An Asymmetric SRAM cell to lower Gate Leakage

Navid Azizi
Department of ECE
University of Toronto
Toronto, Ontario, Canada
nazizi@eecg.utoronto.ca

Farid N. Najm
Department of ECE
University of Toronto
Toronto, Ontario, Canada
f.najm@utoronto.ca

Abstract

We introduce a new Static Random Access Memory (SRAM) cell that offers high stability and reduces gate leakage power in caches while maintaining low access latency. Our design exploits the strong bias towards zero at the bit level exhibited by the memory value stream of ordinary programs. Compared to conventional symmetric high-performance cell, our new cell reduces total leakage by more than 24% in the zero state at high temperature. With one cell design, total cache leakage is reduced by 24% at high temperature with no performance or stability loss. At low temperatures, where gate leakage is dominant, our cell reduces total cache leakage by 43%. We show that the new cell can be combined in an orthogonal fashion with asymmetric dual-V_t cells to lower both gate and subthreshold leakage, reducing total leakage by 45% to 60% with comparable performance and stability.

1 Introduction

As a result of technology trends, leakage (static) power dissipation has emerged as a first-class design consideration in high-performance processor design. Historically, architectural innovations for improving performance relied on exploiting ever larger numbers of transistors operating at higher frequencies. To keep the resulting switching power dissipation at bay, successive technology generations have relied on reducing the supply voltage. In order to maintain performance, however, this has required a corresponding reduction in the transistor oxide thickness to provide sufficient current drive at the reduced supply voltages. At the $70nm$ technology node, CMOS processes will have oxide thicknesses of $1.2nm$ to $1.6nm$ [1]. Since the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) gate tunneling leakage current increases exponentially with a reduced oxide thickness [15], gate tunneling leakage power dissipation will grow to be a significant fraction of overall chip power dissipation in modern, deep-submicron ($< 0.10\mu m$) processes [12] [18]. As the gate oxide thickness gets thinner, gate tunneling leakage could surpass weak inversion and drain-induced- barrier-lowering leakage as the dominant leakage mechanism in future technologies [8] [13] [10].

Since leakage power is proportional to the number of transistors, and given the projected large memory content of future System-on-Chip (SoC) devices (71% of die area by 2005 [3]), it becomes important to focus on SRAM structures such as caches, which comprise the vast majority of on-chip transistors. While the use of high threshold voltage devices provides a straightforward way to reduce subthreshold leakages in SRAM cells [6] [5], it is not easy to reduce gate leakage by simply improving device structure; a high-k dielectric is needed to raise the dielectric constant of the gate insulator and reduce gate leakage, but such high-k dielectrics are not expected to appear until 2007 [1].

Combined circuit- and architecture-level techniques exist that reduce leakage for those parts of the on-chip caches that remain unused for long periods of time (thousands of cycles) [7][11][19]. These techniques reduce the subthreshold and gate tunneling leakage of SRAM cells by gating the supply, and turning off the parts of the cache when they are unused. The mechanisms that identify which cache parts will be unused and that enable leakage reduction incur considerable power and performance overheads that have to be amortized over long periods of time. These methods are not effective when most of the cache is actively used.

We present a novel *asymmetric* SRAM cell design that reduces gate leakage in caches. The new cell exploits the fact that in ordinary programs most of the *bits* in caches are *zeroes* for both the data and instruction streams. It has been shown that this behavior persists for a variety of programs under different assumptions about cache sizes, organization and instruction set architectures, even when perfect knowledge of which cache parts will be left unused for long periods of time is known beforehand [4].

Traditional SRAM cells are composed in a symmetric fashion. Our *asymmetric* SRAM cell design offers lower gate leakage with little or no impact on overall memory access time. In our asymmetric SRAM cell, an extra transistor is added to reduce the voltage across the gate of a leaky transistor to reduce leakage when the cell is storing a zero (the common case). We evaluate our design by simulation, based on a $70nm$ Berkeley Predictive Technology CMOS technology, by adding an empirical gate leakage macro-model. The new cell can provide different performance/leakage/stability characteristics with careful design. The best design reduced total leakage by 24% at high temperatures at 43% at low temperatures with no loss in performance, and comparable stability. Furthermore, the new cell can be combined with the asymmetric dual-V_t cells found in [5] which reduce subthreshold leakage, to reduce both subthreshold and gate leakage. When used in combination with a dual-V_t cell total cache leakage is reduced by 45% to 60% with no loss in performance and comparable stability.
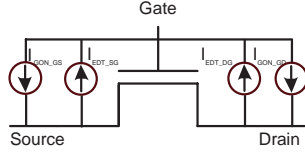
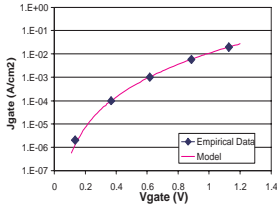Figure 1: Macro Model for Transistor Gate Leakage

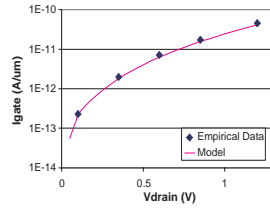**Figure 2: Fit of GON leakage measurements with macro model**

Figure 3: Fit of EDT measurements with macro model

There is, however, a 16.6% increase in cell area.

We make the following contributions: (1) We introduce a novel SRAM cell designed to lower gate tunneling leakage. (2) We evaluate the SRAM cell design and demonstrate that compared to a conventional cell it offers gate tunneling leakage savings while maintaining high performance and comparable noise margins and stability. (3) We show how the cell design can be modified to provide for different performance/leakage/stability characteristics. (4) We show that the new cell can be used in an orthogonal fashion with the dual-V_t cells found in [5] to increase the leakage savings and stability.

The rest of this paper is organized as follows: In section 2, we present some of the background and methodology used in this paper. In section 3, we present the new SRAM cell that lowers leakage. In section 4, certain design issues for the SRAM will be presented. In section 5 we combine the new cell with the dual-V_t cells. Finally, we conclude the paper in section 6.

2 Background and Methodology

2.1 Model for I_{gate}

All results reported in this paper are HSPICE simulation results produced at $110°C$ using Berkeley Predictive Technology Models (BPTM) [2] for a $70nm$ technology. Five different $70nm$ technologies were generated with t_{ox}s ranging from $1.2nm$ to $1.6nm$. The gate tunneling leakage was modeled using four voltage-controlled-current-sources as seen in Fig. 1. Two current sources, $I_{GON_{GS}}$ and $I_{GON_{GD}}$ model the direct tunneling leakage when the transistor is on, and two current source, $I_{EDT_{SG}}$ and $I_{EDT_{SD}}$ model the Edge-Directed-Tunneling (EDT) when the transistor is off [15]. The macro-model is similar to the one found in [9], except that it also includes the EDT current. The macro-model was fitted to industrial data found in [15]. Figs. 2 and 3 show the fit of the macro-model to the data at a $2.0nm$ oxide thickness. One important aspect shown on the plots is that both I_{GON} and I_{EDT} exponentially increase with increased voltage across the gate.

2.2 Leakage in SRAM cells

An SRAM cell is comprised of two inverters (P2, N2) and (P1,N1) and two pass transistors N3 and N4. In the
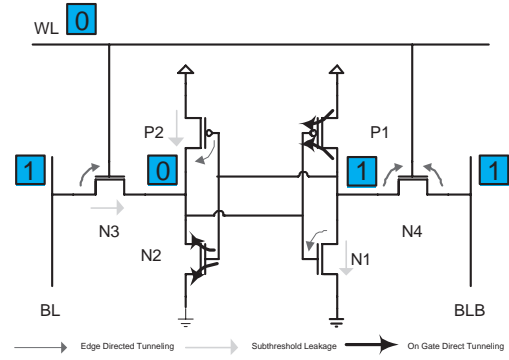

Figure 4: Sources of Leakage in SRAM cells

inactive state, the wordline (WL) is held low so that the two pass transistors are off, isolating the cell from bitline (BL) and bitline-bar (BLB). During this time the bitlines are also typically charged at V_{DD}.

In the inactive state, different transistors may dissipate leakage power, depending on the value stored in the cell. For example, when the cell is storing a '0,' as in Fig. 4, transistors N1, N3 and P2 dissipate subthreshold leakage, transistors N1, N3, N4 and P2 leak due to EDT, and transistors N2 and P1 dissipate on direct tunneling leakage. If the cell were holding a '1', the opposite transistors would be dissipating leakage.

There are multiple sources of gate leakage in an SRAM cell, but this work aims to reduce the gate tunneling current through transistor N2. One reason for this is that 70% of cache bits are zeros, and thus it is more important to reduce leakage in one state [4]. Second, the gate tunneling current through a PMOS transistor is an order of magnitude less than that through an NMOS [18]. Also, EDT is an order of magnitude less than the on-gate leakage [16] [17]. Thus, with these series of simplifications, the gate leakage through N2 is the most important.

While this work aims to reduce the gate leakage in SRAM cells, we have previously developed asymmetric SRAM cells in [5] that reduces subthreshold leakage, and this work can be combined in an orthogonal fashion with the dual-V_t cells to reduce both gate and subthreshold leakage as shown in section 5.

3 The PASS-CELL

Since gate leakage is exponentially related to V_{GS} and V_{GD}, one way to reduce gate leakage is to reduce the voltage on the storage nodes. A reduced V_{DD} lowers the voltage at the storage nodes and thus reduces the gate leakage, as well as subthreshold leakage, in the cell. This technique, however, lowers the stability of the cell, and increases the delay and dynamic power consumption of the cell since V_{DD} must be switched to its nominal value before a read.

Another possibility is to slightly decouple the storage node from the gate of the pull-down transistor, so that the voltage across N2 can be lowered without reducing the voltage at the storage nodes, as seen in Fig. 5. In this Pass Cell (PC), an NMOS pass transistor has been inserted between the right storage node and the gate of N2.

When the cell is holding a '0', which is the common case, the N5 pass transistor enters cutoff when it's source
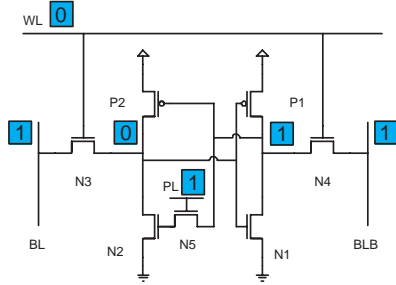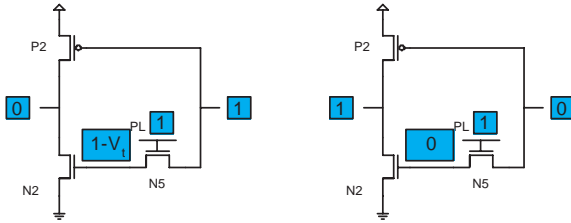
Figure 5: Asymmetric Pass Cell

Figure 6: The PC when holding a '0'

Figure 7: The PC when holding a '1'

Figure 8: PC leakage compared to CC

Figure 9: Increase in Discharge Time on BL

Figure 10: Increase in Cell Flip Times

Figure 11: $I_{\text{trip}}/I_{\text{read}}$ **value of PC cell**

voltage (gate of N2) is $V_{DD} - V_t$. Thus, the voltage across the gate of N2 has been reduced from V_{DD} to $V_{DD} - V_t$, resulting in reduced direct tunneling leakage. Fig. 6 shows the scenario in more detail. Notice that the voltage at the storage nodes is not affected by this design. However, transistor N2's conductance has been reduced, which may affect the performance of the cell, and will be discussed below.

When the cell is holding a '1', as shown in Fig. 7, the N5 pass transistor plays no role is the functioning of the cell as an NMOS is a good conductor of a logic '0'. There is, however, another source of direct tunneling leakage in this state. N5's gate is at V_{DD}, but it's source and drain are both and ground, and thus there is direct tunneling leakage through N5's gate. As will be seen below, since there are many more '0's than '1's, there is still a net leakage reduction.

The rest of this section will explore the leakage, performance, and stability characteristics of the PC.

3.1 Leakage Benefits

The PC was simulated at 110^oC at different values of t_{ox} and it's total leakage savings is shown, as a fraction of the total leakage of a conventional cell (CC), in Fig. 8.

At high values of t_{ox}, the subthreshold leakage dominates, and there is no discernible decrease in total leakage. As t_{ox} is lowered to $1.2nm$, the total leakage when storing a '0' has been reduced by 24% of the CC's total leakage, but the total leakage when holding a '1' has been increased by 10%. Since 70% of SRAM cells are storing a '0' [4], the total leakage of the array is reduced by 13%.

It will be shown below that, once combined with V_{PL} control, the total leakage savings become 24% at $1.2nm$ and once combined with a dual-V_t scheme, these savings become 45% to 60%. A 45% leakage reduction for a cache represents significant improvement.

At high temperatures, subthreshold leakage dominates and thus a 24% total leakage reduction is still a significant improvement given that the PC only reduces the gate leakage. At low temperatures the effect of N5 on reducing gate
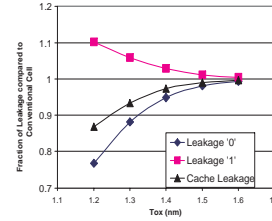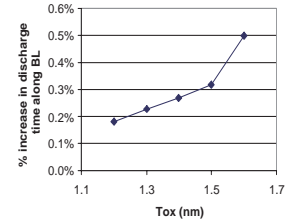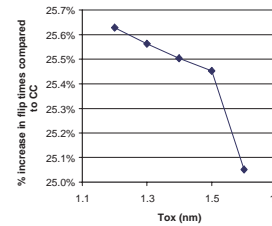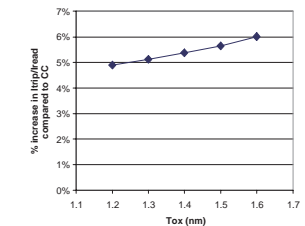
leakage is better seen. At 27^oC at $1.2nm$, the total leakage of the cache would be reduced by 43%.

3.2 Read Performance

The read access time of the PC cell is, however, degraded. When the cell is holding a '0', the bitline discharge along BL takes longer due to N2's lower conductance. The discharge time, which is only a small part of the total read access time, along BL is only 0.2% longer when the t_{ox} is $1.2nm$. When the cell is holding a '1', there is no speed degradation along the BLB discharge path; there is actually a 4% speedup in the BLB discharge path due to the asymmetry in the cell. Fig. 9 shows the BL discharge times.

3.3 Write Performance

With the added pass-transistor in the PC cell, the time to flip the cell has slightly increased. This is due to the increased delay through N5, to turn off N2 when a '1' is being written into the cell. Fig. 10 shows the delay. In the worst case there is a 26% increase in the flip time[1].

3.4 Stability

Another major consideration with the cell design is its stability. There are two interrelated issues: read stability and noise margins [6][14]. Intuitively, read stability indicates how likely it is to invert the cell's stored value when accessing it, and was computed as the ratio of $I_{\text{trip}}/I_{\text{read}}$, where I_{trip} is the current through the pull-down NMOS when the state of the cell is being reversed by injecting an external current I_{test} and where I_{read} is the maximum current through the pass transistor during a read [6]. The static noise margin (SNM) of an SRAM cell is defined as the minimum DC noise voltage necessary to flip the state of the cell [14]. In our case, the stability of all cells was measured by simulation via both the Static Noise Margin (SNM)

---

[1]The flip time is a very small portion of the total write time. A 26% increase is only a $14ps$ increase in the delay
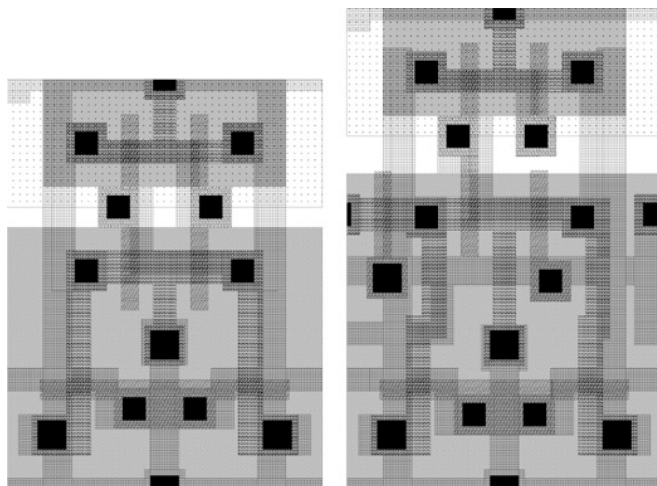
Figure 12: Comparison of Layout

Figure 13: Change in Cell and Cache Leakage due to changing V_{PL}

Figure 14: Change in BL Discharge Time due to changing V_{PL}

and the $I_{\mathrm{trip}}/I_{\mathrm{read}}$ methods. Under both stability tests, the stability was first measured under nominal conditions, assuming no process variations.

Then, to measure stability under process variations, Monte-Carlo analysis was performed to obtain a distribution for the SNM and $I_{\mathrm{trip}}/I_{\mathrm{read}}$. For each cell, 500 scenarios for V_t and length were randomly generated, consistent with their joint distributions, and simulated. The mean and variance of the distribution were then estimated.

The SNM of the PC remains largely unchanged since it is a DC characteristic. The $I_{\mathrm{trip}}/I_{\mathrm{read}}$ measure, on the other hand, improves with the inclusion of N5. Fig. 11 shows the $I_{\mathrm{trip}}/I_{\mathrm{read}}$ of the PC compared to a CC at different t_{ox}'s. At a t_{ox} of $1.2nm$ the PC shows a 5% improvement in $I_{\mathrm{trip}}/I_{\mathrm{read}}$. On the fast side of the cell, which has the limiting $I_{\mathrm{trip}}/I_{\mathrm{read}}$ value, I_{trip} is increased due to the larger capacitance at the storage node. On the slow side of the cell I_{trip} is increased due to the slight decoupling of the positive feedback due to N5.

3.4.1 Process Variations

The Monte-Carlo analysis to determine the stability of the PC compared to the CC was performed with a t_{ox} of $1.2nm$. The mean and standard deviation of the SNM for the PC was virtually unchanged compared to a CC. For the $I_{\mathrm{trip}}/I_{\mathrm{read}}$ measure the mean showed a 10.9% increases, and the standard deviation increased by 29%.

3.5 Area

SRAM cells have a compact layout since they are symmetrical. The addition of one extra transistor increases the area of the cell. Fig. 12 shows the layout of a CC and the PC in a $0.13\mu m$ commercial logic process. The CC has an area of $6.16\mu m^2$, while the PC has an area of $7.18\mu m^2$, an increase of 16.6%.

4 Design with $\mathbf{V}_{PL}$

The reason for designing the PC was to reduce the gate tunneling leakage, and this was accomplished by using a V_t drop from V_{DD} to reduce the direct tunneling leakage. Thus
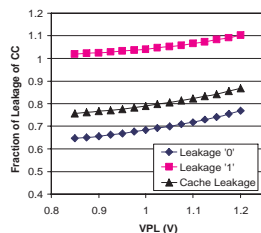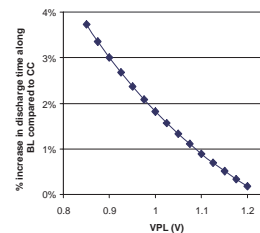
if N5 had higher V_t, then the direct tunneling leakage savings will be increased. This technique would only reduce leakage in the '0' state. If on the other hand the voltage on PL is lowered, leakage in both the '0' and '1' state will be reduced. This section describes the effect on the cell when V_{PL} is varied. All the results shown below are for a cell with a t_{ox} of $1.2nm$.

4.1 Leakage

The effect of lowering V_{PL} on the total cell leakage at high temperature is shown in Fig. 13. Not only does the '0' leakage decrease because of the lower voltage on the gate of N2, but the '1' leakage decreases due to the reduced V_{GS} and V_{GD} on N5. At high temperatures, with a 0.35V decrease in V_{PL}, the total leakage when holding a '0' is reduced by an additional 12%, to a 36% leakage reduction. The total leakage of a cache would be reduced by an additional 11% to a 24% reduction compared to conventional cache. At low temperatures, cache leakage is reduced by an additional 16% to a 43% reduction in total cache leakage.

4.2 Read Performance

A decreased V_{PL} leads to a lower voltage on the gate of N2, and thus the discharge time on BL increases. For example, with V_{PL} dropping by 0.35V, the discharge time is 4% longer than the CC. Fig. 14 shows the increase in discharge times on BL. The discharge time along BLB does not change.

The asymmetry in the discharge times can be used to have fast access times regardless of the value being stored. By using a new sense amplifier and a set of dummy bitlines [5], the read access times of the slow side of asymmetrical cells can be made to match the faster read time. To obtain fast read times irrespective of the data, a new sense amplifier was designed in [5]. The new circuit uses a set of *dummy bitlines*, D and DB, which are connected to a column of cells that all hold a '1'. Thus D and DB are always fast and are used to trigger the reading of a logical '0' thus achieving fast access times when the slow bitline is discharging. Fig. 15 shows the organization of the SRAM array. The use of this organization is advantageous over a single-sided read due to the increased noise immunity that a double-sided read provides. The detailed operation of the sense amplifier can be found in [5].

4.3 Write Performance

When decreasing V_{PL}, the flip times of the PC continues to increase. Fig. 16 shows the results where the flip time is
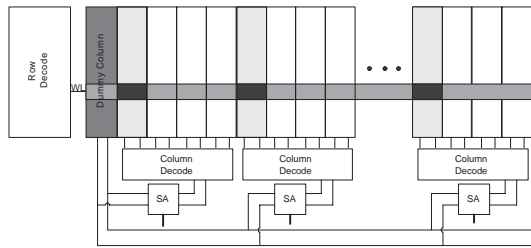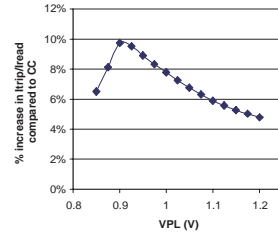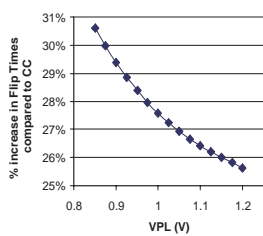
Figure 15: SRAM Organization

Figure 16: Change in Flip Times due to changing V_{PL}

Figure 17: Change in $I_{\mathrm{trip}}/I_{\mathrm{read}}$ due to changing V_{PL}

30% longer than that of a conventional cell's. Again, this change is flip times, which is a small portion of the total write time is unimportant as the read time of the cell is the limiting performance measure.

4.4 Stability

As V_{PL} decreases, and the voltage on the gate of N2 decreases, N2's conductance drops which will impact the stability of the PC. There is still very little change in the SNM of the new cell compared to a conventional cell.

The $I_{\mathrm{trip}}/I_{\mathrm{read}}$ measure, however, is affected by the change in V_{PL}. As V_{PL} is decreased, the I_{trip} on the fast side of the cell, which is the limiting side, increases because of the extra delay through N5, thus increasing $I_{\mathrm{trip}}/I_{\mathrm{read}}$. On the slow side of the cell I_{trip} decreases due to the reduced conductance of N2, thus lowering $I_{\mathrm{trip}}/I_{\mathrm{read}}$. The different sides of the cell experience different stability characteristics. Fig. 17 shows the change in $I_{\mathrm{trip}}/I_{\mathrm{read}}$, and we see that with a $0.275V$ drop in V_{PL} the slow side becomes the limiting side in terms of $I_{\mathrm{trip}}/I_{\mathrm{read}}$. The $I_{\mathrm{trip}}/I_{\mathrm{read}}$ of the PC is still 6% better than that of a conventional cell's.

4.5 Summary

This section examined the effect of decreasing V_{PL} on the various cell parameters. It was found that by decreasing V_{PL} we are trading the increased stability of the PC for reduced leakage. With a 0.35V drop on V_{PL}, the total leakage of the cache at high temperatures reduces from 87% of a conventional cache to 76% of the conventional cache with little change in stability compared to that of a conventional cell.

5 Combining with Dual-V_t

In [5] we have developed a series of asymmetric dual-V_t SRAM cells to lower subthreshold leakage. As t_{ox} becomes thinner, subthreshold leakage will compose a smaller, yet still important part of total static power consumption. In
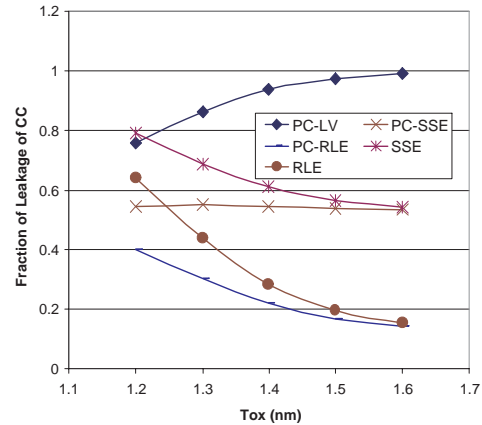

Figure 18: Leakage of dual-V_t cells

this section we combine the PC with the Stability-Speed Enhanced (SSE) and Resized Leakage Enhanced (RLE) asymmetric dual-V_t cells that were developed in [5] to obtain the PC-SSE and PC-RLE cells to lower both gate and subthreshold leakage[2].We show that not only are the leakage savings orthogonal, but so are the performance and stability of the dual-V_t pass-cells.

5.1 Leakage

Fig. 18 shows the total leakage of the various cells. At high t_{ox}'s where subthreshold leakage dominates, the PC-SSE and PC-RLE cells have the same total leakage as the SSE and RLE cells, a 1.8X and 6.6X reduction in total leakage respectively. As gate leakage becomes a more important part of the total leakage, the PC-SSE and PC-RLE cells have better leakage than both the dual-V_t cells that they derived from, and from the single-V_t PC-LV cell. At a t_{ox} of $1.2nm$ the PC-SSE and PC-RLE save an additional 21% and 36% of total leakage respectively compared to the PC cell, and save an additional 24% compared to the dual-V_t cell they derived from.

Fig. 19 shows the total leakage of the various cells at 27^oC. Here it is seen that the combination of the PC and the dual-V_t provides the best leakage savings.

5.2 Read and Write Performance

Since the pass transistor is on the slow side of the cell, the discharge time on the fast side of the cell is unaffected when transforming the SSE to the PC-SSE and when changing the RLE to the PC-RLE. Thus due to the sense amplifier presented in [5], there is no added speed degradation. The PC-LV cell, however, had a slight speedup in the BLB discharge time due to asymmetry inherent in the cell. This slight speedup remains in the PC-SSE cell which is 4% faster than the CC. The RLE cell is 4% slower than the CC cell, but due the added asymmetry of N5, the PC-RLE is now only 1% slower.

By combining the dual-V_t cells and the N5 pass-transistor the flip times for the cells increase. The flip time increase is nearly equal to the separate flip time increases

---

[2]The extra pass transistor is connected to the gate of the pull-down transistor on the slow side of the cell
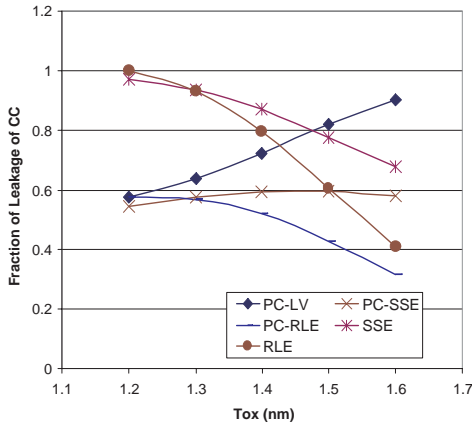
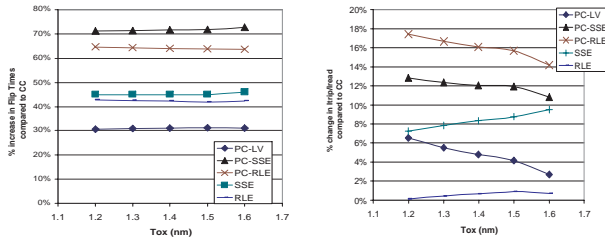**Figure 19: Leakage of dual-V_t cells at low temperatures**

Figure 20: Flip Times of dual-V_t cells   **Figure 21: $I_{\mathrm{trip}}/I_{\mathrm{read}}$ of dual-V_t cells**

associated with the dual-V_t and PC cell design. Fig. 20 shows the increase in flip times. Regardless, the flip time increases are only on the order of tens of picoseconds.

5.3 Stability

The SNM of the SSE and RLE, which are better than that of a conventional cell [5] remain almost unchanged when the cells are converted to the PC-SSE and PC-RLE. Fig. 21 shows $I_{\mathrm{trip}}/I_{\mathrm{read}}$ of the dual-V_t cells. The increased stability in the SSE and RLE cells help to increase the stability of the PC cell when it is transformed to the PC-SSE and PC-RLE cell. At a t_{ox} of $1.2nm$ the $I_{\mathrm{trip}}/I_{\mathrm{read}}$ of the dual-V_t cells is at least 6% better than the PC cell.

6 Conclusion

In this paper, a new asymmetric SRAM cell was proposed to reduce gate leakage in caches. The Pass Cell (PC) reduces direct tunneling leakage through one of the pulldown NMOS transistors because of three key observations. First, gate leakage through a PMOS is an order of magnitude less than through an NMOS [18], EDT tunneling is an order of magnitude less than the direct tunneling leakage [16] [17], and cache-resident memory values of ordinary programs exhibit a strong bias towards zero at the bit level [4].

The PC cell can be combined with dual-V_t design to reduce both subthreshold and gate leakage. There are multiple design possibilities that have different performance/leakage/ stability characteristics. At a t_{ox} of $1.2nm$ and at high temperatures, the best design reduces leakage by 24% of that of a conventional cell with no performance

degradation and comparable stability. At lower temperatures where gate leakage is a larger part of total leakage there is a 43% reduction in total cache leakage. There is, however, a 16.6% increase in cell area.

Furthermore, the leakage savings with this design are orthogonal to leakage savings incurred by turning off parts of the cache or reducing V_{DD} on the cell supply, or allowing the bitline voltage to float.

References

[1] *2002 Inter. Technology Roadmap for Semiconductors*.
[2] http://www-device.eecs.berkeley.edu/~ptm/.
[3] J. Abraham. Overcoming timing, power bottlenecks. *EE Times*, page 58, April 28 2003.
[4] N. Azizi, A. Moshovos, and F. N. Najm. Asymmetric-cell caches: exploiting bit value biases to reduce leakage powerin deep-submicron, high-performance caches. ECE Computer Group TR-01-01-02, University of Toronto, 2002.
[5] N. Azizi, A. Moshovos, and F. N. Najm. Low-leakage asymmetric-cell SRAM. *ISLPED*, 2002.
[6] F. Hamzaoglu et al. Dual V_t-SRAM cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13um technology generation. *ISLPED*, July 2000.
[7] S. Kaxiras et al. Cache decay exploiting generational behavior to reduce leakage power. *ISCA*, July 2001.
[8] A. Keshavarzi, K. Roy, and C. Hawkins. Intrinsic leakage in low power deep submicron CMOS ICs. *International Test Conference*, pages 146–155, 1997.
[9] D. Lee et al. Simultaneious subthreshold and gate-oxide tunneling leakage current analysis in nanometer CMOS design. *ISQED*, pages 287–292, 2003.
[10] M. Rosar, B. Leroy, and G. Schweeger. A new model for the description of gate voltage and temperature dependence of gate-induced drain leakage (GIDL) in the low electric field region. *IEEE Transactions on Electron Devices*, 47(1):154–159, January 2000.
[11] S. H. Yang et al. An integrated circuit/architecture approach to reducing leakage in deep-submicron high-performance I-caches. *HPCA*, January 2001.
[12] S. Song et al. CMOS device scaling beyond 100nm. *IEDM*, pages 235–237, 2000.
[13] S.H. Low et al. Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's. *IEEE Electron Device Letters*, 18(5):209–211, May 1997.
[14] E. S. Sr., F. J. List, and J. Lohstroh. Static-noise margin analysis of MOS SRAM cells. *JSSC*, 22:748–754, Oct 1987.
[15] W.K. Henson et al. Analysis of leakage currents and impact on off-state power consumption for CMOS technology in the 100-nm regime. *IEEE Transactions on Electron Devices*, 47(2):440–447, Feb. 2000.
[16] N. Yang, W. Henson, and J. Wortman. A comparative study of gate direct tunneling and drain leakage currents in N-MOSFETs with sub2-nm gate oxides. *IEEE Transactions on Electron Devices*, pages 1636–1644, Aug. 2000.
[17] Yo-Sheng Ling et al. Leakage scaling in deep submicron CMOS for SoC. *IEEE Transactions on Electron Devices*, 49(6):1034–1041, June 2002.
[18] B. Yu et al. Limits of gate oxide scaling in nano-transistors. *Symposium on VLSI Technology*, pages 90–91, 2000.
[19] H. Zhou, M. C. Toburen, E. Rotenberg, and T. M. Conte. Adaptive mode control: A static-power-efficient cache design. *Inter. Conf. on Parallel Architectures and Compilation Techniques*, September 2001.