SCIENCE CHINA
Information Sciences

• **SPECIAL ISSUE** •

May 2011 Vol. 54 No. 5: 1073–1086
doi: 10.1007/s11432-011-4217-8

Resistance switching for RRAM applications

CHEN Frederick T.[1]*, LEE HengYuan[1], CHEN YuSheng[1,2], HSU YenYa[1], ZHANG LiJie[3], CHEN PangShiu[4], CHEN WeiSu[1], GU PeiYi[1], LIU WenHsing[1], WANG SuMin[1], TSAI ChenHan[1], SHEU ShyhShyuan[1], TSAI MingJinn[1] & HUANG Ru[3]

[1]*Nanoelectronic Technology Division, Electronics and Optoelectronics Research Laboratories, Industrial Technology Research Institute (ITRI), Chutung, Hsinchu, Taiwan, China;*
[2]*Institute of Electronics Engineering, National Tsing Hua University, Taiwan, China;*
[3]*Institute of Microelectronics, Peking University, Beijing* 100084*, China;*
[4]*Department of Chemical and Materials Engineering, Ming Shin University of Science and Technology, Taiwan, China*

Received September 4, 2010; accepted October 19, 2010

Abstract Resistive random access memory (RRAM or ReRAM) is a non-volatile memory (NVM) technology that consumes minimal energy while offering sub-nanosecond switching. In addition, the data stability against high temperature and cycling wear is very robust, allowing new NVM applications in a variety of markets (automotive, embedded, storage, RAM). Based on sudden conduction through oxide insulators, the characteristics of RRAM technology have still yet to be fully described. In this paper, we present our current understanding of this very promising technology.

Keywords random access memory, storage, oxide, breakdown, oxygen, vacancies

Citation Chen F T, Lee H Y, Chen Y S, et al. Resistance switching for RRAM applications. Sci China Inf Sci, 2011, 54: 1073–1086, doi: 10.1007/s11432-011-4217-8

1 Introduction

Resistive random access memory (RRAM) has recently become a promising and popular non-volatile memory (NVM) technology due to its relative simplicity of structure fabrication, as well as its electrical performance [1]. An RRAM structure merely requires a metal-insulator-metal (MIM) stack. In an RRAM array, the RRAM would be coupled to a transistor (as in Figure 1) or a diode for isolation. The insulator may undergo an insulator-metal transition similar to a Mott transition [2], due to a local stoichiometry change, induced by an electric field. The RRAM's MIM structure is far simpler than the deep trench stack needed for DRAM, the multiple magnetic layer stack structure for magneto-resistive random access memory (MRAM), the combination of different-sized features including sub-lithographic heaters for phase change memory (PCM), or even the oxide-nitride-oxide (ONO) dielectric in Flash memory gate stacks. In addition, the selection of materials for RRAM includes transition metal oxides, which are already familiar to workers in the field of high-k gate dielectric, DRAM and MRAM development. The electrical performance of RRAM devices has also generally been superior to other alternative technologies. Table 1 gives a brief comparison of RRAM with other memory technologies under development today.

*Corresponding author (email: fred_chen@itri.org.tw)

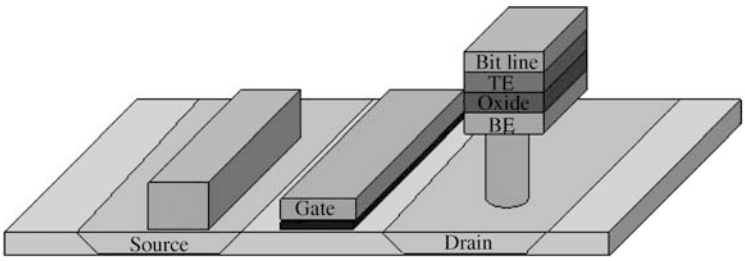

Figure 1 Basic structure of an RRAM. Here a MOSFET is used to access the RRAM, consisting of a bottom electrode (BE), oxide, and top electrode (TE) stacked together.

Table 1 Basic features of memory technologies considered today

	SRAM	DRAM	Flash	PCM	MRAM	FeRAM	RRAM
Mechanism	Voltage latching	Charge storage	Charge storage	Amorphous resistance	Tunneling magneto resistance	Electric dipole field	Localized conduction Interface modification
Special features	>1 transistor	Trench/stack capacitor	ONO Floating gate	Narrow heater	Ferromagnet	Ferroelectric	N/A
Retention	Volatile	Volatile; needs refresh	Non-volatile	Non-volatile	Non-volatile	Non-volatile	Non-volatile
Endurance	10^{17} cycles	10^{10} cycles*	10^5 cycles	10^6 cycles	10^{14} cycles	10^{16} cycles	10^{10} cycles
Capacity	MB	GB	GB	MB	MB	MB	MB
Speed	<ns	ns	μs	100 ns	ns	100 ns	<ns
Energy per bit	pJ	30 fJ/ms	10–100 pJ/page	10 pJ	pJ	<0.1 pJ	<0.1 pJ

*20 years of 64 ms refresh cycles.

While most new NVM technologies are able to surpass NAND and NOR Flash and DRAM in endurance, i.e., the number of operation cycles, some are sensitive to additional external parameters, e.g., PCM is sensitive to temperatures above 150°C due to crystallization, while MRAM is sensitive to the torque from external DC magnetic fields applied for sufficiently long time. Also, among these technologies, only RRAM has thus far been able to demonstrate sub-10 μA switching [3] and sub-nanosecond switching [4]. It should be mentioned that the RRAM category itself contains many different materials systems. Roughly, they may first be categorized according to whether defects in the material (typically oxygen vacancies) or the electrode itself provide the main ionic carriers involved in the switching [5]. The defect-based RRAMs are generally based on transition metal oxides and these may further be divided into those based on perovskite materials and those based on binary metal oxides [6]. The perovskite materials are known to be good oxygen ion conductors in the bulk [7], whereas the binary metal oxides have negligible conductivity in the bulk, and require conducting paths to be formed by an initial electrical stressing step, also known as "forming". After the RRAM is formed, it can be switched between resistance states by two operations: RESET (restoring a high-resistance state from a low-resistance state) and SET (transitioning to low-resistance state from high-resistance state). RRAMs may also be divided according to the RESET polarity with respect to SET. If RESET and SET currents are in the same direction (Figure 2(a)), the

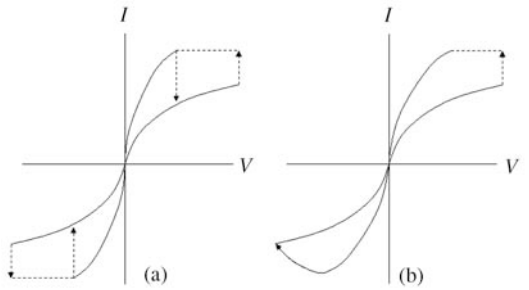

Figure 2 (a) Unipolar operation; (b) bipolar operation.

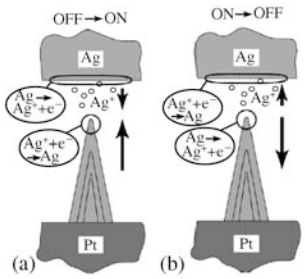

Figure 3 Operating principle of an electrode-based RRAM, or CBRAM. Reproduced from [5].

RRAM is called "unipolar," while if RESET and SET currents are in opposite directions (Figure 2(b)), it is called "bipolar."

The combination of RRAM's attractive features allows it to be targeted for many different applications. The most widely considered application is storage-type memory, typically as NAND flash replacement. The requirement is a minimal cell size of $2F \times 2F = 4F^2$, where F is both the feature size and half-pitch, and furthermore, more than one bit stored in that area. This can be achieved through multi-level cell (MLC) operation [1] or multi-layer stacking on the same substrate. The latter method requires a one-diode, one-resistance (1D1R) approach and has been demonstrated by Unity Semiconductor [8] as well as Samsung [9]. However, it is currently still too difficult to simultaneously achieve high ON/OFF ratio and sufficient current. Data that is frequently accessed instead of being stored is better handled with a smaller array and larger cell size, allowing sufficient current for operation at high speed. In this case, a standard one-transistor, one resistance (1T1R) operation is sufficient, with a cell size of up to 9–12F^2. In addition, other niche applications may allow the replacement of a dual-memory multi-chip package (MCP) with a single RRAM, e.g., combining the high speed of DRAM and the non-volatility of NOR Flash, but with superior endurance.

Much work remains to be done in the development of RRAM technology. First and foremost, the underlying mechanism(s) for the resistance switching must be characterized thoroughly. The mechanisms will differ for different materials as well as for unipolar vs. bipolar switching. Second, the continued demonstration of reliable array operation would further encourage adoption by industry. At this point, the largest array that has been demonstrated was a 64 Mb 4-layer stacked conductive metal oxide (CMOx) memory [8]. This is to be compared with Samsung's 512 Mb PCM demonstrated in 2006 [10], or the recently announced 128 Mb Omneo PCM product from Numonyx [11], as well as the 16 Mb MRAM from Everspin [12], and the 8 Mb FeRAM product already available from Ramtron [13]. It is an indication that RRAM like other alternative NVM technologies, still has some way to go before competing with NAND Flash or DRAM on the high Gb-scale. On the other hand, since the array size is really driven by the application requirement, these developments indicate that these new NVM technologies, especially RRAM, have the potential to create new markets based on their specific behaviors.

2 RRAM types

2.1 Electrode-based RRAM

The electrode-based RRAMs typically have a silver or copper electrode in contact with a material that readily dissolves and conducts ions of the electrode metal [5]. Applying a positive voltage at the source electrode drives ions into the dissolving material (Figure 3(a)). The ions are neutralized upon reaching the inert counter-electrode, forming deposits that are dendritic in nature extending back to the source electrode. Once fully extended the dendrites act as conducting bridges between the source and inert electrodes. Hence the system, also sometimes referred to as a conducting bridge random access memory (CBRAM), is in a low-resistance state. By applying the opposite polarity voltage, the metal bridges are dissolved, releasing ions that drift back to the source electrode (Figure 3(b)). This brings the system back into the high-resistance state. Sometimes, a voltage of the same polarity can disrupt the metal bridge as

well [14, 15].

Many materials have been used as the dissolving material, including chalcogenides (e.g., GeSe) [16], oxides (e.g., SiO_2, WO_x) [17, 18], sulfides [19], and even monatomic substances (e.g., amorphous silicon or carbon) [20, 21], as well as pure vacuum [22]. CBRAM systems as small as 20 nm have been demonstrated [19]. It is clear that to prevent unwanted diffusion into the interlayer dielectric, integration of diffusion barriers, e.g., TaN would be necessary. The thermal stability of the dissolved conducting bridges is also a key factor for data retention.

The switching time of CBRAM for the industrially relevant Cu-SiO_2 system has been studied recently by Bernard et al., indicating a linear dependence on the inverse current (1/I), consistent with charge-injection limited kinetics, as well as an exponential dependence on the inverse voltage (1/V), consistent with a nucleation or electro-crystallization mechanism or even soft breakdown [23]. Based on their study, to switch competitively fast against other memories, i.e., switching time $<$100 ns, the current would need to exceed $\sim$10 mA @ 1V. However, it is not feasible to provide such a large current within a high-density array. Thus, to improve the performance, alternative CBRAMs would need to improve the mobility of the ions injected by the source electrode. This appears to have been attained with the sulfide-based CBRAM of [19].

2.2 Defect-based RRAM

The majority of RRAM's are based on conduction through regions constituted by defects, most commonly oxygen vacancies [1, 24, 25]. In turn, this can be further divided according to the electrode-material interface type. Commonly, in RRAM systems with a Pt electrode in contact with a semiconducting oxide, e.g., a perovskite like PCMO or Nb-doped STO, there is a Schottky barrier, whose height and width are modulated by the concentration of oxygen vacancies, which in turn depends on the uniform drift of oxygen ions under an applied electric field [26]. Most metals, however, react to some degree with the oxides in contact, leading to ohmic interfaces rather than Schottky barriers. The oxygen mobility is also generally quite limited in most binary oxides, which are also more insulating, compared to perovskites. Hence, the forming of oxygen vacancies into conducting paths tends to be non-uniform. Specifically, oxygen ions drift toward the positive electrode (anode) under a sufficiently high electric field. The oxide region near the cathode is therefore more readily depleted of oxygen. Near the anode, oxygen vacancies may form as oxygen enters the anode, but some occasionally recombine with oxygen moving up from the cathode side. This is illustrated in Figure 4, where a narrow conducting path, known as a filament, appears as the occasionally arising connection of oxygen vacancies between electrodes.

The RESET phenomenon, where the filament is disrupted, brings the system back to a higher resistance state. This can be achieved either by a voltage pulse opposite in polarity to that which formed the filament (bipolar mode), or a pulse of the same polarity (unipolar mode). The bipolar mode is favored in asymmetric RRAM systems, particularly where one electrode gets oxygen from a transition metal oxide such as HfO_2 or ZrO_2[1]. Oxygen vacancies are naturally more plentiful near the cathode, and the return of oxygen to the oxide from the anode under the opposite polarity voltage gives ample opportunity for breaking the filament (Figure 5). On the other hand, in more symmetric systems, such as a transition metal oxide sandwiched between two Pt electrodes, polarity dependence is not expected. The RESET process may be driven by internal Joule heating, or electromigration, giving an opportunity for oxygen to enter the filament from the surroundings (Figure 6).

The filament may be restored after a RESET, by re-applying a SET voltage in the same polarity that was used to form the original filament. This voltage is generally lower than the forming voltage, since the electric field necessary for displacing the oxygen ions in re-creating the filament only has to be exerted over the filament gap, not the whole oxide thickness.

Most defect-based RRAMs have recently been focused on transition metal binary oxides. Due to the critical dependence of the switching characteristics on oxygen vacancy generation and movement, key factors affecting selecting RRAM materials are: 1) the conductivity of the oxide; 2) the band gap of the oxide; 3) the ability of a reactive electrode to remove oxygen from the oxide; and 4) the ability of oxygen to move within the oxide. Among the known oxides, HfO_2, ZrO_2, SiO_2, Ta_2O_5, and Al_2O_3 are best

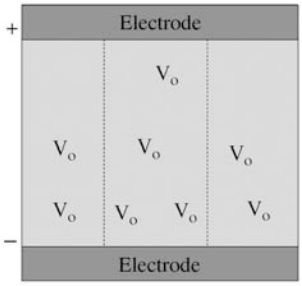

Figure 4 The formation of a filament as an occasional continuous path of oxygen vacancies (indicated by V_o) between the electrodes.

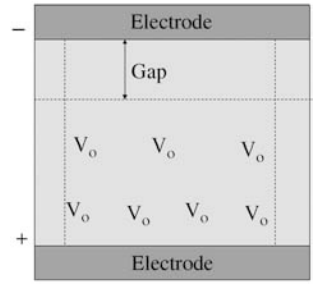

Figure 5 In a bipolar RESET, the oxygen drifts back from the anode to the oxide under the opposite polarity electric field, creating a gap in the filament near the anode.

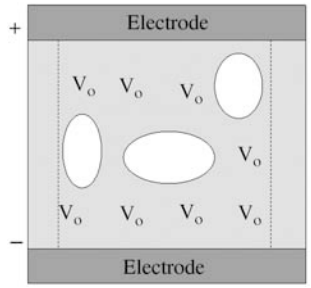

Figure 6 In a unipolar RESET, the oxygen may enter the filament anywhere from the surroundings, facilitated by heating or electromigration. Gaps may form in the filament without the same preference seen in bipolar mode.

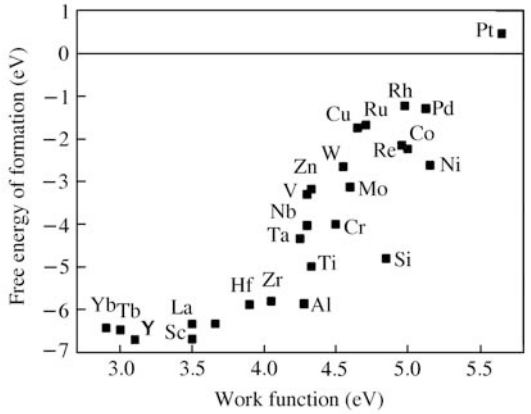

Figure 7 Relative free energy of formation of various metal oxides. Reproduced from [27].

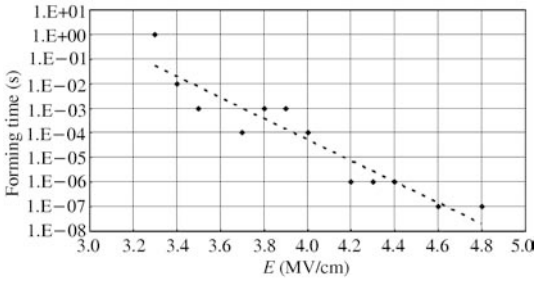

Figure 8 Breakdown, or forming time as a function of electric field for a Ti/HfO_2/TiN RRAM with 10 nm thickness.

known as insulators. However, the band gaps of SiO_2 and Al_2O_3 are relatively wider, leading to larger required forming voltages. The reactivity of an electrode metal with an oxide depends on the difference between the energy of formation of the oxide and that of the electrode oxide. Figure 7 [27] shows the relative energies of formation of oxide for different metals. Platinum is an ideal choice for a non-reactive electrode material, being a noble metal that does not oxidize. Most metal oxides have lower energies of formation, so oxygen is not likely to react with platinum. However, it is expensive and difficult to process in a standard semiconductor process. On the other hand, many oxidizing metals like Hf, Zr, Ti, Al have similar energies of oxidation, in the −5 to −6 eV range. Any of these metals can receive oxygen from and provide oxygen to an oxide of any other of these metals. Ti is possibly the most effective reactive electrode choice, as even after gettering the oxygen, its sub-oxides still allow sufficient conductivity [28], in sharp contrast to Ta or Al. Perovskites, TiO_2 [29], HfO_2 [30] and tetragonal-phase ZrO_2 [31] all have relatively larger oxygen mobility than SiO_2, Ta_2O_5 or Al_2O_3, as evidenced by the thin limiting native oxides on the latter metals, as well as their use as oxygen barriers [32, 33]. An exception may be made for oxides with substantial oxygen vacancy concentration [34]. ITRI recently [1, 35] demonstrated an ideal combination of electrode and oxide, a Ti reactive electrode in contact with HfO_2. The titanium facilitates generation of oxygen vacancies in the HfO_2, leading to formation of HfO_x ($x < 2$). In addition, the dilute presence of oxygen in the Ti does not increase the resistance as the electrode becomes a fairly conductive TiO_x ($x \ll 2$) layer. As a result, the resulting bipolar RRAM structure was able to simultaneously achieve high cycling endurance even with low power.

Due to the oxygen vacancy-based RRAMs being a wider class, and their generally better performance, we will now focus on describing the operation of this category of RRAM. As a final remark, we note that

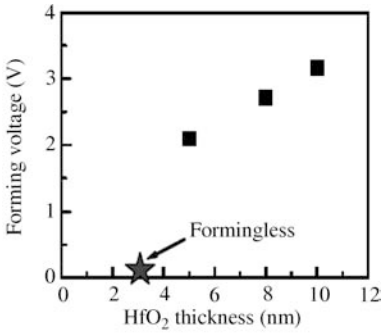

Figure 9 Forming voltage as a function of oxide thickness. Reproduced from [1].

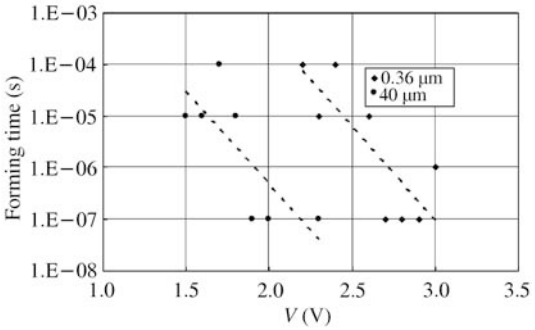

Figure 10 Forming voltage as a function of oxide width for a Ti/HfO_2/TiN RRAM with 5 nm HfO_2 thickness.

defect-based RRAMs need not only be based on oxygen, but can also in some cases be based on nitrogen [36] or even metal dopants [37].

3 Operation of oxygen vacancy-based RRAM's

There are four operations possible with an oxygen vacancy-based RRAM cell: 1) Initial forming of the cell; 2) RESET (increase resistance); 3) SET (decrease resistance); and 4) reading of the cell resistance. The initial forming step is unique to RRAM technology, since it is equivalent to a soft pseudo-breakdown of an insulator, and requires serious consideration for practical implementation. The RESET and SET mechanisms are a partial recovery of a thin layer of the broken-down oxide and re-breakdown of this layer, respectively [35]. Equivalently, the filament is disrupted by the RESET process and re-formed by the SET process. The similarity of forming and SET to breakdown in oxides suggests a possible description of the kinetics of the switching. The kinetics of forming, SET and RESET are important to quantitatively characterize the speed of the transition as well as the stability of the resistance state at a given temperature under a given voltage stress.

3.1 Forming

The forming step is essentially the same as a breakdown of the oxide, except that the current is limited by a series resistor, transistor or diode. Forming requires a sufficiently high electric field across the oxide thickness, applied for a sufficient time [38]. The time to breakdown typically exhibits an exponential dependence on the electric field (Figure 8).

By reducing the oxide thickness, a lower forming voltage would be needed to produce the same electric field. In fact, for sufficiently thin oxide, the forming voltage would go to zero (Figure 9) [1]. The effect of oxide layer width (or area) on forming voltage is to be understood in terms of larger area presenting greater opportunity for potential defect sites to act as forming sites. Figure 10 shows this as a slight upward trend of forming voltage as size decreases.

3.2 RESET

The RESET process is generally used after forming to demonstrate initial ability to switch. As portrayed in Figure 5, the flow of oxygen back from the anode reduces the vacancies in a layer adjacent to the anode, at least partially recovering the oxide in that layer. The voltage applied results in an electric field across the oxide thickness, but the current is expected to be concentrated in the filament(s). Three effects lead to rupture. The first is Joule heating due to the high current density in the filament. This is expected to be more significant for thicker films [39]. For films thinner than the electron inelastic mean free path, there is no opportunity for collision, and furthermore the electrodes are effective heat sinks. The second possible effect is electromigration, also arising from current density. According to Black's law the expected timescale for filament rupture would be inversely proportional to the square of the current

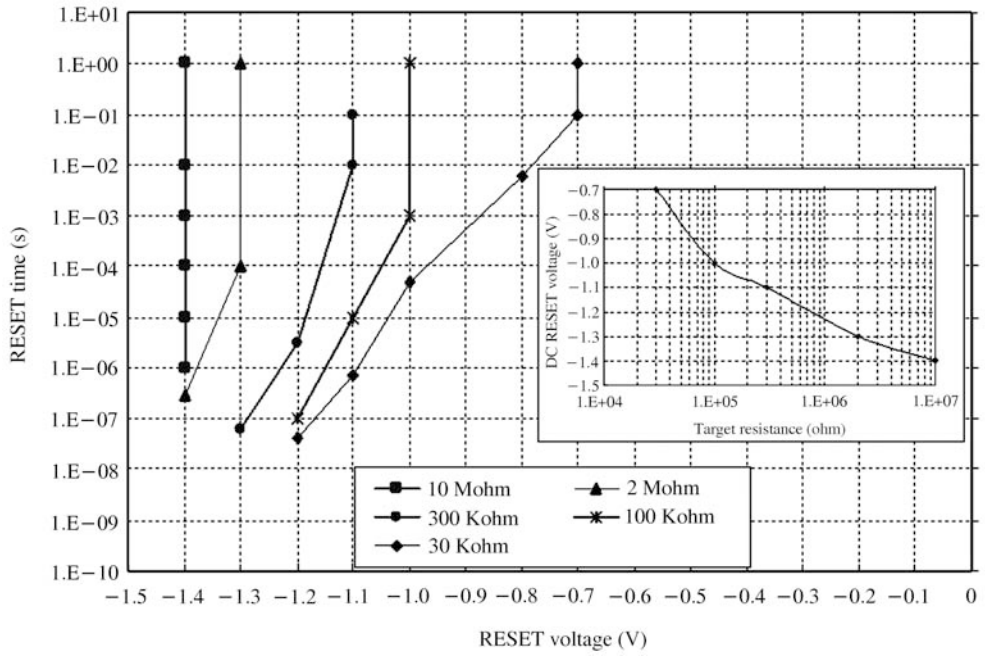

Figure 11 RESET time and voltage to increase resistance from 5 K-ohm to various target values for a Ti/HfO_2/TiN RRAM with 10 nm HfO_2 thickness. Saturation occurs for sufficiently long times. Inset: DC RESET voltage for different target resistances. The data was taken at room temperature.

density [40]. Finally, the RESET could be driven by the reversible flux of O=ions between the oxide and the electrode [34]. Especially in transition metal oxide systems, at a critical concentration of oxygen vacancies, a metal-insulator transition is possible, leading to an abrupt switching between the conducting and insulating states.

The kinetics model for RESET can be very different from that for forming, since RESET is not similar to the breakdown phenomenon at all. For example, different voltage levels can cause resistance to increase at different rates [41]. This analog type of behavior is consistent with increasing oxygen mobility under increasing electric field [42]. If there is negligible oxygen motion below a certain threshold electric field, we would expect the increase in resistance to saturate at some point. This is seen in Figure 11. From the bunching of curves it is also apparent that for higher V_{RESET}, R_{RESET} is more sensitive to V_{RESET}. Unipolar RRAMs must be RESET at voltages less than corresponding SET voltages. Otherwise the overlap between SET and RESET voltage operation windows would make the device unstable. The RESET timescale is often longer, e.g., microseconds, than that for SET, e.g., nanoseconds [43]. Ti doping into NiO was effective in reducing unipolar RESET time down to less than 5 ns [44]. However, even in this case, there is still a dependence of the increase in resistance on the pulse width (Figure 12). The analog nature of RESET vs. the abrupt nature of SET is an interesting aspect of defect-based RRAMs. As will be shown later, this feature of RESET can be used to eliminate errors during array programming.

3.3 SET

The SET process is very similar to the forming process, except that the breakdown only occurs in the recovered oxide region, or the gap in the filament after it has been ruptured. As a result, the SET voltage is always lower than the forming voltage. Commonly, in bipolar RRAMs, the SET voltage is comparable to the RESET voltage in magnitude. Like forming, the SET kinetics is also expected to be exponentially dependent on the electric field. However, unlike forming, the initial RESET condition may correspond to significantly varying gap distances, and different effective dielectric strengths of the gap. Even so, we observed as in Figure 13 that the SET time is still an exponential function of voltage. However, one may imagine that for different initial RESET conditions, the fitted line may shift left or right, reflecting the natural variation in V_{SET}. This would result in a very wide range of times for the high resistance to

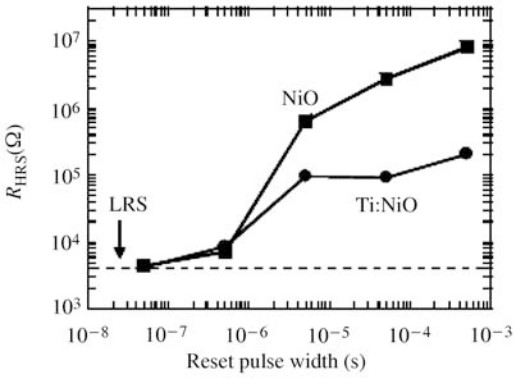

Figure 12 Unipolar RESET rate for NiO-based RRAM. The applied RESET voltage is 1 V, and the plotted HRS resistance is the median of 50 measured values taken during cycling. Reproduced from [43].

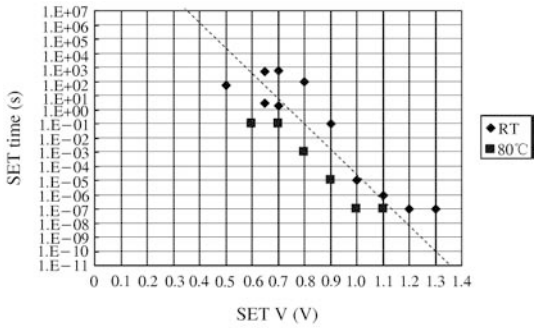

Figure 13 SET time as a function of SET voltage for a Ti/HfO_2/TiN RRAM with 10 nm HfO_2 thickness. RT=room temperature.

abruptly fall to low resistance. In fact, this is indeed observed in the low shape parameter of the Weibull plots of time-to-disturb for the high resistance state [33]. From Figure 13, it may also be predicted that the switching time in the SET process would be much less than one nanosecond. In fact, measurements to date [4] indicate that the switching is faster than 0.3 ns, for both SET and RESET, consistent with both Figures 11 and 13.

A useful characteristic of the SET process is the determination of the SET resistance by the current limit imposed by a series resistor, transistor, or diode [1]. A larger current limit leads to a wider current-carrying area which leads to a lower resistance. The lowest possible current limit that may be used is the highest possible current for the high resistance state. Currents as low as 1 μA have in fact been reported [3]. A general caveat for the use of lower SET currents is that read currents would have to be even lower, which would in turn require longer read times.

3.4 Reading

When reading the resistance of an RRAM cell, care must be taken not to disturb the state which is being read. Generally this is possible if the voltage applied to the RRAM is low enough, not to SET or RESET for a given short enough pulse duration. The reader is again referred to the kinetics described in Figures 11 and 13. The pulse duration can be less than 10 ns [45]. However, the exact voltage upper limit depends not only on the pulse duration but also on the prior RESET or SET condition. Natural variations cause this voltage limit to vary. Conversely for a fixed voltage input, the time at which the resistance changes state can also vary naturally by many orders of magnitude. Consequently, it is a recommended practice to fine-tune the resistance after RESET or SET programming, a practice called verification (to be discussed further below) [35]. Additionally, the temperature at which the RRAM is read should not be too high. Not only does high temperature increase the likelihood of disturbing the state, but it also decreases the resistance of the RESET state, due to trap state activation energy [46, 47].

3.5 Verification

After a RESET or SET operation, the desired resistance may not be attained, due to the natural variation involved with oxygen ion motion. Consequently, verification will be necessary.

As practiced by ITRI, for RESET, the verification consists of assessing the resistance level, then increasing the RESET voltage as needed, and then checking the resistance again. This is required to guarantee the resistance distribution above a lower limit. It is convenient to plot the cumulative distribution function F of the resistance to assess the tightness of the distribution. If the distribution is of the Weibull type, the log-log plot of $-\ln(1-F)$ vs. the resistance would be a line whose slope is the shape parameter. Figure 14 shows how this technique allows the gap between the low and high resistance states in an array to be widened to acceptable levels. SET verification requires more care. Depending on

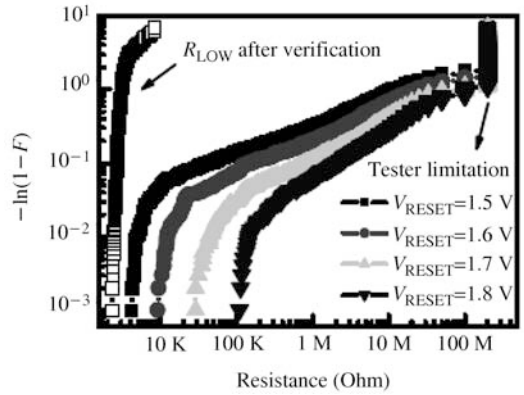

Figure 14 RESET verification to tighten HRS distribution in a 1 kb array. Reproduced from [34].

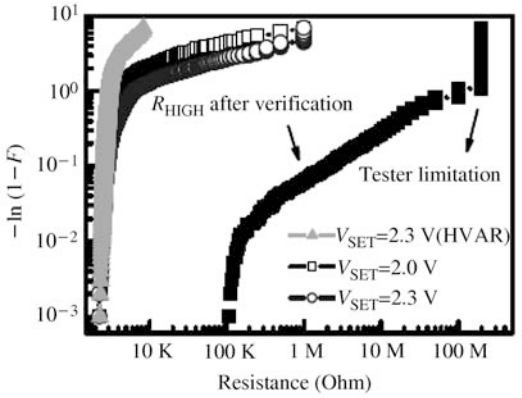

Figure 15 SET verification to tighten LRS distribution in a 1 kb array by higher voltage after RESET (HVAR). Reproduced from [34].

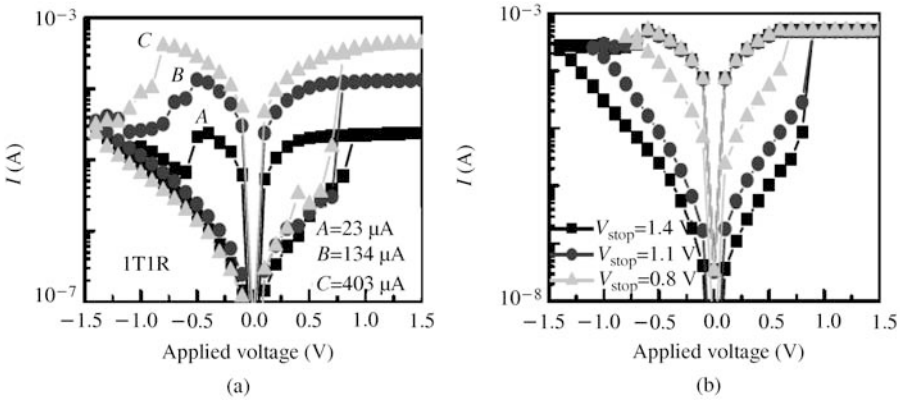

Figure 16 Multi-level RRAM operation by (a) different SET limiting current, or (b) different maximum RESET voltages (V_{stop}). Reproduced from [1].

the initial RESET condition, the required SET voltage may fluctuate. After a first SET attempt, the system resistance abruptly drops to a lower state, but the resistance may not be low enough, i.e., the filament may have re-formed only partially. However, subsequently increasing SET voltage may lead to accidental unipolar RESET [35]. Consequently, the SET attempt can only be safely repeated with a higher voltage after the RRAM is RESET again. Figure 15 shows the result of applying higher voltage after RESET (HVAR), compared to simply ramping the SET voltage.

3.6 Multi-level cell (MLC) operation

The RESET and SET operations described above naturally allow for the possibility of multi-level operation, i.e., the use of more than two states in one cell. Four states correspond to storing two bits, while eight states correspond to storing three bits in one cell. By either modulating the SET current or the RESET voltage, multilevel operation was demonstrated by ITRI in 2008, as shown in Figure 16 [1]. There is a natural tradeoff between the number of bits stored per cell and the required tightness of the resistance distribution, as well as between the number of bits per cell and the maximum power consumption budget of the cell.

4 Selection device considerations

The preceding discussions have been focusing on the nature and operation of the RRAM element itself. However, equally or possibly even more important is the practical integration of the RRAM in a large array, with the use of selection devices to isolate the RRAM cells from one another. The established method is to use a MOSFET in series with the RRAM element, in a so-called one-transistor one-resistor

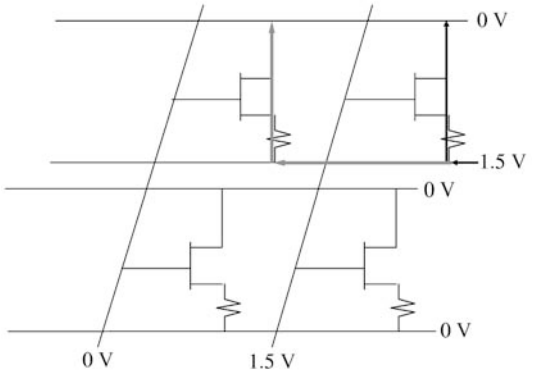

Figure 17 1T1R architecture based on MOSFET. The combined leakage through all the unselected 1T1Rs (black) on a bit line needs to be negligible compared to the read current through the selected 1T1R (gray).

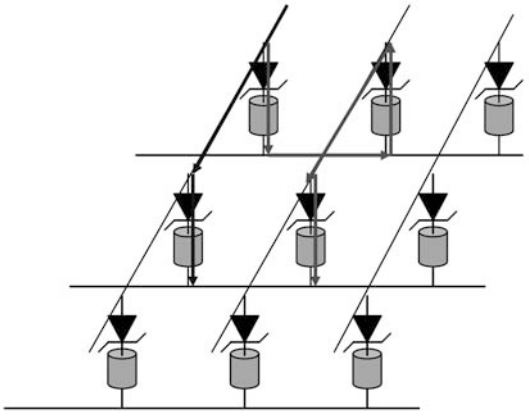

Figure 18 1D1R architecture where the unselected lines are left disconnected to outside world. There are $(M-1)(N-1)$ possible leakage paths consisting of three 1D1R's (middle one is reverse biased) in series.

(1T1R) arrangement. Alternatively, the MOSFET may be replaced by a BJT [48]. Since a transistor requires three terminal contacts for access, the cell size has a lower limit of 6–8 F^2 corresponding to the lateral placement of the three terminals side by side, plus the required isolation. A multi-emitter BJT with shared base and collector offers a local cell size of $4F^2$[49]. Also, by using a two-terminal selection device, i.e., a diode, and arranging these two terminals to be aligned vertically in a 1D1R architecture, the cell size lower limit may be theoretically reduced to $4F^2$. The selection device needs to provide sufficient isolation in a given array size. In other words, the selector's ON/OFF ratio should be large enough, in order that specifically: 1) the OFF resistance allows only negligible leakage current through the RRAM array; and 2) the available ON current should be enough to operate the RRAM device. Other characteristics which could make the selector more attractive are the ability to be stacked in multiple layers, and the ability to conduct current in opposing directions.

Consider a row of 1T1R cells, consisting of an n-MOSFET and an RRAM, with each RRAM connected to the same bit line, and the source of each MOSFET being grounded (Figure 17). To read one cell in this row, a high enough gate voltage is applied to the selected cell's MOSFET to allow current to flow freely through the selected 1T1R. The other cells in the row, being unselected, only pass through a sub-threshold leakage current whose magnitude depends on the bit line voltage, which is effectively the drain voltage. The leakage current is much smaller than the current passing through the selected cell. However, if there are enough unselected cells in parallel with the selected cell, the combined leakage currents from all those cells will outweigh the read current of the selected cell. Most MOSFETs have specified ON/OFF ratios of 10^4–10^5. This suggests that a bit line row cannot be longer than $\sim 10^4$ MOSFET-based 1T1R cells. A similar conclusion would apply to BJT-based 1T1R cells.

1D1R isolation imposes different ON/OFF ratio consideration depending on how the unselected lines are connected. Figure 18 shows the case where the unselected lines are disconnected. For any given selected 1D1R cell, there are $(M-1)(N-1)$ combinations of three unselected 1D1R cells connected in series, where M is the number of rows and N the number of columns in the $M \times N$ array. Furthermore the middle of the three 1D1Rs in series is reversely biased relative to the other two. For a standard silicon diode, in reverse bias, the current drops to a leakage level. This small current, however, has to be multiplied by $(M-1)(N-1)$ to give the total array leakage current.

Consequently, for a Gb array, the unselected cell leakage current should be significantly less than $1/10^9$ of the actual current to be sensed when reading a cell in its highest resistance state. This is an extremely challenging requirement, currently impossible even by the highest industrial standards. Figure 19 shows an alternative case, where the unselected lines are grounded instead of being disconnected. In this case, the middle reverse-biased 1D1R is shorted, leaving only a pair of 1D1R's in series as the representative unselected leakage path. So there are $(M-1)$ paths to ground and $(N-1)$ paths from ground. Assuming $M \gg 1, N \gg 1$, the requirement for an $M \times N$ array would be for the unselected leakage to be much less

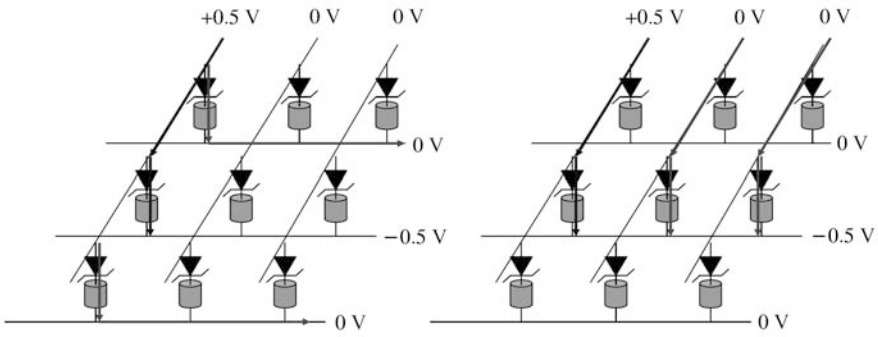

Figure 19 1D1R architecture where the unselected lines are grounded. There are $M-1$ leakage paths to ground and $N-1$ leakage paths back from ground.

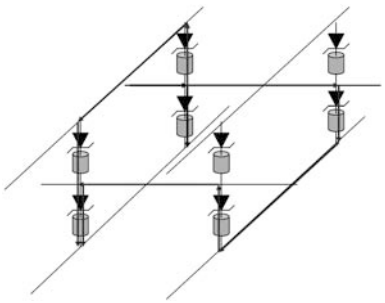

Figure 20 A 3D 1D1R architecture, visualized with two tiers. It is apparent that there are more leakage paths possible with more than one tier.

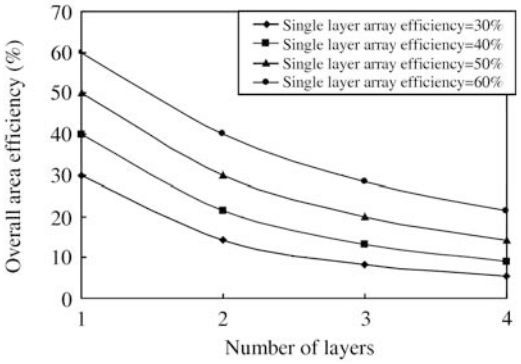

Figure 21 The array efficiency, or the area fraction occupied by a stacked array on a chip, decreases as the number of stacked layers increases.

than $1/M$ or $1/N$ of the actual sensing current for the highest resistance state, whichever is smaller. If our Gb array is arranged as 64 kb×16 kb, for example, the leakage current/read current ratio should always be much less than 1/64000. Furthermore it should be noted that the voltage would be divided between the two unselected 1D1Rs in series. This complicates the analysis, as we must now consider the leakage from a partially ON forward biased diode. Although the required ON/OFF ratio looks much easier to meet, it is still extremely challenging, since the conventional diode reverse bias cannot be applied. Often 1D1R architectures are extended further by implementation of multilayer stacking to form a 3D array, as shown in Figure 20. For a multi-tier 1D1R architecture, the number of possible leakage paths is now much larger. Thus, the ON/OFF ratio requirement will be even more stringent.

As a side comment, we note it is often remarked that a 1D1R architecture requires a unipolar RRAM, since a diode cannot conduct in reverse bias without sacrificing the ON/OFF ratio. However, a special allowance can be made for $p-n$ junctions due to the fact that minority carriers are involved. By directly reversing the bias after an initial forward bias pulse, a very brief transient with high reverse current is possible [50]. Depending on the thickness of the depletion region, the transient timescale can be up to nanoseconds, which is compatible with the high-speed RESET performance of the fastest bipolar RRAMs. Schottky-type diodes, on the other hand, involve only majority carriers, and so would not exhibit as large a reverse transient current, but only a very small displacement current, which can be neglected [51]. A tunneling oxide barrier can also perform the same function as a bidirectional diode. However, the maximum current density is minimal [42].

Ultimately, the viability of the 1D1R architecture will depend on its potential for 3D scaling, in the most cost-effective manner. The cost-effectiveness is determined not just by the array size but also by the area occupied by the controlling circuitry in the periphery. Although stacking additional 1D1R layers increases bit capacity proportionally, the area for controlling the array in the "master layer" must also increase proportionally. Figure 21 shows that this trend leads to increasing chip size, as the array area

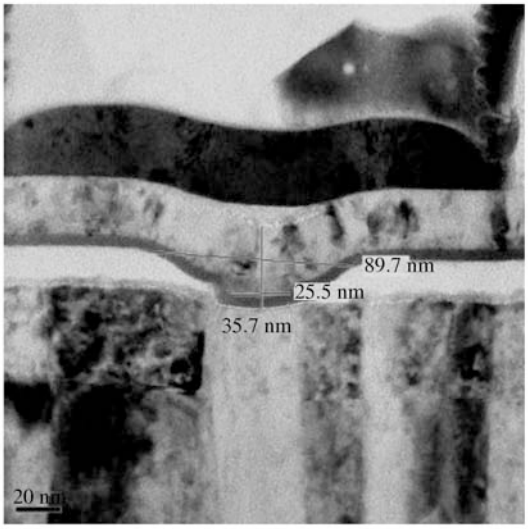

Figure 22 Cross-section of a concave-style RRAM from [34]. The active switching area is 25.5 nm, indicated by the width of the top titanium electrode contact with the HfO_2 layer.

stays the same, with bit capacity increasing vertically, but the periphery area steadily increasing with number of layers. Thus the cost of increasing bit capacity with 3D 1D1R-based architecture requires significant cost reduction to occur for every stacked layer as well as the periphery compared to conventional Flash memory in order to be justified as a cost-effective approach.

5 Outlook

The outlook for RRAM is hopeful as the potential for high speed and low-power switching allows for a wide variety of applications to be supported. Furthermore, scaling of RRAM to less than 30 nm (Figure 22) is feasible [35], and one may even expect that filaments as small as 2 nm are possible [52]. Ironically, while being the simplest memory to fabricate and operate, RRAM may require the most care to implement in products. The challenges that lie ahead for RRAM are:

1) a more thorough understanding of underlying processes for each given type of RRAM;

2) reduction of the variability of the resistance states, i.e., resistance distribution, V_{SET} distribution, V_{RESET} distribution;

3) increasing the number of RRAM switching cycles;

4) increasing the number of bits that can be stored in an RRAM cell;

5) elimination of the voltage and time burden in the forming step required for RRAM; and

6) 3D architectures for RRAM which are cost-effective from the whole system point of view.

Although it is difficult to compete with mainstream memory (i.e., Flash, DRAM, disk drives) on a cost basis, many users would prefer using a memory technology that allows more write cycles than Flash, consumes less power than DRAM, and operates at speeds comparable to state-of-the-art microprocessors. RRAM technology is rapidly becoming the choice that fits all these requirements.

Acknowledgements

This work was supported by the Ministry of Economic Affairs of Taiwan.

References

1 Lee H Y, Chen P S, Wu T Y, et al. Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2-based RRAM. In: IEDM 2008, 2008. 297–300

2 Mott N F. Metal-insulator transition. Rev Mod Phys, 1968, 40: 677–683
3 Wu Y, Lee B, Wong H P S. Ultra-low power Al_2O_3-based RRAM with 1 μA RESET current. In: VLSI-TSA 2010. 2010. 136–137
4 Lee H Y, Chen Y S, Chen P S, et al. Evidence and solutions of Over-RESET problem for HfO_x-based resistive memory with sub-ns switching speed and high endurance. In: IEDM 2010. 460–463
5 Waser R. Resistive non-volatile memory devices. Microelectr Eng, 2009, 86: 1925–1928
6 Sata A. Resistive switching in transition metal oxides. Mater Today, 2008, 11: 28–36
7 Taskin A A, Lavrov A N, Ando Y. Achieving fast oxygen diffusion in perovskites by cation ordering. Appl Phy Lett, 2005, 86: 091910
8 Chevalier C J, Siau C H, Lim S F, et al. A 0.13 μm 64 Mb multi-layered conductive metal oxide memory. In: ISSCC 2010. 2010. 260–261
9 Baek I G, Kim D C, Lee M J, et al. Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application. In: IEDM 2005. 2005. 750–753
10 Oh J H, Park J H, Lim Y S, et al. Full integration of highly manufacturable 512 Mb PRAM based on 90 nm technology. In: IEDM 2006. 2006
11 http://www.numonyx.com/EN-US/MEMORYPRODUCTS/PCM/Pages/P8P.aspx
12 http://www.everspin.com/products.php?hjk=16&a1f3=16Mb
13 Eliason J, Madan S, McAdams H, et al. An 8 Mb 1T1C ferroelectric memory with zero cancellation and mi-cro-granularity redundancy. In: CICC 2005. 2005. 427–430
14 Schindler C, Thermadam S C P, Waser R, et al. Bipolar and unipolar resistive switching in Cu-doped SiO_2. IEEE Trans Electr Dev, 2007, 54: 2762–2768
15 Guan W, Liu M, Long S, et al. On the resistive switching mechanisms of Cu/ZrO_2:Cu/Pt. Appl Phys Lett, 2008, 93: 223506
16 Dietrich S, Angerbauer M, Ivanov M, et al. A nonvolatile 2-Mbit CBRAM memory core featuring advanced read and program control. IEEE J Solid-State Circ, 2007, 42: 839–845
17 Kozicki M N, Gopalan C, Balakrishnan M, et al. A low-power nonvolatile switching element based on cop-per-tungsten oxide solid electrolyte. IEEE Trans Nanotech, 2006, 5: 1–10
18 Thermadam S P, Bhagat S K, Alford T L, et al. Influence of Cu diffusion conditions on the switching of Cu-SiO_2-based resistive memory devices. Thin Solid Films, 2009, 518: 3293–3298
19 Ono K, Kurotshuchi K, Fujisaki Y, et al. Resistive switching ion-plug memory for 32-nm technology node and beyond. In: SSDM 2008. 2008. 1164–1165
20 Jo H, Lu W. CMOS compatible nanoscale nonvolatile resistance switching memory. Nano Lett, 2008, 8: 392–397
21 Zhuge F, Dai W, He C L, et al. Nonvolatile resistive switching memory based on amorphous carbon. Appl Phys Lett, 2010, 96: 163505
22 Kayashima S, Takahashi K, Motoyama M, et al. Control of tunnel resistance of nanogaps by field-emission-induced electromigration. Jap J Appl Phys, 2007, 46: L907–909
23 Bernard Y, Gonon P, Jousseaume V. Resistance switching of Cu/SiO_2 memory cells studied under voltage and current-driven modes. Appl Phys Lett, 2010, 96: 193502
24 Kim K M, Choi B J, Hwang C S. Localized switching mechanism in resistive switching of atomic-layer-deposited TiO_2 thin films. Appl Phys Lett, 2007, 90: 242906
25 Xu N, Gao B, Liu L F, et al. A unified physical model of switching behavior in oxide-based RRAM. In: 2008 Symposium on VLSI Technology, 2008. 100–101
26 Sim H, Seong D J, Chang M, et al. Excellent resistance switching characteristics of Pt/single-crystal Nb-doped $SrTiO_3$ Schottky junction. In: Non-Volatile Semiconductor Memory Workshop, 2006. 88–89
27 Sharia O, Tse K, Robertson J, et al. Extended Frenkel pairs and band alignment at metal-oxide interfaces. Phys Rev B, 2009, 79: 125305
28 Liang C, Terabe K, Hasegawa T, et al. Resistance switching in anodic oxidized amorphous TiO_2 films. Appl Phys Exp, 2008, 1: 064002
29 Hass G, Bradford A P. Optical properties and oxidation of evaporated titanium films. J Opt Soc Am A, 1957, 47: 125–129
30 Ferrari S, Scarel G. Oxygen diffusion in atomic layer deposited ZrO_2 and HfO_2 thin films on Si (100). J Appl Phys, 2004, 96: 144–149
31 Busch B W, Schulte W H, Garfunkel E, et al. Oxygen exchange and transport in thin zirconia films on Si (100). Phys Rev B, 2000, 62: 290–293
32 Kim I, Ahn S D, Cho B W, et al. Microstructure and electrical properties of tantalum oxide thin film prepared by electron-cyclotron resonance plasma-enhanced chemical vapor deposition. Jap J Appl Phys, 1994, 33: 6691–6698
33 Gulino D A. Oxygen barrier coatings for enhanced stability of polyimide composites at elevated temperatures. In: NASA 1994, 1994

34 Wei Z, Kanzawa Y, Arita K, et al. Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism. In: IEDM 2008. 2008. 293–296
35 Chen Y S, Lee H Y, Chen P S, et al. Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity. In: IEDM 2009. 2009. 105–108
36 Kim H D, An H M, Kim K C. Large resistive-switching phenomena observed in $Ag/Si_3N_4/Al$ memory cells. Semicon Sci Tech, 2010, 25: 065002
37 Pandian R, Kooi B J, Palasantzas G, et al. Polarity-dependent reversible resistance switching in Ge-Sb-Te phase-change thin films. Appl Phys Lett, 2007, 91: 152103
38 Cheung K P. A physics-based, unified gate-oxide breakdown model. In: IEDM 1999. 1999. 719–722
39 Chang S H, Chae S C, Lee S B, et al. Effects of heat dissipation on unipolar resistance switching in Pt/NiO/Pt capacitors. Appl Phys Lett, 2008, 92: 183507
40 Black J R. Electromigration—a brief survey and some recent results. In: Proc. IEEE Intl. Reliability Phys. Symposium, 1968. 338–347
41 Lee H Y, Chen Y S, Chen P S, et al. Comprehensive study of read disturb immunity and optimal read scheme for high speed HfOx-based RRAM with Ti layer. In: VLSI TSA 2010. 2010. 132–133
42 Meyer R, Schloss L, Brewer J, et al. Oxide dual-layer memory element for scalable non-volatile cross-point memory technology. In: NVMTS 2008
43 Tseng Y H, Huang C E, Kuo C H, et al. High density and ultra small cell size of contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits. In: IEDM 2009. 2009. 109–112
44 Tsunoda K, Kinoshita K, Noshiro H, et al. Low power and high speed switching of Ti-doped NiO ReRAM under the unipolar voltage source of less than 3 V. In: IEDM 2007. 2007. 767–770
45 Sheu S S, Chiang P C, Lin W P, et al. A 5 ns fast write multi-level non-volatile 1 K bits RRAM memory with advance write Scheme. In: 2009 Symposium on VLSI Circuits. 2009. 82–83
46 Kim M G, Kim S M, Choi E J. Study of transport and dielectric of resistive memory states in NiO thin film. Jpn J Appl Phys, 2005, 44: L1301–L1303
47 McWhan D B, Menth A, Remeika J P. Metal insulator transitions in transition metal oxides. J Phys Colloq, 1971, 32: C1-1079-C1-1085
48 Wang C H, Tsai Y H, Lin K C, et al. 3-Dimensional 4F2 ReRAM cell with CMOS compatible logic process. In: IEDM 2010, in press
49 Servalli G. A 45 nm generation phase change memory technology. In: IEDM 2009. 2009. 113–116
50 Taur Y, Ning T H. Fundamentals of Modern VLSI Devices. Cambridge: Cambridge University Press, 2009. 64–70
51 Singh R, Richmond J. SiC power Schottky diodes in power factor correction circuits. In: Cree Application Note, 2002
52 Krzysztof S, Speier W, Bihlmayer G, et al. Switching the electrical resistance of individual dislocations in single-crystalline $SrTiO_3$. Nat Mater, 2006, 5: 312–320