AsAP7: A 7-nm finFET predictive process design kit

Lawrence T. Clark\textsuperscript{a,}\textsuperscript{b}, Vinay Vashishtha\textsuperscript{a}, Lucian Shifren\textsuperscript{b}, Aditya Gujja\textsuperscript{a}, Saurabh Sinha\textsuperscript{c}, Brian Cline\textsuperscript{c}, Chandarasekaran Ramamurthy\textsuperscript{a}, Greg Yeric\textsuperscript{c}

\textsuperscript{a} Arizona State University, Tempe, AZ 85287, USA
\textsuperscript{b} ARM Inc., San Jose, CA 95134, USA
\textsuperscript{c} ARM Inc., Austin, TX 78735, USA

\textbf{A B S T R A C T}

We describe a 7-nm predictive process design kit (PDK) called the AsAP7 PDK, developed in collaboration with ARM Ltd. for academic use. The PDK is realistic, based on current assumptions for the 7-nm technology node, but is not tied to any specific foundry. The initial version assumes EUV lithography for key layers, a decision based on its present near cost-effectiveness and resulting simpler layout rules. Non-EUV layers assume appropriate multiple patterning schemes, i.e., self-aligned quadruple patterning (SAQP), self-aligned double patterning (SADP) or litho-etch litho-etch (LELE), based on 193-nm optical immersion lithography. The specific design rule derivation is explained for key layers at the front end of line (FEOL), middle of line (MOL), and back end of line (BEOL) of the predictive process modeled. The MOL and BEOL DRC rules rely on estimation of time dependent dielectric breakdown requirements using layer alignments determined with projected machine to machine overlay assumptions, with significant guard-bands where possible. A high density, low-power standard cell architecture, developed using design/technology co-optimization (DTCO), as well as example SRAM cells are shown. The PDK transistor electrical assumptions are also explained, as are the FEOL design rules, and the models include basic design corners. The transistor models support four threshold voltage ($V_{th}$) levels for both NMOS and PMOS transistors. Cadence Virtuoso technology files and associated schematic and layout editing, as well as netlisting are supported. DRC, LVS, and full parasitic extraction is enabled through Mentor Calibre decks.

\textsuperscript{2016} The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Moore’s law has been a self-fulfilling prophecy that has resulted in dramatic improvements in computing power over multiple decades. However, as the 193-nm immersion optical lithography has pushed to its limits and augmented by the use of multiple patterning, cost has become the primary scaling impediment. Extreme ultraviolet (EUV) lithography has failed to be cost-effective through the 14-nm node but may be at the 7-nm node [1–3].

Designers are constantly faced with “what if?” scenarios. Those who must make decisions regarding design collateral for future processes face possibly the most significant challenges in this domain, as the target process is not fully defined. Moreover, what were once purely technology decisions now have increasingly significant impact on the practical usability of a process fabrication technology. Design/technology co-optimization (DTCO) is used to feed the impact of those decisions on the resulting designs back into the technology decision making process [4–7]. It is increasingly important as finFET width discretization and multiple patterning (MP) have increasingly constrained the possible layouts. New metal layers between the front end transistor and back end interconnect flow, appropriately named middle of line (MOL) have been added to provide better cell level connections with restricted patterning capabilities and MP [9].

Dennard (constant field) scaling has given way due to difficulty in reducing $V_{DD}$ due to $I_{on}/I_{off}$ considerations, i.e., vanishing room for threshold voltage ($V_{th}$) scaling while maintaining gate overdrive, and variability [8]. The addition of strain, as well as metal gate and high-k gate dielectrics has allowed performance increases. New transistor architectures, e.g., finFETs, have alleviated short channel effects in extremely scaled technologies and have allowed some, albeit limited, $V_{DD}$ scaling. We thus assume limited $V_{DD}$ scaling from the 14 nm to 7 nm node on the order of 100 mV or less. Notwithstanding the aforementioned improvements, power dissipation has become important to an extent so as to make the use of dark silicon – unpowered regions on a die that prevent it from exceeding the thermal limits – imperative for larger designs [10]. This, and the overwhelming volume of mobile applications...
has driven recent fabrication processes to emphasize optimizing both active and leakage power dissipation over raw performance. These limitations to Moore/Dennard scaling make it increasingly difficult to accurately benchmark new design and semiconductor technologies, as they become increasingly intermingled, and as the scaling pressures push fundamental academic research more closely to the technology development pipeline. The goal of this 7 nm PDK is to enable improved benchmarking of academic research and through better decision-making, accelerate the pipeline of academic research into the semiconductor technology landscape.

1.1. Objective, key assumptions and contribution

Our objective here is the development of a predictive process design kit (PDK) that allows realistic design on a not yet available technology node for academic use. The PDK is not tied to any specific foundry and thus will be inaccurate in some details, but is based on realistic assumptions regarding the lithography and technology capabilities. Our predictive 7-nm PDK, referred to as the ASAP7 PDK for the remainder of the paper, allows design exploration at the 7-nm node, accurately estimating circuit performance, area, and power for a design at that node.

We have assumed EUV for a number of key layers, particularly the middle of line (MOL) and metal 1 through 3. This is done for two reasons. Firstly, the 7-nm node represents the last node where EUV may not require MP solutions [1,2], and is thus the point where its cost effectiveness is maximized. Secondly, layouts using MP on those layers are extremely difficult and tedious [6,11,12], so academic use with those assumptions would be limited, except for the most advanced courses employing high levels of automation. Consequently, the ASAP7 PDK has easily understood design rules that nonetheless accurately capture the state-of-the-art at the 7-nm node.

Except for the EUV assumption, some of the assumptions and resulting rules are conservative when compared to the ITRS roadmap and other projections for this node [13]. For instance, Lin [14] proposes that a 15 nm half-pitch is required for the 7 nm node, just beyond what EUV capabilities are projected to be. Our conservatism is driven by a combination of the emphasis on mobile applications, avoidance of EUV MP, and literature that has indicated that the tightest possible poly pitches may not be essential to achieving the best designs [15].

1.2. Paper organization

Sections 2 and 3 describe the fabrication process constraints that drive the PDK design rules and the assumptions used. Section 4 describes the design rules, tying their values to requirements explained in Section 2. In Section 5 we show example cell library cells and SRAM, as well as how those key structures drove some of the design rule development. Section 6 describes the transistor assumptions, their performance, and the process corner assumptions. Examples validating the PDK are shown throughout. The CAD tools supported are briefly described in Section 7 and the paper is summarized in Section 8. The paper ranges from a tutorial level describing the underlying fabrication tool requirements, to very specific decisions based on capabilities, requirements, and DTCO considerations, regarding design rules and electrical behavior of the PDK.

2. Fabrication limitations

Process steps are divided into three groups, comprising the front end of line (FEOL), middle of line (MOL), and back end of line (BEOL). These encompass the wells and transistors, local interconnect (LI) and metallization layers from M1 to the top metal (M9).

2.1. Overlay

Overlay accuracy is the registration error in aligning a subsequent mask step to the prior one. It determines the worst-case distance between non-self-aligned mask layers [26]. Single machine overlay (SMO) is more accurate, as using the same equipment removes some error terms, but is less efficient and thus more costly. Multiple machine overlay (MMO), which describes the worst-case overlay between scanners is larger.

Lin predicts that typical MMO and SMO values for the 7-nm node must be 1.5 nm and 2 nm, respectively [14]. Other authors and the ITRS roadmap use a variety of values in the range of 2–4 nm [13,16,17]. ASML publishes their current capabilities in [82,83]. The key issue is that multiple patterning puts large stress on the overlay requirements, particularly LELE, which stitches polygons together to create 2-D patterns. This creates the need to use only 1-D patterning, which detrimentally affects the cell circuit density [7,11]. Moreover, pure 1-D layouts make layout of even relatively simple logic gates difficult.

2.2. Critical dimensions and lithography considerations

The critical dimension (CD) of a layer is proportional to the feature size. The inputs are uncorrelated so CD uniformity (CDU) is calculated as a sum of squares, i.e.,

\[
CDU = \frac{CDU_w^2 + CDU_{li}^2 + CDU_{die}^2}{2}
\]

where CDUw, CDUli, and CDUdie are the errors due to dose, focus, and mask, respectively [16,18–23]. The latter is about 1 nm for modern mask technologies. The former terms are equipment dependent.

Mask error enhancement factor (MEEF) is the ratio of the wafer CD error vs. mask error. It is given by

\[
MEEF = \frac{\Delta CD_{wafer}}{\Delta CD_{mask}}
\]

and thus describes the amplification of mask errors during transfer to the wafer [24]. Imaging non-linearity near the resolution limit increases MEEF substantially [25].

Edge placement error encompasses all of these values, summing the square of the overlay, CDU, line edge roughness (LER) and MEEF budgets [21]. It is the determining calculation of the design rules between layers. The required separation is determined by the distance needed for reliability.

2.3. Final layer spacing requirements

Time dependent dielectric breakdown (TDBB) is the primary driver of many of the MOL and BEOL design rules. After accounting for CDU, including LER, overlay errors, and MEEF, the resulting conductor separation must be sufficient to provide adequate lifetime at the expected operating voltages [5,27–29]. The separation is most important for layers with overlay errors, e.g. via to the underlying metal or in the MOL layers, discussed in detail in Section 4. Another key issue is the self-aligned contact to gate separation, which requires adequate spacer thickness and a cap layer on the gate [30].

The resulting yield estimation properly accounts for any edge placement error producing a failure across inter-field, intra-field, and die variations with any failure producing a bad die [21,31]. Designers will recognize this type of analysis as analogous with
3. Process assumptions

In this section we explain the underlying process assumptions. As stated, the proposed process assumes EUV lithography for a number of layers. A primary reason for this assumption was to keep the design rules simple enough for introductory, or at least mid-level design courses. We believe the complexity of MP, especially MP with more than two masks, which would be the EUV alternative for triple patterned or dual patterned with a cut mask at 7 nm, would make the kit difficult for academic use. In particular, determining and drawing such structures by hand is daunting, and we believe best automated. This issue is eased at the higher layers by MP support in modern place and route tools as well as automated decomposition. The basic assumptions are tested on the standard cell and SRAM layouts, i.e., following a DTCO approach. A NAND3 and an inverter based on the resulting standard cell template are shown in Fig. 1, along with their corresponding cross section. Diffusions are connected using M1 whenever possible, but LISD may be used instead in certain scenarios (this is useful to minimize M2 in latches). Because quality diffusion growth is enhanced by having a full fin between gates, fins are assumed to be cut midway through the gate.

Consequently, active (diffusion) breaks require a gate at either side, i.e., double diffusion breaks are required.

3.1. Layers

The FEOL and MOL process cross sections are shown in Figs. 2 and 3. Fig. 2(a) shows the section through the middle of the standard cell, where the local-interconnect gate (LIG) MOL layer contacts the gate. The cross-section through the gate between diffusions comprises Fig. 2(b). Raised source-drain (SD) layers contact the fins through the spacers, with source-drain trench (SDT) contacting the MOL local-interconnect source-drain (LISD) layer [33–35]. The cap on the gate between the spacers follows [30,36] and allows misalignment of the SDT without creating a TDDB limitation to the gate. In the ASAP7 PDK, fins are 32 nm in height and 6.5 nm thick, on a 27 nm pitch. To allow a 1 nm drawn grid, the drawn fin width is rounded to 7 nm.

A replacement high-k metal gate process follows the trend through 14 nm processes [23,34,35,37]. Gates are uniformly spaced on a grid with a contacted poly pitch (CPP) of 54 nm. To accommodate the CPP scaling the spacer thickness is assumed to decrease 1 nm at each node from 14 nm to 7 nm. Spacer formation follows poly gate deposition [38] allowing the use of low-k material in one spacer layer. 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.

**Fig. 1.** Standard cell architecture with 7.5 M2 track height assumed for DTCO analysis. Adjacent NAND3 and inverter FEOL, MOL, and M1 showing the double diffusion break that allows fin cuts under dummy gates. The fin block mask is generated by extending the drawn active regions under the gates. The basic 3-fin NMOS and PMOS standard cell architecture is apparent. Fins that are cut (not in active) are shown as gray. The cross-section comprises the lower right and corresponds to the cut line shown. It does not show LIG, V0, and M1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
or the gate cut fill dielectric, so source/drain growth is on full fins. Moreover, it provides spacer on both sides of SDT for self-alignment regardless of gate cut mask location (see Figs. 1 and 2). The PDK gate height is chosen to allow over 20 nm cap layer thickness (Fig. 2(a)). This thickness provides adequate distance to avoid TDDB after self-aligned contact etch sidewall spacer erosion, accounting for gate metal thickness non-uniformity [30]. Thus referring to Fig. 2(b), the gate cap is nitride, and provides this protection. Dual spacer width is 9 nm.

MOL layers are assumed to be patterned using EUV. We assume etch stops at the top of the pre-metal dielectrics PMD0, PMD1 and PMD2 so that SDT, LIG, and LISD do not penetrate, excepting variations, below their nominal depths. The latter allows LISD to cross over gates, freeing some pressure on M1 for within cell routing, since 2-D polygons are enabled by EUV patterning on LISD. The SDT is patterned separately (Fig. 3). This also enables the 2-D LISD and allows some flexibility in small structures, e.g., SRAM, as shown below. The primary TDDB limitation is however, SDT to LIG, as evident in Fig. 3 at upper right. The grown sources and drains (SD) are assumed to be trapezoidal as shown, but this does not affect the design rules or extraction besides setting the raised SD height.

The BEOL is comprised of nine layers. M1 through M3 are EUV, to allow fine 2-D routing to cell pins. This also allows easier drawn rules for cells and arrays on the first PDK version. Subsequent versions may change these assumptions. Via 1 (V1) through via 3 (V3) are similarly EUV. M4 through M7 are self-aligned double patterned (SADP). While SADP can support limited 2-D routing, we assume 1-D lines for these layers. M8 and M9 and their associated vias are single patterned layers.

Self-aligned via (SAV) is used throughout, following [39–42]. SAV uses the metal patterning hard mask to align the underlying via in one dimension (along the line) greatly easing the lithography requirements [39]. The actual via is patterned to the inner boundary of the hard mask or via mask, which overlaps considerably as illustrated in Fig. 4. Many process cross-sections appear to show zero top layer end-caps allowed for SAV. We allow this in the ASAP7 PDK, as shown, using the top metal hard mask (HMx+1) to define three of the SAV edges (dashed lines), resulting in the cross section at the right of Fig. 4 with the hard mask over the intra-layer dielectric and the barrier layers. The vias are aligned with the metal above, minimizing encroachment to the adjacent metal lines. Rules thus need to comprehend TDDB to the lower layer only.

3.2. Overlay and CDU

Inter-layer design rules, such as spacing and extension, are dependent on the total error in the placement of the masks associated with the layers in question. This total error in placement is given by the variability of the two masks, which can be approximated, assuming independent mask placement, using

$$\sigma_{\text{TEP}} = \sqrt{\sigma_{\text{Error,1}}^2 + \sigma_{\text{Overlay,1}}^2 + \sigma_{\text{Error,2}}^2 + \sigma_{\text{Overlay,2}}^2}.$$  

The $\sigma_{\text{Error}}$ terms for layers 1 and 2 denote the error in placement due to process variations, such as resist variability and LER, while the $\sigma_{\text{Overlay}}$ terms represent the overlay variations. As described earlier, overlay becomes the limiting factor in determining the
design rules. Arnold projected the need to have overlay under 3 nm that implies diffraction for MP due to endcap CD and overlay control between layers [16]. For \(a_{\text{TEP}}\) calculation, we assumed a 3 nm MMO of 1.7 nm, based on EUV scanner capability projections by Van Setten et al. [55]. The selection of this MMO value for design rule determination is one of the few non-conservative assumptions of the PDK and was done to achieve the aggressive spacing values required for patterning EUV layers used in SRAM cells. MMO, instead of SMO is used, assuming cost is important. A 3 nm error in placement, due to process variations alone, was assumed to be 2 nm. Using the aforementioned values for the constituent terms in Eq. (3), results in a \(a_{\text{TEP}}\) value of 3.7 nm. A TDBB required distance of 9 nm was used in the PDK and the design rule values were chosen for the EUV patterned layers to meet this TDBB requirement, despite the worst-case misalignment of 3.7 nm plus 1 nm CDU. The 9 nm distance is expected to be conservative at the 700 mV nominal \(V_{DD}\).

4. ASAP7 characteristics and design rules

4.1. Design rules and drawn polygon conventions

As evident in some examples above (fin and gate), drawn and actual polygons are not what you see is what you get (WYSIWYG) in our PDK. This provides some flexibility in the process assumptions while leaving layouts unchanged and allows the standard cells to be drawn on a relaxed 1 nm grid. It also makes recognizing the layout structures easier in a number of cases. Primarily, we chose to make the drawn diffusions stop short of the gates rather than inside them. This makes them much more evident to the designer, and it is a simple sizing operation at mask generation to move the fin block mask vertical edges, defined by the actual mask, under the gate as needed for the best fin cut. Similarly, SAV masks are derived from via drawn conventionally. The masks are aligned with the upper metal layer, with the overlap to allow larger polygons and overlap for misalignment implied (see the dashed outlines in Fig. 4). We support drawing the fins, although they can be copied since they occupy the same portions in every cell. They are helpful to view the diffusion alignments, but do not need to be viewed to produce layout.

The rules, the actual dimensions, and the underlying assumptions for some key layers are shown in Table 1. The fins are drawn at 7 nm width, with 20 nm spacing, although the actual fin physical dimension is 6.5 nm. Gates are drawn with a 20 nm gate length to stay on a 1 nm grid while the actual length is 21 nm. The compact models account for the difference automatically, so DRC and LVS check for 20 nm drawn.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Lithography</th>
<th>Width/drawn (nm)</th>
<th>Pitch (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>SAQP</td>
<td>6.5/7</td>
<td>27</td>
</tr>
<tr>
<td>Active (horizontal)</td>
<td>EUV</td>
<td>54/16</td>
<td>108</td>
</tr>
<tr>
<td>Gate</td>
<td>SADP</td>
<td>21/20</td>
<td>54</td>
</tr>
<tr>
<td>SDT/LISD</td>
<td>EUV</td>
<td>25/24</td>
<td>54</td>
</tr>
<tr>
<td>LIG</td>
<td>EUV</td>
<td>18/16</td>
<td>54</td>
</tr>
<tr>
<td>VIA0–VIA3</td>
<td>EUV</td>
<td>18/18</td>
<td>25</td>
</tr>
<tr>
<td>M1–M3</td>
<td>EUV</td>
<td>18/18</td>
<td>36</td>
</tr>
<tr>
<td>M4 and M5</td>
<td>SADP</td>
<td>24/24</td>
<td>48</td>
</tr>
<tr>
<td>VIA4 and VIA5</td>
<td>LELE</td>
<td>24/24</td>
<td>34</td>
</tr>
<tr>
<td>M6 and M7</td>
<td>SADP</td>
<td>32/32</td>
<td>64</td>
</tr>
<tr>
<td>VIA6 and VIA7</td>
<td>LELE</td>
<td>32/32</td>
<td>45</td>
</tr>
<tr>
<td>M8 and M9</td>
<td>SE</td>
<td>40/40</td>
<td>80</td>
</tr>
<tr>
<td>VIA8</td>
<td>SE</td>
<td>40/40</td>
<td>57</td>
</tr>
</tbody>
</table>

a Corner to corner spacing as drawn.

4.2. FEOL design rules

In this PDK, we assume fins are patterned over the entire design using self-aligned quadruple patterning (SAQP) and optical immersion lithography. Fin pitch continues scaling at a rate of 0.8 x. In light of the assumed continued scaling of fin pitch, it was decided that the fin height to fin pitch ratio would be too large and fin height was assumed to be unchanged from either the 14 nm or 10 nm process generations. Fin thickness is assumed to decrease one atomic layer (or roughly 0.5 nm) for each node, although this amount of thinning produces minimal changes in device characteristics.

Fin cut is EUV. The drawn active layer horizontal width is 16 nm. The actual mask active layer extends the fins to the CPP, which is 54 nm. As mentioned, this aids layout recognition and the active extension is easily accomplished. The actual raised source/drain is nominally 15 nm, defined by the gap between adjacent gate spacers (see Fig. 1) and accounting for the 21 nm gate actual \(L_g\). In the direction orthogonal to gates, 27 nm active layer minimum dimensions allow single fin devices for low power and SRAM design.

For logic layout, active (diffusion) to well edge vertical spacing is 27 nm, allowing two fin spacing between adjacent NMOS and PMOS devices (Fig. 1). N and P select layers have similar vertical spacing. The minimum horizontal width is 108 nm, greater than one cell pitch, i.e., the fin cut forces the double (dummy) poly diffusion breaks.

The replacement gates are grided across the design, using SADP. Any location where a gate is not needed, it is retained as a dummy gate following convention since the 32 nm node. A gate cut mask is used, as has been common since the 45 nm or 32 nm generation depending on the foundry. Using the published 45–14 nm production process values and trends, the following assumptions are derived. The CPP scaling ratio from the 14 nm to 10 nm technology nodes is 0.85 x, and is 0.9 x for the 10–7 nm nodes. Gate length (\(L_g\)) is assumed to decrease in the 14–10 nm and the 10–7 nm node transitions by 3 nm and 2 nm, respectively.

The CPP is 54 nm. This pitch is not as aggressive as shown in many references [6,13,30], but follows our assumptions that low power is more important, and eases the gate length to ensure good short channel effect (SCE) performance. It is a 15% shrink from published 10 nm CPP [34]. Moreover, DTCO experiments have shown that in some cases, a larger transistor pitch, allowing easier standard cell pin access, provides a better logic density at isopower and performance [15].

SADP is assumed to allow a single fixed gate length (\(L_g\)). Gate work function engineering and/or transistor lightly doped drain (LDD) changes allow multiple threshold voltages (\(V_{th}\)). Unlike published 22-, 16-, and 14-nm SRAM cells [40,46] but rather following [37,47] as we assume the same fin pitch for the entire design, again due to SAQP. This constrains the SRAM cell sizes that can be obtained, as described in Section 5.2.

4.3. MOL design rules

SDT is self-aligned by the gate spacers. However, it must be patterned wide enough to allow complete filling of the 15 nm spacer gap at worst-case gap dimension and SDT layer misalignment. We assume these are +1.5 nm per side and 3.7 nm, respectively. This implies an actual SDT width of over 25 nm. To stay on a 1 nm drawn grid, the rules require a 24 nm drawn width. This would be upsized at mask generation. The corner to corner spacing of SDT to LIG is limiting in the SRAM as shown in Section 5.2. Corner rounding due to lithographic effects may be beneficial in allowing appropriate spacing.

LISD is above SDT in the MOL stack. It can be drawn as narrow
as 16 nm with EUV, with under 15 nm currently proven in experiments [48,49]. Research aimed at the 7 and 5 nm nodes shows that new techniques can resolve smaller features using different mask magnification and anamorphic systems [50,51]. However, LISD is drawn at 24 nm, the same width as SDT, to lower resistance. LIG is drawn at 16 nm to allow dense cell layout with appropriate spacing to LISD and gate end-caps. EUV allows two dimensional layout, allowing LISD to cross over gates. As shown in Fig. 5, we assume that non-minimum LISD, required for SDT coverage allows a smaller line-end to line-end spacing than the 35 nm currently demonstrated for EUV with minimum features. This allows the LISD to LIG power rail connections shown without interference between adjacent standard cells (Fig. 5). This wide metal line-end to line-end rule is 27 nm and is consistent across all EUV layers. In particular LISD benefits when contacting adjacent gates.

4.4. BEOL design rules

We assume self-forming barriers (SFB) with Cu interconnects [37]. These reduce via resistance by up to $3 \times$ as compared to conventional deposited barrier process. Metal and via aspect ratios follow [13] at 2:1. We assume a 1.5 nm TaN barrier layer with Ru liner (seed layer) which provides better resistance due to larger Cu grain size [52–54].

As mentioned self-aligned via (SAV) is assumed at each metal layer. ASAP7 via layers allow diagonally adjacent vias. SAV allows via merging [78,79]. The ASAP7 design rules allow this when the vias are the same width and perfectly aligned, but not for the diagonal case or for SAVs without an end-cap. This capability is particularly helpful in the standard cells, allowing V0 to exist on the power rails without a spacing violation with contact to LISD (Fig. 5). The SAV rules allow non-square vias (Fig. 6) to allow wider metal and lower resistance on power and other critical, e.g., clock routes. This keeps the vias self-aligned to the metal above the via (labeled $M_{k+1}$) as evident. To constrain the sizes, the vias must be minimum in one dimension, as evident. Non-self-aligned vias are not supported in the PDK at present. Note that misalignment may make any via unlanded at the layer below.

M1 through M3 are single patterned EUV. This results in a 36 nm pitch, which is less aggressive than some publications predict at the 7 nm node [7]. The fin to M2 track ratio is commonly referred to as the gear ratio – it is important that these work out to integer (or half-integer) values. The 36 nm M2 pitch provides a good gear ratio with the 27 nm fin pitch and we felt that a 32 nm pitch was insufficient to justify multiple patterning if 36 nm EUV single patterning will suffice [29]. Thus, the kit is most aligned with assumptions made by Mallik et al. [2]. This also allows cell design without gridded M1 and M2, making layout simpler for academic use.

EUV single patterning allows 16 nm width lines (key for LIG to LISD spacing) and 36 nm pitch with NA $= 0.33$ [55]. Based on [55], the initial assumptions for line-end to line and line-end to line-end spacing were 25 nm and 35 nm, respectively. However, as stated earlier, the latter spacing value becomes limiting due to the interaction between LISD when standard cells are abutted along the power rails. Thus, based on the experiments by Setten et al. [31] – demonstrating line-end to line and line-end to line-end spacing down to 20 nm and 30 nm, respectively, with 50 mJ/cm$^2$ – the line-end to line-end spacing value is 31 nm. The aforementioned experiments show significant line-end shortening with a 22 nm space on the mask printing a space approximately 30 nm, but did not use advanced optical proximity correction. MEEF is near one in this regime, with these larger line-end to line and line-end to line-end drawn dimensions. M1–M3 follow the same rules, allowing two-dimensional polygons to ease routing.

SADP metal layers have rules that are a compromise between those required for best lithography, i.e., a pure line and cut approach [12] and what is usable by commonly available automated place and route (APR) tools [43–45]. We assume a single patterned block mask layer.

4.5. Design rule summary

Referring to Table 1, M1 through M3 line and space of 18 nm has been demonstrated as discussed above. Some of the key design rules are summarized in Table 2. The standard cells have very easy gate cut requirements with appropriate crossovers for pass-gates, i.e., a 40 nm gate cut width at the top and bottom of the standard cells, to allow margin to the LIG power rails (see Fig. 5). EUV linewidths down to 16 nm have been demonstrated, and this LIG width is necessary for the gate end to LIG spacing at the power rails of the standard cells. Standard cells require 27 nm (fin pitch)
SDT to cover one fin diffusion while the SRAM design rules stipulate that minimum SDT length be 17 nm.

5. Resulting cell layouts and DTCO considerations

5.1. Standard cell dimensions

We used a dense 7.5 M2 track standard cell baseline for the DTCO analyses when developing the PDK (see Fig. 1). Here, half-integer gear ratio values potentially allow larger power rails. The 27 nm fin and (assumed horizontal) M2 pitch of 36 nm results in an integer gear ratio of one at 6, 7.5, 9, and 12 metal tracks cell height, with 8, 10, 12, and 16 fins, respectively. Eight fins in the 6-track case is too small, as it allows only two fin wide gates and severely limits routing inside cells. The 7.5-track implementation allows dense cells with up to three fins each for NMOS and PMOS devices.

Due to fins under the dummy transistors, single diffusion breaks require the diffusions at both sides to be at the same potential. This is due to the lack of a fin cut in that case – fins connect the diffusions. The designer must explicitly drive both diffusions to the same potential. Making these connections is eased by the two dimensional LISD. This connection is also checked by the DRC rules.

5.2. SRAM cells and SRAM design rules

Unlike published 22, 16, and 14 nm SRAMs, we assume the same fin pitch for the entire design. This constrains the SRAM cell sizes. Processes routinely provide rules that allow closer spacing for SRAM, allowing better density. The highly repeated structures can be carefully tuned using actual manufacturing data, allowing tighter tolerances, but allowing only a few variations. While in commercial processes these SRAM rule changes can run to multiple pages, we endeavored to minimize them, while still allowing relatively dense cells. Various publications have pointed out that the MOL rules are strongly limited by the six transistor (6-T) SRAM requirements [40,46,56–58], so we used SRAM to help drive those DTCO design rule aspects. The difficulty of meeting MOL rules with multiple patterning has also been described [59]. A SRAM DRC layer enables the SRAM design rules and is assumed usable only in repeated memory cells. Foremost among these is the much tighter enclosure of active by N-well and NMOS/PMOS select layer – the nominal required spacing is halved for the SRAMs to allow the smaller NMOS to PMOS distance, with only one intervening fin rather than two as in the standard cells.

In the ASAP7 PDK, the separate SRAM \( V_{DD} \) selection layer forces devices to have SRAM characteristics, as described below. Standard cell devices can be ultra-low leakage using this layer and the associated schematic devices.

On finFET processes, SRAM cells come in three primary configurations. The first uses single-fin PMOS pull up transistors, two-fin NMOS pass-gate transistors, and three-fin NMOS pull down transistors. This SRAM configuration is referred to as the “123 bitcell” for the remainder of the paper. This design most easily meets the key requirements for write-ability and read stability. The access device can overpower the PMOS pull up to reliably write. The \( \beta \)-ratio of the pull down and access transistors provide a reasonably low risk in the potential of the logic 0 storage node providing read current for read stability. The second SRAM cell option uses a single-fin NMOS pass-gate device and two-fin NMOS pull down devices (referred to as the “112 bitcell”). Due to the close ratios of the NMOS and PMOS devices, this cell requires write assist. Lastly, the “111 bitcell” is densest, using a single fin for each device. This design provides poor write-ability and read stability, requiring circuit level assist techniques, e.g., reduced word line (WL) voltages with respect the cell \( V_{DD} \) to read, and write assist such as sub-\( V_{DD} \) bit line (BL and BLN) drive or column \( V_{DD} \) collapse. Since the “123 bitcell” is actually taller than our assumed standard cells and thus not dense, we focus on rules required for the others.

The “112 bitcell” on the ASAP7 PDK is illustrated in Fig. 7, along with the fin cut (block) and gate cut masks showing how rounding affects the actual fins and gates as patterned. Diffusions are separated by a fin (cut out) between NMOS or NMOS to PMOS devices. Following [59], we extend the access device gate to cover the fin cut defining the break from pull down to access NMOS device. This keeps sharp fin edges due to block mask curvature under gate (note that this is redundant with the continuous gate spacer assumption in Section 3.1). Additionally, it guarantees a gate spacer adjacent to the SDT, so SDT is self-aligned on both left and right, regardless of spacer and gate cut polygons. This assumption removes some gate cut to SDT rules that would otherwise be required. Dense SRAM and register file cells require minimum gate end to end spacing, which requires a 17 nm wide gate cut layer. The SRAMs have their cell top and bottom cell boundaries centered on a fin, rather than the middle of the fin spacing, for density. This necessitates a 0.5 nm drawn grid in that dimension. The LISD line-end to line-end rules are pushed to 31 nm to support this cell. Well to fin spacing is 17 nm. This allows a 5 nm misalignment while maintaining a large gap to the nearest sub-fin.

The “111 bitcell” follows the layout of Fig. 7, but has a single continuous active for the NMOS diffusion, resulting in an eight fin cell height.

6. Transistor electrical behavior

6.1. Scaling assumptions

All values used in the BSIM-CMG SPICE models [60] are derived from publically available sources [36,40,41,61–71] with straightforward assumptions based on historical trends. As above, we used conservative assumptions. Drive current is assumed to increase 15% node to node from 14 nm to 7 nm, which is consistent with the trends showing diminished \( I_{dsat} \) improvement over time. Linear current was assumed to be 4.5 \( \times \) smaller than the saturation current, in line with modern devices.

Regarding short channel effects (SCE) we assume that the subthreshold slope (SS) continues to approach the theoretical room temperature limit of 60 mV/decade, while drain induced barrier lowering (DIBL) is around 30 mV, following the best to date finFET published result. Both of these are eased by the relatively conservative 54 nm CPP allowing longer \( L_{gns} \). Capacitance is calculated using a method similar to that shown in [36] where the individual capacitances are broken up into simple 2-D parallel plate capacitors. While not based on TCAD, the method is accurate enough to estimate the compact model capacitances within 10%.

While compound transistor channels (fins) have been suggested for future technology nodes [72], CMOS devices with silicon channels are assumed for our PDK. This decision is based on research showing lack of compelling electrostatic improvement with Ge PMOS channels due to their higher dielectric constant. Additionally, these devices exhibited a significantly higher \( I_{gns} \) leakage floor due to band-to-band tunneling and greater gate delay when variability was evaluated [73]. We assume that strain trends are maintained and that this results in an approximately 10:9 NMOS to PMOS drive ratio. This follows trends reported for major foundries from 32 nm planar to 16 and 14 nm finFETs, where PMOS strain appears to be easier to obtain. Moreover, recent publications, e.g., [35,41] have reported greater PMOS than NMOS \( I_{dsat} \) values for some devices. This has led to the suggestion that PMOS rather than NMOS bit cell access transistors may be preferred in
such advanced technologies [84].

Since the use of multiple threshold voltages has become essential to meet both performance and standby power constraints in modern system-on-chip (SoC) designs, four devices are supported in the ASAP7 PDK. These are SLVT, LVT, RVT, and SRAM in decreasing order of drive strength, for both NMOS and PMOS (Fig. 8). As mentioned, all use the same drawn gate length. Low leakage devices are achieved mainly with work function engineering while the SRAM devices have very low leakage via both a work function change and the removal of the LDD implant. The LDD removal increases the effective channel length \( L_{\text{eff}} \), reduces GIDL and additionally, reduces the overlap capacitance. The SRAM \( V_{\text{th}} \) transistors may also be valuable for use in retention latches and other low-standby power circuits.

6.2. Models

To support realistic performance and power analysis on a variety of circuits, fast, typical, and slow corner models are provided. The basic transistor parameters at the typical (TT) corners are shown in Tables 3 and 4 for the NMOS and PMOS devices, respectively. The fast–fast (FF) and slow–slow (SS) process corners affect \( I_{\text{off}} \) and \( I_{\text{dsat}} \) as shown in Fig. 9. The threshold voltage behavior measured from simulation by measuring the gate voltage at constant drain current (VTC) [80], at typical, FF and SS corners comprises Fig. 10. \( I_{\text{off}} \) tracks the \( I_{\text{dsat}} \) very closely as evident in Table 3. Excellent subthreshold slope factors, nearing ideal, provides good \( I_{\text{on}}/I_{\text{off}} \) ratios for all devices.

Parasitic extraction models are based on the Cu/Ru materials (Section 4.4) and properly account for inter- and intra-layer dielectrics and spacing [74–77]. Pyzyna et al. described the behavior of Cu resistivity in highly scaled interconnects down to 28 nm pitches [81]. The Cu resistivity tracks with the grain sizes based on the drawn line width and the layer thicknesses following their data. Presently, air gap is not assumed in the extraction models.

7. CAD tools supported by the ASAP7 PDK

The kit is implemented in Cadence Virtuoso for schematic and layout entry. Gate cut follows gate, with a black background,
making gates appear conventional and easing recognition for the designer. Otherwise the appearance and use is completely conventional, including the active and via appearance as drawn. As mentioned above, BSIM-CMG models are used allowing flexibility for simulation, although only SPICE netlisting is currently supported.

Design rule checks (DRC) and layout vs. schematic (LVS) comparisons use Calibre nm-DRC and nm-LVS, respectively. We used Calibre XCalibrate to generate Calibre parasitic extraction (PEX) rules. Appropriate settings avoid double counting of S/D capacitance values, as verified by non-physical layout extractions with and without SDT and LISD layers present.

8. Summary

This paper has presented an overview of the ASAP7 7-nm PDK.

We have presented the key assumptions and attempted to justify the basic design rules and electrical performance for the predictive process. The PDK is based on realistic assumptions for the 7 nm technology node, which have been described in the paper. The rationale for key assumptions has also been described. The initial version assumes EUV lithography for some layers on FEOL through BEOL, a decision based on its present near cost-effectiveness and much simpler layout rules. A high-density low-power standard cell architecture, developed using design/technology co-optimization (DTCO), as well as an example SRAM bitcell and the impact on the design rules have also been discussed.

Acknowledgments

The authors thank the students in the spring 2015 special topics class at ASU who contributed to this effort by researching the cited material, setting up the initial Cadence schematic and layout capabilities, writing the netlisting, initial DRC and LVS decks, and providing initial cell library layouts on the PDK rules for DTCO. We thank Mentor Graphics for their generous and extensive training at ASU on SVRF rule writing for DRC and LVS.

References


Table 3

NMOS typical corner parameters (per fin) at 25 °C.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SRAM</th>
<th>RVT</th>
<th>LVT</th>
<th>SLVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idss (μA)</td>
<td>28.57</td>
<td>37.85</td>
<td>45.19</td>
<td>50.79</td>
</tr>
<tr>
<td>Ioff (nA)</td>
<td>13.07</td>
<td>18.13</td>
<td>23.56</td>
<td>28.67</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>0.25</td>
<td>0.17</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>0.27</td>
<td>0.19</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>SS (mV/decade)</td>
<td>62.44</td>
<td>63.03</td>
<td>62.90</td>
<td>63.33</td>
</tr>
<tr>
<td>DIBL (mV/V)</td>
<td>19.23</td>
<td>21.31</td>
<td>22.32</td>
<td>22.55</td>
</tr>
</tbody>
</table>

Table 4

PMOS typical corner parameters (per fin) at 25 °C.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SRAM</th>
<th>RVT</th>
<th>LVT</th>
<th>SLVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idss (μA)</td>
<td>26.90</td>
<td>32.88</td>
<td>39.88</td>
<td>45.60</td>
</tr>
<tr>
<td>Ioff (nA)</td>
<td>11.37</td>
<td>14.08</td>
<td>18.18</td>
<td>22.64</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>0.004</td>
<td>0.023</td>
<td>0.230</td>
<td>2.410</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>-0.20</td>
<td>-0.16</td>
<td>-0.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>-0.22</td>
<td>-0.19</td>
<td>-0.13</td>
<td>-0.07</td>
</tr>
<tr>
<td>SS (mV/decade)</td>
<td>64.34</td>
<td>64.48</td>
<td>64.44</td>
<td>64.94</td>
</tr>
<tr>
<td>DIBL (mV/V)</td>
<td>24.10</td>
<td>30.36</td>
<td>31.06</td>
<td>31.76</td>
</tr>
</tbody>
</table>

Fig. 9. Transistor I_on vs. I_off characteristics at process corners at nominal VDD=0.7 V.

Fig. 10. Transistor Vt (as measured with VTC) vs. process corner extracted from HSPICE at VDD=0.7 V.


