New category of ultra-thin notchless 6T SRAM cell layout topologies for sub-22nm

Randy W. Mann, Benton H. Calhoun
Department of Electrical and Computer Engineering, University of Virginia,
Charlottesville, VA, 22904
E-mail: <rwm3p, bcalhoun>@virginia.edu

Abstract

The extent to which the 6T SRAM bit cell can be perpetuated through continued scaling is of enormous technological and economic importance. Understanding the growing limitations in lithography, design and process technology, coupled with the mechanisms which drive systematic mismatch, provides direction in identifying more optimum solutions. We propose an alternative, ultra-thin (UT) SRAM cell layout topology as a means to address many of the challenging bit cell design constraints facing the most advanced CMOS process technologies today. Compared to the industry standard 6T topology, the newly proposed cell offers: 1) a lower bit line capacitance, 2) reduced M1 complexity and 3) notchless design for improved resistance to alignment induced device mismatch.

Keywords

SRAM, variation, SNM, Write Margin, manufacturability, 6T bit cell, yield, technology scaling

1. Introduction

Although the 6T SRAM cell topology has changed in previous technology nodes, the success and industry-wide use of today's 6T SRAM bit cell topology is evident in the ubiquitous use in the advanced VLSI industry at 65nm and below [1] [2] [3] [4] . Despite the wide scale use of this bit cell, it faces emerging challenges as scaling continues. Beyond 22nm, it is certainly less clear if the planar 6T cell will maintain its dominant role in microprocessor cache, ASIC, and mobile computing applications. This will depend on many factors such as continued advances in lithography, the successful incorporation of circuit assist methods [5], improved manufacturing practices for SRAM, and emerging technology options to address variation.

Additionally, the bit cell design may continue to evolve and adapt to the lithographic capabilities and constraints. The central question addressed in this investigation is, given the widespread use and acceptance of the topology in use today as the optimal solution, “Do competitive 6T alternative topologies exist for 22nm and beyond, and if so, what might they be?” In this paper, we propose a new 6T bit cell topology for future nanoscale SRAM technologies.

1.1. Constraints for future nanoscale 6T bit cell

The desired attributes for the next generation bit cell topology would include high density, reduced lithographic and manufacturing complexity, low bit line capacitance, and elimination or reduction of the sources of systematic mismatch. At 22nm, the use of 193nm immersion lithography and double patterning will be employed by the leading advanced silicon providers to meet the aggressive layout dimensions required. For nodes below 22nm, extreme ultra violet (EUV) with a wavelength of 13.4nm will be phased in for the most critical layers. These changes may serve as a driving force for continued evolution of the 6T layout topology.

The need to reduce variation from lithographic sources will continue to drive geometric simplicity, pattern regularity, fixed pitch regulations and increased reliance on optical proximity correction (OPC) algorithms [6] [7]. The use of double patterning is now commonly practiced for the gate level, and pitch doubling techniques are being developed with renewed emphasis. Mask costs continue to increase with each new node and the need for improved overlay or alignment tolerances drive increased costs of the stepper tools.

Printing the SRAM cell shared contact and conventional contact using the same mask level has been highly challenging and becomes more so as scaling continues. Elimination of right angles and jogs in the printed gate structures has been adopted for image control and integrity. Additional restrictions on gate direction and pitch are commonly implemented to provide further image fidelity. These factors converge to provide constraints on the cell designs for future technologies. These evolving constraints are becoming more restrictive with each node and effectively limit the viable set of 6T topologies for future nodes.

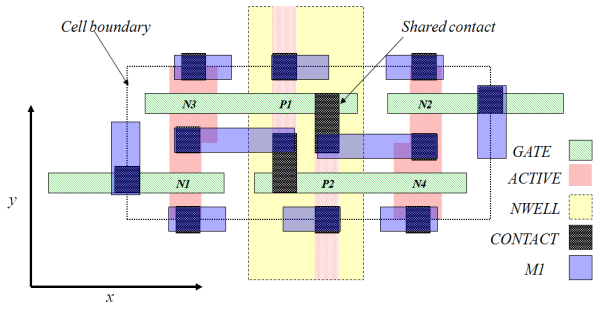


Figure 1: Type 4 6T layout. Depicts simplified M1 layer pattern [8], where the 'L' shaped pattern used in prior generations is eliminated to further simplify the required pattern.

For the industry standard 6T bit cell topology, there are several areas that are becoming more difficult with continued

scaling. Two areas specifically highlighted are: 1) the metal 1 (M1) pattern required (Fig.1) for this bit cell topology retains the relatively complex orthogonal directionality of the short lines [8], and 2) the jogs or notches in the active silicon region, used to achieve a desired pull down to pass gate ratio for cell stability during a read access, are subject to significant rounding.

Table-1: Variations of the inverter layouts and SRAM cell layouts.

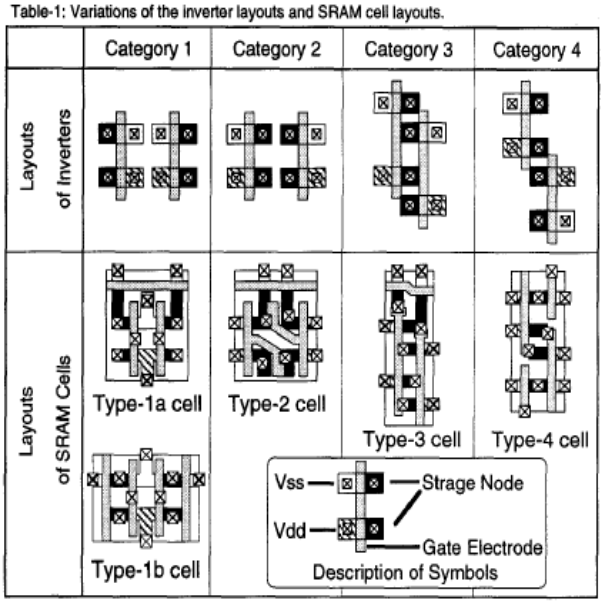


Figure 2: Summary of 6T cell layout topologies (© IEEE '98) [9].

2. 6T cell layout options beyond 22nm

Given the growing lithography restrictions with scaling and the known 6T topologies [9], as shown in Fig. 2, only two existing 6T topology options appear viable for further development. They are the topology that is currently the industry standard 6T cell (type 4) and a variant of type 1, Fig. 3. These two options, following the naming convention of Ishida [9], are shown in Fig. 1 (type 4) and Fig. 3 (type 1).

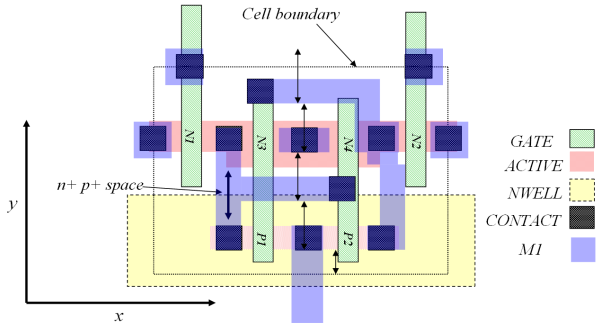


Figure 3: Alternative layout topology following Fig. 2, category 1, (type 1) that offers reduced lithographic pattern complexity for active and gate layers. With much larger area and increased complexity at M1, it offers elimination of the shared contact feature.

New category of ultra-thin notchless 6T SRAM…

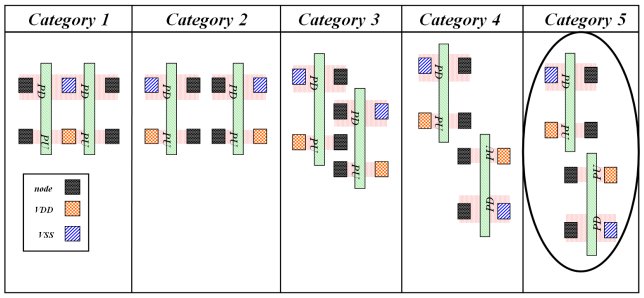
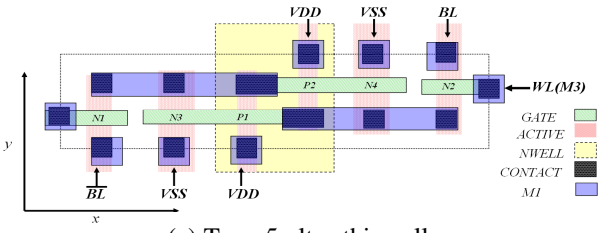
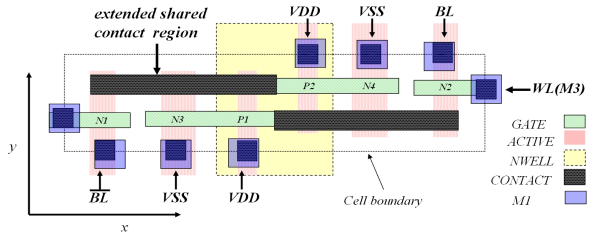


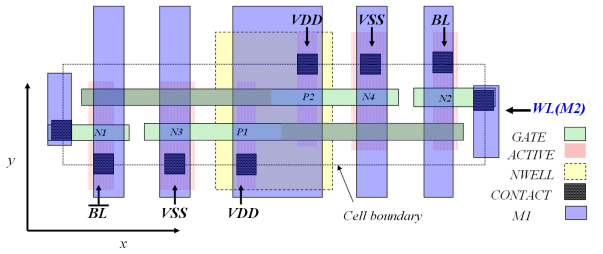
Figure 4: A new category 5 is shown in relation to previously defined categories for the 6T cell inverter layout options. The cross coupled inverters are now shifted so that the gate of the second inverter is in line with the contacts of the first inverter.



(a) Type 5 ultra-thin cell



(b) Type 5e ultra-thin cell



(c) Type 5b ultra-thin cell

Figure 5: Three versions (5, 5e and 5b) of the new, (ultra thin cell) short-bitline, notchless category 5 bit cell topology.

In light of this, a re-examination of Ishida's four base layout categories may be useful to determine if additional suitable base category alternatives may exist. In this work, we propose that the four categories may be expanded to five as shown in Fig. 4. A new base category is achieved by shifting the placement of the cross coupled inverters so that the gate of the second inverter is in line with the contacts of the first inverter [10]. This new category provides a third viable 6T cell topology option, consistent with the deeply

scaled CMOS lithographic restrictions and exhibiting many of the desired characteristics for further investigation.

The full 6T layout configuration for this category 5 ultra-thin (UT) cell is shown in Fig. 5. As shown in Fig. 5, the shared contact now runs parallel with the gate for the UT type 5 cell categories rather than orthogonal to the gate as for the conventional (type 4) cell. There are potentially several advantages for future generation technologies with this new layout topology. First, the metal 1 (M1) complexity is reduced to unidirectional routing further simplifying the required pattern compared the type 4 cell. Second, the cell height is further reduced (in the bit line direction) which allows for a reduced bit line capacitance. Third, the jogs or notches in the active silicon region are eliminated. This third point and its potential importance will be explored in the next section.

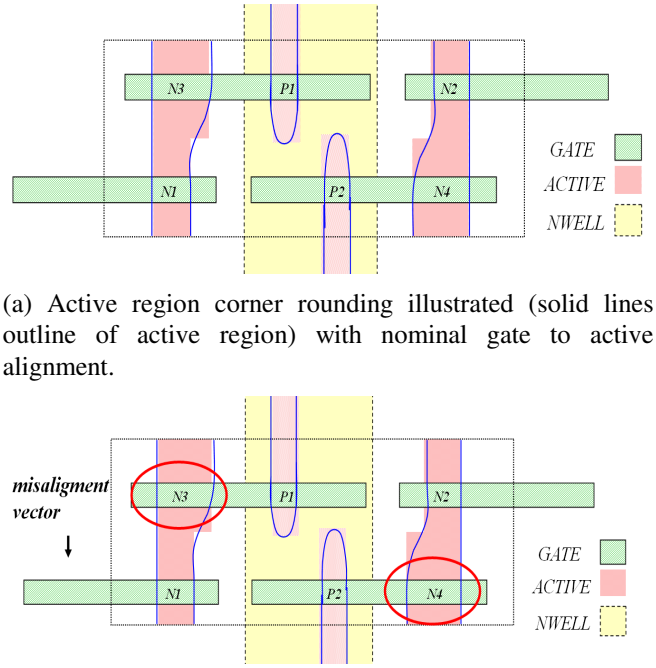
The shared contact is expected to have a similar complexity level as the type 4 where the shared contact 'bar' and conventional contact 'square' regions are printed using the same mask. An alternative, type 5e, layout scheme extends the shared contact across diffusion regions of the opposite inverter. This alternative layout Fig. 5(b), requiring an elongated 'shared contact' may offer unique options that will be discussed in section 5.

3. Layout dependent device variation in SRAM

Sources of mismatch in dense nanoscale SRAM devices due to variations in channel doping (both random and systematic) may be attributed to the use of pushed design rules and alignment sensitive doping variation sources such as halo shadowing, lateral implant straggle [11]. The general subject of non-random variation in dense SRAM devices may be further expanded to include the geometric sources of mismatch. These arise from the non-ideal environment associated with pushed design rules, variation in alignment and additional lithography effects such as corner rounding and line end foreshortening. These effects are layout topology dependent and can also contribute to the overall mismatch in the dense bit cell devices. Accounting for these additional components, the total variance is then expressed more fully as:

$$\sigma^2_{Vt_{TOTAL}} = \sigma^2_{Vt,DF} + \sigma^2_{Vt,GWF} + \sigma^2_{Vt,LER} + \sigma^2_{Vt,Weff} + \sigma^2_{Vt,Leff} \quad (1)$$

where the first term, $\sigma^2_{Vt,DF}$, captures the variation in channel doping due to both random and sources of systematic variation. The second term, $\sigma^2_{Vt,GWF}$, captures the variation associated with the gate work function. The last three terms in (1) capture the physical or geometrical variation. While line edge roughness (LER) plays a role in the ideal logic mismatch, the last two terms are typically neglected due to the proximity assumptions of the drawn ideal mismatch structures. As illustrated in Fig. 6, this is not always the case for the dense SRAM devices.



(a) Active region corner rounding illustrated (solid lines outline of active region) with nominal gate to active alignment.

(b) With misalignment the PD NMOS devices become geometrically mismatched due to corner rounding effects associated with the jog.

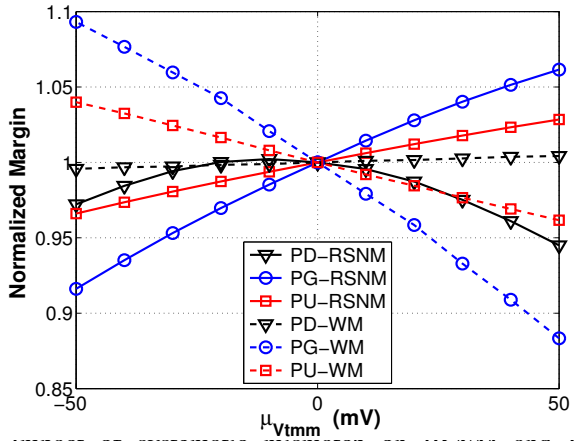
Figure 6: Illustration showing impact of gate misalignment on the device geometries. The devices circled exhibit different width characteristics and the width of N3 is effectively less than that of N4.

The geometry of the right (N4) and left (N3) devices, Fig. 6, become increasingly dissimilar as a function of alignment in the conventional 6T (type 4) cell [3] [4]. Additional variation in the Leff (not shown) can arise from similar arguments when line end foreshortening coupled with corner rounding are captured.

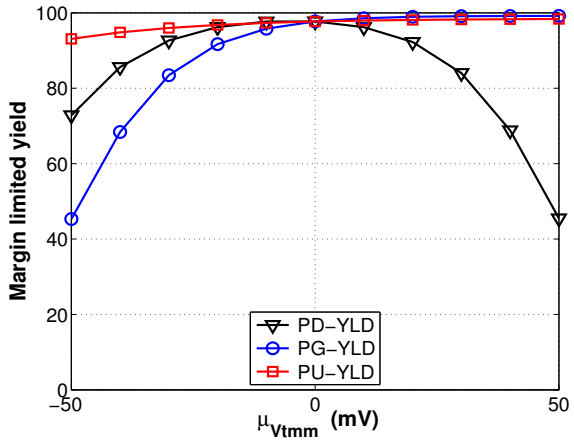
3.1. Yield impact of non-random mismatch

Normal random variation in device threshold mismatch within the bit cell device pairs is anticipated with an expected value or mean of zero. When non-random sources of mismatch introduce a mean shift in *Vtmm*, where ($\mu_{Vtmm} \neq 0$), an impact on the functional noise margins (RSNM), or write margin (WM) may be observed. Margin simulations were conducted using a commercially available 45nm LP technology. The impact of μ_{Vtmm} on the mean RSNM and WM is plotted in Fig. 7(a).

The yield (simulated at nominal voltage and relatively small array size) associated with relatively small deviations from the expected mean of zero begins to roll off quickly. This indicates that the notched N1/N3 (N2/N4) active silicon can rapidly affect yield unless alignment tolerances are sufficiently tightened below 22nm.



(a) Impact of systematic mismatch on RSNM and WM (normalized).



(b) Impact of systematic mismatch on margin limited yield for 2 Megabit SRAM.

Figure 7: Impact of ($\mu_{Vtmm} \neq 0$) on both RSNM and WM and margin limited yield. Simulations performed using on commercial 45nm LP technology SRAM models without the impact of increased variance. PD, PG, PU refer to the pull down NMOS (N3/N4) the pass gate NMOS (N1/N2) and the pull up PMOS (P1/P2) respectively as shown in Fig. 5.

4. Estimating 6T bit cell dimensions

By using the set of 'pushed' (sub-DRC) layout rules, given in Tables 1 and 2, optimized for the type 4 layout, the bit cell area for this topology may be estimated for comparison purposes. The rules will be applied to calculate the cell area for the industry standard topology (type 4) cell and then the new bit cell option for comparison purposes.

While some deviation will be expected as technologies evolve, the rules are expressed as function of the technology node (λ) to capture the effect of scaling. Although these pushed rules are consistent with those used in industry, some differences will exist between technology suppliers to allow optimization of yield and parametric values as desired.

Table 1: SRAM bit cell design rule scaling assumptions

Design rule	symbol	Dimensions (λ)
Gate to contact space	(GC)	0.7
Gate past active	(GPA)	1
Gate tip to tip	(TT)	1
Gate contact to active	(GCA)	1
Contact size	(CW)	1.4
Contact space	(CS)	1.4
p+ to p+ space	(AA)	1.7
n+ to p+ space	(NP)	1.8
M1 pitch	(M1P)	2.8

Table 2: SRAM cell device design rule scaling assumptions

Cell device	symbol	Dimensions (λ)
Pull down NMOS width	(Wpd)	2.5
Pull down NMOS length	(Lpd)	0.9
Pull up PMOS width	(Wpu)	1.4
Pull up PMOS length	(Lpu)	0.9
Pass gate NMOS width	(Wpg)	1.7
Pass gate NMOS length	(Lpg)	1.0

We use (Wpd/Lpd), (Wpg/Lpg), (Wpu/Lpu) to refer to the width and length of the pull down NMOS, pass gate NMOS and pull up PMOS devices respectively. The dimension X_4 for design topology 4, illustrated in Fig. 1, becomes:

$$X_4 = 2 \cdot (\tfrac{1}{2}(TT) + (GPA) + \max(Wpd, Wpg) + (NP) + Wpu + \tfrac{1}{2}(AA)) \quad (2)$$

and the dimension (Y4) is:

$$Y_4 = 2(CW) + 4(GC) + \max(Lpd, Lpu) + Lpg \quad (3)$$

Using the substitutions provided in Table 1 the bit cell area for topology 4 is expressed as a function of device dimensions and technology node dimension:

$$A_4 = (8.3 + 2 \cdot \max(Wpd, Wpg) + 2(Wpu))\lambda \cdot (5.6 + \max(Lpd, Lpu) + Lpg)\lambda \quad (4)$$

Given the assumptions provided in Table 2, the type 4 cell height (Y_4) is estimated to be approximately 7.5λ. This dimension is an important metric since it dictates the bit line length for this topology. Following the same set of pushed scaling rules, the X_5 dimension is estimated to be approximately:

$$X_5 = 2 \cdot (\tfrac{1}{2}(CW) + (GCA) + (Wpg) + (GPA) + (TT) + (GPA) + (Wpd) + (NP) + (Wpu) + \tfrac{1}{2}(AA)) \quad (5)$$

and the dimension (Y_5) is calculated to be:

$$Y_5 = 2 \cdot [\tfrac{1}{2}(CW) + (GC) + \tfrac{1}{2}(2 \cdot (CW) + (CS) - ((CW) - \max(Lpd, Lpu)))] \quad (6)$$

Using consistent assumptions the Y_5 estimate of 6.5λ represents a 13% reduction in the bitline length over the array. This directly corresponds to improved access speed. The estimated cell area is:

$$A_5 = 6.5\lambda \cdot (14.7\lambda + 2\lambda \cdot (Wpu + Wpg + Wpd)) \quad (7)$$

Using scaled and equivalent device dimensions a comparison of the calculated bit cell area results in 168.35λ^2 for cell type 5, and 142.45λ^2 for type 5e, compared to the 120λ^2 estimated for the type 4 cell. The limiting design rules used to calculate the type 5 cell dimensions are highlighted in Fig. 8.

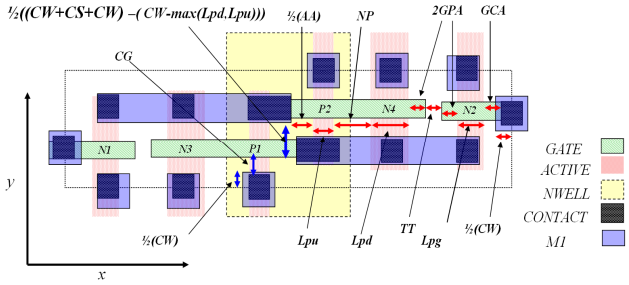


Figure 8: Type 5 6T layout with the area limiting rule assumptions highlighted.

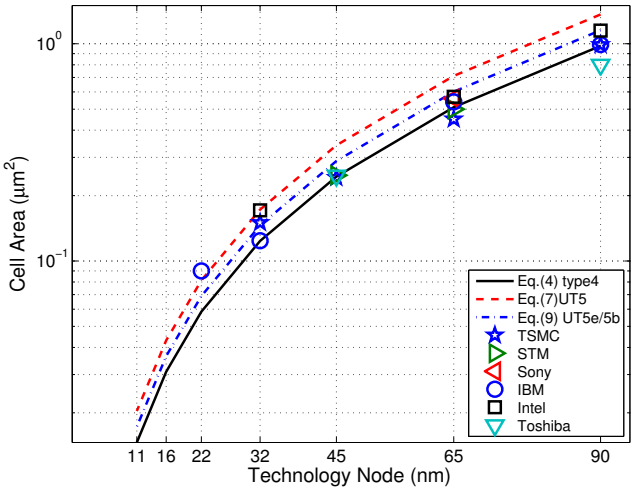


Figure 9: Calculated area for topology 5 cell across multiple technology nodes.

The second layout method which utilizes the extended shared contact (referred to as type 5e) is used to illustrate the potential area improvement that could be obtained by using a pitch doubling technology. While the assumed X_5 value will remain equivalent to the the type 5, the Y_5 value could be further reduced by:

$$Y_{5e} = 2(\tfrac{1}{2}(CW) + (GC) + \max(Lpd, Lpu, Lpg) + \tfrac{1}{2}(GS)) \quad (8)$$

The area for the type 5e becomes:

$$A_{5e} = 5.5\lambda \cdot (14.7\lambda + 2\lambda \cdot (Wpu + Wpg + Wpd)) \quad (9)$$

which provides a further reduction in the BL dimension with (Y_{5e}) equal to 5.5λ.

5. Summary

The UT cell (type 5e) example demonstrates the potential for further development. The potential synergy with the replacement gate process option used by some to combine the gate and shared contact patterning is clearly and interesting possibility for further exploration. By replacing the three 'in-line' shared contacts shown in Fig. 8 with shared buried contacts (patterning the gate and shared contact in one step) is an area for further investigation. If the shared contact layer is separate and isolated from the conventional contact, the cell can be wired very simply with VDD, VSS, BL, and BLB running vertically (y-direction) using the M1 layer and with the WL on M2 completing the 6T design running horizontally. This would result in a completely wired SRAM array requiring only two levels of metal (M1 and M2). This is an area for future exploration since the advantage of reduced required metal levels to complete the cell could offer savings in cost and free up M3 wiring channels across the array for logic routing.

Using the same set of pushed SRAM layout rules, the newly defined bit cell does not achieve the density calculated for the type 4 layout. This is partially due to the fact that the pushed rules used today are clearly optimized for the type 4 layout topology. The calculated cell areas based on the equations given here and published 6T bit cell areas are shown in Fig. 9. It is not clear if the deviation from 120λ^2 evident in published conventional cell sizes is driven purely by W and L up-sizing or if lithography limitations are playing a larger role. This deviation from the traditional scaled area may indicate that the type 4 layout is hitting limitations in scaling, which will renew interest in alternative topologies such as proposed here.

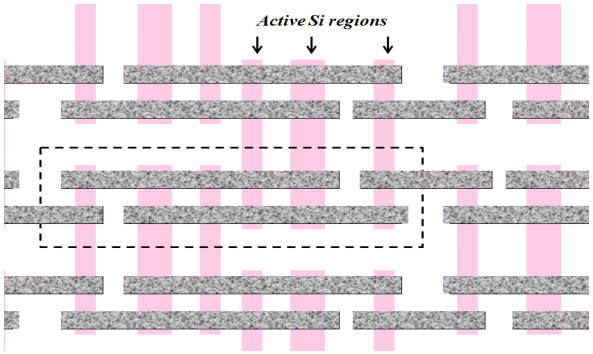
A comparison of bit cell metrics that highlight the key differences by cell type is given in Table 3. The bit cell area, BL length (L_{BL}), and number of required metal levels is summarized. Because the number of contacts required per cell is also a metric of interest, this metric highlights an additional advantage of the type 5 topology.

Table 3: SRAM bit cell metric comparison by cell type

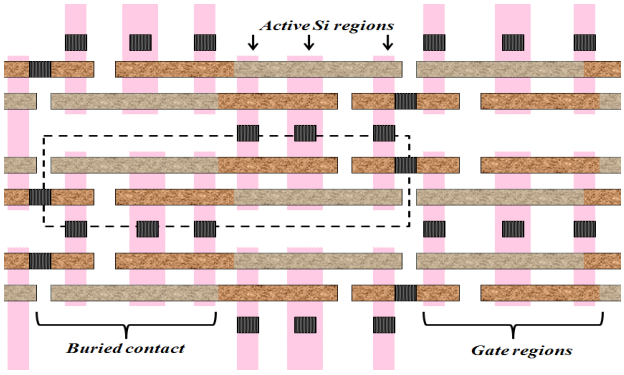
	Cell Type			
Metric	**4**	**5**	**5e**	**5b**
Number contacts/cell	6	8	4	4
Number shared contacts/cell	2	2	2e	2b
Cell area (λ^2)	120	168	142	142
L_{BL} (λ)	7.5	6.5	5.5	5.5
Number metal levels	3	3	3	2

The new 6T cell layout allows active silicon, gate, M1 and M2 to be printed as a series of straight unidirectional lines across the array, as shown in Fig. 10, eliminating the need for complex shapes corners and jogs, reducing systematic mismatch in the pull down NMOS devices as a result of the elimination of jogs in the active silicon. Active silicon, gate, and M1 may be completed with a cut mask

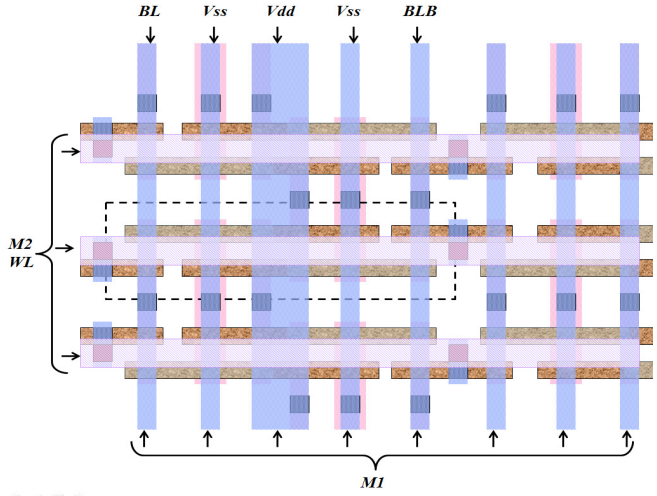
layer. An improvement of 13% or more in read access delay may be realized due to reduction in the bit line length.



(a) Array segment showing patterned notchless unidirectional active silicon (vertical) and partially patterned gate lines (horizontal).



(b) After gate cut and replacement gate processing with buried and conventional contacts.



(c) Fully completed array wiring with unidirectional, vertical M1 and horizontal M2 word lines.

Figure 10: (a) Top view of array segment showing active silicon and continuous gate pattern. (b) After gate cut and replacement gate processing with buried and conventional contacts. (c) Completed array through M2. Bit cell boundary is highlighted in (a-c) by dashed line rectangle.

6. Conclusions

As layout dimensions continue to be reduced, lithographic considerations will impose additional constraints on the layout of future nanoscale SRAM layout. Previously identified 6T bit cell topologies offer few promising alternatives for further exploration beyond 22nm. A new 6T topology is proposed in this work which may offer improved compatibility with future lithography restrictions and provide additional advantages over the existing type 4 topology. Based on this analysis, an area penalty of approximately (18-40%) will need to be weighed against the advantages of reduced alignment sensitive geometric mismatch, improved performance through reduced BL capacitance and reduced lithographic complexity. As systematic variation sources become worse with scaling, we anticipate a reduced area penalty for the UT category 5 cell options.

9. References

[1] Z. Luo, et al., "High performance and low power transistors integrated in 65nm bulk CMOS technology," *IEDM*, 2004, pp. 661–664.

[2] P. Bai, et al., "A 65nm logic technology featuring 35nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57μm^2 SRAM cell," *IEDM* , 13–15 Dec. 2004, pp. 657–660.

[3] K.-L. Cheng, et al., "A highly scaled, high performance 45 nm bulk logic CMOS technology with 0.242 μm^2 SRAM cell," *IEDM*, 10–12 Dec. 2007, pp. 243–246.

[4] Y. Wang, et al., "A 4.0 Ghz 291 mb voltage-scalable SRAM design in a 32 nm high-k + metal-gate CMOS technology with integrated power management," *J. Solid-State Circuits*, vol. 45, no. 1, pp. 103 –110, jan. 2010.

[5] R. Mann, et al., "Impact of circuit assist methods on margin and performance in 6T SRAM," *Solid-State Electronics*, vol. 54, no. 11, pp. 1398 – 1407, Nov 2010.

[6] L. Liebmann, et al., "Simplify to survive, prescriptive layouts ensure profitable scaling to 32nm and beyond," *Proc. SPIE,* vol. 7275, 2009.

[7] T. Jhaveri, et al.,, "Co-optimization of circuits, layout and lithography for predictive technology scaling beyond gratings," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 29, no. 4, pp. 509 –527, april 2010.

[8] B. Haran, et al., "22 nm technology compatible fully functional 0.1 μm^2 6T SRAM cell," *IEDM*, 15-17 Dec. 2008, pp. 1–4.

[9] M. Ishida, et al.,"A novel 6T-SRAM cell technology designed with rectangular patterns scalable beyond 0.18 µm generation and desirable for ultra high speed operation," *IEDM* , 6–9 Dec. 1998, pp. 201–204.

[10] U.S. Patent Application 61/365,962 "Improved Dense 6T Cell SRAM Layout Structure and Related Method".

[11] T. B. Hook, et al., "Lateral ion implant straggle and mask proximity effect," *IEEE Transactions on Electron Devices,* vol. 50, no. 9, pp. 1946–1951, Sept. 2003.