.

[opeb] http://openeda.org.

[PPL11] C.-H. Park, D. Z. Pan, and K. Lucas. "Exploration of VLSI CAD researches for early design rule evaluation." In *Proc. Asia and South Pacific Design Automation Conference*, pp. 405–406, 2011.

[PSS03] L. Pileggi, H. Schmit, A.J. Strojwas, P. Gopalakrishnan, V. Kheterpal, A. Koorapaty, C. Patel, V. Rovner, and K.Y. Tong. "Exploring regular fabrics to optimize the performance-cost trade-off." In *Proc. IEEE/ACM Design Automation Conference*, 2003.

[Rai06] R. Raina. "What is DFM & DFY and why should I care?" In *IEEE Intl. Test Conference ITC'06*, pp. 1–9, October 2006.

[Rap] *Raphael Interconnect Analysis tool v2006 from Synopsys.*

[RIX94] A. Rubio, N. Itazaki, X. Xu, and K. Kinoshita. "An approach to the analysis and detection of crosstalk faults in digital VLSI circuits." **13**(3):387–395, March 1994.

[RRC07] E. Roseboom, M. Rossman, F.-C. Chang, and P. Hurat. "Automated full-chip hotspot detection and removal flow for interconnect layers of cell-based designs." pp. 65210C–65210C–9, 2007.

[Sak93] T. Sakurai. "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs." *IEEE Trans. on Electron Devices*, **40**(1):118–124, January 1993.

[SBS07] R. Singhal, A. Balijepalli, A. Subramaniam, F. Liu, S. Nassif, and Y. Cao. "Modeling and analysis of non-rectangular gate for post-lithography circuit simulation." In *Proc. of Design Automation Conference*, pp. 823–828, 2007.

[SCK91] A. Srinivasan, K. Chaudhary, and E.S. Kuh. "RITUAL: a performance driven placement algorithm for small cell ICs." In *Computer-Aided Design, 1991. ICCAD-91. Digest of Technical Papers., 1991 IEEE International Conference on*, pp. 48 –51, nov 1991.

[She05] Y. Sheu et al. "Modeling well edge proximity effect on highly-scaled MOSFETs." In *Proc. of IEEE CICC'05*, pp. 831–834, 2005.

[Shi08] W. Shiu et al. "Spacer double patterning technique for sub-40nm DRAM manufacturing process development." In *Proc. of SPIE Lithography Asia*, p. 71403Y, 2008.

[Siv11] S. Sivakumar. "EUV lithography: Prospects and challenges." In *Proc. Asia and South Pacific Design Automation Conference*, p. 402, jan. 2011.

[SLC08a] M. C. Smayling, H. Liu, and L. Cai. "Low k1 logic design using gridded design rules." In *Proc. of SPIE*, p. 69250B, 2008.

[SLC08b] M. C. Smayling, H. Liu, and L. Cai. "Low k1 Logic Design using Gridded Design Rules." In *Proc. SPIE*, volume 6925, p. 69250B, 2008.

[Soc10] B. Socha. "Method, Program Product and Apparatus for Performing a Model Based Coloring Process for Geometry Decomposition for Use in a Multiple Exposure Process.", 2010.

[Sou81] J. Soukup. "Circuit layout." *Proceedings of the IEEE*, **69**(10):1281 – 1304, oct. 1981.

[SS07] K. Suzuki and B. W. Smith. *Microlithography: science and technology*. Boca Raton: CRC Press, 2 edition, 2007.

[SSD10] M. C. Smayling, R. J. Socha, and M. V. Dusa. "22nm logic lithography in the presence of local interconnect." In *Proc. SPIE*, p. 764019, 2010.

[SSL05] R. Socha, X. Shi, and D. LeHoty. "Simultaneous source mask optimization (SMO)." In *Proc. SPIE*, volume 5853, pp. 180–193, 2005.

[SSS05] R. S. Shelar, S. S. Sapatnekar, P. Saxena, and X. Wang. "A predictive distributed congestion metric with application to technology mapping." *IEEE Trans. CAD*, **24**:696–710, 2005.

[SSW08] A. R. Subramaniam, R. Singhal, C. Wang, and Y. Cao. "Design rule optimization of regular layout for leakage reduction in nanoscale design." In *Proc. Asia and South Pacific Design Automation Conference*, pp. 474–479, 2008.

[Sta83] C. H. Stapper. "Modeling of integrated circuit defect sensitivities." *IBM Journal of Research and Development*, **CAD-3**(3):549–557, 1983.

[SW97] M. Stoer and F. Wagner. "A simple min-cut algorithm." *J. ACM*, **44**(4):585–591, Jul 1997.

[SYP06] S. X Shi, P. Yu, and D. Z. Pan. "A unified non-rectangular device and circuit simulation model for timing and power." In *Proc. IEEE/ACM Int'l Conf. on Computer-Aided Design*, 2006.

[Tag10] T. Taghavi et al. "New placement prediction and mitigation techniques for local routing congestion." In *IEEE Intl. Conf. on Computer-Aided Design*, pp. 621–624, 2010.

[TC11] X. Tang and M. Cho. "Optimal layout decomposition for double patterning technology." In *Intl. Conf. on Computer-Aided Design*, pp. 9 –13, Nov 2011.

[TGK08] A. Tritchkov, P. Glotov, S. Komirenko, E. Sahouria, A. Torres, A. Seoud, and V. Wiaux. "Double-patterning decomposition, design compliance, and verification algorithms at 32nm HP." **7122**:p. 71220S–1, 2008.

[TY06] X. Tang and X. Yuan. "Technology migration techniques for simplified layouts with restrictive design rules." In *IEEE/ACM Intl. Conf. on Computer Aided Design*, pp. 655–660, November 2006.

[TYY08] C. F. Tseng, C. C. Yang, E. Yang, T. H. Yang, K. C. Chen, and C. Y. Lu. "A comprehensive comparison between double patterning and double patterning with spacer on sub-50nm product implementation." In *Proc. of SPIE*, p. 69241Y, 2008.

[UCL] "UCLA_DRE v1.4." http://nanocad.ee.ucla.edu/Main/DownloadForm.

[VBR08] E. Vreugdenhil, H. Benten, L. Reijnen, G. Dicker, J. Gemmink, and F. Bornebroek. "Yield Aware Design of gate layer for 45 nm CMOS-ASIC using a high-NA dry KrF systems." In *Proc. of SPIE*, p. 69250D, 2008.

[Wak07] S. Wakamoto et al. "Improved overlay control through automated high-order compensation." In *SPIE Metrology, Inspection, and Process Control for Microlithography XXI*, p. 65180J, 2007.

[Wal09] T. Wallow. "Logic Double Patterning at Pitches Below 80 nm." Presentation in Sokudo Lithography Forum Semicon West, July 2009. http://www.sokudo.com/event/images/090715/SOKUDO_LBF2009_GLOBALFOUNDARIES.pdf.

[WBG04] J. Westra, C. Bartels, and P. Groeneveld. "Probabilistic congestion prediction." In *Proceedings of the 2004 international symposium on Physical design*, ISPD '04, pp. 204–209, New York, NY, USA, 2004. ACM.

[WCH09] R. Wang, CY Chiang, W. Hsu, R. Yang, J. Chiu T. Shih, J. Chen, and W. Lin. "Overlay improvement by ASML HOWA 5th alignment strategy." In *Proc. SPIE*, volume 7520, p. 752023, 2009.

[WIY07] S. Wakamoto, Y. Ishii, K. Yasukawa, A. Sukegawa, S. Maejima, A. Kato, J. C. Robinson, B. J. Eichelberger, P. Izikson, and M. Adel. "Improved overlay control through automated high-order compensation." In *Proc. SPIE*, volume 6518, p. 65180J, 2007.

[XC09] Y. Xu and C. Chu. "GREMA: Graph reduction based efficient mask assignment for double patterning technology." In *Intl. Conf. on Computer-Aided Design*, pp. 601–606, Nov. 2009.

[YCS06] J. Yang, L. Capodieci, and D. Sylvester. "Layout verification and optimization based on flexible design rules." In *Proc. SPIE*, volume 6156, p. 61560A, 2006.

[Yeh09] C. W. Yeh et al. "Contact formation with extremely low proximity effect by double patterning technology." In *Proc. of SPIE*, p. 72731E, 2009.

[YLC10] J.-S. Yang, K. Lu, M. Cho, K. Yuan, and D. Z. Pan. "A new graph theoretic, multi-objective layout decomposition framework for double patterning lithography." In *Proc. Asian and South Pacific Design Automation Conference*, pp. 637–644, 2010.

[YMH05] X. Yuan, K. W. McCullen, F-L. Heng, R. F. Walker, J. Hibbeler, R. J. Allen, and R. R. Narayan. "Technology migration technique for designs with strong RET-driven layout restrictions." In *ACM/IEEE Intl. Symp. on Physical Design*, pp. 175–182, 2005.

[Yon08] I. Yoneda et al. "Study of nanoimprint lithography for applications toward 22nm node CMOS devices." In *Proc. SPIE*, volume 6921, p. 692104, 2008.

[YP08] J. Yang and D. Z. Pan. "Overlay aware interconnect and timing variation modeling for double patterning technology." In *Intl. Conf. on Computer-Aided Design*, pp. 488–493, November 2008.

[YP09] K. Yuan and D. Z. Pan. "WISDOM: wire spreading enhanced decomposition of masks in double patterning lithography." In *IEEE Intl. Conf. on Computer-Aided Design*, pp. 32–38, 2009.

[YRS07] Y. Yamamoto, R. Rigby, and J. Sweis. "Multi-layer reticle (MLR) strategy application to double-patterning/double- exposure for better overlay error control and mask cost reduction." In *Proc. of SPIE Photomask Technology*, p. 67302X, 2007.

[YS07] B. Yenikaya and A. Sezginer. "A rigorous method to determine printability of a target layout." In *Proc. SPIE*, volume 6521, p. 652112, 2007.

[YYP09] K. Yuan, J-S. Yang, and D. Z. Pan. "Double patterning layout decomposition for simultaneous conflict and stitch minimization." In *Intl. Symp. on Physical Design*, pp. 107–114, 2009.

[YYZ11] B. Yu, K. Yuany, B. Zhang, D. Ding, and D. Z. Pan. "Layout decomposition for triple patterning lithography." In *Intl. Conf. on Computer-Aided Design*, pp. 1–8, 2011.

[ZBR08] R. Zayani, R. Bouallegue, and D. Roviras. "Levenberg-Marquardt Learning Neural Network for Adaptive Predistortion for Time-Varying HPA with Memory in OFDM Systems." In *Proc. of 16th European Signal Processing Conference*, 2008.

[ZCY08] Y. Zhang, J. Cobb, A. Yang, J. Li, K. Lucas, and S. Sethi. "32nm design rule and process exploration flow." In *Proc. SPIE*, volume 7122, p. 71223Z, 2008.

[ZFT05] J. Zhu, F. Fang, and Q. Tang. "Calligrapher: a new layout-migration engine for hard intellectual property libraries." *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, **24**(9):1347–1361, 2005.