Performance Evaluation of SRAM Cells in 22nm Predictive CMOS Technology

David Hentrich, Erdal Oruklu and Jafar Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, Illinois, 60616
Email: erdal@ece.iit.edu

***Abstract*— Static Random Access Memory (SRAM) units are often directly integrated onto the same die with the microprocessors and influence the design metrics significantly. SRAM often consumes large percentages of the die size and their leakages significantly contribute to the static power dissipation of those chips. The main objective of this article is to characterize the speed and power consumption of five different SRAM cells in a predictive high performance 22nm transistor process and in a predictive low power 22nm transistor process. The five types of studied cells are traditional 6T, gated-ground 7T, full Self-Controlled Voltage Level (SVL) 12T, SVL 9T Footed, and SVL 9T Headed. The simulation results indicate that the timing behavior of SRAM cells are largely the same but power dissipation, leakage power in particular, vary significantly in 22nm technology. The gated-ground 7T cells are deemed superior in the high performance process, while traditional 6T cells are deemed the best in the low power process.**

I. Introduction

Static Random Access Memory (SRAM) continues to be one of the most fundamental and vitally important memory technologies today. Because they are fast, robust, and easily manufactured in standard logic processes, they are nearly universally found on the same die with microcontrollers and microprocessors. The demand for ever larger and ever faster microprocessors has aggressively driven the scaling of transistor geometries down, density of SRAM cells up, and speed of SRAM cells up. However, along with these benefits, significant challenges have emerged that must be addressed at every transistor technology shrink. The two primary challenges (among others) are that the rail voltages are decreased to prevent thermal destruction to the transistors and that the current leakage through the transistors increases. Decreased rail voltages result in slower transistors (although this is offset by the reduction in capacitance afforded by smaller transistors). Increased current leakage occurs due to the necessary lowering of the transistor threshold voltages (V_t) and the reduction in the thickness of the gate insulators. Discussions concerning the adjustment of rail voltages, V_t, and sizing as they relate to SRAM timing and power can be found in [1] and [2].

This paper details the results of a study that evaluates the speed and power performance for various SRAM cells at the 22nm CMOS technology node. The HSPICE simulations utilize both the high speed and the low power predictive 22nm transistor models provided by the Nanoscale Integration and Modeling (NIMO) Group at Arizona State University [3] [4].

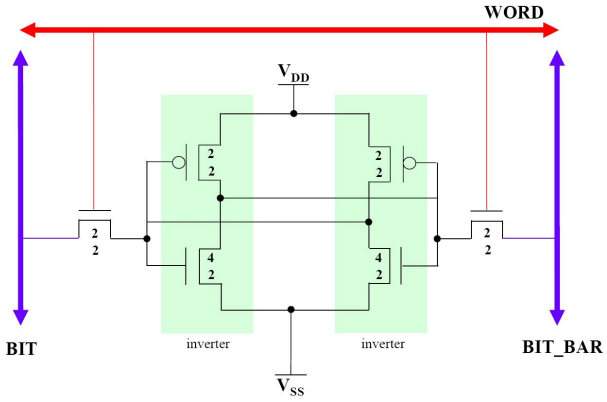

Fig. 1. Traditional 6T SRAM Cell Schematic

The NIMO group's Predictive Technology Model (PTM) work can be found at http://www.eas.asu.edu/~ptm. All of the 22nm PTM models used in this study assume the usage of metal gates, high-k dielectric gate insulators, and strained silicon construction.

II. SRAM Cell Designs

Five SRAM cells are evaluated at the 22nm technology node:

1) 6T Traditional SRAM Cell
2) 7T Gated-ground SRAM Cell
3) 12T SVL SRAM Cell with header and footer
4) 9T SVL SRAM Cell with footer only
5) 9T SVL SRAM Cell with header only

All of the above cells are variations of the traditional six transistor (6T) SRAM cell. However, the cells with more than six transistors are modified to reduce leakage currents. This study focuses on SRAM cell behavior at 22nm process relative to each another. The first studied cell is the 6T SRAM cell [5] [6]. Currently, it is the most prevalent type of SRAM cell in production. Figure 1 shows the schematic of the 6T cell evaluated in this study.

The 6T SRAM cell is two cross-coupled inverters (i.e. two transistors per inverter) arranged in a logic loop with a pair of pass transistors. The pass transistors are located on opposite sides of the cell from one another and couple their side of the SRAM cell to the associated bitline. When the WORD line is

978-1-4244-3355-1/09/$25.00©2009 IEEE

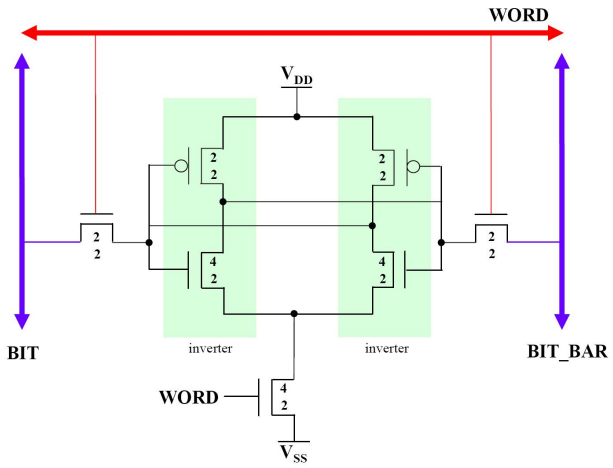

Fig. 2. 7T SRAM Cell Schematic

low, the pass transistors are off and the inverters are isolated from the bitlines. In this state, the inverters can hold the bit for as long as the correct voltages remain on the V_{DD} and V_{SS} rails. When the WORD line is high, the internal inverters are exposed to the bitlines for either *reading* or *writing*. In the case of *reading*, the cell (with an external amplifier) drives the physical bitline signals (BIT and BIT_BAR) to the values held in the respective sides of the SRAM cell. In the case of *writing*, the physical bitline signals are externally driven to the intended values for each side of the cell. The inverters inside the cell are overwhelmed and take on the respective bitline values.

In the 6T cell in Figure 1, all of the transistors have a channel length of 22nm (2λ). The width of the inverter PMOS transistors is 22nm (2λ) and the width of the inverter NMOS transistors is 44nm (4λ). The width of the NMOS pass transistors are sized as small as possible (22nm width) to minimize their capacitance contribution to the bitlines. The inverter NMOS transistors are sized larger than the pass transistors to increase the stored bit stability during read events [5].

The second cell investigated is the seven transistor (7T) *gated-ground* SRAM cell depicted in Figure 2. This cell has an additional NMOS transistor placed in the ground path of a traditional 6T SRAM cell to reduce leakage while the cell is in standby mode [7] [8] [9]. The bottom transistor is intended to cutoff the ground path while the cell is in standby mode to eliminate the leakage paths through the inverter NMOS sources. Practice has shown that the cell retains its value during standby even in the absence of a ground rail. In this study, the bottom transistor's gate is connected to the WORD line. The bottom transistor is sized identically to the inverter NMOS transistors to match their current carrying capacity.

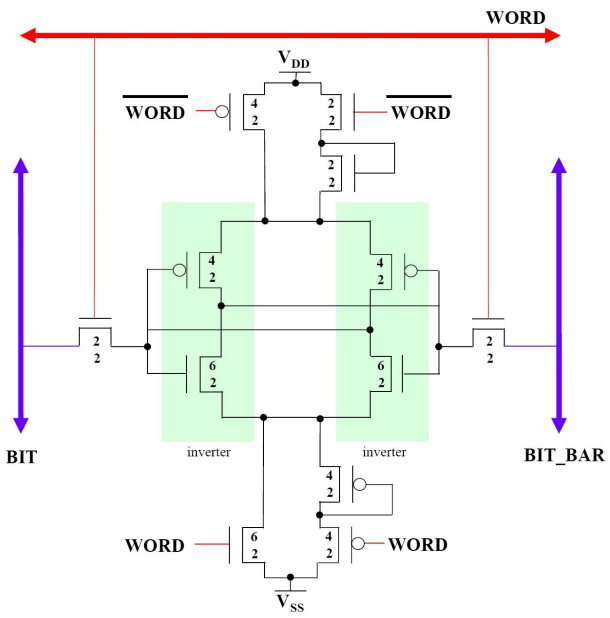

Fig. 3. Full SVL 12T SRAM Cell Schematic

The next three cells are variants of the Self-Controlled Voltage Level (SVL) circuit as applied to SRAM [10]. SVL circuits are meant to be placed in the V_{DD} and/or V_{SS} paths of a circuit to select between full rail voltage when the underlying circuit is in use and a reduced rail voltage when the underlying circuit is in standby. This technique can be applied to a wide range of circuits. SVL circuits are merely a way of dynamically adjusting the voltage to a circuit based on whether it is being utilized or not.

The third cell in this study is a 12T SRAM cell with SVL circuits placed in both the V_{DD} and V_{SS} paths of the SRAM cells as shown in Figure 3. The inverter NMOS and PMOS transistors were increased in size to speed up the cell. This size increase is intended to increase the speed of the cell by increasing the current capacity of the transistors to offset the extra capacitance created by the extra transistors.

The fourth cell is shown in Figure 4 and is a 9T subset of the previous 12T SVL SRAM cell. In this case, the SVL circuit is only applied to the bottom V_{SS} rail (thus, it is called a *footed SVL SRAM cell*). The PMOS transistors at the top of the cell are sized like the original 6T SRAM cell, while the bottom NMOS transistors are sized for the SVL circuit on the V_{SS} rail.

The fifth and final cell is shown in Figure 5. Like the previous cell, it has a single SVL circuit. However, in this case, the SVL circuit is in the top V_{DD} circuit path. As a result, it is called a *headed SVL SRAM cell*. The NMOS inverter transistors are sized like the 6T cell, while the PMOS transistors are sized for the SVL circuit.

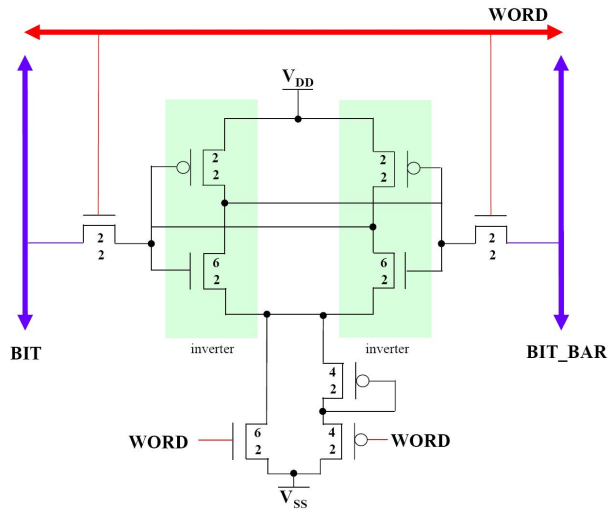

Fig. 4. Footer-only SVL 9T SRAM Cell Schematic

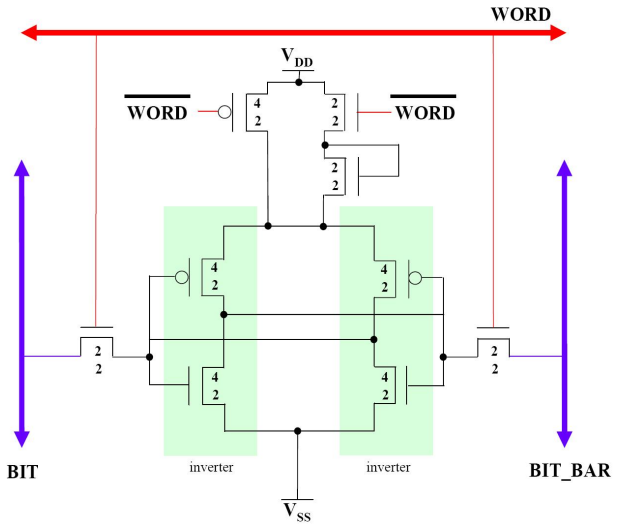

Fig. 5. Header-only SVL 9T SRAM Cell Schematic

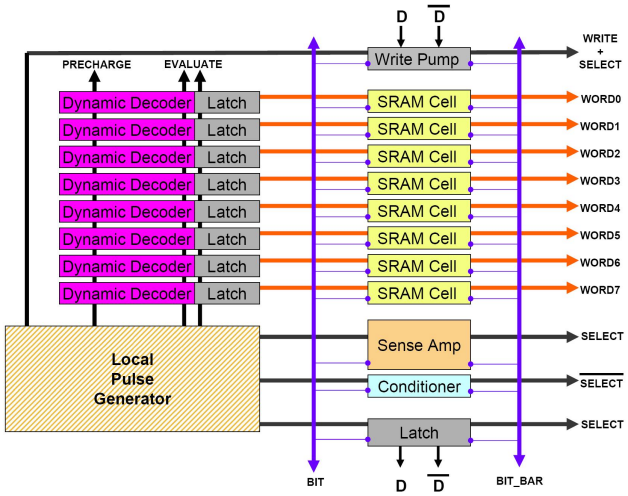

Fig. 6. Framework Block Diagram

III. Methodology

A. Test Framework

To understand the speed and power dissipation behavior of these SRAM cells at the 22nm technology node, a single SRAM bitline framework has been constructed in SPICE using the high performance 22nm PTM transistor models. Figure 6 shows a block diagram of the test framework. The left side of the diagram is devoted to address decoding and control signaling. The right side is a traditional 8 cell bitline block.

There are essentially two timing phases when accessing a bit. The first is the *evaluate* phase where the bit cell address is decoded. The bitline block is idle during this time and both physical bitline signals (BIT and BIT_BAR) are held at $V_{DD}/2$. The second is the *select* phase where the bitline block is accessed. The bitline is composed of a write pump, a conditioner, a sense amplifier, a latch, and eight SRAM cells.

The write pump is a circuit that sets an individual SRAM cell in the bitline to a particular value. The conditioner is a circuit that drives both physical bitline signals to $V_{DD}/2$ when the entire bitline block is in standby (idle). This is both the neutral position for the sense amplifier and a condition that will not upset the contents of an SRAM cell at the beginning of the read process. The sense amplifier is a circuit that assists the SRAM cells to drive the physical bitlines to the values reflected by the contents of the cell during a *read*. The sense amplifier used in this design is a traditional voltage sense amplifier. The latch is a mechanism that holds the value of the last bitline transaction while the bitline is unselected.

The five SRAM cell designs are evaluated in two separate transistor technology models from ASU's NIMO group: 22nm PTM HP V2.1 (High Performance) and 22nm PTM LP V2.1 (Low Power). This results in ten scenarios where both timing and power are measured. Every simulation contained exactly eight SRAM cells of the same type. The simulations were conducted in Synopsys HSPICE version A-2007.12.

It should be noted that the V_{DD} for both the high performance model SRAM cells and the lower power model SRAM cells is 0.8 V. This is the nominal voltage for the 22nm high performance transistor models. However, the nominal voltage for the 22nm low power transistor models is 0.95 V. The low power transistor model SRAM cells are evaluated at the lower V_{DD} of 0.8 V because the HSPICE test framework (i.e. everything except the SRAM cells) is implemented using only the 0.8 V high performance transistor models.

The results of this study are ideal. Neither interconnects nor cross-talk are modeled. Additionally, the models are run at 25 C under perfect voltage rail conditions. The results are only intended to be used as a general barometer of how various SRAM cells will behave in a 22nm process.

B. Measurement Criteria

Two kinds of measurements are considered in this evaluation: *speed* and *power*. For purposes of this study, the speed of an SRAM cell transaction is defined as the time between

TABLE I

PTM HIGH PERFORMANCE (HP) 22NM V2.1 SRAM CELL TRANSACTION TIMES IN PICOSECONDS (PS)

	6T	7T	SVL 12T	SVL 9T Footed	SVL 9T Headed
t_{read0}	97	108	105	104	99
t_{read1}	124	133	131	130	126
$t_{write0\rightarrow0}$	33	33	32	32	33
$t_{write0\rightarrow1}$	42	42	43	43	42
$t_{write1\rightarrow0}$	37	35	37	37	37
$t_{write1\rightarrow1}$	41	41	40	40	40

TABLE II

PTM LOW POWER (LP) 22NM V2.1 SRAM CELL TRANSACTION TIMES IN PICOSECONDS (PS)

	6T	7T	SVL 12T	SVL 9T Footed	SVL 9T Headed
t_{read0}	201	210	200	198	192
t_{read1}	219	256	240	237	229
$t_{write0\rightarrow0}$	32	24	24	22	23
$t_{write0\rightarrow1}$	41	27	27	27	28
$t_{write1\rightarrow0}$	34	23	25	22	25
$t_{write1\rightarrow1}$	40	20	27	19	18

the $SELECT$ line reaching 50% (0.4 V) to the time the last physical bitline signal (BIT or BIT_BAR) reaches its 90% position (this is 0.72 V if the line is going high or 0.08 V if the line is going low). SRAM cell transaction power was measured using the HSPICE *MEASURE* command across the SRAM *idle*, *read*, and *write* times. The power was measured using root-mean-square (RMS) as opposed to simple averaging. It was noticed during experimentation that there was often a wide difference between the power values computed by simple averaging versus the powers measured using RMS. RMS is a more robust way of taking power measurements because it reflects all of the power dissipation components across a circuit, regardless of the direction of the current or sign of the voltage.

IV. RESULTS AND ANALYSIS

The individual *read* and *write* times for the SRAM cells implemented in the high performance models can be found in Table I. The transaction times for the cells in the low power models can be found in Table II. Both tables are expressed in picoseconds.

Likewise, the RMS power measurements for the above high performance model and low power model *idle*, *read*, and *write* times can be found in Tables III and IV, respectively. Both tables are expressed in nanowatts.

A. Timing Analysis

In all the SRAM cells, whether they are implemented in the high performance or the low power transistor models, the *read* times are much longer than the *write* times. This is due to the different drive mechanisms behind a *read* and a *write*. In the case of a *read*, the SRAM cell itself is driving the physical bitline signals until the sense amp starts to help. In the case of a *write*, the write pump is driving the physical bitline signals. The *writes* are much faster than the *reads* because the write pump has a larger current drive capability than the SRAM cells.

TABLE III

PTM HIGH PERFORMANCE (HP) 22NM V2.1 SRAM CELL RMS POWER IN NANOWATTS (NW)

	6T	7T	SVL 12T	SVL 9T Footed	SVL 9T Headed
$p_{leakage}$	4.05	0.53	0.63	0.86	0.83
p_{read0}	2181	1807	2018	2030	2186
p_{read1}	1563	1344	1475	1494	1582
$p_{write0\rightarrow0}$	841	491	497	565	682
$p_{write0\rightarrow1}$	4874	4220	5870	5488	5116
$p_{write1\rightarrow0}$	5219	4590	6575	5987	5722
$p_{write1\rightarrow1}$	352	272	260	300	335

TABLE IV

PTM LOW POWER (LP) 22NM V2.1 SRAM CELL RMS POWER IN NANOWATTS (NW)

	6T	7T	SVL 12T	SVL 9T Footed	SVL 9T Headed
$p_{leakage}$	0.010	0.017	0.034	0.028	0.020
p_{read0}	372	276	340	339	369
p_{read1}	337	272	323	324	355
$p_{write0\rightarrow0}$	92	45	43	58	89
$p_{write0\rightarrow1}$	1485	1637	1775	1768	2112
$p_{write1\rightarrow0}$	1338	1068	1259	1022	1403
$p_{write1\rightarrow1}$	37	20	16	13	30

1) High Performance (HP) PTM Implementation: The *write* times are equivalent for all the cells implemented in the high performance transistor models. This is expected since the SRAM cells are sized such that their bit switching voltage thresholds are very closely matched to one another.

On the other hand, the *read* times vary by 18% between the fastest cell and the slowest cell. The fastest cell is the traditional 6T cell, followed in-order from fastest to slowest by the headed SVL, the footed SVL, the full SVL, and the 7T cell. In general, increasing the number of transistors slows the cells down due to the increase in capacitance and the stacking effect of transistors (i.e. transistors located physically farther away the rails see less voltage, and thus, don't switch as quickly). The order from fastest to slowest cells indicates that the presence of additional transistors on the ground rail is the most significant factor in slowing down an SRAM cell *read*. This suggests that the SRAM cell inverter NMOS sizing is critical to the *read* speed of an SRAM cell at the 22nm technology node.

2) Low Power (LP) PTM Implementation: Unlike the high performance transistor SRAM cell implementations, the *write* timing performance amongst the different SRAM cells implemented in the low power process is not the same. In this case, the 6T SRAM cell is approximately 50% slower than any of the other cells. With the exception of the 6T SRAM

cell, the *write* timings of the low power transistor SRAM cells are approximately the same. This seems to indicate the the transistor stacking along the rails in the low power transistor technology substantially weakens an SRAM cell, allowing it to more easily be overwhelmed by the write pump.

The *read* times for the low power transistor SRAM cells are generally twice as long as the *read* times for the cells implemented with the high power transistors. However, within the low power transistor SRAM cells, the relative *read* speed profiles between the cells are identical to the relative *read* speed profiles within the high power transistor SRAM cells (see section IV-A.1). Once again, the 6T is the fastest cell and the 7T is the slowest cell.

B. Power Analysis

1) High Performance (HP) PTM Implementation: For the high performance transistor implementations, every 7T and SVL SRAM cell implementation provides a large improvement in leakage power over the traditional 6T SRAM cell. This indicates that the dominant leakage component in the high performance transistor implementations is the drain-to-source leakage path, because this is the leakage path that the 7T and SVL SRAM cells are designed to reduce. Of all the cells, the 7T cell has the least power dissipation. This is due to the gate-to-body leakage (the other dominant, well-known leakage path at this transistor technology level [2] [11]). The 7T cell in this implementation has the fewest transistors, hence, the smallest gate-to-body leakage component.

The *read* power dissipation between all the cells are close between all the high performance transistor cell measurements. The 7T cell dissipates approximately 16% less power than the traditional 6T SRAM cell. The SVL 12T and SVL 9T footed cells dissipate approximately 6% less power than the traditional SRAM cell while the SVL 9T headed cell dissipates approximately the same power as the 6T cell. The common thread is that the circuits with feet (i.e. stacked transistors in the ground path reducing the voltage drops across the cell inverters) dissipate less power than a standard 6T SRAM cell.

When writing the same value that a high performance transistor SRAM cell already contains (i.e *write_same*), the 7T and SVL cells all dissipate less power than the 6T cell. In this case, the 12T SVL cell dissipates the least power, followed very closely in performance by the 7T gated ground cell. In this case, the header and footer both deliver a visible power dissipation benefit. Like *read* power dissipation in the previous example, it appears that the footer reduces power more than the header for the same reasons. The 7T cell outperforms the footed SVL 9T cell due to the fewer and smaller transistors on the ground rail side of the cell.

Writing a value to a high performance transistor cell that is different than what the cell already contains (i.e. a *write_flip*) is the most power expensive operation that can be performed on an SRAM cell. It is a few multiples greater than the equivalent *read* operation. The power dissipation associated with a *write* has historically been recognized as the dominant power transaction in an SRAM cell [12]. In the case of the high performance transistors, the 7T cell dissipates almost 13% less power than the 6T cell. However, all of the SVL cells dissipate more power than the 6T cell during a *write_flip* due to their short-circuit current capacity.

2) Low Power(LP) PTM Implementation: For the low power transistor implementations, the 6T cell dissipates less leakage power than any of the other cells. The designs of the 7T and SVL cells are meant to counteract source-to-drain leakage, however, these designs only increased leakage. This suggests that the dominant leakage component in the low power transistors is gate-to-body leakage instead of source-to-drain leakage in the high performance transistors. Given this, the 6T cell dissipates the least leakage power because it has generally fewer and smaller transistors than the 7T and SVL cells.

The power dissipation during *read* transactions in the low power transistor cells is approximately six times less than the power dissipations in the same respective cells in the high performance transistor process. However, the low power transistor SRAM cell *reads* generally behave the same relative to one another as in the high power transistor SRAM cells (i.e. the footed cells dissipate less power than the non-footed cells, with the 7T cell dissipating the least power). The voltage drops across the stacked transistors in the feet of the cells diminishes the short-circuit power that occurs during a *read* by reducing the total voltage drop across the NMOS transistors in the SRAM cell inverters.

During *write_same* events, the low power transistor cells consume approximately ten times less power than their high performance transistor counterparts. Once again, the low power transistor cells behave the same relative to one another as they do in the high performance transistor process. The 12T SVL consumes the least power followed very closely by the 7T cell. Again, both headers and footers deliver reduced power consumption due to the transistor stacking effect reducing the voltage drop across the SRAM cell inverters.

Like the high performance transistor SRAM cell implementations, the 7T cell dissipates the least power during a *write_flip* transaction in the low power transistors. However, the footed 9T SVL dissipates slightly less power than the 6T cell. The 12T SVL cell dissipates more power than the 6T cell, while the 9T SVL headed cell dissipates even more power than the 12T SVL cell. In the low power transistor process, SRAM cell footers deliver a power dissipation improvement. However, the presence of a header increases the *write_flip* transaction short circuit switching power.

V. Conclusion

This study focused on the timing and power behavior for various kinds of SRAM cells at the 22nm node using HSPICE simulations and the predictive transistor models. In summary, the 6T cell is the fastest cell to *read* in both the high performance and low power processes. However, it has the same *write* speed as the other cells in the high performance process, but is slower than the other cells in the low power process. In contrast, the 7T cell is generally the slowest cell

to *read*, but generally among the fastest to *write*. Overall, the timings of the cells are very close together. Timing does not appear to be a clear differentiator between any of the studied cells at the 22nm technology node.

From a power perspective, the 7T cell performs the best during *read* and *write* transactions in both the high performance and low power transistor implementations. From a leakage perspective, the 7T cell performs phenomenally better than the 6T cell in the high performance process. However, the 6T cell leaks less than the 7T cell in the low power transistor implementations.

Since leakage is generally the most important concern (i.e. every cell leaks all the time, while only a minuscule number of cells are generally *read* or *written* at any one time) and cell timing is not viewed as a big differentiator, the overall conclusion is that 7T SRAM cells should be used when utilizing 22nm high performance transistors, while traditional 6T SRAM cells should be used on 22nm low power transistor designs.

REFERENCES

[1] E. Morifuji, T. Yoshida, M. Kanda, S. Matsuda, S. Yamada, and F. Matsuoka, "Supply and threshold-voltage trends for scaled logic and SRAM MOSFETs," *IEEE Transactions on Electron Devices*, vol. 53, no. 6, pp. 1427–1432, June 2006.

[2] E. Morifuji, D. Patil, M. Horowitz, and Y. Nishi, "Power optimization for SRAM and its scaling," *IEEE Transactions on Electron Devices*, vol. 54, no. 4, pp. 715–722, April 2007.

[3] Y. Cao, T. Sato, M. Orshansky, D. Sylvester, and C. Hu, "New paradigm of predictive MOSFET and interconnect modeling for early circuit simulation," *IEEE 2000 Custom Integrated Circuits Conference*, pp. 201–204, May 21-24 2000.

[4] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Transactions on Electron Devices*, vol. 53, no. 11, pp. 2816–2823, November 2006.

[5] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, 2nd ed. Pearson Education, Inc., 2003.

[6] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 5th ed. Oxford University Press, 2004.

[7] A. Agarwal, H. Li, and K. Roy, "DRG-Cache: A data retention gated-ground cache for low power," *Proceedings of the 39th Design Automation Conference*, pp. 473–478, June 10-14 2002.

[8] ——, "A single-Vt low-leakage gated-ground cache for deep submicron," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 2, pp. 319–328, February 2003.

[9] K.-S. Yeo and K. Roy, *Low-Voltage, Low-Power VLSI Subsystems*. McGraw-Hill, 2005.

[10] T. Enomoto, Y. Oka, and H. Shikano, "A self-controllable voltage level (SVL) circuit and its low-power high-speed CMOS circuit applications," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1220–1226, July 2003.

[11] R. M. Rao, J. L. Burns, A. Devgan, and R. B. Brown, "Efficient techniques for gate leakage estimation," *Proceedings of the 2003 International Symposium on Low Power Electronics and Design (ISLPED'03)*, pp. 100–103, August 25-27 2003.

[12] K. Kanda, H. Sadaaki, and T. Sakurai, "90% write power-saving SRAM using sense-amplifying memory cell," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 6, pp. 927–933, June 2004.