Abstract—The spin-transfer torque random access memory (STT-RAM) has been widely investigated as a promising candidate to replace the static random access memory (SRAM) as on-chip cache memories. However, the existing STT-RAM cell designs can be used for only single-port accesses, which limits the memory access bandwidth and constrains the system performance. In this work, we propose the design solutions to provide dual-port accesses for STT-RAM. The area increment by introducing an additional port is reduced by leveraging the shared source-line structure. Detailed analysis on the performance/reliability degradation caused by dual-port accesses and the corresponding design optimization are performed. We propose two types of dual-port STT-RAM cell structures having 2 read/write ports (2RW) or 1-read/1-write port (1R/1W), respectively. Comparison shows that a 2RW STT-RAM cell consumes only 42% of area of a dual-port SRAM. The 1R/1W design further reduces 7.7% of cell area under same performance target.

I. INTRODUCTION

The continuously increasing demand on system performance and functionality in recent years has greatly stimulated the development of Chip-Multiprocessor (CMP) and System-on-Chip (SoC). Consequently, the large instruction and data exchange among different memory hierarchies makes the memory accesses more and more critical. Often a memory array receives multiple requests from one or many cores at the same time. The single-port memory which grants access to one request and stalls all the others can lead to significant performance degradation. Therefore, the dual-port or multi-port memory to reduce access conflicts and provide high memory bandwidth becomes a popular approach [1][2][3].

The traditional static random access memory (SRAM) as on-chip memory is facing severe challenges at the scaled technology node such as large cell size, higher leakage power and vulnerability to soft-errors [4]. On the other side, spin-transfer torque random access memory (STT-RAM) has demonstrated great potential to replace SRAM in near future and attracted much attention from both academia and industry world [5][6]. By storing the data as the relative magnetic direction of magnetic tunnel junction (MTJ), STT-RAM provides high density, fast access speed, zero standby power, as well as hardness to radiation-injected soft-errors.

However, all the previous STT-RAM designs can support only single-port access [7][8]. For example, the popular one-transistor-one-MTJ cell structure contains only one set of word-line (WL), bit-line (BL), and source-line (SL), which makes dual-port access impossible. Considering the fact that writing to a STT-RAM cell takes longer time than programming a SRAM cell, the stall of the pending accesses of STT-RAM will become even more severe, especially when the port is occupied by write operations. Therefore, the dual-port or multi-port STT-RAM cell design is necessary to enhance the system performance.

In dual-port SRAM designs [1][9], the additional port access is implemented by adding two extra access transistors and one set of WL/BL to the six-transistor cell design. However, as we shall show in Section III, the same design method cannot be applied to STT-RAM design for the extremely large area overhead. By leveraging the shared source-line array structure [10][11], we propose a STT-RAM design solution that supports dual-port accesses by paying a small cell area. In our design, each STT-RAM cell has two BLs and a memory array shares a single grounded SL. To the best knowledge of authors, it is the first STT-RAM design that enables the dual-port accesses.

To meet the different access requirements of various applications, two types of designs are presented. In a 2RW STT-RAM cell, both data access ports can support read and write operations. In contrast, 1R/1W STT-RAM has one read-only port and one write-only port. Separating the read and write accesses reduces the size requirement of access transistor, therefore the even smaller cell area can be achieved. Furthermore, we analyze the reliability of the proposed structures and present the design and layout optimization techniques for density improvement. Our results show that the area of a 2RW STT-RAM cell is only 42% of a dual-port SRAM’s area. Compared to a single-port STT-RAM cell, the area overhead introduced by the extra port access is 39%. The 1R1W design with the constrained access flexibility can further save 7.7% of cell area and reduce the probability of read-disturbance.

II. STT-RAM BASICS

The basic storage element in STT-RAM is magnetic tunneling junction (MTJ). Conceptually, an MTJ contains three layers as shown in Figure 1(a): two ferromagnetic layers are respectively named as reference layer and free layer, which are separated by an oxide barrier, e.g. MgO. The magnetization direction of the reference layer is fixed, but the magnetization direction of the free layer can be switched through a spin polarized current [12]. For example, a large current injected from the free layer to the reference layer can switch the magnetization direction of the free layer to be parallel to that
of the reference layer, and vice versa. When the magnetization directions of the two ferromagnetic layers are parallel (P) or anti-parallel (AP), the MTJ demonstrates a low- or high-resistance state, representing logic ‘0’ or ‘1’, respectively.

Figure 1(b) illustrates the most popular STT-RAM cell structure consisting of one NMOS transistor and one MTJ (1T-1J) [12][7]. The NMOS transistor, named as the access transistor, connects to the MTJ’s reference layer and controls the accessibility of the MTJ. Since there is only one set of WL, BL and SL, this cell structure can only be used for single-port memory design. The MTJ pillar has a very small area so the NMOS transistor determines the area of a STT-RAM cell. In other words, a small transistor is expected for high density. However, the MTJ switching performance strongly relies on the switching current [13]. Reducing transistor size reduces switching current through MTJ and hence degrades the write performance.

In this work, we use 65nm CMOS technology [14] with a 65nm × 130nm in-plane MTJ model calibrated against the experimental data [8]. The switching behavior of the MTJ is modeled based on the Landau-Lifshitz-Gilbert equation [13]. The detailed parameters are listed in Table I.

III. DUAL-PORT STT-RAM DESIGN CHALLENGES

Figure 2(a) illustrates a typical SRAM design with two sets of read/write ports [1][9]. Compared to a single-port SRAM cell with six transistors, two more transistors (M1 and M2) associated with the wordline control (WLB) and the data access connections (BLB and WLB) of the second port, are inserted.

By following the same design concept, Figure 2(b) shows a dual-port STT-RAM cell with four transistors and one MTJ (4T-1J). Here, a duplicate pair of BL and SL provide the access through the second port. Compared to the single-port STT-RAM cell in Figure 1(b), three additional transistors (M1, M3, and M4) are needed for access control. Note that in a STT-RAM array, the BL and SL are usually shared by entire column. For single-port cells, only one memory cell within a column can be activated at a time. Therefore, one transistor at SL terminal is sufficient to control the accessibility to one cell per column. In contrast, a dual-port array may simultaneously access two cells within one column through Port-A and Port-B, respectively, as illustrated in Figure 3(a). Determined by the operation type and data pattern, the two concurrent accesses could have the different BL voltages. Thus, M1 and M3 are necessary to isolate BLA and BLB from each other.

Due to degraded biasing condition, the 4T-1J dual-port STT-RAM cell is functionally correct by paying significant area overhead compared to the 1T-1J single-port design. As shown in Figure 1(b), a conventional 1T-1J STT-RAM encounters VGS degradation induced by the voltage drop on MTJ only in write-1 operations, which constrains the switching current through MTJ. The 4T-1J dual-port STT-RAM has a symmetric cell structure: along an access path, e.g., from BLA to SLA, two transistors M1 and M2 are turned on and connected side by side of the MTJ. No matter in write-1 or write-0 operations, one of them suffers from VGS degradation, as shown in Figure 3(b). In other words, the biasing condition of the access transistors in the 4T-1J cell is much worse than that of a 1T-1J design. We have to enlarge all the access transistors to provide sufficient MTJ switching current in write operations.

Figure 4(a) shows the relation between the write-1 current and the size of the access transistors in the 4T-1J design, assuming all the four transistors are of the same size. Here, the write-1 operation dominates the transistor size selection because of the asymmetric P → AP and AP → P switching currents of the MTJ device used in this work (refer Table I). To obtain the write time of 10ns, the access transistors' width is approximately 1400nm. Integrating four such large transistors into one memory cell leads to a cell area of 575F², which

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>SIMULATION PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>65nm</td>
</tr>
<tr>
<td>VDD</td>
<td>1.2V</td>
</tr>
<tr>
<td>MTJ geometry</td>
<td>65nm × 130nm</td>
</tr>
<tr>
<td>Rf/RAP</td>
<td>1.88/3.77kΩ</td>
</tr>
<tr>
<td>AP→P Switching Current</td>
<td>112μA</td>
</tr>
<tr>
<td>P→AP Switching Current</td>
<td>142μA</td>
</tr>
</tbody>
</table>

1 The minimum channel length is 60nm.
2 At 10ns switching time.
is even bigger than that of the dual-port SRAM design (e.g., 233F² reported in [9]). It is not acceptable to adopt such a large STT-RAM design for on-chip applications.

IV. STT-RAM DESIGN WITH TWO READ/WRITE PORTS

A. Design Concept

Previously, the shared SL for single-port STT-RAM array has been proposed by Zhao et al. to increase array density [10]. It also has been used to balance the write-0 and write-1 performance [11]. The basic design concept is that all the cells on the same row share the same SL, then all the SLs are connected together and grounded (GND/0V).

In this work, we propose to reduce the cell area of dual-port STT-RAM design by utilizing the shared SL structure. Figure 5 depicts the STT-RAM design with two read and write ports (2RW). Please note in [10], the grounded (0V) SL is connected to the transistor, but in the proposed 2RW design, the grounded SL is connected to the MTJ in order to support Dual-Port. The write-1 operation requires a switching current from SL (GND) to BL, so a negative voltage (V_{BLN}) need be applied to BL. Such V_{BLN} can be generated using level converter[15].

The 2RW cell design can significantly reduce the cell area compared to 4T-1J. First, since the SL is always connected to GND, isolating SLs of different memory cells is no longer necessary. The transistors used for SL access control in STT-RAM cell can be removed. Only two transistors M1 and M2 remain to enable/disable the access to Port-A and Port-B, respectively. Thus, the number of transistors reduces to half of the 4T-1J dual-port design. Second, the width of access transistors can greatly decrease because only one transistor exists along the current path between BL and SL. Figure 4(b) shows the relation between the write-1 current and the size of the access transistors in the proposed 2RW STT-RAM design.

The required transistor width to achieve the 10ns write time is 585nm, which is only ~42% of the access transistor size of the 4T-1J STT-RAM cell. The area of a proposed 2RW cell is approximately 21% of the 4T-1J design.

B. Reliability Analysis

The voltage of the shared SL (V_S) in the single-port STT-RAM array may not be ideal 0V due to the existence of the parasitic resistance (R_S) [10]. Figure 6 illustrates the scenario. When turning on WL and applying a certain voltage to BL, the variation on V_S exists and induces degradation on both read and write performance. For example, if V_S is higher than ideal 0V, the actual voltage drop across the BL and SL reduces. Consequently, the write-1 current becomes lower than the projected value obtained under the ideal condition. In read operations, a higher/lower V_S can decrease/increase the read-0/read-1 current. The reduced difference between read-0 and read-1 currents could result in more read errors. V_S variation can also leads to higher possibility of read disturbance, i.e., unwanted ‘0’→‘1’ switch when reading a cell which stored ‘0’ [7]. A negative V_S will increase the read-0 current (I_{R0}) and bring it closer to the P→AP switching current (I_{W1}).

For the proposed 2RW STT-RAM design, the impact of the V_S variation becomes even more severe. First, the V_S variation increases as the number of cells being accessed grows. When both ports access the cells on the same row as illustrated in Figure 6, the number of cells doubles compared to that of single-port STT-RAM array. So a larger V_S variation is expected. Moreover, we notice that in the single-port STT-RAM, the read operations have a lower V_S variation than the write operation. This is because the write requires a bigger voltage amplitude applied to BL (|V_B|) and the only port can perform either write or read access. However, for the 2RW STT-RAM design, it is possible that the read and write are conducted simultaneously through the two sets of ports. The interaction in between degrades the V_S variation of read operations. Third, the value of V_S is also affected by the MTJ resistance states of the cells being accessed. When the MTJ is at high resistance state, the V_S is more reluctant to be disturbed by V_B.

Here, we use n_{A/B,R/W1/W0,H/L} to represent the number of the cells under certain access pattern. The subscript A/B indicates Port-A or Port-B access. R/W1/W0 describes the operation modes, including read, write-1, or write-0. H/L represents the high or low resistance states of MTJ. For example, n_{B,W1,L} is the number of the cells that are with
low MTJ resistance and conducting write-1 operations through Port-B.

Without loss of generality, we studied the current through a 2RW STT-RAM cell when it is accessed through Port-A. Table II summarizes its worst-case current and the corresponding access patterns in read and write operations. In the experiment, we assume a SL is shared by 32 columns, and each port accesses only 8 cells by using column selection, which is very common to support set-associative cache. The R of such setup is set to 27.5Ω according to [10]. The worst-case scenario happens when all the 16 cells being accessed fall on the same row. For comparison purpose, the currents under the ideal condition when VS is exactly 0V are also presented.

The simulation results show that write-1 and write-0 currents drop from 142μA and 112μA projected under the ideal condition to 130μA and 96μA in the worst scenario, respectively. The write current degradation means the design cannot meet the target of a 10ns switching time. The situation for read operation is even worse: ideally the read-0 current is 18.6μA more than the read-1 current. However, in the worst-case combination, the read-0 current turns to be less than that of the read-1 current, which can result in read decision errors. One possible way to solve this is to increase the read voltage (Section IV-C). The margin between read-0 current and P→AP switching current (IW1−I10) reduces from 94.4μA to 83.7μA under worst-case, which indicate higher possibility of read disturbance. Please note the “worst-case” for the IW1−I10 occurs when I10 reach its highest value.

In Table II, we also show the results when disabling Port-B, which is indeed equivalent to single-port access. The results show that the second set of access ports results in 8μA degradation on both write-1 and write-0 currents in the worst-case condition. The difference between read-0 and read-1 currents dramatically drops 17.3μA due to the interaction between read and write in dual-port accesses.

C. The Cell Configuration and The Operating Setup

Previously we demonstrate that the variation of VS is exaggerated by the dual-port access, which must be considered when determining the access transistor size in cell design and setting up the operating conditions, i.e., the read and write voltages.

For the given MTJ device in Table I, the write-1 operation is critical in transistor size selection. To compensate the current degradation under the worst-case access pattern, we have further increase the transistor width. The simulation result in Figure 7(a) shows that to maintain the write-1 current at 142μA in the worst-case condition, the access transistor grows to 715nm in width.

The negative voltage (VBLN) for write-0 operations also needs to be adjusted to compensate the impact of VS variation. With the access transistor width of 715nm, Figure 7(b) shows that |VBLN| should increase to 0.58V to obtain the 112μA write-0 current in the worst-case condition. Figure 7(c) demonstrates the relation between the read voltage (VRead) and the current difference in read-1 and read-0 operations. The negative value of current difference indicates that the read-0 produces a smaller current than the read-1, which will result in inevitable read decision error. Increasing VRead can significantly improves the read current difference. On the other hand, the higher read-0 current can increase the chance of read disturbance.

D. Layout Design

Figure 8(a) shows the layout of the proposed 2RW STT-RAM cell, where λ is half of the feature size (F). Based on the analysis in the previous section, the access transistor width is 715nm (11F). The two access transistors in one cell can share the diffusion area, which is connected to the MTJ.
V. STT-RAM Design with 1-Read/1-Write Port

A. Design Concept

Some dual-port SRAM designs restrict the port functionality [2][3]: one support read operations only and the other is for writes only. Such designs with 1-read/1-write port (1R1W) can alleviate the degradation of static noise margin, compared to 2RW design.

Similarly, the 1R1W design concept can be applied to the dual-port STT-RAM to reduce the impact of $V_S$ variation. Figure 9 illustrates the access pattern when constraining the port functionality to 1R1W. Not like writes through two port aggravate the $V_S$ variation in 2RW STT-RAM, a write in 1RW design can be accompanied to only a read through the other port. Since the read voltage is much lower than the write voltage, $V_S$ reduces compared to the 2RW case. Moreover, the positive $V_{READ}$ tends to bring $V_S$ to the positive direction, which actually improves the write-0 current strength. Therefore, the worst-case access patterns for the write operations in 1R1W STT-RAM are redefined as shown in Table III. The patterns for read operations remain the same as the 2RW design in Table II.

B. Transistor Sizing and Operating Voltage

Benefiting from the improved worst-case write current, the 1R1W design can shrink the transistor sizes to achieve the same write performance as the 2RW design. For example, if assuming the two access transistors are of the same dimension and setting $V_{READ}$ to 0.24V, our simulation shows that the transistors can reduce to 670nm.

Moreover, if utilizing the different sizes to the read access transistor ($W_R$) and the write access transistor ($W_W$), the design could be further reduced. On one hand, the increased resistance induced by a smaller $W_R$ helps reduce the $V_S$ variation, which in turn alleviates the sizing requirement for $W_W$. On the other hand, the smaller read access transistor degrades the read current difference $I_{Rdiff}$, which could lead to more read errors. To maintain $I_{Rdiff}$ when decreasing transistor sizes, we can increase $V_{READ}$, which however exaggerates the $V_S$ variation.

Fig. 9. Access pattern of 1R1W.

For a given $W_R$, we proposed the following design flow to obtain the minimum $W_W$ and hence the most area-efficient configuration:

Step 0: Randomly choose a $V_{READ}$.

Step 1: With the given $W_R$ and $V_{READ}$, sweep $W_W$ till it meets the write-1 current target in the worst-case pattern.

Step 2: Find the $V_{BLN}$ to achieve the write-0 current target under the worst-case configuration, when $W_R$, $V_{READ}$ and $W_W$ are fixed.

Step 3: Get the $V_{READ}$ to achieve the $I_{Rdiff}$ target for the given $W_W$, $V_{BLN}$ and $W_R$.

Step 4: Repeat the iteration from Step 1 to Step 3 until $W_W$ and $V_{READ}$ converge to certain values.

Fig. 10 shows the minimum $W_W$ and the corresponding $V_{READ}$ under different $W_R$. Here, we set the targeted $I_{Rdiff}$ as 10μA and the write time as 10ns for both write-0 and write-1. The result shows that reducing $W_R$ from 660nm to 540nm helps relax the sizing requirement of $W_W$ due to the increased equivalent resistance of read access transistor. However, $W_W$ starts to increase when further decreasing $W_R$ because the higher $V_{READ}$ becomes the dominating factor. As the width of the cell layout is determined by $W_W$, the smallest 1R1W STT-RAM cell can be obtained when $W_W = 660nm$ and $W_R = 540nm$. The corresponding $V_{READ}$ and $V_{BLN}$ are 0.27V and −0.53V, respectively.

TABLE III

<table>
<thead>
<tr>
<th>Worst-Case Pattern</th>
<th>Write-1</th>
<th>Write-0</th>
</tr>
</thead>
</table>

* Assuming Port-A is write only and Port-B is read only.
C. Comparison of 2RW and 1R1W STT-RAM Designs

We compared the proposed 2RW and 1R1W STT-RAM designs by following the worst-case design methodology and the results are summarized in Table IV. Thanks to the smaller transistors, the cell area of a 1R1W STT-RAM cell is only 92.3% of that of the 2RW design. The amplitude of V_{BLN} is smaller too. The reduced transistors and \(|V_{BLN}|\) indicates that the 1R1W design has the less write current in the non-worst-case condition and hence consumes less write energy than the 2RW design. Interestingly, although the V_{READ} is higher for 1R1W, the worst-case difference between read-’0’ current and P \rightarrow AP switching current (I_{W1} – I_{R0}) is still improved, which indicates lower possibility of read disturbance. This is because smaller W_{W} and V_{BLN} reduce the V_{S} drift toward negative direction, which is the main reason for the excessive I_{R0}. In summary, 1R1W cell can achieve smaller area, less energy waste and smaller possibility of read disturbance, with the cost of restricted port functionality.

VI. THE RELATED WORKS

Dual-port SRAM design has been widely used in order to satisfy the increasing demand on memory bandwidth. The 2RW SRAM is a common style [1][9]. The major design challenge is the degradation of static noise margin caused by the additional WL during “common-row-different-column” access. One common solution is to isolate the read port and write port, i.e., 1R1W, present in [2][3].

Nearly all the previous works on STT-RAM focused on the single-port designs, such as the most popular 1T-1J STT-RAM [12][7]. The cell structure with two transistors (2T-1J) have been presented by Chung [8]. However, the main motivation was to enhance the writability and array density. The two transistors are controlled by the same WL and hence the design still has only one port.

TABLE IV

<table>
<thead>
<tr>
<th></th>
<th>2RW</th>
<th>1R1W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst-case write time</td>
<td>10ns</td>
<td></td>
</tr>
<tr>
<td>Worst-case (I_{R0})</td>
<td>10(\mu)A</td>
<td></td>
</tr>
<tr>
<td>Transistor Width</td>
<td>both 71.5nm</td>
<td>W_W = 660nm, W_R = 540nm</td>
</tr>
<tr>
<td>Cell Size</td>
<td>100F^2</td>
<td>92.3F^2</td>
</tr>
<tr>
<td>Area overhead over Single-Port STT</td>
<td>39%</td>
<td>28%</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

In this work, we firstly propose the dual-port STT-RAM design, which can provide higher data bandwidth. We propose to leverage the shared SL design to simplify the cell structure and reduce the memory cell area. Two types of the dual-port STT-RAM design, 2RW and 1R1W, are presented. Furthermore, the related design issues, including reliability, cell configuration, operating setup, and layout techniques, have been considered and discussed.

VIII. ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1116684, CNS-1116171, and CNS-1149654. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

REFERENCES