An Improved Read/Write Scheme for Anchorless NEMS-CMOS Non-volatile Memory

Anh Tuan Do¹, Karthik G. Jayaraman¹, Vincent Pott², Geng L. Chua², Pushpapraj Singh², Kiat S. Yeo and Tony T. Kim¹
¹VIRTUS, School of EEE, Nanyang Technological University, Singapore 639798
²Institute of Microelectronics, A*STAR, Singapore 117685

Abstract—We proposed a NEMS-based anchorless memory structure with two stable mechanical states (Up and Down) for retaining data even at high operating temperature (>200°C) [5]. Compared to the conventional anchored devices, the anchorless structure offers better scalability and can operate at lower supply voltage, which is more desirable for integration with CMOS processes. This work addresses several issues in the previous NEM memory array implementation such as shuttle oscillation due to electrostatic pendulum and non-polarity write operation. We propose a 128 × 128 NEM memory array consisting of NEM memory cells and CMOS read/write control circuits. Our proposed read/write scheme with the two-level gate control eliminates the non-polarity write issue and achieves 32% power and 14% delay improvements when compared to the previous control scheme.

I. INTRODUCTION

Commercial nonvolatile memory (NVM) devices typically fail at high temperature (HT). The market-dominant Flash technology demonstrates remarkable performances in terms of density and reliability but cannot permanently store data if the chip is exposed to HT (typically, T > 200°C) [1]. NVM devices have been investigated to overcome this but they either suffer from physical data loss [2, 3] or poor scalability [4]. To alleviate this issue, we proposed a nano-electro-mechanical (NEM) anchorless nonvolatile memory device (ANVM) [5]. In this device, only electrostatic actuation and adhesion forces control the switching of the cell and thus offer a better scalability. However, the performance evaluation (power consumption, delay, variations, etc) of this newly proposed device in an array implementation has not been adequately addressed.

This work proposes a new control scheme to address the drawbacks of the previous read/write scheme in [6] and at the same time to improve its power efficiency and variation tolerance. In the next section, a brief introduction on the ANVM and its prior read/write operation will be presented. Section III proposes a read/write control scheme with an array of 128 rows × 128 columns. We explain the proposed two-level gate control and the analysis results of the ANVM integrated with a commercial 180 nm CMOS process technology in section IV, followed by the conclusion.

II. NEMS-BASED ANCHORLESS MEMORY

A. Memory cell

Fig. 1 illustrates the cross section of the ANVM cell. It has three terminals (D: Drain, S: Source, G: Gate) and a shuttle that can be actuated between the G and the D-S terminals [5]. Depending on the relative voltage differences between the terminals, the ANVM cell has three distinctive operation modes, namely “standby”, “read” and “write”.

During standby, the shuttle is attached to either G (Down state – Fig. 1(a)) or D/S (Up state – Fig. 1(b)) by the Van der Waal adhesive force (Fad). At the same time, all three terminals should be maintained at the same potential (e.g. VG = VS = VD = 0) to prevent unwanted shuttle actuation. By engineering the smoothness of the shuttle and terminal surfaces, the adhesive force stabilizes the shuttle at the two states until the voltage difference between G and D/S exceeds a threshold voltage (VD).

When new data is written into the cell, the shuttle must be switched from G to S/D or vice versa. Assume that the shuttle is at the Down state as shown in Fig. 1(a). By keeping the D/S voltages at ground (VD = VS = 0) and raising VG, Charges (Q) will be accumulated at the top surface of the shuttle:

\[ Q = \varepsilon_0 A \frac{VG - VD}{2d_{gap}} \]  

Where \( \varepsilon_0 \) is the dielectric constant of vacuum, \( A \) is the surface area of the shuttle, \( V_{G-D/S} \) is the G-to-D/S voltage, \( 2d_{gap} \) is the distance between the top surfaces of the shuttle to the D/S terminals. This induces an electrostatic force for the shuttle:
If $F_e > F_{ad}$, the shuttle will be lifted and accelerated towards the D/S terminals. The required gate voltage to break the adhesive force ($F_{ad}$) is called the programming voltage ($V_P$) of the ANVM, which depends on several factors such as $d_{gap}$ and smoothness of the surfaces [5].

During read, a potential difference is developed across the D and S terminals ($V_{DS} > 0$) while the gate terminal is kept at ground. If the shuttle is at the Up state, current ($I_{DS} > 0$) will be detected between the D and S terminals (Fig 1(b)). Otherwise, no current flows between D and S. By sensing the $I_{DS}$, one can deduce the internal state of the ANVM. Table I summarizes the operating principle of the ANVM:

### Table I. Operating Principle of the Conventional ANVM

<table>
<thead>
<tr>
<th>Operation</th>
<th>Condition</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching down</td>
<td>$V_B = V_G = V_D = V_P$</td>
<td>Pull-out voltage, $V_e &gt; V_{ad}$</td>
</tr>
<tr>
<td>Switching up</td>
<td>$V_B = V_G = V_D = V_S = 0$</td>
<td>Switching to both drain and source</td>
</tr>
<tr>
<td>Read</td>
<td>$V_B = V_D, V_G = V_S = 0$</td>
<td>$V_E$: Read voltage ($V_{CE} &gt; V_E$), Down-state: $I_{DS} = 0$, Up-state: $I_{DS} &gt; 0$</td>
</tr>
<tr>
<td>Hold</td>
<td>$V_B = V_D, V_G = V_S = 0$</td>
<td>Non-volatility: adhesion increases with temperature</td>
</tr>
</tbody>
</table>

B. Issues in previous array implementation

The above operating principle of the ANVM presents a few flaws when implemented in an array structure. First, if the gate voltage is maintained at above $V_P$ too long, the shuttle will arrive at the intended terminal and return to the other terminal(s). This phenomenon is called "shuttle oscillation" due to electrostatic pendulum. Thus quantitative analysis of the write delay must be conducted to prevent the shuttle oscillation while maintaining good write operation. Second, the electrostatic force developed between the shuttle and G or D/S at the beginning of the write period is always attractive force, regardless of the polarity of the $V_{GDS}$. This indicates that the shuttle always toggles as long as $|V_{GDS}| > |V_e|$. For example, when the shuttle is at the "Down" state and $V_G = V_S = 0$, $V_D = V_P$, it will be switched up. This means that $V_{GDS} = V_P$ does not ensure the write polarity. Therefore, an improved read/write control is required to correctly program the data into the NEM memory.

III. PROPOSED READ/WRITE CONTROL SCHEME

Fig. 2 illustrates the architecture of the proposed ANVM. During standby, worldlines (WLs) and bitlines (BLs) are kept at ground to minimize leakage power. A read operation starts by triggering the READ EN signal. Coupled with the row address decoder, it will subsequently drive one selected WL to $V_{DD}$ to turn on the access transistors ($M_{A1}$ and $M_{A2}$) of the cells in the selected row. Simultaneously, READB signal goes low to turn on $P_1$ and turn off $N_1$ and $N_2$, leaving BLB floating. If the shuttle is at the "UP" state (i.e. attached to the D/S terminals), BLB is connected to BL through $M_{A1}$, $M_{A2}$ and the NEMS device. Otherwise, there is an open circuit between BL and BLB. Therefore, $P_1$ will pull up BLB conditionally, depending on the state of the accessed shuttle. Since there is a $V_{th}$ drop from $V_{DD}$ to BLB due to the NMOS access transistors, we use an output buffer at the bottom of each BLB to quickly obtain the full CMOS output voltage. Fig. 3 demonstrates the waveforms of several nodes during two read operations of our ANVM at 1.8 V supply voltage. It achieves a read delay of 2.5 ns at 27$^\circ$C.

The rising edge of WRITE EN signal indicates the start of the write operation (Fig. 2, 4). Due to the requirements of the Read-Pulse generator (RPG) and the Write control circuit (WCC), WRITE EN must be synchronized with the rising edge of CLK. RPG generates a "Read-pulse" to retrieve the data from the accessed cell (Fig. 4). After a small read delay, "OLD_DATA" will be available to WCC. If "OLD_DATA" and "NEW_DATA" are different, WCC produces a "Toggle" to enable the GATE signal. If not, nothing will happen. The "GATE" signal lasts until the end of the write cycle and is disabled at the falling edge of WRITE EN signal (Fig. 5). This write scheme ensures that cell data is kept untouched if the write data is the same as the existing one (see the first write cycle in Fig. 5). Statistically, it also saves 32% of write
Figure 4. Zoomed in waveforms of several nodes at the beginning of a write cycle. The cell is toggled after checking that old and new data are different.

Figure 5. Waveforms of the proposed write scheme in three consecutive write cycles. Cell content is flipped only in the second and the third cycle where new and old data are different.

Figure 6. Power and delay of the read operation versus temperature variations.

(a)

(b)

Figure 7. Delay of the write operation versus different gate voltages (top) and gate pulse widths (bottom).

IV. PERFORMANCE EVALUATION WITH TWO-LEVEL GATE CONTROL FOR WRITE

A. Read/Write delay

Fig. 6 shows the read power and delay of the proposed design using 1.8 V supply. When temperature changes from 0°C to 300°C, the read delay increases from 2.5 ns to 3.8 ns (52%). The memory can operate at a maximum read frequency of around 250 MHz. Its power consumption is only 86 µW at 300°C. Note that this work focuses more on the write operation thus both read power and delay can be further improved if advanced techniques such as latch-based sense amplifiers are used.

The write delay of the ANVM is much longer (around 550 ns at the nominal condition), mainly because of the flying time of the shuttle, governed by its mass and electrostatic force. Fig. 7(a) illustrates the simulated write delay over different gate voltages. Although increasing gate voltage can speed up this process, it also affects the reliability of the MOS devices and consumes more energy. Fig. 7(b) shows how the delay changes with different gate pulse width (V_G = 6V). Note that when the gate is shut down before the shuttle reaches its destination, it continues to fly (with no acceleration and thus takes longer time) thanks to its velocity. Note that gravity is insignificant in this context. Therefore, it is more desirable to maintain the gate signal to minimize the flying time. However, the gate signal should be cut off to prevent the shuttle oscillation, which will be explained with details in the next section.

B. Shuttle oscillation

Once the shuttle arrives at its destination (e.g. switching from G to D/S), all charges that were accumulated on the top surface of the shuttle (i.e. Q in Eq. (1)) will be discharged to that terminal (i.e. D/S). Immediately after that, the shuttle will accumulate charge again on the bottom surface but with opposite sign. Since V_G is already higher than the programming voltage (V_P), the shuttle will be lift-off and
starts to fly down. This situation is referred to as shuttle oscillation and is illustrated in Fig. 8 when the gate voltage is maintained longer than necessary.

C. Two-level gate control

To avoid the shuttle oscillation, the gate voltage must be shut down before the shuttle arrives at its destination. However, with process variations, some shuttles have shorter flight times than others, as shown in Fig 9. The gate voltage thus must be turned off before the time T_A. That will slow down all other shuttles that have longer flight time than T_A, especially the slowest ones. To mitigate this issue while ensuring no oscillation, a novel two-level gate control is proposed as shown in Fig. 10. V_{G1} (> V_P) is maintained until T_1 (T_1 is slightly shorter than T_A) and is switched to V_{G2} (< V_P) for a duration of T_2. In Fig. 10, the gate is kept at 6V for 300 ns and switched to 4.5 V for another 300 ns. Compared to the single-level gate pulse, this scheme ensures that no oscillation occurs and improves the write delay by 14%.

Fig. 11 analyses the write delay of a typical shuttle with different T_1 and T_2 combinations. In general, a longer T_1 is more desirable to have a short write delay.

more desirable to have a short write delay. The maximum V_{G1} is primarily limited by the coupling effect in the unselected NEM devices. A higher V_{G1} can be used by decreasing the coupling effect from G to D/S with larger access transistor sizes.

V. CONCLUSION

This work presents a NEMS-CMOS integration of the ANVM for data storage at high temperature (>200 °C). To address the issues in the previous control scheme, we propose novel control schemes for reliable NEM memory operation. The proposed write operation ensures that correct data is programmed into the memory cells while saving 32% of the write power. Comprehensive simulations have been performed to analyze the write delay, shuttle oscillation phenomenon and coupling effects between the G and the D/S terminals of the NEMS devices. Finally, the two-level gate control eliminates shuttle oscillation without performance degradation. The utilization of two levels improves the write performance by 14 % and variation tolerance.

REFERENCES