

Robust 6T Si tunneling transistor SRAM design

Xuebei Yang and Kartik Mohanram

Department of Electrical and Computer Engineering, Rice University, Houston

xbyang@rice.edu kmram@rice.edu

Abstract

SRAMs based on tunneling field effect transistors (TFETs) consume very low static power, but the unidirectional conduction inherent to TFETs calls for special care when designing the SRAM cell. In this work, we make the following contributions. (i) We perform the first study of 6T TFET SRAMs based on both n-type and p-type access transistors and determine that only inward p-type TFETs are suitable as access transistors. However, even using inward p-type access transistors, the 6T TFET SRAM achieves only the write *or* the read operation reliably. (ii) In order to improve the reliability of 6T TFET SRAMs, we perform the first study of four leading write-assist (WA) and four leading read-assist (RA) techniques in TFET SRAMs. We conclude that the 6T TFET SRAM with GND lowering RA is the most reliable 6T TFET SRAM during write and read, and we verify that it is also robust under process variations. It also achieves the best performance and reliability, as well as the least static power and area, in comparison to other existing TFET SRAM structures. Further, it not only has comparable performance and reliability to the 32nm 6T CMOS SRAM, but also consumes 6–7 orders of magnitude lower static power, making it attractive for low-power high-density SRAM applications.

1. Introduction

With technology scaling, static power has emerged as a significant component of total circuit power consumption, contributing more than 50% of the total power in processors at the current technology node [1]. The majority of static power in processors is contributed by static random access memory (SRAM) [2, 3], which is widely used in processors in caches, buffers, reservation stations, etc. and occupies the majority of the processor chip area [3, 4]. Therefore, various methods, such as sleep transistors, multiple threshold voltage, and virtual ground [2, 5–7] have been proposed to reduce the static power of SRAM. However, the fundamental limit of the subthreshold swing of MOSFETs is 60mV/dec at room temperature [8], which limits the static power reduction that can be achieved. To further lower SRAM static power, new device structures with lower subthreshold swing are being widely explored.

Recently, tunneling field effect transistors (TFETs henceforth) have attracted strong interest in low power applications [9–13]. In comparison to traditional MOSFETs, the most exciting highlight is that TFETs exhibit lower subthreshold swing and lower off current. A subthreshold swing of 52.8mV/dec and off current of 10^{-14} A/ μ m have been experimentally demonstrated in [9, 12], and even lower values have been predicted in simulation studies [11]. Furthermore, since the fabrication of Si TFETs is fully compatible with the fabrication process for MOSFETs [10], Si TFETs can leverage the prior investment in CMOS technology. These advantages make TFETs very attractive in low-power SRAM designs as a substitute for MOSFETs. However, unlike traditional MOSFETs whose source and drain are interchangeable, the source and drain

of TFETs are determined during fabrication. As a result, TFETs are only suitable to conduct current in one direction. This unique characteristic of TFETs, referred to as unidirectional conduction, poses a significant challenge in the design of the access transistors of the 6T TFET SRAM since the access transistors are required to conduct current in both directions.

In this work, we make the following two contributions. First, whereas previous studies have only investigated n-type access transistors, we perform the first study of the 6T TFET SRAM based on both n-type and p-type access transistors. We observe that inward p-type access transistors result in the only configuration that achieves successful write and read operations while consuming low static power. However, in comparison to the 6T CMOS SRAM, the write and read margins of the 6T TFET SRAM are low and it cannot achieve high write and read margins simultaneously.

Our second contribution addresses the limitation imposed by the unidirectional conduction property and improves write/read reliability of the 6T TFET SRAM. Existing works on this topic have all focused on proposing new unconventional TFET SRAM structures [14, 15]. However, these approaches introduce large overhead in either area or static power. In contrast, we propose to keep the standard 6T TFET SRAM structure and to utilize one write assist (WA) *or* read assist (RA) technique to improve the write *or* read reliability. By not adopting new SRAM structures, the overhead of static power consumption and area can be avoided or kept at the minimum level. However, the biggest challenge for this approach is the selection of the best WA or RA technique. Since no prior work has investigated the effectiveness of WA and RA techniques in TFET SRAMs, we first perform a complete study of the effectiveness of four leading WA and four leading RA techniques on TFET SRAMs. We investigate the effectiveness of each WA/RA technique to improve write/read reliability and use Monte-Carlo simulation to further examine their behavior under process variations. We observed that the most effective design strategy is to size the 6T TFET SRAM to achieve robust write with the cell ratio $\beta \approx 0.6$, and to use GND lowering RA to improve read reliability.

We finally compare the proposed 6T TFET SRAM with inward p-type access transistors and GND lowering RA to previously reported TFET SRAMs [14, 15] and the 32nm 6T CMOS SRAM. Our proposed 6T TFET SRAM has better performance and reliability than existing TFET SRAMs [14, 15], occupying 10–15% less area and consuming at least 4 orders of magnitude lower static power than the TFET SRAMs in [14] and [15], respectively. Further, it not only has comparable performance and reliability to the 32nm 6T CMOS SRAM, but also consumes 6–7 orders of magnitude lower static power, making it attractive for low-power high-density SRAM applications.

Our TFET design was implemented and simulated using Sentaurus TCAD. The TFET has a leakage current of the order of 10^{-17} A/ μ m at $V_{DS} = 1$ V, which is 6 orders of magnitude lower than the 32nm MOSFET [16]. The I - V and C - V performance data are extracted for a range of device parameters and operating conditions. The I - V and C - V performance are then stored in two dimensional lookup tables, which are used by Verilog-A to implement the circuit simulation model.

The authors acknowledge Prof. Vijay Narayanan, Prof. Suman Datta, and Dheeraj Mohata at the Penn State University for help with TCAD modeling of TFETs, and Dr. Vikas Chandra at ARM R&D for helpful discussions. This research was supported by NSF CAREER Award CCF-0746850.

This paper is organized as follows. Section 2 provides a background for TFETs. Section 3 presents a complete study of the 6T TFET SRAM based on various types of access transistors. We then compare four leading WA and four leading RA techniques in Section 4 and determine the best 6T TFET SRAM design. The proposed 6T TFET SRAM is compared with existing TFET SRAMs and the 6T CMOS SRAM in Section 5. Section 6 is a conclusion.

2. TFET background

Figure 1 presents the device structure of the Si n-type and p-type TFETs (nTFET and pTFET henceforth) considered in this work. Both TFETs have a channel length of 32nm, gate underlap of 2nm, drain/source doping of 10^{20} dopants/cm³, and channel doping of 10^{15} dopants/cm³. A 2nm-thick HfO₂ with dielectric constant 25 is used as the gate insulator. All the parameters are comparable to the values in previous TFET studies, e.g., [11, 13]. As shown in the figure, the main difference between TFETs and MOSFETs is that the source and drain of a TFET are inversely doped. nTFETs consist of a p-doped source, an n-doped drain, and an intrinsic or slightly n-doped channel, whereas pTFETs consist of an n-doped source, a p-doped drain, and an intrinsic or slightly p-doped channel.

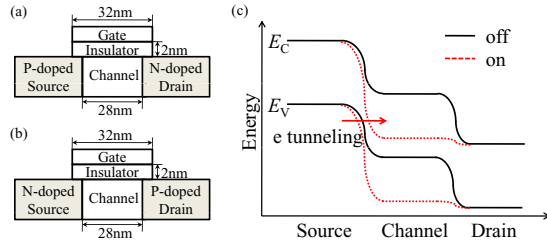


Figure 1: Device structure of (a) n-type TFET (nTFET) and (b) p-type TFET (pTFET). (c) nTFET energy band structure.

Figure 1(c) presents the energy band diagram of an nTFET, which will be used to illustrate the operating principle of TFETs. The operation of TFETs depends mainly on band-to-band tunneling current. In the off state, the conduction band E_C in the channel is higher than the valence band E_V at the source side, and the depletion region associated with the source to channel tunnel junction is thick, as shown by the solid line. As a result, the transmission probability for electron tunneling is low, resulting in low off current. In the on state, however, E_C in the channel is lowered below E_V at the source and the depletion region becomes thinner, as shown by the dotted line. As a result, electrons can tunnel from E_V at source to E_C inside the channel, resulting in band-to-band tunneling and a significant increase in the on current. Compared to conventional MOSFETs, whose current depends on thermionic conduction, the tunneling current of TFETs is more sensitive to the terminal biases, and hence TFETs have a much lower subthreshold swing.

In this work, the TFETs are simulated in Sentaurus TCAD using a non-local tunneling model. The non-local tunneling model simulates the spatial charge transfer across the tunneling barrier by considering the potential profile along the entire tunneling path. In Figure 2(a), we plot the I_{DS} - V_{GS} behavior of an nTFET and a pTFET for $V_{DS}=1V$ and $V_{DS}=-1V$, respectively. The gate work function is modulated to obtain an on current of $10^{-4} A/\mu m$ and an off current of $10^{-17} A/\mu m$. We also present the symbols for nTFETs and pTFETs. The arrow denotes the direction of current flow for TFETs under forward bias, which is from drain to source for nTFETs and from source to drain for pTFETs.

In Figure 2(b), we plot the I_{DS} - V_{GS} behavior of an nTFET when the drain and the source are switched; i.e., the p-doped contact becomes the drain while the n-doped contact becomes the source. We

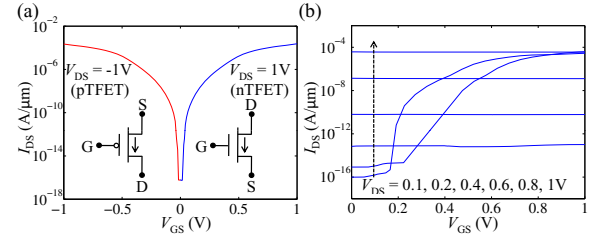


Figure 2: (a) The I - V performance for a pTFET and an nTFET under forward bias. The TFETs have an on current of $10^{-4} A/\mu m$ and an off current of $10^{-17} A/\mu m$. (b) The I - V performance of an nTFET under reverse bias. (The drain and the source are switched.) Note that (i) the gate loses control over the channel conduction at high V_{DS} , and (ii) the on current is much smaller than the on current under forward bias condition except for V_{DS} close to 1V or 0V.

refer to this bias condition as the reverse bias condition. It can be observed that while the current increases with the increase in gate voltage at low V_{DS} , for high V_{DS} , the gate has lost control over the drain current and the TFET does not behave as a transistor. It is also observed that the on current under the reverse bias condition is much smaller than the on current under the forward bias condition, except for V_{DS} close to 1V or 0V. These observations indicate that the drain and the source of a TFET cannot be switched and TFETs are only suitable to conduct current in one direction. We refer to this property of TFETs as unidirectional conduction.

In this work, the I - V and C - V TFET data are stored in two-dimensional lookup tables, which are then used by Verilog-A to implement a lookup table based model for circuit simulation. Since there is currently no compact model for TFETs, this method provides an efficient and accurate way to model emerging devices.

3. 6T TFET SRAM study

In Figure 3(a), we present the circuit structure of the 6T CMOS SRAM cell. The data in the SRAM cell is stored at nodes q and qb, which are connected by two cross-coupled inverters. Two transistors M3 and M6, called the access transistors, serve to write and read the data. Recently, TFET-based SRAMs have attracted strong interest for low-power applications. In the previous section, we have shown that the leakage current of TFETs is of the order of $10^{-17} A/\mu m$, which is 6 orders of magnitude lower than the leakage current in 32nm CMOS. Therefore, TFET SRAMs are projected to consume significantly less static power than CMOS SRAMs. However, different from CMOS, TFETs exhibit unidirectional conduction characteristics as explained above. While the transistors in the cross-coupled inverters (M1, M2, M4, M5) always conduct current in one direction, the access transistors need to conduct current in both directions. Therefore, special attention must be paid to choose the appropriate TFETs as the access transistors. For the 6T TFET SRAM, due to the inherent unidirectional conduction, there are four possible choices of access transistors: inward nTFETs/pTFETs and outward nTFETs/pTFETs, as shown in Figure 3(b)–(e). Whereas existing studies on TFET SRAMs have considered only inward/outward nTFET access transistors, we perform a comprehensive study of the 6T TFET SRAM based on all four types of access transistors. In the rest of this section, we use the two key metrics of static power and cell stability to evaluate the different access transistor configurations. The default supply voltage V_{DD} is 0.8V in this work, which is consistent with the values used in the literature.

Static power: During hold condition, both the bitlines are traditionally clamped at V_{DD} or slightly below V_{DD} (weakly clamped) [17].

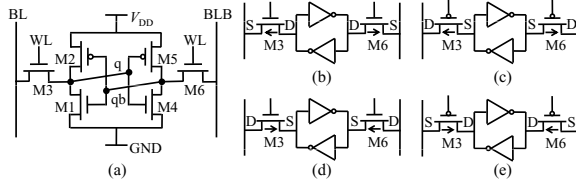


Figure 3: (a) 6T CMOS SRAM design. (b)–(d) are four types of access transistors for 6T TFET SRAM. (b) Outward n-type. (c) Outward p-type. (d) Inward n-type. (e) Inward p-type.

For outward pTFET and nTFET access transistors, one of the two access transistors will be under reverse bias. For example, consider outward nTFETs as shown in Figure 3(b). The source of M3 is connected to the bitline while the drain is connected to q. Therefore, the source voltage is higher than the drain voltage during hold condition if q stores ‘0’. As shown in Figure 2, the drain current of a TFET under reverse bias is usually much larger than the off current under forward bias regardless of V_{GS} . As a result, the 6T TFET SRAM based on outward access transistors will consume very high static power. In contrast, for inward access transistors, neither M3 or M6 is reverse biased during hold, so the static power consumption will be much lower. In our simulations, the 6T TFET SRAM based on outward access transistors consumes 5 and 9 orders of magnitude higher static power than the TFET SRAM based on inward access transistors at supply voltage of 0.6V and 0.8V, respectively. Since reducing static power is the motivation for TFET-based SRAMs, we conclude that outward TFETs are not suitable as the access transistors. (Note that if an SRAM architecture allows both bitlines to be clamped to ground instead of V_{DD} during hold condition, outward TFETs should be used as the access transistors instead of inward TFETs.)

Cell stability: Since outward TFETs have been shown to be unsuitable as the access transistors, we investigate only the cell stability of the 6T TFET SRAM based on inward access transistors. For a conventional SRAM cell, the read failure and write failure are two major sources of cell failure. The read failure is characterized by a flip in the state during read operation, whereas the write failure is defined by the failure to flip the cell state during write operation. In this work, the cell stability during read operation is quantified by the dynamic read noise margin (DRNM), which is the minimum voltage difference over time between q and qb during read [18]. The cell stability during write operation is characterized by the critical width of the wordline pulse (WL_{crit}), which is the minimum time required for the wordline pulse to flip the state of q and qb [19]. In contrast to prior work based on static read and write margins, this approach captures the dynamic behavior of read and write operation, and hence is more accurate.

In Figure 4(a) and 4(b), the DRNM and WL_{crit} of the 6T TFET SRAM based on inward nTFET and pTFET access transistors are plotted for different cell ratio β , which is the ratio of the width of nTFETs in the inverter and the access transistor. Note that for inward nTFET and pTFET, the WL_{crit} is infinite for all β and for $\beta > 1$, respectively, and is not plotted in these regions in Figure 4(b). Since infinite WL_{crit} indicates a write failure, we conclude that inward nTFETs cannot be used as the access transistors, and inward pTFETs are the only suitable choice in that they can provide both low static power and successful read/write operation.

In Figure 4, we also plot the DRNM and WL_{crit} of the 6T CMOS SRAM, simulated using the 32nm PTM model [16]. For DRNM, the 6T TFET SRAM based on inward pTFET access transistors (6T inpTFET SRAM henceforth) and the 6T CMOS SRAM do not have a large difference at large β . However, for small β , the 6T CMOS SRAM significantly outperforms the 6T inpTFET SRAM because

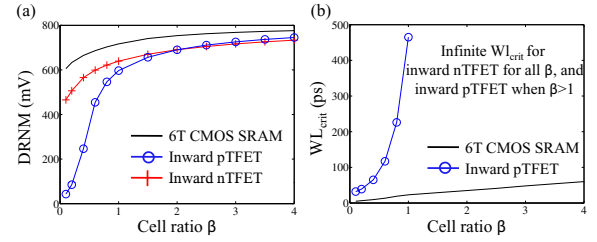


Figure 4: (a) DRNM and (b) WL_{crit} comparison for the 6T TFET SRAM using inward nTFET/pTFET access transistors and the 6T CMOS SRAM.

pTFET access transistors are more likely to pull up the node storing ‘0’ during read, especially when the access transistors are much larger than the pull-down transistors. For WL_{crit} , the 6T inpTFET SRAM is inferior to the 6T CMOS SRAM for all β . This inferiority can be mainly attributed to the fact that in the CMOS SRAM, both the access transistors are conducting during write operation, yet only one access transistor is conducting in the 6T inpTFET SRAM due to the unidirectional conduction, as shown in Figure 5. Therefore, the minimum time required to flip the state will be longer for the 6T inpTFET SRAM. Since this problem arises from the unidirectional conduction inherent to TFETs, the same problem remains when using inward nTFET or outward access transistors. We also observe that the value of β has a much larger effect on the 6T TFET SRAM than the 6T CMOS SRAM. This is because in the 6T CMOS SRAM, the major obstacle during write is the pull-up device, yet for the 6T TFET SRAM based on inward access transistors, only the access transistor connected to the node that initially stores ‘0’ is conducting, so the major obstacle during write is the pull-down device. Therefore, since the size of pull-up device does not change as β increases, the 6T CMOS SRAM does not see a significant change in WL_{crit} , but the WL_{crit} of the TFET SRAM based on inward access transistors changes greatly.

In summary, we conclude that: (i) only inward pTFETs are suitable as the access transistors for the 6T TFET SRAM, and (ii) even using inward pTFET access transistors, the 6T TFET SRAM will still fail to achieve both write and read operation reliably.

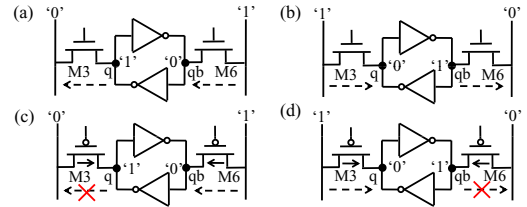


Figure 5: (a) and (b) are the current flow during write operation for the 6T CMOS SRAM. (c) and (d) are the current flow during write operation for the 6T inpTFET SRAM.

4. Exploration of WA and RA techniques

In order to address the inability of the 6T TFET SRAM to achieve both write and read operation reliably, previous works have focused only on proposing novel unconventional TFET SRAM structures. However, these novel structures sacrifice either the area or the static power consumption of the TFET SRAM. For example, in [14], a 7T TFET SRAM with separate read port was introduced, yet it incurs an area increase of 10–15%. In [15], an asymmetric 6T TFET SRAM that incorporates a modified version of GND raising WA was proposed. However, unless the SRAM architecture allows both bitlines to float instead of being clamped at V_{DD} during hold condition, the asymmetric 6T TFET SRAM will incur a large increase in static power (e.g., 4 orders of magnitude at $V_{DD}=0.5V$).

In this work, we propose to address this limitation as follows. The standard 6T TFET SRAM structure is preserved in this work, using inward pTFETs as the access transistors. However, we propose to utilize one WA *or* RA technique to improve the write *or* read operation. Specifically, the TFET SRAM is first sized to ensure reliable read *or* write operation, and then one WA *or* RA technique will be utilized to further enable reliable write *or* read operation, thereby achieving both write and read operation successfully. By not adopting new TFET SRAM structures, the overhead of static power and area can be avoided or kept at the minimum level. One challenge, however, is the selection of the most effective WA or RA technique. In CMOS SRAMs, various WA and RA techniques have been proposed and studied [20, 21]. However, since TFETs are inherently different from CMOS, the results of the studies on the CMOS SRAM cannot be applied to the TFET SRAM directly. Therefore, in this work, we evaluate the effectiveness of four leading WA and four leading RA techniques on the write and read stability improvement of the proposed 6T inpTFET SRAM.

4.1 Write assist techniques

In this subsection, we explore the effectiveness of four leading WA techniques [20]: V_{DD} lowering, GND raising, wordline lowering, and bitline raising. Note that for the traditional CMOS SRAM using nMOS access transistors, wordline voltage should be raised to assist write operation, yet since the 6T TFET SRAM uses pTFETs as the access transistors, it is the wordline lowering that could assist write. Also note that when adopting WA techniques, the cell ratio β of the 6T TFET SRAM must be larger than 1, so that the read operation can be achieved reliably in all cases. For the sake of fair comparison, the same percentage (30% of V_{DD}) of increase or decrease in voltage levels is imposed for all WA techniques.

V_{DD} lowering: Figure 6(a) presents the timing relationships for V_{DD} lowering WA. A lowered V_{DD} reduces the strength of the cross-coupled inverters with respect to the access transistor, thereby helping the write process. In SRAM design, V_{DD} lowering is usually implemented via a second power supply or on-chip regulator.

GND raising: Figure 6(b) presents the timing relationships for GND raising WA. Similar to V_{DD} lowering, a raised GND also reduces the strength of the cross-coupled inverters with respect to the access transistor, therefore helping the write process. Similar to V_{DD} lowering, GND raising is usually implemented via a separate ground or on-chip regulator.

Wordline lowering: Figure 6(c) presents the timing relationships for wordline lowering WA. In contrast to V_{DD} lowering and GND raising, wordline lowering improves the write operation by increasing the gate voltage of the access transistors, thereby increasing its drive strength. In SRAM design, wordline lowering can be implemented by charge pump or by capacitive coupling.

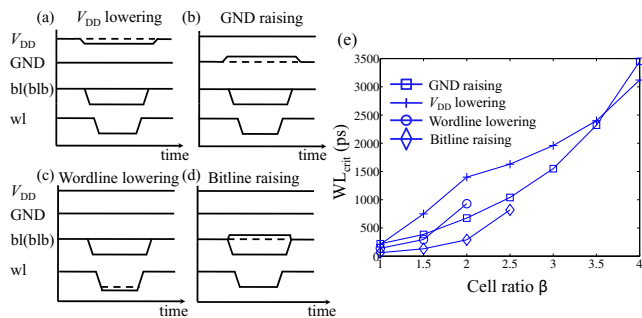


Figure 6: (a)–(d) The timing relationships for four WA techniques. (e) Simulation results for WA techniques.

Bitline raising: Figure 6(d) presents the timing relationships for bitline raising WA. By using bitline raising, the drain-source voltage of the conducting access transistor is raised, thereby increasing its drive strength. Similar to wordline lowering, bitline raising can be implemented by charge pump or by capacitive coupling.

Figure 6(e) compares the effectiveness of different WA techniques. It can be observed that at low β , wordline lowering and bitline raising are better than V_{DD} lowering and GND raising, indicating that increasing the strength of the access transistors is more effective than reducing the strength of the cross-coupled inverters. However, as β increases, the advantage of wordline lowering and bitline raising quickly vanishes. For $\beta > 2.5$, both the techniques fail to help achieve successful write. In contrast, V_{DD} lowering and GND raising can still enable successful write operation for large β because the pull-down devices are the main obstacle during write. As the size of the pull-down devices increases, it is more effective to reduce the strength of pull-down devices than to increase the strength of access transistors.

4.2 Read assist techniques

We next explore the effectiveness of four RA techniques: V_{DD} raising, GND lowering, wordline raising, and bitline lowering. In adopting RA techniques, the cell ratio β is kept below 1 so that the write operation can be achieved reliably in all cases. For fair comparison, the percentage of increase or decrease in voltage for all RA techniques are still kept at 30% of V_{DD} .

V_{DD} raising: Figure 7(a) presents the timing relationships for V_{DD} raising RA. As V_{DD} increases, the strength of the cross-coupled inverters also increases with respect to the access transistor, and hence the cell is more stable during read operation. In SRAM design, V_{DD} raising is usually implemented via a second power supply or on-chip regulator.

GND lowering: Figure 7(b) presents the timing relationships for GND lowering RA. Similar to V_{DD} raising, a lowered GND also increases the strength of the cross-coupled inverters with respect to the access transistor, thereby improving the read stability. In SRAM design, GND lowering is usually implemented via a second power supply or on-chip regulator.

Wordline raising: Figure 7(c) presents the timing relationships for wordline raising RA. By raising the wordline voltage, the gate-source voltage of the access transistors decreases, so the drive strength of the access transistor is reduced, therefore improving the read stability. In SRAM design, wordline raising can be implemented by charge pump or by capacitive coupling.

Bitline lowering: Figure 7(d) presents the timing relationships for bitline lowering RA. Similar to wordline raising, bitline lowering reduces the strength of the access transistor by reducing both the gate and drain voltage, hence improving the read stability. In SRAM design, bitline lowering can be implemented by charge pump or by capacitive coupling.

Figure 7(e) presents the effectiveness of different RA techniques. It can be observed that at large cell ratio β , V_{DD} raising and GND lowering achieve higher read stability, indicating that it is more effective to increase the strength of the cross-coupled inverters than to decrease the strength of the access transistors. As β decreases, the access transistor becomes larger, and it becomes more effective to reduce the strength of the access transistors than to increase the strength of the cross-coupled inverter. For $\beta < 0.4$, wordline raising achieves the best read stability.

In Figure 8, we present the comparison between all the WA and RA techniques. DRNM is set as the x-axis and the achieved WL_{crit}

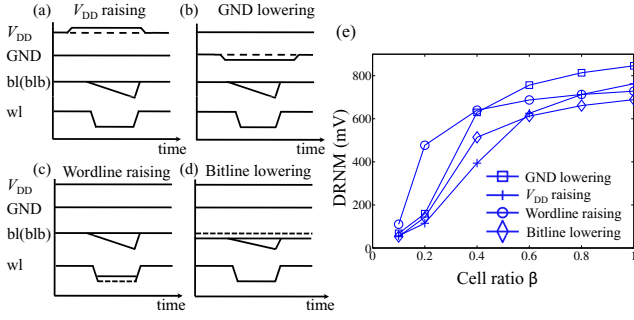


Figure 7: (a)–(d) The timing relationships for four RA techniques. (e) Simulation results for RA techniques.

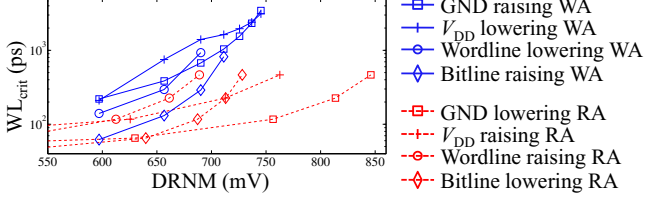


Figure 8: Comparison of WA and RA techniques.

for each WA or RA technique at corresponding DRNM is plotted on the y-axis. In order to achieve both write and read reliably, the DRNM should be large and the WL_{crit} should be small, so we are interested in the curve that is closest to the lower-right corner of the figure. Based on this criterion, we conclude that GND lowering RA is most effective among all the WA and RA techniques.

4.3 Impact of process variation

As the dimension of TFETs enters the nanometer region, process variations are inevitable during fabrication. Therefore, it is important to investigate the impact of process variations on the WA and RA techniques. In this work, we restrict the variations in TFETs to the gate insulator thickness, while other important sources of process variations for MOSFETs such as channel length variation and random dopant fluctuations are not considered because (i) previous work [13] has shown that channel length variation has negligible effects on TFETs, and (ii) the effects of random dopant fluctuations on TFETs are also anticipated to be limited due to the lightly doped channel in TFETs. As reported in prior work, the gate insulator thickness can be controlled to within 5% using novel fabrication techniques [13], and we restrict the study of variations in gate insulator thickness to $\pm 5\%$ in this paper.

The impact of process variations on WA and RA techniques are analyzed using Monte-Carlo simulations. For WA techniques, the cell ratio β is set to 2, and the WL_{crit} simulation results are presented in Figure 9(a)–(c). We observe that WL_{crit} varies greatly under process variations for GND raising, V_{DD} lowering, and bitline raising, while wordline lowering even sees infinite WL_{crit} (write failure) under process variations (the simulation results are hence not shown in the figure). In contrast, the DRNM is hardly influenced by process variations in Figure 9(d). For RA techniques, the cell ratio β is set to 0.6, and the DRNM simulation results are presented in Figure 10(a)–(d). We observe that for all RA techniques, the DRNM is minimally impacted by process variations. It is also observed from Figure 10(e) that the impact of process variations on WL_{crit} is much smaller than the case using WA techniques, mostly due to the much stronger access transistors.

In summary, we conclude that the 6T TFET SRAM should be sized at small β value to ensure reliable write operation, complemented by GND lowering RA to improve the read reliability.

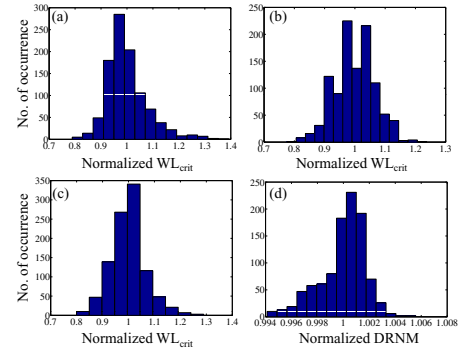


Figure 9: Impact of process variations on WL_{crit} for (a) GND raising, (b) V_{DD} lowering, and (c) bitline raising. (d) Impact of process variations on DRNM when the SRAM is sized to use WA techniques.

($\beta \approx 0.6$ is a good value to achieve the best tradeoff between WL_{crit} and DRNM.) The 6T inPTFET with GND lowering RA not only enjoys sufficient write and read margin, but also shows strong immunity to process variations. Some drawbacks of this design, however, do exist. There is dynamic power overhead to generate lowered GND and lowered DRNM for half-selected cells due to the small β . Fortunately, various techniques have been proposed in CMOS SRAMs to effectively mitigate these effects, such as the segmented GND technique [7] and the weakly-clamped bitline during hold [17]. These techniques can also be adopted by our design, and hence we believe these shortcomings can be readily addressed.

5. Results comparison

In order to further examine the effectiveness of the proposed 6T inPTFET SRAM with GND lowering RA, in this section, we compare the performance characterized by the write/read delay, the reliability characterized by the DRNM and WL_{crit} , the static power, and the area of the 6T inPTFET with ground lowering (β set to 0.6) with the 6T CMOS SRAM, asymmetric 6T TFET SRAM [15], and the 7T TFET SRAM [14]. We use the 32nm low-power PTM model [16] for 6T CMOS SRAM simulation, and use the same TFET presented in Section 2 to simulate the asymmetric 6T TFET SRAM and 7T TFET SRAM. V_{DD} is varied from 0.5V to 0.9V, which is the preferred operating voltage range for TFETs.

Performance: The write and read delay comparison is shown in Figure 11. For the write delay, the 6T CMOS SRAM has smaller delay than all the TFET SRAMs over most V_{DD} . This advantage comes from the fact that CMOS can conduct current in both directions. Among the three TFET SRAMs, the proposed 6T inPTFET SRAM with GND lowering RA has the minimum delay except for $V_{DD} = 0.5V$. This advantage can be attributed to the fact that it is sized to favor the write operation. For the read delay, the RA technique helps the proposed 6T inPTFET SRAM with GND lowering RA to achieve the minimum delay under low V_{DD} . For $V_{DD} > 0.8V$, the 6T CMOS SRAM has the smallest read delay, yet the proposed 6T TFET SRAM still outperforms other TFET SRAMs.

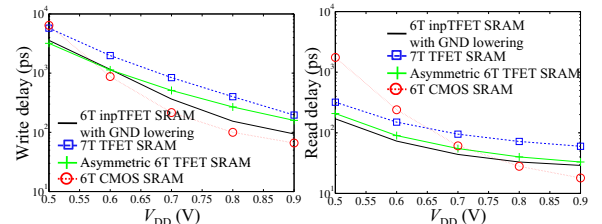


Figure 11: Comparison of (a) write and (b) read delay

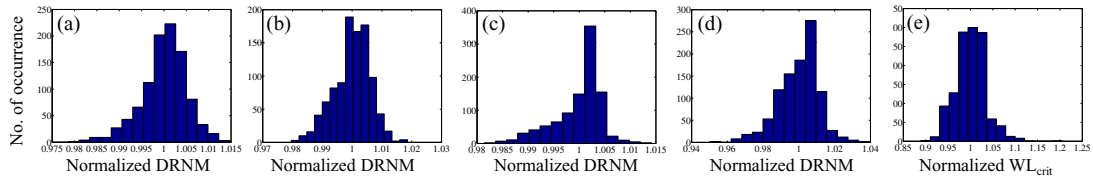


Figure 10: Impact of process variations on DRNM for (a) GND lowering, (b) V_{DD} raising, (c) wordline raising, and (d) bitline lowering. (e) Impact of process variations on WL_{crit} when the SRAM is sized to use RA techniques.

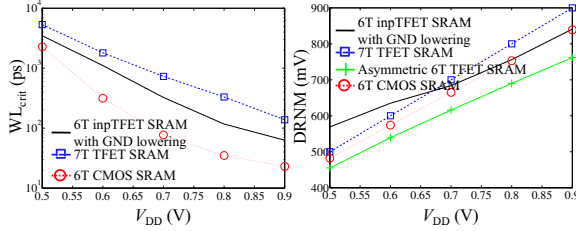


Figure 12: Comparison of (a) write and (b) read margins

Reliability: The WL_{crit} and DRNM comparison is shown in Figure 12. Notice that WL_{crit} for the asymmetric 6T TFET SRAM cannot be defined since it does not have the separatrix, so in Figure 12(a), there are only three curves. It can be observed that all TFET SRAMs have larger WL_{crit} than the 6T CMOS SRAM due to the inherent unidirectional conduction of TFETs. Among TFET SRAMs, the proposed 6T inptFET SRAM with GND lowering RA has the smallest WL_{crit} . For the DRNM, the 7T TFET SRAM has the highest DRNM at high V_{DD} due to the use of an additional read buffer. However, for V_{DD} under 0.7V, the 6T inptFET SRAM with GND lowering RA achieves the highest DRNM.

Static power: Due to the inherent low leakage of TFETs, the 6T inptFET SRAM with GND lowering RA and the 7T TFET SRAM consume the same static power, which is 6–7 orders of magnitude lower than the 6T CMOS SRAM. Note that although the 7T TFET SRAM uses outward TFETs as the access transistors, its write bitlines and read bitlines are separate, and the write bitlines can be set to 0 during hold condition to avoid excessive leakage. In contrast, unless the SRAM architecture allows both bitlines to float instead of being clamped at V_{DD} during hold, the usage of outward access transistors causes the asymmetric 6T TFET SRAM to consume much higher static power (4 orders of magnitude higher than the proposed 6T inptFET SRAM and the 7T TFET SRAM at $V_{DD} = 0.5V$ for example).

Cell Area: The 6T CMOS SRAM, the asymmetric 6T TFET SRAM and the proposed 6T inptFET SRAM with GND lowering RA have the minimum number of transistors, and hence occupy the least area. The 7T TFET SRAM uses an extra transistor and will introduce an unavoidable area increase of 10–15% [14].

In summary, the proposed 6T inptFET SRAM with GND lowering RA achieves the best performance and reliability compared to existing TFET SRAMs. Further, it consumes at least 4 orders of magnitude lower static power than the asymmetric 6T TFET SRAM, while reducing area by 10–15% over the 7T TFET SRAM. Compared to the 6T CMOS SRAM, the proposed 6T inptFET SRAM with GND lowering RA has comparable performance and reliability, but consumes 6–7 orders of magnitude lower static power.

6. Conclusions

In this paper, we perform the first study of the 6T TFET SRAM based on both n-type and p-type access transistors and determine that only inward pTFETs are suitable as access transistors. We also perform a comprehensive study of four leading write assist (WA)

and four leading read assist (RA) techniques to improve write or read stability of 6T TFET SRAMs. We observe that the best design strategy is to size the transistors for a $\beta \approx 0.6$ to favor the write operation, and to use GND lowering RA to improve the read stability. Using this design strategy, the proposed 6T TFET SRAM achieves the best performance and reliability as well as the minimum static power and area in comparison to existing TFET SRAMs. It further consumes 6–7 orders of magnitude lower static power than the 6T CMOS SRAM with comparable performance and reliability, making it attractive for low-power high-density SRAM applications.

References

- [1] *International Technology Roadmap for Semiconductors*. <http://public.itrs.net/>.
- [2] K. Zhang *et al.*, “SRAM design on 65-nm CMOS technology with dynamic sleep transistor for leakage reduction,” *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 895–901, 2005.
- [3] A. Pavlov and M. Sachdev, *CMOS SRAM Circuit Design and Parametric Test in Nano-scaled Technologies*. Springer, 2008.
- [4] C. Molina *et al.*, “Non redundant data cache,” in *Int. Symp. Low Power Electronics and Design*, pp. 274–277, 2003.
- [5] F. Hamzaoglu *et al.*, “Dual-Vt SRAM cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13 μ m technology generation,” in *Intl. Symp. Low Power Electronics and Design*, pp. 15–19, 2000.
- [6] R. W. Mann *et al.*, “Ultralow-power SRAM technology,” *IBM Journal of Research and Development*, vol. 47, pp. 553–566, 2003.
- [7] M. Sharifkhani *et al.*, “Segmented virtual ground architecture for low-power embedded SRAM,” *IEEE Trans. VLSI Systems*, vol. 15, pp. 196–205, 2007.
- [8] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI devices*. Cambridge University Press, 2009.
- [9] W. M. Reddick *et al.*, “Silicon surface tunnel transistor,” *Applied Physics Letters*, vol. 67, pp. 494–496, 1995.
- [10] P. F. Wang *et al.*, “Complementary tunneling transistor for low power applications,” *Solid-State Electronics*, vol. 48, pp. 2281–2286, 2004.
- [11] Q. Zhang *et al.*, “Low-subthreshold-swing tunnel transistors,” *IEEE Electron Device Letters*, vol. 27, pp. 297–300, 2006.
- [12] W. Y. Choi *et al.*, “Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec,” *IEEE Electron Device Letters*, vol. 28, pp. 743–745, 2007.
- [13] S. Saurabh *et al.*, “Estimation and compensation of process induced variations in nanoscale tunnel field effect transistors (TFETs) for improved reliability,” *IEEE Trans. Device and Materials Reliability*, vol. 11, pp. 1–21, 2010.
- [14] D. Kim *et al.*, “Low power circuit design based on heterojunction tunneling transistors (HETTs),” in *Intl. Symp. Low Power Electronics and Design*, pp. 219–224, 2009.
- [15] J. Singh *et al.*, “A novel Si-tunnel FET based SRAM design for ultra low-power 0.3V V_{DD} applications,” in *Asia and South Pacific Design Automation Conference*, pp. 181–186, 2010.
- [16] Y. Cao *et al.*, “New paradigm of predictive MOSFET and interconnect modeling for early circuit design,” in *Custom Integrated Circuits Conference*, pp. 201–204, 2000.
- [17] C. T-Chuang *et al.*, “High-performance SRAM in nanoscale CMOS: Design challenges and techniques,” in *IEEE Intl. Workshop on Memory Technology, Design, and Testing*, pp. 4–12, 2007.
- [18] W. Dehaene *et al.*, “Embedded SRAM design in deep deep submicron technologies,” in *European Solid State Conference*, pp. 384–391, 2007.
- [19] J. Wang *et al.*, “Analyzing static and dynamic write margin for nanometer SRAMs,” in *Int. Symp. Low Power Electronics and Design*, pp. 129–134, 2008.
- [20] V. Chandra *et al.*, “On the efficacy of write-assist techniques in low voltage nanoscale SRAMs,” in *Proc. Design, Automation and Test in Europe*, pp. 345–350, 2010.
- [21] R. W. Mann *et al.*, “Impact of circuit assist methods on margin and performance in 6T SRAM,” *Solid-State Electronics*, vol. 54, pp. 1398–1407, 2010.