

# Low Write-Energy STT-MRAMs using FinFET-based Access Transistors

Alireza Shafaei, Yanzhi Wang, and Massoud Pedram

Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089

{shafaeib, yanzhiwa, pedram}@usc.edu

**Abstract**—Spin-Transfer Torque Magnetic RAM (STT-MRAM) technology requires a high current in order to write data into memory cells, which gives rise to large access transistors in conventional MOS-accessed cells. On the other hand, FinFET devices offer higher ON current and denser layout compared with planar CMOS transistors. This paper thus proposes the design of an energy-efficient STT-MRAM cell which utilizes a FinFET access transistor. To assess the performance of the new cell, optimal layout-related parameters of the FinFET access transistor and the MTJ are analytically derived in order to minimize the STT-MRAM cell area. Afterwards, detailed cell- and architecture-level comparisons between FinFET- vs. MOS-accessed STT-MRAMs are performed. According to the comparison results, while the area of the MOS-accessed STT-MRAM increases significantly under 3ns write pulse width ( $\tau_w$ ), the FinFET-based design can effectively function under  $\tau_w = 2ns$ , at the cost of slight increase in the memory area. Hence, the FinFET-accessed STT-MRAM offers denser area and higher energy efficiency compared with the conventional MOS-accessed counterpart.

## I. INTRODUCTION

Non-volatility, high endurance, low leakage and CMOS compatibility are attractive features of STT-MRAMs which have turned STT-MRAMs into a strong candidate for low power and high performance memory designs [11], [15], [18]. The storage element of an STT-MRAM cell is a magnetic device which is accessed through an NMOS transistor. Unfortunately, the magnetic element requires a high current in order to change its stored data. Accordingly, the NMOS transistor should be scaled-up to support such current levels, resulting in a large access transistor.

STT-MRAM cells are denser than their SRAM counterparts [3], but due to the requirement of having large access transistors, they are not competitive with the flash and DRAM designs in terms of their density. The access transistor of an STT-MRAM cell may be 56 times larger than that of the magnetic storage element [20], and 9 times larger than the entire DRAM cell [13]. To alleviate this effect, a cross-point architecture was proposed [20], which amortizes the cost of large access transistors by sharing them among a number of storage elements. However, this architecture requires more sophisticated and specially-designed peripheral circuitry.

On the other hand, the structure of a FinFET device [17] allows more effective gate control (and less control by source and drain terminals) on the channel, which mitigates short channel effects and enhances the ON/OFF current ratio and soft-error immunity [10]. A FinFET device is also more compact than the CMOS counterpart, since its channel width can be adjusted by the fin height without impacting the device area. Moreover, FinFETs offer higher ON current under the same channel width compared with CMOS transistors. In other words, to obtain a specific current requirement, a FinFET with smaller width than that of the CMOS transistor can be used.

Since both STT-MRAM and FinFET technologies are compatible with the CMOS process [7], [17], utilizing a FinFET device instead of a CMOS access transistor in the STT-MRAM cell emerges as a potential solution to achieve a denser cell layout. This paper thus presents the design of a FinFET-accessed STT-MRAM cell. Layout of this new cell under various FinFET-specific parameters is analyzed in order to find the optimal cell design with minimum area for different process- and geometry-related variables. The impact of process variation on the access transistor and the magnetic element has also been investigated. Comparison with an optimized MOS-accessed design [6] shows that under practical process technologies and considering process variations, FinFET-accessed STT-MRAM cell is 34% and 85% smaller for 5ns and 3ns write pulse widths, respectively.

In order to assess the proposed cell in memory designs, FinFET support is added to NVSim [4], a circuit-level model for emerging nonvolatile memories. Accordingly, performance, energy consumption, and area of the FinFET-accessed STT-MRAM for various memory subarray designs are calculated and reported. Detailed comparison with MOS-accessed cell design proves the effectiveness of utilizing a FinFET device as the access transistor in STT-MRAM designs. In particular, in the range of write current requirements, the area of the FinFET-accessed STT-MRAM design varies slightly, whereas write latency and write energy consumption which are the main issues associated with STT-MRAMs reduce significantly. More precisely, the FinFET-accessed STT-MRAM design can effectively function under 2ns write pulse width, with a slight increase in the memory area.

The rest of the paper is organized as follows. Section II provides basic concepts of STT-MRAMs and FinFET devices. Layout of the FinFET-accessed STT-MRAM cell is analyzed in Section III. Cell- and architecture-level comparisons of FinFET- vs. MOS-accessed STT-MRAMs are presented in Section IV. Finally, Section V concludes the paper.

## II. BASIC CONCEPTS

In the following section, we briefly review related concepts in STT-MRAM cells and FinFET devices.

### A. STT-MRAM

An STT-MRAM cell is composed of a magnetic storage element, also known as the *magnetic tunneling junction* (MTJ), which is serially connected to an *access transistor* as shown in Figure 1(c). The MTJ contains two ferromagnetic layers, *fixed* and *free* layers, which are separated by an insulating layer. While the fixed layer has a fixed magnetic direction, the free layer's magnetic direction can be programmed to be either *parallel* (Figure 1(a)) or *anti-parallel* (Figure 1(b)) to the magnetic direction of the fixed layer, resulting in low ( $'0'$ ),  $R_P$ , or high ( $'1'$ ),  $R_{AP}$ , resistance states, respectively.

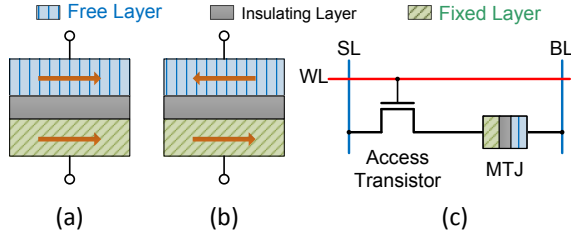


Fig. 1. MTJ in (a) parallel (low resistance,  $R_P$ ), and (b) anti-parallel (high resistance,  $R_{AP}$ ) states. (c) STT-MRAM cell.

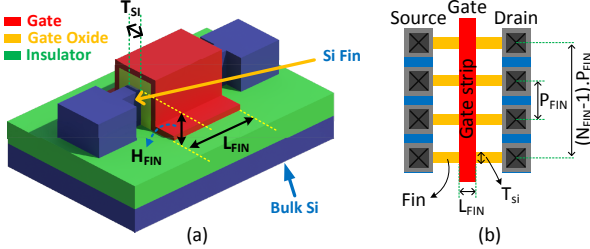


Fig. 2. FinFET device: (a) structure, (b) layout.

The access transistor is controlled by a word-line (WL), which itself is activated by the row address decoder, and provides a mechanism to write into or read from a desired memory cell. For writing into a cell, a high *write current* ( $I_W$ ) is needed to successfully change the state of the MTJ cell (i.e., changing the free layer's direction). This is achieved by applying a large voltage difference between source-line (SL) and bit-line (BL). Based on the polarity of this write voltage, MTJ is set to '0' (for positive voltage) or '1' (for negative voltage). Moreover, the minimum current that can change the state of the MTJ is called the *critical switching current*,  $I_C$ . Hence, in order to ensure the write-ability, we should have  $I_W \geq I_C$ .

To read from a cell, a small voltage difference between SL and BL is applied. The resultant *read current* ( $I_R$ ) is then compared to a reference current to determine the state of the MTJ (higher current reads-out '0', whereas lower current reads-out '1'). The read voltage should be small enough such that it does not change the free layer's magnetic direction (to ensure read stability,  $I_R < I_C$ ), but large enough such that it produces a distinguishable current between low and high resistances.

### B. FinFET Devices

FinFET device is a quasi-planar double-gate transistor [17]. This structure allows FinFETs to enhance the energy efficiency, ON/OFF current ratio, and soft-error immunity compared with bulk CMOS counterparts. As a result, FinFET technology is currently viewed as the substitute for the bulk CMOS for technology nodes from 32 nm and below [10], [12].

Figure 2(a) shows the structure of a three-terminal FinFET device. The main component is the *fin* which provides the channel for conducting current when the device is switched on. This vertical fin is surrounded by the gate, and hence a more efficient control over the channel is achieved which in turn helps to suppress short channel effects.

Key geometric parameters of a FinFET are related to the fin which include the height ( $H_{FIN}$ ), Silicon thickness ( $T_{SI}$ ), and

length ( $L_{FIN}$ ) of the fin (Figure 2(a)). The effective channel width of a single fin,  $W_{min}$ , is thus (approximately) equal to

$$W_{min} \approx 2 \times H_{FIN}. \quad (1)$$

Increasing the width (strength) of a FinFET device is achieved by connecting more fins in parallel. Hence, the layout area of a FinFET device is proportional to the number of fins.

### III. FINFET-ACCESSED STT-MRAM CELL

To provide the high write current in an STT-MRAM cell, the width of the access transistor should be increased, leading to an access transistor with larger area than the MTJ [20]. Therefore, the area (density) of the STT-MRAM cell is determined by the access transistor width as well as the process design rules. Replacing the MOS-based access transistor with a FinFET device is beneficial for two reasons: (1) The area (footprint) of a FinFET device may be reduced by increasing  $H_{FIN}$ , which is along the  $z$ -axis. (2) FinFET offers higher ON current than CMOS transistors under the same channel width. This means that for the same write current of the STT-MRAM cell, FinFET requires smaller channel width compared with the CMOS counterpart. More details on the layout of the FinFET-accessed STT-MRAM cell are presented below.

Layout of a three-terminal FinFET device with four fins is shown in Figure 2(b) [1]. A single strip is used for the gate terminal. Moreover, source (and also drain) terminals of multiple fins are connected together through a metal 1 wire to make a wider FinFET device. A critical process-related geometry in Figure 2(b) is the *fin pitch*,  $P_{FIN}$ , which is defined as the minimum center-to-center distance of two adjacent parallel fins. The value of  $P_{FIN}$  is determined by the underlying FinFET technology. More precisely, there are two types of FinFET technologies: (1) *Lithography-defined* technology where lithographic constraints limit the fin pitch spacing, and (2) *spacer-defined* technology which relaxes the constraints on  $P_{FIN}$ , and obtains  $2\times$  reduction in the value of  $P_{FIN}$  at the cost of a more elaborate and costly lithographic process [2]. The area of a FinFET device is thus proportional to  $(N_{FIN} - 1) \cdot P_{FIN}$ .

Major process-related FinFET geometries for 32nm and 45nm technologies are reported in Table I. As mentioned earlier, the cell area is inversely proportional to  $H_{FIN}$ , but the value of  $H_{FIN}$  is bounded by the  $H_{FIN}/T_{SI}$  parameter whose practical values range from 1 to 4 (depending on the etching technology). Accordingly,  $H_{FIN}$  cannot be increased arbitrarily. Table I also includes process design rules which are similar for FinFET and CMOS technologies (their difference is in the fin fabrication, which does not influence design rules [1]). Values of design rules are taken from [6] for comparison with CMOS baseline layout described next.

Gupta et al. [6] proposed an optimized layout for a MOS-accessed STT-MRAM cell, which will serve as our CMOS baseline. Following [6], the gate strip is shared among all cells in a row. Moreover, SL contacts of two neighboring cells in a column are shared, and thus half of the SL contact and spacing is considered in height calculations. Additionally, SL and BL are routed through a distinct metal layer, which connects the access transistor to the MTJ. Some possible layouts of FinFET-accessed STT-MRAM cells are shown in Figure 3.

**Cell Width.** Width of the STT-MRAM cell is constrained by metal-related design rules or the number of fins, which are referred to as *metal-constrained* (MC) and *fin-constrained* (FC) widths, respectively. Since SL and BL are vertically routed through two parallel metal wires, the cell width may be limited

TABLE I. PROCESS-RELATED FINFET GEOMETRIES AND DESIGN RULES

Parameter	Value in 32nm	Value in 45nm	Comment	Reference
$L_{FIN}$	35nm	45nm	Fin length	[1]
$T_{SI}$	23nm	30nm	Silicon thickness	
$H_{FIN}/T_{SI}$	{1, 2, 3, 4}	{1, 2, 3, 4}	Determines the fin height	
$P_{FIN}$ (lithography)	80nm	120nm	Fin pitch in lithography-defined technology	
$P_{FIN}$ (spacer)	40nm	60nm	Fin pitch in spacer-defined technology	
$W_M$	$3\lambda = 48\text{nm}$	$3\lambda = 67.5\text{nm}$	Minimum width of metal wires	[6]
$W_{M2M}$	$3\lambda = 48\text{nm}$	$3\lambda = 67.5\text{nm}$	Minimum spacing between metal wires	
$W_C$	$2\lambda = 32\text{nm}$	$2\lambda = 45\text{nm}$	Minimum contact size	
$W_{G2C}$	$2\lambda = 32\text{nm}$	$2\lambda = 45\text{nm}$	Minimum spacing between gate and contact	

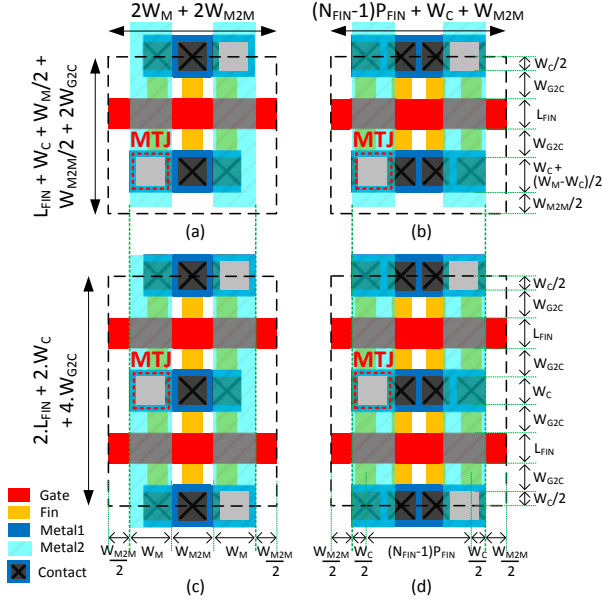


Fig. 3. Layout of the STT-MRAM cell for (a) single-finger, metal-constrained, (b) single-finger, fin-constrained, (c) two-finger, metal-constrained, and (d) two-finger, fin-constrained FinFET access transistors.

by metal width and metal spacing design rules (Figure 3(a) and Figure 3(c)):

$$W_{MC} = 2 \times W_M + 2 \times W_{M2M} = 2(3\lambda) + 2(3\lambda) = 12\lambda. \quad (2)$$

On the other hand, the number of fins is another factor that can limit the cell width. Given that a FinFET transistor with width of  $W$  is needed, the number of fins,  $N_{FIN}$ , is given by

$$N_{FIN} = \lceil \frac{W}{W_{min} \cdot N_f} \rceil, \quad (3)$$

where  $N_f$  is the number of fingers if a multi-finger device is used. For a fin-constrained STT-MRAM cell, the width is given by (Figure 3(b) and Figure 3(d)):

$$W_{FC} = (N_{FIN} - 1) \cdot P_{FIN} + W_C + W_{M2M} = (N_{FIN} - 1) \cdot P_{FIN} + 5\lambda. \quad (4)$$

As a result, the maximum of  $W_{MC}$  and  $W_{FC}$  determines the cell width,  $W_{Cell}$ :

$$W_{Cell} = \max(W_{MC}, W_{FC}) = \max(12\lambda, (N_{FIN} - 1) \cdot P_{FIN} + 5\lambda) = \begin{cases} 12\lambda & \text{if } N_{FIN} \leq \lceil 1 + \frac{7\lambda}{P_{FIN}} \rceil \\ (N_{FIN} - 1) \cdot P_{FIN} + 5\lambda & \text{otherwise} \end{cases} \quad (5)$$

**Cell Height.** The number of fingers is the main factor that determines the cell height. The height of a single-finger STT-MRAM cell, as shown in Figures 3(a) and 3(b), is given by

$$H_{1f} = L_{FIN} + W_C + W_M/2 + W_{M2M}/2 + 2 \times W_{G2C} = L_{FIN} + 9\lambda, \quad (6)$$

whereas the height of a two-finger cell, as shown in Figure 3(c) and Figure 3(d), can be computed as follows:

$$H_{2f} = 2 \times L_{FIN} + 2 \times W_C + 4 \times W_{G2C} = 2 \times L_{FIN} + 12\lambda. \quad (7)$$

In general, the height of an STT-MRAM cell with  $N_f$  fingers is calculated by:

$$H_{Cell} = \begin{cases} N_f \cdot H_{2f}/2 = N_f \cdot (L_{FIN} + 6\lambda) & \text{if } N_f \text{ is even} \\ (N_f - 1) \cdot H_{2f}/2 + H_{1f} = N_f \cdot (L_{FIN} + 6\lambda) + 3\lambda & \text{if } N_f \text{ is odd} \end{cases} \quad (8)$$

**Process-related and Design Variables.** According to (3), (5), and (8), values of  $H_{FIN}/T_{SI}$ ,  $P_{FIN}$ ,  $\lambda$ , and  $N_f$  dictate the cell area.  $H_{FIN}/T_{SI}$ ,  $P_{FIN}$ , and  $\lambda$  are process-related variables imposed by the technology, whereas  $N_f$  is a design variable of the cell.

Decreasing  $P_{FIN}$  and  $\lambda$ , and increasing  $H_{FIN}/T_{SI}$  always reduce the cell area (i.e., improve the cell density). In order to observe the effect of each of these process-dependent variables, we consider a baseline FinFET-accessed STT-MRAM with  $W = 250\text{nm}$ ,  $N_f = 1$ , and  $H_{FIN}/T_{SI} = 1$  in 45nm lithography-defined technology. The cell area is reduced by 43% when scaling down from 45nm to 32nm, and by 42% when the spacer-defined technology is used. Furthermore, compared to  $H_{FIN}/T_{SI} = 1$ , setting  $H_{FIN}/T_{SI}$  to 2 and 4 reduces the cell area by 50% and 60%, respectively. On the other hand, increasing  $N_f$  causes the cell height to be dominated by the design rules overhead. Thus, FinFET-accessed STT-MRAM cells with  $N_f \geq 3$  never appear as the minimum area under any transistor width values.

**Minimum Cell Area.** Figure 4(a) shows the area of an STT-MRAM cell using single- and two-finger FinFET devices under various transistor width values. Results are based on the 32nm spacer-defined technology with  $H_{FIN}/T_{SI} = 2$ . The optimal MOS-accessed STT-MRAM cell [6] is also included in the figure for comparison. Additionally, the piecewise linear behavior of FinFET-accessed cells is due to the width quantization property of FinFET devices (i.e., the FinFET width can only take discrete values).

As can be seen in Figure 4(a), the cell width is initially constrained by metal-related design rules when the transistor width is relatively small. As a result, single- and two-finger STT-MRAM cells have the same width. However, since  $H_{1f} < H_{2f}$ , single-finger STT-MRAM cell achieves the minimum area. At the point specified by the blue dashed line in the figure, which will be referred to as the *threshold transistor width* ( $W_{th}$ ), the area of the two-finger STT-MRAM cell

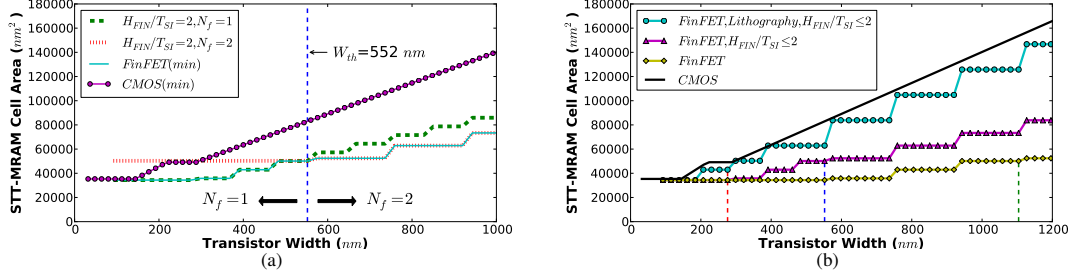


Fig. 4. STT-MRAM cell area as a function of transistor width. (a) Finding minimum area for FinFETs under fixed  $H_{FIN}/T_{SI}$  value. Initially, single-finger device is smaller. Higher than the threshold width, two-finger device obtains the minimum area. (b) Comparing minimum area of MOS- [6] vs. FinFET-accessed cells. FinFET under strict ( $\circ$ ), moderate ( $\Delta$ ), and aggressive ( $\diamond$ ) technologies achieve smaller area than CMOS. Vertical dashed lines on both figures point to the threshold transistor width of the corresponding plot.

becomes smaller than that of the single-finger device. At  $W_{th}$ , the single-finger device becomes fin-constrained, whereas the two-finger device is still metal-constrained, and hence  $W_{th}$  is obtained as follows.

$$H_{1f} \cdot W_{FC} = H_{2f} \cdot W_{MC},$$

which results in

$$W_{th} = W_{min} \cdot \left(1 + \frac{\lambda}{P_{FIN}} \cdot \left(12 \cdot \frac{H_{2f}}{H_{1f}} - 5\right)\right). \quad (9)$$

Therefore, for a given  $H_{FIN}/T_{SI}$  value, if  $W \leq W_{th}$  then using a single-finger STT-MRAM cell leads to the minimum cell area; otherwise, a two-finger cell has to be employed to obtain the minimum area.

The minimum achievable cell areas for the following three cases are presented in Figure 4(b): (1) lithography-defined FinFET technology is used and the value of  $H_{FIN}/T_{SI}$  is limited to at most 2, (2) constraint on FinFET technology is relaxed but still  $H_{FIN}/T_{SI} \leq 2$ , and (3) no constraint is imposed on FinFET-specific parameters. As shown in the figure, under the same transistor width, the area of the FinFET-accessed STT-MRAM cell in all three cases is smaller than the area of the MOS-accessed cell. In particular, for the range of transistor widths specified in the figure ( $0 < W \leq 1200$ nm), cases (1), (2), and (3) on average reduce the MOS-accessed cell area by 11%, 37%, and 48%, respectively. The minimum area in case (3) is achieved by an aggressive value of  $H_{FIN}/T_{SI} = 4$ , using the spacer-defined FinFET technology.

#### IV. COMPARISON RESULTS

In this section, we provide detailed comparisons of FinFET- vs. MOS-accessed STT-MRAMs at cell and architecture levels.

##### A. Cell-level Comparison

So far, we compared the cell area of FinFET- vs. MOS-accessed STT-MRAMs under the same transistor width. However, for a certain width, a FinFET device delivers higher ON current than the CMOS counterpart. This is due to the improved gate control mechanism in FinFET devices, which mitigates short channel effects and hence enhances the ON/OFF current ratio. On the other hand, an STT-MRAM cell requires a specific write current in order to change its internal state, which is the main requirement of the cell (the access transistor width is the effect of this requirement). Hence, area comparison of FinFET- vs. MOS-accessed STT-MRAM cells should be done under the same write current, as presented next.

**Write Current.** We measure the write current of FinFET- and MOS-accessed STT-MRAM cells using 32nm Predictive

Technology Model (PTM) [19] in HSpice for various transistor widths. For this purpose, we considered the worst-case scenario for writing into the cell. That is, the SL and WL are connected to  $V_{dd}$ , BL to  $GND$ , and the MTJ resistance is high [4]. Results are plotted in Figure 5(a). Under the specified range of transistor widths, FinFET-accessed cell delivers on average 28% higher current than the MOS-accessed cell. Additionally, while the maximum write current delivered by the MOS-accessed cell is  $50\mu A$ , FinFET-accessed counterpart can generate a maximum write current of  $62\mu A$  (24% higher).

**Process Variation.** We also considered the effect of process variations on the FinFET and CMOS transistors, as well as on the MTJ. Both CMOS and FinFET devices have non-negligible process variations, i.e., *random dopant fluctuation* (RDF) and *line edge roughness* (LER) for CMOS devices, and *work function variation* and LER for FinFETs. However, FinFETs experience less significant process variation than CMOS devices because the channel can be undoped [9].

Various types of process variations will induce variation in the threshold voltage  $V_{th}$ , and thereby the ON current and switching speed. We assume 2.5% variation in the  $V_{th}$  of FinFETs, 5% variation in  $V_{th}$  of CMOS devices, and 10% variation in the MTJ resistance values, which are common values in the previous work [8], [16]. Considering process variations, the worst-case ON current for both FinFET- and MOS-accessed cells are shown in Figure 5(a). However, for the desired range of ON currents, the width of the FinFET access transistor does not increase significantly, and still a relatively small FinFET device is needed.

**Cell Area.** More compact layout and higher ON current are two features of FinFET devices that lead to significant area reduction in STT-MRAM cells. This is shown in Figure 5(b), which plots the cell area of FinFET- vs. MOS-accessed STT-MRAMs, with and without considering process variations, for possible write current requirements from  $32\mu A$  up to  $50\mu A$ . A FinFET-accessed cell with the minimum width in 32nm PTM produces  $44.8\mu A$ , so for currents between  $32\mu A$  and  $44.8\mu A$  a minimum width FinFET is adopted. As can be seen, the cell area of the MOS-accessed STT-MRAM rises dramatically as current increases above  $40\mu A$ , whereas for FinFET-accessed STT-MRAM cell, even after considering the effect of process variation, the area remains unchanged.

To be more specific, we present cell area for various STT-MRAM write pulse widths in Table II. Each write pulse width requires a specific write current to ensure a successful write operation. For each write pulse width, we also report the access transistor width, the cell area and its aspect ratio (defined as cell height divided by cell width) along with the



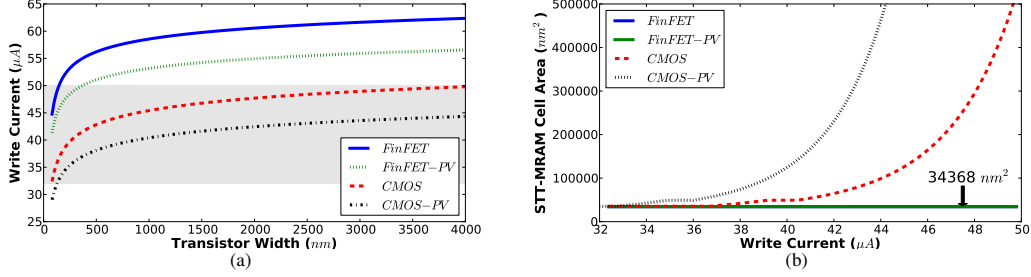


Fig. 5. (a) Write current of FinFET- vs. MOS-accessed STT-MRAM cells for different transistor widths in 32 nm PTM. Shaded region highlights the range of currents delivered by CMOS. (b) STT-MRAM cell area as a function of write current. FinFET-accessed STT-MRAM shows a steady cell area for the specified range of write currents. On both figures, *PV* denotes process variation.

leakage current for FinFET- and MOS-accessed STT-MRAM cells. Leakage currents are calculated using 32nm PTM by connecting gate to  $GND$ , and drain to  $V_{dd}$ . For write pulse widths between  $10ns$  and  $3ns$ , area of the MOS-accessed STT-MRAM cell rises from  $34.5F^2$  ( $F$  for the feature size) to  $220.1F^2$  ( $6\times$  increase). However, for the same range of write pulse widths, FinFET-accessed STT-MRAM cell has a steady cell area of  $33.6F^2$ , which improves CMOS results by 30% on average, and up to 85% (for write pulse width of  $3ns$ ).

Considering worst-case of process variations, the maximum write currents of FinFET- and MOS-accessed cells drop to  $44.3\mu A$  and  $56.6\mu A$ , respectively. Accordingly, for write pulse widths between  $10ns$  and  $4ns$ , area of the MOS-accessed STT-MRAM cell changes from  $49.0F^2$  to  $336.5F^2$ , whereas the FinFET-accessed STT-MRAM still has the steady cell area of  $33.6F^2$ .

Results of Table II are obtained for  $H_{FIN}/T_{SI} = 2$ , as  $H_{FIN}/T_{SI} = 4$  produces the same results for this range of currents. The effect of  $H_{FIN}/T_{SI} = 4$  can be seen in higher currents. For instance, considering  $60\mu A$ , the cell area of  $H_{FIN}/T_{SI} = 2$  is  $102.3F^2$ , whereas in the case of  $H_{FIN}/T_{SI} = 4$ , the cell area reduces to  $61.4F^2$  (40% improvement).

**Reliability Metrics.** Write-ability and read stability of an STT-MRAM cell are measured by *write margin*,  $WM = (I_W - I_C)/I_C$ , and *read disturb margin*,  $RDM = (I_C - I_R)/I_C$ , respectively [6]. For STT-MRAMs, *cell tunneling magneto-resistance*,  $CTMR$ , is also defined which measures the distinguish-ability of low and high states of the MTJ, and is obtained by  $CTMR = (R_{AP} - R_P)/(R_P + R_{AC})$ , where  $R_{AC}$  is the resistance of the access transistor [6].

If FinFET and CMOS access transistors deliver the same write current then they will have the same reliability metrics. However, FinFET transistors may be oversized for two reasons. (1) For write currents smaller than  $44.8\mu A$ , an oversized FinFET is adopted, as  $44.8\mu A$  is the minimum current delivered by the FinFET cell in 32nm PTM. (2) Width quantization property may result in an oversized FinFET device. In these cases, the FinFET delivers a write current higher than what was required. This higher write current improves  $WM$  and  $CTMR$  metrics, but degrades  $RDM$ , which in turn can be improved by appropriate choice of read voltage. Hence, the FinFET-accessed STT-MRAM cell needs a smaller read voltage to match the  $RDM$  of the CMOS counterpart.

### B. Architecture-level Comparison

We integrated the FinFET-accessed STT-MRAM cell into various subarray designs to evaluate its impact on the area, leakage power, as well as read and write latencies and energy

consumptions at a higher level of the memory system. A memory subarray contains  $r \times c$  memory cells, where  $r$  and  $c$  denote the number of rows and columns of the subarray, respectively. The subarray also includes peripheral circuitry such as the row address decoder, bit-line equalizer, sense amplifier, and column multiplexers. For this purpose, we added the FinFET support to NVSim [4], which is a circuit-level model for emerging nonvolatile memories, as follows.

FinFET-specific process parameters are extracted from MASTAR tool [5] for 32nm double-gate technology. Capacitance calculations are borrowed from the BSIM-CMG, which includes compact models for multi-gate devices [14]. Moreover, ON and OFF currents of FinFET and CMOS transistors are calculated in HSpice using 32nm PTM [19] for temperatures from  $300K$  to  $400K$ . Furthermore, STT-MRAM cell configurations are updated based on the results of Table II. For FinFET devices, spacer-defined technology with  $H_{FIN}/T_{SI} = 2$  is assumed.

Table III compares FinFET and CMOS STT-MRAM designs for various subarray sizes considering different write pulse widths ( $\tau_w$ ), and a temperature of  $320K$ . Based on this table, FinFET design on average improves area, read and write latencies, read and write energy consumptions, and leakage power of the CMOS counterpart by 116%, 75%, 27%, 14%, 22%, and 43%, respectively. Specifically, for  $\tau_w = 3ns$ , on average  $3.5\times$  area reduction is achieved.

Moreover, for  $4ns \leq \tau_w \leq 5ns$ , the same FinFET cell is used, which results in same results (except for those related to write operation) in both cases. However, access transistor width increases in  $\tau_w = 3ns$ , and hence a slightly larger row decoder is needed. The interesting case is  $\tau_w = 2ns$ , where compared to  $\tau_w = 5ns$ , on average  $1.1\times$  increase in the area leads to  $2.4\times$  and  $1.6\times$  decrease in write latency and energy consumption (major issues associated with STT-MRAMs), respectively.

According to our simulations, row address decoder has the largest share on the leakage power. On the other hand, sense amplifiers and bit-lines are the main contributors to the dynamic power (due to high activity factors). As a result, for memory subarrays with the same number of cells (e.g. 64KB), increasing the number of rows (from 128 to 512) increases the leakage power ( $2.0\times$  on average), but decreases read and write energy consumptions ( $3.7\times$  on average).

Hence, the FinFET design not only improves the results of the CMOS counterpart, but also shows a better scalability in write operation where decrease in  $\tau_w$  (for  $2ns \leq \tau_w \leq 5ns$ ) results in write latency and energy consumption reductions as well. Detailed comparisons provided in this section prove the effectiveness of FinFET-accessed STT-MRAM cell as a

TABLE II. FinFET- vs. MOS-ACCESSED STT-MRAM CELL COMPARISON. FinFET RESULTS ARE OBTAINED FOR SPACER-DEFINED PROCESS WITH  $H_{FIN}/T_{SI} = 2$ . CMOS RESULTS ARE BASED ON [6]. MOREOVER,  $F$  IS THE FEATURE SIZE.

	Write Pulse Width (ns)	Write Current ( $\mu A$ ) [4] [21]	Access Transistor Width ( $F$ )		Cell Area ( $F^2$ )		Cell Aspect Ratio ( $height/width$ )		Leakage Current ( $\mu A$ )	
			FinFET	CMOS	FinFET	CMOS	FinFET	CMOS	FinFET	CMOS
10	36.11	2.5	4.1	33.6	34.5	0.93	0.96	0.023	0.041	
9	36.64	2.5	4.5	33.6	34.5	0.93	0.96	0.023	0.045	
8	37.29	2.5	5.0	33.6	37.3	0.93	0.89	0.023	0.050	
7	38.13	2.5	5.8	33.6	41.8	0.93	0.79	0.023	0.058	
6	39.25	2.5	7.1	33.6	48.0	0.93	1.33	0.023	0.073	
5	40.82	2.5	9.8	33.6	51.2	0.93	1.25	0.023	0.101	
4	43.18	2.5	17.1	33.6	80.3	0.93	0.80	0.023	0.178	
3	47.11	3.2	52.0	33.6	220.1	0.93	0.29	0.030	0.543	
2	54.96	11.4	N/A §	35.0	N/A	0.90	N/A	0.106	N/A	
1	78.51	N/A †	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

§ Maximum write current of MOS-accessed STT-MRAM cell in 32 nm PTM is 49.8  $\mu A$ .

† Maximum write current of FinFET-accessed STT-MRAM cell in 32 nm PTM is 62.4  $\mu A$ .

TABLE III. FinFET- vs. MOS-ACCESSED STT-MRAM COMPARISON FOR VARIOUS MEMORY SUBARRAY DESIGNS.  $r$  AND  $c$  DENOTE THE NUMBER OF ROWS AND COLUMNS IN THE SUBARRAY, RESPECTIVELY. STT-MRAM WRITE PULSE WIDTH (ns) IS SPECIFIED BY  $\tau_w$ .

r	c	$\tau_w$	Area ( $\mu m^2$ )		Latency (ns)				Energy (pJ)				Leakage Power (mW)	
			FinFET	CMOS	Read		Write		Read		Write		FinFET	CMOS
					FinFET	CMOS	FinFET	CMOS	FinFET	CMOS	FinFET	CMOS		
64	64	5	751	849	1.13	1.15	5.04	5.06	0.80	0.84	1.08	1.12	0.18	0.25
64	64	4	751	957	1.13	1.16	4.04	4.07	0.80	0.87	0.92	1.00	0.18	0.35
64	64	3	794	1605	1.13	1.26	3.04	3.17	0.80	1.01	0.77	0.99	0.20	0.51
64	64	2	<b>894</b>	N/A	1.14	N/A	<b>2.05</b>	N/A	0.82	N/A	<b>0.64</b>	N/A	0.33	N/A
128	128	5	1676	2044	1.14	1.17	5.06	5.08	1.59	1.67	2.18	2.29	0.50	0.71
128	128	4	1676	2453	1.14	1.21	4.06	4.12	1.59	1.72	1.86	2.05	0.50	0.72
128	128	3	1705	4963	1.14	1.56	3.06	3.46	1.59	1.99	1.57	2.09	0.54	1.00
128	128	2	<b>1885</b>	N/A	1.16	N/A	<b>2.07</b>	N/A	1.63	N/A	<b>1.33</b>	N/A	0.86	N/A
128	512	5	5214	6524	1.22	1.44	5.14	5.35	6.29	6.56	8.65	9.08	0.99	1.02
128	512	4	5214	8389	1.22	1.89	4.14	4.80	6.29	6.75	7.41	8.09	0.99	1.10
128	512	3	5214	17954	1.24	6.94	3.15	8.85	6.30	7.73	6.21	8.19	0.99	1.03
128	512	2	<b>5591</b>	N/A	1.40	N/A	<b>2.31</b>	N/A	6.42	N/A	<b>5.23</b>	N/A	1.24	N/A
256	256	5	4416	5962	1.17	1.25	5.09	5.14	3.17	3.32	4.48	4.82	1.46	1.80
256	256	4	4416	7809	1.17	1.38	4.09	4.26	3.17	3.43	3.86	4.38	1.46	1.83
256	256	3	4459	17218	1.17	2.69	3.10	4.55	3.17	3.98	3.26	4.67	1.60	1.93
256	256	2	<b>5031</b>	N/A	1.23	N/A	<b>2.13</b>	N/A	3.25	N/A	<b>2.85</b>	N/A	2.19	N/A
512	128	5	4224	5780	1.17	1.26	5.12	5.23	1.60	1.70	2.38	2.68	1.77	2.64
512	128	4	4224	7654	1.17	1.31	4.12	4.27	1.60	1.78	2.07	2.52	1.77	2.68
512	128	3	4296	17659	1.18	1.72	3.13	3.49	1.60	2.13	1.78	2.92	1.94	3.71
512	128	2	<b>4806</b>	N/A	1.22	N/A	<b>2.17</b>	N/A	1.66	N/A	<b>1.66</b>	N/A	3.21	N/A
512	512	5	13138	18449	1.26	1.53	5.19	5.35	6.32	6.70	9.48	10.65	3.30	3.57
512	512	4	13138	26173	1.26	1.99	4.19	4.80	6.32	6.98	8.23	9.97	3.30	3.60
512	512	3	13138	63884	1.27	7.11	3.20	8.85	6.34	8.28	7.07	11.46	3.31	3.80
512	512	2	<b>14255</b>	N/A	1.45	N/A	<b>2.31</b>	N/A	6.53	N/A	<b>6.55</b>	N/A	4.33	N/A

suitable replacement for the MOS-accessed counterpart.

## V. CONCLUSIONS

We proposed an STT-MRAM cell design that utilizes a FinFET access transistor. We compared the proposed cell design with conventional MOS-accessed cells in terms of area, reliability metrics, leakage power, as well as read and write latencies and energy consumptions. Our comparison results show that STT-MRAM designs significantly benefit from FinFET-accessed cells. In particular, in the range of write current requirements, the area of the FinFET-accessed STT-MRAM design varies slightly, whereas write latency and energy consumption which are the main issues associated with STT-MRAMs reduce significantly.

## ACKNOWLEDGMENT

This research is supported by grants from the PERFECT program of the Defense Advanced Research Projects Agency and the Software and Hardware Foundations of the National Science Foundation.

## REFERENCES

- [1] M. Alioto. Comparative Evaluation of Layout Density in 3T, 4T, and MT FinFET Standard Cells. *IEEE Trans. on VLSI Systems*, 19(5):751–762, 2011.
- [2] Y.-K. Choi, T.-J. King, and C. Hu. Nanoscale CMOS Spacer FinFET for the Terabit Era. *IEEE Electron Device Letters*, 23(1):25–27, 2002.
- [3] X. Dong, X. Wu, G. Sun, Y. Xie, H. Li, and Y. Chen. Circuit and Microarchitecture Evaluation of 3D Stacking Magnetic RAM (MRAM) as a Universal Memory Replacement. In *45th DAC*, pages 554–559, 2008.
- [4] X. Dong, C. Xu, Y. Xie, and N. Jouppi. NVSim: A Circuit-level Performance, Energy, and Area Model for Emerging Nonvolatile Memory. *IEEE Trans. on CAD*, 31(7):994–1007, 2012.
- [5] International Technology Roadmap for Semiconductors. The Model for Assessment of CMOS Technologies and Roadmaps (MASTAR). Available: <http://www.itrs.net/models.html>.
- [6] S. Gupta, S. P. Park, N. Mojmumder, and K. Roy. Layout-aware Optimization of STT MRAMs. In *DATE'12*, pages 1455–1458, 2012.
- [7] C. Lin et al. 45nm Low Power CMOS Logic Compatible Embedded STT MRAM utilizing a Reverse-Connection 1T/1MTJ Cell. In *IEEE International Electron Devices Meeting (IEDM)*, pages 1–4, 2009.
- [8] B. Liu, M. Ashouei, J. Huisken, and J. de Gyvez. Standard Cell Sizing for Subthreshold Operation. In *49th DAC*, pages 962–967, 2012.
- [9] T. Matsukawa et al. Comprehensive Analysis of Variability Sources of FinFET Characteristics. In *Symposium on VLSI Technology*, pages 118–119, 2009.
- [10] E. Nowak et al. Turning Silicon on Its Edge. *IEEE Circuits and Devices Magazine*, 20(1):20–31, 2004.
- [11] S. P. Park, S. Gupta, N. Mojmumder, A. Raghunathan, and K. Roy. Future Cache Design using STT MRAMs for Improved Energy Efficiency: Devices, Circuits and Architecture. In *49th DAC*, pages 492–497, 2012.
- [12] T. Sairam, W. Zhao, and Y. Cao. Optimizing FinFET Technology for High-Speed and Low-Power Design. In *17th GLSVLSI*, pages 73–77, 2007.
- [13] T. Schloesser et al. 6F<sup>2</sup> Buried Wordline DRAM Cell for 40nm and Beyond. In *IEEE International Electron Devices Meeting (IEDM)*, pages 1–4, 2008.
- [14] V. Sriramkumar et al. BSIM-CMG 107.0.0 Technical Manual. Available: [http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG\\_LR](http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG_LR)
- [15] G. Sun, X. Dong, Y. Xie, J. Li, and Y. Chen. A Novel Architecture of the 3D Stacked MRAM L2 Cache for CMPs. In *15th HPCA*, pages 239–249, 2009.
- [16] Z. Sun, H. Li, Y. Chen, and X. Wang. Variation Tolerant Sensing Scheme of Spin-transfer Torque Memory for Yield Improvement. In *ICCAD'10*, 2010.
- [17] S. Tang et al. FinFET - A Quasi-Planar Double-Gate MOSFET. In *IEEE International Solid-State Circuits Conference (ISSCC)*, pages 118–119, 2001.
- [18] W. Xu, H. Sun, X. Wang, Y. Chen, and T. Zhang. Design of Last-level On-Chip Cache using Spin-Torque Transfer RAM (STT RAM). *IEEE Trans. on VLSI Systems*, 19(3):483–493, 2011.
- [19] W. Zhao and Y. Cao. New Generation of Predictive Technology Model for Sub-45nm Early Design Exploration. *IEEE Trans. on Electron Devices*, 53(11), 2006.
- [20] W. Zhao et al. High Density Spin-Transfer Torque (STT)-MRAM based on Cross-Point Architecture. In *4th IEEE International Memory Workshop (IMW)*, 2012.
- [21] Personal communication with NVSim authors.