Temperature-adaptive voltage scaling for enhanced energy efficiency in subthreshold memory arrays

Ranjith Kumar a,*, Volkan Kursun b

a Advanced Design/Digital Circuits Technology Integration, Intel Corporation, Hillsboro, OR 97124, USA
b Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

A R T I C L E   I N F O

Article history:
Received 13 March 2008
Received in revised form 6 February 2009
Accepted 24 February 2009
Available online 8 April 2009

Keywords:
Adaptive body bias
Dynamic supply voltage scaling
Noise immunity
Process variations
Reversed temperature dependence
SRAM Subthreshold leakage
Subthreshold memory
Supply voltage tuning
Temperature variations
Threshold voltage tuning

A B S T R A C T

Static random access memory (SRAM) circuits optimized for minimum energy consumption typically operate in the subthreshold regime with ultra low-power-supply voltages. Both the read and the write propagation delays of a subthreshold memory circuit are significantly reduced with an increase in the die temperature. The excessive timing slack observed in the clock period of constant-frequency subthreshold memory circuits at elevated temperatures provides new opportunities to lower the active-mode energy consumption. Temperature-adaptive dynamic supply voltage tuning (TA-DVS) technique is proposed in this paper to reduce the high-temperature energy consumption of ultra low-voltage subthreshold SRAM arrays. Results indicate that the energy consumption can be lowered by up to 32.8% by dynamically scaling the supply voltage at elevated temperatures. The impact of the temperature-adaptive dynamic supply voltage scaling technique on the data stability of the subthreshold SRAM bit-cells is presented. The effectiveness of the TA-DVS technique under process parameter and supply voltage variations is evaluated. An alternative technique based on temperature-adaptive reverse body bias (TA-RBB) to exponentially reduce the subthreshold leakage currents at elevated temperatures is also investigated. The active-mode energy consumption characteristics of the two temperature-adaptive voltage tuning techniques are compared.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The embedded memory occupies the majority of the total chip area in the current state-of-the-art high-performance integrated circuits. The amount of on-chip memory is expected to continue to increase for enhancing the performance of future generations of portable devices and high-performance processors. To achieve higher reliability and longer battery lifetime in portable applications, the power consumed by the static random access memory (SRAM) arrays should be reduced. Scaling the supply voltage enhances the energy efficiency primarily by reducing the dynamic switching energy. The optimum supply voltage that provides minimum energy consumption is typically observed in the subthreshold region, as reported in [1,2].

Integrated circuits with ultra low-voltage power supplies are highly sensitive to process and temperature variations [3,7]. The absolute value of the MOSFET threshold voltage degrades and the thermal voltage is enhanced as the temperature increases [11,12]. A small increase in the die temperature exponentially enhances the subthreshold leakage current [7]. Contrary to the standard higher-voltage circuits designed for high-speed, low-voltage circuits optimized for minimum energy operate faster when the die temperature increases [4,14].

Variations in the die temperature are caused by the imbalanced switching activity and the corresponding non-uniform heat dissipation across a die and/or the fluctuations in the environmental temperature. In integrated circuits designed for minimum energy consumption, the formation of local hot-spots is unlikely on the die. The on-chip temperature gradients induced by imbalanced switching activity are therefore typically small across the die of a low-voltage integrated circuit. Die temperature fluctuations due to the variations in the ambient temperature however can cause significant fluctuations in the speed and the power characteristics of ultra low-voltage circuits. For example, the ambient temperatures for integrated circuits employed in robotic explorations vary from −180 to 486 °C [5]. Similarly, ultra low-power sensor-net modules in security applications are designed for functionality at a temperature range of −25 to 125 °C [6].

Dynamic supply voltage scaling technique is primarily used for reducing the active-mode power consumption of an integrated circuit (IC) by exploiting the variations in the computational workload [7,8]. Alternatively, adaptive body bias technique reduces both the active and the standby mode power consumption.
by dynamically adjusting the device threshold voltages depending on
the variations of the workload and the circuit activity [7,9]. In
this paper, a new temperature-adaptive dynamic supply voltage
tuning technique is proposed for reducing the active-mode energy
consumption by exploiting the excessive timing slack produced in
the clock period of subthreshold SRAM circuits at elevated
temperatures. The high-temperature energy efficiency is enhanced
while maintaining a constant clock frequency by dyna-
mically scaling the supply voltage of a subthreshold memory
circuit with the fluctuations of the die temperature. The supply
voltage that lowers the energy consumption without degrading
the circuit speed at increased temperatures is identified for an
SRAM array in the TSMC 180 nm CMOS technology [10]. The
impact of the temperature-adaptive dynamic supply voltage
scaling (TA-DVS) technique on the data stability of the subthresh-
hold SRAM bit-cells is evaluated. The effectiveness of the TA-DVS
technique under process parameter and supply voltage variations
is explored. An alternative technique based on temperature-
adaptive threshold voltage tuning through reverse body bias is
also investigated. The active-mode energy consumption charac-
teristics of the two temperature-adaptive voltage tuning techni-
ques are compared.

The paper is organized as follows. The effects of temperature
fluctuations on the device and circuit characteristics are examined
in Section 2. The sizing constraints in a subthreshold SRAM cell
and the design of a 64-bit × 64-bit ultra low-voltage SRAM array
are discussed in Section 3. A methodology to identify the supply
voltage providing minimum energy consumption in the standard
constant-Vth and constant-frequency SRAM arrays is presented in
Section 4. The new temperature-adaptive supply and threshold
voltage scaling techniques for dynamically reducing the energy
consumed at high die temperatures are described in Section 5. The
influence of the temperature-adaptive dynamic supply voltage
scaling scheme on the data stability of the memory cells is
examined in Section 6. The effectiveness of the TA-DVS scheme
under process parameter and supply voltage variations is
evaluated. Finally, some conclusions are provided in Section 7.

## 2. Low-voltage device and circuit behavior under
temperature fluctuations

The effects of temperature fluctuations on the device and
circuit characteristics are reviewed in this section. An increase in
the die temperature degrades the absolute values of threshold
temperature [12]. Threshold voltage degradation with tempera-
ture tends to enhance the drain current because of the increase in
gate overdrive |VGS–Vth|. Alternatively, degradation in carrier
mobility tends to lower the MOSFET drain current [4,13,14].

Effective variation of MOSFET drain current is determined by the
variation of the dominant device parameter when the tempera-
ture fluctuates. Gate overdrive and carrier mobility variations due
to temperature fluctuations at different supply voltages for devices in a 180 nm CMOS technology are listed in Table 1.

For devices operating at the nominal supply voltage (VDD =
1.8 V), variations in gate overdrive are smaller as compared to
carrier mobility fluctuations when the temperature is increased
from 25 to 125 °C, as listed in Table 1. The MOSFET drain current is,
therefore, reduced following the degradation of carrier mobility as
the temperature is increased, as shown in Fig. 1 [4,13].

The sensitivity of gate overdrive to temperature fluctuations is
enhanced at scaled supply voltages, as listed in Table 1 [4,13]. For a
particular lower supply voltage (VDD < 1.8 V), the temperature
fluctuation induced gate overdrive variation completely counter-
balances the carrier mobility variation, thereby providing tem-
terature variation insensitive MOSFET drain current, as shown in
Fig. 1 [4,13]. Further scaling of the supply voltage reverses the
temperature-dependent speed characteristics of CMOS circuits.
The enhanced variations of the gate overdrive voltage begin to
determine the propagation delay fluctuations with the tempera-
ture. Low-voltage integrated circuits therefore operate faster
when the die temperature increases [4]. When the supply voltage
is scaled below the threshold voltages of the transistors, the
devices operate in the weak inversion region (subthreshold
regime). The switching current in the weak inversion region is
the subthreshold leakage current. Subthreshold leakage current is
extremely sensitive to temperature fluctuations [7]. Degradation
in the device threshold voltages and the enhancement of the
thermal voltage exponentially increase the subthreshold leakage
currents when the die temperature increases [7].

### 3. Sizing of a subthreshold SRAM bit-cell

In this section, the sizing constraints for the stability and the
functionality of a conventional super threshold (VDD > Vth) SRAM
cell are reviewed. The sizing constraints for the robust operation
of the subthreshold SRAM circuits (VDD < Vth) are then distin-
guished. The simulation setup for a 64-bit × 64-bit SRAM array is
also presented.

#### Table 1
Gate overdrive and carrier mobility variations at different supply voltages.

<table>
<thead>
<tr>
<th>Supply voltage (V)</th>
<th>Temperature (°C)</th>
<th>Gate overdrive (V)</th>
<th>Carrier mobility (× 10⁻³ m²/V s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PMOS</td>
<td>NMOS</td>
</tr>
<tr>
<td>1.8</td>
<td>25</td>
<td>-1.34</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>-1.41</td>
<td>1.39</td>
</tr>
<tr>
<td>Variation (%)</td>
<td></td>
<td>5.37</td>
<td>4.95</td>
</tr>
<tr>
<td>1.1</td>
<td>25</td>
<td>-0.64</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>-0.71</td>
<td>0.69</td>
</tr>
<tr>
<td>Variation (%)</td>
<td></td>
<td>11.28</td>
<td>10.48</td>
</tr>
<tr>
<td>0.7</td>
<td>25</td>
<td>-0.24</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>-0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Variation (%)</td>
<td></td>
<td>30.39</td>
<td>30.01</td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>-0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>-0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>Variation (%)</td>
<td></td>
<td>198.88</td>
<td>249.31</td>
</tr>
</tbody>
</table>
A commonly used six transistor (6T) SRAM bit-cell is shown in Fig. 2. A single word-line (WORD) and both true and complementary bit-lines (BL and BLB) are utilized with the standard 6T-SRAM bit-cell. The bit-cell is composed of a pair of cross-coupled inverters (P1/N1 and P2/N2) and two access transistors for the bit-lines (A1 and A2), as shown in Fig. 2. Both the true and the complementary versions of the data are stored in the cross-coupled inverters (on node1 and node2).

Regardless of a read or a write cycle, both bit-lines (BL and BLB) are periodically pre-charged to \( V_{DD} \). Without loss of generality, in the following discussion it is assumed that node1 and node2 are initially at 0 V and \( V_{DD} \), respectively, as shown in Fig. 2. The word-line is raised for a read operation. BL is pulled down through A1 and N1. The current flowing through N1 raises the voltage on node1 while BL is being discharged. If the node1 voltage rises above the switching threshold voltage of the P2/N2 inverter, the data stored in the SRAM cell is flipped. To prevent the loss of data during a read operation, the pull-down transistors in the cross-coupled inverters (N1 and N2) must be stronger as compared to the access transistors (A1 and A2) [15]. Alternatively, when a ‘0’ is to be written to node2, BLB is pulled low by the write driver. After BLB is discharged, the word-line is asserted. A ‘0’ is forced onto node2 through A2. P2 opposes the high-to-low transition of node2. A1 and A2 must be stronger as compared to P1 and P2 to be able to flip the state of a memory cell with brute-force through the bit-line access transistors [15,19,20].

Static noise margin (SNM) is the maximum amount of noise that is tolerated at the data storage nodes of an SRAM cell [3,16]. The voltage transfer characteristics (VTC) of two cross-coupled inverters are shown in Fig. 3. The resulting curve is called the “butterfly curve”. The SNM is the length of the largest square that can be embedded inside the lobes of a butterfly curve, as illustrated in Fig. 3 [3,16,19,20].

As described in [15,16,19,20], for the memory circuits operating in the super threshold (strong inversion) region, the read static noise margin increases with an increase in the memory cell ratio. The read noise margin to the SRAM cell ratio is diminished. As the supply voltage is scaled to the subthreshold region, the dependence of the data stability on the SRAM cell ratio becomes negligible.
The reason for the diminishing sensitivity of the read static noise margin to the transistor sizes at scaled supply voltages is identified next. The switching current at ultra low voltages is the subthreshold leakage current. The subthreshold leakage current produced by a MOSFET is [7]

\[
I_{\text{leak}} = \frac{\mu W_{\text{eff}} C_{\text{OX}}}{L_{\text{eff}}} V_{T}^2 (V_{GS}-V_{TH})/nV_{T}(1-e^{-V_{GS}/V_{T}}),
\]

where \( I_{\text{leak}} \), \( \mu \), \( W_{\text{eff}} \), \( C_{\text{OX}} \), \( L_{\text{eff}} \), \( V_{T} \), \( V_{GS} \), \( V_{DS} \), and \( n \) are the subthreshold leakage current, carrier mobility, effective transistor width, oxide capacitance per unit area, effective channel length, threshold voltage, thermal voltage, gate-to-source voltage, drain-to-source voltage, and subthreshold swing coefficient, respectively. For devices operating in the weak inversion region, the switching current is exponentially dependent on the voltage levels, as given by (1). Alternatively, increasing the device width produces only a linear increase in the switching current. A linear change in the subthreshold switching current has a relatively small impact on the voltage transfer characteristics. The sensitivity of the read noise margin to the memory cell ratio is therefore negligible in an ultra low supply voltage subthreshold memory circuit. In this paper, all the devices in the subthreshold SRAM cells are sized identical (\( W_{N1} = W_{N2} = W_{P1} = W_{P2} \)), since increasing the memory cell ratio does not provide a significant enhancement in data stability.

The carrier mobility of a PMOS device is lower as compared to an NMOS device since a hole is heavier as compared to an electron. The total weak inversion current produced by a PMOS device is, therefore, smaller as compared to an NMOS device with similar physical dimensions (width, length, and \( T_{\text{ox}} \)) and similar voltage difference across the device terminals, as given by (1). Therefore, even when the bit-cell devices are sized equal, the access transistors (A1 and A2) are stronger than the pull-up devices (P1 and P2), thereby satisfying the necessary condition for write ability.

The performance and the power consumption of the subthreshold logic circuits are affected significantly by the process parameter variations [3,7]. To suppress the fluctuations of the threshold voltage due to the process parameter variations, devices can be sized larger than the minimum width allowed in a given technology [3]. The minimum device width allowed in this TSMC 180 nm CMOS technology is 220 nm. In this paper, the widths of all the devices in the SRAM bit-cells are sized 360 nm for robustness against process variations.

The layout of the 6T-SRAM bit-cell drawn using the design rules of the TSMC 180 nm CMOS technology [15] is shown in Fig. 5. The dimensions of the layout of a single bit-cell are 49 \( \lambda \times 16 \lambda \), where \( \lambda \) is 90 nm. The width of each device is 4\( \lambda \). The bit-lines and the word-lines are routed using the metal layers 2 and 3, respectively. The per cell word-line resistance (\( R_{\text{WORD}} \)) and capacitance (\( C_{\text{WORD}} \)) estimated using the sheet layers 2 and 3, respectively. The per cell word-line resistance (\( R_{\text{BL/BLB}} \)) and capacitance (\( C_{\text{BL/BLB}} \)) are 0.225 \( \Omega \) and 0.334 \( fF \), respectively. The bit-lines and the word-lines are modeled as \( \pi \)-networks. The simulation setup of the 64-bit \( \times 64 \)-bit SRAM array is shown in Fig. 6. The read and write propagation delays reported in this paper are measured for the SRAM bit-cell farthest from the word-line driver and the read driver, as shown in Fig. 6. The read and write circuitry used with this SRAM array are shown in Fig. 7 [15].

4. Supply voltage optimization for minimum energy consumption

The impressive growth of the mobile products industry and the growing interest in the self-sustaining integrated systems with energy-scavenging capability has produced significant interest in
ultra low-power circuit design. Power consumption of CMOS circuits can be lowered by employing several techniques as described in [7–9]. In this section, a design methodology for minimizing the energy consumption of CMOS circuits is described.

The two primary sources of power dissipation in CMOS circuits are the static power, which results from leakage currents of the MOSFETs, and the dynamic power, which results from the switching activity. The energy consumed per clock cycle is

\[
\text{Energy}_{\text{total}} = \text{Energy}_{\text{switching}} + \text{Energy}_{\text{leakage}},
\]

\[
\frac{\text{Energy}_{\text{switching}}}{V_D^2},
\]

\[
\text{Energy}_{\text{leakage}} = I_{\text{leakage}} V_D T,
\]

where \(\text{Energy}_{\text{total}}\), \(\text{Energy}_{\text{switching}}\), \(\text{Energy}_{\text{leakage}}\), \(I_{\text{leakage}}\), \(V_D\), and \(T\) are the total energy consumed per clock cycle, the total dynamic switching energy per clock cycle, the total leakage energy per clock cycle, the total leakage current, the supply voltage, and the clock period, respectively. The energy efficiency of an integrated circuit can be enhanced by scaling the power supply voltage [4]. Supply voltage scaling quadratically reduces the dynamic switching energy, as given by (3). Scaling the supply voltage, however, can also increase the total leakage energy per cycle due to the increase in the clock period, as given by (4) [1]. The total energy consumed by an IC, therefore, has a minimum as the supply voltage is scaled.

Standard ICs are designed to operate with a constant supply voltage (constant-\(V_{DD}\)) at a constant frequency (constant-\(f_s\)) under different environmental conditions. One criterion for determining an appropriate supply voltage is to minimize the energy consumption of an IC. An algorithm that optimizes the supply voltage of a standard constant-\(V_{DD}\) (with no supply voltage scaling capability) and constant-\(f_s\) IC for achieving minimum energy consumption is illustrated in Fig. 8, assuming a \(T_1 \rightarrow T_2\).
temperature spectrum. \(V_{\text{DD-nom}}\), \(iDD_{\text{-min}}\), and \(V_{\text{step}}\) are the nominal supply voltage, the supply voltage below which the circuit malfunctions, and the voltage scaling resolution, respectively. \(V_{\text{DD-nom}}\) is technology dependent (1.8 V for a 180 nm CMOS technology) and \(V_{\text{step}}\) is assumed to be 10 mV in this paper. In the first iterative part of the algorithm, the supply voltage is scaled with a voltage resolution of \(V_{\text{step}}\). The highest constant clock frequency for which the functionality is guaranteed within the entire temperature spectrum is identified for each supply voltage. In the second part of the algorithm, the energy consumed by the circuit is measured at various temperatures of interest for each pair of supply voltage and the corresponding highest achievable clock frequency. From the measured energy consumption, the constant supply voltage that achieves minimum energy at a specific temperature (within the temperature spectrum) is identified assuming a standard constant-\(V_{\text{DD}}\) and constant-\(f_s\) circuit operation.

The methodology used in this paper to measure the maximum frequency (\(f_{\text{max}}\)) achievable with the SRAM array at a specific supply voltage and temperature is illustrated in Fig. 9. The signals for the read and the write operations in an SRAM bit-cell are shown in Fig. 9. The initial voltages of node1 and node2 in the SRAM bit-cell are assumed to be 0 V and \(V_{\text{DD}}\), respectively, as shown in Fig. 2. First, a 0 V is written into the SRAM bit-cell. The BLB is discharged by applying 0 V on DATA_IN during the write operation. Node1 transitions from 0 V to \(V_{\text{DD}}\), as shown in Fig. 9. The contents of the SRAM bit-cell are flipped. The write operation is followed by a read operation. Following this sequence, another write operation is performed. During the second write operation, DATA_IN is maintained at \(V_{\text{DD}}\), BL is discharged. Node1 transitions from 0 V to \(V_{\text{DD}}\) during the write operation, as shown in Fig. 9. Finally, the most recent data is read from the selected bit-cell of the memory array. The voltage at DATA_OUT rises during the read operation indicating that the data stored in the SRAM bit-cell corresponds to logic 1.

The SRAM array is initially operated at a low clock frequency \(f < f_{\text{max}}\) where \(f_{\text{max}}\) needs to be determined. \(\text{Time}_1\) (\(\text{Time}_2\)) is the time taken for the rising (falling) node1 voltage to cross 0.9*\(V_{\text{DD}}\) (0.1*\(V_{\text{DD}}\)) after the rising WORD signal crosses 0.1*\(V_{\text{DD}}\) (0.1*\(V_{\text{DD}}\)). Similarly, \(\text{Time}_3\) is the time taken for the rising DATA_OUT signal to cross 0.9*\(V_{\text{DD}}\) after the rising WORD signal crosses 0.1*\(V_{\text{DD}}\). A 20% margin is added to the maximum of \(\text{Time}_1\), \(\text{Time}_2\), and \(\text{Time}_3\) to provide a timing slack against parameter variations and clock-skew. The maximum clock frequency achievable with a synchronous memory circuit operating at a specific supply voltage and temperature is

\[
f_{\text{max}} = \left(\frac{2^\max(\text{Time}_1, \text{Time}_2, \text{Time}_3)}{C_3} + 0.2^\max(\text{Time}_1, \text{Time}_2, \text{Time}_3)\right)^{-1}. \tag{5}
\]

To find the highest achievable constant clock frequency at a particular supply voltage, the maximum achievable frequencies

\[
\text{Start} \\
V_{\text{DD}} = V_{\text{DD-nom}} \\
\text{Find and record the worst-case (lowest)} \\
\text{frequency in the temperature range of } T_1 \text{ to } T_2 \\
V_{\text{DD}} = V_{\text{DD-step}} \\
\text{no} \\
V_{\text{DD}} = V_{\text{DD-min}} \\
\text{yes} \\
V_{\text{DD}} = V_{\text{DD-nom}} \\
T = T_{\text{specific}} (T_{\text{specific}} \in (T_1 \to T_2)) \\
\text{For the highest constant frequency achievable with } V_{\text{DD}}, \text{ find and} \\
\text{record the energy consumption at } T_{\text{specific}} \\
V_{\text{DD}} = V_{\text{DD-step}} \\
\text{yes} \\
V_{\text{DD}} = V_{\text{DD-min}} \\
\text{no} \\
\text{Report the constant supply voltage that provides the} \\
\text{minimum energy consumption at } T_{\text{specific}} \\
\text{stop}
\]

Fig. 8. Flow-chart for identifying the supply voltage that achieves minimum energy consumption at a specific temperature (\(T_{\text{specific}}\)) for a standard constant-\(V_{\text{DD}}\) and constant-\(f_s\) IC.
At the higher supply voltages (such as the nominal operating frequency is also dependent on the supply voltage of the circuit. The propagation delays (read and write) observed at the lowest temperature for \( V_{DD} = 0.85 \text{V} \). \( V_{DD,25} \) is the constant supply voltage applied to a standard CMOS circuit (without any voltage tuning capability) for achieving minimum energy consumption at 25 °C.

The long channel threshold voltage of the n-channel and the p-channel devices in this 180 nm CMOS technology is 0.48 and 0.45 V, respectively [10]. The supply voltage providing minimum energy (\( V_{DD,25} \)) is lower than the threshold voltage of the devices in this technology (the SRAM array, therefore, operates in the subthreshold regime), as listed in Table 2 [12]. The switching current at these ultra low supply voltages is the subthreshold leakage current. As described in Section 2, subthreshold leakage current is extremely sensitive to temperature fluctuations. Absolute value of the threshold voltage degrades and the thermal voltage is enhanced as the temperature increases [11,12]. A small change in the die temperature exponentially alters the subthreshold leakage current, as given by (1). The reversal in the temperature-dependent propagation delay characteristics coupled with the high sensitivity of the circuit speed to the temperature fluctuations provides opportunities for reducing the energy consumption without degrading the clock frequency at elevated die temperatures in an ultra low supply voltage memory array.

### Table 2

<table>
<thead>
<tr>
<th>( V_{DD} ) (V)</th>
<th>Max. frequency at 25°C (MHz)</th>
<th>Max. frequency at 125°C (MHz)</th>
<th>Worst-case frequency (MHz)</th>
<th>Energy consumption at the worst-case frequency and 25°C (pJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.80</td>
<td>521.7</td>
<td>392.7</td>
<td>392.7</td>
<td>22.60</td>
</tr>
<tr>
<td>1.70</td>
<td>486.2</td>
<td>388.7</td>
<td>388.7</td>
<td>20.30</td>
</tr>
<tr>
<td>1.60</td>
<td>465.3</td>
<td>374.7</td>
<td>374.7</td>
<td>17.91</td>
</tr>
<tr>
<td>0.86</td>
<td>180.2</td>
<td>176.2</td>
<td>176.2</td>
<td>4.84</td>
</tr>
<tr>
<td>0.85</td>
<td>173.3</td>
<td>173.6</td>
<td>173.3</td>
<td>4.72</td>
</tr>
<tr>
<td>0.84</td>
<td>167.9</td>
<td>168.1</td>
<td>167.9</td>
<td>4.59</td>
</tr>
<tr>
<td>0.30</td>
<td>0.1572</td>
<td>2.0204</td>
<td>0.1572</td>
<td>0.776</td>
</tr>
<tr>
<td>0.29</td>
<td>0.1311</td>
<td>1.7986</td>
<td>0.1311</td>
<td>0.728</td>
</tr>
<tr>
<td>0.28</td>
<td>0.1019</td>
<td>1.5882</td>
<td>0.1019</td>
<td>0.711</td>
</tr>
<tr>
<td>0.27</td>
<td>0.0867</td>
<td>1.3319</td>
<td>0.0867</td>
<td>0.716</td>
</tr>
<tr>
<td>0.26</td>
<td>0.0648</td>
<td>1.1164</td>
<td>0.0648</td>
<td>0.729</td>
</tr>
</tbody>
</table>

Supply voltage below which the circuit exhibits reverse temperature dependence

\( V_{DD,25} \) Minimum energy at 25°C

<table>
<thead>
<tr>
<th>( V_{DD,25} )</th>
<th>Minimum energy at 25°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.30</td>
<td>0.1572</td>
</tr>
<tr>
<td>0.29</td>
<td>0.1311</td>
</tr>
<tr>
<td>0.28</td>
<td>0.1019</td>
</tr>
<tr>
<td>0.27</td>
<td>0.0867</td>
</tr>
<tr>
<td>0.26</td>
<td>0.0648</td>
</tr>
</tbody>
</table>

Results are for a 64-bit \( \times \) 64-bit SRAM array in a 180nm CMOS technology.

### 5. Techniques for high-temperature energy reduction

In this section, the previously proposed conventional voltage scaling and body bias techniques are briefly discussed. Two new temperature-adaptive dynamic voltage tuning techniques for enhancing the high-temperature active-mode energy efficiency of the subthreshold SRAM circuits are then introduced.

The operational load for an integrated circuit tends to have peak performance requirements followed by idle periods [7]. Maintaining the full computational capacity at all times, despite the reduction of the throughput requirements with the variations of the workload, wastes significant amount of energy. Dynamic supply voltage scaling technique exploits the variations in the computational workload by dynamically adjusting the supply voltage and the clock frequency of a synchronous system. The primary objective of dynamic supply voltage scaling technique is to provide high throughput during the execution of only the computation-intensive tasks while saving energy during the rest of the time by lowering the supply voltage and the operating clock frequency. The dynamic voltage scaling technique is primarily aimed at reducing the active-mode power consumption of an integrated circuit.

Alternatively, the adaptive body bias techniques utilize the bulk terminal to dynamically modify the threshold voltages of the devices during circuit operation. Depending upon the polarity of the voltage difference between the source and the body terminals \( (V_{DBL}) \), the threshold voltage can be either increased or decreased as compared to a zero body-biased transistor. Device threshold voltages can be increased by applying reverse body bias in the standby mode in order to reduce the subthreshold leakage current produced by idle circuits [9]. Furthermore, the dynamic supply voltage scaling and adaptive body bias techniques can also be used to compensate for the die-to-die and within-die process parameter variations, thereby enhancing yield [7].

In ultra low-voltage circuits, such as the subthreshold memory arrays, temperature gradients due to imbalanced switching activity within a die are typically small. The primary source of the die temperature fluctuations in low-voltage embedded memory circuits are the variations in the ambient temperature. Changes in the ambient temperature tend to uniformly affect all

\( (f_{\text{max}}) \) at the extremes of the die temperature spectrum \((T_1 \text{ and } T_2)\) are measured using the described procedure. The smaller of the two frequencies is the highest constant frequency that can be maintained by the circuit within the entire temperature spectrum \((T_1 \rightarrow T_2)\) at the particular supply voltage.

The results of the algorithm applied to a 64-bit \( \times \) 64-bit SRAM array are listed in Table 2. The die temperature spectrum is assumed to be from 25 to 125 °C. The standard constant supply voltage for achieving minimum energy consumption at 25 °C is reported. As listed in Table 2, the temperature that determines the highest operating frequency is also dependent on the supply voltage of the circuit. At the higher supply voltages (such as the nominal-\( V_{DD} = 1.8 \) V), the SRAM circuit operates slower when the die temperature increases. The maximum achievable (worst-case) frequency is therefore determined by the \( f_{\text{max}} \) at the highest temperature. Alternatively, as the supply voltage is scaled, the worst-case clock frequency shifts to the lowest operating temperature, as listed in

Fig. 9. Signals indicating the read and the write operations in an SRAM bit-cell.
the devices in an IC. At elevated die temperatures, the leakage currents as well as the circuit speed are enhanced. Increased leakage power, in turn, further enhances the heat dissipation and elevates the die temperature. This positive feedback between the die temperature, the leakage current, and the total power consumption significantly reduces the battery lifetime in portable devices, accelerates the degradation of the device/circuit reliability due to excessive heating, and can even cause thermal runaway in extreme environments despite the relatively low supply voltage. New temperature-adaptive design methodologies are, therefore, highly desirable to enhance the reliability and energy efficiency of ultra low-voltage circuits operating at environments subject to significant temperature fluctuations.

Integrated circuits are typically designed for guaranteed functionality at the estimated worst-case process and environmental parameter corners. In constant-\(V_{DD}\) and constant-\(f_c\) circuits optimized for minimum energy, the worst-case speed is observed at the lowest operating temperature, as listed in Table 2. The lowest temperature therefore determines the achievable maximum clock frequency. As the die temperature increases, the subthreshold logic circuits operate faster, thereby producing significant timing slack in the constant clock period.

In this paper, temperature-adaptive supply and threshold voltage tuning techniques are proposed to dynamically adjust the speed of the synchronous memory arrays based on the die temperature. The primary objective of the proposed temperature-adaptive schemes is to lower the active-mode energy consumption by exploiting the excessive timing slack produced in the clock period at high die temperatures while maintaining a constant clock frequency across the entire die temperature spectrum. The objective can be achieved by either dynamically scaling the power supply voltage or dynamically increasing the device threshold voltages through reverse body bias at elevated temperatures. The temperature-adaptive supply voltage tuning and the temperature-adaptive reverse body bias (TA-RBB) techniques are presented in Sections 5.1 and 5.2, respectively.

5.1. Temperature-adaptive dynamic supply voltage scaling

The temperature-adaptive dynamic supply voltage scaling technique is presented in this section. All the primary components of power consumption in a CMOS circuit, namely dynamic switching, short-circuit, and leakage power are significantly reduced by scaling the supply voltage. The read and the write propagation delays of an SRAM array are also strongly dependent on the supply voltage [7]. Scaling the supply voltage reduces the power consumed by a circuit at the cost of degraded circuit performance.

The CLK and DATA_OUT of the 64th column in a 64-bit \(\times\) 64-bit SRAM array operating at the constant-\(V_{DD-25}\) with various temperatures are shown in Fig. 10. \(V_{DD-25}\) for the SRAM array is 0.28 V, as listed in Table 2. The clock frequency is fixed at 101.9 KHz in a standard CMOS circuit (determined by the lowest operating temperature), the highest constant-\(f_c\) that can be maintained at all die temperatures, as listed in Table 2. The circuit operates faster at elevated temperatures, thereby producing excessive timing slack in the constant clock period, as shown in Fig. 10. At elevated temperatures the total energy consumption also significantly increases due to the increase in the subthreshold leakage current [7].

The significant timing slack in the clock period can be exploited to reduce the active-mode energy consumption at elevated temperatures. With the proposed technique, the supply voltage of the circuit is dynamically scaled below \(V_{DD-25}\) while maintaining the constant clock frequency as the die temperature increases. The supply voltage of the TA-DVS SRAM circuit is tuned until the high-temperature performance at the scaled supply voltage matches the low-temperature performance of the standard constant-\(V_{DD}\) and constant-\(f_c\) subthreshold logic circuit operating at \(V_{DD-25}\). Unlike the conventional work-load adaptive dynamic voltage scaling techniques [78], a new die temperature adaptive dynamic voltage scaling technique is proposed in this paper for tuning the supply voltage of the SRAM array based on the fluctuations of the die temperature and the circuit speed.

A system with temperature-adaptive supply voltage tuning capability is illustrated in Fig. 11. The low-temperature-\(V_{DD}\) and the target operating frequency (\(f_{error}\)) of the SRAM array for achieving minimum energy consumption are determined at the lowest operating temperature according to the algorithm illustrated in Fig. 8. A ring oscillator providing a replica of the critical delay path of the entire SRAM array is employed to track the fluctuations of the circuit delay with the variations of the ambient temperature at a specific supply voltage. A relatively uniform temperature is assumed across the die with this technique. Note that the uniform die temperature assumption is typically satisfied with the ultra low-voltage subthreshold logic circuits that cannot produce local hot-spots. The ring oscillator translates the variations in the die temperature to a specific clock frequency (\(f_{clock}\)) for a specific power supply voltage generated by the DC–DC converter. As the die temperature increases, the ring oscillator frequency (\(f_{clock}\)) also increases due to the enhanced weak inversion currents produced by the MOSFETs. The ring oscillator frequency is compared to the target clock frequency (\(f_{error}\)), generating a frequency error signal (\(f_{error}\)). The pulse width modulator using this error signal generates the control signals for the DC–DC converter to either modify or maintain the output voltage. The power supply voltage is thereby, dynamically tuned based on the variations of the die temperature using the closed loop feedback circuitry shown in Fig. 11.

The propagation delays (read and write) and the energy consumption of the SRAM array at various power supply voltages and temperatures are listed in Table 3. For all the supply voltages listed in Table 3, the outputs achieve at least a 0.1 \(V_{DD}\) to 0.9 \(V_{DD}\) voltage swing (condition for functionality). Max_delay is the maximum of the read and the write propagation delays at a given supply voltage and temperature. At \(V_{DD-25}\) (\(V_{DD} = 0.28\)), the max_delay reduces from 3.72 to 0.22 \(\mu\)s when the temperature increases from 25 to 125°C, as listed in Table 3. To exploit this slack in the clock period, the supply voltage can be scaled with the proposed TA-DVS technique while maintaining a constant...
frequency. The supply voltage can be scaled down to 0.21 V at the highest temperature while maintaining the circuit delay lower than the lowest temperature propagation delay ($\text{max}_{\text{delay}}$), as listed in Table 3. Supply voltage scaling reduces the energy consumption of a circuit. With the proposed TA-DVS technique, the high-temperature energy consumption is reduced by up to 32.8% (from 12.93 to 8.69 pJ) without degrading the clock frequency of the SRAM array in this 180 nm CMOS technology. At supply voltages below 0.21 V, the outputs do not achieve $0.1V_{\text{DD}} - 0.9V_{\text{DD}}$ voltage swing (condition for functionality).

5.2. Temperature-adaptive body bias

An alternative voltage tuning technique based on temperature-adaptive body bias (TA-RBB) is presented in this section. Similar to the dependence on the supply voltage, the propagation delay of a circuit is also strongly dependent on the device threshold voltages [7]. The absolute value of the threshold voltages degrade and the thermal voltage is enhanced as the temperature increases, thereby simultaneously enhancing the circuit speed and the subthreshold leakage currents at elevated temperatures [7,11]. In ultra low-voltage subthreshold memory circuits exhibiting reversed temperature dependence, the threshold voltage of devices is dynamically increased through reverse body bias at elevated temperatures [7,11]. To exponentially reduce the leakage current without degrading the clock frequency, the device threshold voltages are increased until the high-temperature performance of the SRAM array with TA-RBB matches the worst-case performance of a standard zero body-biased SRAM array operating at a constant-$V_{\text{DD}}$. Unlike the conventional body bias techniques aimed at altering the device threshold voltages based on the variations of the workload and circuit activity, the presented temperature-adaptive body bias technique alters the threshold voltages of the devices based on the fluctuations of the die temperature and the circuit speed.

The temperature-adaptive reverse body bias technique is illustrated in Fig. 12. Integrated circuits with the TA-RBB technique operate with a constant supply voltage. The supply voltage and the target operating frequency ($f_{\text{target}}$) are determined according to the algorithm illustrated in Fig. 8. For minimum energy consumption at 25 °C, the supply voltage of the circuit is fixed at $V_{\text{DD}} - 25$. A ring oscillator providing a replica of the critical delay path of the entire SRAM array translates the die temperature to a specific clock frequency ($f_{\text{clock}}$) for a specific set of body bias voltages produced by the PMOS and NMOS body bias generators. Note that a relatively uniform temperature is assumed across the die with this technique. The ring oscillator frequency ($f_{\text{clock}}$) is compared with the target operating frequency ($f_{\text{target}}$) and a frequency error signal ($f_{\text{error}}$) is generated. Using this error signal, the body bias generators either modify or maintain the body bias voltages applied to the devices in the integrated circuit. The device threshold voltages are, thereby, dynamically tuned based on the die temperature variations for maintaining a constant circuit speed across the entire die temperature spectrum.

The propagation delays (read and write) and the energy consumption of the SRAM array at various body bias voltages and temperatures are listed in Table 4. At $V_{\text{DD}-25}$ ($V_{\text{DD}} = 0.28$), the $\text{max}_{\text{delay}}$ of the memory circuit is reduced when the temperature increases. To exploit the available timing slack, the proposed

### Table 3
Read delay, write delay, and energy consumption of the SRAM array at various temperatures and power supply voltages.

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>$V_{\text{DD}}$ (V)</th>
<th>Read delay (μs)</th>
<th>Write delay (μs)</th>
<th>$\text{Max}_{\text{delay}}$ (μs)</th>
<th>Energy consumption (pJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.28</td>
<td>3.72</td>
<td>2.95</td>
<td>3.72</td>
<td>0.711</td>
</tr>
<tr>
<td>125</td>
<td>0.28</td>
<td>0.22</td>
<td>0.20</td>
<td>0.22</td>
<td>12.93</td>
</tr>
<tr>
<td></td>
<td>0.27</td>
<td>0.23</td>
<td>0.25</td>
<td>0.23</td>
<td>12.33</td>
</tr>
<tr>
<td></td>
<td>0.26</td>
<td>0.26</td>
<td>0.30</td>
<td>0.30</td>
<td>11.63</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.27</td>
<td>0.34</td>
<td>0.34</td>
<td>11.05</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0.28</td>
<td>0.42</td>
<td>0.42</td>
<td>10.46</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>0.30</td>
<td>0.50</td>
<td>0.50</td>
<td>9.91</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.32</td>
<td>0.59</td>
<td>0.59</td>
<td>9.30</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>0.33</td>
<td>0.72</td>
<td>0.72</td>
<td>8.69</td>
</tr>
</tbody>
</table>

$\text{Max}_{\text{delay}}$ is the maximum of the read and the write propagation delays at a specific supply voltage and temperature.
TA-RBB technique is applied to the SRAM array. For the memory circuits in this 180 nm CMOS technology, the applicable high-temperature reverse body bias voltage without degrading the clock frequency is up to 0.37 V with the TA-RBB technique, as listed in Table 4. The energy consumption of the SRAM array for the various body bias voltages are also listed in Table 4. Results indicate that the high-temperature energy consumption with this TA-RBB technique increases by up to 1.47 $\times$ (from 12.93 to 19.02 pJ).

The reason for the higher energy consumption at elevated temperatures in a TA-RBB SRAM array is illustrated here with a p-channel MOSFET in a 180 nm CMOS technology. The switching current at ultra low-power-supply voltages is the subthreshold leakage current. A p-channel MOSFET operating in the subthreshold regime with the drain biased at 0 V and the gate and source terminals biased at 0.28 V ($V_{dd,25}$ for the SRAM array, as listed in Table 2) is shown in Fig. 13. The currents observed at the different terminals of this PMOS transistor for various body bias voltages ($V_{SB}$) are listed in Table 5 along with the total power consumption of the device.

Applying reverse body bias increases the device threshold voltage thereby reducing the subthreshold leakage current [7]. Applying reverse body bias, however, also increases the junction leakage currents due to the enhanced band-to-band tunneling [7]. As listed in Table 5, even for a small reverse body bias voltage ($|V_{SB}| = 0.04$ V), the leakage current through the body diodes increases by up to 68.8% (from 160.4 to 270.8 pA) as compared to a zero body-biased transistor. For the relatively higher reverse body bias voltage applied to the TA-RBB SRAM array ($|V_{SB}| = 0.37$ V), the increase in the substrate current dominates the reduction in the weak inversion current, thereby increasing the total power consumed by the individual devices, as listed in Table 5.

6. Effectiveness of the TA-DVS technique

In this section, the reliability of the subthreshold memory circuits with the TA-DVS technique is evaluated. The influence of the temperature-adaptive dynamic supply voltage tuning technique on the data stability of the subthreshold memory cells is

---

**Table 4**

Read delay, write delay, and energy consumption of the SRAM array at various temperatures and body bias voltages.

| Temperature (°C) | $|V_{SB}|$ (V) | Read delay (µs) | Write delay (µs) | $\text{Max}_\text{delay}$ (µs) | Energy consumption (pJ) |
|----------------|--------------|----------------|----------------|----------------|------------------|
| 25             | 0.0          | 3.72           | 2.95           | 3.72           | 0.711            |
| 125            | 0.0          | 0.22           | 0.20           | 0.22           | 12.93            |
|                | 0.17         | 0.77           | 0.65           | 0.77           | 13.30            |
|                | 0.21         | 1.03           | 0.84           | 1.03           | 14.23            |
|                | 0.25         | 1.38           | 1.08           | 1.38           | 15.30            |
|                | 0.29         | 1.86           | 1.38           | 1.86           | 16.47            |
|                | 0.33         | 2.52           | 1.77           | 2.52           | 17.72            |
|                | 0.37         | 3.45           | 2.45           | 3.45           | 19.02            |
|                | 0.41         | 4.74           | 2.85           | 4.74           | 20.42            |

$\text{Max}_\text{delay}$ is the maximum of the read and the write propagation delays at a specific supply voltage and temperature.
presented in Section 6.1. The effectiveness of the proposed TA-DVS scheme under the presence of process parameter and environmental variations is evaluated in Section 6.2.

6.1. Influence of the TA-DVS technique on the noise margins

In this section, the influence of the proposed temperature-adaptive dynamic supply voltage scaling scheme on the noise margins of the SRAM cell is evaluated. The hold static noise margin and the read static noise margin of the SRAM bit-cells (all the devices in the bit-cells are sized the same) at different supply voltages and temperatures are listed in Table 6. The read static noise margin of the SRAM bit-cells operating at $V_{DD,25} (V_{DD} = 280 \text{ mV})$ at both 25 and 125 °C is 35 mV, as listed in Table 6.

With the proposed dynamic supply voltage tuning technique, the power supply voltage of the SRAM array can be scaled down to 210 mV when the temperature increases to 125 °C. Both the hold and read static noise margins are reduced when the supply voltage is scaled. With the proposed technique, the read static margin of the SRAM bit-cell is reduced by up to 19 mV (from 35 to 16 mV) at high temperatures, as listed in Table 6. Provided that this degradation in the read static noise margin can be tolerated, the proposed temperature-adaptive dynamic voltage scaling scheme can be employed to reduce the high-temperature power consumption by up to 32.8%, as discussed in Section 5.

6.2. Impact of the process parameter and supply voltage variations

Subthreshold logic circuits are highly sensitive to variations in the process parameters, the supply voltage, and the operating temperature [13,7]. Both the performance and the energy consumption of integrated circuits are altered due to the fluctuations of the circuit parameters [7,17,18]. The impact of the parameter variations on the proposed TA-DVS technique is evaluated in this section.

Random and systematic fluctuations in the channel length ($L_{\text{CATE}}$), the doping concentration ($N_{\text{CH}}$), and the gate-oxide thickness ($T_{\text{OX}}$) cause variations in the threshold voltage of a MOSFET. Fluctuation in the threshold voltage alters the performance and the power consumption (both dynamic and leakage power consumption) of a circuit. In this paper, the variations in the performance and the energy consumption due to the process variations in the channel length ($L_{\text{CATE}}$), the doping concentration ($N_{\text{CH}}$), and the gate-oxide thickness ($T_{\text{OX}}$) are evaluated. Each parameter is assumed to have an independent normal Gaussian statistical distribution with a three-sigma variation of 10% [18].

Another important source of noise in CMOS integrated circuits is the power supply noise [7]. Integrated circuits are typically designed to meet the performance specifications at a voltage 10% lower than the nominal supply voltage to account for the supply voltage variations [7]. In this paper, the supply voltage is assumed to have an independent normal Gaussian statistical distribution with a three-sigma variation of 10%.

Monte-Carlo simulations are run to evaluate the performance and the energy consumption fluctuations in circuits with the TA-DVS technique. The delay versus energy consumption plots for the 64-bit x 64-bit SRAM arrays operating at $V_{DD,25}$ and the optimized high-temperature supply voltage with the TA-DVS technique under the presence of process parameter and environmental variations are discussed in Section 6.2.

### Table 5

Post-layout current measured at the different terminals of the PMOS device for various body bias voltages.

<table>
<thead>
<tr>
<th>$V_{SB}$ (V)</th>
<th>$V_{SB}$ (V)</th>
<th>$I_G$ (pA)</th>
<th>$I_L$ (pA)</th>
<th>$I_D$ (pA)</th>
<th>$I_S$ (pA)</th>
<th>$I_B$ (pA)</th>
<th>Device total power consumption (pW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>0.00</td>
<td>-228.1</td>
<td>0.0</td>
<td>-67.6</td>
<td>160.4</td>
<td>26.0</td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>0.04</td>
<td>-209.2</td>
<td>0.0</td>
<td>61.7</td>
<td>270.8</td>
<td>103.9</td>
<td></td>
</tr>
<tr>
<td>0.36</td>
<td>0.08</td>
<td>-195.8</td>
<td>0.0</td>
<td>109.5</td>
<td>305.3</td>
<td>140.6</td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.12</td>
<td>-186.3</td>
<td>0.0</td>
<td>129.8</td>
<td>316.1</td>
<td>162.8</td>
<td></td>
</tr>
<tr>
<td>0.44</td>
<td>0.16</td>
<td>-179.5</td>
<td>0.0</td>
<td>140.1</td>
<td>319.5</td>
<td>179.8</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.20</td>
<td>-174.6</td>
<td>0.0</td>
<td>146.1</td>
<td>320.6</td>
<td>194.8</td>
<td></td>
</tr>
<tr>
<td>0.52</td>
<td>0.24</td>
<td>-171.9</td>
<td>0.0</td>
<td>150.0</td>
<td>321.0</td>
<td>208.9</td>
<td></td>
</tr>
<tr>
<td>0.56</td>
<td>0.28</td>
<td>-168.5</td>
<td>0.0</td>
<td>152.8</td>
<td>321.2</td>
<td>222.7</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.32</td>
<td>-166.6</td>
<td>0.0</td>
<td>154.7</td>
<td>321.3</td>
<td>236.1</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6

Hold and read static noise margins of the SRAM bit-cells operating at different power supply voltages and temperatures.

<table>
<thead>
<tr>
<th>$V_{DD}$ (mV)</th>
<th>Hold noise margin (mV)</th>
<th>Read noise margin (mV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25 °C</td>
<td>125 °C</td>
</tr>
<tr>
<td>250</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>260</td>
<td>81</td>
<td>80</td>
</tr>
<tr>
<td>270</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td>280</td>
<td>91</td>
<td>90</td>
</tr>
<tr>
<td>290</td>
<td>95</td>
<td>94</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>310</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>320</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>330</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>340</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>350</td>
<td>125</td>
<td>125</td>
</tr>
</tbody>
</table>

Fig. 13. A PMOS device in the TSMC 180 nm CMOS technology. The gate and source terminals are biased at 0.28 V. Temperature = 125 °C. The device is reverse body biased by applying a voltage higher than 0.28 V to the body terminal.
The 3-sigma offset of the high-temperature (125°C) energy consumption of the SRAM arrays (propagation delay at $V_{DD} = V_{DD-25}$ and temperature = 25°C is 3719 ns, as listed in Table 3). The smaller high-temperature propagation delay in the TA-DVS memory arrays indicate that there is sufficient timing slack in the constant clock period even in the presence of parameter variations.

The mean and the standard deviation of the high-temperature (125°C) energy consumption of the standard constant-$V_{DD}$ memory circuits operating at $V_{DD-25}$ are 12.98 pJ and 555 fJ, respectively. The 3 standard deviation (3-sigma) offset of the lowest energy consumption in these circuits is (mean−3*standard deviation) 11.32 pJ. Alternatively, the mean and the standard deviation of the high-temperature (125°C) energy consumption of the SRAM arrays with TA-DVS are 8.85 pJ and 425 fJ, respectively. The 3-sigma offset of the highest energy consumption in the circuits with the TA-DVS technique is (mean+3*standard deviation) 10.12 pJ. These results indicate that the highest possible energy consumption of a circuit with TA-DVS is still lower than the lowest possible energy consumption of a standard constant-$V_{DD}$ circuit when the parameter fluctuations are considered. The effectiveness of the proposed temperature-adaptive dynamic voltage scaling technique for enhancing the high-temperature energy efficiency in subthreshold memory arrays is therefore maintained in the presence of process parameter and supply voltage variations.

7. Conclusions

In ultra low-power-supply voltage subthreshold memory arrays, the circuit performance is enhanced with the increased temperature. The excessive timing slack observed in the clock period at elevated temperatures provides new opportunities to lower the active-mode energy consumption without violating the constant clock frequency requirement. Temperature-adaptive dynamic supply voltage tuning technique is proposed in this paper to reduce the high-temperature energy consumption of ultra low-voltage subthreshold memory circuits.

The temperature-adaptive supply voltage scaling technique dynamically adjusts the power supply voltage of a circuit based on the die temperature fluctuations. The high-temperature energy consumed with the temperature-adaptive voltage scaling technique is reduced by up to 32.8% as compared to the minimum energy achievable with a standard constant-$V_{DD}$ and constant-frequency 64-bit × 64-bit SRAM array in a 180 nm CMOS technology. An alternative technique based on temperature-adaptive reverse body bias that dynamically tunes the threshold voltages of the devices based on the fluctuations of the die temperature and the circuit speed is also evaluated in this paper. Temperature-adaptive dynamic supply voltage tuning technique is shown to be very effective to reduce the high-temperature energy consumption without degrading the clock frequency in subthreshold memory circuits operating at ultra low-power-supply voltages.

References


