Predictive Process Design Kits for the 7 nm and 5 nm Technology Nodes

by

Vinay Vashishtha

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2019 by the
Graduate Supervisory Committee:

Lawrence T. Clark, Chair
David R. Allee
Umit Ogras
Jae-sun Seo

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

Recent years have seen fin field effect transistors (finFETs) dominate modern complementary metal-oxide-semiconductor (CMOS) processes, [1][2], e.g., at the sub-20 nm technology nodes, as they alleviate short channel effects, provide lower leakage, and enable some continued V_{DD} scaling. However, a realistic finFET-based predictive process design kit (PDK) that supports investigation into both circuit and physical design, encompassing all aspects of digital design, for academic use has been unavailable. While the finFET-based FreePDK15 was supplemented with a standard cell library, it lacked full physical verification (LVS) and parasitic extraction at the time [3][4]. Consequently, the only available sub-45 nm educational PDKs are the planar CMOS based Synopsys 32/28 nm and FreePDK45 (45 nm PDK) [5][6]. The cell libraries available for those processes are not realistic since they use large cell heights, in contrast to recent industry trends. Additionally, the SRAM rules and cells provided by these PDKs are not realistic. Because finFETs have a 3D structure, which affects transistor density, using planar libraries scaled to sub-22 nm dimensions for research is likely to give poor accuracy.

Commercial libraries and PDKs, especially for advanced nodes, are often difficult to obtain for academic use, and access to the actual physical layouts is even more restricted. Furthermore, the necessary non-disclosure agreements (NDAs) are un-manageable for large university classes and the plethora of design rules can distract from the key points. NDAs also make it difficult for the publication of physical design as these may disclose proprietary design rules and structures.

This work focuses on the development of realistic PDKs for academic use that overcome these limitations. These PDKs, developed for the N7 and N5 nodes, even

before 7 nm and 5 nm processes were available in industry, are thus *predictive*. The predictions have been based on publications of the continually improving lithography, as well as estimates of what would be available at N7 and N5. For the most part, these assumptions have been accurate with regards to N7, except for the expectation that extreme ultraviolet (EUV) lithography would be widely available, which has turned out to be optimistic.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Lawrence T. Clark for his guidance throughout the course of my MS and PhD programs. He has been a constant source of inspiration and his advice has been invaluable to me.

I also thank my committee members Prof. David Allee, Prof. Umit Ogras, and Prof. Jae-Sun Seo. I am very grateful to my graduate advisors Sno Kleespies and Lynn Pratte for being exceedingly supportive and helpful. My thanks also to Jenna Snowberger for all her help. I am deeply obliged to the ASU Graduate College for the Completion Fellowship which was immensely helpful to me.

My sincere thanks to Aditya Gujja, Ankit Mittal, Ankita Dosi, Anudeep Reddy Gogulamudi, Bhargav Reddy Vemireddy, Chad Farnsworth, Chandarasekaran Ramamurthy, Christopher Lieb, Lovish Masand, Manoj Vangala, Nanda Kishore Babu Vemula, Parshant Rana, Parv Sharma, Punit Shah, Ramanjaneyulu Gudipati, Sai Bharadwaj Medapuram, Sai Chaitanya Jakkireddy, Sandeep Shambhulingaiah, Shivangi Mittal, Srivatsan Chellapa, Sushil Kumar, and Yitao Chen. This work would not have been possible without their help. I am also deeply indebted to my parents for everything.

Finally, I would like to express my gratitude to Dangdang Shao for her unwavering support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

1 INTRODUCTION

1.1 Outline

This work describes the developmental details and experiments related to the finFET based ASAP7 and ASAP5 predictive PDKs for the 7 nm (N7) and 5 nm (N5) complementary metal-oxide-semiconductor (CMOS) process nodes, respectively. Chapter 2 provides a discussion on salient lithography considerations, which include metrics for overlay, mask errors and other effects that limit layer patterning dimensions are described. Thereafter, modern liquid immersion optical lithography and its use in multiple patterning (MP) techniques that extend it beyond the standard 80 nm pitch limit are discussed. This sets the stage for a discussion of extreme ultraviolet (EUV) lithography, which can be used to pattern features at approximately 32 nm pitch in a single exposure, but at a high capital cost and low throughput.

Design-technology co-optimization (DTCO) has been required for recent CMOS process nodes to ensure that the very limited possible structures that can be practicably fabricated are usable to build real designs. Thus, a key part of a process development is not just to determine transistor and interconnect structures that are lithographically possible, but also ensuring that successful designs can be built with those structures.

When a process is still in development, designers are faced with “what if?” scenarios, where the process developers naturally wish to limit the process complexity. However, excessive limitations may make design overly difficult or lacking the needed density to make a new process node worthwhile. Those who must make decisions regarding cell architecture for future processes face significant challenges, as the target process is not fully defined. Specifically, bends in diffusions and gates have become increasingly untenable owing to the bend/corner rounding extent approaching the same order as patterned layer critical dimension (CD). To address these issues, middle of line (MOL) layers have been introduced to connect source drain regions and replace or augment some structures such as gate cross overs. Decisions that were once purely up to the technology developers increasingly affect design possibilities. Consequently, DTCO is used to feed the impact of such process structure support decisions on the actual designs back into the technology decision making process [7]–[10]. This discussion is carried out in Chapter 3 by separating the front end of line (FEOL), middle of line (MOL), and back end of line (BEOL) portions of the process. The latter comprise the global interconnect metallization while the former two comprise the transistors, contacts and local interconnect, respectively.

Chapter 4 discusses DTCO impact on standard cells, together with an overview of the ASAP7 standard cell library architecture. Chapter 5 describes the SRAM DTCO as well as array development and performance in the ASAP7 predictive PDK. DTCO is increasingly important, as finFET width discretization and multiple patterning (MP) constrain the possible layouts that have implications on standard cells as well as SRAM cells. Consequently, determining design rules progressed in this predictive PDK development by setting rules based on the equipment capabilities, designing cell layouts to use them, and iterating the rules based on the outcomes. Furthermore, the DTCO treatment additionally includes incorporating learning from using the cells originally derived in [11] and [12] for realistic SRAM array designs and large digital designs using automated place and route (APR) tools. Chapter 8 summarizes the work.

2 LITHOGRAPHY CONSIDERATIONS

Photolithography, hereinafter referred to simply as lithography, in a semiconductor industry context, refers to a process whereby a desired pattern is transferred to a target layer on the wafer through use of light. Interconnect metal, via, source-drain regions and gate layers in a CMOS process stack are a few examples of the patterns defined, or “printed”, using lithography.

A simplified pattern transfer flow is as follows. From among the pattern information that is stored in an electronic database file (GDSII) corresponding to all the layers of a given integrated circuit (IC) design, the enlarged pattern, or its photographic negative, corresponding to a single layer is inscribed onto a photomask or reticle. The shapes on the photomask, hereinafter referred to as the mask, define the regions that are either opaque or transparent to light. Light from a suitable source is shone on the mask through an illuminator, which modifies the effective manner of illumination, and passes through the transparent mask regions. Thereafter, light passes through a projection lens, which shrinks the enlarged pattern geometries on the mask to their intended size and exposes the photoresist that has been coated on the wafer atop the layer to be patterned. The photoresist is developed to either discard or retain its exposed regions corresponding to the pattern. This is followed by an etch that removes portions of the target layer not covered by the photoresist, which is then removed, leaving behind the intended pattern on the layer. Both lines and spaces can be patterned through this approach with some variations in the process steps.

Lithography plays a leading role in the continued scaling, which is the industry’s primary growth driver, as it determines the extent to which feature geometries can be shrunk in successive technology nodes. Lithography is one of the most expensive and complex procedures in semiconductor manufacturing, with mask manufacturing being the most expensive processing steps within lithography [13]. Both complexity and the number of masks used for manufacturing at a node affect the cost, and an increase in either of these can increase the cost to the point of becoming the limiting factor in the overall cost of the product.

As in any other manufacturing process, the various lithography steps also suffer from variability. The lithographic resolution determines the minimum feature dimension, called critical dimension (CD), for a given layer and is based on the lithography technique employed at a particular technology node. Design rules (DRs) constitute design guidelines to minimize the effects from mask manufacturability issues, the impact of variability as well as layer misalignment, and ensure printed pattern fidelity to guarantee circuit operation at good yield. Ascertaining these DRs thus requires consideration of the following lithography-related metrics that can cause final printed pattern on a layer to deviate from the intent and/or result in reliability issues.

2.1 Lithography Metrics and Other Considerations for Design Rule Determination

2.1.1 Critical Dimension Uniformity (CDU)

Critical dimension uniformity (CDU) relates to the consistency in the dimensions of a feature printed in resist. CD variations arise due to a number of factors—wafer temperature and photoresist thickness, to name a few. It is given by,

$$CDU = \frac{\sqrt{CDU_E^2+CDU_F^2+CDU_M^2}}{2}, \tag{1}$$

where the CDU_E, CDU_F, and CDU_M are the CD variation due to dose, focus, and mask variations [14]. The required CDU is typically calculated as 7% of the target CD requirement, but the modern scanner systems continue to push the envelope beyond that requirement. The 3σ CDU for 40 nm isolated and dense lines can be as small as 0.58 nm and 0.55 nm [15], respectively, for ASML's TWINSCAN NXT:1980Di optical immersion lithography scanner released in 2016. For the ASAP7 PDK, we assumed a CDU of 2 nm for optical immersion lithography, which is in line with [16]. For extreme ultraviolet lithography (EUVL) patterned layers, we assumed the CDU to be 1 nm, which is close to the 1.2 nm CDU estimated by Van Setten et al. [17] and the later CDU specification of 1.1 nm for ASML's TWINSCAN NXE:3400B EUV scanner [18].

2.1.2 Overlay

Overlay refers to the positional inaccuracy resulting from the misalignment between two subsequent mask steps and denotes the worst-case spacing between two non-self-aligned mask layers [19]. Single machine overlay (SMO) refers to the overlay arising from both layers being printed on the same machine (scanner), which results in better alignment accuracy and thus smaller overlay. However, using the same machine for two layers or masks is slower from a processing perspective, and consequently, more expensive. Matched-machine overlay (MMO) refers to that arising from two successive layers being printed on different machines, resulting in a larger value than SMO. As two separate scanners are employed, the overall processing rate in an assembly line setting is faster.

Lin predicted 3σ SMO and MMO values at N7 to be 1.5 nm and 2 nm, respectively [20]. ASML's TWINSCAN NXT:1980Di optical immersion lithography scanner released in 2016 has a 3σ SMO and MMO of 1.6 nm and 2.5 nm, respectively [21], while its Twinscan NXE:3400B EUV scanner released in 2017 has 3σ SMO and MMO of 1.4 nm and 2 nm, respectively [18]. For the PDK we assumed a 3σ MMO of 3.5 nm for optical immersion lithography, based on [22]. We assumed a 3σ MMO of 1.7 nm for the EUVL, based on the estimates by [17].

2.1.3 Mask Error Enhancement Factor (MEEF) and Edge Placement Error (EPE)

Mask error enhancement factor refers to the ratio of wafer or resist CD error to the mask CD error and is given as,

$$MEEF = \frac{\Delta CD_{wafer}}{\Delta CD_{mask}}. \tag{2}$$

Thus, it denotes the amount by which errors on the mask are magnified when they are transferred to the wafer and it depends on mask, optics, and the process. Its effects are more pronounced near the resolution limit for a specific patterning technique [23]. Features such as the metal line-ends or tips are typically more adversely affected in optical immersion lithography (193i) systems. Van Setten et al. found the MEEF for 193i patterned layers to range from five to seven, but found it to be nearly one for EUVL patterned layers [17].

The edge placement error (EPE) gives the deviation in edge placement of one layer relative to another, while accounting for both CDU and overlay contributions. For two layers, each patterned through single exposure steps, the EPE is given as,

$$EPE = \sqrt{(\frac{3\sigma CDU_{layer1}}{2})^2 + (\frac{3\sigma CDU_{layer2}}{2})^2 + (3\sigma Overlay_{layer1-2})^2}\,. \tag{3}$$

2.1.4 Time-Dependent Dielectric Breakdown (TDDB)

The primary design rule limiter for metal layers is time dependent dielectric breakdown (TDDB). At the very small fabrication dimensions, very high electric fields are generated not just in the gate dielectric, but in all isolating dielectrics between metals. A key issue in the DTCO process is determining the worst-case spacing between any two metal structures with misalignment, so that the resulting process is reliable against TDDB. TDDB occurs due to the presence of a large (although not as large as in gate dielectrics) electric field between two conductors over a long duration. Its severity is more readily pronounced in conductor layers with large overlay issues, for instance between a via and metal at disparate voltage, or in the middle-of-line (MOL) layers [24]. Obviously, sharp edges exacerbate the fields and are thus also an important issue.

Although layer self-alignment can alleviate the TDDB to some extent, it does not guarantee complete mitigation and necessitates other measures. One such case is the self-aligned raised source-drain contact to gate separation, that must be increased through the addition of extra spacer thickness and gate cap [25]. This is partially in anticipation of some erosion of the self-aligning spacer material. The final separation between two layers must therefore be not based on just overlay and CDU, i.e. EPE, but also on the TDDB requirement, which is related to the expected potential differences between the structures. For the PDK, we assumed a 9 nm spacing requirement for TDDB prevention, a value similar to that assumed by Standiford et al. [24]. Given that operating voltages are well below 1 V V_{DD}, this is conservative, which hopefully covers for any other small errors in the analysis.

2.2 Single Exposure Optical Immersion Lithography

The conventional lithography resolution limit, which determines the CD, is given by the Rayleigh equation as follows [26] by,

$$CD = k_1 \frac{\lambda}{\mathrm{NA}}, \tag{4}$$

where λ is the illumination source wavelength, NA is the projection lens numerical aperture. NA is given as

$$NA = n_1 \sin\theta, \tag{5}$$

where θ is the maximum angle of the light diffracted from transparent mask regions, that can be captured by the lens, and n_1 is the refractive index of the material between the projection lens and wafer. The value of processing factor k_1 in eq. (4) depends on the illumination method and resist process.

The term optical lithography has become nearly synonymous with the use of ArF light sources in the industry, employed since the 90 nm technology node [27]. The use of water to boost the NA leads to the technique being termed as optical immersion lithography. NA for the present 193i toolsets is 1.35 and the k_1 value is 0.28. The present set of values for these terms are a result of enhancements over the years, arising from resist process improvements and resolution enhancement techniques such as optical proximity correction (OPC), off-axis illumination (OAI), and source mask optimization (SMO) among others, each enabling a smaller CD at successive technology nodes. Ultimate limits for NA and k_1 are 1.35 and 0.25, respectively, but operating at these limits is challenging [20]. Thus, as it stands, CD or half pitch for the layers patterned using a single exposure (SE) in 193i is about 40 nm. The technical specifications for ASML's

TWINSCAN NXT:1980Di optical immersion lithography scanner suggests that it can attain a resolution of about 38 nm [21].

2.3 Multi-Patterning (MP) Approaches

The use of 193i single exposure, to achieve CD targets for all the patterned layers ended at the 22 nm technology node. This marked a severe restriction to continuing with the scaling trends for the subsequent nodes. Overcoming this limitation requires the use of multi-patterning (MP) techniques.

2.3.1 Litho-Etch^x (LE^x)

One of the most straightforward approaches to MP involves using multiple independent lithography and etch steps, where one litho-etch (LE) step refers to patterning shapes on a given layer through single exposure and etch step. The technique is termed LE^x, where x represents the number of litho-etch steps. It applies to any light source and is not specific to just 193i. LE^2, or LELE, technique is of more immediate use since it is used to pattern the target shapes just below the SE patterning limit.

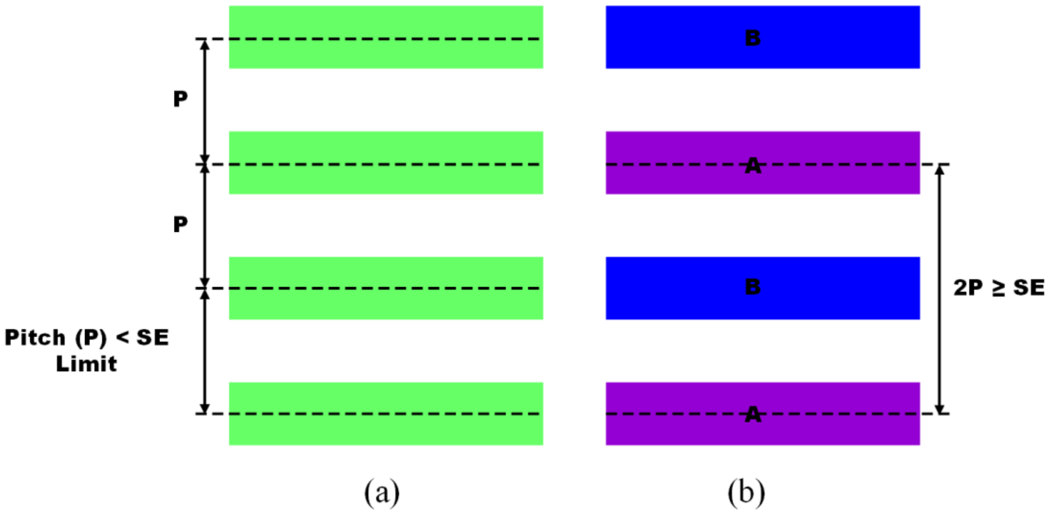


Fig. 2.1. Litho-etch-litho-etch (LELE) multi-patterning (MP) approach. (a) Target layout shapes with inter-shape pitch below the single exposure (SE) patterning limit. (b) Target layout after decomposition of shapes into separate colors. Same-color shapes are at a pitch above the 193i SE patterning limit.

To prepare a design for the LELE process, the design layout (Fig. 2.1 (a))—containing target shapes at a pitch that is smaller than the single exposure limit—is decomposed into two separate layers, e.g. A and B, with different "colors" (Fig. 2.1(b)). This decomposition or "coloring", must produce shapes assigned to a specific color layer at a pitch that can be patterned through a single exposure, thereby "splitting" the pitch. Consequently, the two color layers, resulting from the LELE decomposition step, correspond to two masks that are used in consecutive LE steps to pattern all the shapes on a single layer. This approach results in the LELE steps defining the line CD. Yet another approach involves specifying the space CD through LELE steps, instead of the line CD. While the process steps are simpler for the latter, the layout decomposition is more complex compared to the former approach.

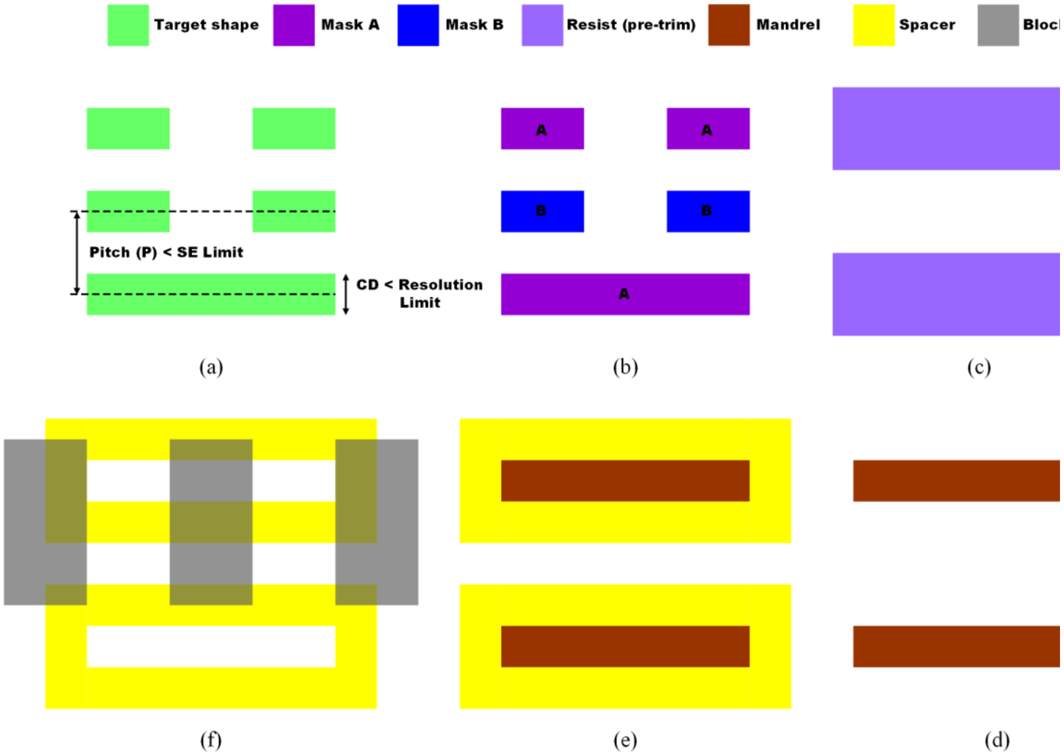
The same basic principles used for LELE can be extended to LE^x process with an *x* value that is larger than two, where *x* denotes the number of distinct colors and corresponding masks. As multiple masks are used to pattern a single layer, the cost associated with LE^x is higher than the single exposure lithography and increases with the number of masks. Unsurprisingly, the greater the number of LE steps to pattern a single layer, the higher the complexity and overlay concerns. Moreover, misalignment between steps must be considered, and jagged edges can result. LE^x is prone to odd cycle conflicts, whereby decomposition may result in a coloring conflict when more than two shapes geometries exist so as to preclude topology patterning through single exposure. Such odd-cycle conflicts will be discussed in Section 2.3.3. Stitching can alleviate these odd-cycle conflicts to some extent by patterning a contiguous shape through different exposures. It requires that the disparate fragments of the same shape have a reasonable overlap to counter edge placement error. Stitching does not completely mitigate odd-cycle conflicts in all topologies. 1-D patterns are specially challenging from this perspective.

2.3.2 Self-Aligned Multiple Patterning

Self-aligned multiple patterning (SAMP) represents another MP approach that seeks to limit the mask-defined line or space CD, thereby reducing the overlay error as compared to LE^x. The technique derives its name from spacers that are deposited along the sidewalls of a mask-defined one-dimensional (1-D) or bi-directional (2-D) line, are thus self-aligned to it. These spacers subsequently define the layer as the actual mask. Self-aligned double patterning (SADP) and self-aligned quadruple patterning (SAQP) are two of the more common forms of SAMP technique and denote whether the pitch is split by a factor of two or four, respectively. In the latter case, a first spacer is used to produce two second spacer, i.e., pitch splitting, that is used to pattern the actual lines.

SAMP can be broadly categorized into "spacer positive tone" and "spacer negative tone" process flows [13]. In the former, the spacers define the dielectric isolation or space between the lines, therefore, the process is also called "spacer-is-dielectric" (SID). It allows for multiple line and space CDs. In the latter process flow, also called "spacer-is-metal" (SIM), the spacers define the line. However, the latter only allows for two line widths and allows more variability in the intra-layer line spacing, which is a reliability risk for TDDB.

Fig. 2.2 shows a generic SADP SID flow. Similar to LE^x, decomposition is also necessary for the SADP process (Fig. 2.2(b)). In decomposition, one of these two masks (e.g. mask A) is selected as a candidate for a derivative photomask, using which the shape is patterned on the resist using a single exposure (Fig. 2.2(c)). The resist is trimmed in the event the target line CD is under the single exposure resolution limit, followed by an etch and resist strip for mandrel formation (Fig. 2.2(d)) [28]. The mandrel is a sacrificial



(a) (b) (c)

(f) (e) (d)

Fig. 2.2. A "spacer-is-dielectric" (SID) or "spacer positive tone" self-aligned double patterning (SADP) process. (a) Target layout shapes with inter-shape pitch and line CD below the single exposure (SE) patterning limit. (b) Target layout after decomposition of shapes into separate colors. (c) Resist patterning (pre-trim) with CD obtainable through single exposure resolution limit. (d) Mandrel formation (post resist trim and etch). (e) Sidewall spacer deposition. (f) Spacer and block mask defined trench formation for line patterning. (g) Mandrel cross section. (h) Mandrel and spacer cross section after spacer deposition. (i) Spacer cross section after mandrel strip. Block mask is overlaid on top of the spacer to denote the regions where metal will not be deposited, but is not indicative of the process.

feature, around which the spacers are deposited as shown in Fig. 2.2(e). Note that the result includes loops. A second photomask, called the block mask, is then used in conjunction with the spacers to “block” the regions where the feature should not be present (Fig. 2.2(f)). In the case of metal lines, the remaining regions define the trenches in a damascene process and subsequently define the line widths. SAQP process involves two spacer deposition steps, where the first set of spacers serve as mandrels for the second set of spacers. This is evident in fin examples below.

Note that the decomposition criteria for SAMP is different than LE^x, since the mandrel is continuous and any discontinuities in it can be patterned using the block mask, unlike LE^x, where the lines with such discontinuities must be patterned through a separate exposure, resulting in a potential odd-cycle conflict for certain topologies. It must be also noted that for SAMP, shapes with different colors do not correspond to separate lithography masks, as deposition of other colored shapes is through spacer-deposition and the number of masks employed is lower than the number of decomposition colors.

2.3.3 Multiple Patterning Approach Comparison

Before selecting a particular MP technique for layer patterning, consideration must be given to the variability, complexity, and the cost associated with it, and the way in which it differs from another MP technique based on these metrics. Being the simplest and most commonly used multiple patterning techniques, LELE and SADP are contrasted here to determine their favorability as the preferred MP technique for a given layer. In terms of cost, the SADP process is more expensive than LELE due to sequential etch and deposition steps [16]. Liebmann et al. put the 193i LELE and 193i SADP normalized wafer costs at 2.5× and 3×, respectively, of the 193i SE cost [10]. However, the lower variability and resulting smaller values for similar design rules that determine design density weigh in favor of SADP.

When the LELE steps define the line CD and as the two line populations are distinct, their CDU are generally uncorrelated [29]. Any overlay between the two masks affects the space CD. When LELE steps define the space CD instead, the CDU between the two space populations is uncorrelated and overlay error affects the line CD. Thus, notwithstanding its use to define either the line or space CD, the CDU is entangled with overlay in the LELE process. The edge placement error (EPE) in LELE can be calculated as,

$$EPE_{LELE} = \sqrt{(\frac{3\sigma CDU_{line1}}{2})^2 + (\frac{3\sigma CDU_{line2}}{2})^2 + (3\sigma Overlay_{line1-2})^2}\ . \quad (6)$$

for metal lines. On the other hand, in an SADP process, as the line or space CD is mostly defined through a single mask and spacer, the overlay or misalignment does not play a significant role in CD determination unless the target shape edges that determine CD are defined by the block mask. Consequently, block mask edge definition should be avoided. The edge placement error (EPE) in SADP, for spacer defined features, can thus be given as,

$$EPE_{SADP} = \sqrt{(3\sigma CDU_{litho})^2 + (3\sigma spacer_{left\_edge} + 3\sigma spacer_{right\_edge})^2}\ . \quad (7)$$

The spacer edge related terms in eq. (7) are associated with the spacer 3σ CDU, which can be as small as 1 nm [13]. The absence of overlay due to the absence of block mask defined edges, together with the small spacer CDU, gives SADP an advantage over LELE in terms of EPE control for spacer-defined edges, assuming greater than 3 nm LELE overlay and well controlled SADP spacer CDU of about 1 nm. However, if LELE overlay can be made small enough (~ 2.5 nm), then the EPE for the LELE overlay influenced edges approach that for SADP spacer-defined edges, assuming that the remaining EPE constituent terms remain unchanged [30]. As mentioned earlier, ASML's TWINSCAN NXT:1980Di 193i scanner has a 3σ MMO of 2.5 nm, but this is accompanied by a change in the CDU value, so that the advantage still lies with SADP. Finally, the spacers do not suffer from much line edge roughness (LER) [31], which is another advantage that SADP offers over LELE. From an electrical perspective, the larger lithography-related variations for LELE patterned layers translate into larger RC variations, which can be nearly twice as much as that for SADP patterned layers employed for critical net routing, primarily due to capacitance variations [32]. Thus, a choice of LELE creates a potential plethora of metal corners for designers to deal with.

The impact of topology-imposed design rule constraints must also be considered when selecting a MP technique, so as to limit any density penalty. Double patterning can typically overcome the large tip-to-tip (T2T) or tip-to-side (T2S) spacing requirements inherent to single exposure lithography by avoiding these features from being assigned to the same color. The SE T2T or T2S spacing values are larger than the minimum SE-defined width to spatially accommodate the hammerheads used for optical proximity correction (OPC) applied to ensure pattern fidelity [33]. Large spacing also prevents tips from shortening due to bridging by ensuring sufficient contrast, as the regions between tips have low image contrast. However, even with double patterning, certain feature topologies, such as gridded metal routes with discontinuities, can result in T2T or T2S features on the same mask [13]. This occurs because shapes on adjacent routing tracks must be colored alternatingly, which forces the shapes along the same track to be the same color.

For LELE patterned layer as illustrated in Fig. 2.3(a). This necessitates the use of DR values related to SE T2T spacing, *x*, that are even larger than the minimum SE-defined width and nullify the advantage of double exposure.

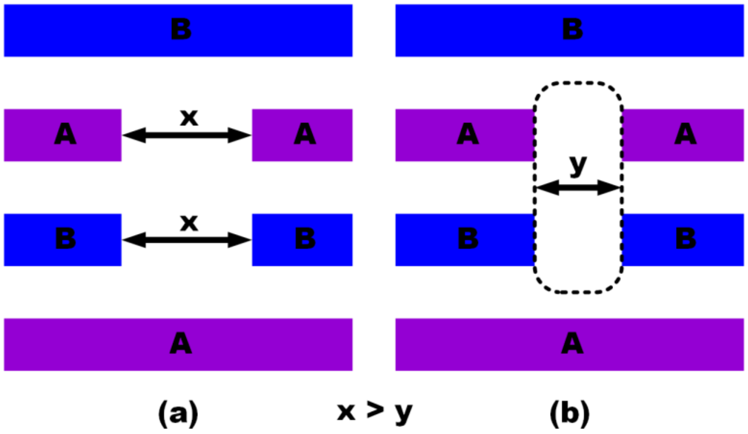
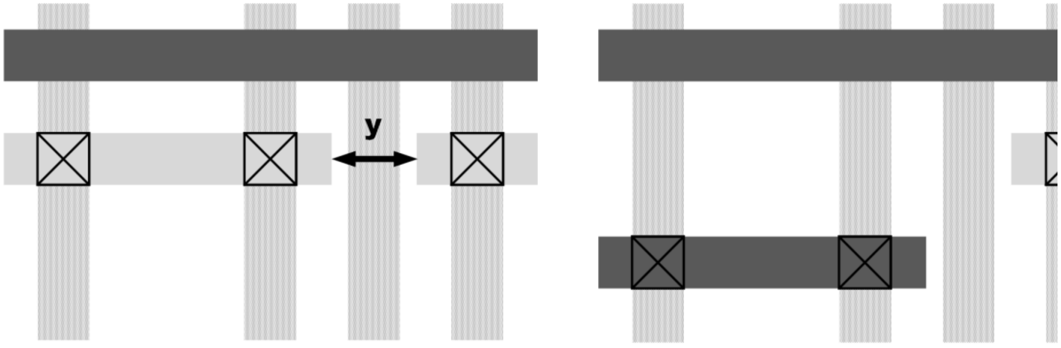


Fig. 2.3. Comparison of tip-to-tip (T2T) spacing cases for different double patterning approaches. (a) Scenario where T2T spacing, x, is single exposure limited for LELE patterned layers are not too uncommon for designs with 1D, gridded routing. (b) T2T spacing, y, for SADP patterned layer is defined by the block or cut mask (dotted polygon). Being a line-like feature, the block/cut mask width is similar to the single exposure-defined line width, which is smaller than the spacing required for single exposure-defined tips (after [34]*).*

By comparison, the T2T and T2S features in SADP are block/cut mask defined (see Fig. 2.3(b)). Although also SE-defined, the block/cut mask width is similar to the SE-defined line width, *y*, instead of the larger T2T or T2S SE spacing requirement.

As identified by Ma, et al., this enables SADP patterned metal routes on the same track to connect to pins on a lower metal interdigitating another lower metal route as shown in



(a) (b)

Fig. 2.4. Pin connection scenarios for double patterned layers. (a) Connections to next neighboring pins can be made through metals on the same track by leveraging the smaller block mask defined T2T spacing for SADP patterned metal layer. (b) The same becomes impossible with LELE patterned metal layer due to larger T2T spacing requirements for same color shapes and different tracks must be utilized, which results in density penalty.

Fig. 2.4(a), while different tracks must be used as in Fig. 2.4(b) to realize the same connections for LELE patterned metal routes [32]. Thus, SADP can lower the density penalty in certain design scenarios.

Furthermore, as shown in Fig. 2.5(a), LELE can produce odd cycle conflicts during mask decomposition into two colors. Such a conflict requires increasing the spacing between the features to the SE T2T value (*x*) in order to resolve the conflict. In contrast, SADP is largely free of odd cycle conflicts for 1-D topologies with equal metal width along routing tracks, as evident in Fig. 2.5(b).

However, odd cycle conflicts may arise during mask decomposition into two colors for SADP, when unequal metal widths or 2-D features are used as shown in Fig. 2.5(c). Thus, the simpler SADP is adopted for BEOL in the ASAP7 PDK. Some simple DRs, such as limiting line widths to specific values and pitches, makes automatic decomposition possible. This is handled by the Calibre DRC flows automatically, to simplify usage.

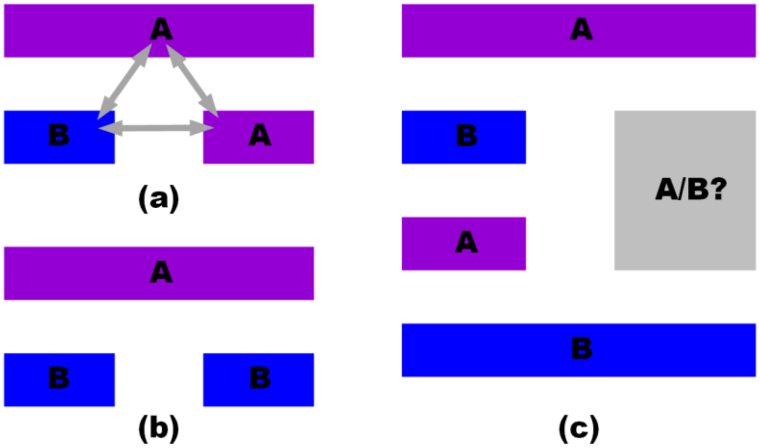


Fig. 2.5. (a) Even the most common topologies suffer from LELE odd cycle conflicts. (b) SADP is largely free of such conflicts, given correct coloring and considering 1-D equal width metals. (c) Conflicts with SADP may arise for wide metals used in power or clock routing (after [34]*).*

2.4 Extreme Ultra-Violet Lithography (EUVL)

2.4.1 EUVL Necessity

Keeping in line with the area scaling trends requires manufacturing capability enhancement through technological advancement, with pitch scaling through photolithographic improvements constituting the major effort. This increases the wafer cost initially by nearly 20-25% at a new technology node [35]. As a particular technology node matures, the subsequent process and yield optimization bring down the wafer cost. These improvements, together with the increased transistor density, eventually result in the overall cost reduction per transistor. 193i MP adoption to pattern increasingly small

FEOL and BEOL pitches has led to a larger than conventional wafer cost increment. This is due to an increased number of photomasks and process steps required to pattern a given layer with MP—hitherto patterned through a single LE step. MOL layer and finFET introduction has led to further cost escalation as they constitute critical layers that require MP. The transistor cost reduction trend has slowed down over the previous few nodes, to which MP has contributed to some extent. This contribution will become even larger with an increase in layers patterned through MP techniques with each new node, resulting in a further slow-down in the per transistor cost-reduction trend. Thus, MP can potentially undermine the cost-effectiveness associated with transitioning to a new node.

In addition to cost, MP is also challenging from a variability perspective, which leads to stringent CDU and overlay considerations, as discussed in Section 2.3. Accommodating these requires guard-banding the projected pitch/spacing targets for relaxed process margins. Patterning a critical BEOL layer through MP is one such case. Patterning a 2-D shape using LE^x may necessitate stitching, which complicates overlay requirements and requires that the pitch target be relaxed to ensure sufficient shape overlap. An alternative is using SADP patterning, which is less amenable to 2-D shapes, and a 1-D patterning approach using SADP becomes the other choice [36]. This detrimentally affects the cell circuit density, but permits better manufacturability [37]. Mallik et al. estimate a 16% and 5-15% lower area penalty for 2-D, instead of 1-D, metal layers for SRAM cell and standard cells, respectively [38]. A pure 1-D approach makes layout of even relatively simple logic gates more difficult and results in poor input pin accessibility.

Extreme ultra-violet (EUV) lithography can mitigate some of the MP related issues. It uses a 13.5 nm light wavelength, instead of 193 nm used for ArF immersion lithography. This enables patterning the features at a much smaller pitch and resolution, so that a single EUV exposure suffices for patterning at the target pitch only attainable through multiple exposure with 193i, thus, greatly simplifying the design rules. Since academic use was a primary goal of the PDK, opting for simpler EUV rules was a key consideration, even at the risk of optimism in EUV availability. However, we are presently working on additions and libraries to support pure 1-D metallization scenario.

2.4.2 EUVL Description and Challenges

ASML's EUVL scanners have a 0.33 NA and can operate with a processing factor k_1 of nearly 0.4 [39][40]. Using these values, together with a 13.5 nm wavelength, in equation (4) gives a CD of approximately 16 nm. Improvements have led the CD to be reduced to 13 nm [18] [40]. An NA improvement to 0.5 can lead to a further CD reduction to 8 nm [41].

EUVL differs from 193i in a number of ways that make it a more challenging patterning approach and have contributed to the delay in EUVL being production-ready. Light at EUV wavelength is generated as follows. A droplet generator releases tin droplets, which are irradiated by a laser to create plasma containing highly ionized tin that emits light at 13.5 nm wavelength that is gathered by a collector for further transmission [42]. Light in the EUV spectrum is absorbed in the air, which necessitates manufacturing under vacuum. It is also absorbed by nearly all materials, which precludes optical lens use to prevent excessive energy loss. Instead, reflective optics, i.e. mirrors, are used. These mirrors have a reflectivity of around 70% and an EUV system can contain over ten such mirrors, resulting in only around 2% of the optical transmission to reach the wafer [42]. Inefficiency in the power source reduces the transmitted optical power even further, thus creating a demand for a high-power light source. It is also desirable for the photoresist to have a high sensitivity to EUV light.

Production throughput for a photolithography system, given in wafers per hour (WPH), is closely associated with cost-effectiveness. To a considerable extent, it relies on the optical power transmitted to the wafer, and thus on the EUV light source power. It also depends on the amount of time the system is available for production, i.e. system availability. System downtime adversely affects the cost. For EUVL patterning to be cost-effective, a source power exceeding 250 W is desired for over 100 WPH throughput at a 15 mJ/cm^2 photoresist sensitivity [35]. Currently, the source power is around 205 W and droplet generator, hitherto a major factor in EUVL system unavailability, has become a smaller concern in more modern EUVL systems [43]. These and other improvements have brought EUVL system close to the HVM production goals by increasing the throughput to 125 WPH [40]. Collector lifetime is the biggest contributor to the system unavailability at the moment and a number of other issues must be surmounted for further cost-effectiveness [43]. Overall, EUVL systems continue to improve and are being deployed by foundries for production at N7 and N5 [44][45].

2.4.3 EUVL Advantages

Both cost and complexity concerns become the deciding factors in choosing single exposure EUV over 193i MP approaches. The EUV mask cost alone is approximately 1.5× that of a 193i mask [38]. Other operational expenses bring up the EUVL cost to nearly 3× of 193i single exposure [10]. The number of masks used in each technology node have increased almost linearly up until N10, but the continuation of MP use will result in an abrupt departure from this trend at N7 [46]. The issue is compounded by an increase in the associated process steps, which further adds to the cost. Liebmann et al. estimate the normalized LE^2, LE^3, SADP, and SAQP cost to be 2.5×, 3.5×, 3×, and 4.5× that of 193i SE, respectively. Dicker et al. estimate a 50% patterning cost reduction with EUV as opposed to SAQP [46]. They also estimate a faster time to yield and time to market with EUVL due to cycle time reduction as compared to 193i MP approaches that suffer from large learning cycle time—as large as 30% compared to 2-D EUV, thus improving EUVL cost-effectiveness.

EUVL single exposure also reduces the process complexity by virtue of reduced overlay. The small EUV wavelength allows the processing factor k_1 to be relatively large—in the range 0.4-0.5. This enables EUVL to have a high contrast, given by normalized aerial image slope (NILS), than 193i and allows features to be printed with higher fidelity [40][44]. The high feature fidelity, better corner rounding, and a single exposure use with EUVL cause fewer line and space CD variations. Consequently, metals and vias patterned through EUVL have more uniform sheet resistance [44]. They also have lower capacitance as compared to SADP patterned shapes, as EUV single exposure obviates dummy fills and metals cuts. These improvements contribute to improved scalability and better performance [43].

Mallik et al. estimate the normalized wafer cost to increase by 32% N10 and by a further 14% at N7 without EUV insertion at these nodes [35]. They also estimate a 27% cost reduction, as compared to the latter case, due to EUVL use at N7 for critical BEOL layer patterning with a 150 WPH throughput as a best-case scenario. Ha et al. put the number of mask reduction at N7 due to EUVL use at 25% [44]. Dicker et al. estimate over 40% cost per function reduction in moving from N10 to N7, and further to N5 as a consequence of EUV insertion for critical BEOL layers [46]. Thus, EUVL deployment at these nodes will likely help ensuring the economic viability of process node transitions.

2.5 Patterning Cliffs

Patterning cliffs mark the pitch limits for a given lithography technique or MP approach. Table 2.1 summarizes these pitch limits for the metal layers [41][47]. It must be noted that EUV scanners are being continuously refined and their capabilities may vary, resulting in different the final patterning cliffs. However, Table 2.1 gives good rule of thumb values.

Patterning Technique	Minimum Pitch (nm)
193i	80
193i LELE	64
193i LELELE	45
193i SADP	40
EUV SE (2-D, NA=0.33)	36
EUV SE (1-D, NA=0.33)	26
EUV SE (2-D, NA=0.55)	22
193i SAQP	20
EUV SE (1-D, NA=0.55)	16

Table 2.1 Minimum feature pitch for a particular patterning technique.

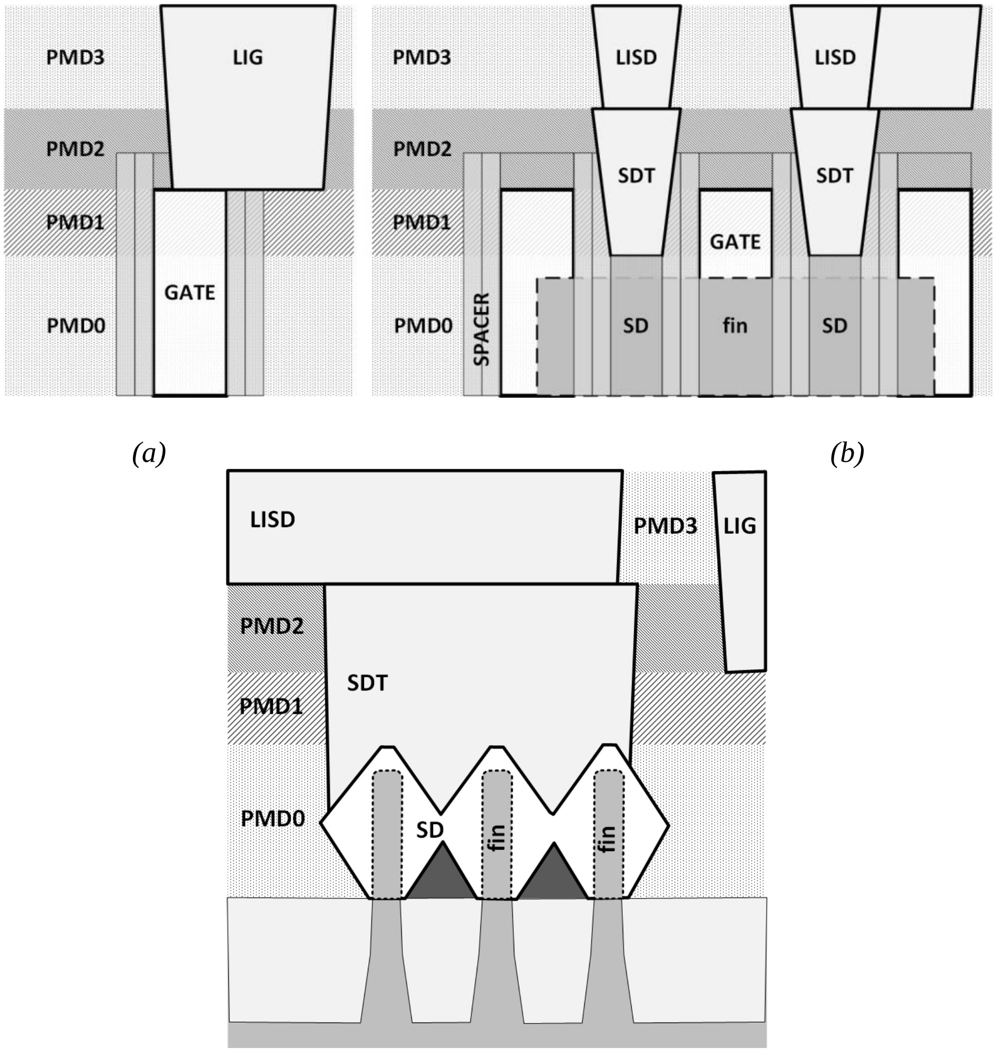
3 PATTERNING ASSUMPTIONS AND DESIGN RULE FORMULATION

3.1 Front End of Line (FEOL) and Middle of Line (MOL) Layers

For ASAP7, the transistors are assumed to be fabricated using a standard finFET type process: a high-K metal gate replaces an initial polysilicon gate, allowing different work functions for NMOS and PMOS, as well as different threshold voltages (V_{th}) [2][16][48][49]. Fins are assumed to be patterned at a 27 nm pitch and have a 7 nm drawn (6.5 nm actual) thickness. The layer active is drawn so as to be analogous to the diffusion in a conventional process and encloses the fins—over which a raised source-drain is grown—by 10 nm on either side along the direction perpendicular to the fin run length [11]. The drawn active layer differs from the actual active layer, which is derived by extending it halfway underneath the gates—perpendicular to the fins. The actual active layer, therefore, corresponds to the fin 'keep' mask, with its horizontal extent marking the place where fins are cut and its vertical extent denoting the raised source-drain regions.

Gates are uniformly spaced on a grid with a relatively conservative 54 nm contacted gate pitch (CGP). Gates are 20 nm wide (21 nm actual). Spacer formation follows poly gate deposition [50]. Cutting gate polysilicon with the gate cut mask, in a manner that keeps the spacers intact with a dielectric deposition following, ensures that fin cuts are buried under gates or the gate cut fill dielectric, so source/drain growth is on full fins. A double diffusion break (DDB) is assumed to be required to keep fin cuts under the gate. Recently announced processes have removed the DDB requirement, improving standard cell density [51]. A 20 nm gate cap layer thickness is assumed. This thickness provides adequate distance to avoid TDDB after self-aligned contact etch sidewall spacer erosion, accounting for gate metal thickness non-uniformity [25]. The spacer width is 9 nm.

The resulting FEOL and MOL process cross section comprises Fig. 3.1. Fig. 3.1(a) shows the (trapezoidal) source/drains grown on the fins. The MOL layers can be used for



(a) *(b)*

(c)

Fig. 3.1. FEOL and MOL cross sections. (a) LIG connection to the gate. (b) LISD to SDT and SDT to source-drain (SD) connection. LISD location in the stack allows it to cross over gates, so as to be used for routing. (c) Fin and SD cross section. LIG is shown here to illustrate its necessary separation from LISD and SDT. Sub-fin and shallow trench isolation (STI) are evident underneath PMD0 (after [11]*).*

functions typically reserved for the first interconnect metal (M1) layer [2] and serve to lower M1 routing congestion, thereby improving standard cell pin accessibility [52]. The connection to the MOL LISD layer is through the source drain trench (SDT) contact layer. The minimum SDT vertical width of 17 nm is required in an SRAM cell, which necessitates patterning using extreme ultraviolet lithography (EUVL). It has a 24 nm drawn horizontal width, with the actual width being 25 nm. This width is larger than the 15 nm gap between the spacers, so as to ensure complete gap coverage and contact with RSD, despite the 5 nm 3σ edge placement error for EUV.

LISD provides the means of connecting RSD regions to power rails and other equipotential RSDs within a standard cell. It is drawn at the same horizontal width as SDT for lower resistance when connecting to RSD through SDT. LISD may also be used for routing purpose within a standard cell at 18 nm width and 36 nm pitch, as it can pass over gates—further lowering M1 usage. The layer thicknesses are defined by different dielectric layers, so that appropriate etch stops can be used. This implies bidirectional, i.e. 2-D, LISD routing, which, when combined with width and pitch assumptions, means that EUVL must be used for patterning. LISD connects to M1 through via 0 (V0), which is another MOL layer.

The local interconnect gate (LIG) layer is used for connecting gates to M1 through V0 and for power delivery to standard cells by connection to LISD where they intersect. The minimum LIG width of 16 nm is dictated by the LIG power rail spacing from the gate. The width value implies that LIG is also patterned using EUVL. Recent advances in EUVL have demonstrated resolutions lower than 18 nm [41][53] even for 2-D patterns, but the PDK restricts this value as the minimum line width for 2-D layers, in accordance with a more conservative 36 nm pitch for bidirectional EUVL [38]. However, smaller line width is permitted for LIG as its usage are limited to unidirectional, i.e. 1-D, patterns. LIG connects to the gate through the cap layer in Fig. 3.1(b), which cuts through a standard cell between the NMOS and PMOS devices. Fig. 3.1(c) shows the same view, but at the fins, so the source/drains are illustrated perpendicular to that in Fig. 3.1(a). As mentioned, the fin cuts occur under a dummy gate, which comprises ½ of a DDB. The SDT is self-aligned to the gate spacer as shown. MOL rules turn out to be limiting for both standard cells and SRAM. The details are described below. More details and the transistor electrical behavior is presented in [11].

3.2 Back End of Line (BEOL) Layers

The ASAP7 PDK assumes nine interconnect metal layers (M1-M9) for routing purpose and corresponding vias (V1-V8) to connect these metals.

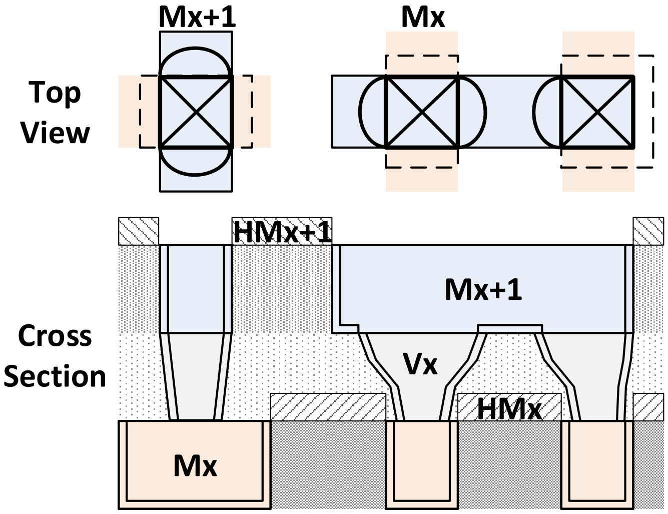


Fig. 3.2. Representative BEOL cross section. The dotted lines represent the actual self-aligned via (SAV) masks top view. Vias with (left two) and without end-cap (rightmost) are shown. Arcs along Mx+1 length denote via widening at the non-hard mask edges, evident in the cross section for the two vias at the right (after [11]*).*

Fig. 3.2 shows a representative BEOL stack cross section, comprising of a lower metal layer (Mx), via (Vx), and an upper metal layer (Mx+1). Following the industry trend [2], all BEOL layers assume copper (Cu) interconnects. Metal and vias have a 2:1 aspect ratio, in line with the ITRS roadmap [54]. Table 3.1 enumerates thickness of the interconnect layers. Fig. 3.2 (bottom) also shows the barrier layers that increasingly consume the damascene trench, increasing resistance.

In the ASAP7 PDK, metals have the same thickness as the corresponding inter-layer dielectric (ILD), but the vias are thicker than the ILD by 10%—an amount corresponding to the assumed HM thickness. Actual processes at N7 include as many or more than 14 metals, with more metal layers at each thickness and pitch value, culminating in 2 layers that are much thicker and wider at the top than in our PDK. The layers here are representative of all but the very thick top layers, which are primarily for power distribution. We had initially not foreseen PDK use for large die power analysis, but are considering adding layers to make the PDK more amenable to full die power analysis.

Metal/Via Layer	Thickness (nm)
M1-M3	36
V1-V3	39.6
M4-M5	48
V4-V5	52.8
M6-M7	64
V6-V7	70.4
M8-M9	80
V8	88

Table 3.1 BEOL layer thickness and metal pitches.

3.3 SAV and Barrier Layer

In a typical via-first flow, the ILD corresponding to a via sustains damage and erosion—caused by dry etch and cleaning steps, respectively, during both via and metal patterning. The metal width itself is also affected, since via formation goes through the metal damascene trench. This results in via widening that can cause shorts to adjacent, non-equipotential metal lines [55][56]. This issue is exacerbated by small metal pitches at lower technology nodes. Consequently, in the ASAP7 design rules, a self-aligned via (SAV) formation flow as described by Brain et al., is assumed, whereby vias are patterned after the upper interconnect metal layer is patterned on a hard mask (HM) that is relatively unaffected by the via etch [56]. The HM greatly limits via widening perpendicular to the its edges, although some widening occurs along the metal as evident in Fig. 3.2 (top). The resulting via edges are delineated by the upper metal HM, i.e, via self-alignment. Via mask edges tend to extend outward where they are defined by the via rather than hard mask. Nonetheless, the via is perfectly aligned perpendicular to the upper metal direction despite via and upper metal overlay errors. The dashed lines in the Fig. 3.2 (top) show that the actual via mask overlaps so that the HM defines the width even with misalignment. However, for simplicity the vias are drawn conventionally in the PDK. They are sized as part of the DRC flows.

Barrier materials, such as tantalum nitride (TaN), are required at the Cu and inter-layer dielectric (ILD) interface to prevent Cu diffusion. Thickness of the barrier—composed of more resistive TaN—does not scale commensurately with the interconnect scaling. This, together with the diffuse electron scattering at interfaces, causes a greater increase in line resistance than is expected as a consequence of scaling [57]. Additionally, the presence of

TaN at via interface has the undesirable effect of increased resistance. This contribution is incorporated for resistance value calculation during process cross section specification in the Mentor Calibre MIPT file that is used for generation of rule decks employed by Calibre PEX/xACT 3D parasitic extraction tools [49]. The metal resistivity, as specified for extraction purpose, is calculated based on [58],

$$\rho = \rho_0 \frac{3}{8} C(1-p) \left(\frac{1}{h}+\frac{h}{A}\right) \lambda + \rho_0 \left[1-\frac{3}{2}\alpha+3\alpha^2-3\alpha^3 \ln\left(1+\frac{1}{\alpha}\right)\right]^{-1}, \tag{8}$$

where the ρ_0 denotes the bulk resistivity and C is a geometry based constant. p is the electron collision specularity with surfaces, λ is the bulk electron mean free path, h is the line height, and A is the cross section area. α is given as $\lambda R/[G(1-R)]$, where R is the electron reflection coefficient at the grain boundaries and G is the average grain size. The first and second terms are the resistivity due to surface electron scattering and grain boundary scattering, respectively. The latter dominates for the 7 nm node [58].

3.4 EUV Lithography Assumptions and Design Rules

In addition to some of the FEOL, i.e., fin cut, and MOL layers, in the PDK EUVL is also assumed for patterning M1-M3 and vias corresponding to these metals, i.e., V1 through V3. The choice of an EUVL assumption and the accompanying M1-M3 pitch of 36 nm is based on the premise that this pitch may be attained using single EUV exposure [Mallik14]. Meeting the same target using optical immersion lithography requires the use of multiple patterning (MP) techniques, such as SAQP with LE or LELE block mask, which also pushes towards all 1-D topologies. While EUVL is costlier than optical immersion lithography when considering a single exposure, the use of multiple masks in SAQP with block means that the MP approach becomes nearly as expensive as EUVL due to expensive mask tooling and associated processing steps [10]. It is noteworthy that several ITRS target pitch values at N7 are at or over the EUV single exposure cliff. As a result, beyond N7, MP will be required even for EUV lithography defined layers. With N7 at the cutoff, we felt EUVL would be appropriate.

As EUVL permits the single mask use, employing it simplifies the design process by circumventing the issues related to MP, such as complicated cell pin optimization issues due to SADP and odd-cycle conflicts as a consequence of LELE use [32]. As mentioned earlier, EUVL use at a slightly relaxed pitch of 36 nm also permits 2-D routing, which has the effect of further simplifying both standard cell and SRAM cell design. Thus, given similar cost and reduced design complexity, the PDK assumed EUVL over 193i MP schemes for a number of layers that complicate standard cell design, viz. SDT, LISD, LIG, V0, and M1. However, subsequent iterations of the PDK may revise this choice, so as to use SAQP for 1-D M1-M3, as a consequence of delays in EUV readiness for high volume manufacturing. While the aggressive pitch value of 36 nm that EUV affords are not required for M2 and M3, we considered these layers to be EUVL patterned due to the ease that this assumption lends when routing to standard cells during auto-place-and-route (APR). Having the vertical M3 match the M2 also allows better routing density vs. a choice of say, the gate pitch for vertical M3. This also allows relative flexibility in metal directions. We foresee vertical M2 as a better choice for 1-D cells.

A 3σ edge placement error of 5 nm for two EUV layers, when determining inter-layer DRs, was calculated assuming a 3σ mixed machine overlay (MMO) of 1.7 nm for the EUV scanner, a 3σ error in placement of 2 nm due to process variations and through additional guard-banding [17]. Patterned M1-M3 lines have a 36 nm 2-D pitch and minimum line width of 18 nm is enforced by the DRs. T2T spacing for narrow lines is 31 nm following Van Setten et al. [39], while wider lines can have a smaller tip-to-tip spacing at 27 nm [59]. As per the PDK DRs, lines narrower than 24 nm are considered thin lines and those wider than this value are considered as wide lines. This threshold value was determined based on the minimum LISD width, since LISD routes near the power rails become the limiting cases for T2T spacing. A moderate T2S spacing of 25 nm follows the results demonstrated by Van Setten et al. [17]. Corner-to-corner EUV metal spacing of 20 nm enables via placement to metals on parallel tracks at the minimum possible via spacing of 26 nm with 5 nm EUV upper metal end-cap to allow full enclosure.

3.5 MP Optical Immersion Lithography Assumptions and Design Rules

Metal interconnect layers above the intermediate metal layers, i.e. M3, are assumed to be patterned using 193 nm optical immersion lithography and MP. The metal pitches are the same as the thickness values described in Table 3.1 (the 2:1 aspect ratio). The pitch values for M4-M5, and M6-M7 correspond to the targets defined by Liebmann et al. for 1.5× and 2× metal, respectively [10]. The same metal pitch ratios do not apply to our PDK since the 1× metal pitch is 36 nm to ensure 2-D routing instead of the roadmap value of 32 nm. Moreover, recent foundry releases seem to be tending to the more conservative 36 nm pitch as well. Incrementing the pitch in multiples other than 0.5× does not have any design implication as proven by our DTCO APR experiments. The metal and via aspect ratios are in line with ITRS roadmap projections.

3.5.1 Patterning Choice

The M4-M5 pitch target of 48 nm can be attained using either SADP or LE^3, and that of 64 nm for M6-M7 using either SADP or LELE. LE^3 was dismissed outright for M4-M5 patterning, since it is costlier than SADP [10], and even though the lower LELE cost is appealing for M6-M7 patterning, SADP has lower EPE, LER and hence RC variations as mentioned in Section 2.3.3. Therefore, we chose SADP over LELE for patterning the layers M4-M6. Furthermore, Ma et al. also found that LELE fails TDDB reliability tests at 64 nm pitch with 6 nm overlay while SADP appears reliable [13]. Although we used a 3.5 nm overlay for 193i patterned layers that would allay the TDDB severity, we still assumed SADP patterning for M6-M7 so as to be more cautious, given the TDDB concern in addition to the other aforementioned SADP advantages. We chose SID over SIM as it permits multiple metal widths and spaces, which is beneficial for clock and most importantly, power routing. Patterning 2-D shapes is possible in SADP but presents challenges, as certain topologies, such as the odd-pitch U and Z constructs, may either not be patterned altogether or contain block mask defined metal edges that result in overlay issues and adversely affect metal CD [32]. Consequently, we chose to restrict M4 through M7 1-D (straight line) routing in the DRs.

The ASAP7 PDK supports design rule checks (DRCs) based on actual mask decomposition into two different masks (colors) for the purpose of SADP. The mask decomposition is performed as part of the rule decks using Mentor Graphics Calibre multi-patterning (MP) solution. An automated decomposition methodology [60] is employed and does not require coloring by the designer, which greatly simplifies the design effort. As mentioned, we consider this key in an academic environment. To the best of our knowledge, this is the first educational PDK that offers design rules based on such a decomposition flow for SADP. In contrast, the FreePDK15 also used multi-colored DRs [3], but required decomposition by a designer by employing different colored metals for the same layer.

3.5.2 SADP Design Rule Formulation

DRs must ensure that shapes patterned using the two photomasks, viz. the block and the mandrel, can be resolved. This entails writing DRCs in terms of the derived photomasks and perhaps even showing these masks to the designer. However, fixing DRCs by looking at these masks, and the spacer, is both non-intuitive and confusing. In the flows here, the colors are generated automatically, and the flows can produce the block and mandrel masks created by metal layer decomposition. We formulated restrictive design rules (RDRs) that ensure correct-by-construction metal topologies and guarantee resolvable shapes created using mandrel and block mask referring only to the metals as drawn, so that the rules are color agnostic. Nonetheless, to validate the DRCs, the mandrel, block mask, and spacer shapes are completely derived in our separate validation rule decks [34]. All DRCs were written using Mentor standard verification rule format (SVRF) and CalibreDRV was used to execute the DRCs.

3.5.2.1 Basic and Grid-Based SADP Mask Decomposition Criteria

The decomposition criteria for assigning shapes to separate A or B masks/colors for further SADP masks and spacer derivation is as follows. The basic criterion is assignment of two shapes at a pitch below the 193i supportable 80 nm pitch to different masks. SADP uses a continuous mandrel (where possible) throughout the design extent to define the non-mandrel metal edges by spacers deposited around the mandrel. If the basic decomposition criterion is used solely, with spacing checked only between the shape edges sharing a common parallel run length, then isolated shapes can result in color assignments that incorrectly produce mandrels on adjacent routing tracks (as shown in Fig. 3.3(a)). Thus, routing grid-based decomposition is used as well. Single CD, i.e. minimum width (1×), metals are decomposed based on a grid in the PDK, as they are the most numerous. Two grids (A and B) are created and single CD metals are assigned to mask A or B depending on the grid they lie on, which leads to correct decomposition (see Fig. 3.3(b)).

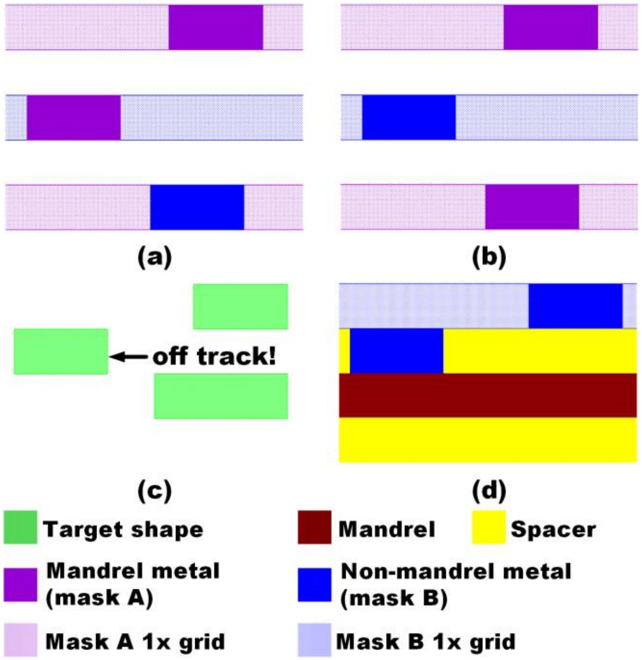


Fig. 3.3. (a) Insufficient decomposition criterion results in conflicts, (b) that we limit by assigning 1× metals on alternating tracks to different masks. (c) Example target off-track metal shapes that (d) are not patterned correctly. Figures generated by executing the PDK DRCs in CalibreDRV (after [34]).

Any off-track single CD metal routes, such as the one in Fig. 3.3(c), are flagged to enforce APR compatibility and correct single CD/1× grid-based decomposition. The PDK allows offsetting single CD A and B grids from the origin if required. The same grids must be enforced for DRCs as are used in the APR flow.

3.5.2.2 Decomposition Criteria to Support Wide Metals Patterned Using SADP

Despite the two aforementioned criteria, wide metals—typically used in power routing from the top level and custom blocks—that result in a pitch below 80 nm, cause incorrect

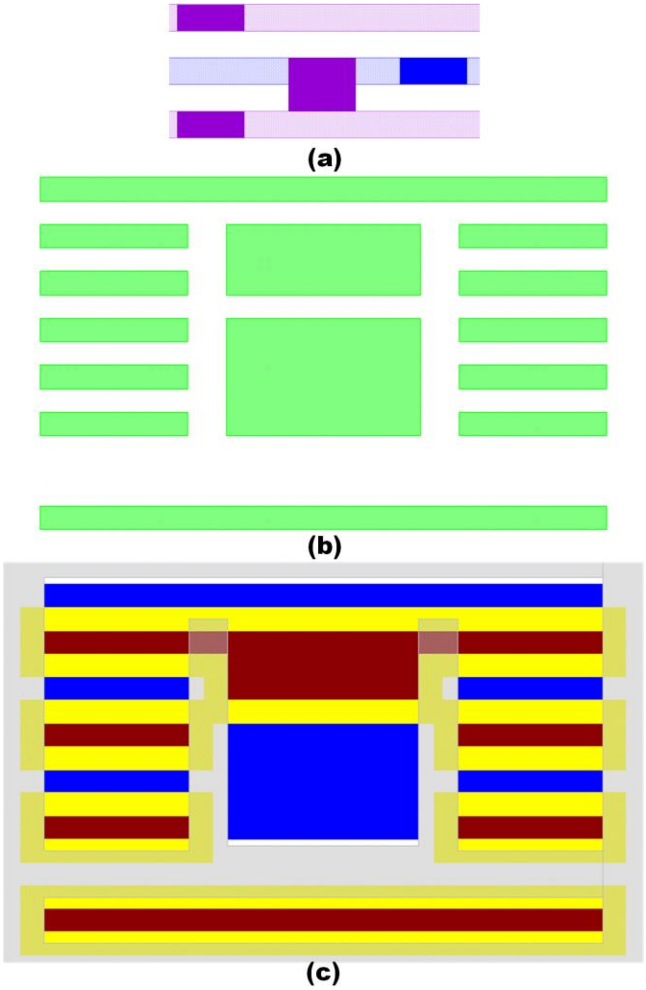


Fig. 3.4. (a) Incorrect decomposition from 1x metal interaction with wide metal. (b) Target shapes with wide and 1× metals. (c) SADP layers with block mask in grey. Figures generated by executing the PDK DRCs in CalibreDRV. (after [34]).

decomposition (Fig. 3.4(a)). Such cases are essentially odd cycle conflicts if viewed from a perspective that treats the single CD metal grid as a layer with regards to which neighboring wide metals must be decomposed. Although Calibre can detect odd cycle conflicts, it does so for drawn and not derived layers. Thus, another decomposition criterion assigns single CD metal grids and neighboring wide metals to different colors if they are at a separation resulting in a pitch less than 80 nm. Supplementary DRCs flag any single CD metal grids that touch, but do not intersect, the wide metals along the routing direction. Moreover, they also stipulate that the minimum spacing between same mask wide metal and single CD grid be equal to at least three spacer widths, to maintain the single CD grid.

Fig. 3.4(c) shows the successful decomposition of target shapes in Fig. 3.4(b) with both single CD and wide metals. Metal widths that are even integer multiples of minimum width are disallowed, to prevent decomposition resulting in mandrel spacing below 193i resolution limit (Fig. 3.4(a)) or a mandrel touching a non-mandrel shape. Odd width metals can be represented by $w = 2n+1$ ($n \in Z$). An odd value of n results in metal constructs that suffer from odd-cycle conflicts if the wide metal is flanked by single CD metals along either adjacent routing tracks (Fig. 2.5(c)). Thus, although the 3 CD metal in Fig. 3.4(b) would have resulted in a conflict, the unique topology allows correct decomposition due to the vacant track above the bottommost track and since the 5 CD metal is not pre-assigned to a color. In Fig. 3.4(c) the topmost non-mandrel single CD metal and the lower non-mandrel 5 CD metal (both in blue) are not adjoined by a mandrel forming metal on one side. The block mask (grey) defines metal edges. Additionally, the block mask can be sized so that it does not abut the drawn target edge. Note the white space between the block and non-mandrel metals in Fig. 3.4(c), so that it only touches the drawn edge when misaligned by the worst-case block mask edge placement error (EPE). This prevents high resistance narrow wires from being formed due to mask misalignment.

3.5.2.3 Restrictive SADP Design Rules

While varying metal spacing can be produced using SID, such cases are prevented by employing restrictive DRs to lower the occurrences of block mask defined edges, as mentioned. Fig. 3.5(a) shows a case where a non-minimum width space separates a wide metal (to the right) from single CD metals. Since the spacer deposition width is uniform, this results in the metal edges adjoining the non-minimum space being defined by the block mask. Thus, the PDK assumes equal spacing for SADP metals (24 nm for M4 and M5, and 32 nm for M6 and M7). Fig. 3.5(b) shows scenarios dictating minimum T2T rules. The minimum T2T same mask spacing is 40 nm, which is the ½ pitch of the block mask.

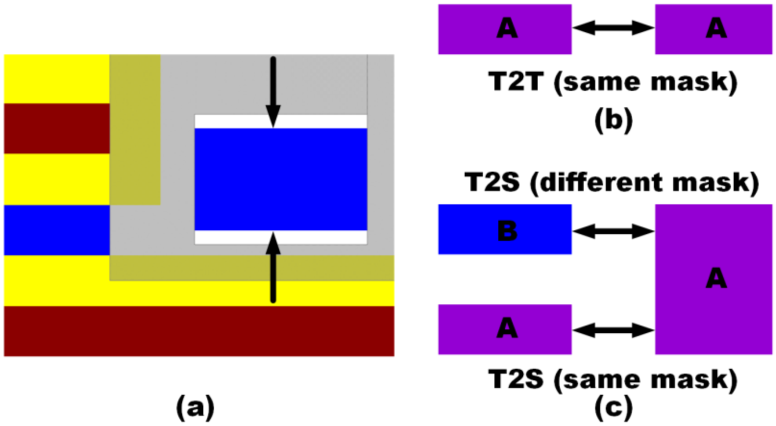


Fig. 3.5. (a) Target shapes with non-equal metal spacing. Arrows indicate block mask defined edge. Figure generated by executing the PDK DRCs in CalibreDRV. (b) Same mask T2T spacing rule. Different 1x metals along same tracks are prohibited. (c) Same and different mask T2S rules. Equal rule values facilitate rectangular block mask creation (after [34]).

Minimum T2T different mask spacing is not required, since single CD metal along a routing track have the same color and those wider than or equal to 2 CD metals are not considered as tip features. Non-integer multiple metal widths are flagged as off-grid. The minimum T2S same mask spacing is 40 nm—again, block mask-defined. Although the minimum T2S different mask spacing can be 24 nm (i.e., spacer-defined), a 40 nm spacing is used, as a change in neighboring topology can cause tip and side features to end up on the same mask, creating a cycle of DRC fix-and-run. Moreover, this restricts the design to topologies similar to that in Fig. 3.3(b), so as to limit the number of block mask bends that may create unresolvable shapes. Minimum different mask side-to-side (S2S) spacing is spacer-defined. The minimum same mask S2S spacing equals three spacer widths. Although it is possible to have a smaller value for this rule, the larger value restricts the same mask metals at odd number of spacers, thereby ensuring on-grid routing and equal spaces throughout. This is another instance of restrictive DRs producing correct by construction layout. The restrictive DRs also facilitate experiments with pure line and cut BEOL strategies.

3.5.2.4 Block Mask Related SADP DRCs

Fig. 3.6 shows the critical block mask width and spacing scenarios. The Z-shape block mask may result in sub-lithographic feature if the T2T spacing between the metals on adjacent tracks is inadequate. A DR value of 40 nm ensures the minimum block mask width. The two block masks to the right represent a T2T spacing scenario for the block mask, which requires larger spacing than the block mask width. Thus, a DR stipulates a 44 nm parallel run length between metals on adjacent routes.

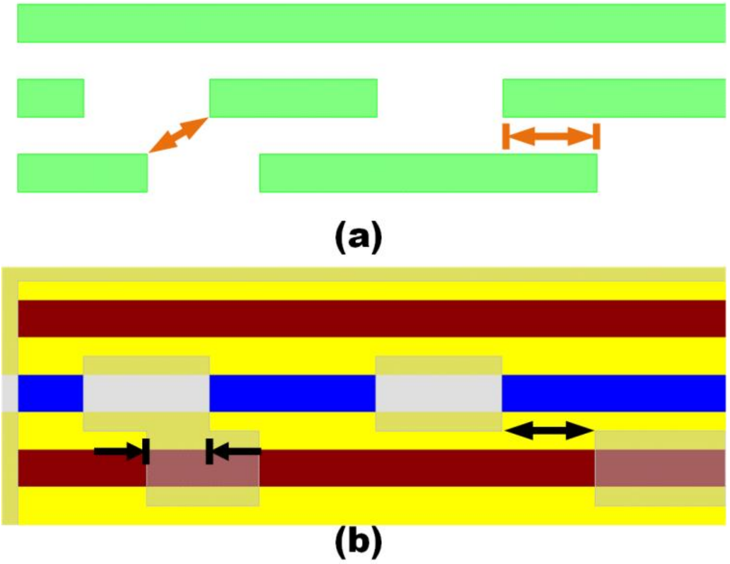


Fig. 3.6. (a) DRs stipulate minimum adjacent track T2T spacing and parallel run length to ensure (b) sufficient width and spacing, respectively, on the block mask. Figures generated by executing the PDK DRCs in CalibreDRV. (after [34]).

3.5.2.5 SADP Design Rule Summary

By mandating gridded APR layout and specific design rules, we have demonstrated that the design need not be color aware at the APR stage. The Calibre decomposition, using the outlined DRC flow, including mandrel and block mask generation, shows that the resulting masks can be used for patterning while limiting manufacturability issues. In

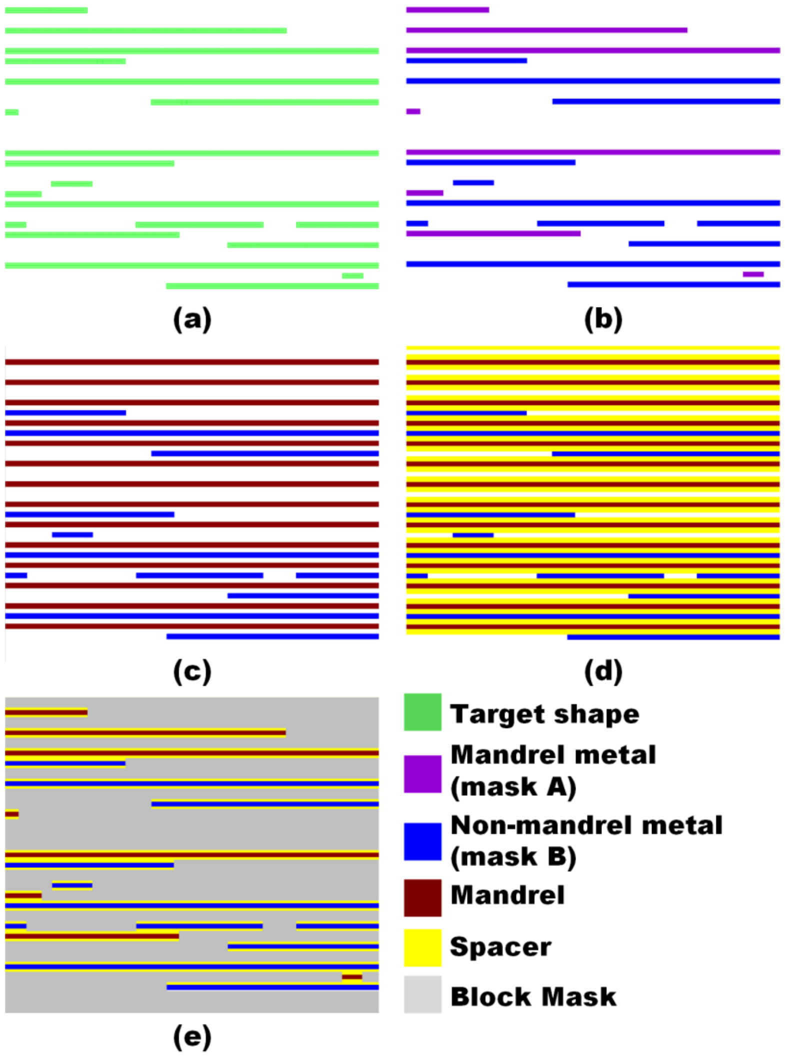


Fig. 3.7. APR block M4 decomposition with fully gridded routing. (a) Initial 1-D M4. (b) Target decomposition into two colors. (c) Mandrel formation along color A. (d) Spacer formation around mandrel. (e) Block mask creation using spacer and target metal. Its derivation ensures line widths to be delineated by spacers, i.e., SADP friendly layout. Figures generated by executing the PDK DRCs in CalibreDRV. (after [34]).

the event of errors, the intermediate results guide the designer to the structure that causes the issue. Fig. 3.7 shows the full DRC decomposition of the various SADP mask generation steps in a representative section of an APR block.

This sub-section has described DTCO for the PDK BEOL layers. We chose SADP as the most cost-effective, low variability approach. The DTCO has been driven by full APR of representative digital blocks including designs with up to 8000 flip-flops and by confirming that the masks derived via Calibre result in lithographically valid base (mandrel and block) masks.

4 CELL LIBRARY ARCHITECTURE

A standard cell library has become the usual way to construct general digital circuits, by synthesizing an HDL (typically Verilog) behavioral description and then performing automated placement and routing (APR) of the gates and interconnections. This chapter describes the ASAP7 cell library. We also explain the cells provided and discuss changes to the design rules based on APR results.

4.1 Gear Ratio and Cell Height

As in planar CMOS with gridded gates, which was introduced by Intel at the 45 nm node and other foundries at 28 nm, the selection of standard cell height in a finFET process is also application specific. Low power ASICs tend to use short cells for density, and high-performance systems, e.g., microprocessors, have historically used taller cells. The cell height constrains the number of usable fins per transistor and thus the drive strength as well as the number of M1 tracks for cell internal routing. Assuming horizontal M2, an adequate number of M2 pitches, with sufficient M2 track access to cell pins, are also required. Thus, the horizontal metal (here M2) to fin pitch ratio, known as the gear ratio, becomes an important factor.

The ¾ M2 to fin pitch ratio in the ASAP7 predictive PDK used for our libraries allows designing standard cells at 6, 7.5, 9 or 12 M2 tracks and with 8, 10, 12, or 16 fins fitting in the cell vertical height, respectively. Multiple foundries have mentioned "fin depopulation" on finFET processes, i.e., using fewer fins per gate over time. This appears primarily driven by diminishing need for higher speed, increasingly power constrained designs, high drive per fin, MP lithography considerations, and the near 1:1 PMOS-to-NMOS drive ratio, as well as power density considerations. Moreover, for academic use and teaching cell design issues, a large number of tracks makes tall cells trivial to design and their excessive drive capability makes them unsuitable for low power applications. The two fin wide transistors in 6-track library cells have limited M1 routes and pin accessibility, but the best density. 1-D layout, with two metals layers for intra-cell routing, is well suited for 6-track cells, but is not very amenable for classroom use. Therefore, we chose a 7.5-track cell library (see Fig. 4.1), which is in line with other publications targeted at N7 [10].

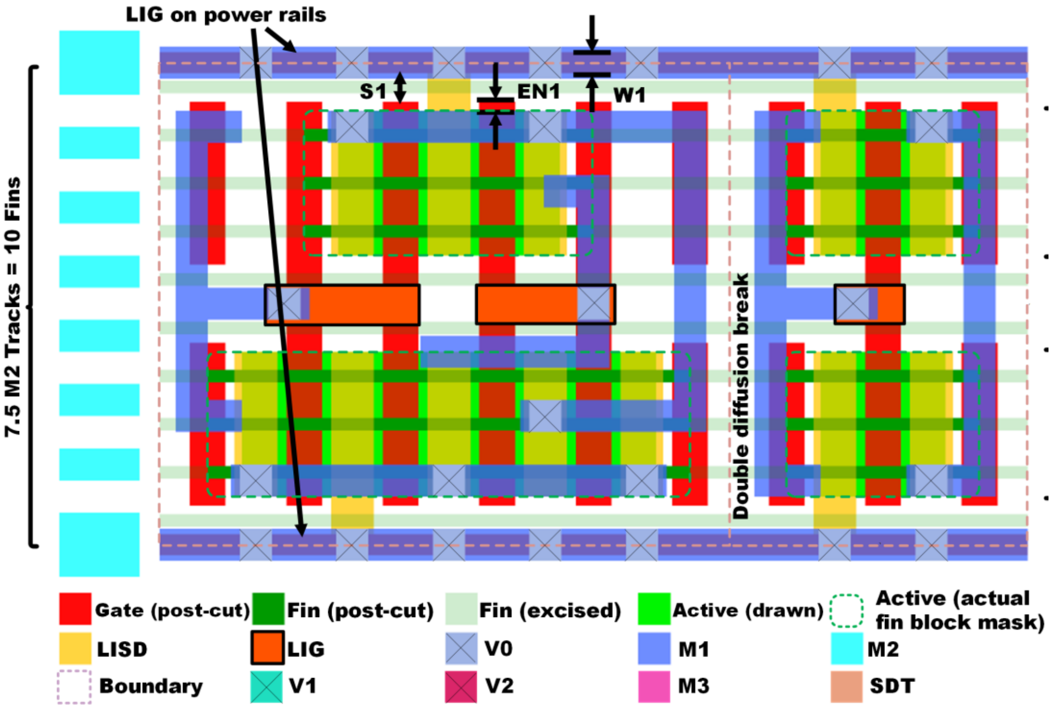


Fig. 4.1. A 7.5 M2 track standard cell template, with 3-fins per device type, shows the FEOL, MOL, and M1 in adjacent NAND2 and inverter. A double diffusion break is required between the cells. Fins are tucked underneath the gate when breaking the diffusion. Fins that are cut away are in light green. S1 = minimum LIG to GATE spacing, EN1 = minimum GATE endcap, W1 = minimum LIG width (after [12]).

This choice also allows wider M2 power follow rails at the APR stage, which are preferred for robust power delivery. However, wide M2, while supportable (and indeed required for SAQP M2) requires non SAV V1 on the power follow rails. This feature is not currently supported.

Utilizing all of the 10 fins in a 7.5 track library is not possible as sufficient separation is required between the layers SDT and LISD that connect to the transistor source-drain. Also, transistors must have sufficient enclosure by the select layers used for doping. The minimum transistor separation thus dictates that the middle two fins in a standard cell may not be used, as apparent in Fig. 4.1, that shows a NAND2 and inverter cell adjacent to each other. Fig. 4.1 also illustrates the cell boundaries producing a DDB between the cells. The fin closest to the power rails cannot be used either. Thus, each transistor has three fins and this choice is aided by the nearly equal PMOS-to-NMOS ratio. Notwithstanding the number of fins, the transistors have sufficient drive capability due to the large per-fin drive current value [11]. ASAP7 standard cells are classified based on their drive strengths, so that those with three fin transistors are said to have a 1× drive. Those will less drives strength have fraction values, e.g. a one fin transistor corresponds to a p33× (0.33×) drive strength.

4.2 Fin Cut Implications

Again, referring to Fig. 4.1, the vertical extent of the drawn active layer in the PDK denotes the raised source-drain region. As evident from Fig. 3.1 (b), Fig. 4.1, and Fig. 4.2, the actual active (fin 'keep' mask) extends horizontally halfway underneath the non-channel forming (dummy) gates, where fins are actually cut. Implementing the fin cuts in such a manner necessitates a double diffusion break (DDB) requirement for the cases where diffusions at disparate voltages must be separated. Therefore, all standard cells end in a single diffusion break (SDB), resulting in a DDB upon abutment with other cells to ensure fin separation. Apart from the cell boundaries, an SDB may exist elsewhere in a cell if equipotential diffusion regions are present on either side of a dummy gate and is equivalent to shorted fins underneath such a gate. Nonetheless, this is a useful structure, particularly in the latch and flip-flop CMOS pass-gates.

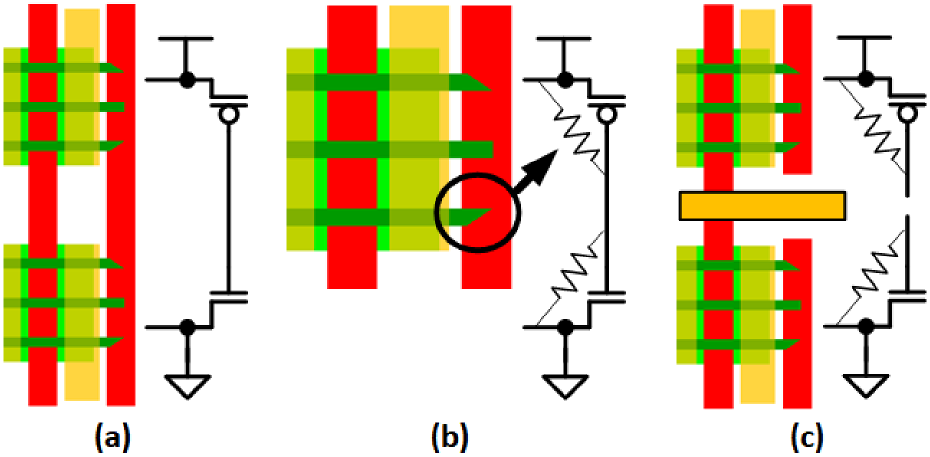


Fig. 4.2. Dummy gate at DDB (a). Sharp fin edges arise due to mask rounding when cutting fins. This creates a TDDB scenario between the fins and dummy gate (b). This is avoided where possible by cutting the dummy gates (c) (after [12]).

As shown in Fig. 4.2(a), block mask rounding effects, when implementing fin cuts, create sharp fin edges that have high charge density and consequently, a high electric field [12]. This creates a possible TDDB issue between the fin at the cut edge and the dummy gate (Fig. 4.2(b)), potentially causing leakage through the gate oxide. To mitigate this, in the libraries provided with the PDK, the dummy gates are cut at the middle in the standard cells to preclude an electrical path between the PMOS and NMOS transistors through the dummy gate. The gate cut also allows longer LIG for nearby routing without attaching to the cut dummy gate, as evident in Fig. 4.2(c).

SAQP requires fixed width and constrains the fin spacing (Fig. 4.3). Assuming an EUVL fin cut/keep mask enables the 7.5T library as designed (Fig. 4.3(a)). If MP is

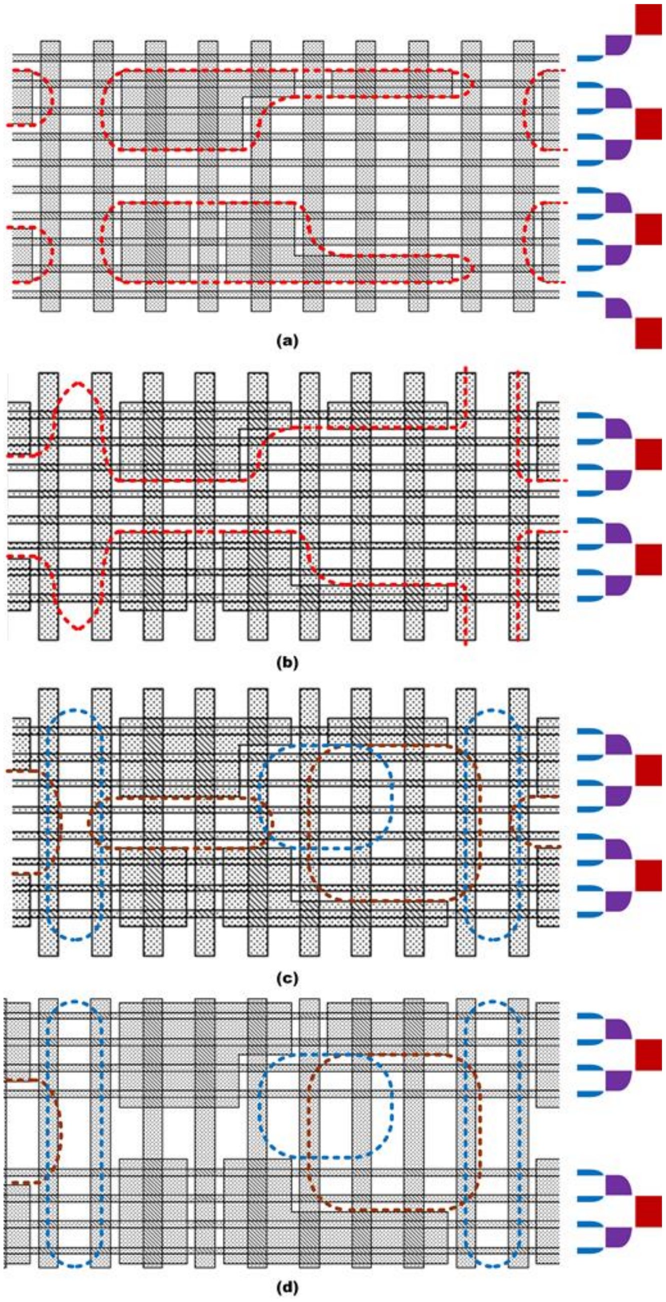


Fig. 4.3. The original EUV fin keep mask assumptions and SAQP spacer/mandrels (a). Another fin keep option that allows larger patterns (b). LELE fin cut (c) may be problematic, but an even number of fins eases this (d) (after [12]).

assumed, LELE cut/keep is complicated as in Fig. 4.3(b) and (c), where the middle fins are excised. An even number of fins for PMOS and NMOS transistors allows the SAQP mandrel to completely define the fin locations, greatly easing the subsequent fin masking as illustrated in Fig. 4.3(d). A 6T (or 6.5T) or 9T library, which has an even number of fins per device at two or four, is more amenable to SAQP fin patterning than the 7.5T library we primarily used for DTCO.

4.3 Standard Cell MOL Usage

LISD is primarily used to connect diffusion regions to power rails and other equipotential diffusions within the cell through M1, somewhat relieving M1 routing congestion. However, connections to the M1 power rails cannot be completed using LISD alone, since that would require extending LISD past the cell boundary so as to allow sufficient V0 landing on LISD, or 2-D layout that would have issues with the T2S spacing. Larger LISD overlap (past the cell boundary) would violate the 27 nm LISD tip-to-tip (T2T) spacing by interaction with the abutting cell LISD that is not connected to the power rails. Therefore, an LIG power rail is used within the cells to connect LISD to M1, as LISD and LIG short upon contact, but being different layers, do not present 2-D MEEF issues. The PDK DRs require a 4 nm vertical gate endcap past active and a 14 nm LIG to gate spacing (S1). These rules restrict the LIG power rail to its 16 nm minimum width (W1). This width does not provide fully landed V0, but the LIG power rail is fully populated with V0 to ensure robustness and low resistance.

LISD can also serve a secondary role of a routing layer within a cell, but is used sparingly, as local interconnects favor tungsten over copper, and therefore suffer from higher resistivity [47]. However, LISD routing is helpful in complex cells, especially sequential cells, due to M1 routing congestion and allows us to limit M2 track usage. Fig. 4.4 shows a D-latch, in which LISD use to connect diffusions across an SDB of a constituent tristate inverter becomes necessary, since M1 cannot be used due to intervening M1 routes or their large T2T and T2S spacing requirements. The SDBs arise due to the inability to accommodate two gate contacts along a single gate track, which also necessitates a M1 crossover. SDBs are used in the cell CMOS pass-gate (crossover) comprised of a CLKN-CLKB-CLKN gate combination evident in the schematic at the bottom of Fig. 4.4. Double diffusion breaks are used in the cell, as all diffusions cannot be shared.

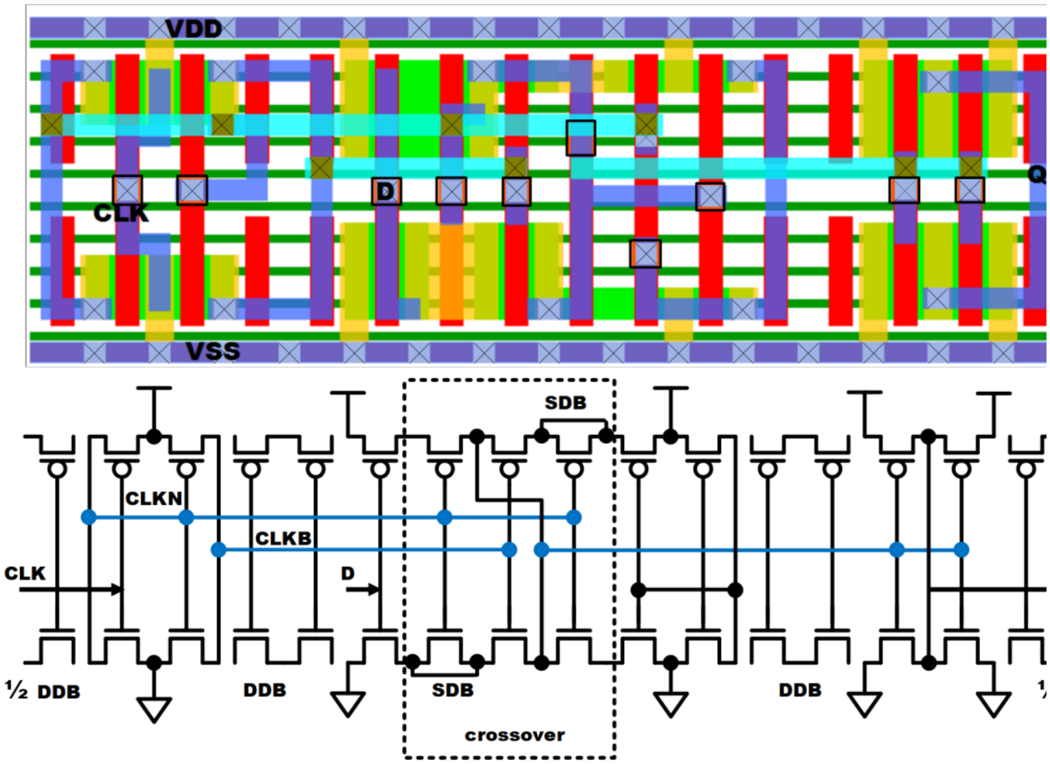


Fig. 4.4. Cell layout of a transparent high D-latch. Double and single diffusion breaks are shown. The latter require LISD crossovers. Fins shown are prior to the cut. (after [12]).

4.4 Standard Cell Pin and Signal Routing

The cell library architecture emphasizes maximizing pin accessibility, since it has a direct impact on the block density after APR. Where possible, standard cell pins are extended, so as to span at least two, and preferably three horizontal M2 routing tracks. This ensures that even standard cells with a large number of pins, such as the AOI333 in Fig. 4.5, have all of their M1 pins accessible to multiple M2 tracks. Fig. 4.5 demonstrates that the use of staggered M2 routes to allow accessing the cell pins by M3. Good pin access is forced by our use of the M1 templates for cell design [Vash17d]. The M1 template was employed for the rapid cell library development, as the pre-delineated metal constructs take the M1 design rules into consideration, which aids in DRC error free M1 placement. Most, but not all of the M1 routing in the standard cells is template based.

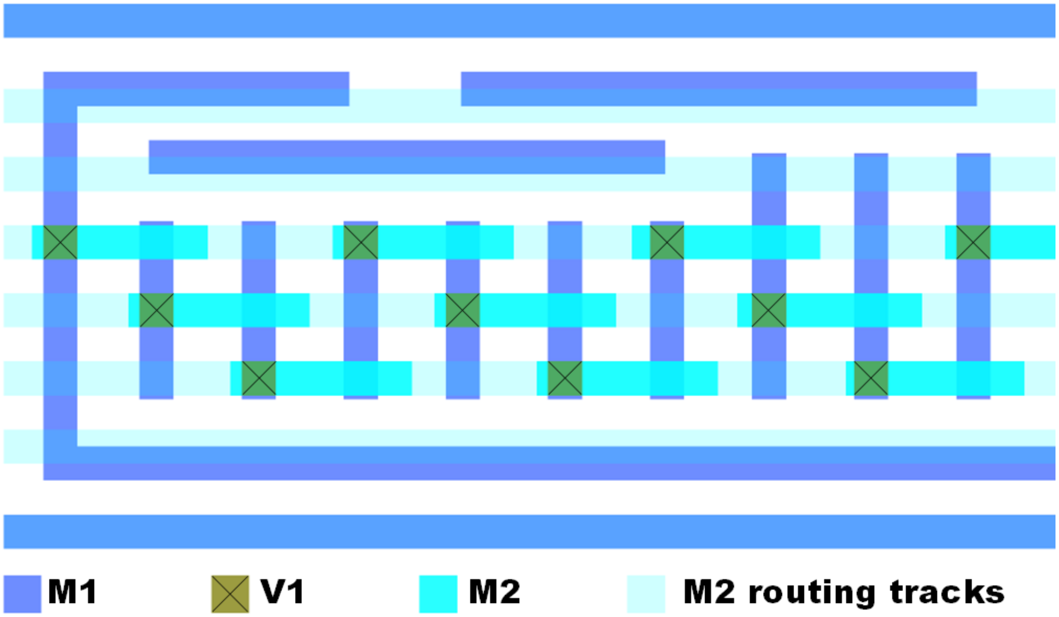


Fig. 4.5. An AOI333 cell with all of its input M1 pins connected to M2. The staggered M2 allows further connections to M3. Note that this version does not use the bends on M1, following the post-APR DTCO changes. These vias may be slightly unlanded (after [12]).

Referring to Fig. 4.6, the non-power M1 lines closest to the cell boundary are horizontal, which alleviates the T2S DR spacing requirement for such routes as they constitute a side, rather than a tip feature. This approach was used for the initial cell version in the library, as it allowed better V1 landing from M2. However, this feature was removed after extensive APR validation, as mentioned below. The 7.5 track tall standard cell results in equidistant M2 routing tracks that are all 18 nm wide, except for wider spacing of 36 nm around the M2 power follow-rails. The cells accommodate seven horizontal M1 routes that are arranged as two groups of three equidistant tracks between the M1 power rails, with a larger spacing at the center of the cell where vertical M1 is required for pins.

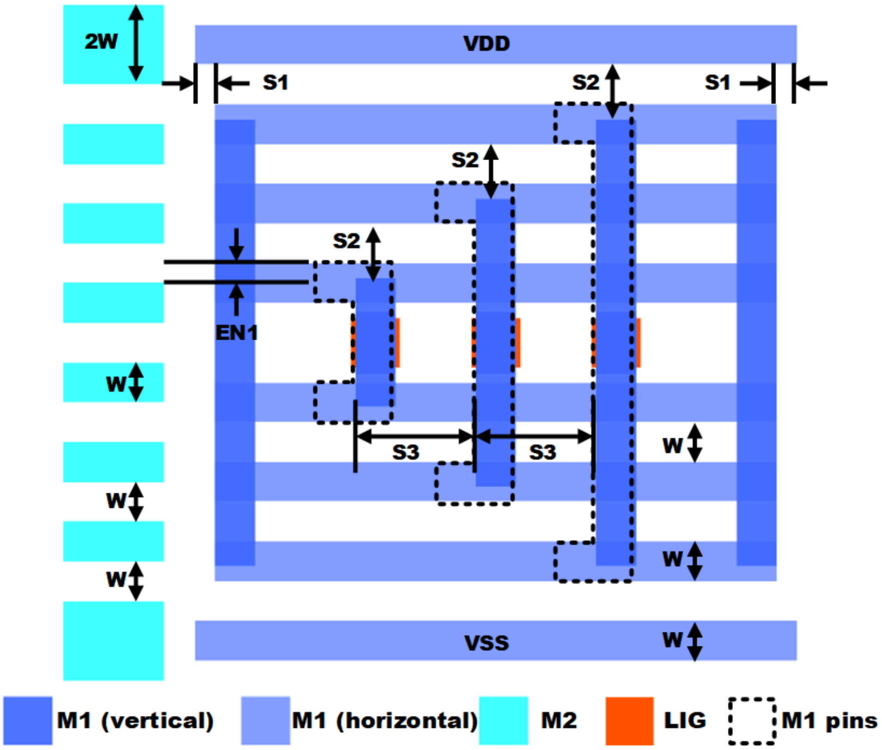


Fig. 4.6. M1 template to ensure and maximize pin access. S1 = 0.5 × M1 side-to-side (S2S) spacing; S2 = M1 tip-to-side spacing; S3 = gate pitch; EN1 = V1 enclosure by M1 pin (after [12]).

Accounting for the extra M1 spacing near the center, instead of the power rails, helps to maximize pin access through M1 pin extension past the M2 tracks adjacent to the power rails. The vertical M1 template locations, corresponding to the pins, match the CGP. Originally, they always ended at a spacing of 25 nm from the nearest horizontal M1 constructs that they do not overlap, so as to honor the M1 T2S spacing. However, L-shaped pins formed using solely vertical M1 constructs on either one of their ends, do not provide the stipulated 5 nm V1 enclosure necessary when connecting to an M2 track. Consequently, the final topology in Fig. 4.6, which does not allow full V1 landing to input pins is adopted for releases after the 1st, as described below.

Connecting to M1 near the power rails is enabled by via merging. While the example presented here is for V0, all layers support via merging. The presence of V0 on the power rails may engender a case similar to that in Fig. 4.7(a), where the unmerged via mask spacing 'd' may be smaller than the lithographically allowed minimum metal side-to-side spacing 'y'. Merging these vias (Fig. 4.7 (b)) in such cases permits the patterning of V0 containing M1 tracks at minimum spacing. It is enabled by keeping the vias on grid as evident in Fig. 4.7 (c).

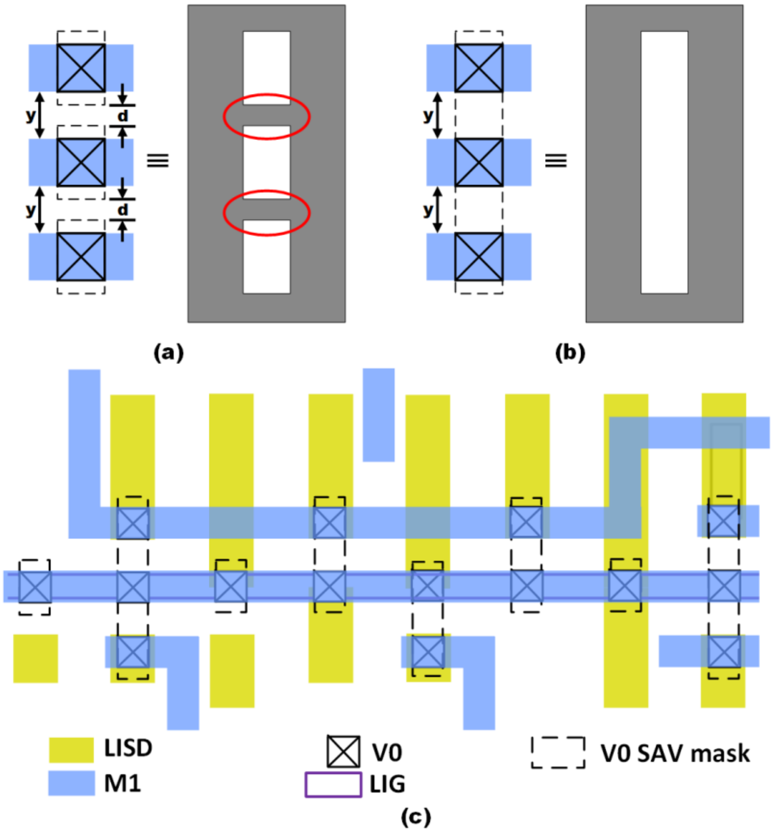


Fig. 4.7. (a) Sub-lithographic features (ellipses) on SAV mask. (b) Merging SAV masks precludes such features. (c) Merging enables M1 signal routes to be at the minimum spacing from the power rails, which would otherwise have been not possible due to un-merged SAV mask shapes (after [12]).

4.5 Library Collaterals

A cell library requires Liberty timing characterization files (compiled into .db for Synopsys tools) as well as Layout Exchange Format (LEF) of the cell pins and blockages (FRAM for Synopsys). We have focused primarily on Cadence collateral, since we are most familiar with those tools. The technology LEF provides basic design rule and via construction information to the APR tool. This is a key enabling file for high quality, i.e., low DRC count APR results. Additional files are the CDL (spice) cell netlists. We provide Calibre PEX extracted, and LVS versions of the netlists. The former allow full gate level circuit simulation without needing to re-extract the cells. Data sheets of the library cells are also provided. These are generated automatically by the Liberate cell characterization tool. Mentor Graphics Calibre is used for DRC, LVS and parasitic RC extraction. The parasitic extraction decks allow accurate circuit performance evaluation.

4.6 DTCO Driven DR Changes Based On APR Results

The original M1 standard cell template shown in Fig. 4.8(a) above was meant to provide fully landed V1 on M1 in all cases. However, this meant that where an L shape could not be provided, the track over that M1 portion of the pin was lost. APR tools generally complete the routing task, resulting in DRC violations when the layout cannot be completed without producing one. The APR exercises illuminated some significant improvements that could be made, as well as pointing out some substandard cells, which have been revised. A key change to the library templates is illustrated here.

Based on reviewing APR results described below, the V1 landing design rule was changed to 2 nm, which does not provide full landing with worst-case misalignment. However, most standard cell pins are inputs and thus drive gates. We decided that gaining the useful pin locations was desirable for multiple reasons. Firstly, a single gate load does not present much capacitance, so the extra RC delay from a misaligned (partially landed) via will not significantly affect delay. Secondly, some gates, notably those with a large number of inputs such as the AOI22 cell, have considerably better pin access, as shown in Fig. 4.8(b). Initially, two of the pins only had one usable M2-M1 intersection at the middle. These also conflicted for the same M2 track. The improved layout is simpler, and has lower capacitance, as well as moving the pins to a regular horizontal grid identical to the 54 nm gate grid. The cell output nodes still require 2-D routing to reach the correct diffusions. This maintains the original full 5 nm landing for the output node that drives a much larger load and is subject to self-heating.

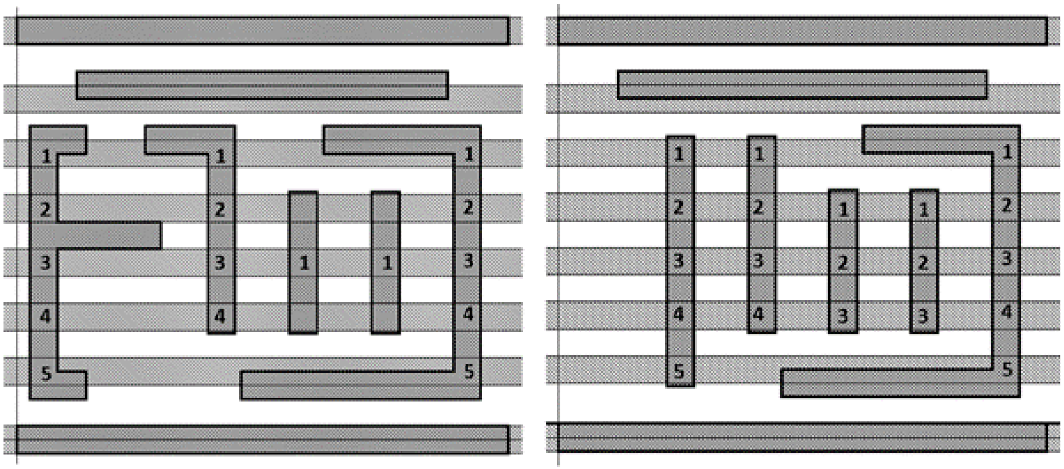


Fig. 4.8. The original AOI333x1 M1 layout and M2 track overlay (a) and the new layout (b) showing the improved access by relaxing the V1 to M1 overlap rule from 5 nm (full landing) to 2 nm.

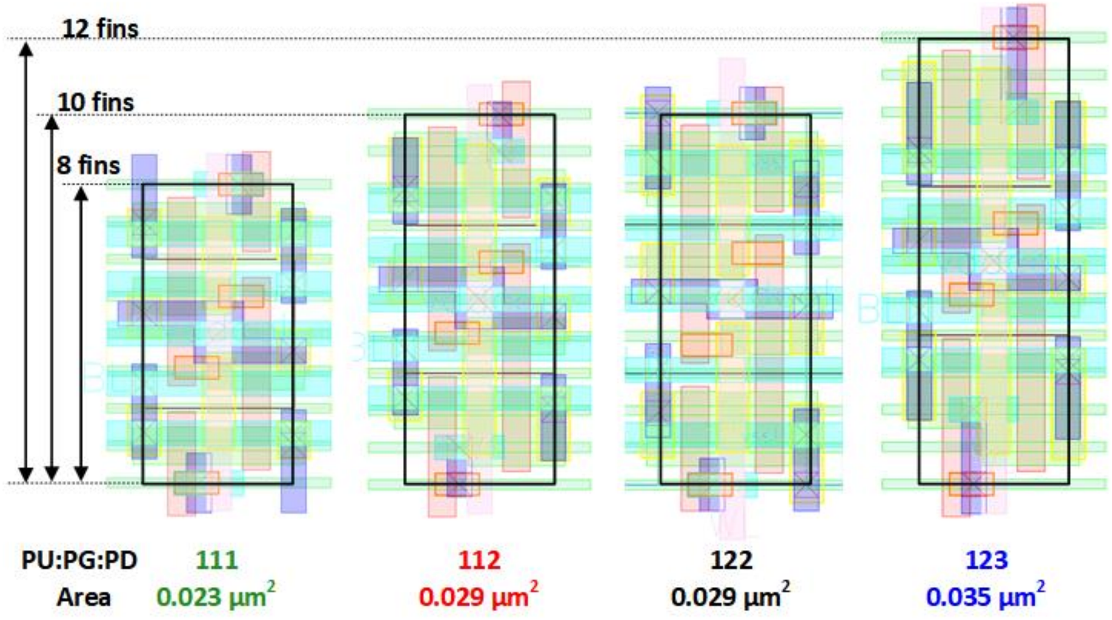
5 SRAM DESIGN WITH ASAP7

SRAMs are essential circuit components in modern digital integrated circuits. Due to their ubiquity, foundries provide special array rules that allow smaller geometries than in random logic for SRAM cells to minimize their area. SRAM addressability makes them ideal vehicles for defect analysis to improve yield in early production. Moreover, running early production validates the issues arising from the tighter design rules. One focus of this chapter is the manner in which SRAMs affected the DTCO analysis for the ASAP7 design rules. They turn out to be more limiting, and thus more important than the previously discussed cell library constructs.

Due to use of the smallest geometries possible, SRAMs are especially prone to random microscopic variations that affect SRAM cell leakage, static noise margin (SNM), read current (speed) and write margin. Consequently, they also provide a place to discuss the PDK use in statistical analysis, as well as our assumptions at N7. Historically, providing the needed 6-T SRAM transistor drive ratios for the needed cell write-ability and read stability has been provided by very careful sizing of the constituent transistors, i.e., the pulldown is largest, providing a favorable ratio with the access transistor to provide read stability, while the pull up PMOS transistor is smallest, so that the access transistor can overpower it when writing. Improving PMOS vs. NMOS strain has led to near identical, or in some literature greater PMOS drive strengths, which in combination with discrete finFET sizing, requires read or write assist techniques for a robust design. The yield limiting cases occur for cells that are far out on the tail of the statistical distribution, due to the large number of SRAM cells used in a modern device. Consequently, statistical analysis is required.

5.1 FinFET Implications and Fin Patterning

Besides the constraint on transistor width to discrete fin count, MP techniques, e.g., SADP or SAQP, further complicate the allowed cell geometries. On finFET processes, SRAM cells are divided into classes based on the ratios of the pullup (PU), pass gate/access (PG), and pulldown (PD) ratios, represented by PU, PG, and PD fin counts. The different sized cells that we used for DTCO in the PDK comprise Fig. 5.1.

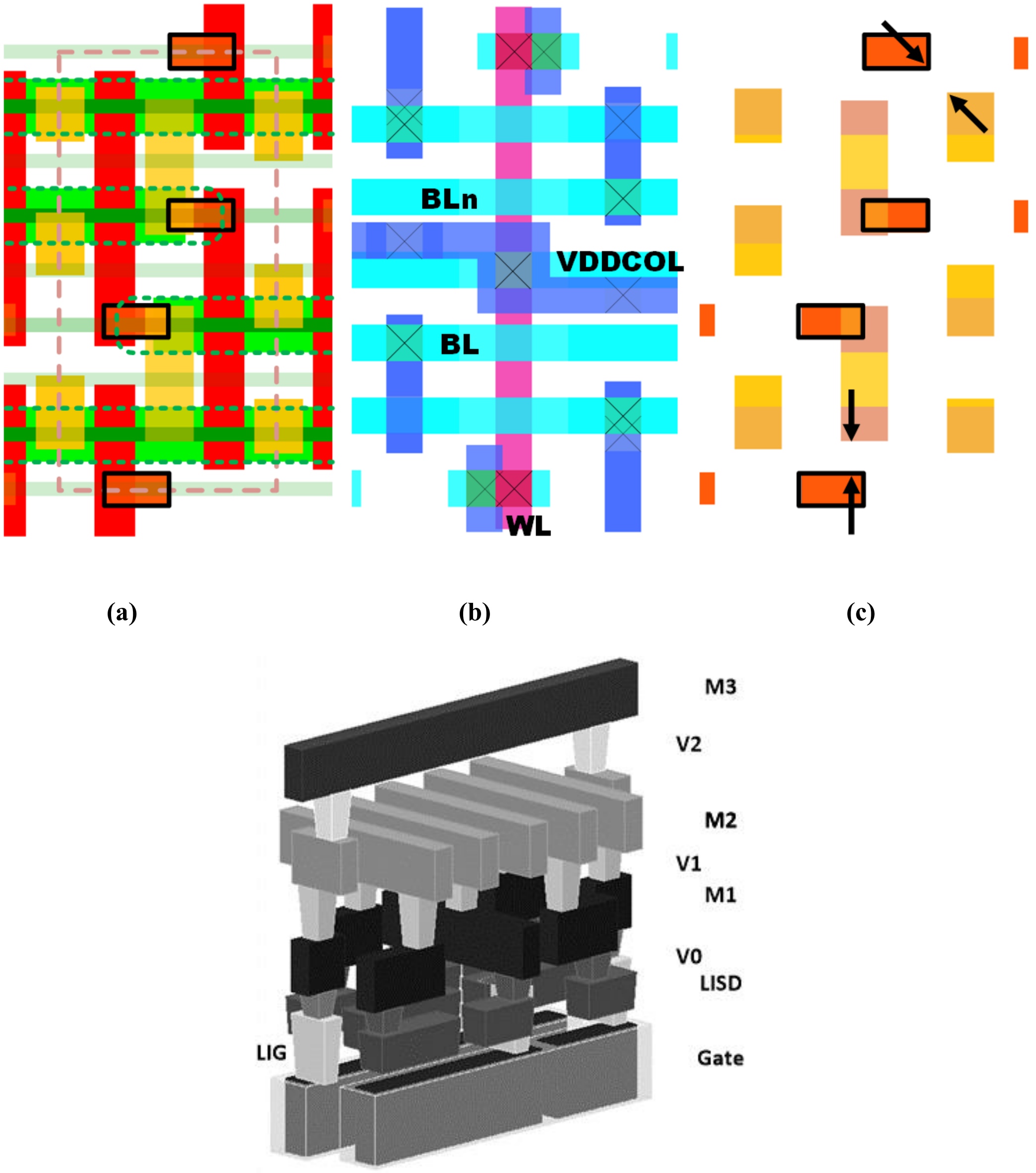


(a) (b) (c) (d)

Fig. 5.1. Layouts for 111 (a), 112 (b), 122 (c) and 123 (d) ASAP7 SRAM cells. 112 is same footprint as 122 (after [61]*).*

For instance, the smallest cell is 111 (Fig. 5.1(a)) and has nominally equal drive strength for each device. The cell that most easily meets the read stability and write-ability requirements outlined above is thus the 123 cell whose layout is illustrated in Fig. 5.1(d). As in the standard cells, there must be adequate spacing between the NMOS and PMOS devices for well boundaries, as well as active region separation. Thus, at least one fin spacing is lost between adjacent NMOS transistors in separate cells and between the NMOS and PMOS devices. The 112 and 122 cells comprise Fig. 5.1(b) and (c), respectively.

The SRAM DR active mask spacing, which is optimistically set at a single fin, is evident in Fig. 5.2 (a) and (c). The single fin spacing allows the SAQP fin patterning to



(a) **(b)** **(c)**

(d)

Fig. 5.2. ASAP7 111 SRAM cell layout showing fin, active, SDT, gate and gate cut layers (a), M1 and M2 as well as via 1 (b) and active and MOL layers with arrows to indicate critical design rules (c) (after [12]. 3-D cell view with the correct metal aspect ratios (d). The gate spacer is shown as transparent, but fins are not shown.

be uniform across the ASAP7 die. However, the fin patterning can be changed (and often is on foundry processes) by adjusting the mandrels. The EUV assumption drives some of the metal patterning. Referring to Fig. 5.2(b), note that the M1 is not 1-D, in that the cell V_{DD} connection zigzags through the array on M1. This M1 is redundant with the straight M2 route. This also provides full M1 landing for the via 1 (V1) connection in each cell. The zigzag is forced by the MEEF constraints described earlier (the M1 lines vertical in Fig. 5.2 (b) that connect the V_{SS} and BL to the MOL). Since the MOL provides a great deal of the connections, M1 BL designs are also possible and given the slow actual roll out of EUV, are probably dominant at the 10 nm and possibly the 7 nm nodes. However, they are not compatible with the horizontal M2 direction outlined above for ASAP7. Consequently, we focus here on M2 BL designs, which maintain the same metal directions that are used in the standard cell areas across the SRAM arrays. The cell gate layers through metallization are illustrated in Fig. 5.2(d) which emphasizes the high aspect ratios of modern metallization. The fins are omitted from the Fig. 5.2(d).

5.2 Statistical Analysis

AVT has long been used as a measure of local mismatch. AVT is a transistor channel area normalized $\sigma\Delta V_T$, where $\sigma\Delta V_T$ is the V_T difference variance as measured between nearby, identical transistors [62][63]. Thus,

$$AVT = \sigma\Delta V_T \times\sqrt{A} \tag{9}$$

$$\sigma\Delta V_T = \sigma(|V_{T1} - V_{T2}|). \tag{10}$$

In non-fully depleted devices, the dopant atoms are proportional to the area but vary statistically as random dopant fluctuations that in aggregate affect the transistor V_{th}, and dominate the mismatch. Other parameters affect the matching, but can often be driven out by improvements in the manufacturing processes and design. SRAM memories use the smallest devices that can be fabricated (in this case a single fin). Consequently, SRAMs are strongly affected and statistical yield analysis is critical to the overall design [64]. As a result, SRAM yield falls off rapidly as V_{DD} is lowered towards V_{DDmin}, the minimum yielding SRAM V_{DD}. This is due to increasing variability in the drive ratios as the transistor gate overdrive $V_{GS} - V_{th}$ diminishes, due to V_{th} variations. In metal gate and finFET devices, the variability is primarily due to fin roughness and metal gate grain size variations [64] [65]. Nonetheless, due to its ease of measurement and historical significance, AVT is still used to characterize mismatch [63]. FinFET devices, while having different sources of variability since they are fully depleted, continue to use the AVT rubric for analysis. In general, AVT tracks the inversion layer thickness, including the effective oxide thickness. It improved by nearly ½ with the advent of high-K metal gate processes, which eliminated the poly depletion effect, and again with finFETs, as they are fully depleted [63]. The primary sources of variability in finFETs turn out to be metal gate grain size and distribution, followed by fin roughness and oxide thickness and work function variations. We use AVT = 1.1 mV.µm, which we chose as a compromise between the NMOS and PMOS values published for a finFET 14-nm process [66]. Literature has reported pretty steady values across key device transistor fabrication technologies, i.e., poly gate, metal gate, and finFET. Of course, the overall variability increases in our 7 nm SRAM transistors vs. the 14 nm devices, since the channel area is less.

For sense amplifier analysis, we use a conventional Monte-Carlo (MC) approach, applying different differentials to a fixed variability case to find the input referred offset, then applying the next fixed variability, to reach the desired statistics. 1k MC points are used to determine the offset mean and sigma. For SRAM cell analysis, we use a combination of Monte-Carlo and stratified sampling [67]. In general, unless a large number of simulations are used, Monte-Carlo is not adequate for estimations at the tails of the distributions. It is, however, adequate to estimate the standard deviation σ and mean μ of the distributions. We confirm the needed σ/μ using the stratified sampling approach, which applies all of the statistical variability at each circuit (sigma) strata, to all the possible combinations of the transistors within that strata, using a full factorial. Thus, 728 simulations are performed, one for each possible 1, 0, or -1 impact on each combination of devices at each strata, attempting to find failures in that circuit strata (at that circuit sigma). As an example, if the circuit sigma is distributed evenly across six devices, then there is minimal variation in the circuit. However, if all of the circuit level variability is applied to just the access and pulldown transistors, and in opposite fashion, then the read stability may be jeopardized. In this manner the technique evaluates the tail of the circuit distribution directly and efficiently. At each strata, the transistor variation applied changes with the number of non-zero variations, following the circuit σ_{DEVICE} as

$$\sigma_{DEVICE} = \frac{\sigma_{CELL}}{\sqrt{\Sigma a_i^2}} \tag{11}$$

The technique has been successful at predicting the SRAM yield for foundry processes [67].

5.3 SRAM Cell Design and DTCO Considerations

5.3.1 MOL Patterning

Referring to Fig. 5.2(c) the key DTCO limitation is the corner of LISD and SDT to LIG spacing as indicated by the arrows at the upper right. EUV single patterning would result in some rounding, which helps by reducing the peak electric field and increasing the spacing slightly. As before, we use conservative assumptions. Another result of maintaining adequate spacing for TDDB with worst-case misalignment of the LISD/SDT and LIG layers is that the SDT does not fully cover the active areas, evident at the top and bottom of the NMOS stacks (Fig. 5.2(a)). This led us to separate the LISD and LIG drawn layers. We believe that given the late introduction of EUV, it will most likely be used for MOL layer patterning, as well as vias and cuts. The former is due to projections that MOL layers may require as many as five block masks for SRAM at N7, which makes the expense of EUV more favorable [68]. Foundry presentations have also shown MOL experimental EUV results.

5.3.2 1-D Cell Metallization

The 111 cell (Fig. 5.2) has 1-D M2 and M3, potentially making it very amenable with a 1-D cell library, and as mentioned, M1 BL designs are compatible with horizontal M1 1-D cell library architectures. The 2-D cell layouts of the 122 cell are shown in Fig. 5.3 (a)-(d). The similarity to the overall architecture of the 111 cell is apparent. The same MOL limitations exist, but the two fin NMOS devices make the overall source/drain connections to the NMOS stacks considerably better, as they have more coverage and lower resistance at worst-case layer misalignments. The EUV M2 assumption allows narrow metals with wide spacing. This in turn reduces line to line coupling and overall capacitance. Other widths are possible.

Fig. 5.3 (f)-(g) show 1-D metal impact on the cell design at the M2 through MOL layers. Here, M1 and M2 are produced by 1-D stripes using SADP or SAQP, respectively, and then cut to produce separate metal segments. The cuts are assumed the same as SAV, i.e., 16 nm. Cuts are aligned to each other where possible. Wider M2 is required given the SID assumptions. The SAQP mandrel and 1st and 2nd level spacers to produce this layout

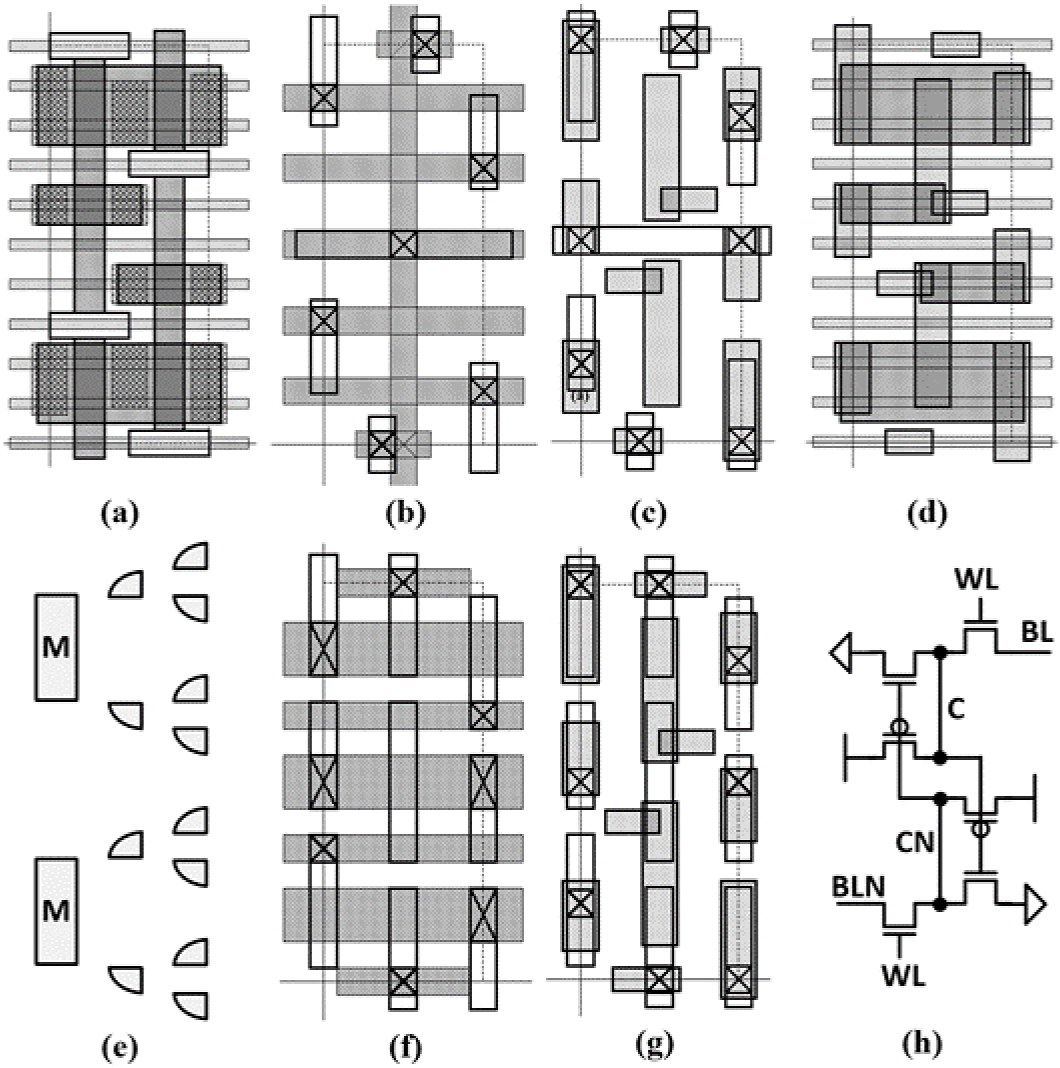


Fig. 5.3 ASAP7 122 SRAM cell layout showing fin, active, SDT, gate and gate cut, active and SDT layers (a) M3, V2, M2, V1 and M1 layers (b), M1, V0 and MOL (c). Fin, active and MOL (d). The SAQP mandrel and spacers produced for SID M2 (e) and the resulting M2, V1, and M1 layers (f). M1 through MOL (g) and the schematic (h).

is shown in Fig. 5.3(e). The spacers walk across the cells, but the wide, thin, wide, thin repetitive pattern is easily produced. The lines/cuts 1-D metal approach requires a dummy M1 in the middle of the cell, evident in Fig. 5.3(f) and (g). The SAV V1 are wider, but still aligned by the M2 hard mask, following the convention used for the APR power routes. Note that these vias are un-landed on M1. Fig. 5.3(h) is provided as an aid to the transistor layout pattern for readers who are not readily familiar with theses modern, standard SRAM layouts.

5.3.3 Stability and Yield Analysis

Read mode static noise margin (SNM) analysis following [69] at the typical process corner is remarkably similar for the four cells [61]. The 112 cell has the best SNM due to the 1:2 PG to PD ratio, followed by the 123 cell with its 2:3 ratio. At TT the 122 cell has the lowest value, but it is very close to the 111 cell.

	Cell Type	111	112	122	123
Margin Type	Quantity				
Read SNM	μ	105.2	116	100.4	106.6
	σ	11.2	9.48	8.3	7.5
	μ/σ	9.4	12.2	12.1	14.2
Hold SNM	μ	119.0	126.1	126.1	128.8
	σ	8.2	6.2	6.2	5.4
	μ/σ	14.4	20.3	20.3	23.9
Write Margin	μ	68.1	69.3	103.6	98.4
	σ	20.9	23.8	18.3	15.8
	μ/σ	3.3	2.9	5.7	6.2

Table 5.1 Mean, variation, and mean/sigma of read SNM, hold SNM, and write margin simulations for 111, 112, 122, and 123 SRAM cells at nominal VDD = 0.5 V (hold at VDD = 350 mV) and 25ºC (after [61]).

Referring to Table 5.1, the read SNM σ is inversely proportional to the number of fins as expected. The read SNM overall quality can be ranked by the standard μ/σ of each, since it gives an indication of the sigma level at which margin vanishes. Using this metric, the largest (123) cell is the best, as expected. The 111 cell is the worst, also as expected, entirely due to greater variability. The 112 and 122 cells are nearly even, with the latter making up for a lower baseline SNM with lower variability. The lithography for the 122 cell is easier, and as shown below, it has better write margins, so it becomes the preferred cell. With the possible exception of the 111 cell, all cells have good read margins.

In the arrays presented here, a CMOS Y-multiplexer allows a single tristate write circuit per column and eliminates the need for separate write enables [61]. Write margin is essentially the BL voltage at which the cell writes under the worst-case voltage and variability conditions. For this analysis, we use an approach that includes the write driver and Y-multiplexer variability, since the write signal passes through it. The analysis thus includes not only the write-ability of the cell, but also the ability of the column circuits to drive the BL low. MC generated V_{th} offsets are applied to the cell, Y-mux, and write driver transistors. The write driver input swings from V_{SS} to V_{DD}, sweeping the driver output voltage low. We tie the gate of the opposing PMOS transistor in the SRAM cell low, so that the circuit ratio varies only with the write driver voltage. The margin is the BL difference between the point at which the far SRAM node (CN) rises past $V_{DD}/2$ and the lowest BL voltage that can be driven. Table 5.1 shows that the 122 cell has good margin.

The high sigma margins are confirmed by the stratified sampling approach as shown in Fig. 5.4. The first fails occur at the sigma predicted in Table 5.1, after 2.9 σ/μ for the 112 cell, and after 5.7 (with some added margin) for the 122 cell. While five sigma margins have been suggested in academic papers [70] and this value has been used in commercial designs at the column group level, we consider it inadequate at the cell level since a column group can have as many as 8×256 cells. The conclusion is that the preferred 122 cell, which has good density and read stability, would benefit from write assist. The hold margin controls V_{DDmin}, as well as limiting write assist using reduced V_{DD} or raised V_{SS}. Our target V_{DDmin} is 0.5 V, so we use V_{DDcol} of 350 mV as the hold SNM evaluation point to provide guard band. The results are also shown in Table 5.1. Low SRAM transistor I_{off} allows good I_{on} to I_{off} ratio at low voltage, providing large hold SNM at 350 mV V_{DD} μ/σ.

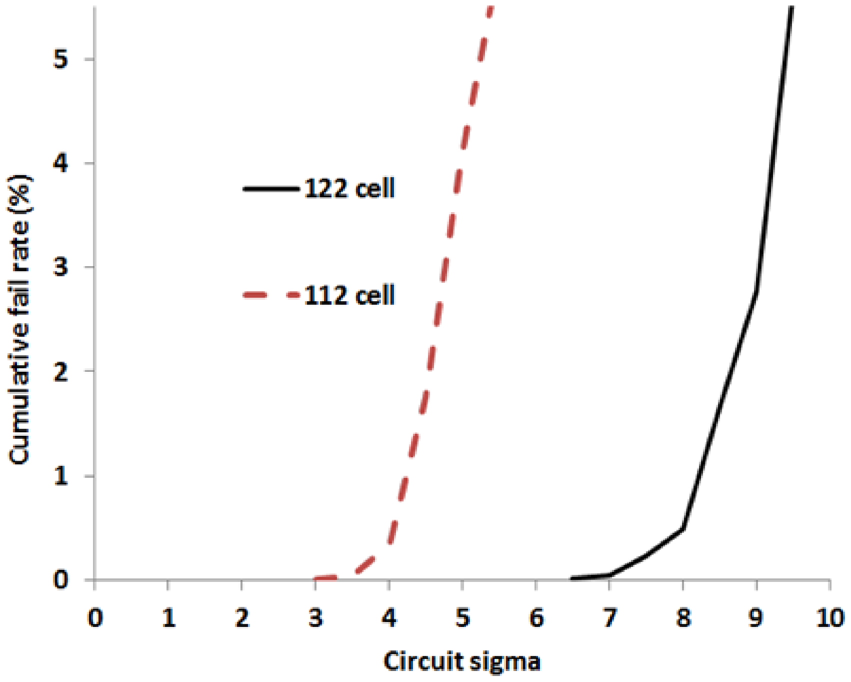


Fig. 5.4. Comparing the failure rates of the 112 and 122 SRAM cell write margins using the stratified sampling approach.

5.4 Array Organization and Column Design

Further analysis comprehends more than one SRAM cell, so we proceed by describing the column group, which is the unit that contains the write, sense, and column multiplexing circuits, as well as multiple columns of SRAM above and below it. The

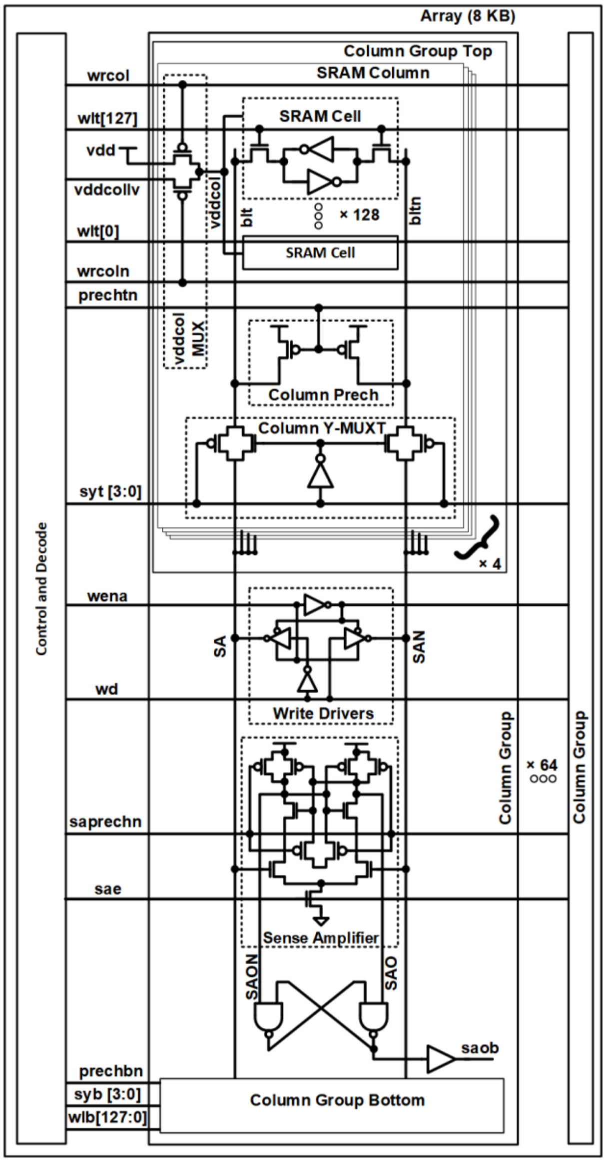


Fig. 5.5. Column group circuits shown in the context of the SRAM array. Four SRAM columns are attached at the top and bottom of the sense and I/O circuits, sharing a common CMOS pass gate Y-mux for reads and writes.

basic circuit is shown in Fig. 5.5, including the “DEC” style sense amplifier. Typical number of columns in a column group are four or eight columns of SRAM per sense amplifier, so we (somewhat arbitrarily) chose four.

This sense circuit is chosen since it is naturally isolated from the BLs during writes and one of the authors prior experience has shown it to have similar mismatch, but simpler timing than the sense amplifier using just cross-coupled inverters. Note that changing the array size merely requires changing the height of the SRAM columns and the number of column groups included in the array. As shown in Fig. 5.5, the differential sense amplifier drives a simple set-reset (SR) latch to provide a pseudo-static output from the array. The SR latch has a differential input, so it will be well behaved, and does not need a delayed clock that a conventional D-latch would in this function. There is thus no race condition on the sense precharge.

The sense amplifier input referred offset voltage was determined by SPICE MC simulation. As noted above, the simulations apply between -100 and 100 mV to V(SA) – V(SAN) in 1 mV increments for each random (MC chosen) amplifier transistor mismatch selection. The point at which the amplifier output changes direction is the input-referred offset for that mismatch selection. The results produce a mean offset less than 1 mV, which is near the expectation of zero, also indicating no serious systematic offsets due to the layout.

The distribution is Gaussian, as indicated by the SA input-referred offset voltage quantile plot shown in Fig. 5.6. The input referred offset standard deviation is 16.5 mV. Using the aforementioned five sigma offset for the sense, 82.6 mV of voltage difference is required at the sense nodes for correct operation. We guard band this up to 100 mV since extra signal is required to provide adequate speed—the sense operation with no residual offset starts the circuit in the metastable state and can produce significant output delays.

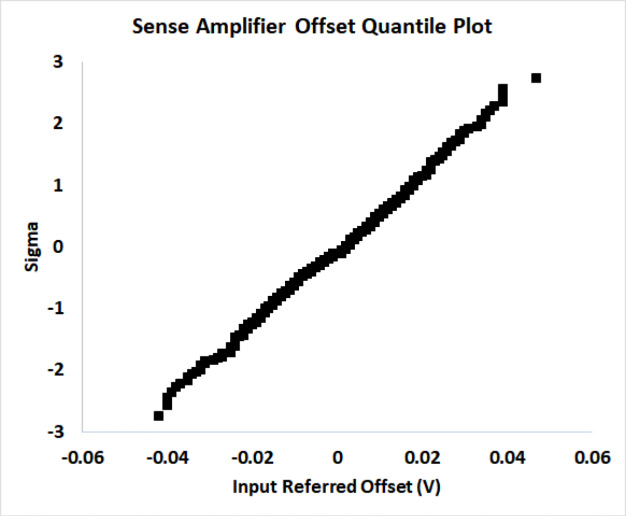
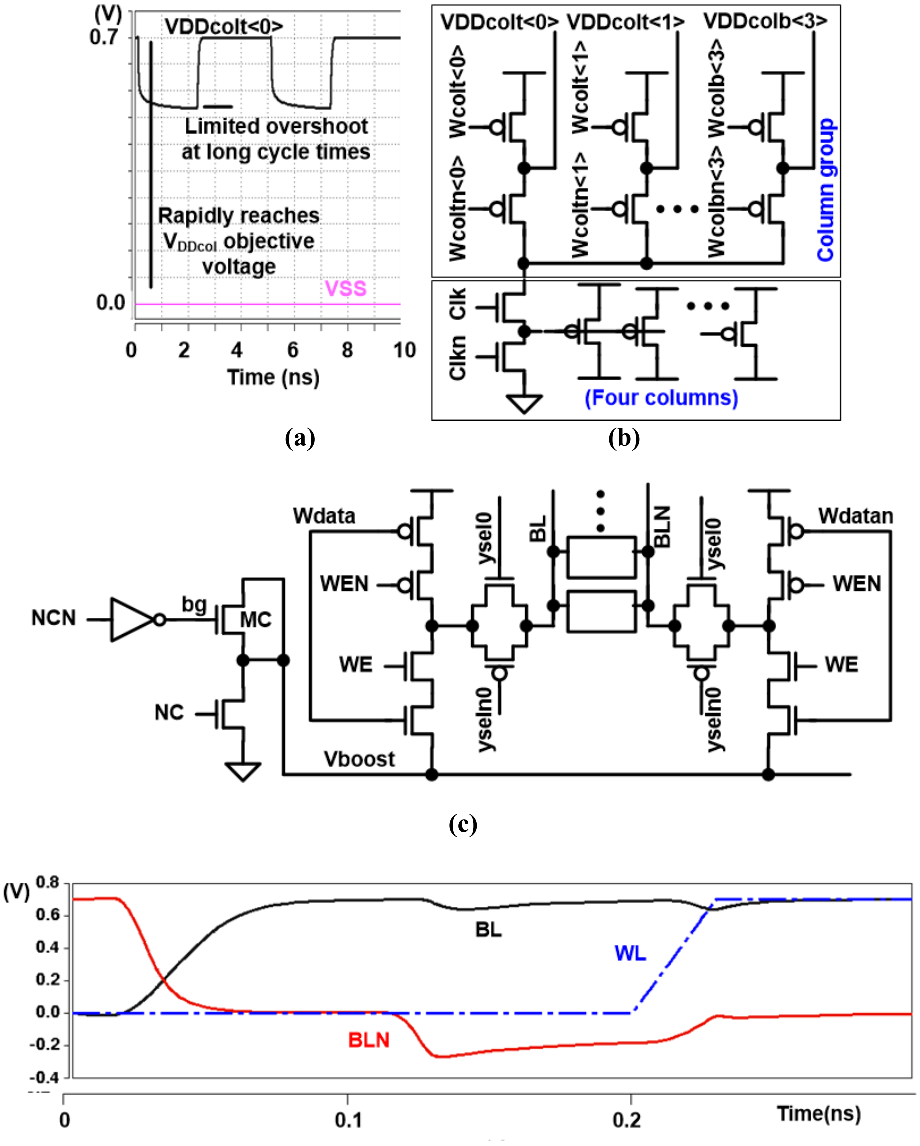


Fig. 5.6. Quantile plot of the sense amplifier input-referred offset voltage determined by Monte-Carlo simulations (after [61]).

5.5 Write Assist

The limitations of transistor width and lengths possible given the MP and other limitations essentially requires the need for read or write assist techniques in the ASAP7 SRAM memory designs. Thus, circuit design techniques provide the needed statistical yield margins instead of the cell geometries [71]. While many assist approaches have been published, the impact on the overall SRAM size, as well as the variability of the assist techniques themselves, drove the choice for our ASAP7 SRAM arrays.

We evaluated two approaches. First, driving the V_{DD} of the column being driven



(a) (b)

(c)

(d)

Fig. 5.7. Simulation waveforms of the bitcell column supply voltage modulation using our write assist technique (a). Write assist circuit (b). BL write assist circuit (c). Negative BL waveforms for BL write assist (d) (after [61]).

low as shown in Fig. 5.7 (a). Previous designs have used timed pull-downs to V_{SS}. In the SRAMs here, we use charge sharing V_{DDcol} voltage generation (see Fig. 5.7 (b)), as in [71]. The resulting column V_{DD} voltage is capacitively matched to the SRAM columns and thus tracks corners well. However, this scheme adds eight poly tracks to the top and bottom of the column sense/write circuits.

This area is recovered by using a negative BL write assist that provides the write driver a low supply of less than V_{SS}. The same charge sharing circuit is used, with minor polarity changes, to drive a negative BL as shown in see Fig. 5.7 (c). We see excellent margin improvement with -150 mV on the low driven BL (see Fig. 5.7 (d)). The CMOS BL multiplexer is unchanged and passes the negative voltage well. Leakage increases at lower voltages. For either scheme, sufficient capacitance is provided by eight SRAM width columns, which are integrated into the dummy cells at the left and right of each array. The circuit occupies the same columns in the read/write/IO circuit height.

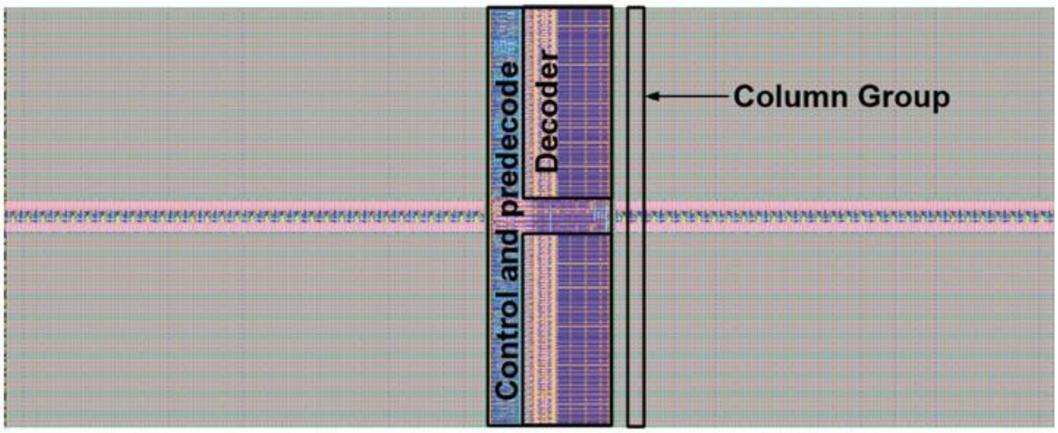


Fig. 5.8. 8 kB array with 128 cells per BL (64-bit words, 84% area efficiency) array example layout (after [61]).

5.6 Array Energy, Performance, and Area

The 8 kB (64k-bit) array (layout shown in Fig. 5.8) occupies 2267 µm^2 with X and Y dimensions of 74.8 and 30.4 µm, respectively. Array efficiency (total area)/(cell area) is 84.2%. For comparison, arrays with 32 cells per BL have array efficiencies of 67% and 64% for 64 and 32-bit words, respectively. For the negative BL write assist, the 128 cell/BL array efficiency improves to 86 %.

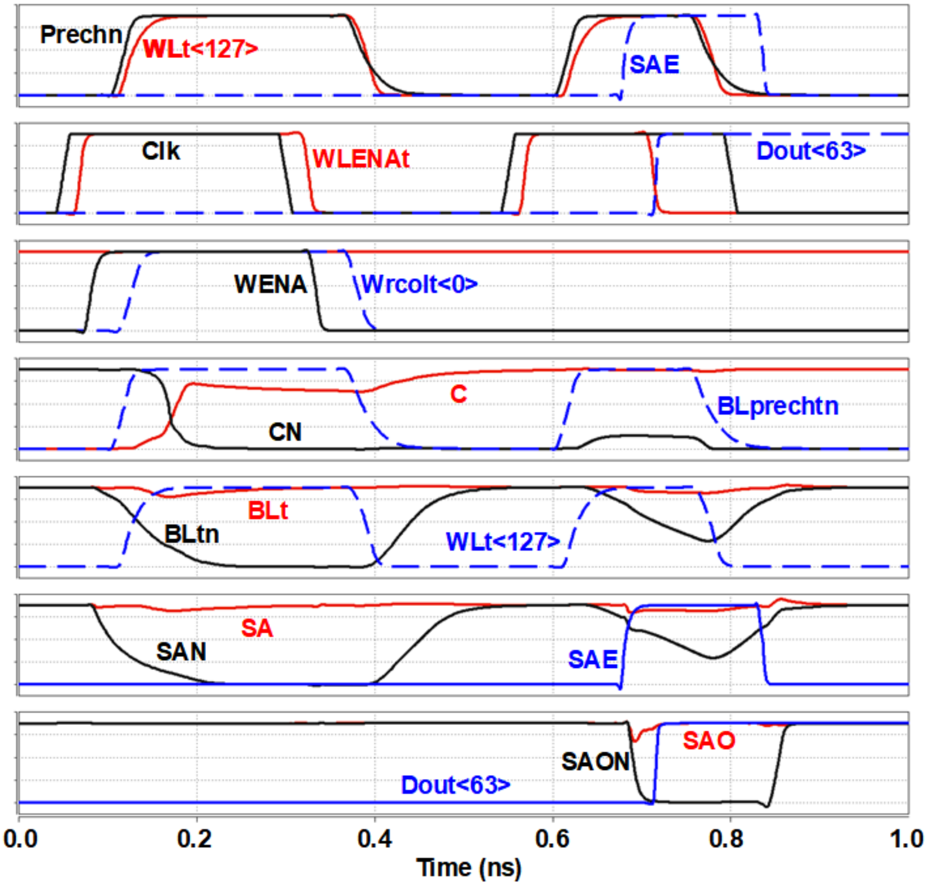


Fig. 5.9. PEX extracted write and read operation at TT corner at 2 GHz clock rate. This array has 128 cells per BL (after [61]).

At the typical (TT) process corner 2 GHz operation is obtained with considerable timing slack (see Fig. 5.9). The address setup time dominates the inputs, requiring 11.5 ps

t_{SETUP} to the block clock input. WLENA is 18.6 ps after the clock. The WLENA to WL delay is 57 ps and generation of a BL differential of 115 mV requires 55 ps. The sense, latch and output drive take 35.5 ps from SAE. Consequently, the total access time at the TT corner and nominal VDD = 0.7 V, defined as the address $t_{SETUP} + t_{CLK2Q}$ of the data, is 166 ps. The 8kB array energy per operation, again at the TT corner and nominal voltage, is 0.25 pJ/op and 0.23 pJ/op for write and read cycles, respectively.

6 TRANSISTOR SIMULATIONS FOR FINFET AND GATE ALL-AROUND FETS FOR ASAP7 AND ASAP5

In this chapter, bulk-Si finFET, horizontal gate-all-around (GAA) nanowire and nanosheet field-effect transistors are compared for 7 nm and 5 nm technology nodes. The performance of these transistors is assessed through 3-D technology computer-aided design (TCAD) simulations. The BSIM SRAM transistor compact model for the ASAP5 PDK, which corresponds to a 5 nm technology node, is created by calibrating to the aforementioned TCAD simulations. The RVT, LVT, and SLVT ASAP5 transistors compact models are derived through V_t adjustment to yet another compact model calibrated to TCAD simulations, as outlined in Section 6.4.

FinFETs have successfully enabled continued CMOS scaling past the conventional planar MOSFETs by providing the required current drive at a higher current density, while reducing drain induced barrier lowering (DIBL), threshold voltage (V_t) roll-off, subthreshold leakage, i.e. improved overall electrostatics through gate control of the channel on three sides [1], [72]–[75]. The undoped or fully depleted fins mitigate random dopant fluctuation (RDF) [45][76], and consequently, the V_t roll-off [73]. Foundry publications have indicated finFET use at the 7 nm technology node (N7) [44][75][76]. However, their continued use at the 5 nm node (N5) is uncertain due to the degradation in their electrostatic performance at shorter gate lengths (L_g) [77]. The idea of using gate-all-around field-effect transistors (GAAFETs), such as nanowire (NW) FET and nanosheet (NSH) FET, for continued scaling has been explored by some for their better electrostatic characteristics as compared to the finFET, owing to an even greater gate control of channel due to the wrap-around gate [78][79].

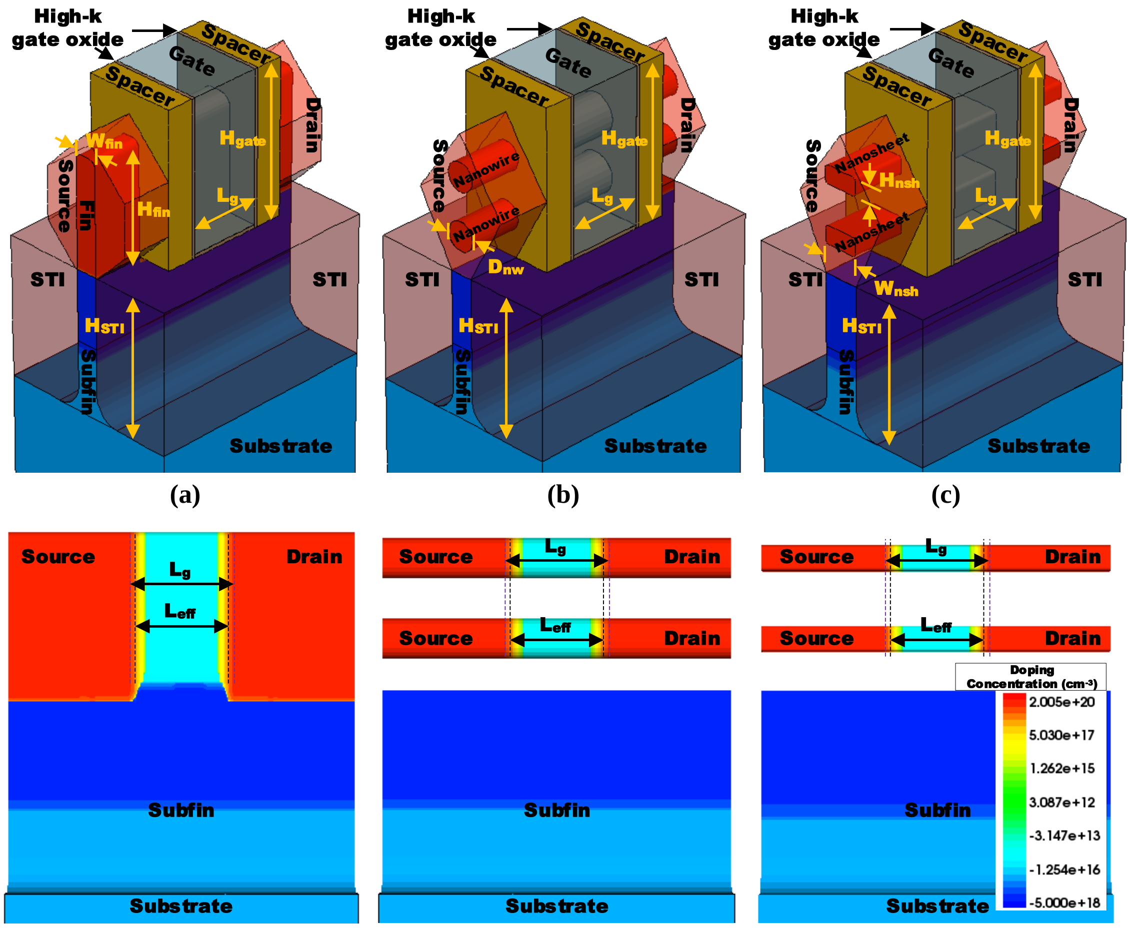
The ASAP7 BSIM compact models were based on assumptions consistent with scaling trends but were not based on TCAD results. For ASAP5, we pursued investigations involving 3-D TCAD device simulations to create BSIM compact models. The transistor parameters were adjusted to be congruous with our initial ASAP7 compact models and geometries. The structure and parameter assumptions for the aforementioned finFET simulations served as a baseline for TCAD simulations employing horizontal NWFETs and NSHFETs. This allowed us to further ascertain the extendibility of finFETs to N5 and also determine the GAAFET characteristics at both N7 and N5.

6.1 FinFET and GAAFET Design Parameters

Synopsys Sentaurus Device [80] is used to simulate the finFET and NWFET structures for L_g ranging from 23 nm to 13 nm. To include carrier mobility effects, the extended Canali high field velocity saturation model [81] and the thin layer mobility model [82] together with the inversion and accumulation layer mobility model are employed [83]. The latter comprehends both the Philips unified mobility model [84] and the enhanced Lombardi models [85]. Recombination models used include the Shockley-Reed-Hall model [86], [87], the Auger model [88], and Hurkx band-to-band tunneling models [89]. Models are also included to account for strain effects on band structure deformation [90], carrier mobility, conduction band and valence band density of states [91], [92]. Other models used include the Slotboom band gap narrowing model [93] and density gradient quantization model [94].

6.2 FinFET Design

The simulated finFET structure dimensions are the same as those assumed in the ASAP7 finFET compact models and front-end-of-the-line stack specifications [11]. A non-aggressive 21 nm L_g had allowed for low SS (< 63.4 mV/dec) and low DIBL (< 22.6 mV/V) in the aforementioned models. Subsequent publications from TSMC and Alliance employed an aggressive 15 nm L_g at N7, but suffered from poor DIBL of 40 mV/V and 89 mV/V, respectively [45], [95]. Thus, the initial L_g assumption is maintained for the baseline finFET structure used in the simulations. Fig. 6.1(a) shows the simulated finFET



(a) (b) (c)

(d) (e) (f)

Fig. 6.1. Simulated (a) n-finFET, (b) n-NWFET (NW=2), and (c) n-NSHFET (NSH=2) 3-D views with 21 nm L_g. Cross section depicting doping concentration in the (d) n-finFET, (e) n-NWFET (NW=2), and (f) n-NSHFET (NSH=2) along the fin length.

structure 3-D view.

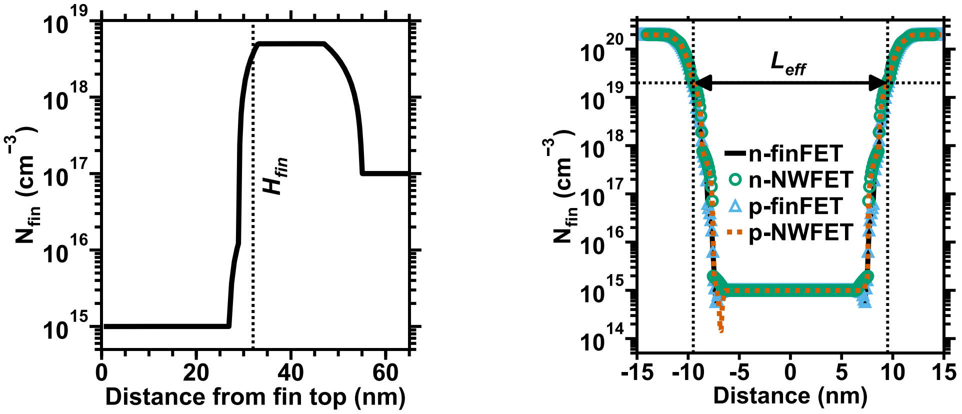
The high-k gate oxide is a 1 nm thick HfO_2 layer, and a 0.6 nm native SiO_2 layer separates it from the Si fin. This results in a 0.78 nm equivalent oxide thickness (EOT), which is in accordance with 0.8 nm EOT for N7 [96]. The gate trench sides are coated with a 1 nm thick HfO_2 layer during its deposition at the trench bottom, prior to the replacement metal gate (RMG) insertion. The sidewall high-k gate layer abuts 8 nm wide Si_3N_4 gate sidewall spacers that are used for source/drain (S/D) self-alignment. Consequently, for the transistors with 54 nm CGP and 21 nm L_g, the S/D epi regions span 15 nm along the fin length. Table 6.1 summarizes doping concentrations and other relevant design parameter for finFET, NWFET, and NSHFET with a 21 nm L_g.

	n-finFET	p-finFET	n-NWFET	p-NWFET	n-NSHFET	p-NSHFET
L_g (nm)	21	21	21	21	21	21
H_{fin} (nm)	32	32	38	38	38	38
$W_{fin}/D_{nw}/[W_{nsh}/H_{nsh}]$(nm)	6.5	6.5	8	8	10/5	10/5
W_{eff} (nm)	70.5	70.5	50.25	50.25	56.56	56.56
H_{STI}/H_{subfin}	40	40	40	40	40	40
H_{gate}	44	44	44	44	44	44
EOT (nm)	0.78	0.78	0.78	0.78	0.78	0.78
N_{sub} (cm^{-3})	1×10^{17}	1×10^{17}	1×10^{17}	1×10^{17}	1×10^{17}	1×10^{17}
N_{subfin} (cm^{-3})	5×10^{18}	5×10^{18}	5×10^{18}	5×10^{18}	5×10^{18}	5×10^{18}
N_{fin} (cm^{-3})	1×10^{15}	1×10^{15}	1×10^{15}	1×10^{15}	1×10^{15}	1×10^{15}
$N_{S/D}$ (cm^{-3})	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}
S/D Doping Gradient (nm/dec)	1	1	1	1	1	1

Table 6.1 FinFET, NWFET, and NSHFET Nominal Design Parameters.

Fig. 6.1(d) shows the doping concentration for various regions within the simulated NMOS finFET. To avoid sub-surface punch-through, the subfin region close to the fin is highly doped (5×10^{18} cm^{-3}). The fin has a low doping concentration (1×10^{15} cm^{-3}) to alleviate RDF [97]. This, as evident in Fig. 6.2(a), results in a steep retrograde doping profile. The peak S/D region doping concentration is 2×10^{20} cm^{-3}, which diminishes with a 1 nm/dec gradient from S/D to the channel.

The effective gate length (L_{eff}) is defined as the gate length demarcated by a 2×10^{19} cm^{-3} peak doping concentration, similar to Zhang et al. [98]. Thus, the L_{eff} is 19 nm for a 21 nm L_g so that the 1 nm fin regions, which underlay the metal gate edges on its either sides, constitute the gate drain overlap. Fig. 6.2(b) shows the net doping concentration along the fin length. Contacts to the S/D epi regions are assumed to be tungsten. Uniaxial 1.6 GPa tensile and 2 GPa compressive stress are added to the NMOS and PMOS, respectively, to increase ON-state currents in the linear region (I_{dlin}) and saturation region (I_{dsat}).



(a) **(b)**

Fig. 6.2. (a) Doping concentration profile along the fin depth in the n-finFET. (b) Doping concentration profile along the fin/NW length in n-finFET, n-NWFET, p-finFET, p-NWFET. L_g *is 21 nm.*

6.3 GAAFET Design

The NWFET and NSHFET structures simulated with TCAD are similar to the simulated finFET to facilitate like-for-like comparison for a given L_g and L_{eff}. Fig. 6.1 (b) and Fig. 6.1 (c) show the simulated n-NWFET and n-NSHFET structure 3-D views, respectively. As evident, the fin pitch would be increased for the NSHFET, but the same

24 nm pitch can support both the finFET and NWFET. Doping concentration for the various device regions in the n-NWFET and n-NSHFET is evident from Fig. 6.1 (e) and Fig. 6.1 (f), respectively. Both GAAFET variants have the same high-k gate oxide, gate sidewall spacer, and S/D dimensions as well as the same shallow trench isolation (STI) and subfin heights as the finFET. Stress and doping concentrations are also kept the same—the latter is evident in Fig. 6.2 (b), wherein the NSHFET doping concentration profile has been precluded for brevity.

Diameter (nm)	7	8
I_{dsat} (µA/fin)	37.26	43.50
I_{off} (pA/fin)	20.16	19.80
V_t (V)	0.204	0.207
DIBL (mV/V)	9.23	10.76
SS (mV/dec)	60.176	61.095

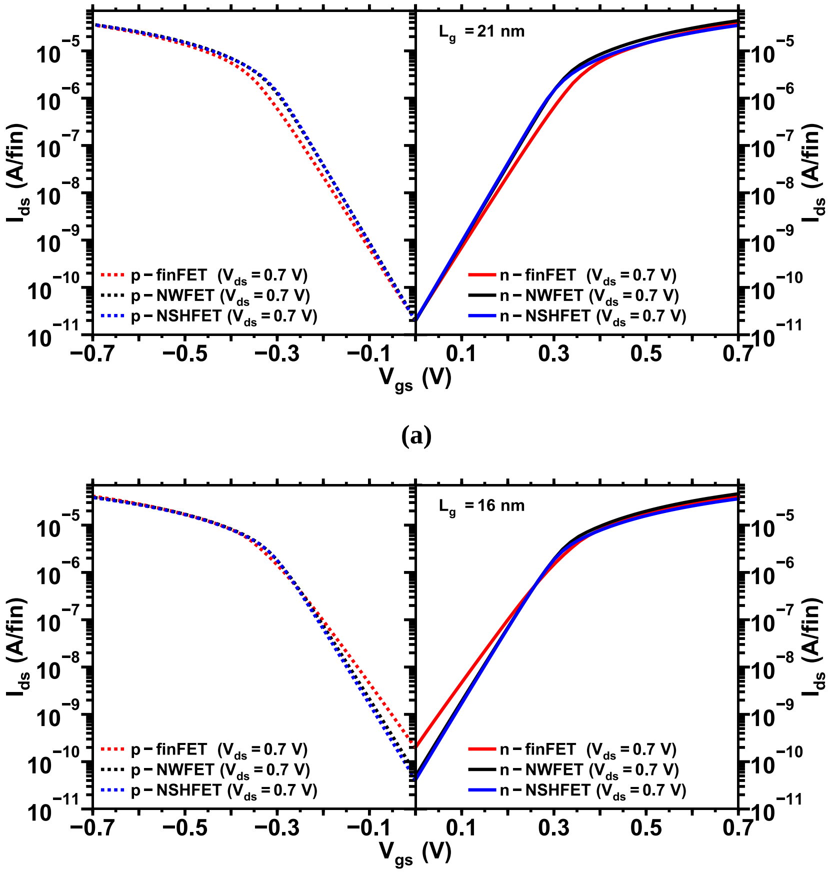
Table 6.2 Nanowire diameter effect on the RVT n-NWFET simulated at 21 nm L_g and 19 nm L_{eff} at 300 K.

NWFETs and NSHFETs with two vertically stacked horizontal NWs and NSHs—constituting a single fin—are simulated. NW being analogous to a fin in a standard cell context, must have a diameter (D_{nw}) that is comparable to W_{fin} in order to keep in line with the scaling trends. We chose an 8 nm D_{nw} instead of 7 nm for the NWFET due to the larger I_{dsat} at comparable, I_{off}, subthreshold slope (SS) and DIBL for the former diameter (see Table 6.2), despite some estimates for D_{nw} to be 7 nm at N5 [99], [100]. A 5 nm NSH height is used for good gate control of the channel following [100] and the NSH width is assumed to be 10 nm with a 1 nm edge rounding, resulting in a 56.56 nm W_{eff}.

6.4 FinFET and Nanowire GAAFET Performance Comparison

A 0.7 V nominal supply voltage (V_{DD}) is used for the simulations, in line with our initial assumptions and is kept constant for varying L_g values [11]. The metal gate workfunction (WF) is adjusted in the TCAD simulations for both n- and p-finFET with 21 nm L_g, in order to match the 20 pA/fin OFF-state leakage current (I_{off}) as specified for the ASAP7 regular V_t (RVT) finFET compact model [11] with the same L_g. The same I_{off} specification is also used to determine the WF for NWFET and NSHFET, both NMOS and PMOS, with 21 nm L_g. Instead of tuning the WF for individual L_g for transistor performance optimization, the WF values thus ascertained are kept constant for varying L_g to enable like-for-like comparisons. Given this approach, all simulated FETs, except for those with 21 nm L_g, do not correspond to our regular V_t (RVT) specification ($I_{off} \approx$ 20 pA/fin) and instead have considerably larger I_{dsat} and I_{off} for L_g smaller than 21 nm.

Fig. 6.3 (a) and (b) show the simulated (at 300 K) I_{ds} vs. V_{gs} characteristics of the n-type and p-type finFET, NWFET, and NSHFET at 21 nm and 16 nm gate length. The former L_g corresponds to the ASAP7 specification while the latter to the ASAP5, which is in line with other L_g estimate for N5 [78]. These L_g specifications also correspond to 19 nm and 14 nm L_{eff}, respectively. Despite the same I_{off}, the poor finFET subthreshold slope



(a)

(b)

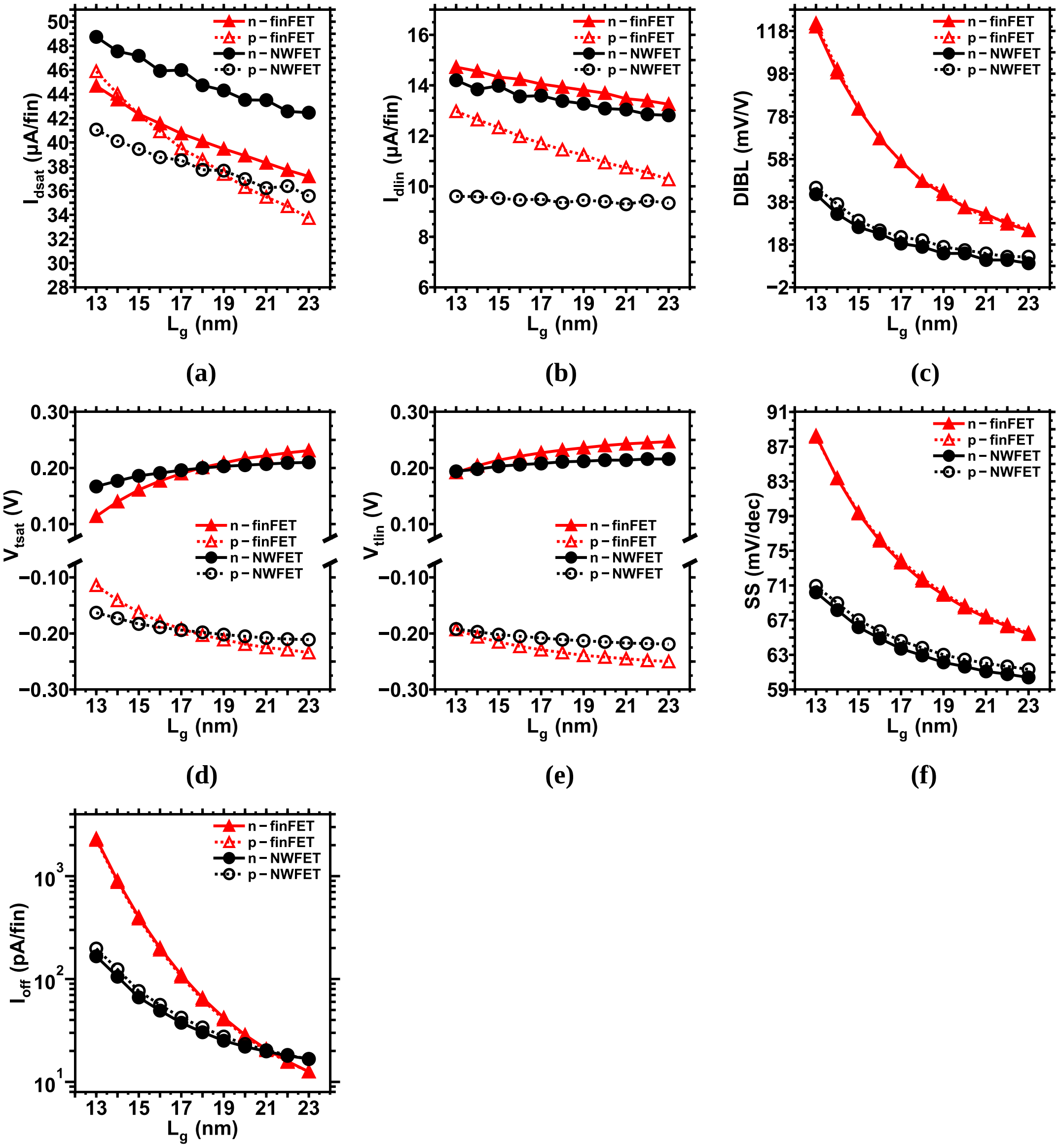
Fig. 6.3. Simulated Ids-Vgs characteristics for the finFETs, NWFETs, and NSHFETs at (a) 21 nm L_g, and (b) 16 nm L_g. The simulated NWFETs and NSHFETs comprise two nanowires/nanosheets per fin and all simulations are at 300 K.

(SS) as compared to the GAAFETs is evident from Fig. 6.3(a). The degradation in finFET electrostatics with L_g scaling to N5 specification becomes more apparent in Fig. 6.3(b) from further deteriorated SS and an I_{off} that is 4-5× larger than that for the GAAFETs.

Fig. 6.4 compares finFET and NWFET I_{dsat}, I_{dlin}, threshold voltage in the saturation (V_{tsat}) and linear (V_{tlin}) regions, I_{off}, SS, and DIBL. The same characteristics for NWFET and NSHFET are plotted on Fig. 6.5 for better legibility.

Fig. 6.4 (a) shows that while the n-NWFET provides substantially larger I_{dsat} across all L_g as compared to n-finFET, the p-NWFET and p-NSHFET lag behind the p-finFET in the sub-19 nm L_g range encompassing N5 and N3 gate lengths. Heretofore, p-finFETs with larger I_{dsat} than n-finFETs have been discussed [101], and we observe a similar trend that is evident in Fig. 6.4 (a) for L_g < 15 nm. However, the n-NWFET is consistently stronger than the p-NWFET across the entire L_g range and the NMOS-to-PMOS I_{dsat} ratio at 16 nm L_g is approximately 1:0.84. The larger I_{dsat} for the n-NWFET, as compared to the p-NWFET, carries over to the high V_t devices—classified as SRAM V_t—simulated with TCAD for maximized I_{on}/I_{off} ratio through WF tuning (see Fig. 6.6 (a)). Consequently, NWFET use at N5, instead of finFETs with potentially stronger PMOS than NMOS, obviates the departure from established circuit design approaches such as changing the SRAM access gates from NMOS pass gates to transmission gates as proposed by Jeong et al. [101].

For DIBL computation, a 50 mV V_{ds} is used for the linear operation region. As evident in Fig. 6.4 (a), finFETs exhibit 2-3× larger DIBL than the NWFETs. Indeed, the 23.08 mV/V and 24.62 mV/V DIBL for n-NWFET and p-NWFET, respectively, at the



(a) (b) (c)

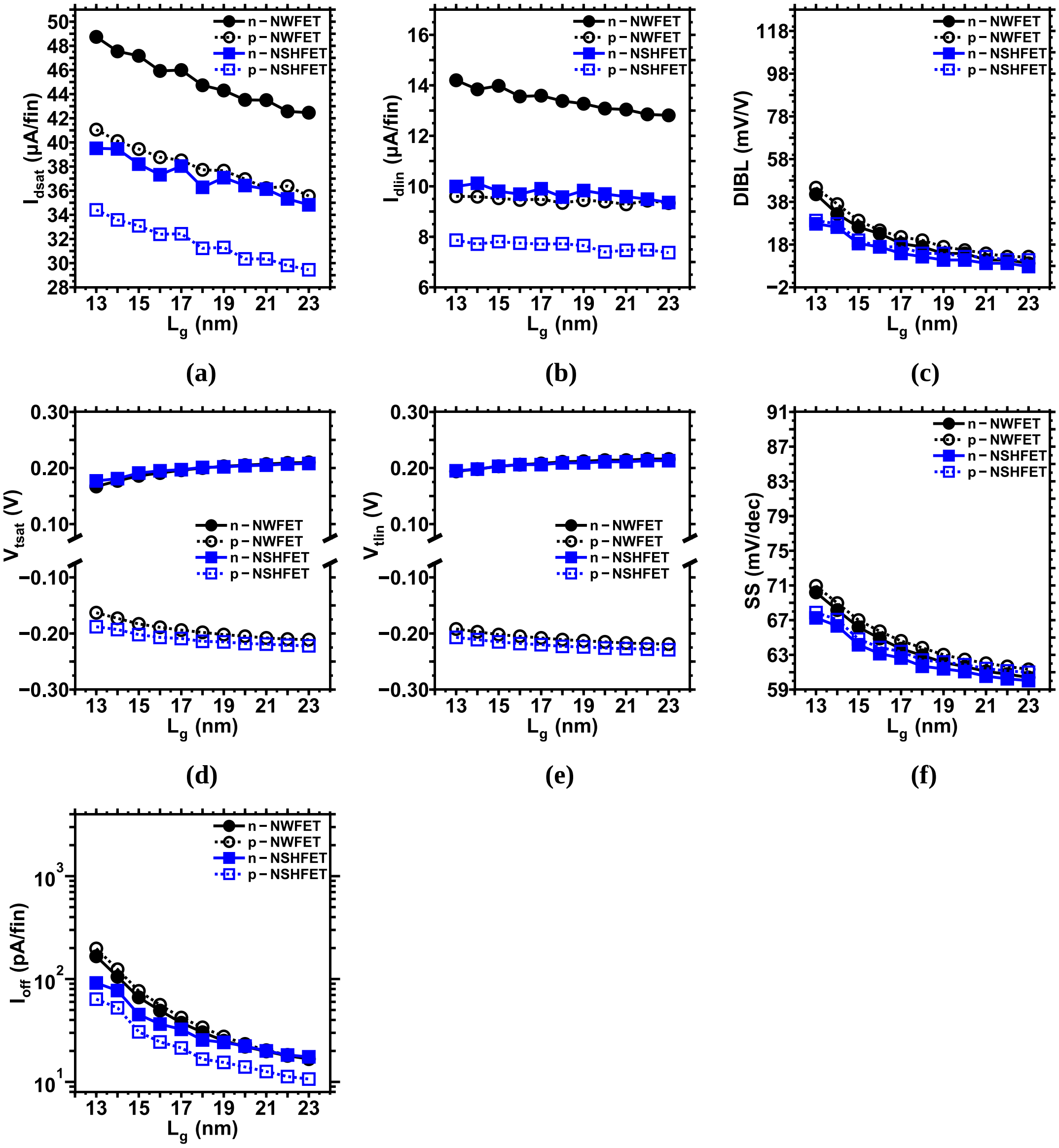
(d) (e) (f)

(g)

Fig. 6.4. Simulated L_g variation effect on (a) I_{dsat}, (b) I_{dlin}, (c) DIBL, (d) V_{tsat}, (e) V_{tlin}, (f) SS, and (g) I_{off} for the n-type and p-type finFET and NWFET. The simulated NWFETs comprise two nanowires per fin and all simulations are at 300 K.

ASAP5 L_g is even smaller than that of 32.31 mV/V and 30.77 mV/V for n-NWFET and p-NWFET, respectively, at the ASAP7 L_g.

The lower GAAFET DIBL substantially alleviates the V_t roll-off as compared to the



(a) (b) (c)

(d) (e) (f)

(g)

Fig. 6.5. Simulated L_g variation effect on (a) I_{dsat}, (b) I_{dlin}, (c) DIBL, (d) V_{tsat}, (e) V_{tlin}, (f) SS, and (g) I_{off} for the n-type and p-type NWFET and NSHFET. The simulated NWFETs and NSHFETs comprise two nanowires/nanosheets per fin and all simulations are at 300 K.

finFETs (see Fig. 6.4 and Fig. 6.5 (d) and (e)). The threshold voltage is ascertained by computing the gate voltage at a 50 nA constant drain current.

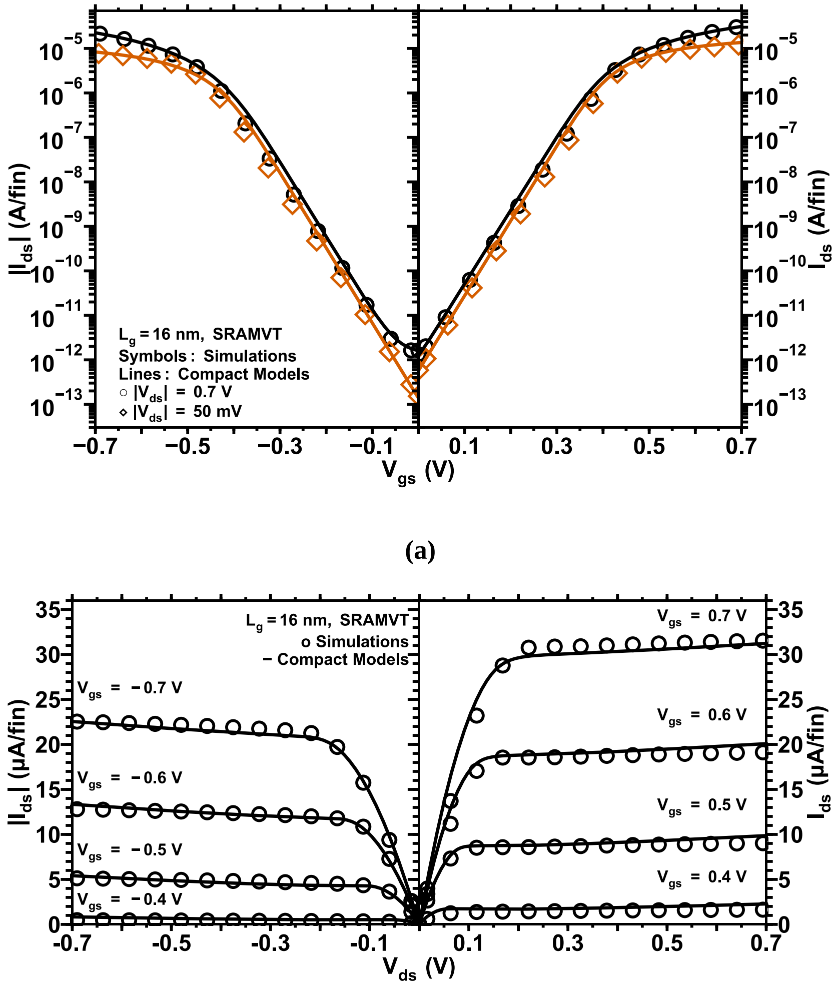
GAAFETs show well behaved SS increment with L_g reduction (see Fig. 6.4 (f) and Fig. 6.5 (f)), with the SS increase from ASAP7 to ASAP5 being 6.21% and 5.94% for the n-NWFET and p-NWFET, respectively. The same increase in the SS is 13.16% for both n-finFET and p-finFET, resulting in a nearly 1.16-1.17× larger SS at the ASAP5 Lg specification as compared to the NWFETs.

The V_t roll-off mitigation in the GAAFETs results in a more gradual I_{off} increase with L_g decrease as compared to the finFETs, which is evident in Fig. 6.4 (g) and Fig. 6.5 (g). At the ASAP5 target L_g of 16 nm, the finFET I_{off} is 4.06× and 3.43× that of the NWFET for NMOS and PMOS, respectively. This indicates towards the NWFET suitability to replace finFETs at N5 to enable high drive strength while maintaining or even improving the leakage performance.

Although the NSHFETs demonstrate better electrostatic and leakage performance as compared to the NWFETs, they provide lower drive current than the latter, thus prompting us to choose NWFETs for ASAP5.

6.5 ASAP5 BSIM Compact Model Calibration and Performance

BSIM compact models are calibrated to the results from the n- and p-NWFET TCAD simulations at 16 nm L_g. The metal gate WF is changed in these compact models



(a)

(b)

Fig. 6.6. (a) I_{ds}-V_{gs} and (b) I_{ds}-V_{ds} characteristic of the n-NWFET and p-NWFET SRAM compact models and the simulations to which they were calibrated at 16 nm L_g and 14 nm L_{eff}. The simulated NWFETs comprise two nanowires per fin and all simulations are at 300 K.

to derive the ASAP5 BSIM compact models wherein each V_t device type has the same I_{off} per micrometer as that for the corresponding ASAP7 V_t device type. This procedure is used to derive the regular threshold voltage (RVT), low threshold voltage (LVT), and super-low threshold voltage (SLVT) ASAP5 BSIM compact models for both the n-NWFET and p-NWFET. However, separate TCAD simulations are used for calibrating the NWFET SRAM BSIM compact models to ensure the device characteristic accuracy, since memories constitute significant leakage component in an IC. The metal gate WF in the TCAD simulations for these devices is tuned for the lowest possible I_{off} while

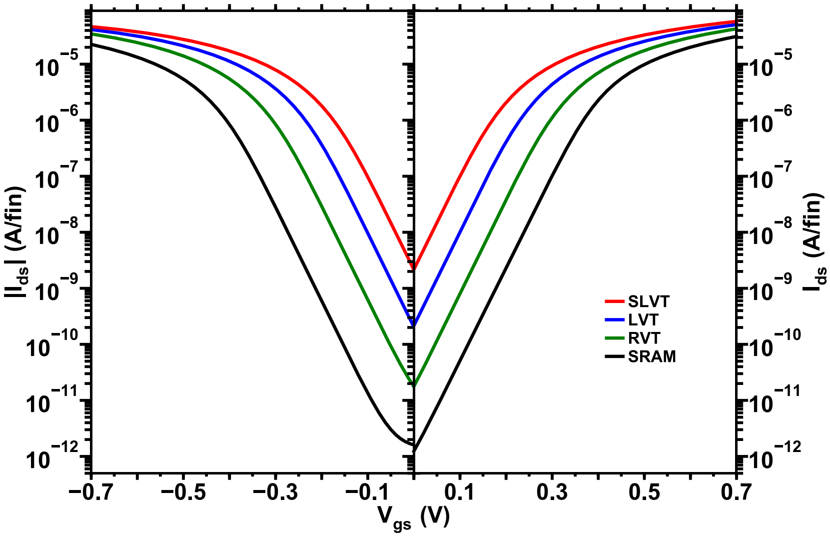


Fig. 6.7. ASAP5 n-NWFET and p-NWFET BSIM compact model I_{ds}-V_{gs} characteristics for different threshold voltage devices. The simulated NWFETs comprise two nanowires per fin and all simulations are at 25 °C.

maximizing the I_{on}/I_{off} ratio. Both the I_{ds} versus V_{gs} and I_{ds} versus V_{ds} characteristics from the TCAD simulations and the corresponding BSIM compact model fits are shown in Fig. 6.6.

The SPICE simulation results at the typical corner for the various ASAP5 BSIM compact models, viz. SRAM, RVT, LVT, and SLVT, are shown in Fig. 6.7. The relevant parameters for the ASAP5 NMOS and PMOS transistors at the typical corner are shown in Table 6.3 and Table 6.4, respectively. It is observed that the p-NWFET I_{dsat} is nearly 80% that of the corresponding n-NWFET for all, except the SRAM devices. For the latter, the p-NWFET I_{dsat} is approximately 72% that of the n-NWFET.

Parameter	SRAM	RVT	LVT	SLVT
I_{dsat} (µA)	31.22	42.70	50.87	57.91
I_{off} (pA)	1.24	17.76	215.67	2172.8
V_{tsat} (V)	0.280	0.206	0.140	0.081
V_{lin} (V)	0.293	0.221	0.155	0.096
DIBL (mV/V)	18.74	22.80	22.80	22.80
SS (mV/dec)	61.32	60.86	60.86	60.89

Table 6.3 ASAP5 n-NWFET (L_g = 16 nm, L_{eff} = 14 nm) typical corner parameter (per fin) at 25 °C.

The I_{off} definition for ASAP5 is consistent with ASAP7 in that it reduces by an order of magnitude for each V_t device type change from SLVT to SRAM in the ascending order of threshold voltage.

Parameter	SRAM	RVT	LVT	SLVT
$\lvert I_{dsat}\rvert$ (µA)	22.54	34.65	41.47	46.71
$\lvert I_{off}\rvert$ (pA)	1.60	17.72	204.31	2143.8
V_{tsat} (V)	-0.315	-0.214	-0.143	-0.082
V_{lin} (V)	-0.330	-0.230	-0.159	-0.098
DIBL (mV/V)	22.55	24.53	24.53	24.53
SS (mV/dec)	61.40	61.48	61.50	61.50

Table 6.4 ASAP5 p-NWFET (L_g = 16 nm, L_{eff} = 14 nm) typical corner parameter (per fin) at 25 °C.

7 ASAP5 INTERCONNECT STACK AND CELL ARCHITECTURE

7.1 Front End of Line (FEOL) and Middle of Line (MOL) Layers

As described in Chapter 6, we chose NWFETs over finFETs and NSHFETs for the ASAP5 PDK. Table 7.1 summarizes the salient width/pitch and lithography assumptions for ASAP5. Fin pitch scaled by 0.7× from N22 to N14 [102][74] and by 0.8× from N14 to N10 [51]. Assuming a 0.88× fin pitch scaling factor with respect to ASAP7 in order to conform with the slowdown in scaling trend, nanowires (NWs) are considered to be patterned at a 24 nm pitch and an 8 nm diameter using SAQP. The active layer is drawn following the same convention as ASAP7. However, unlike ASAP7, the actual active layer in ASAP5 does not extend half-way underneath the gate but rather terminates at the gate edge. This is done to be in line with a recent foundry publication demonstrating single diffusion break (SDB) [51], implying that the fins are cut at the gate edge, i.e., self-aligned by the spacers..

Layer	Width (nm)	Pitch (nm)	Lithography
Gate	16	40	193i SADP
Fin	8	24	193i SAQP
COG-COSD	13	40	EUV SE (2-D, 0.55 NA)
M1-M5	13	26	EUV SE (2-D, 0.55 NA)
V1-V5	13	26	EUV SE (2-D, 0.55 NA)
M6-M9	20	40	193i SADP
V6-V9	20	40	193i SADP
M10-M12	40	80	193i SE
V10-V12	40	80	193i SE

Table 7.1 Width, pitch, and lithography assumptions for ASAP5 layers.

Although foundry publications have demonstrated 48 nm CGP at N7 [45] and 54 nm CGP at N10 [51], the latter being the same as that for ASAP7, the CGP scaling is likely to slow down at N5, with some projecting a 42 nm value [78]. However, we assume a more aggressive 40 nm CGP since this allows for pitch upscaling in cell layouts, which is easier than pitch downscaling, in the event of a subsequent change in pitch assumption. The gates are assumed to be patterned at this pitch using SADP at a uniform spacing and have a 16 nm gate length (L_g), as outlined in Chapter 6.

Fig. 7.1 and Fig. 7.2 show the ASAP5 FEOL and MOL cross sections. Unlike in ASAP7, the SDT is mandatory over the raised source-drain (RSD) regions and does not

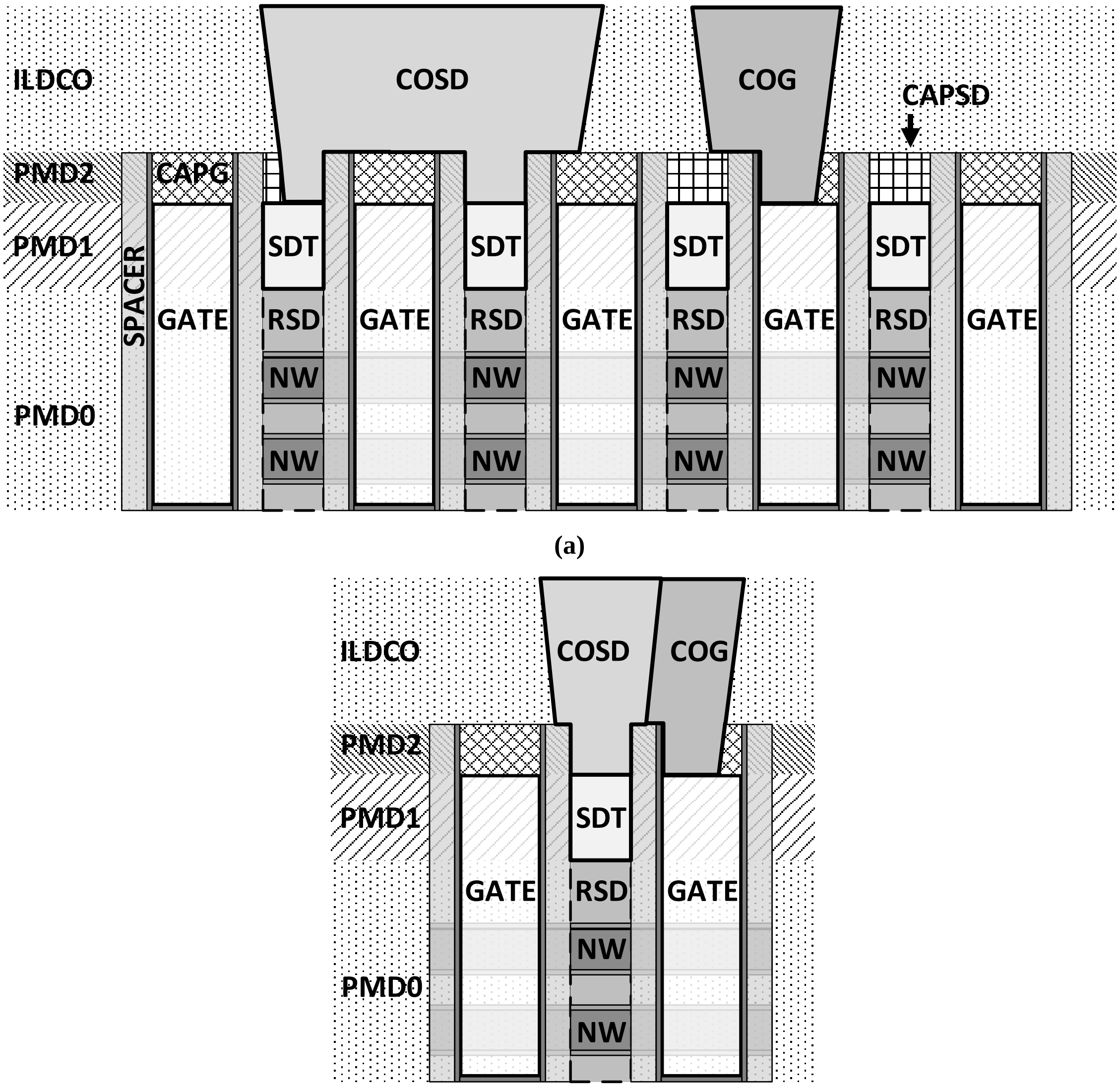


Fig. 7.1. ASAP5 MOL and FEOL stack cross sections. (a) COSD to SDT and COG to gate connections. Both COSD and COG can cross over the Gate and SDT, respectively. (b) COSD and COG can be shorted in cells us as the SRAM to advantage.

cut through the gate sidewall spacers. ASAP5 forgoes the ASAP7 LISD and LIG local interconnect layers and instead assumes two contact layers, namely contact over source-drain (COSD) and contact over active gate (COG). COSD connects SDT to M1, whereas COG provides contact between M1 and gate over active device area and is one of the key features enabling density improvement that has been demonstrated in a foundry publication [51]. The process for COG and COSD patterning and deposition is assumed after that demonstrated by Pethe et al. [103] and is as follows. First, COSD pattern is transferred atop the dielectric layer ILDCO (see Fig. 7.1), followed by ILDCO and trench cap—CAPSD in Fig. 7.1—layer selective etching to create opening for contacts to the SDT. The gate spacers and gate cap—CAPG in Fig. 7.1—remain unetched as a consequence of the selective etch action and due to the gate cap (CAPG) material being different than the trench cap (CAPSD) material. Etching in this manner has the advantage of enabling COSD routing over gates (not shown). Furthermore, by having a 10 nm tall CAPG and a 5 nm wide gate sidewall spacer, the selective etching prevents TDDB (assuming a 5 nm TDDB spacing) despite the COSD mask being misalignment, thus alleviating any spacing requirements between these layers. Note that the COSD fill material is not yet deposited. After this, the same process is repeated with COG patterns to etch CAPG to create openings for contacts to the gate, with the spacer and CAPSD remaining unetched due to selective etching which also prevents TDDB case despite COG mask misalignment. It is only after this step that the fill material, which is same for COG and COSD, is deposited. The steps outlined for COG and COSD are interchangeable.

Both COSD and COG layers are drawn to horizontally overlap the gate and spacer (empty space between SDT and gate in cell layouts), respectively, by 3 nm. This corresponds to the EUV edge placement error (EPE) and ensures maximization of the COSD/COG fill contact area over SDT/gate, i.e. fully COSD/COG landing over SDT/gate. Fig. 7.1 (a) shows an exaggerated case where the COSD left edge and COG right edge are not drawn in the aforementioned manner so that they are not completely landed in the case of misalignment. Correspondingly, the other COSD and COG edges overlap CAPG and CAPSD, respectively, but do not contact to the underlying layers, which prevents a TDDB scenario despite the misalignment. Fig. 7.1 (b) shows the manner in which COG and COSD can be shorted. This structure is used to our advantage in the SRAM cell.

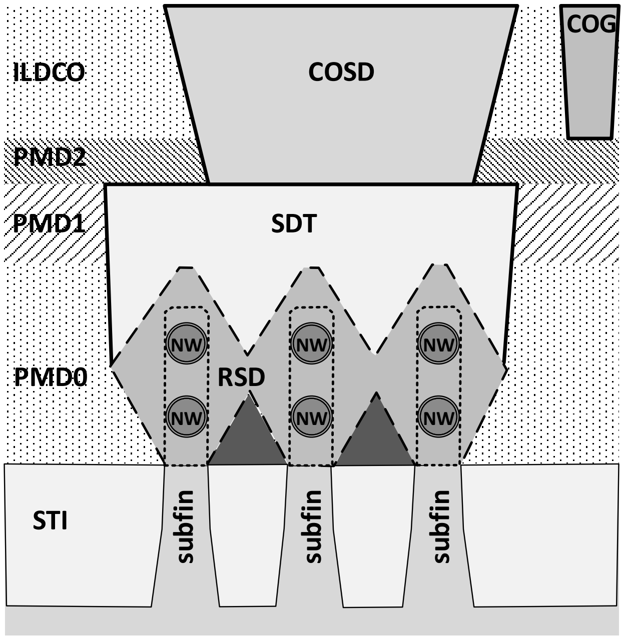


Fig. 7.2. NW and raised S/D cross section showing SDT and COSD connection. Ensuring a vertical spacing between SDT and COG eases the SDT-to-COG spacing.

In ASAP7, a V0 is required to de-couple the LIG and LISD layers from M1 to permit LISD routing crossing M1 tracks and M1 routing crossing LIG that connect multiple gates in the middle of the cell. In ASAP5, limiting the LISD use for routing and moving the COG over the active regions obviates a V0 layer, thereby simplifying both the process and design rules. Fig. 7.2 shows the manner in which keeping the trench cap intact with respect to the COG layer ensures a vertical separation between the two layers, easing the SDT-to-COG spacing requirement, as opposed to the SDT-to-LIG spacing requirement shown in Fig. 3.1. The 13 nm COG/COSD CD allows for connectivity to the M1 layer that is patterned at the same CD.

7.2 Back End of Line (BEOL) Layers

The ASAP5 PDK assumes 12 interconnect metal layers (M1-M12) for routing purpose and corresponding vias (V1-V12) to establish connection between these metals. ASAP5 vias are fully self-aligned (FSAV) and differ from the self-aligned vias (SAV) used for ASAP7. A fully self-aligned Vx is not merely aligned with respect to Mx+1, which was the case in ASAP7, but also with respect to Mx. The FSAV process follows after [104]–[106]. As shown in Fig. 3.2 (top view, right), an ASAP7-like process that assumes partially self-aligned vias still suffers from via spreading along the Mx+1 length. This creates a TDDB scenario and can even create shorts to the neighboring Mx—adjacent to the Mx to which Vx must connect—in the event of excessive misalignments. Furthermore, in the case when the Vx connects perpendicular Mx and Mx+1, a Vx mask that is only drawn at a larger CD perpendicular to Mx+1 can only tolerate the misalignment in the same direction and not perpendicular to the Mx direction. This results in Vx not landing completely on Mx when misalignment occurs, thereby increasing the via resistance due to a smaller contact area. A FSAV alleviates the TDDB and shorting scenarios to neighboring Mx by ensuring sufficient spacing and also maximizes the contact area by ensuring full Vx landing on Mx. This is achieved in the following manner. Alignment of FSAV Vx to Mx is enabled by recessing Mx with respect to the dielectric layer surrounding Mx. This raised ILD works in much the same manner as the gate/SDT do in our MOL process and even in the case of a Vx misalignment perpendicular to the Mx length, the height difference between the raised ILD and Mx ensures sufficient Vx to Mx separation. This alleviates the Vx to neighboring Mx TDDB and shorting scenario and in turn allows a wider Vx mask CD perpendicular to the Mx length that does not engender these scenarios. The latter ensures full Vx landing on Mx in case of a misalignment perpendicular to Mx, thereby maximizing Vx-to-Mx contact area. As the Vx CD being printed is larger, it also reduces the litho variability. In a FSAV process, the Vx mask shape is derived by extending the Vx visible layer edges corresponding to the Mx and Mx+1 layers and when the Vx connects perpendicular Mx and Mx+1, the Vx mask shape can be a square if the extension amount is equal.

ASAP7 supported M1-M3 layers were assumed to be patterned EUVL with a 0.33 NA source, which enabled 2-D shapes on these layers that facilitated easier classroom use. We follow the same approach for ASAP5 and to enable lower metal pitch scaling as well as 2-D shapes, we assume the M1-M5 layers to be patterned using EUVL with a 0.55 high-NA source [107]–[109] at a 13 nm half-pitch [41][47]. The V1-V5 are also assumed to be patterned using the same technique instead of EUVL with a 0.33 NA source since NILS reaches the minimum usable value of two at 18 nm half-pitch [110]. As we require vias at a minimum spacing of 13 nm, this necessitates employing EUVL patterning with a 0.55 NA source as it can print dense contact/via hole patterns at an 11 nm half pitch with a NILS value greater than two [108]. Furthermore, even though EUVL LELELE can reach the 11 nm half-pitches with a 0.33 NA source, the throughput, i.e. wafers per hour, is greatly reduced as compared to a single EUVL exposure with a 0.55 NA source [111]. M6-M9 and V6-V9 are assumed to be patterned using 193i SADP whereas M10-M12 and V10-V12 are assumed to be patterned using 193i SE.

7.3 Interconnect Parasitic Extraction

As outlined in Section 3.3, the choice of barrier layer and metal fill play a crucial role in managing the increase in interconnect resistance due to the highly resistive unscaled barrier layers and the non-linear metal fill resistivity increase. In previously released versions of ASAP7, we incorporated contribution from the barrier resistivity but did not comprehend its effects on the line resistivity due to the changes in barrier sidewall and bottom thickness as a result of changes in the interconnect trench slope angle, trench aspect ratio, and trench CD. Furthermore, while we computed the resistivity change with CD for copper, we did not do so for the tungsten MOL layers. For the newer version of ASAP7 and ASAP5, we used a more detailed interconnect parasitic resistance analysis that incorporates the effects of fill and barrier on the line resistivity due to changes in the trench slope angle, barrier sidewall and bottom thickness. We also comprehended the fill and trench pinch-off scenarios when critical CDs are reached. Furthermore, we looked at several difference metal fill and barrier configurations to determine the ASAP5 interconnect stack.

Publication for a foundry N10, with similar minimum metal pitch targets as ASAP7, described replacing copper in the interconnect stack with cobalt to reduce the via resistance by 2× reduction and also bring about a reduction in electromigration [51]. Thus, we included copper, tungsten, and cobalt as fill metal candidates and ruthenium (Ru), titanium nitride (TiN), and tantalum nitride (TaN) as barrier layer candidates in our analysis.

Although cobalt can be deposited directly on a dielectric without a barrier layer, its adhesion to the latter is weak and causes delamination [112]. A 1 nm thin layer of TiN has been found to prevent Co delamination and diffusion into the dielectric [112]. We compute the copper resistivity using the classic Fuchs-Sondheimer, Mayadas-Shatzkes (FS-MS) model after Pyzyna et al. [58] as well as its extension (FS-MS extension) after Hu et al. [113]. The cobalt resistivity is determined through the FS-MS extension Hu et al. [113] and the tungsten resistivity by using the classic FS-MS after Steinhogl et al. [114].

The trends for line resistivity as a function of the line/trench CD are shown in Fig. 7.3. This includes contributions from both the metal fill and barrier layers. Fig. 7.4 shows

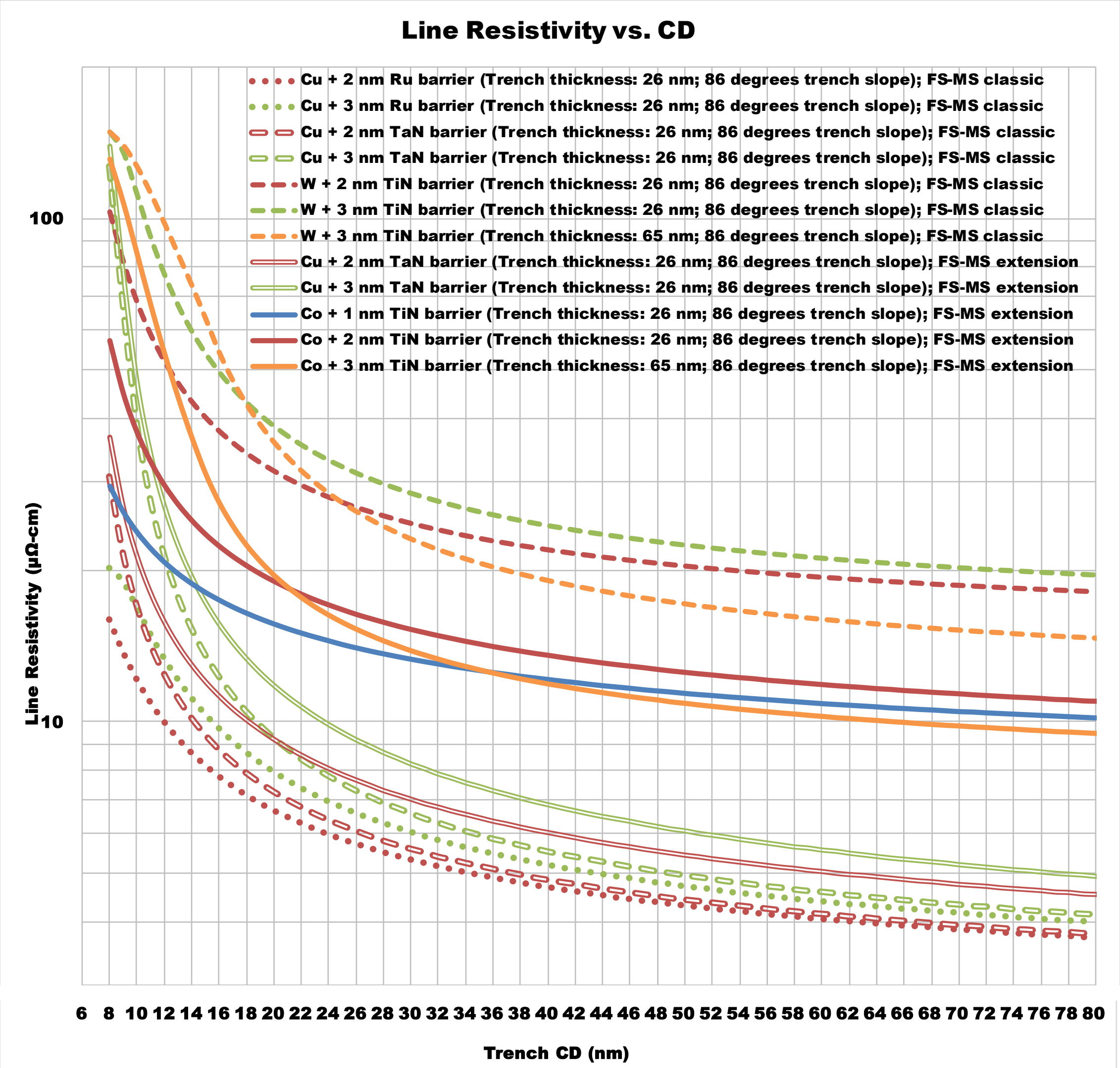


Fig. 7.3. Line resistivity versus CD.

change in metal fill resistivity as a function of trench CD. Note that in the latter, the fill does not occupy the entire trench but rather the area not occupied by the barrier, thus accurately tracking the effect of barrier sidewall and bottom thickness on the fill resistivity with changes in the trench slope, aspect ratio, and trench CD, even though the barrier resistivity contribution becomes insignificant—apparent in Fig. 7.4 from the overlapping copper fill resistivity values for the same computation method and barrier thickness. The resistivity is plotted for a 2:1 aspect ratio (AR) assuming a 13 nm trench

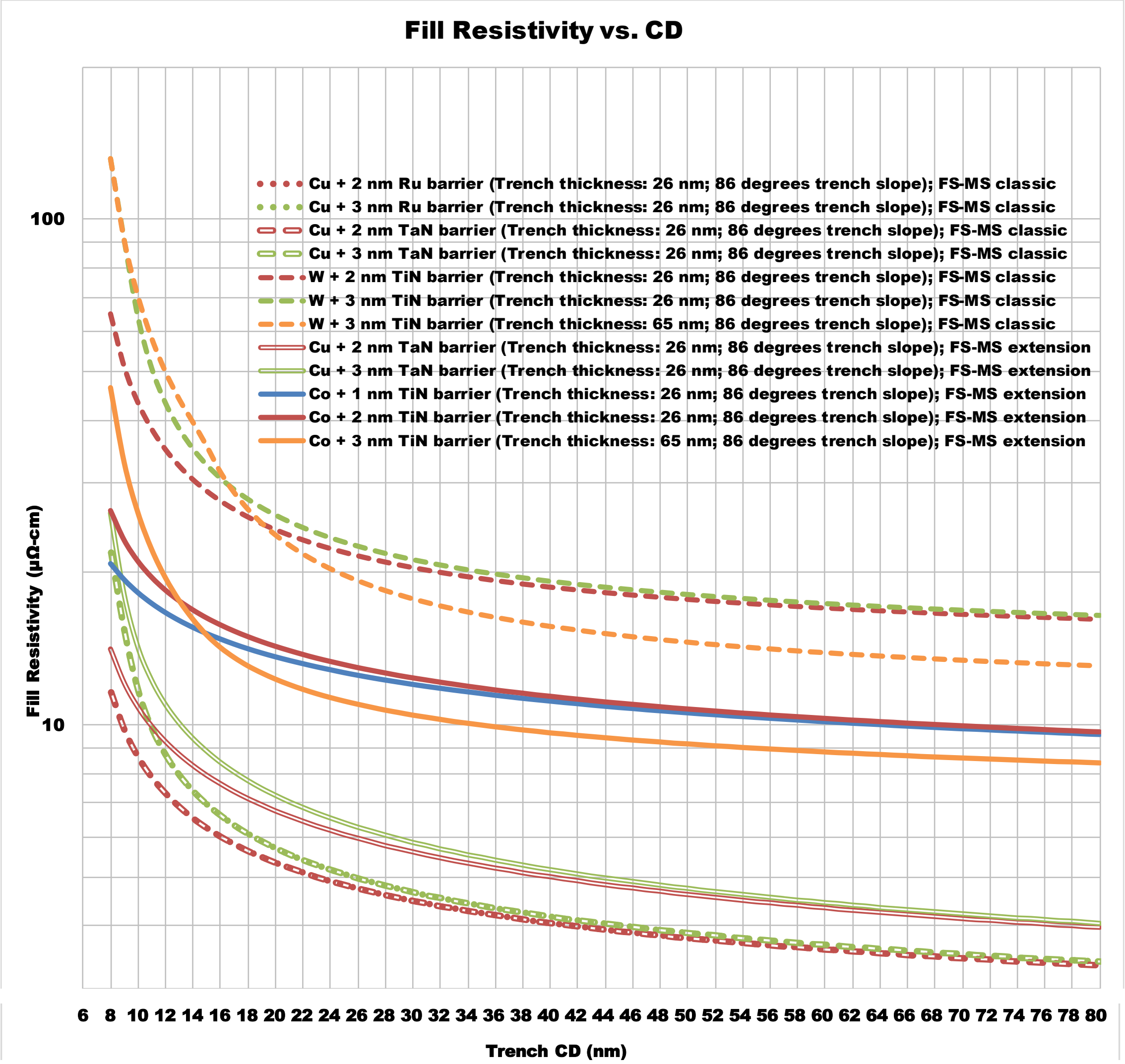


Fig. 7.4. Fill resistivity versus CD.

CD corresponding to the ASAP5 minimum metal CD target, as well as a 5:1 AR for the same trench CD. The former results in a 26 nm trench thickness and the latter in a 69 nm trench thickness.

Even though copper has a smaller bulk resistivity than cobalt [113], for the copper and cobalt resistivities calculated using FS-MS extension, it is evident from Fig. 7.4 that the rate of increase in copper fill resistivity is greater than that in cobalt fill resistivity with trench CD shrink. When the barrier resistivity contribution is incorporated in the analysis as well, the need for thicker and more resistive barriers use with copper, as compared to cobalt, results in larger line resistivity for lines/trenches containing copper fill with TaN barrier than that for lines containing cobalt fill and TiN barrier, which is evident in Fig. 7.3.

The contribution of barrier resistivity alone to the line resistivity as a function of CD is plotted in Fig. 7.5. Negative values indicate a larger fill resistivity than the line resistivity, which results from the barrier alleviating the high fill material resistivity by helping in current conduction. The larger contribution from the highly resistive TaN as

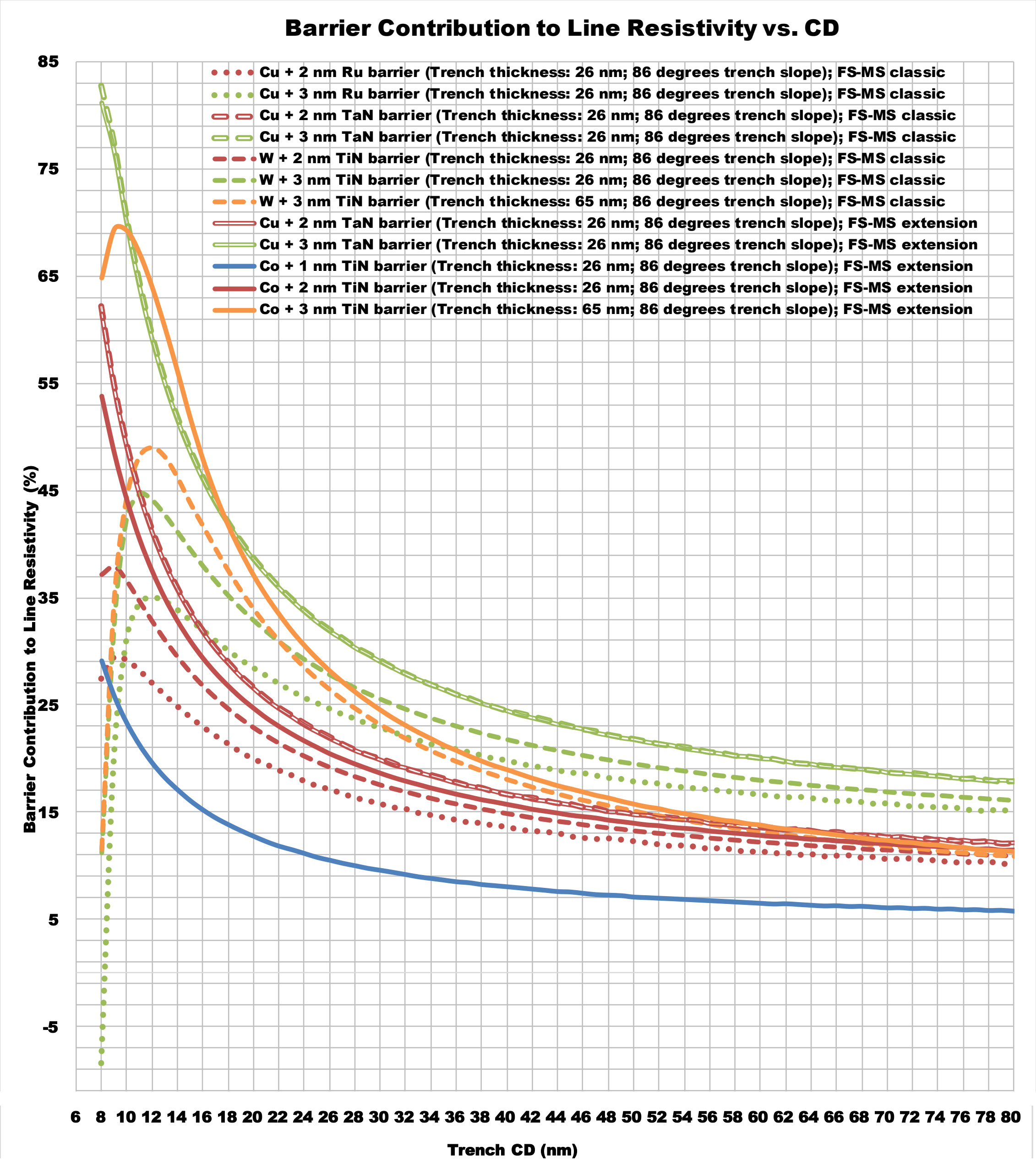


Fig. 7.5. Barrier contribution to line resistivity versus CD.

compared to TiN is apparent. Although Ru is a good candidate for TaN replacement as a barrier layer for copper, which is evident by the large resistivity increase different between Ru and TaN barriers at small trench CDs and thicker barriers (see Fig. 7.3), its adaption for this purpose in foundry N7 and N5 is as yet unclear. However, we still assumed Ru barrier for the interconnects in ASAP7, which is optimistic.

The 5:1 high AR wires have lower resistivities than the 2:1 AR wires at CDs larger than 36 nm. However, the resistivities for high AR wires starts to increase at smaller CDs, eventually exceeding that of a wire with 2:1 AR at a 18 nm CD for tungsten with a 3 nm TiN barrier, as seen in Fig. 7.3, at 14 nm CD for copper with a 3 nm TaN barrier, and also at 14 nm CD for cobalt with a 3 nm TiN barrier, all of these at an 86° trench slope. The last two cases have been excluded from Fig. 7.3 to improve legibility. Thus, both metals and vias have a 2:1 aspect ratio in ASAP5. The ASAP5 MOL layers, viz. COG and COSD, have a 13 nm CD and a 3:1 aspect ratio, resulting in a 39 nm COG and COSD trench depth. This is useful as it allows for 10 nm (including an additional 5 nm guard-banding) gate/SDT cap for TDDB prevention.

Even though the tungsten fill resistivity is exceedingly high (see Fig. 7.3), heretofore, MOL layers in the interconnect stack have used tungsten fill since it enables good step coverage and gap-fill even in highly scaled MOL interconnects and contacts with large aspect ratios [115]. We had also assumed tungsten fill in ASAP7 following this convention. However, the lower cobalt fill resistivity combined with the convenience of a thinner barrier layer that suffices to prevent diffusion into the dielectric and FEOL layers [112]. Thus, we assume cobalt fill for the ASAP5 SDT, COSD, and COG layers instead of tungsten fill. Auth et al. [51] have demonstrated that replacing tungsten with cobalt reduces the contact resistance by up to 60% in a foundry process.

From the BEOL perspective, where copper and cobalt are two of the metal fill candidates, cross over points favoring cobalt metal fill arise between a realistic process employing cobalt fill with a 1 nm TiN barrier and a realistic process using copper fill with a 3 nm TaN barrier at 15 nm trench CD and at 9 nm trench CD for an optimistic process employing copper fill with a 2 nm barrier layer, as seen in Fig. 7.3. Note that in our analysis both 83° and 86° trench slopes were used and considering an 83° trench slope with 3 nm barrier layer, the fill width at the trench bottom is 0 nm, i.e. fill pinch-off, at and below 14 nm trench CD. The pinch-off with an 86° slope and a 3 nm thick barrier layer occurs at and below 9 nm trench CD. This implies that a better process that uses barriers smaller than or equal to 3 nm and preferably with a steeper trench angle will be required for an N5 process with minimum metal CD below 15 nm. Since a realistic process involving copper fill with a 3 nm thick TaN barrier layer is more resistive than a realistic cobalt with a 1 nm TiN barrier layer, we chose the latter for M1-M5 and the corresponding V1-V5, since the routing on these layers will frequently be at the minimum metal/via width of 13 nm. Although a more optimistic process involving copper fill with a 2 nm thick TaN barrier layer is less resistive than a realistic cobalt with a 1 nm TiN barrier layer, we still assume cobalt fill for lower electromigration after Auth et al. [51] who demonstrated that replacing copper with cobalt resulted in a 5-10× improvement in electromigration in a foundry process. For M6-M12 and corresponding V6-V12 in ASAP5, copper fill with a 1 nm ruthenium barrier is used for line resistivity calculations following our ASAP7 assumptions, which is once again optimistic.

7.4 Cell Library Architecture

For the ASAP5 cell library, we assumed a 6.5-track cell (see Fig. 7.6) library that allows for double CD, i.e. twice the minimum metal width, M2 power rails that ensure robust power delivery. A taller cell library than this greatly increases the power density

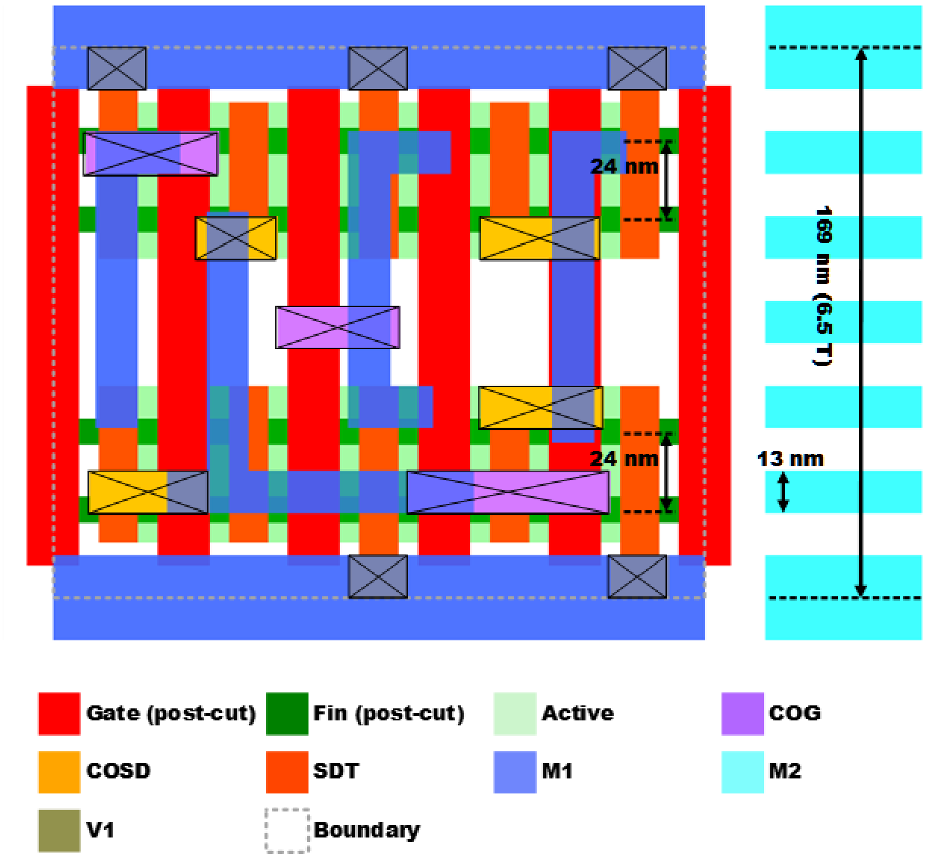


Fig. 7.6. A 6.5 M2 track ASAP5 standard cell, with 2-fins per device type. The cell layout shows the FEOL, MOL, and M1 layers in an AND2 cell. Unlike ASAP7, a single diffusions break is required between the cells.

while a shorter one, such as a 5.5 or 5 track one, greatly reduces the number of cell internal routing track which adversely affect the cell library richness—as described in Chapter 4. A 6.5 track cell library allows for two fin wide transistors for both device types and has five M2 routing tracks. The COG and COSD contact layers fall on the same M2 routing track grid, which simplifies cell design by permitting a template like approach. The cells end on a single diffusion break (SDB) which further maximizes cell density.

The connectivity between the M1 power rail and source-drain (S/D) regions in the ASAP5 standard cells is established using the SDT-COSD stack, as evident in Fig. 7.6. Connecting to the power rails using this scheme means that unlike ASAP7, where LIG on the power rails created a spacing requirement to the gate tips in the standard cells, no such COSD-to-gate spacing requirement exists in ASAP5 at the standard cell power rails. This is because both the SDT and COSD are guaranteed to not create TDDB scenarios with the gate layer, since the SDT is self-aligned with respect to the low-k gate sidewall spacer. This ensures sufficient spacing from the gate, whereas the COSD does not short to the gate due to the gate cap that also ensures sufficient gate-to-COSD vertical separation for TDDB alleviation. Also, since the COSD connecting to the M1 power rail is only

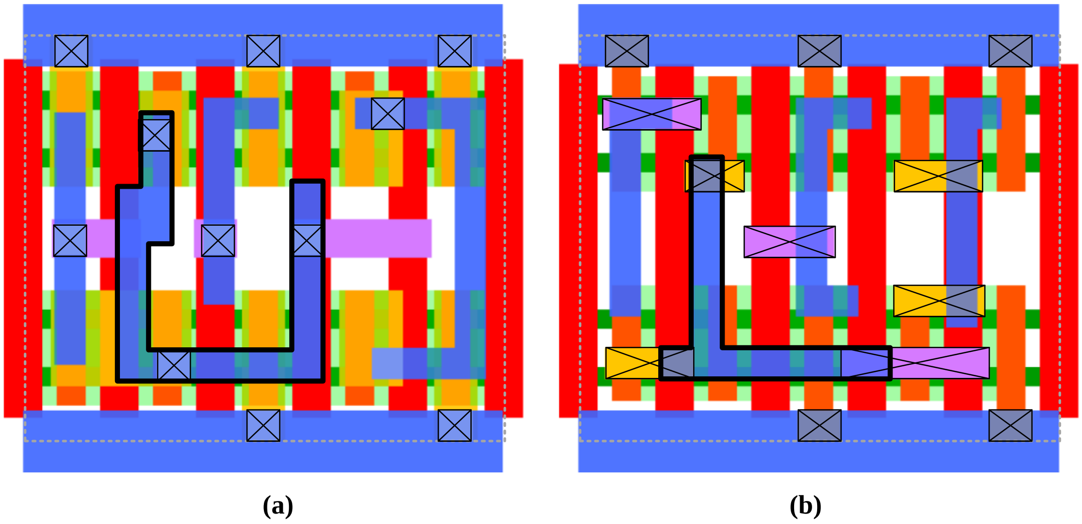


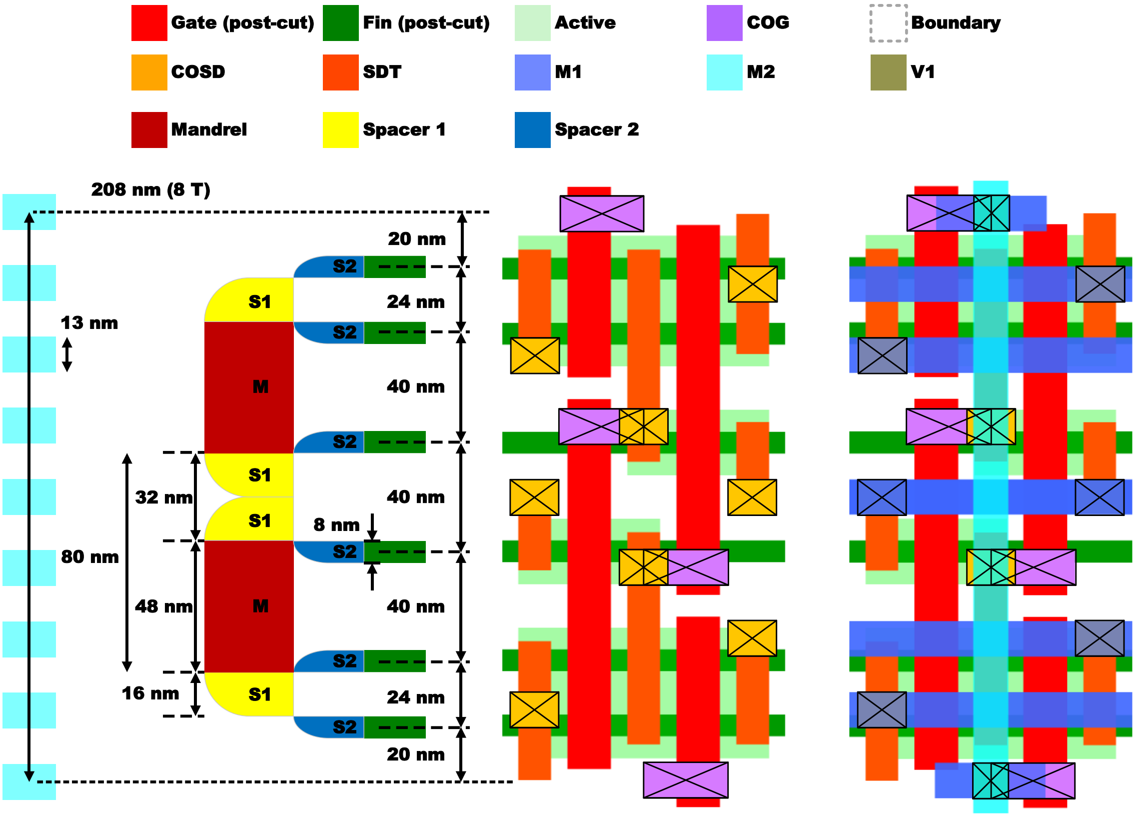
Fig. 7.7. (a) A 6.5-track AND2 cell, designed with an older ASAP5 version with an MOL stack (LIG, LISD, and V0) similar to ASAP7. (b) AND2 cell with the ASAP5 COG and COSD stack. Emboldened M1 routes in both cells illustrate the routing simplification in the latter cell when contact over active gate layer (COG) is used.

drawn to half the M1 power rail width, it is always separated from the non-equipotential SDT in the vertically adjoined cell by a spacing that is equal to single M1 CD and SDT-to-M1 spacing at the cell edge.

Fig. 7.7 (a) shows 6.5 track cells designed for ASAP5 with an ASAP7-like MOL stack without gate contact over the active region and Fig. 7.7 (b) shows the same cell with the new ASAP5 MOL stack without V0 and with a gate contact over the active region. It can be seen that using the COG layer over the active results in greatly simplified routing on an internal node. This can also help to maximize M2 access to M1 pins in certain cases. The route is also shorter, which results in lower line resistance. The elimination of V0 also means that in some cases access to M1 pins cannot be increased. However, the penalty is a single track and is typically limited to just a few pins per cell, which can be offset to some extent by the pin access maximization due to the routing simplification from COG use.

7.5 SRAM Cell Architecture

Fig. 7.8 shows a 122 SRAM cell designed for ASAP5. The fins are pattered at a non-uniform pitch using SAQP, as shown in Fig. 7.8 (a). The NMOS transistors are at a 24 nm pitch and the PMOS transistors are at a 40 nm pitch to each other as well as the NMOS transistors. SAQP results in eight fins in a typical case, but the fin patterning at the center of the SRAM is assumed to be suppressed by abutting the two first level spacers, as shown in Fig. 7.8 (a). This approach is not robust as a gap between the spacers can result in the creation of an undesired fin/nanowire. Another approach involves



(a) (b) (c)

Fig. 7.8. (a) Nanowire (NW) patterning scheme for ASAP5 122 SRAM cell. The first level spacers are merged to preclude NW formation. 122 SRAM cell layout (b) without M1 and M2 for clarity and (c) with M1 and M2.

improving the robustness by precluding the undesired fin formation through overlapping of the spacers, such that the spacing between the mandrels is 24 nm. However, this results in 216 nm tall cell which is not an integer multiple of half M1 CD that ensure equal M1 spacing.

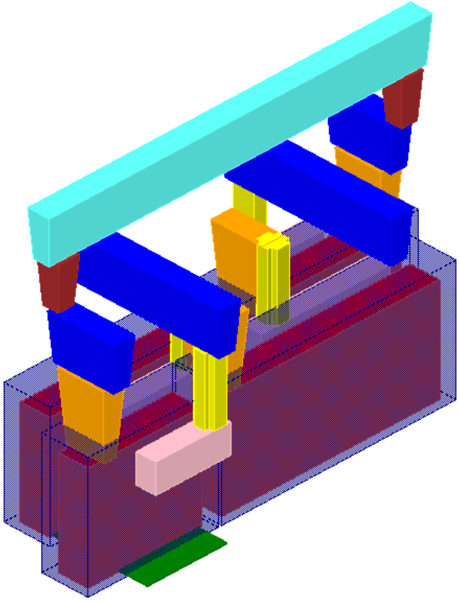


Fig. 7.9. ASAP5 122 SRAM cell 3-D cross section showing the FEOL, MOL and BEOL layers but does now show nanowires. The gate spacer is shown as transparent.

The FEOL and MOL layers for the cell can be seen in Fig. 7.8 (b). The PMOS and NMOS S/D regions in the inverters are shorted using SDT routing, which is then shorted to the other inverter gate by using a structure comprising shorted COG and COSD. Fig. 7.8 (c) shows all the layers in the cell. The BL and power rails are in M1 whereas the WL is in M2. The 122 SRAM cell thus designed is eight tracks tall and is 0.57× the area of a 122 SRAM cell in ASAP7. The LIG-to-SDT spacing that was critical in ASAP7 SRAM cells is greatly eased in this ASAP5 122 SRAM cell by the fact that the COG layer does not short with SDT due to the presence of the aforementioned S/D cap layer.

8 SUMMARY

This dissertation has described the ASAP7 and ASAP5 PDK development. It discussed the ASAP7 and ASAP5 device electrical performance characteristics. The dissertation also provided a basic overview of lithography variability concerns, double patterning approaches, and how they compare with each other. Furthermore, it covered EUV lithography basic and related challenges. PDK details, such as patterning choices for the various layers, cell library architecture, and DTCO considerations for developing the architecture in addition to the SRAM cells were discussed as well. The dissertation described the APR experiments and SRAM designs based on ASAP7, which demonstrated its suitability for research into various VLSI circuit and system design related aspects.

The ASAP7 PDK has been deployed in graduate-level VLSI courses at the Arizona State University since 2015. The PDK is available for free to Universities. It has received attention from research groups and faculty at a number of other Universities, which indicates towards a high likelihood of its adoption in the classroom. Thus, we hope that both ASAP7 and ASAP5 PDKs will fulfill their development intent of enabling sub-10 nm CMOS research in academia at a much larger scale and beyond a few University research groups with access to advanced foundry PDKs.

REFERENCES

[1] S.-Y. Wu *et al.*, “A 16nm FinFET CMOS technology for mobile SoC and computing applications,” in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 9.1.1-9.1.4.

[2] C. Lin *et al.*, “High performance 14nm SOI FinFET CMOS technology with 0.0174µm^2 embedded DRAM and 15 levels of Cu metallization,” in *2014 IEEE International Electron Devices Meeting*, 2014, pp. 3.8.1-3.8.3.

[3] K. Bhanushali and W. R. Davis, “FreePDK15: An Open-Source Predictive Process Design Kit for 15nm FinFET Technology,” in *Proceedings of the 2015 Symposium on International Symposium on Physical Design - ISPD ’15*, 2015, pp. 165–170.

[4] M. Martins *et al.*, “Open Cell Library in 15nm FreePDK Technology,” *Proc. 2015 Symp. Int. Symp. Phys. Des. - ISPD ’15*, pp. 171–178, 2015.

[5] R. Goldman, K. Bartleson, T. Wood, K. Kranen, V. Melikyan and E. Babayan, “32/28nm Educational Design Kit: Capabilities, deployment and future,” in *2013 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*, 2013, pp. 284–288.

[6] J. E. Stine *et al.*, “FreePDK: An Open-Source Variation-Aware Design Kit,” in *2007 IEEE International Conference on Microelectronic Systems Education (MSE’07)*, 2007, pp. 173–174.

[7] R. Aitken, “Physical Design and FinFETs (Slide),” *Int. Symp. Phys. Des.*, pp. 65–68, 2014.

[8] B. Chava *et al.*, “Standard cell design in N7: EUV vs. immersion,” vol. 9427, p. 94270E, 2015.

[9] L. W. Liebmann and R. O. Topaloglu, “Design and Technology Co-Optimization Near Single-Digit Nodes,” pp. 582–585, 2014.

[10] L. Liebmann, A. Chu and P. Gutwin, “The daunting complexity of scaling to 7NM without EUV: pushing DTCO to the extreme,” 2015, vol. 9427, no. February, p. 942702.

[11] L. T. Clark *et al.*, “ASAP7: A 7-nm finFET predictive process design kit,” *Microelectronics J.*, vol. 53, pp. 105–115, Jul. 2016.

[12] V. Vashishtha, M. Vangala and L. T. Clark, “ASAP7 predictive design kit development and cell design technology co-optimization: Invited paper,” in *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2017, vol. 2017-Novem, pp. 992–998.

[13] Y. Ma *et al.*, “Decomposition strategies for self-aligned double patterning,” vol. 7641, pp. 76410T-76410T–13, 2010.

[14] T. Chiou, M. Dusa, A. C. Chen, D. Pietromonaco, R. O. Way and S. Jose, “Lithographic challenges and their solutions for critical layers in sub- 14nm node logic devices,” vol. 8683, no. 59, pp. 1–15, 2013.

[15] R. de Graaf, S. Weichselbaum, R. Droste, M. McLaren, B. Koek and W. de Boeij, “NXT:1980Di immersion scanner for 7nm and 5nm production nodes,” vol. 9780, p. 978011, 2016.

[16] T. Vandeweyer *et al.*, “Immersion Lithography and Double Patterning in Advanced Microelectronics,” vol. 7521, no. 0, pp. 1–11, 2010.

[17] E. van Setten *et al.*, “Imaging performance and challenges of 10nm and 7nm logic nodes with 0.33 NA EUV,” 2014, vol. 9231, p. 923108.

[18] “ASML Twinscan NXE:3400B.” [Online]. Available: http://web.archive.org/web/20170612174252/https://www.asml.com/products/systems/twinscan-nxe/twinscan-nxe3400b/en/s46772?dfp_product_id=10850.

[19] I. Servin *et al.*, “Mask contribution on CD and OVL errors budgets for double patterning lithography,” in *25th European Mask and Lithography Conference*, 2009, vol. 7470, no. 0, p. 747009.

[20] B. J. Lin, “Optical lithography with and without NGL for single-digit nanometer nodes,” vol. 9426, p. 942602, 2015.

[21] “TWINSCAN NXT:1980Di.” [Online]. Available: http://web.archive.org/web/20180828184022/https://www.asml.com/products/systems/twinscan-nxt/en/s46772?dfp_product_id=10567.

[22] “TWINSCAN NXT:1970Ci.” [Online]. Available: https://web.archive.org/web/20170222015809/https://www.asml.com/asml/show.do?lang=EN&ctx=46772&dfp_product_id=8036.

[23] K. T. Yeh and W. A. Loong, “Simulations of mask error enhancement factor in 193 nm immersion lithography,” *Japanese J. Appl. Physics, Part 1 Regul. Pap. Short Notes Rev. Pap.*, vol. 45, no. 4 A, pp. 2481–2496, 2006.

[24] K. Standiford and C. Bürgel, “A new mask linearity specification for EUV masks based on time dependent dielectric breakdown requirements,” vol. 8880, pp. 1–7, 2013.

[25] S. Demuynck *et al.*, “Contact module at dense gate pitch technology challenges,” *2014 IEEE Int. Interconnect Technol. Conf. / Adv. Met. Conf. IITC/AMC 2014*, pp. 307–310, 2014.

[26] T. Ito and S. Okazaki, “Pushing the limits of lithography,” *Nature*, vol. 406, no. 6799, pp. 1027–1031, Aug. 2000.

[27] L. Liebmann, V. Gerousis, P. Gutwin, M. Zhang, G. Han and B. Cline, “Demonstrating production quality multiple exposure patterning aware routing for the 10NM node,” vol. 9053, p. 905309, 2014.

[28] K. Oyama *et al.*, “Sustainability and applicability of spacer-related patterning towards 7nm node,” 2015, vol. 9425, p. 942514.

[29] W. H. Arnold, “Toward 3nm overlay and critical dimension uniformity: an integrated error budget for double patterning lithography,” *Proc. SPIE*, vol. 6924, pp. 692404-692404–9, 2008.

[30] W. Jung *et al.*, “Patterning with amorphous carbon spacer for expanding the resolution limit of current lithography tool,” *Spie*, vol. 6520, no. 2007, pp. 1–9, 2007.

[31] K. Oyama, S. Natori, S. Yamauchi, A. Hara and H. Yaegashi, “CD error budget analysis for self-aligned multiple patterning,” 2012, vol. 8325, p. 832517.

[32] Y. Ma, J. Sweis, H. Yoshida, Y. Wang, J. Kye and H. J. Levinson, “Self-Aligned Double Patterning (SADP) Compliant Design Flow,” vol. 8327, pp. 832706-832706–13, 2012.

[33] B. Wong, F. Zach, V. Moroz, A. Mittal, G. Starr and A. Kahng, *Nano-CMOS Design for Manufacturabililty*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.

[34] V. Vashishtha, A. Dosi, L. Masand and L. T. Clark, “Design technology co-optimization of back end of line design rules for a 7 nm predictive process design kit,” in *2017 18th International Symposium on Quality Electronic Design (ISQED)*, 2017, pp. 149–154.

[35] A. Mallik *et al.*, “The economic impact of EUV lithography on critical process modules,” 2014, p. 90481R.

[36] J. Ryckaert *et al.*, “Design Technology co-optimization for N10,” in *Proceedings of the IEEE 2014 Custom Integrated Circuits Conference*, 2014, pp. 1–8.

[37] K. Vaidyanathan, R. Liu, L. Liebmann, K. Lai, A. J. Strojwas and L. Pileggi, “Design implications of extremely restricted patterning,” *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 13, no. 3, p. 031309, Oct. 2014.

[38] A. Mallik, J. Ryckaert, A. Mercha, D. Verkest, K. Ronse and A. Thean, “Maintaining Moore’s law: enabling cost-friendly dimensional scaling,” 2015, vol. 9422, p. 94221N.

[39] E. van Setten *et al.*, “Patterning options for N7 logic: prospects and challenges for EUV,” vol. 9661, p. 96610G, 2015.

[40] M. van de Kerkhof *et al.*, “Enabling sub-10nm node lithography: presenting the NXE:3400B EUV scanner with improved overlay, imaging, and throughput,” vol. 10143, p. 101430D, 2017.

[41] J. van Schoot, K. van Ingen Schenau, C. Valentin and S. Migura, “EUV lithography scanner for sub-8nm resolution,” in *Proc. SPIE*, 2015, vol. 9422, p. 94221F.

[42] G. Tallents, E. Wagenaars and G. Pert, “Optical lithography: Lithography at EUV wavelengths,” *Nat. Photonics*, vol. 4, no. 12, pp. 809–811, Dec. 2010.

[43] S.-S. Kim *et al.*, “Progress in EUV lithography toward manufacturing,” vol. 10143, p. 1014306, 2017.

[44] D. Ha *et al.*, “Highly manufacturable 7nm FinFET technology featuring EUV lithography for low power and high performance applications,” in *2017 Symposium on VLSI Technology*, 2017, pp. T68–T69.

[45] R. Xie *et al.*, “A 7nm FinFET technology featuring EUV patterning and dual strained high mobility channels,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, vol. 12, no. c, pp. 2.7.1-2.7.4.

[46] G. Dicker, D. de Bruin, B. Peterson, P. Wöltgens, B. Sluijk and P. Jenkins, “Getting ready for EUV in HVM,” vol. 9661, p. 96610F, 2015.

[47] S. M. Y. Sherazi *et al.*, “Architectural strategies in standard-cell design for the 7 nm and beyond technology node,” *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 15, no. 1, p. 013507, Feb. 2016.

[48] K. Seo *et al.*, “A 10nm Platform Technology for Low Power and High Performance Application Featuring FINFET Devices with Multi Workfunction Gate Stack on Bulk and SOI,” pp. 5–6, 2014.

[49] K. Schuegraf, M. C. Abraham, A. Brand, M. Naik and R. Thakur, “Semiconductor logic technology innovation to achieve sub-10 nm manufacturing,” *IEEE J. Electron Devices Soc.*, vol. 1, no. 3, pp. 66–75, 2013.

[50] H. Hody, V. Paraschiv, D. Hellin, T. Vandeweyer, G. Boccardi and K. Xu, “Gate double patterning strategies for 10-nm node FinFET devices,” *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 14, no. 1, p. 014504, Mar. 2015.

[51] C. Auth *et al.*, “A 10nm high performance and low-power CMOS technology featuring 3 rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 29.1.1-29.1.4.

[52] W. Ye, B. Yu, D. Z. O. Pan, Y.-C. Ban and L. Liebmann, “Standard Cell Layout Regularity and Pin Access Optimization Considering Middle-of-Line,” *Proc. 25th Ed. Gt. Lakes Symp. VLSI - GLSVLSI ’15*, pp. 289–294, 2015.

[53] J. T. Neumann *et al.*, “Imaging performance of EUV lithography optics configuration for sub-9nm resolution,” 2015, vol. 9422, p. 94221H.

[54] “ITRS 2.0 (2017).” [Online]. Available: https://web.archive.org/web/20170712035039/http://www.itrs2.net/.

[55] M. R. Baklanov, P. S. Ho and E. Zschech, *Advanced Interconnects for ULSI Technology*. Chichester, UK: John Wiley & Sons, Ltd, 2012.

[56] R. Brain *et al.*, “Low-k interconnect stack with a novel self-aligned via patterning process for 32nm high volume manufacturing,” in *2009 IEEE International Interconnect Technology Conference*, 2009, vol. M, pp. 249–251.

[57] S. Im *et al.*, “Scaling analysis of multilevel interconnect temperatures for high-performance ICs,” *IEEE Trans. Electron Devices*, vol. 52, no. 12, pp. 2710–2719, 2005.

[58] a Pyzyna *et al.*, “Resistivity of copper interconnects beyond the 7 nm node,” in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015, vol. 1, no. 1, pp. T120–T121.

[59] J. Mulkens, M. Hanna, H. Wei, V. Vaenkatesan, H. Megens and D. Slotboom, “Overlay and edge placement control strategies for the 7nm node using EUV and ArF lithography,” *Proc. SPIE*, vol. 9422, p. 94221Q, 2015.

[60] F. G. Pikus, “Decomposition Technologies for Advanced Nodes,” 2016.

[61] V. Vashishtha, M. Vangala, P. Sharma and L. T. Clark, “Robust 7-nm SRAM design on a predictive PDK,” in *Proceedings - IEEE International Symposium on Circuits and Systems*, 2017, pp. 5–8.

[62] M. J. M. Pelgrom, A. C. J. Duinmaijer and A. P. G. Welbers, “Matching properties of MOS transistors,” *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[63] K. J. Kuhn *et al.*, “Process technology variation,” *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, 2011.

[64] Y. Liu, K. Endo and O. Shinichi, "On the Gate-Stack Origin Threshold Voltage Variability in Scaled FinFETs and Multi-FinFETs Background and Purpose FinFET Fabrication by Nano Wet Etching V t Variability in Scaled FinFETs Separation of V t Variation Sources."

[65] T. Matsukawa *et al.*, "Comprehensive analysis of variability sources of FinFET characteristics," *VLSI Technol. 2009 Symp.*, pp. 118–119, 2009.

[66] M. D. Giles *et al.*, "High sigma measurement of random threshold voltage variation in 14nm Logic FinFET technology," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2015, vol. 2015-Augus, no. mV, pp. T150–T151.

[67] L. T. Clark, S. Leshner and G. Tien, "SRAM cell optimization for low AVT transistors," in *International Symposium on Low Power Electronics and Design (ISLPED)*, 2013, vol. 1, pp. 57–63.

[68] S. Sakhare *et al.*, "Layout optimization and trade-off between 193i and EUV-based patterning for SRAM cells to improve performance and process variability at 7nm technology node," 2015, vol. 9427, p. 94270O.

[69] E. Seevinck, F. J. List and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987.

[70] M. Qazi, K. Stawiasz, L. Chang and A. P. Chandrakasan, "A 512kb 8T SRAM Macro Operating Down to 0.57V With an AC-Coupled Sense Amplifier and Embedded Data-Retention-Voltage Sensor in 45nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 85–96, Jan. 2011.

[71] V. Chandra, C. Pietrzyk and R. Aitken, "On the efficacy of write-assist techniques in low voltage nanoscale SRAMs," *Des. Autom. Test Eur. Conf. Exhib. (DATE), 2010*, vol. 1, pp. 345–350, 2010.

[72] C. Auth, "22-nm fully-depleted tri-gate CMOS transistors," in *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*, 2012, pp. 1–6.

[73] C. C. Wu *et al.*, "High performance 22/20nm FinFET CMOS devices with advanced high-K/metal gate scheme," in *2010 International Electron Devices Meeting*, 2010, pp. 27.1.1-27.1.4.

[74] S. Natarajan *et al.*, "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *2014 IEEE International Electron Devices Meeting*, 2014, pp. 3.7.1-3.7.3.

[75] H.-J. Cho *et al.*, "Si FinFET based 10nm technology with multi Vt gate stack for low power and high performance applications," in *2016 IEEE Symposium on VLSI Technology*, 2016, vol. 2016-Septe, pp. 1–2.

[76] T. B. Hook, “Fully depleted devices for designers: FDSOI and FinFETs,” in *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*, 2012, no. Figure 3, pp. 1–7.

[77] P. Raghavan *et al.*, “5nm: Has the time for a device change come?,” in *2016 17th International Symposium on Quality Electronic Design (ISQED)*, 2016, pp. 275–277.

[78] H. Mertens *et al.*, “Vertically stacked gate-all-around Si nanowire transistors: Key Process Optimizations and Ring Oscillator Demonstration,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 37.4.1-37.4.4.

[79] S. Barraud *et al.*, “Performance and design considerations for gate-all-around stacked-NanoWires FETs,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, no. 001, pp. 29.2.1-29.2.4.

[80] “Sentaurus Device User Guide.” Synopsys Inc., Mountain View, California, 2015.

[81] C. Canali, G. Majni, R. Minder and G. Ottaviani, “Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature,” *IEEE Trans. Electron Devices*, vol. 22, no. 11, pp. 1045–1047, Nov. 1975.

[82] S. . Reggiani, E. . Gnani, A. . Gnudi, M. . Rudan and G. . Baccarani, “Low-Field Electron Mobility Model for Ultrathin-Body SOI and Double-Gate MOSFETs With Extremely Small Silicon Thicknesses,” *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2204–2212, Sep. 2007.

[83] Syed Aon Mujtaba, “Advanced Mobility Models For Design And Simulation of Deep Submicrometer MOSFETS,” Stanford University, 1995.

[84] D. B. M. Klaassen, “A unified mobility model for device simulation,” in *International Technical Digest on Electron Devices*, 1990, pp. 357–360.

[85] M. N. Darwish, J. L. Lentz, M. R. Pinto, P. M. Zeitzoff, T. J. Krutsick and Hong Ha Vuong, “An improved electron and hole mobility model for general purpose device simulation,” *IEEE Trans. Electron Devices*, vol. 44, no. 9, pp. 1529–1538, 1997.

[86] D. J. Roulston, N. D. Arora and S. G. Chamberlain, “Modeling and Measurement of Minority-Carrier Lifetime versus Doping in Diffused Layers of n +-p Silicon Diodes,” *IEEE Trans. Electron Devices*, 1982.

[87] J. G. Fossum and D. S. Lee, “A physical model for the dependence of carrier lifetime on doping density in nondegenerate silicon,” *Solid State Electron.*, 1982.

[88] R. Häcker and A. Hangleiter, “Intrinsic upper limits of the carrier lifetime in silicon,” *J. Appl. Phys.*, vol. 75, no. 11, pp. 7570–7572, Jun. 1994.

[89] G. A. M. Hurkx, D. B. M. Klaassen and M. P. G. Knuvers, “A new recombination model for device simulation including tunneling,” *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 331–338, 1992.

[90] G. L. B. G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*. New York: John Wiley & Sons, Inc., 1974.

[91] M. V. Fischetti and S. E. Laux, “Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys,” *J. Appl. Phys.*, vol. 80, no. 4, p. 2234, Jun. 1998.

[92] V. Sverdlov, E. Ungersboeck, H. Kosina and S. Selberherr, “Effects of shear strain on the conduction band in silicon: An efficient two-band k·p theory,” in *ESSDERC 2007 - 37th European Solid State Device Research Conference*, 2007, pp. 386–389.

[93] D. B. M. Klaassen, J. W. Slotboom and H. C. de Graaff, “Unified apparent bandgap narrowing in n- and p-type silicon,” *Solid. State. Electron.*, vol. 35, no. 2, pp. 125–129, Feb. 1992.

[94] M. G. Ancona and H. F. Tiersten, “Macroscopic physics of the silicon inversion layer,” *Phys. Rev. B*, vol. 35, no. 15, pp. 7959–7965, May 1987.

[95] S. Wu *et al.*, “A 7nm CMOS platform technology featuring 4 th generation FinFET transistors with a 0.027μm^2 high density 6-T SRAM cell for mobile SoC applications,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 2.6.1-2.6.4.

[96] V. Moroz, J. Huang and R. Arghavani, “Transistor design for 5nm and beyond: Slowing down electrons to speed up transistors,” in *2016 17th International Symposium on Quality Electronic Design (ISQED)*, 2016, pp. 278–283.

[97] N. Agrawal, Huichu Liu, R. Arghavani, V. Narayanan and S. Datta, “Impact of Variation in Nanoscale Silicon and Non-Silicon FinFETs and Tunnel FETs on Device and SRAM Performance,” *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1691–1697, Jun. 2015.

[98] X. Zhang *et al.*, “Analysis of 7/8-nm bulk-Si FinFET technologies for 6T-SRAM scaling,” *IEEE Trans. Electron Devices*, vol. 63, no. 4, pp. 1502–1507, 2016.

[99] T. Huynh-Bao *et al.*, “A Comprehensive Benchmark and Optimization of 5-nm Lateral and Vertical GAA 6T-SRAMs,” *IEEE Trans. Electron Devices*, vol. 63, no. 2, pp. 643–651, Feb. 2016.

[100] D. Jang *et al.*, “Device Exploration of NanoSheet Transistors for Sub-7-nm Technology Node,” *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2707–2713, Jun. 2017.

[101] J. Jeong, F. Atallah, H. Nguyen, J. Puckett, K. Bowman and D. Hansquine, “A 16nm configurable pass-gate bit-cell register file for quantifying the VMIN advantage of PFET versus NFET pass-gate bit cells,” in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, 2015, pp. 1–4.

[102] C. Auth *et al.*, “A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors,” in *2012 Symposium on VLSI Technology (VLSIT)*, 2012, vol. m, no. 2003, pp. 131–132.

[103] Pethe, Abhijit Jayant, et al. “Gate contact structure over active gate and method to fabricate same,” U.S. Patent No. 9,461,143. Oct. 2016.

[104] J.-H. Franke, M. Gallagher, G. Murdoch, S. Halder, A. Juncker and W. Clark, “EPE analysis of sub-N10 BEOL structures with Coventor’s SEMulator3D,” vol. 10145, p. 1014529, 2017.

[105] G. Murdoch *et al.*, “Feasibility study of fully self aligned vias for 5nm node BEOL,” in *2017 IEEE International Interconnect Technology Conference (IITC)*, 2017, pp. 1–4.

[106] B. D. Briggs *et al.*, “Fully aligned via integration for extendibility of interconnects to beyond the 7 nm node,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, no. c, pp. 14.2.1-14.2.4.

[107] B. Bilski *et al.*, “High-NA EUV imaging: challenges and outlook,” in *35th European Mask and Lithography Conference (EMLC 2019)*, 2019, no. August, p. 42.

[108] E. van Setten *et al.*, “High NA EUV lithography: Next step in EUV imaging,” in *Extreme Ultraviolet (EUV) Lithography X*, 2019, vol. 1095709, no. March, p. 5.

[109] D. De Simone and G. Vandenberghe, “Printability study of EUV double patterning for CMOS metal layers,” in *Extreme Ultraviolet (EUV) Lithography X*, 2019, no. August, p. 21.

[110] J. van Schoot *et al.*, “High-numerical aperture extreme ultraviolet scanner for 8-nm lithography and beyond,” *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 16, no. 04, p. 1, Oct. 2017.

[111] J. van Schoot *et al.*, “High-NA EUV lithography enabling Moore’s law in the next decade,” in *International Conference on Extreme Ultraviolet Lithography 2017*, 2017, no. October 2017, p. 30.

[112] N. Bekiaris *et al.*, “Cobalt fill for advanced interconnects,” in *2017 IEEE International Interconnect Technology Conference (IITC)*, 2017, pp. 1–3.

[113] C.-K. Hu *et al.*, “Electromigration and resistivity in on-chip Cu, Co and Ru damascene nanowires,” in *2017 IEEE International Interconnect Technology Conference (IITC)*, 2017, pp. 1–3.

[114] W. Steinhögl *et al.*, “Tungsten interconnects in the nano-scale regime,” *Microelectron. Eng.*, vol. 82, no. 3-4 SPEC. ISS., pp. 266–272, 2005.

[115] H. Xiao, *Introduction to Semiconductor Manufacturing Technology*, no. 2nd. SPIE, 2012.