SRAM Cell Leakage Control Techniques for Ultra Low Power Application: A Survey

Pavankumar Bikki, Pitchai Karuppanan

Department of Electronics and Communication, Motilal Nehru National Institute of Technology, Allahabad, India
Email: bikkipavan@gmail.com

Abstract

Low power supply operation with leakage power reduction is the prime concern in modern nano-scale CMOS memory devices. In the present scenario, low leakage memory architecture becomes more challenging, as it has 30% of the total chip power consumption. Since, the SRAM cell is low in density and most of memory processing data remain stable during the data holding operation, the stored memory data are more affected by the leakage phenomena in the circuit while the device parameters are scaled down. In this survey, origins of leakage currents in a short-channel device and various leakage control techniques for ultra-low power SRAM design are discussed. A classification of these approaches made based on their key design and functions, such as biasing technique, power gating and multi-threshold techniques. Based on our survey, we summarize the merits and demerits and challenges of these techniques. This comprehensive study will be helpful to extend the further research for future implementations.

Keywords

Body Biasing, Gate Leakage, Junction Leakage, Power Gating, Multi-Threshold, SRAM Cell, Sub-Threshold Leakage

1. Introduction

To accomplish high-density chip, ultra-low power dissipation, and high performance, complementary metal oxide semiconductor (CMOS) devices have been scaled since last 30 years. As a result, the propagation delay time has been reduced by 30% per technology leading to the microprocessor performance being doubled every two years. Scaled technology has reduced the supply voltage to obtain low power consumption [1] [2] [3] [4]. Additionally, a scaled technology also has reduced device parameters such as threshold voltage, channel length and gate oxide thickness. However, the scaled technology has two drawbacks.
First, a low-$V_{TH}$ device has an exponential increase in sub-threshold leakage. Sub-threshold leakage rises by ten times for every 0.1-volt decrease of the threshold voltage [5] [6] [7]. The second problem is the reduction of worst-case performance due to threshold variation at lower supplies [8] [9] [10]. As technology scales down, leakage current in a sub-micron region becomes more significant and is comparable with the dynamic power dissipation. Figure 1(a) shows the full chip leakage power dissipation based on the international technology roadmap for semiconductor (ITRS) [11] [12]. Various components affecting the sub-threshold leakage, gate leakage, and junction leakage are depicted in Figure 1(b). However, finding and modelling of the several leakage mechanisms are essential for evaluation and minimization of leakage current for low power application [13] [14].

![Figure 1](image_url)

**Figure 1.** (a) Full chip power dissipation based on ITRS. (b) Leakage currents components.
Generally, device non-conducting current ($I_{off}$) depends on the supply voltage, threshold voltage, length of the channel, surface/channel doping profile, drain/source junction depth and gate oxide thickness [15]. For long channel devices $I_{off}$ mainly originates from the drain-source reverse bias junctions. Short-channel device needs low power supply in order to reduce power dissipation [16] [17]. Hence, the reduced threshold voltage causes exponential increase in $I_{off}$ current due to the weak-inversion region [18] [19].

A conventional 6T SRAM cell consists of two inverters connected back to back and two access NMOS transistors as shown in Figure 2(a). The SRAM cell leakage versus technology scaling is shown in Figure 2(b). It shows that the

![Figure 2.](image)

(a)

(b)

**Figure 2.** (a) Leakage phenomena in basic SRAM cell (b) SRAM Leakage current versus technology node [24].
leakage current exponential increased with technology scaling below 90 nm. Sub-threshold leakage arises whenever the transistor gate voltage is zero and the drain-source voltage is non-zero. Hence, leakage current can arise within the bit cell or on the bit line access paths [20] [21]. The transistor leakage in SRAM cell essentially depends on 1) the data value stored in the cell, 2) the logic data level of the word line and 3) the type of operation being performed. Bit line leakage current is a very small amount as compared to total cell leakage. However, the bit line leakage is a very important factor, as the bit line leakage can affect the constancy of the memory stored in the cell [22] [23]. The total leakage current is given by

\[
I_{\text{total-leak}} = I_{\text{sub-th}} + I_{\text{gate}} + I_{\text{junction}}
\]

(1)

In order to bring the classification of leakage minimization approaches, we analyzed based on their fundamental design and mechanism. A brief summary of different leakage control schemes with their merits and demerits along with the limitations by using these schemes is presented. This paper will be helpful for researches to work towards emerging power-efficient memory designs for ultra-low power applications. The rest of the paper is organized as follows. Section 2 presents the origin of leakage current in a short-channel device. Various biasing techniques for leakage control SRAM are discussed in Section 3. Emerging power gating techniques for low power SRAM designs are presented in Section 4. Asymmetrical SRAM designs with multi-threshold transistor are described and comparisons of various low power techniques are tabulated in Section 5. Finally, the survey paper concludes in Section 6.

2. Leakage Currents

Basic leakage currents in a MOS transistor are as shown in Figure 3. Three types of leakage currents are presented when the device channel is in a non-conducting state, which are sub-threshold leakage current \( I_{\text{sub-th}} \), gate-induced drain leakage \( I_{\text{GIDL}} \) and punch-through leakage \( I_{\text{Punch}} \) [25]. Two other leakage currents that are independent of the conduction of the device are gate tunneling current \( I_{\text{tunnel}} \) and pn junction reverse bias leakage \( I_{\text{junction}} \) due to a band to band tunneling effect [26] [27].
2.1. Sub-Threshold or Weak Inversion Current

Sub-threshold or weak inversion current is present between the source and the drain terminals of the MOSFET transistor when the gate voltage is less than the threshold voltage of the device \((V_{GS} < V_{TH})\). Weak inversion region has small minority carrier concentration and it varies along with the length of the channel. The drift current dominates in the strong inversion region and the diffusion current dominates in weak inversion region [28] [29] [30]. Let us assume that the source of the NMOS is zero \((V_{G} < V_{TH})\) and the drain to source voltage is greater than 1 V. In this case, weak inversion current at \(V_{DS}\) falls completely due to the existence of reverse-biased subtract-drain pn junction [10] [31]. The NMOS transistor Gate-source voltage versus drain current has presented in Figure 4, which represents the sub-threshold and junction currents when the threshold voltage is 0.6 V. Drain current for NMOS transistor function in weak-inversion region is given by

\[
I_{DS}^{(\text{weak inversion})} = I_{on} \cdot e^{(V_{GS}/V_t)}
\]

where \(\eta\) is sub-threshold slope coefficient; \(V_t\) is the thermal voltage equal to \(q/\sqrt{K}\); and \(I_{on}\) is \(\mu_c C_{ox} (W/L) V_t^2 e^{-1.8}\).

2.2. Gate-Induced Drain Leakage

Gate-induced drain leakage (GIDL) occurs due to band to band tunneling at gate-drain overlap region under the strong electric field. Electrical field increases with the decreasing the depletion layer at the surface. The GIDL is more in case of single-diffused drain as compared to double-diffused drain. At low subtract voltage the minority carrier accumulated into drain depletion region under the gate region that forms the path for the GIDL current [32] [33]. Further the GIDL current increases exponential with thinner gate oxide thickness, low \(V_{TH}\) and the higher potential difference between gate and drain [34] [35]. A simple mathematical analysis is presented for the effect of gate work function in the electric

![Figure 4](image_url). Gate-source voltage vs. drain current.
field at the gate-drain overlap in Equation (3).

\[
E_{\text{Total}}^2 = \frac{V_{DG}}{T_{OX}} \left( 1 - \frac{V_{FB} - \psi_S}{V_{DG}} \right)^2 + \frac{V_{DG}^2}{\gamma h}
\]  

(3)

Based on the tunneling theory, the \(I_{\text{GIDL}}\) current can be written as

\[
I_{\text{GIDL}} = A E_{\text{GIDL}}^2 \exp\left( -\frac{B}{E_{\text{GIDL}}} \right)
\]

(4)

where \(\psi_S\) is the potential drop across the silicon for BTBT; \(h\) is the parameter interrelated to junction depth, \(E_{\text{GIDL}}\) the electric field accountable for \(GIDL\) current; \(A\) and \(B\) are constants that aid in tunneling \[36\] \[37\].

2.3. Punch-Through Current

Punch-through current occurs in small-geometry MOS transistors due to the nearness of the source and drain depletion regions as these junctions spread into the short-channel (i.e., when \(x_{ds} + x_{ad} = L\)). Since the short-channel doping concentration remains constant as a result, the boundaries between the depletion regions are reduced \[38\] \[39\]. In submicron technology, the surface region was highly doped as compared to the substrate and the depletion region extended towards the substrate which rises the punch-through current below the surface \[40\] \[41\]. Punch-through current can be reduced with higher substrate doping, small oxides, modest junctions and preferably with long channels.

In short-channel devices, the potential barrier is governed by both the drain-to-source voltage \(V_{DS}\) and the gate-to-source voltage \(V_{GS}\). When the drain voltage is increased, it decreases the potential barrier in the short channel, leading to drain-induced barrier lowering (DIBL). Carriers are injected into the channel surface from the source region independent of the gate voltage \[42\]. The drain and source junction widths are expressed as

\[
x_{nd} = \sqrt{\frac{2e_{\text{Si}}}{qN_j}} \left( V_{DS} + \varnothing_{as} + V_{SS} \right)
\]

(5)

\[
x_{ds} = \sqrt{\frac{2e_{\text{Si}}}{qN_j}} \left( \varnothing_{as} + V_{DS} \right)
\]

(6)

The sub-threshold surface diffusion current \(I_{\text{Sdiff}}\) for short channel at its saturation level \(\left| V_{ds} \right| > \frac{4kT}{q} \approx 0.1V\) can be expressed as \[43\].

\[
I_{\text{Sdiff}} \propto D n_i^2 e^{\left( \varnothing_{ss} \right)} | I_{\text{eff}} |
\]

(7)

\[
\Delta \varnothing_s = \Delta \varnothing_{ss} \left( V_{gs}, V_{hs} \right) + m \left| V_{ds} \right|
\]

(8)

where \(D\) is the surface diffusion constant of minority carriers; \(n_i\) is the intrinsic carrier concentration; \(\Delta \varnothing_{ss}\) is the surface-band-bending (a function of \(V_{gs}, V_{hs}\)) and \(m\) is a dimensionless constant.

Two other leakage currents that are independent of the conduction of the de-
vice are gate tunneling current and pn junction reverse bias leakage due to a band to band tunneling effect ($I_{BTBT}$) [26] [27].

2.4. Gate Tunneling Current

Gate tunneling current occurs due to the high electric field formed across the small gate oxide layer. Gate tunneling current depends on the device structure and biasing conditions. At high electric field, tunneling of electrons take place from gate to bulk and also from bulk to gate region through the gate oxide layer (i.e the quantum-mechanical wave function of a charged carrier [44]). Highly charged electrons can easily enter into or through the oxide layer due to small width of potential barrier. It increases the gate current [45] [46]. On the other hand, if the gate voltage is less than zero ($V_g < 0$) the charged electrons in $n$-poly-silicon can easily tunnel into or through the gate oxide and form a gate current. Gate tunneling current depends on the device structure and biasing conditions. Gate tunneling current in a scaled device contains a gate-to-channel current ($I_{gc}$) and edge direct tunneling currents ($I_{gso}$ and $I_{gdo}$) [47] [48].

\[ I_g = I_{gc} + I_{gso} + I_{gdo} \]  

(9)

Gate tunneling current is divided into mainly two types, Fowler-Nordheim (FN) tunneling and direct tunneling. In FN tunneling mechanism, electrons tunnel straight into the conduction band through the forbidden band gap of the oxide layer [49] [50]. The direct tunneling current form between the source-drain extension and the gate overlap. Direct tunneling occurs mainly due to the electrons and holes tunneling, the electron tunnel from the conduction band and the valence band while the holes tunnel from valence band. It is more sensitive to the gate oxide thickness [51]. The FN current expression represents the tunneling through the triangular potential barrier and is valid for $V_{ox} > \phi_{ox}$, where $V_{ox}$ denotes the voltage drop across the oxide. The current density of FN tunneling is expressed as

\[ J_{FN} = \frac{q^3 E_{ox}^2}{16\pi^2 h \phi_{ox}} \exp \left( \frac{-4\sqrt{2m\phi_{ox}}^{3/2}}{3hqE_{ox}} \right) \]  

(10)

where $E_{ox}$ is electric field across the gate oxide; $\phi_{ox}$ is the potential barrier height for electrons and $m$ is the mass of an electron in the conduction band [52] [53].

2.5. Junction Leakage

Generally, body-to-drain and body-to-source junction are in reverse bias, which produces a pn junction reverse bias leakage current. In small-geometry devices, highly doped $p$ and $n$ regions cause a band-to-band tunneling (BTBT) leakage. The pn junction reverse bias leakage current primarily originates from minority carrier diffusion and drift current near the depletion region. Secondly, in reverse biased pn junction, the depletion region generates electron-hole pair. The reverse bias pn junction leakage current is mainly dependent on the doping concentration and depletion junction area [54] [55]. The BTBT current density is
given by

\[ J_{B-B} = A \frac{E_{V_{app}}}{E_g^{1/2}} \exp \left( -B \frac{E_g^{1/2}}{E} \right) \]  \hspace{1cm} (11)

where \( m \) is the mass of an electron; \( E_g \) is the energy-band gap; \( V_{app} \) is the applied reverse bias; \( E \) is the electric field at the junction; \( q \) is the electronic charge; and \( h \) is \( 1/2\pi \) times Planck’s constant.

3. Biasing Techniques

Generally the body terminal of the transistors is connected to the supply voltage (\( V_{DD} \) in the case of PMOS and GND in the case of NMOS) in order to control a threshold voltage. In the case of reverse body bias (RBB), the body voltage of PMOS is greater than the \( V_{DD} \) and the body voltage of NMOS is less than \( V_{GND} \). This results in an increase in depletion layer width. The higher gate voltage is required to create an inversion layer by increasing the threshold voltage [56] [57] [58]. Similarly, in forward body bias (FBB) the body voltage of PMOS is less than the \( V_{DD} \) and the body voltage of NMOS is greater than \( V_{GND} \). As a result, the depletion layer width is reduced. The smaller gate voltage creates the inversion layer thereby decreasing threshold voltage. Consequently FBB has fast transition [59] [60]. Analysis of NMOS body bias versus threshold voltage and leakage current is as shown in Figure 5(a) and Figure 5(b). PMOS body bias versus threshold voltage and leakage current is as shown in Figure 5(c) and Figure 5(d). These analysis shows that the FBB has lower \(-V_{TH}\), it increase the leakage current exponential and RBB has higher \(-V_{TH}\) it reduce the leakage current. Various biasing ideas to design high \(-V_{TH}\) PMOS and NMOS transistors given in Table 1 and the arrows represent the sub-threshold current. In bulk MOSFETs, the threshold voltage (\( V_{TH} \)) is given by

\[ V_{TH} = V_{TH0} + \gamma \left( \sqrt{2\phi_b - V_{BS}} - \sqrt{2\phi_i} \right) \]  \hspace{1cm} (12)

\[ \gamma = \frac{t_{ox}}{E_{ox}} \sqrt{2e_{si}qN_A} \]  \hspace{1cm} (13)

\[ \phi_b = \frac{kT}{q} \ln \left( \frac{N_d}{N_i} \right) \]  \hspace{1cm} (14)

where \( V_{TH} \) is the threshold voltage of the device without body bias; \( \gamma \) is the coefficient of body effect; \( t_{ox} \) is the gate oxide thickness; \( E_{ox} \) is the dielectric constant of the silicon dioxide; \( e_{si} \) is the permittivity of silicon; \( \phi_b \) is the device surface potential on strong inversion; \( N_d \) is the doping concentration density of the body; \( N_i \) is the carrier concentration in intrinsic silicon; \( k \) is Boltzmann’s constant; \( q \) is the electric charge and \( T \) is the absolute temperature.

3.1. Reverse Body Bias (RBB)

Since the mid-1970s, RBB has been widely used in memory cells, in order to reduce the latch-up problem and memory data damage [62] [63]. RBB reduces the
Figure 5. NMOS body voltage versus (a) Threshold voltage (b) Leakage current. PMOS body voltage versus (c) Threshold voltage (d) Leakage current.
Table 1. Biasing methods for high-$V_{TH}$ PMOS and NMOS [61].

<table>
<thead>
<tr>
<th>Control voltage(s)</th>
<th>NMOS</th>
<th>PMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_G$ reverse biasing</td>
<td>$V_{DD}$</td>
<td>$V_{DD}$</td>
</tr>
<tr>
<td>$V_{GS}$</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>$V_{GS}$ reverse biasing</td>
<td>$V_{DD}$</td>
<td>$V_{DD}$</td>
</tr>
<tr>
<td>$V_{GS}$ offset gate driving</td>
<td>$V_{DD}$</td>
<td>$V_{DD}$</td>
</tr>
<tr>
<td>$V_{GS}$ offset source driving</td>
<td>$V_{DD}$</td>
<td>$V_{DD}$</td>
</tr>
<tr>
<td>$V_{TH}$ reduction</td>
<td>$V_{DD}$</td>
<td>$V_{DD}$</td>
</tr>
</tbody>
</table>

sub-threshold leakage while increasing the body voltage. Hence, an optimized RBB reduces the total leakage current. However, the junction leakage is influenced by substrate BTBT rather than the surface BTBT which is also called GIDL [64]. Novel techniques are required to reduce the substrate BTBT leakage in order to use the advantage of RBB. Furthermore, RBB is ineffective for leakage control in short channel devices and also for low-$V_{TH}$ devices functions at high/room temperature [65] [66].

3.2. Forward Body Bias (FBB)

Reverse body bias has larger drain-substrate depletion layer to minimizes the short channel effects. However, threshold variation occurs across a die due to larger depletion layer. Moreover, the short channel devices have lower body coefficient ($\gamma$), and the channel potential is more affected by drain than by the body due to the DIBL. The short channel effect and DIBL are more sensitive to the low-$V_{TH}$ transistors [67]. Hence the range of body biasing is motivated from RBB to FBB. FBB reduces the threshold voltage of high-$V_{TH}$ devices and improves the circuit performance due to smaller switching capacitance. The FBB devices form larger junction capacitance due to reduced depletion width across the source and drain region [68] [69].
However, FBB increases the leakage current due to a source to a body junction is in forward bias. At lower- $V_{\text{TH}}$ the short circuit current is increase due to larger junction and gate capacitance. Hence, for optimized design operation at a maximum temperature of 110°C, the desired forward body voltage is 450 mV, with a tolerance of ±50 mV [70].

3.3. Dual-$V_{\text{TH}}$ Technology

Further improvements in short channel and DIBL can be achieved through dual-$V_{\text{TH}}$ technology. A dual-$V_{\text{TH}}$ device switches between low-$V_{\text{TH}}$ and high-$V_{\text{TH}}$. The low-$V_{\text{TH}}$ device can be formed by applying FBB and high-$V_{\text{TH}}$ device with zero body steps. The complexity associated with a dual-$V_{\text{TH}}$ process is reduced because it has a critical fabrication masking process. Need for additional device fabrication with various threshold voltages are eliminated [71].

3.4. VTCMOS Technique

In variable threshold CMOS (VTCMOS), a threshold voltage is controlled by the body biasing technique. Different threshold voltages can be achieved by using a self-body biasing transistor (SBT). In active mode, nearly zero body bias is used to achieve high speed. In standby mode, RBB is used to raise threshold voltage, thus minimizes the leakage current. Moreover, in active mode, a little FBB is applied to increase the performance speed while minimizing the short-channel effect. In VTCMOS, additional circuitry is required for body biasing which increases chip area [72] [73].

3.5. Clamping Diode

Lijun Zhang et al. proposed a source biasing technique to reduce leakage currents in SRAM cell, as shown in Figure 6. High threshold NMOS transistor M7
is connected between SRAM source line \( V_{SL} \) and ground terminals. In active mode, M7 will turn on when gate voltage is high and its resistance is small, \( V_{SL} \) is equal to the ground and SRAM cell functions in the traditional manner. In standby mode, M7 is turned off, the source voltage \( V_{SL} \) increases which reduces the gate and sub-threshold leakage currents. The problem of floating voltage can be resolved by a clamping diode (PMOS transistor gate connected to its drain terminal) which is placed in parallel with the M7 [74].

### 3.6. Stacking Technique

The NMOS stacking transistors are realized with 16 nm, node voltage and leakage current as shown in Figure 7(a) and Figure 7(b). The stacking transistors provide the self-reverse biasing effect:

- Drain-to-source voltage decreases, leading to reduced DIBL current and the sub-threshold leakage current.
- Gate-to-source voltage is less than zero, thus decreasing sub-threshold current exponentially.

![Figure 7](image)

**Figure 7.** (a) Stacked transistor (b) leakage current curve.
Substrate-to-source voltage is negative, thus increasing the threshold voltage due to body effect, resulting in a decrease in sub-threshold current.

N. K. Shukla et al. proposed a novel P4-SRAM cell consisting of a stacked transistor as shown in Figure 8. When the word line is low, the cell functions as conventional 6T SRAM cell. When the word line is high, transistor P3 and P4 will turn off as a result, the self-reverse biasing (SRB) of series connected transistors will reduce the sub-threshold and gate leakage currents [75] [76].

4. Power Gating Techniques

Another efficient leakage reduction technique generally used in industry is power gating. In power gating scheme, the leakage currents are almost minimized by introducing the external header and footer transistors. These transistors eliminate the path exiting between $V_{DD}$ to ground when the devices are in quiescent mode [77] [78].

4.1. Power Gating Scheme with a Sleep Transistor

To reduce the leakage mechanism in SRAM cell, M. Powell et al. proposed a power gating scheme with a sleep transistor connected in the ground path called Gated-$V_{DD}$ as shown in Figure 9. This design basic principle is to introduce an additional NMOS transistor in the leakage current path present in between the power supply to a ground [79] [80]. The additional transistor is fundamentally “gating” the cell’s power supply by “switch on” in the active mode and “switch off” in an idle mode. Gated-$V_{DD}$ scheme significantly reduces the leakage current by maintaining the performance compensation of low power supply and threshold voltage [81] [82].

![Figure 8. P4-SRAM cell design.](image-url)
4.2. Data Retention Gated-Ground (DRG)

The drawback of the SRAM cell with power gating scheme is the determination of the stored data and the ground gated transistor increases the pull-down (PD) path resistance. To resolve this problem Amit Agarwal et al. proposed a data retention gated-ground (DRG) SRAM cell design to achieve low leakage power, as shown in Figure 10. The ground gated transistor is controlled by the external signal which is connected to the word-line. The leakage current significantly reduced through fundamentally “gating” the cell’s power supply by “switch on” in the active mode and “switch off” in an idle mode [83].

4.3. N-Control SRAM Cell with Gated-V$_{DD}$

Praveen Elakkumanan et al. proposed an N-control SRAM cell with gated-V$_{DD}$ and dual threshold voltage to achieve more reduction in leakage power dissipation, as shown in Figure 11. The SRAM cell is designed with low threshold transistor to attain high speed. Additional sleep transistors are designed with high-V$_{TH}$ transistor to minimize leakage power, high-V$_{TH}$ transistor are show with doted circles [84].

4.4. Diode-Connected PMOS Bias Transistor

Ankur Goel et al. proposed a power gating scheme along with post-silicon trim of the SRAM cell, as shown in Figure 12(a). This technique provides many ways to trim the source voltage across SRAM cell ranging from 50 to 150 mV. Sector-based power gating has been presented which allows the leakage current reduction while SRAM is in active mode [85]. The PMOS transistor acts as a diode by connecting the source and gate terminal together. This process is termed as the self-biasing technique and it controls the virtual ground [86].

However, the self-biasing technique used in nanometre technology suffers from three problems. First, an additional self-biasing transistor (SBT) requires
**Figure 10.** DRG SRAM cell design.

**Figure 11.** NC-SRAM cell design.

**Figure 12.** The SRAM cell with (a) diode-connected PMOS bias transistor (b) programmable bias transistor (c) active feedback with Op-Amp-based control.
extra area which is around 5%. Second, these schemes fail to carry out the correct mechanism for the modified source voltage. Third, increasing the virtual ground voltage has serious consequences such as negative bias temperature instability (NBTI)/positive bias temperature instability (PBTI) [87] [88] [89].

4.5. Programmable Bias Transistor

Kevin Zhang et al. proposed a programmable bias transistor (PBT) to control the SRAM virtual ground in standby mode, as shown in Figure 12(b). PBT design is more effective in solving the post-silicon trimming issues and it has good control in a virtual ground. PBT provides two main advantages. First, the transistor biasing settings (virtual ground voltage) can be adjusted to achieve maximum leakage current reduction. Second, various biasing settings can be randomly chosen at various power supplies to provide an effective design under dissimilar voltage environments. However, they have some issues regarding the die-to-die (D2D), within die (WID) and induced temperature variations [90].

4.6. Active Feedback with Op-Amp-Based Control

To overcome the variations in a die, another efficient technique has been proposed with an active feedback op-amp, as shown in Figure 12(c). A sleep transistor with active feedback op-amp bias reduces the standby power of the last level caches (LLCs) under all circumstances throughout the lifespan of the processor [91]. The main disadvantage of this technique is the DC power consumed by the op-amp that needs to be replicated along with each data. Further, this technique increases the SRAM cell area and it is noticeable in CMOS nanometre technology.

4.7. Diode and on Transistor Interchangeable Technique

Suhwan Kim et al. proposed a novel design which consists of “on transistor” and “power gating diode”. This design is termed as “diode and on transistor interchangeable technique”. This technique functions in two modes, light sleep mode and shutdown mode. In light sleep mode SRAM cell, data will be retained and voltage around SRAM memory must be greater than the minimum data retention voltage. It is also referred to as leakage saving mode which is shown in Figure 13(a). In shutdown mode, SRAM memory data need not necessarily be recalled and the virtual ground level will be raised to attain enhanced leakage power saving. The shutdown mode design is as shown in Figure 13(b), which has modified control logic transistor. The modified design can be operated in both light sleep and shutdown modes [92].

5. Multi-Threshold Design

Generally, SRAM cell contains six transistors with symmetrical configuration and normally they maintain same threshold voltage. MTCMOS is one of the prominent topologies to minimize the leakage power [93] [94]. By using an appropriate multi-threshold transistor result in a low leakage SRAM cell without
degrading the read/write performance. A suitable solution for having the low power SRAM cell with high performance and constancy is to implement asymmetrical designs [95] [96].

5.1. Asymmetrical Design

Asymmetrical cells can be designed in two different methods. The first approach is to have all original 6T-cell with different threshold voltages for selected transistors. The second approach is to add an extra transistor to the traditional 6T-cell. Asymmetrical SRAM cells designs are mainly based on the state of the cell, choose a state “0” or “1” and change the necessary transistor with high- $V_{TH}$ transistors in order to minimize the leakage current. N. Azizi et al. proposed several designs of asymmetrical SRAM cells, with main focus on state “0”, as it plays important role in storing the data. These architectures reduce the leakage current in the “0” state and probably they do not affect the “1” state.

5.2. Basic Asymmetric (BA)

The first design is named as Basic Asymmetric (BA) as shown in Figure 14(a). In this design N1, N4, and P2 transistors are replaced with high- $V_{TH}$ transistors, which is shown in doted circles. This SRAM cell reduces leakage power by 70X when the cell stores “0” and maintains the same leakage in case “1” is stored. However, the read access time is reduced (the bit-line discharge time is increased w.r.t traditional cell) due to high-$V_{TH}$ transistors (N1 and N4). P1 and P2 transistor do not affect the read access time since the bit lines are discharged through two NMOS transistors.

An asymmetrical SRAM cell transistor P1 is modified with high-$V_{TH}$ as shown in Figure 14(b). This design reduces the leakage current by 1.6X when the cell stored data is “1” as compared to traditional design and the BA design. It reduces the leakage power reduction by 70X when the cell stores data “0”. As another modification, transistor N2 is designed with high-$V_{TH}$ as shown in Figure 14(c).
Asymmetric cells have different discharge timings for bit-line and bit-line-bar. In this case, sense amplifier should match the read access time of slow side to the fast access side. This design further minimizes the leakage power by $7X$ when cell stored data “1”. To reduce the read access time delay, an asymmetrical SRAM cell needs to be modified.

The special precharge SRAM cell is as shown in Figure 14(d), where only P1, N2 transistors are replaced with high- $V_{TH}$ and the remaining with low- $V_{TH}$ transistors. Initially bit-line is precharge to $V_{DD}$ but in steady state bit-line needs to be kept at ground. This design attains leakage reduction $83X$ when the cell is stores data “0” and no reduction in leakage when cell stored data “1”. The bit-line times decreased by $12X$ as compared to traditional cell [97] [98].

### 5.3. Effect of MTCMOS on RBL Sensing

A T Do et al. have proposed a decoupled 8T SRAM cell to achieve larger bit-line data sensing margin which is used in ultra-low power SRAM cells due to enhanced stability. The read word line performs the read operation by enabling N7 and the read bit lines are conditionally discharged according to the SRAM cell data. However, due to aggressive scaling of device parameters, the read bit-line leakage current increase which is almost comparable with the SRAM cell read current. The read bit-line data discharging depends on the cell read access current and read bit-line leakage current [99] [100]. In order to overcome these limitations, a leakage control bit-line scheme was proposed in read access path...
with a high-\(V_{TH}\) and low-\(V_{TH}\) transistors which are shown in Figure 15. In the super-threshold region, high-\(V_{TH}\) transistors design achieves good \(I_{on}\) and \(I_{off}\) current ratio. Low-\(V_{TH}\) transistor design had larger read bit-line swing along with smaller read bit-line data sensing margin due to sub-threshold leakage current [101] [102].

5.4. Bit Line Boosting Current Scheme

Bo Wang et al. proposed a 9T-MTCMOS SRAM cell, which consists of additional 3T read access bit-line, as shown in Figure 16(a). This design reduces the leakage current while increasing read bit-line sensing margin. A 9T SRAM cell consists of a basic 6T cell and the read access port contains 3T NMOS. All transistors in 6T-SRAM cell are designed with high-\(V_{TH}\) which are shown in thick lines and the read access bit-line 3T NMOS are designed with low-\(V_{TH}\) to achieve very high performance through improved bit-line sensing margin. However, the cell bit line leakage is more than the cell read access current at higher temperatures [103] [104].

A. Teman et al. have proposed a 9T SRAM cell with a unique structure of supply feedback methodology for low voltage operation. In this 9T SRAM design, the supply voltage feedback transistor gate terminal is connected to the latch output to weaken the pull-up network of SRAM cell in a write operation, as shown in a Figure 16(b). The main advantage of this feedback is to maintaining

![Figure 15. RBL sensing (a) high-\(V_{TH}\) (b) low-\(V_{TH}\)](image)
improved write margins and functionality, even though the PMOS transistors are stronger than the NMOS transistors [105] [106].

6. Conclusion

In this paper, the origin of leakage currents in a short-channel device has been discussed when a device is in conduction and non-conduction state. The various leakage control techniques proposed for low power SRAM cell architecture are presented. Classification of leakage minimization approaches analyzed based on their fundamental design and mechanism, such as biasing techniques, power gating techniques and multi-threshold techniques. The biasing techniques focus on changing the threshold voltage in order to control sub-threshold leakages. RBB design has reduced the leakage but it has consequences in terms of performance. Optimized FBB design can help to achieve high performance with low power dissipation. The key emphasis of power gating technique is on leakage path that exists in the circuit. Extra circuitry is added to create the virtual ground in leakage path to control the leakage currents. The multi-threshold design has the concept of asymmetrical design with a high-$V_{TH}$ transistor in leakage path and a low-$V_{TH}$ transistor for fast access. All these techniques have
Table 2. Comparison of various low power techniques.

<table>
<thead>
<tr>
<th>Technique</th>
<th>No. of MOS’s</th>
<th>Merits</th>
<th>Demerits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse body biasing</td>
<td>6T</td>
<td>$I_{\text{sub-th}} \downarrow$</td>
<td>$I_{\text{func}} \uparrow$, Large transition delay</td>
<td>Sensitive to $V_{TH}$ variation &amp; GIDL</td>
</tr>
<tr>
<td>Forward body bias</td>
<td>6T</td>
<td>Small switching capacitance, speed ↑</td>
<td>Process complexity, Area overhead</td>
<td>Sensitive to SCE &amp; DIBL</td>
</tr>
<tr>
<td>Source-biasing</td>
<td>6T</td>
<td>$I_{\text{sub-th}}$, $I_{\text{Gate}}$ ↓</td>
<td>Delay penalty ($\tau_{PD}$ ↑)</td>
<td>Impact on soft error rate (SER)</td>
</tr>
<tr>
<td>Dual-$V_{TH}$</td>
<td>6T</td>
<td>Improved short-channel, DIBL</td>
<td>Switching</td>
<td>Need of FBB</td>
</tr>
<tr>
<td>VTCMOS</td>
<td>6T</td>
<td>Low leakage, high speed</td>
<td>Additional circuitry need to control the threshold</td>
<td>Need to control the body bias</td>
</tr>
<tr>
<td>Clamping diode</td>
<td>8T</td>
<td>$I_{\text{sub-th}}$, $I_{\text{Gate}}$ ↓</td>
<td>Delay penalty ($\tau_{PD}$ ↑)</td>
<td>Floating voltage at $V_{SL}$</td>
</tr>
<tr>
<td>Stacking body biasing (SRB)</td>
<td>8T</td>
<td>$I_{\text{sub-th}}$, $I_{\text{Gate}}$ ↓</td>
<td>Delay increase, Area overhead</td>
<td>$V_{TH}$↑ due to body effect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power gating techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground gated, sleep transistor</td>
</tr>
<tr>
<td>Data retention gated-ground</td>
</tr>
<tr>
<td>N-control with gated-$V_{TH}$</td>
</tr>
<tr>
<td>Diode-connected</td>
</tr>
<tr>
<td>Programmable bias transistor</td>
</tr>
<tr>
<td>Active feedback op-amp</td>
</tr>
<tr>
<td>Light sleep mode</td>
</tr>
<tr>
<td>Shutdown mode design</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multi-threshold design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Asymmetric</td>
</tr>
<tr>
<td>High-$V_{TH}$ on RBL Sensing</td>
</tr>
<tr>
<td>Low-$V_{TH}$ on RBL Sensing</td>
</tr>
<tr>
<td>Bit-line boosting</td>
</tr>
<tr>
<td>Supply Feedback approach</td>
</tr>
</tbody>
</table>
the single objective of reducing the leakage in the nano-scale era. Most of these techniques focused on the sub-threshold leakage minimization. However, some of the techniques emphasize on retaining the data during the standby mode. A brief summary of different leakage control schemes with their merits and demerits along with the limitations by using these schemes is presented in Table 2.

The low power SRAM cell designs presented in this paper will be helpful for researchers to work towards emerging power-efficient memory designs for ultra-low power applications. We presented only three main techniques for a basic SRAM cell. Simulation analysis has been shown for basic leakage currents in a device; these are the limitations of this paper.

References

https://doi.org/10.1007/BF01250732


https://doi.org/10.1023/A:1008373903657


https://doi.org/10.1109/TVLSI.2012.2198678

https://doi.org/10.1109/TVLSI.2008.2005413


https://doi.org/10.1145/313817.313908

https://doi.org/10.1109/TVLSI.2003.821549


Components of Shallow Silicided p+/-n Junction for Quarter- and Sub-Quarter-Micron MOSFET’s. IEEE Transactions on Electron Devices, 45, 1848-1850. https://doi.org/10.1109/16.704389


IEEE Transactions on Very Large Scale Integration Systems, 13, 349-357. https://doi.org/10.1109/TVLSI.2004.842903


https://doi.org/10.1109/soc.2003.1241450


https://doi.org/10.1109/JSSC.2004.842846

https://doi.org/10.1109/JSSC.2006.888357


https://doi.org/10.1109/5.371965

https://doi.org/10.1109/4.400426


https://doi.org/10.1109/4.913744


### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTBT</td>
<td>Band-to-band tunneling</td>
</tr>
<tr>
<td>BL</td>
<td>Bit line</td>
</tr>
<tr>
<td>BLB</td>
<td>Bit line bar</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary metal oxide semiconductor</td>
</tr>
<tr>
<td>DIBL</td>
<td>Drain-induced barrier lowering</td>
</tr>
<tr>
<td>DRG</td>
<td>Data retention gated-ground</td>
</tr>
<tr>
<td>D2D</td>
<td>Die to die</td>
</tr>
<tr>
<td>FBB</td>
<td>Forward body biasing</td>
</tr>
<tr>
<td>GIDL</td>
<td>Gate-induced drain leakage</td>
</tr>
<tr>
<td>$I_{BTBT}$</td>
<td>Band to band tunneling effect</td>
</tr>
<tr>
<td>$I_{gc}$</td>
<td>Gate-to-channel current</td>
</tr>
<tr>
<td>$I_{GIDL}$</td>
<td>Gate-induced drain leakage</td>
</tr>
<tr>
<td>$I_{Punch}$</td>
<td>Punch-through leakage</td>
</tr>
<tr>
<td>$I_{sdif}$</td>
<td>Sub-threshold surface diffusion current</td>
</tr>
<tr>
<td>$I_{Sub-th}$</td>
<td>Sub-threshold leakage current</td>
</tr>
<tr>
<td>Low-$V_{TH}$</td>
<td>Low threshold voltage</td>
</tr>
<tr>
<td>NBTI</td>
<td>Negative bias temperature instability</td>
</tr>
<tr>
<td>PBTI</td>
<td>Positive bias temperature instability</td>
</tr>
<tr>
<td>PBT</td>
<td>Programmable bias transistor</td>
</tr>
<tr>
<td>PD</td>
<td>Pull-down</td>
</tr>
<tr>
<td>RBB</td>
<td>Reverse body bias</td>
</tr>
<tr>
<td>RBL</td>
<td>Read word line</td>
</tr>
<tr>
<td>SBT</td>
<td>Self-biasing transistor</td>
</tr>
<tr>
<td>SCE</td>
<td>Short-channel effect</td>
</tr>
<tr>
<td>SNM</td>
<td>Signal to noise margin</td>
</tr>
<tr>
<td>SRAM</td>
<td>Static random access memory</td>
</tr>
<tr>
<td>SRB</td>
<td>Self-reverse biasing</td>
</tr>
<tr>
<td>VTCMOS</td>
<td>Variable threshold CMOS</td>
</tr>
<tr>
<td>$V_{sc}$</td>
<td>Virtual source line</td>
</tr>
</tbody>
</table>
Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/
Or contact cs@scirp.org