



ELSEVIER

# Design and Analysis of a 32nm PVT Tolerant CMOS SRAM Cell for Low Leakage and High Stability

Sheng Lin, Yong-Bin Kim<sup>\*</sup>, Fabrizio Lombardi

*Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Ave, Boston MA 02115, USA*

**Elsevier use only:** Received date here; revised date here; accepted date here

## Abstract

A novel nine transistor (9T) CMOS SRAM cell design at 32nm feature size is presented to improve the stability, power dissipation, and delay of the conventional SRAM cell along with detailed comparisons with other designs. An optimal transistor sizing is established for the proposed 9T SRAM cell by considering stability, energy consumption, and write-ability. As a complementary hardware solution at array-level, a novel write bitline balancing technique is proposed to reduce the leakage current. By optimizing its size and employing the proposed write circuit technique, 33% power dissipation saving is achieved in memory array operation compared with the conventional 6T SRAM based design. A new metric that comprehensively captures all of these figures of merit (and denoted to as SPR), is also proposed; under this metric, the proposed 9T SRAM cell is shown to be superior to all other cell configurations found in the technical literatures. The impact of the process variations on the cell design is investigated in detail. HSPICE simulation shows that the 9T SRAM cell demonstrates an excellent tolerance to process variations comparing with the conventional SRAM cells.

*Keywords:* Memory cell design; Leakage current; Nanotechnology; Stability; Process-Voltage-Temperature variation

## 1. Introduction

Advances in chip design using CMOS technology have made it possible to design very dense chips that deliver high performance at low power consumption. To achieve these objectives, the feature size of CMOS devices has been dramatically scaled to smaller dimensions. Over the last few years, devices at 45nm have been manufactured and the deep sub-micron/nano range of 32nm is foreseen to be reached in the near future as technology continues to scale down.

Power and density have become the key limitations in many designs as nanoscale devices are becoming a reality at rapid pace. Today's high performance integrated circuits consume more than 40% of the total active mode power due to leakage currents. Furthermore, leakage is the only source of power consumption in idle circuit. For the foreseeable future, SRAM will likely remain the

embedded memory technology of choice for many microprocessors and systems on chips (SoC) due to its speed and compatibility with standard logic processes [1]. With the advent of SoC, the design of power efficient SRAM structures has become highly desirable. One of the most effective approaches to meet this objective is to design SRAM cells that operate in ultra-low power mode. The decrease in supply voltage reduces the dynamic power quadratically and the leakage power is reduced linearly to the first order [2]. However, with an aggressive scaling in technology as predicted by the Industry Technology Roadmap, substantial problems have already been encountered when the conventional six transistors (6T) SRAM cell configuration is utilized at an ultra-low power supply, because this cell shows poor stability at very small feature sizes. Moreover, the small static noise margin makes the memory susceptible to radiation and errors in data retention. SRAM cell configurations with more than six transistors to reduce leakage power and improve stability have been proposed in [3] - [5]. An 8T

<sup>\*</sup> Corresponding author. Tel.: +1-617-373-2919; fax: +1-617-373-8970; e-mail: ybk@ece.neu.edu.

cell [3] employs two more transistors to access the read bitline. Two additional transistors (thus yielding 10T cell designs) are employed in [4] [5] to reduce the leakage current. However, those previous researches still do not provide a complete viable solution for low power application on nanoscale technology due to leakage power, cell area overhead, and stability issues.

In this paper, a nine transistors (9T) SRAM cell is presented, and the presented 9T SRAM cell is more amenable than previous configurations to small feature sizes encountered in deep sub-micron/nano bulk CMOS technology. Simulations have been performed using Berkeley Predictive Technology Model at 32nm [22]. Compared with the 8T and 10T cells of [3] [4] [5], the proposed 9T scheme offers significant advantages in terms of power consumption and leakage. An optimal transistor sizing is found for the proposed 9T SRAM cell to increase stability and critical charge, thus tolerating soft errors. An innovative precharging and bitline balancing scheme for the write operation of the 9T SRAM cell is utilized for maximum standby power savings in a memory array. A novel metric that comprehensively captures all of these figures of merit is also proposed and used for comparison.

## 2. Review of SRAM cells

The conventional SRAM cell consists of 6 transistors (6T) as shown in Fig. 1. Issues regarding process variations and power supply voltages have been reported for this type of cell [4]. The conventional 6T SRAM cell has been found to be rather unstable for deep sub-micron/nano scale technology. This cell fails to meet the operational requirements due to the low read Static Noise Margin (SNM). Several configurations have been proposed to improve the SNM by adding separate read access structures to the original 6T configuration, thus making the read SNM equal to the hold SNM [3]-[5]. When the conventional SRAM cell is in the read operation, the pass gate is turned on and pulls the node that stores the logic "0" (for example, the node identified by 'qb' in Fig. 1) to a non-zero value. This decreases the read SNM, especially when a low power supply voltage is utilized. Fig. 2 shows the hold and read butterfly plot for a 6T SRAM cell at 0.6V. As shown in this figure, the read SNM is very low and is not acceptable for most memory designs.

To address the reduced read SNM problem, the read and write operations are separated by adding read access structures to the original 6T cell, thus increasing the transistor count to eight. As the read current does not significantly affect the cell value, the read stability of the 8T cell [3] is dramatically increased compared with the original 6T SRAM cell. By using this cell, the read SNM is determined by the two cross-coupled inverters. The

worst-case stability condition encountered previously in a 6T SRAM cell is avoided and a high read SNM is retained. Therefore, the 8T cell has a higher read SNM than the 6T SRAM cell. However, for the 8T structure, the read bitline leakage is significant, especially in the deep sub-micron/nano ranges. When the column is not accessed, the leakage current through the read access cell may cause a severe voltage drop at the read bitline, thus errors may appear at the output. Since it has not yet been possible to design a high-density SRAM using 8T cells, this has led to an investigation of other cell configurations such as the 10T structures in [4] [5] and the 9T structure as proposed in this paper. To prevent the leakage current from the read bitline, a PMOS transistor is added to the read access circuit in the 10T structures. However, by adding a PMOS to reduce the bitline leakage, the 10T SRAM cell suffers from a larger cell area and standby power consumption.

## 3. Proposed 9T SRAM cell

### 3.1. 9T cell scheme

The initial design of this 9T structure has been introduced in [6] [23]; Fig. 3 shows the proposed 9T SRAM cell. Similarly to the 8T cell of [3], the configuration from M1 to M6 is unchanged (same as in the 6T SRAM cell). "Write access" to the cell occurs through the write access transistors and from the write bitlines, WBL and WBLB. Read access to the cell is through the read access transistor and controlled by the read wordline, RWL. The read bitline, RBL, is precharged prior to the read access. The wordline for read is also distinct from the write wordline. The read SNM margin is maintained by retaining the write access circuit. For the 8T cell in [3], a leakage problem has been observed on the read bitline, this may cause data to change during the read operation, i.e. the bitline must be kept high, but it may drop to a low level due to the bitline leakage of other unaccessed cells. This problem limits the 8T cell to low-density applications. In the 9T cell, by adding a NMOS transistor (MN9) between MN7 and MN8, the bitline leakage is significantly reduced by the so-called "stack effect". The reduced bitline leakage makes it possible to have more SRAM cells on a bitline for high-density SRAM designs. Simulation results in [23] have shown that 512 bitcells can be connected to the same bitline for 9T SRAM cell case while only 32 8T SRAM cells can be connected to one bitline at 0.6V power supply. With more bitcells on one bitline, less peripheral circuits are needed in the 9T SRAM array design. Therefore, the area of the 9T SRAM block is reduced.

For high density memory design, the SRAM cell should be sized as small as possible. However, for correct

operation a sizing constraint is applied to the conventional 6T SRAM cell shown in Fig. 1, i.e. the pull-down to the pass gate transistors ratio must be greater than 1.2 to avoid the read-upset problem [7]. However, in the 9T SRAM cell case, when the read is enabled ( $RWL=1$ ), the read bitline (RBL) is conditionally discharged through the pull-down transistors MN7, MN9, and MN8 depending on the data stored at 'qb'. The cell node is isolated from the bitline during the read operation, and it retains the hold mode SNM. Therefore, the transistor ratio between MN3 and MN1 can be decreased to achieve better performance. As the pull-down transistors (MN3 and MN4) are the largest transistors, a significant amount of power is saved by scaling down these transistors. As the loading capacitance of the access transistors (MN1 and MN2) decreases, the write delay will also be decreased. At the same time, the write-ability will be strengthened as the ratio between the access transistor and the pull-down transistor is increased.

The power-delay product is commonly used to show the impact of decreasing the pull-down transistors to improve the performance of the SRAM cell. Fig. 4 shows the power-delay product of the 9T SRAM cell at 0.6V power supply voltage for various transistor sizes. As the transistor size of MN3 decreases, the ratio between MN3 and MN1 decreases, thus both power consumption and write delay of the SRAM cell decrease due to the reduction in size of the pull-down transistor and the loading capacitance of the access transistors. Fig. 4 also shows the relationship between the MP5/MN1 ratio and the power-delay product. When the pull-up transistor MP5 or MP6 is large in Fig. 3, 'q' or 'qb' node can change faster from "0" to "1" at the expense of large power consumption. As the size of MP5 decreases, the power consumption decreases at the expense of circuit performance. Fig. 4 shows that the least value of the power-delay product of the SRAM cell is achieved when the MP5/MN1 ratio is between 0.83 and 1.17.

Although the read-disturb problem is solved by separating the read and write operations, the half-select disturb occurs when the wordline is on (while the column select is off) in the 6T, 8T, 9T and 10T SRAM arrays. This leads to another disturb similar to the read-disturb problem. If the half-select disturb problem is not solved in the 9T SRAM array, then the size of the pull-down transistors MN3 and MN4 cannot be reduced for better power-delay product performance because the read-disturb still exists. A hardware-based approach proposed in [8] utilizes local write wordlines that are only selected when the write control for the selected block is on. Fig. 5 shows the conceptual circuit scheme and waveforms of the proposed approach. The local write wordline WWL is generated by the global write wordline signal WWLB and the block select signal BSB. The write access transistors are only accessed when the block is selected for write (BSB is low). Therefore, the local write wordlines are

only selected when the write control for the selected block is on, which avoids disturbing un-selected cells on an accessed row for a write operation. This scheme can be applied also to the 9T SRAM array to eliminate the half-select disturb problem.

Compared with the 6T differential SRAM, the 9T SRAM cell requires a single ended read port due to the separate read buffer. A single ended read operation requires a larger voltage swing at the read bitline and it provides a larger noise margin. For the 6T SRAM cell array, the differential bitlines must be precharged to  $V_{dd}$  by the starting time of clock cycle and one of the bitlines needs to be discharged in every read cycle. However, for the single ended read bitline, the bitline is discharged only when state "1" is read out. Therefore, the transient probability of the bitline is reduced to half of the conventional SRAM, which reduces the switching power of the single ended bitline. In the SRAM design with differential bitlines, the sense point is set to 50mV, which is significantly lower than the sense point of the single end bitline (half  $V_{dd}$ ). However, most cells will discharge the bitline at more than 50mV. [24] shows that the average voltage difference between two bitlines is 80% of  $V_{dd}$ . Therefore, the dynamic power of the differential SRAM, for example a 6T SRAM cell array, is higher than a single ended SRAM. The simulation results of [24] have shown that the readout power of the differential SRAM array is 25% higher than the single end SRAM array. Furthermore, as the 9T SRAM array allows more bitcells on one bitline, the area and power consumption of the peripheral circuits of the 9T SRAM array are significantly lower than for the 6T SRAM array.

The Read operation is important for high performance SRAM design. A 512-row, 128-column cell array has been designed to measure the read access time. For the 6T cell, the read access time is the time required for developing 50mV bitline differential voltage after the wordline is turned on during a read operation [25]. For the 9T cell, the read access time is the time required for discharging the bitline voltage to half  $V_{dd}$  after the wordline is turned on during a read operation. HSPICE simulation shows that the read access time of the 6T SRAM cell is 74.60ps while the read access time of the 9T SRAM cell is 82.04ps (at 0.6V power supply voltage [23]). The single ended configuration in the proposed 9T SRAM cell results in a 10% degradation of readout time.

### 3.2. Data stability and write-ability

#### 3.2.1. Static noise margin and write-ability

The pull-down transistors (MN3 and MN4) cannot be scaled down too much due to the stability considerations. Proper data retention strength is necessary for the SRAM cell, especially for the proposed 9T cell to operate at ultra-low power supply. The stability of SRAM cell is

usually represented by the SNM (the SNM is defined as the maximum value of DC noise voltage that can be tolerated by the SRAM cell without changing the stored bit [9]). The static noise margin (SNM) can be graphically found on the butterfly curve shown in Fig. 6. Its drawback is the inability to measure the SNM with automatic inline testers directly because the SNM still has to be derived by mathematical manipulation of the measured data after measuring the butterfly curves of the cell. An alternative method to characterize the SRAM stability is based on the N-curve of the cell measured by inline testers [11] [12]. The typical N-curve of the SRAM cell is measured by reading the input current at the internal storage node and by sweeping the storage node voltage from 0V to  $V_{dd}$ .

Two common metrics for the SRAM cell static noise margin are the static voltage noise margin (SVNM) and the static current noise margin (SINM), and they are found on the N-curve in Fig. 6. At three points (A, B, and C) of the N-curve, the current at the internal storage node ‘q’ is zero. A and C correspond to the two stable points of the butterfly curve, while B corresponds to the meta-stable point on the N-curve. When the points A and B coincide, then the cell is at the edge of stability. The voltage difference between the points A and B indicates the maximum tolerable DC noise voltage at node ‘qb’ prior to changing its content. This metric is referred to as SVNM.

Table 1

Noise margin comparison at 0.9V Power Supply, Slow Corner, and 100 °C Temperature

| SRAM cell                      | SNM     | SVNM  | SINM    |
|--------------------------------|---------|-------|---------|
| 6T SRAM cell<br>(MN3/MN1=1.33) | 28.3mV  | 208mV | 5.89uA  |
| 8T SRAM cell<br>(MN3/MN1=1.33) | 395.9mV | 360mV | 87.09uA |
| 9T SRAM cell<br>(MN3/MN1=1.00) | 367.7mV | 371mV | 68.15uA |

The peak current between A and B can also be used to characterize the stability of the cell. This metric is referred to as the SINM. SINM is defined as the maximum value of DC current that can be injected in the SRAM cell prior to a change in its content. The voltage is swept from 0V to  $V_{dd}$  at node ‘qb’ in Fig. 3 and the N-curves of the 9T SRAM cell at different MN3/MN1 ratios are extracted. Finally, the same voltage sweep is applied to the node ‘qb’ of the 6T cell during read operation. Fig. 7 shows the N-curves of the SRAM cells for different transistor size ratios. As shown in Fig. 7, when the MN3/MN1 ratio is 1.00, the SINM of the 9T cell is larger than the read SINM of the 6T cell (the larger SVNM of the 9T cell is retained due to the separate read and write operations). Table 1 shows the worst case noise margin

comparison between the 6T SRAM cell with MN3/MN1=1.33, the 8T SRAM cell with MN3/MN1=1.33, and the 9T SRAM cell with MN3/MN1=1. The 6T SRAM cell has very small noise margin in terms of SNM, SVNM, and SINM, while the noise margins of the 8T and 9T SRAM cells are higher even in the worse case.

The N-curve is used for assessing the write-ability of the cell, because it provides a measure for the driving current in the write operation [12]. The write-trip current (WTI) is the negative peak between the points B and C when the cell is accessed and both bitlines are kept at  $V_{dd}$ . Therefore, it specifies the amount of current required for writing, this is the current margin of the cell when its content changes. The smaller the absolute value of WTI is, the easier to write into the SRAM cell. Fig. 8 shows that the absolute value of WTI decreases as the pull-up transistor MP5 decreases, i.e. it is easier to write a “0” into the SRAM cell with smaller pull-up transistors.

A good write-ability of the SRAM cell ensures the write drivers and access transistors to overpower the load inside the cell. To increase the write-ability, the pull-up transistors MP5 and MP6 should be sized as small as possible. However, a good write-ability means that the data-holding capability of the SRAM cell degrades (possibly leading to soft errors), and this may make SRAM cells rather unstable at low power supply. The dynamic stability analysis described in the next section can be used to find the optimum pull-up transistor sizing for robust operation with respect to both good write-ability and resilience to soft errors.

### 3.2.2. Soft errors and dynamic stability

In this section, soft error tolerance and dynamic stability are analyzed for the proposed SRAM cell.

When a SRAM cell is holding data, a carbon or nitrogen atom can release a neutron from the atom's nucleus if a cosmic ray collides with oxygen. If such neutron then collides with a silicon atom, it can split the atom's nucleus into smaller charged particles. Moreover, if the silicon atom is in a semiconductor memory, the charged particles can change the contents of a memory cell, yielding a so-called “soft error.” [13] With technology scaling, the soft-error rate is expected to be significantly higher for deep submicron/nano SRAMs due to the lower  $V_{dd}$  and smaller node capacitance.

The robustness of SRAMs against soft errors can be assessed by considering the critical charge,  $Q_{crit}$  [14].  $Q_{crit}$  is the minimum amount of charge necessary to disturb the sensitive node of a SRAM cell. It exhibits an exponential relationship with the soft error rate. Therefore,  $Q_{crit}$  should be as high as possible to limit the soft error rate.

To determine the critical charge for a SRAM cell, a current generator is applied (through HSPICE) to the storage node of the SRAM cell (‘q’ and ‘qb’ in Fig. 3) as an equivalent noise source to the transient noise when the

cell is holding data. The pull-up PMOS transistor (MP5) is significantly weaker than the driver NMOS (MN3) due to the lower W/L ratio and low mobility, this makes the node storing a “1” weaker and more susceptible to soft errors than the node storing a “0”. The critical charge,  $Q_{crit}$ , is estimated only at specific nodes having a low  $Q_{crit}$ . For the 6T SRAM cell in Fig.1, the nodes ‘q’ and ‘qb’ have the same  $Q_{crit}$  due to symmetry in the structure. However, in the 9T SRAM cell proposed in Fig.3, the node with the lowest  $Q_{crit}$  is ‘q’ due to asymmetry in the 9T cell configuration.

Therefore, a current generator is applied to the node storing a “1” to determine the critical charge  $Q_{crit}$  in the worst case. The minimum total charge  $Q_{crit}$  can be found by finding the integrated current applied to the node storing a “1” to change the content of the cell. Fig. 8 shows the critical charge of the 9T SRAM cell for different MP5/MN1 ratios.

The tolerance of a SRAM cell to soft errors is enhanced by increasing the pull-up transistor size. It is important to find the pull-up transistor size by considering the conflicting constraint of write-ability and dynamic stability of the cell. As shown in Fig. 8, if a good write-ability is desired, the MP5/MN1 ratio should be as small as possible. However, if good soft error immunity is desired, the MP5/MN1 ratio should be as large as possible. However, MP5/MN1 is related to other design metrics such as power delay product, critical charge, write-trip current, and area. In this paper, to find the optimal design points considering all the design metrics mentioned above using the simulation data shown in Fig. 4 and Fig. 8, the proposed 9T SRAM cell uses the MP5/MN1 ratio of 0.67 to retain a high critical charge (compared to a cell with very weak pull-up transistors) and a low absolute value of WTI (compared to a cell with very strong pull-up transistors).

For the proposed 9T SRAM cell at 32nm feature size, this section has established the transistor size ratios (between the pull-up PMOS, the pull-down NMOS, and the access transistors) for low-power, best static and dynamic stability, and write-ability. These figures of merit (power, stability, and performance) are optimized in the proposed 9T SRAM cell when MP5/MN1 = 0.67 and MN3/MN1 = 1.00. Fig.9 shows the 9T SRAM cell with the transistor sizes found in this section.

### 3.3. Area

The layouts of the conventional 6T and 9T SRAM cells are drawn based on MOSIS deep sub-micrometer design rules [21] as shown in Fig. 9. For the 6T SRAM cell shown in Fig.1, the ratio of 1.33 is used for M3 and M1, and M1 and M5 have the same sizes as MN1 and MP5 shown in Fig.3 and Fig.9. In Fig.10, 3x2 SRAM cell arrays of 6T and 9T SRAM cells are shown to demonstrate that the SRAM cell can be integrated into an

array design. General scaling has been applied to the MOSIS rule to scale the area of the SRAM cells to 32nm feature size by a factor of  $1/S^2$ . With the scaling factor, the area of the 6T cell layout at a 32nm feature size is  $0.1899\mu\text{m}^2$ , while the area of the proposed 9T SRAM cell with optimal sizing at 32nm feature size is  $0.2331\mu\text{m}^2$ . The area of the proposed 9T SRAM cell is increased by 22% comparing to 6T SRAM cell when the previously described transistor sizing process is used. The area increase is not due to the additional read bit line but to the additional transistors. The area of the proposed 9T SRAM cell is exactly same as the 8T SRAM cell.

In the traditional differential SRAM, a large area overhead is accounted due to the differential sense amplifiers (compared with the single ended design). In the single ended design, the readout circuit only consists of an inverter and a compensation keeper [24]. Furthermore, the proposed 9T SRAM cell is designed for low-power supply and high-density. As reported in [23], both 10T cell [4] and the proposed 9T cell can be used in high-density memory designs to reduce the area overhead due to the peripheral circuitry. Compared with its 10T cell counterparts, the 9T SRAM cell has an area reduction of 16.7%, which makes the proposed 9T cell configuration viable for high-density design. Therefore, in the proposed 9T SRAM design, less peripheral circuits are required and the overall area is saved.

## 4. Standby power reduction

In a short channel device, the drain-induced barrier lowering (DIBL) has a significant impact on the subthreshold current. In a short-channel device, the source and drain depletion width in the vertical direction and the source drain potential have a strong effect on the band bending over a significant portion of the device. Consequently, the threshold voltage and the subthreshold current of short-channel devices vary with the drain bias. This effect is referred to as DIBL [15] [16]. When a high drain voltage is applied to a short-channel device, it lowers the barrier height, resulting in a further decrease of the threshold voltage. As the threshold voltage decreases, the subthreshold leakage current increases exponentially. DIBL is enhanced at high drain voltages and shorter channel lengths. When the 9T SRAM cell in Fig.3 stores “0” (this is the worst-case leakage current scenario for the 9T SRAM cell), the sub-threshold current of MN1 is very high due to the large voltage difference between WBL and node ‘q’. This leakage current can be reduced by lowering the voltage at WBL.

In a conventional 6T SRAM design, following a write operation, both bitlines must be restored to  $V_{dd}$  to ensure a successful read operation. The write amplifier circuitry of [7] ensures that the selected bitline is back to a “high” value by generating a negative pulse to precharge the

selected bitline high after driving the bitline low to write “0” into the SRAM cell. Therefore, both bitlines (WBL and WBLB in Fig. 3) will be restored to a “high” state after the write operation. When the SRAM cell stores a “0”, the voltage difference between the drain and source of MN1 is  $V_{dd}$ , i.e. a large subthreshold current from WBL to ground will be present when the SRAM cell is in a standby mode.

Table 2

Power dissipation comparison at 0.6V Power Supply and Room Temperature

| SRAM cell   | Slow Corner | Typical Corner | Fast Corner |
|---|-------------|----------------|-------------|
| 6T SRAM cell array  | 220.54nW    | 832nW          | 3.92uW      |
| 8T SRAM cell array  | 318.72nW    | 946nW          | 4.05uW      |
| 9T SRAM cell array without proposed bitline balancing circuit | 296.90nW    | 903nW          | 3.99uW      |
| 9T SRAM cell array with proposed bitline balancing circuit    | 176.26nW    | 547nW          | 2.81uW      |

The subthreshold current from WBL to ground can be reduced exponentially by lowering the voltage at WBL in the standby mode. In this paper, the 9T cell configuration is complemented with a solution at circuit-level for bitline balancing. The new “write bitline balancing” circuitry is shown in Fig. 11. Fig. 12 shows the simulation results at a power supply voltage of 0.6V. When the WR\_EN signal is enabled, the NMOS transistor NM1 is turned off, and a negative pulse is generated to precharge the bitline (that is going to be in the “high” state for a fast write). There is sufficient time to precharge the high rising bitline to a proper “high” state prior to accessing the SRAM cell because the WR\_EN signal always arrives faster than the WWL and SEL signals. After the write operation, the WR\_EN signal is removed from the column and NM1 is turned on to balance the voltage at both bitlines to half the value of  $V_{dd}$ . Therefore, the sub-threshold current from WBL to ground decreases exponentially due to the reduced  $V_{ds}$  of the access transistor MN1. Furthermore, for the “write amplifier” circuitry in [7], the bitline voltage drops in the standby mode due to the leakage in the SRAM cell and this will increase the write delay. By employing the proposed write circuitry, the bitline leakage problem is significantly reduced. Therefore, this scheme provides an efficient solution to leakage at array-level.

For comparative evaluation, a 1x128 SRAM cell array operating at power supply voltage of 0.6V is used to assess the power dissipation of 6T and 9T cell arrays in standby mode. All the simulation results are obtained by using Berkeley Predictive Technology Model at 32nm

[22]. The transistor sizes of the 6T and 9T SRAM cell are the same as discussed in Section III. As shown in Table 2, the 9T SRAM array with the proposed write bitline balancing circuit achieves a significant power reduction at the three different process corners, i.e. power savings of 33%, 20%, and 29% at typical, fast, and slow corners, respectively. Therefore, the subthreshold leakage is reduced both at cell-level and at array-level (by using the novel balancing circuitry as proposed in a previous section).

## 5. New metric for memory cells

To compare the different memory cell configurations found in the technical literature with the 9T cell of this paper, a novel metric is introduced in this section. This is required to comprehensively assess the performance (as a function of the delay), stability (with respect to noise) and power dissipation within a single yet comprehensive metric that relies on the simulation results of the previous sections.

It has been shown in [12] that SVN<sub>M</sub> and SIN<sub>M</sub> must be both used when considering the static stability of a SRAM cell. Therefore, SVN<sub>M</sub> and SIN<sub>M</sub> can be multiplied together, thus yielding a comprehensive figure of merit as the Static Power Noise Margin (SPNM), as first proposed in [12]. The Power Delay Product (PDP) is an important and often used to measure and compare the performance of the electronic circuits quantitatively. For SRAM cell, the standby power consumption is of major concern while the write delay gives an indication of the cell’s write ability. Therefore, the PDP for the SRAM cell of this paper is the product of the standby power and the write delay.

The proposed new metric combines the above figures of merit. For a high stable design, the SPNM should be as high as possible. However, for high performance and low power, the PDP should be as low as possible. The proposed new metric is given by dividing the SPNM by the PDP and this is referred to as the SPNM to PDP Ratio (SPR) [26]. The SPR (in units of 1/Secs) can be expressed as follows:

$$SPR = \frac{SVNM * SINM}{Power * Delay} \quad (1)$$

Equation (1) implies that the SPR provides a metric for high stability, high performance (low delay) and low power in memory cells. The decrease in write delay implies that it is easier to write data into the SRAM cell. Therefore, the stability of the SRAM cell is decreased. An improvement in delay performance usually requires larger transistors, thus increasing power consumption. So, it is often needed to comprehensively assess performance as a function of delay and stability with respect to noise, and use power dissipation as a single metric. The proposed

SPR provides a comprehensive metric for a SRAM cell in terms of delay, stability, and power. The proposed SPR also contains information on the cell's area. As the area of the memory cell increases, the power delay product of the memory cell will increase due to the increase of node capacitance, causing a decrease in SPR. It is important to note that a small increase in area changes little the combined power-delay product for a cell. However, as the area gets further increased, the power delay product will increase.

Simulations have been performed using Berkeley Predictive Technology Model at 32nm [22]. Fig.13 gives the SPRs of the proposed 9T SRAM cell, a conventional 6T SRAM cell, the 8T SRAM cell of [3], the 8T SRAM cell of [3] with the proposed bitline balancing circuit of Fig.11, the 9T SRAM cell of [17] and the 9T SRAM cell of [17] with the same optimized 32nm transistor sizing (as established previously) at 0.6V power supply, typical corner and room temperature. For the 8T SRAM cell, the cell size is the same as the 9T SRAM cell in Fig.3 and Fig.9 but without MN9. Among the SRAM cells, the proposed 9T SRAM cell with optimum transistor sizing and the novel write bitline balancing circuit has the best SPR, i.e. the 9T SRAM cell achieves high stability as well as low power and high performance. Compared with the proposed 9T SRAM cell, a 6T SRAM cell has low power, but also low SNM and SINM. 8T SRAM cell shows lower SPR due to its high standby power. The utilization of the proposed bitline balancing circuit to this cell configuration improves the SPR, but it is still lower than for the 9T cell. It should be noted that the 9T SRAM cell in [17] achieves a high SNM and extra low power but it has a low SINM and a low circuit speed, thus achieving a low SPR.

Fig. 13 confirms that the proposed 9T cell configuration has the best SPR, hence attaining low power, high stability, and low delay within the comprehensive metric provided by the SPR.

## 6. Impact of PVT variations

Systematic and random variations in process, supply voltage, and temperature (PVT) are posing a major challenge to nano-scale integrated circuit design. The demand for low power causes supply voltage scaling, thus making voltage variations a significant design challenge. Furthermore, the need for growth in operational frequency occurs at high junction temperatures and within die temperature variations. Fig. 14 shows the static power and write delay product for the proposed 9T SRAM cell using the write circuitry in Fig. 11 and the conventional 6T SRAM cell design of [7] with inter-die variations. Simulation results show that the 9T SRAM cell has a lower power delay product as well as less sensitivity due

to power and temperature variations because the proposed 9T SRAM cell and bitline balancing scheme reduce the leakage power up to 33% without degrading the speed of the memory cell. With the bitline balancing scheme, the subthreshold leakage of the 9T SRAM cell array is significantly reduced. The subthreshold leakage component takes a significant portion of the total power; therefore, the power delay product variation of the 9T SRAM cell is also reduced since the subthreshold leakage is very sensitive to temperature and voltage change.

As CMOS technology scales down into the nano ranges, process variations are a serious concern due to uncertainty in the device and interconnects characteristics. Process variations negatively impact the speed, stability, and power consumption of traditional SRAM designs [1]. In this section, the stability of SRAM cells is evaluated in the presence of process variations, i.e. intra-die variation and local mismatches, in the channel length and the threshold voltage of the transistors. Fig. 15 shows the power-delay product of the 6T SRAM cell and the 9T SRAM cell with variations in mobility, threshold voltage, and velocity saturation, respectively. Fig. 16 shows the power-delay product of the 6T SRAM cell and the 9T SRAM cell with variations in channel length. At 32nm feature size, channel length and threshold voltage variations have the most significant impact on cell performance while mobility and velocity saturation have the least impact. Monte Carlo analysis on the stability of the SRAM cells shows that the 6T cell could fail at 0.6V power supply, typical corner, and room temperature in the presence of process variations while the 9T cell has SNM of 240mV and SINM of 23.61uA at the worst case. Therefore, the 9T cell achieves excellent stability in the presence of process variations.

Monte Carlo simulation is usually used to investigate the process variation effect. However, at 32nm feature size, the SRAM parameter distributions are not Gaussian and Monte Carlo analysis using a Gaussian distribution is not accurate. Therefore, Importance sampling is used in this paper to increase the accuracy of the Monte Carlo analysis [27] [28]. The Importance sampling takes more samples from the tails of the parameter distributions. In this paper, a simple form of Importance sampling is used to perform process variation analysis for 9T and 6T SRAM cells at 32nm feature size. A distribution of the parameters with a higher standard deviation than the standard Gaussian distribution is used to draw more samples from the tails of the distribution [27]. The Importance sampling simulation in this paper uses a wider distribution of 8-sigma for the threshold voltage and the effective channel length of all transistors in the SRAM cells while Monte Carlo simulation uses 3-sigma distribution. The impact of the threshold voltage variations, i.e., process variations, on the power consumption of the individual SRAM cell has been assessed by the importance sampling method. The

threshold voltage distribution is modeled as a  $\pm 20\%$  Gaussian distribution with variation of threshold voltage at a  $\pm 8$ -sigma level. Simulation results are shown in Fig.17, where the power consumption of the 6T SRAM cell decreases as the average threshold voltage increases. In the 9T SRAM cell, the power consumption variation caused by the variation in the threshold voltage is significantly smaller than for the 6T SRAM cell case because the subthreshold leakage current is significantly reduced. Fig. 18 shows the distribution of the power delay product due to process variations, i.e. intra-die variation and local mismatches, including transistor geometry variations, threshold variation, and mobility variation. Unlike the power consumption Monte Carlo analysis due to process variation, more parameters need to be modeled for more accurate power delay product Monte Carlo analysis. Parameter variations are modeled using a  $\pm 20\%$  Gaussian distribution with variations at a  $\pm 8$ -sigma level for all the parameters. All power delay product results are normalized with respect to the arithmetic mean of the power delay product of the proposed 9T SRAM cell. As shown in Fig. 18, the power delay product variance of the proposed 9T SRAM cell is also lower compared to the distribution of the 6T SRAM cell. The bitline voltage is reduced during standby mode by the balancing circuit in Fig. 11. Consequently, for the 9T SRAM cell, low power consumption and a considerable mitigation for the impact of process variations are accomplished.

## 7. Conclusion

In this paper, a new 9T cell has been proposed, and this new 9T SRAM cell achieves a lower power consumption compared with 8T and 10T cells. This 9T cell is best suited for high density memory design in the deep-submicro/nano range of 32nm.

## References

- [1] B.H. Calhoun, Yu Cao, Xin Li, Ken Mai, L.T. Pileggi, R.A. Rutenbar, K.L. Shepard, "Digital circuit design challenges and opportunities in the era of nanoscale CMOS", in Proceedings of the IEEE; Volume 96, Issue 2, Feb 2008, pp. 343 – 365.
- [2] J.P. Kulkarni, K. Kim, K. Roy, "A 160 mV robust schmitt trigger based subthreshold SRAM", in Solid-State Circuits, IEEE Journal of; Volume 42, Issue 10, Oct. 2007, pp. 2303 - 2313.
- [3] L. Chang, D.M. Fried, J. Hergenrother, J.W. Sleight, R. H. Dennard, R.K. Montoye, L. Sekaric, S.J. McNab, A.W. Topol, C.D. Adams, K.W. Guarini, W. Haensch, "Stable SRAM cell design for the 32 nm node and beyond", in VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on, June 2005, pp. 128 - 129.
- [4] B.H. Calhoun and A.P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation", in Solid-State Circuits, IEEE Journal of, Volume 42, Issue 3, March 2007, pp. 680 - 688.
- [5] T. Kim, J. Liu, J. Keane, C.H. Kim, "A high-density subthreshold SRAM with data-independent bitline leakage and virtual ground

This paper has presented a new design with an optimal sizing for the 9T cell by considering the N-curve and increasing the critical charge (for tolerating soft errors). An optimal pull-down to access transistor ratio has been found by considering cell stability and performance. Critical charge and the write-trip current are utilized to determine the optimum pull-up to access transistor ratio. Along with transistor sizing/scaling, an innovative bitline balancing scheme has been utilized to reduce the leakage current at memory array-level. Hence in this paper, memory design is addressed at both cell and array levels. Simulation results have shown that this write scheme for the proposed 9T SRAM cell-based array achieves a substantial reduction in power consumption at a typical process corner compared with a conventional 6T SRAM cell based array. HSPICE simulation results confirm that the proposed scheme achieves a 33% power saving compared to the conventional SRAM array based on the 6T cell configuration. The simulation results have also confirmed that the proposed cell is highly stable and resilient to soft errors.

A new comprehensive metric for SRAM cells (denoted as SPR) has then been proposed. SPR is a composite performance measure in terms of high stability, low power and low delay. The proposed 9T SRAM cell has the best SPR among all cell configurations found in the current literature.

Finally, the impact of parameter variations on the power delay product has been analyzed; the simulation results show that the 9T SRAM cell offers superior energy and stability performance in the presence of PVT variations. In conclusion, this paper demonstrates that the novel 9T SRAM cell is a viable solution for highly dense and low power memory design at the nano range of 32nm.

- replica scheme", in Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International, Feb 2007, pp. 330 - 331.
- [6] S. Lin, YB Kim, F. Lombardi, "A 32nm SRAM design for low power and high stability", in 51st Midwest Symposium on Circuits and Systems, 2008. Aug 2008, pp. 422 - 425.
- [7] A. Chandrakasan, W.J. Bowhill, F. Fox, "Design of high-performance microprocessor circuits", IEEE Press, 2000, pp. 285 - 308.
- [8] R. Joshi, R. Houle, K. Batson, D. Rodko, P. Patel, W. Huott, R. Franch, Y. Chan, D. Plass, S. Wilson, P. Wang, "6.6+ GHz Low Vmin, read and half select disturb-free 1.2 Mb SRAM", 2007 IEEE Symposium on VLSI Circuits, June 2007, pp. 250 – 251.
- [9] E. Seevinck, F.J. List, J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells", in Solid-State Circuits, IEEE Journal of Volume 22, Issue 5, Oct 1987, pp. 748 - 754.
- [10] B.H. Calhoun and A.P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS", in Solid-State Circuits, IEEE Journal of, Volume 41, Issue 7, July 2006, pp. 1673 - 1679.
- [11] C. Wann, R. Wong, D.J. Frank, R. Mann, Shang-Bin Ko, P. Croce, D. Lea, D. Hoyniak, Yoo-Mi Lee, J. Toomey, M. Weybright, J.

- Sudijono, "SRAM cell design for stability methodology", in Proc. IEEE VLSI-TSA, April 2005, pp. 21 - 22.
- [12] E. Grossar, M. Stucchi, K. Maex, W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies", in Solid-State Circuits, IEEE Journal of, Volume 41, Issue 11, Nov. 2006, pp. 2577 - 2588.
- [13] R. Baumann, "Soft errors in advanced computer systems", in IEEE Design & Test of Computers, vol. 22, May-Jun. 2005, pp. 258-266.
- [14] Shah M. Jahinuzzaman, Mohammad Sharifkhani, Manoj Sachdev, "Investigation of process impact on soft error susceptibility of nanometric SRAMs using a compact critical charge model", in Quality Electronic Design, 2008. ISQED 2008. 9th International Symposium on, March 2008, pp. 207 - 212.
- [15] A. Agarwal, S. Mukhopadhyay, A. Raychowdhury, K. Roy, C.H. Kim, "Leakage power analysis and reduction for nanoscale circuits", in Micro, IEEE, Volume 26, Issue 2, March-April 2006, pp. 68 - 80.
- [16] K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits", in Proceedings of the IEEE, Volume 91, Issue 2, Feb 2003, pp. 305 - 327.
- [17] Zhiyu Liu, V. Kursun, "High read stability and low leakage cache memory cell", in IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007, May 2007, pp. 2774 - 2777.
- [18] R. W. Mann, W. W. Abadeer, M. J. Breitwisch, O. Bula, J. S. Brown, B. C. Colwill, P. E. Cottrell, W. G. Crocco, Jr., S. S. Furkay, M. J. Hauser, T. B. Hook, D. Hoyniak, J. M. Johnson, C. H. Lam, R. D. Mih, J. Rivard, A. Moriwaki, E. Phipps, C. S. Putnam, B. A. Rainey, J. J. Toomey, M. I. Younus, "Ultralow-power SRAM technology", in IBM J. Res & Dev., vol. 47, no. 5/6, Sep./Nov. 2003, pp. 553 - 566.
- [19] R. Venkatraman, R. Castagnetti, O. Kobozeva, F.L. Duan, A. Kamath, S.T. Sabbagh, M.A. Vilchis-Cruz, J.J. Liaw, Jyh-Cheng You, S. Ramesh, "The design, analysis, and development of highly manufacturable 6-T SRAM bitcells for SoC applications", in Electron Devices, IEEE Transactions on; Volume 52, Issue 2, Feb 2005, pp. 218 - 226.
- [20] Chung-Kuan Tsai, M. Marek-Sadowska, "Analysis of process variation's effect on SRAM's read stability", in Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on; March 2006, pp. 8.
- [21] The MOSIS service, <http://www.mosis.org/Technical/Designrules/scmos/scmos-main.html>.
- [22] Berkeley predictive technology model website, <http://www.eas.asu.edu/~ptm/>.
- [23] S. Lin, Y. B Kim, F. Lombardi, "A low leakage 9t sram cell for ultra-low power operation", in 18th ACM Great Lakes Symposium on VLSI, 2008. May 2008, pp. 123 - 126.
- [24] H. Noguchi, S. Okumura, Y. Iguchi, H. Fujiwara, Y. Morita, K. Nii, H. Kawaguchi, M. Yoshimoto, "Which is the best dual-port SRAM in 45-nm process technology? — 8T, 10T single end, and 10T differential", in Proceedings of ICICDT 2008, June 2008, pp. 55 - 58.
- [25] J.P. Kulkarni, K. Kim, K. Roy, S, "A 160 mV Robust Schmitt Trigger Based Subthreshold SRAM", in Solid-State Circuits, IEEE Journal of, Volume 42, Issue 10, Oct 2007, pp. 2303 - 2313.
- [26] S. Lin, Y. B Kim, F. Lombardi, "Design of a CNTFET-Based SRAM Cell by Dual-Chirality Selection", accepted, IEEE Transactions on Nanotechnology.
- [27] T.S. Doorn, E.J.W. ter Maten, J.A. Croon, A. Di Bucchianico, O. Wittich, "Importance Sampling Monte Carlo Simulations for Accurate Estimation of SRAM yield", in 34th European Solid-State Circuits Conference, 2008. Sept 2008, pp. 230 - 233.
- [28] R. Kanj, R. Joshi, S. Nassif, "Mixture Importance Sampling and its Application to the Analysis of SRAM Designs in the Presence of Rare Failure Events", in Proceedings of the 43rd annual Design Automation Conference, 2008. July 2008, pp. 69 - 72.

## Appendix A. Figure Captions

Fig. 1. Worst-case stability condition of the conventional 6T SRAM cell.

Fig. 2. Hold and read butterfly plot of the conventional 6T SRAM cell at power supply voltage of 0.6V.

Fig. 3. Proposed 9T SRAM cell for low-power operation

Fig. 4. Power-delay product plot of the 9T SRAM cell

Fig. 5. Local write wordline generation scheme

Fig. 6. Butterfly plot and N-curve of the SRAM cell

Fig. 7. N-curve of the 9T SRAM cell with different MN3/MN1 ratios

Fig. 8. Write-trip current and critical charge of the 9T SRAM cell

Fig. 9. Proposed 9T SRAM cell with transistor sizes in W/L (nm)

Fig. 10. Layouts of a 3x2 cell array of a) the 6T, b) the proposed 9T SRAM cells

Fig. 11. Proposed write bitline balancing circuitry

Fig. 12. Simulation results for the write operation

Fig. 13. SPRs of different SRAM cells

Fig. 14. Power-delay product of the SRAM cells at different temperatures and power supplies for a) proposed 9T cell; b) 6T cell

Fig. 15 Power-delay product of the SRAM cells with the variations of mobility, threshold voltage, and velocity saturation for a) proposed 9T cell; b) 6T cell

Fig. 16 Power-delay product of the SRAM cells with channel length variations

Fig. 17. Distribution of the SRAM cell power dissipation due to threshold voltage variations for a) 6T cell; b) proposed 9T cell

Fig. 18. Power-delay product distributions of the 6T and the proposed 9T cell due to process variations

